

RESEARCH

Open Access



Development and validation of an interpretable machine learning model for predicting the risk of hepatocellular carcinoma in patients with chronic hepatitis B: a case-control study

Linghong Wu¹, Zengjing Liu², Hongyuan Huang¹, Dongmei Pan², Cuiping Fu³, Yao Lu³, Min Zhou⁴, Kaiyong Huang¹, TianRen Huang^{5*} and Li Yang^{1*}

Abstract

Background The aim of this study was to develop and internally validate an interpretable machine learning (ML) model for predicting the risk of hepatocellular carcinoma (HCC) in patients with chronic hepatitis B (CHB) infection.

Methods We retrospectively collected clinical data from patients with HCC and CHB treated at the Fourth Affiliated Hospital of Guangxi Medical University from January 2022 to December 2022, including demographics, comorbidities, and laboratory parameters. The datasets were randomly divided into a training set (361 cases) and a validation set (155 cases) in a 7:3 ratio. Variables were screened using Least Absolute Shrinkage and Selection Operator (LASSO) and multifactor logistic regression. The prediction model of HCC risk in CHB patients was constructed based on five machine learning models, including logistic regression (LR), K-nearest neighbour (KNN), support vector machine (SVM), random forest (RF) and artificial neural network (ANN). Receiver operating characteristic (ROC) curve, calibration curve and decision curve analysis (DCA) were used to evaluate the predictive performance of the model in terms of identification, calibration and clinical application. The SHapley Additive exPlanation (SHAP) method was used to rank the importance of the features and explain the final model.

Results Among the five ML models constructed, the RF model has the best performance, and the RF model predicts the risk of HCC in patients with CHB in the training set [AUC: 0.996, 95% confidence interval (CI) (0.991–0.999)] and internal validation set [AUC: 0.993, 95% CI (0.986–1.000)]. It has high AUC, specificity, sensitivity, F1 score and low Brier score. Calibration showed good agreement between observed and predicted risks. The model yielded higher positive net benefits in DCA than when all participants were considered to be at high or low risk, indicating good clinical utility. In addition, the SHAP plot of the RF showed that age, basophil/lymphocyte ratio (BLR), D-Dimer, aspartate

*Correspondence:

TianRen Huang
tianrenhuang@sina.com
Li Yang
yangli8290@hotmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **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/>.

aminotransferase/alanine aminotransferase (AST/ALT), γ -glutamyltransferase (GGT) and alpha-fetoprotein (AFP) can help identify patients with CHB who are at high or low risk of developing HCC.

Conclusion ML models can be used as a tool to predict the risk of HCC in patients with CHB. The RF model has the best predictive performance and helps clinicians to identify high-risk patients and intervene early to reduce or delay the occurrence of HCC. However, the model needs to be further improved through large sample studies.

Keywords Chronic hepatitis B, Hepatocellular carcinoma, Machine learning, Predictive model

Introduction

Liver cancer is one of the top 10 malignancies listed by the World Health Organization as a serious threat to human health and safety. According to statistics (Globocan 2022), approximately 865,273 new cases of liver cancer are reported worldwide annually, ranking it sixth amongst malignant tumours. Approximately 757,906 people are estimated to die from liver cancer, making it the third most common malignant tumour [1]. China has a high incidence of liver cancer, accounting for about 50% of the world's liver cancer cases. China has an estimated 367,657 new cases and 316,544 deaths from liver cancer each year, making it the fourth and second largest country in terms of morbidity and mortality from malignant tumours, respectively [2]. Hepatocellular carcinoma (HCC) is the most common primary liver cancer, accounting for 75–85% of all cases [3]. Hepatitis B virus (HBV) infection is a major risk factor for HCC in China. About 69.9% of Chinese HCC patients reportedly have a background of HBV infection [4]. China is the country with the largest HBV-infected population, accounting for about one-third of the world's infected population [5]. In China, approximately 90 million people are infected with HBV, and an estimated 300,000 people die from HBV-related diseases each year [6]. Additionally, HBV infection increases the risk of cirrhosis and HCC, and according to a cohort study, patients with HBV infection have a 16.1-fold increased risk of HCC [7]. HCC is also highly latent in the early stage, and only about 36% of Chinese patients are preliminarily diagnosed and eligible for treatment in the early stage. The remaining 9% and 55% are in the intermediate and late stages, respectively [8]. Partial hepatectomy is one of the main treatments of HCC. However, the rate of local recurrence can be as high as 70% at 5 years after primary hepatectomy [9]. Therefore, early and accurate prediction of the risk of HCC in HBV-infected individuals is urgently needed.

Risk-prediction models aid in the early detection, diagnosis, and treatment of diseases in high-risk populations and have been applied to cardiovascular diseases and tumours [10, 11]. Several models can reportedly predict the risk of HCC in patients with chronic hepatitis, including GAG-HCC, CU-HCC, REACH-B, REACH-B II, LSM-HCC, PAGE-B, REACH-B, etc [7, 12]. However, the above models primarily focus on the population from

Korea, Hong Kong, and Taiwan, or the Caucasian population. Studies on inland samples of the Chinese population are lacking. The applicability of these models to Chinese populations is uncertain because Chinese populations have different living environments, lifestyles, socioeconomic status, dietary cultures, and genetic characteristics from those in other parts of the world. Models such as the LSM-HCC risk model [13] have predictors that are difficult to detect in the clinical practice or contain too many items. Consequently, the calculation becomes more complicated and unsuitable for clinical practice. Finally, the aforementioned risk-prediction models are constructed mostly using traditional statistical methods. Given that the interaction mechanism between the risk factors of liver cancer is non-linear, the traditional linear regression model and logistic regression models are inadequate to solve the collinearity problem [14], and artificial intelligence can improve this problem to a certain extent. Currently, artificial intelligence including machine learning (ML) and deep learning is being used in medical research and practice. HCC risk prediction also needs to make full use of artificial intelligence for diagnosis, prognosis, and treatment to improve prediction accuracy. Therefore, using artificial intelligence to establish an HBV-related liver cancer risk-prediction model based on Chinese population with high accuracy and convenient clinical application is urgent.

ML is one form of artificial intelligence. It involves algorithmic methods that allow machines to learn how to solve problems without having to write specific programmes [15]. Over the past decade, ML applications in medicine have exploded, particularly in oncology [16]. As a complex, diverse, and prevalent group of diseases, cancers present challenging diagnostic problems and abundant data across multiple modalities, making clinical oncology a strong area for ML. Wang et al. [17] used ML algorithms to build a prostate cancer risk-prediction model based on common clinical indicators to provide evidence for diagnosing prostate cancer. Hou et al. [18] built a distant-metastasis prediction model for papillary thyroid cancer based on nine ML models, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN). By comparing the performance of the models, they found that the RF model had the best predictive ability and selected

the best predictive model. Recently, Korean researchers used a ML algorithm to build a PLAN-B model including 10 factors of cirrhosis, age, platelet count, type of NAs, gender, alanine aminotransferase(ALT), HBV DNA level, albumin, bilirubin, and HBeAg status in 6,051 patients with chronic hepatitis B(CHB) treated with entecavir (ETV) or tenofovir fumarate (TDF). It was validated in the Korean cohort (C-statistic of 0.79, 95% confidence interval (CI): 0.78–0.80) and the Caucasian cohort (C-statistic of 0.81, 95% CI: 0.79–0.83) and was found to have better predictive ability for HCC than previously reported models [19]. The above study demonstrates the feasibility of applying ML to disease prediction. As mentioned earlier, several models can predict the risk of developing HCC in patients with CHB. However, many studies have not yet used ML techniques or have models that lack the interpretability to adequately capture the complex relationships between variables and provide clinically actionable explanations. Today's medical infrastructure generates a significant amount of data. Therefore, more sophisticated methods rely on this data to develop more accurate models. Selecting clinically accessible predictors based on HBV cases in China is urgently needed. Combining ML technology can help develop interpretable models, which can improve the accuracy of liver cancer prediction and establish an efficient liver cancer risk-prediction model for CHB patients that is convenient for clinical practice.

The present study aimed to develop and validate an interpretable ML model for predicting the risk of HCC in patients with CHB. The SHAP method was used to clarify the importance of the features and to interpret model predictions and thus determine the practical significance of the model for predicting the occurrence of HCC. To construct the model, this study enrolled patients with HBV-related HCC and patients with CHB infection as study subjects and systematically collected epidemiological survey data (e.g., demographic information: gender, age, family history, history of smoking, and alcohol consumption), comorbidities (liver cirrhosis, diabetes, and hypertension), and clinical serological test data (routine blood tests, liver function, coagulation function, alpha-fetoprotein(AFP), HBV-DNA, etc.). Nearly 63 indicators were subjected to inter-variable descriptive and univariate analyses. After screening the predictors by LASSO-logistic regression, five ML models were constructed. The effects of the predictive models were compared to obtain the optimal model and to interpret the model. A prediction model was then established to assess the risk of HCC in the clinic and predict the risk of HCC in patients to implement prompt interventions and thus provide a reference basis for the prevention and treatment of HCC.

Methods

Information on research participants

This study was a case-control type. We collected clinical data of 393 patients with HCC and 263 patients with CHB virus infection admitted to the Fourth Affiliated Hospital of Guangxi Medical University from January to 12, 2022 through the inpatient electronic medical-record system. Inclusion criteria for patients with HCC were as follows: (1) based on the Guidelines for Diagnosis and Treatment of Primary Liver Cancer (2024 Edition) [20], patients were diagnosed as primary HCC, consistent with the diagnosis of HCC in the 10th Revision of the International Classification of Diseases (ICD-10:C22.0); (2) HBV-associated HCC; (3) complete clinical data; (4) patients had BCLC stages A, B and C. Exclusion criteria were as follows: (1) combined with primary tumours of other systems infection with other hepatotropic viruses; (2) accompanied by autoimmune liver disease, hepatolenticular degeneration, drug-induced liver disease accompanied by other serious diseases; (3) samples for laboratory indicators were collected after treatment (such as surgery, radiofrequency ablation, chemotherapy, or immunotherapy) had started; (4) patients not diagnosed with HCC for the first time; (5) under the age of 18 years. Patients diagnosed with CHB having complete clinical data and receiving antiviral therapy such as ETV or TDF were selected as the control group. In the end, 516 subjects were identified, including 273 cases and 243 control cases. The samples were randomly divided into a training set and a validation set using a 7:3 stratification, with 361 subjects in the training set and 155 subjects in the validation set. Details of the study design are shown in Fig. 1. This work was conducted in accordance with the Declaration of Helsinki and was reviewed by the Ethics Committee of the Fourth Affiliated Hospital of Guangxi Medical University (approval number: KY2021070). This study was retrospective and all data were anonymised, so informed consent was not required.

Data collection

Based on the purpose of this research and with reference to relevant clinical experience and the characteristics of patient information, the following data were collected: (1) demographic information including gender, age, height, and weight; (2) daily living habits including history of smoking and alcohol consumption; (3) chronic disease conditions including liver cirrhosis, diabetes, hypertension, cerebral infarction, coronary heart disease, arrhythmia, and renal insufficiency; (4) routine blood tests including white blood cell, neutrophil (NE), lymphocyte (Lym), monocyte, eosinophil, and red blood cell counts, haemoglobin, hematocrit, mean corpuscular volume, RBC distribution width-SD(RBC-SD), RBC distribution width-CV (RBC-CV), platelet count, platelet clotting

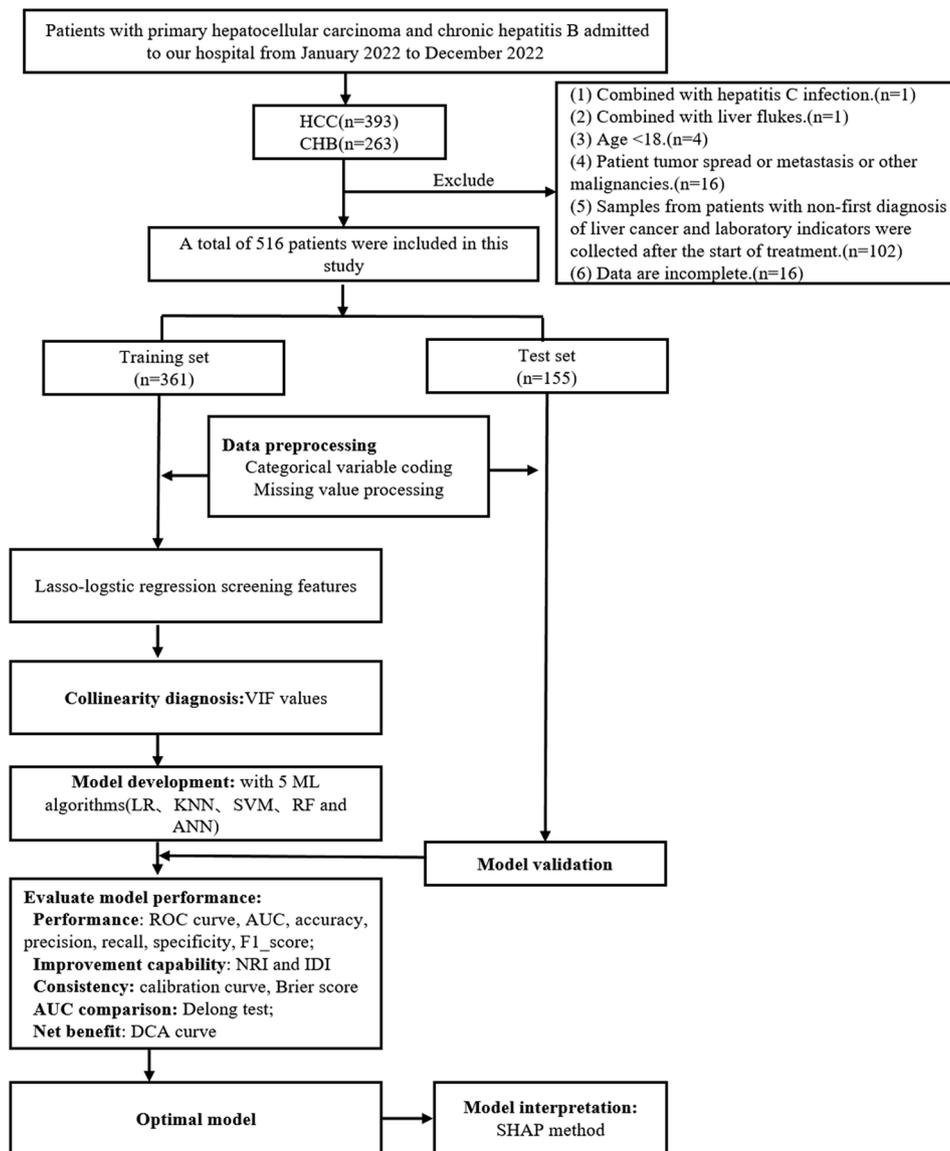


Fig. 1 Flow chart of the study design

time, platelet distribution width, and mean Corpuscular Volume; (5) liver and kidney functions including alanine aminotransferase (ALT), aspartate aminotransferase (AST), AST/ALT ratio, total bilirubin, direct bilirubin, indirect bilirubin, total bile acid, total protein, albumin globule ratio (A/G), alkaline phosphatase, and γ -glutamyl transferase (GGT); (6) coagulation function including prothrombin time, international normalised ratio, activated partial thromboplastin time, thrombin time, fibrinogen, and D-Dimer; (7) HBV-DNA and AFP.

Calculation of relevant indicators

Relevant indicators were calculated as follows: body mass index (BMI) = weight/height², NE-Lym ratio(NLR), platelet-Lym ratio(PLR), NE-monocyte-Lym ratio(NMLR), systemic immune-inflammation index(SII),

Lym-monocyte ratio(LMR), NE-Lym ratio(NPR), platelet-Lym ratio (PAR), eosinophil-Lym ratio (ELR), basophil-Lym ratio (BLR), and red-cell distribution width-platelet count ratio (RPR), which were derived from peripheral blood to assess the inflammatory status of the body. The formulas were as follows: NLR=NE/Lym, PLR=platelet/Lym, NMLR=NE×monocyte/Lym, RPR=RDW-SD/platelet, SII=platelet count × NE count/Lym count, LMR=Lym count/monocyte count, NPR=NE count/platelet count, PAR=platelet count/Lym count, ELR=eosinophil count / Lym count, BLR=basophil count/Lym ratio, and RPR=RBC distribution width/platelet.

Outcome indicator

The main outcome index of this study was whether the patient had HBV-related HCC. Based on available clinical indicators of hospitalised patients with CHB and HCC, five predictive models were developed and internally validated using ML algorithms to assess the risk of HCC in patients with CHB.

Data pre-processing and feature selection

All laboratory indicators were continuous and unclassified. Before building the model, the presence of data outliers for continuous variables was checked and replaced by the corresponding extreme values. Categorical data were processed using coding methods. To avoid data leakage, the above data pre-processing was performed on the training and validation sets, respectively. To improve data utilisation, variables with more than 30% data loss were excluded before data imputation. Data groups with a missing rate of less than 30% were processed by replacing normal distributions with the mean, non-normal distributions with the median, and non-numerical data with the mode.

Additionally, LASSO regression analysis was adopted to screen for potential risk predictors by using non-zero coefficient variables and the best penalty parameters of the model. The LASSO regression analysis was cross-validated 10 times to select penalty parameters, and predictors were initially screened when model error was minimised. The predictors screened by LASSO regression were further analysed by multi-factor logistic regression to determine the most important clinical factors. To assess collinearity between variables, the variance inflation factor (VIF) was calculated for the predictors obtained by logistic regression. According to the collinearity criterion, if $VIF > 2$, the variable is eliminated. All retained variables meet the standard of $VIF \leq 2$, indicating that the model has no significant collinearity. The predictive indicators modelled for this study were included. This method improved the predictive accuracy of the ML model and prevented overfitting.

Model development, evaluation, and interpretation

The study was performed according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines for the development and validation of predictive models [21]. Five ML models were used to predict HCC risk in patients with CHB, including LR, KNN, SVM, RF, and ANN. To optimise the prediction model, we obtained the final hyperparameters of each model on the optimal feature subset based on 10 rounds of 10× cross-validation combined with the default hyperparameter grid search of the 'caret' package. Validation sets were not used in the model tuning phase and were used only for model

evaluation after the completion of model selection and training procedures. The training set was pre-processed, and the method of Combined Random Undersampling and Synthetic Minority Oversampling Technique was used to solve the problem of sample imbalance [22]. Finally, the model was re-fitted to the validation set using the optimal feature subset and final hyperparameters (based on 10 rounds of 10× internal cross-validation).

To evaluate the performance of our model, we used confusion-matrix metrics including accuracy, precision, recall, specificity, F1 score, area under the curve (AUC), and Brier score. The Brier score was a measure of the degree of deviation between the predicted and the actual results. A lower Brier score corresponded with better performance of the prediction model [23]. The DeLong test was used to determine whether significant differences existed amongst the AUC values of different models. Continuous-net reclassification improvement (NRI) and integrated discrimination improvement (IDI) evaluated the ability of different models to improve classification efficiency, $NRI > 0$, $IDI > 0$ was positive improvement. The calibration curve was used to reflect the degree of agreement between the predicted probability and the actual result. DCA was used to evaluate the net benefit of the model at different thresholds. According to the performance of the above evaluation indices on the training set and the validation set, the best prediction model was selected.

Finally, to address the opacity of ML algorithms and their inability to facilitate clinical interpretation, we applied the SHAP method to interpret the output of the final model by calculating the contribution of each variable to the prediction. This explanatory approach provided two explanations: global interpretation of the model at the feature level and local interpretation at the individual level. Global interpretation described the overall functionality of the model. Local interpretation helped understand the decision-making mechanism of the model by calculating and displaying the contribution of each feature to the predicted outcome of a single sample. SHAP waterfall plots were used to show the contribution of each feature to the model's prediction of HCC for specific patients.

Statistical analysis

SPSS 22.0 software (IBM, NY, USA) and R4.3.1 software (R Foundation for Statistical Computing, Vienna, Austria) were used for statistical analyses. R software was primarily used with the 'calibrate', 'rms', 'e1071', 'neuralnet', 'pROC', 'caret', 'ggplot2', and 'randomForest' packages. Measures conforming to a normal distribution were expressed as the mean ± standard deviation ($\bar{x} \pm s$). Considering that HBV-DNA levels were highly skewed, they were log-transformed (\log_{10}) before

analysis. Afterwards, the variables displayed a normal distribution and were expressed as the mean \pm standard deviation ($\bar{x} \pm s$). Quantitative information that did not conform to a normal distribution even after transformation was expressed as median and quartiles [$M (P_{25}, P_{75})$]. Count information was expressed as frequency, rate, or constitutive ratio. Normal measurements were compared between the two groups using t-tests, whereas non-normal measurements were compared using non-parametric tests. The chi-squared test was used to compare two groups of categorical information. All variables in the training set were included in lasso regression for preliminary screening of predictors, followed by multi-factor logistic regression for screening important predictors. Multicollinearity tests were performed on these indicators, and indicators with a VIF < 2 were included in the model analysis. The receiver operating characteristic (ROC) curve was plotted, and the discriminative ability of the five models was assessed in combination with the AUC. To calculate the 95% CI, the 1000-fold bootstrap was used. The Brier score (ranging from 0 to 1) was used to measure the model performance by calculating the difference between the predicted probability and the actual outcome. A closer value to 0 meant better calibration effect, thereby assessing the model calibration. DeLong test was used to assess the performance of the two prediction models by comparing their AUC values. The NRI and IDI were used to assess the ability of different models to improve classification efficiency. NRI > 0 and IDI > 0 indicated positive improvement. Calibration and decision curves were used to evaluate the agreement of model predictions and actual values with the net clinical benefit of the model at different threshold probabilities, respectively. The RF model explains our use of the Python programming language (version 3.7 Python Software Foundation, Wilmington, DE). Statistical significance was set at $p < 0.05$ (two tailed).

Results

Basic characteristics of research participants

A total of 516 subjects were included, namely, 273 cases in the case group and 243 cases in the control group. Amongst them, 391 were males (75.8%) and 125 were females (24.2%). The mean age was (50.07 ± 13.95) years. The demographic and clinical characteristics of all patients are shown in Table 1. Subjects were randomly divided into the training set ($n = 361$) and the internal validation set ($n = 155$) in a 7:3 ratio, with no statistical significance between the two groups ($p > 0.05$). Additionally, there were significant differences existed between the case group and the control group in terms of age, gender, AFP, comorbidities (hypertension, diabetes, cerebral infarction, liver cirrhosis, arrhythmia, and splenomegaly), BMI, NLR, PLR, NMLR, SII, LMR, NPR, PAR, BLR, RPR,

white blood cell count, NE, Lym, monocytes, eosinophils, basophils, red blood cell count, hemoglobin, hematocrit, RDW-SD, RDW-CV, platelet, platelet volume, prothrombin time, international normalized ratio, thrombin time, fibrinogen, D-Dimer, total protein, albumin, A/G, alkaline phosphatase, ALT, AST/ALT, GGT, and HBV-DNA ($p < 0.05$), as shown in Table 1.

LASSO-logistic regression for screening feature variables

All the indicators included in the training set were used as independent variables to explore the independent risk factors associated with HCC. LASSO regression was used for variable screening. As illustrated in Fig. 2A, with an increase in the regularisation parameter λ , the regression coefficients of each variable exhibited a tendency towards zero, and the number of variables with non-zero coefficients also declined. When the minimum value of λ ($\lambda = 0.1005025$, $\text{Log}\lambda = -4.8878$) and the minimum value of λ ($\lambda = 0.02261307$, $\text{Log}\lambda = -3.7892$) were represented by thick lines, the final selection of the minimum value of λ (1 times the standard error) was identified as the optimal value (Figure B), and 15 non-zero coefficient predictor variables were selected. They included gender, age, liver cirrhosis, BLR, Lym count, red blood cell count, thrombin time, fibrinogen, D-Dimer, albumin, ALT, AST/ALT, GGT, AFP, and LOGHBV-DNA.

To further control the influence of confounding factors, the selected variables were included in multivariate logistic regression analysis. Results showed that age (OR = 1.138, 95% CI 1.083–1.210, $p < 0.0001$), BLR (OR = 3.783, 95% CI 1.901–8.343, $p = 0.0004$), D-Dimer (OR = 2.160, 95% CI 1.438–3.574, $p = 0.0007$), AST/ALT (OR = 3.889, 95% CI 1.388–10.907, $p = 0.00110$), GGT (OR = 1.005, 95% CI 1.001–1.010, $p = 0.0301$), and AFP (OR = 113.166, 95% CI 20.639–1027.827, $p < 0.0001$) were independent risk factors for HCC (Table 2). The calculated VIF was all less than 2, indicating that no cross-collinearity existed between the variables (Table 2).

Development of a model and comparison of performance

Five ML models based on LR, KNN, SVM, RF, and ANN were constructed to predict the risk of developing HCC. The RF model performed well in terms of prediction and calibration capabilities (Figs. 3A and 4A), with ROC results showing an AUC of 0.996 (95% CI: 0.991–0.999) and a Brier score of 0.025 (95% CI: 0.013–0.039) in the training set. The AUC of the ANN model was 0.994 (95% CI: 0.989–0.999), followed by the SVM model (AUC = 0.982 (95% CI: 0.970–0.994)), the LR model (AUC = 0.966 (95% CI: 0.950–0.982)) and the KNN model (AUC = 0.894 (95% CI: 0.862–0.926)). The accuracy (0.978), precision (0.966), recall rate (0.988), specificity (0.968), and F1 score (0.977) of the RF model were the highest, as shown in Table 3. DeLong test results

Table 1 Comparison of demographic and clinical characteristics of patients with HCC and CHB, as well as comparison between the training set and the validation set

| Characteristic | Overall (n=516) | Chronic hepatitis(n=243) | Hepatocellular carcinoma(n=273) | p-value | Training set(n=361) | Validation set(n=155) | p-value |
|---------------------------|-------------------------|--------------------------|---------------------------------|---------|-------------------------|-------------------------|---------|
| Age, Mean (SD) | 50.07 (13.95) | 42.29 (11.12) | 56.99 (12.49) | < 0.001 | 50.00 (13.80) | 50.21 (14.32) | 0.875 |
| Gender(%) | | | | < 0.001 | | | 0.831 |
| female | 125 (24.2) | 85 (35.0) | 40 (14.7) | | 86 (23.8) | 39 (25.2) | |
| male | 391 (75.8) | 158 (65.0) | 233 (85.3) | | 275 (76.2) | 116 (74.8) | |
| AFP (%) | | | | < 0.001 | | | 0.572 |
| <400 µg/L | 369 (71.5) | 239 (98.4) | 130 (47.6) | | 255 (70.6) | 114 (73.5) | |
| ≥400 µg/L | 147 (28.5) | 4 (1.6) | 143 (52.4) | | 106 (29.4) | 41 (26.5) | |
| Hypertension(%) | | | | 0.003 | | | 0.743 |
| No | 458 (88.8) | 227 (93.4) | 231 (84.6) | | 322 (89.2) | 136 (87.7) | |
| Yes | 58 (11.2) | 16 (6.6) | 42 (15.4) | | 39 (10.8) | 19 (12.3) | |
| Diabetes (%) | | | | 0.041 | | | 0.189 |
| No | 464 (89.9) | 226 (93.0) | 238 (87.2) | | 320 (88.6) | 144 (92.9) | |
| Yes | 52 (10.1) | 17 (7.0) | 35 (12.8) | | 41 (11.4) | 11 (7.1) | |
| Cerebral infarction(%) | | | | 0.025 | | | 0.593 |
| No | 505 (97.9) | 242 (99.6) | 263 (96.3) | | 352 (97.5) | 153 (98.7) | |
| Yes | 11 (2.1) | 1 (0.4) | 10 (3.7) | | 9 (2.5) | 2 (1.3) | |
| Liver cirrhosis(%) | | | | < 0.001 | | | 1.000 |
| No | 291 (56.4) | 193 (79.4) | 98 (35.9) | | 204 (56.5) | 87 (56.1) | |
| Yes | 225 (43.6) | 50 (20.6) | 175 (64.1) | | 157 (43.5) | 68 (43.9) | |
| Coronary heart disease(%) | | | | 0.083 | | | 1.000 |
| No | 499 (96.7) | 239 (98.4) | 260 (95.2) | | 349 (96.7) | 150 (96.8) | |
| Yes | 17 (3.3) | 4 (1.6) | 13 (4.8) | | 12 (3.3) | 5 (3.2) | |
| Arrhythmia(%) | | | | 0.021 | | | 1.000 |
| No | 482 (93.4) | 234 (96.3) | 248 (90.8) | | 337 (93.4) | 145 (93.5) | |
| Yes | 34 (6.6) | 9 (3.7) | 25 (9.2) | | 24 (6.6) | 10 (6.5) | |
| Splenomegaly(%) | | | | < 0.001 | | | 0.888 |
| No | 423 (82.0) | 216 (88.9) | 207 (75.8) | | 297 (82.3) | 126 (81.3) | |
| Yes | 93 (18.0) | 27 (11.1) | 66 (24.2) | | 64 (17.7) | 29 (18.7) | |
| Renal insufficiency(%) | | | | 0.381 | | | 0.458 |
| No | 496 (96.1) | 236 (97.1) | 260 (95.2) | | 349 (96.7) | 147 (94.8) | |
| Yes | 20 (3.9) | 7 (2.9) | 13 (4.8) | | 12 (3.3) | 8 (5.2) | |
| BMI(kg/m ²) | | | | 0.008 | | | 0.785 |
| <18.5 | 17 (3.3) | 7 (2.9) | 10 (3.7) | | 13 (3.6) | 4 (2.6) | |
| 18.5~23.9 | 321 (62.2) | 140 (57.6) | 181 (66.3) | | 220 (60.9) | 101 (65.2) | |
| 24~27.9 | 136 (26.4) | 66 (27.2) | 70 (25.6) | | 97 (26.9) | 39 (25.2) | |
| ≥28 | 42 (8.1) | 30 (12.3) | 12 (4.4) | | 31 (8.6) | 11 (7.1) | |
| Smoking(%) | | | | 0.568 | | | 0.567 |
| No | 435 (84.3) | 202 (83.1) | 233 (85.3) | | 307 (85.0) | 128 (82.6) | |
| Yes | 81 (15.7) | 41 (16.9) | 40 (14.7) | | 54 (15.0) | 27 (17.4) | |
| Alcohol use(%) | | | | 0.872 | | | 0.254 |
| No | 423 (82.0) | 198 (81.5) | 225 (82.4) | | 301 (83.4) | 122 (78.7) | |
| Yes | 93 (18.0) | 45 (18.5) | 48 (17.6) | | 60 (16.6) | 33 (21.3) | |
| NLR, Median [IQR] | 2.43 [1.74, 4.11] | 2.13 [1.57, 2.62] | 3.52 [2.21, 5.58] | < 0.001 | 2.32 [1.74, 3.91] | 2.58 [1.80, 4.43] | 0.445 |
| PLR, Median [IQR] | 120.83 [97.69, 181.11] | 119.07 [93.09, 150.08] | 141.18 [101.35, 205.81] | < 0.001 | 124.46 [99.25, 180.65] | 119.07 [91.51, 182.15] | 0.377 |
| NMLR, Median [IQR] | 2.82 [2.08, 4.70] | 2.45 [1.85, 3.02] | 3.94 [2.59, 6.42] | < 0.001 | 2.72 [2.03, 4.43] | 2.99 [2.10, 5.20] | 0.373 |
| SII, Median [IQR] | 453.65 [316.65, 789.88] | 453.65 [315.58, 526.19] | 646.26 [324.32, 1184.25] | < 0.001 | 453.65 [320.72, 783.78] | 453.65 [287.44, 858.89] | 0.999 |
| LMR, Median [IQR] | 2.67 [1.78, 3.54] | 3.14 [2.54, 4.11] | 2.02 [1.41, 3.00] | < 0.001 | 2.70 [1.82, 3.58] | 2.58 [1.64, 3.31] | 0.102 |
| NPR, Median [IQR] | 0.02 [0.01, 0.03] | 0.02 [0.01, 0.02] | 0.02 [0.02, 0.04] | < 0.001 | 0.02 [0.01, 0.03] | 0.02 [0.02, 0.03] | 0.117 |
| PAR, Median [IQR] | 5.08 [3.75, 6.22] | 5.24 [4.31, 6.02] | 4.63 [3.23, 6.72] | 0.018 | 5.06 [3.81, 6.15] | 5.15 [3.63, 6.56] | 0.663 |

Table 1 (continued)

| Characteristic | Overall (n=516) | Chronic hepatitis(n=243) | Hepatocellular carcinoma(n=273) | p-value | Training set(n=361) | Validation set(n=155) | p-value |
|---|-------------------------|--------------------------|---------------------------------|---------|-------------------------|-------------------------|---------|
| ELR, Median [IQR] | 0.07 [0.04, 0.10] | 0.07 [0.04, 0.09] | 0.07 [0.03, 0.12] | 0.997 | 0.07 [0.04, 0.10] | 0.07 [0.04, 0.10] | 0.562 |
| BLR, Median [IQR] | 0.02 [0.01, 0.03] | 0.01 [0.01, 0.02] | 0.02 [0.01, 0.03] | < 0.001 | 0.01 [0.01, 0.03] | 0.02 [0.01, 0.03] | 0.167 |
| RPR, Median [IQR] | 0.21 [0.17, 0.29] | 0.20 [0.18, 0.25] | 0.23 [0.17, 0.35] | < 0.001 | 0.21 [0.18, 0.29] | 0.20 [0.17, 0.30] | 0.995 |
| WBC, Median [IQR] | 6.37 [5.01, 7.67] | 6.37 [5.10, 6.93] | 6.62 [4.93, 8.49] | 0.006 | 6.37 [4.93, 7.43] | 6.37 [5.19, 8.02] | 0.107 |
| Neutrophil, Median [IQR] | 3.81 [2.85, 5.21] | 3.81 [2.74, 4.19] | 4.19 [2.95, 6.02] | < 0.001 | 3.81 [2.86, 5.10] | 3.81 [2.84, 5.58] | 0.27 |
| Lymphocyte, Median [IQR] | 1.48 [1.10, 1.88] | 1.79 [1.39, 2.01] | 1.29 [0.92, 1.67] | < 0.001 | 1.48 [1.10, 1.85] | 1.47 [1.10, 1.92] | 0.754 |
| Monocyte, Median [IQR] | 0.57 [0.45, 0.72] | 0.56 [0.43, 0.63] | 0.62 [0.46, 0.81] | < 0.001 | 0.57 [0.44, 0.69] | 0.59 [0.46, 0.78] | 0.017 |
| Eosinophil, Median [IQR] | 0.10 [0.05, 0.17] | 0.12 [0.06, 0.17] | 0.08 [0.04, 0.17] | 0.001 | 0.10 [0.05, 0.17] | 0.10 [0.05, 0.17] | 0.699 |
| Basophil, Median [IQR] | 0.02 [0.01, 0.04] | 0.02 [0.01, 0.03] | 0.03 [0.01, 0.04] | 0.001 | 0.02 [0.01, 0.03] | 0.02 [0.02, 0.04] | 0.096 |
| RBC, Median [IQR] | 4.46 [4.05, 4.90] | 4.73 [4.40, 5.06] | 4.20 [3.70, 4.67] | < 0.001 | 4.50 [4.07, 4.90] | 4.43 [4.00, 4.92] | 0.484 |
| Hemoglobin, Median [IQR] | 126.00 [110.00, 142.00] | 130.00 [120.00, 147.50] | 123.00 [106.00, 138.00] | < 0.001 | 125.00 [111.00, 142.00] | 126.00 [109.50, 141.50] | 0.744 |
| Hematocrit, Median [IQR] | 0.41 [0.37, 0.48] | 0.44 [0.39, 0.48] | 0.39 [0.33, 0.46] | < 0.001 | 0.41 [0.37, 0.48] | 0.41 [0.35, 0.46] | 0.309 |
| MCV, Median [IQR] | 89.10 [85.77, 94.00] | 88.40 [86.50, 93.00] | 89.70 [85.10, 94.30] | 0.205 | 89.00 [85.80, 93.90] | 89.60 [85.80, 94.15] | 0.764 |
| Mean hemoglobin content, Median [IQR] | 29.55 [28.28, 31.40] | 29.30 [28.80, 30.95] | 29.70 [27.40, 31.60] | 0.239 | 29.60 [28.20, 31.40] | 29.50 [28.45, 31.25] | 0.875 |
| Mean hemoglobin concentration, Median [IQR] | 329.56 [323.00, 337.00] | 329.56 [325.00, 336.00] | 329.04 [322.00, 339.00] | 0.680 | 329.56 [324.00, 337.00] | 329.56 [322.00, 337.00] | 0.430 |
| RDW-SD, Median [IQR] | 43.35 [40.20, 47.70] | 43.00 [39.80, 44.65] | 46.00 [40.90, 50.20] | < 0.001 | 43.35 [40.20, 47.50] | 43.35 [40.55, 47.90] | 0.581 |
| RDW-CV, Median [IQR] | 13.85 [12.50, 15.10] | 13.40 [12.40, 14.25] | 14.30 [12.80, 16.00] | < 0.001 | 13.70 [12.50, 14.90] | 13.94 [12.70, 15.90] | 0.127 |
| Platelet, Median [IQR] | 204.00 [157.75, 245.00] | 213.13 [179.50, 236.00] | 192.00 [133.00, 247.00] | 0.006 | 203.09 [158.00, 241.00] | 213.00 [152.50, 250.00] | 0.844 |
| Mean Platelet Volume, Median [IQR] | 10.33 [9.80, 10.80] | 10.33 [9.80, 10.70] | 10.37 [9.80, 10.80] | 0.284 | 10.33 [9.80, 10.80] | 10.33 [9.75, 10.70] | 0.576 |
| Platelet Volume Fraction, Median [IQR] | 0.21 [0.18, 0.24] | 0.22 [0.20, 0.24] | 0.21 [0.15, 0.25] | < 0.001 | 0.21 [0.18, 0.24] | 0.21 [0.16, 0.24] | 0.547 |
| Platelet Distribution Width, Median [IQR] | 11.80 [10.60, 12.62] | 11.99 [10.85, 12.60] | 11.79 [10.40, 12.70] | 0.056 | 11.90 [10.70, 12.60] | 11.79 [10.55, 12.70] | 0.643 |
| P-LCR, Median [IQR] | 27.45 [22.87, 31.20] | 27.45 [23.55, 31.00] | 27.55 [22.50, 31.70] | 0.509 | 27.45 [22.90, 31.30] | 27.45 [22.75, 30.70] | 0.493 |
| PT, Median [IQR] | 11.60 [10.70, 12.50] | 11.30 [10.50, 11.90] | 11.80 [11.00, 13.00] | < 0.001 | 11.50 [10.80, 12.30] | 11.60 [10.70, 12.85] | 0.348 |
| INR, Median [IQR] | 1.01 [0.93, 1.10] | 0.98 [0.91, 1.04] | 1.03 [0.96, 1.15] | < 0.001 | 1.01 [0.93, 1.09] | 1.02 [0.93, 1.13] | 0.358 |
| APTT, Median [IQR] | 27.30 [24.40, 29.80] | 27.30 [24.90, 28.65] | 27.30 [24.00, 30.40] | 0.619 | 26.80 [24.20, 29.70] | 27.71 [24.90, 30.00] | 0.090 |
| TT, Median [IQR] | 20.30 [19.30, 21.10] | 20.66 [19.50, 21.25] | 20.10 [19.00, 21.00] | < 0.001 | 20.30 [19.30, 21.10] | 20.30 [19.20, 21.15] | 0.762 |
| Fibrinogen, Median [IQR] | 2.55 [2.09, 3.07] | 2.38 [1.94, 2.63] | 3.07 [2.35, 3.45] | < 0.001 | 2.53 [2.09, 3.07] | 2.64 [2.08, 3.07] | 0.669 |
| D-Dimer, Median [IQR] | 1.62 [1.43, 3.18] | 1.62 [1.62, 1.62] | 2.68 [0.79, 4.24] | < 0.001 | 1.62 [1.56, 3.24] | 1.62 [1.37, 3.16] | 0.742 |
| Total Protein, Median [IQR] | 68.80 [64.88, 72.25] | 70.05 [67.10, 73.05] | 66.85 [63.00, 71.40] | < 0.001 | 69.00 [64.70, 72.60] | 68.20 [65.10, 72.20] | 0.796 |
| Albumin, Median [IQR] | 40.35 [34.60, 43.90] | 40.68 [37.55, 43.75] | 37.80 [32.20, 44.40] | < 0.001 | 40.50 [34.60, 43.90] | 40.10 [34.45, 43.85] | 0.476 |
| A/G, Median [IQR] | 1.36 [1.00, 1.83] | 1.36 [1.00, 1.69] | 1.35 [1.00, 3.31] | 0.021 | 1.37 [1.00, 1.82] | 1.34 [1.01, 1.83] | 0.776 |
| Total Bilirubin, Median [IQR] | 15.85 [10.20, 36.18] | 13.60 [9.40, 40.46] | 17.30 [10.60, 32.16] | 0.421 | 15.70 [10.20, 37.00] | 15.90 [10.10, 32.16] | 0.805 |

Table 1 (continued)

| Characteristic | Overall (n=516) | Chronic hepatitis(n= 243) | Hepatocellular carcinoma(n=273) | p-value | Training set(n=361) | Validation set(n=155) | p-value |
|----------------------------------|------------------------|---------------------------|---------------------------------|---------|------------------------|------------------------|---------|
| Direct Bilirubin, Median [IQR] | 8.10 [5.00, 24.55] | 6.90 [4.50, 31.14] | 9.10 [5.80, 24.43] | 0.121 | 7.90 [5.10, 24.90] | 8.30 [4.95, 24.43] | 0.761 |
| Indirect Bilirubin, Median [IQR] | 6.30 [3.90, 9.20] | 6.40 [3.80, 9.20] | 6.00 [4.00, 8.30] | 0.289 | 6.40 [4.00, 9.20] | 6.00 [3.80, 9.05] | 0.429 |
| ALT, Median [IQR] | 52.00 [26.00, 161.00] | 105.00 [31.50, 205.20] | 41.00 [24.00, 69.59] | < 0.001 | 54.00 [27.00, 166.00] | 49.00 [24.00, 135.00] | 0.275 |
| ALP, Median [IQR] | 105.00 [78.75, 165.25] | 101.00 [70.50, 111.50] | 136.00 [88.00, 192.00] | < 0.001 | 104.00 [79.00, 157.00] | 108.00 [79.00, 170.50] | 0.566 |
| AST, Median [IQR] | 64.50 [32.00, 134.12] | 72.00 [27.00, 134.12] | 61.00 [37.00, 129.78] | 0.883 | 66.00 [33.00, 134.12] | 61.00 [31.00, 134.12] | 0.712 |
| AST/ALT, Median [IQR] | 1.10 [0.80, 2.00] | 0.90 [0.60, 1.00] | 1.60 [1.10, 2.40] | < 0.001 | 1.10 [0.80, 1.90] | 1.20 [0.80, 2.10] | 0.389 |
| GGT, Median [IQR] | 87.79 [42.00, 197.25] | 81.00 [28.00, 102.00] | 158.00 [61.00, 253.00] | < 0.001 | 87.79 [43.00, 209.41] | 87.79 [36.50, 166.00] | 0.360 |
| Total Bile Acids, Median [IQR] | 20.10 [6.50, 40.88] | 23.00 [6.25, 40.88] | 19.40 [6.80, 33.10] | 0.101 | 19.20 [6.50, 40.88] | 23.10 [6.90, 40.88] | 0.387 |
| HBV-DNA, Median [IQR] | 4.45 (1.76) | 4.17 (1.56) | 4.69 (1.88) | 0.001 | 4.17 [3.12, 6.36] | 4.17 [2.60, 6.36] | 0.734 |

AFP, Alpha-fetoprotein; BMI, Body mass index; NLR, Neutrophil-lymphocyte ratio; PLR, Platelet-lymphocyte ratio; NMLR, Neutrophil-monocyte-lymphocyte ratio; SII, Systemic immune-inflammation index; LMR, Lymphocyte-monocyte ratio; NPR, Neutrophil-lymphocyte ratio; PAR, Platelet-lymphocyte ratio; ELR, Eosinophil-lymphocyte ratio; BLR, Basophil-lymphocyte ratio; RPR, Red-cell distribution width-platelet count ratio; WBC, White blood cell; RBC, Red blood cell; MCV, Mean corpuscular volume; RDW-SD, Red cell distribution width - standard deviation; RDW-CV, Red cell distribution width - coefficient of variation; P-LCR, Platelet large cell ratio; PT, Prothrombin time; INR, International Normalized Ratio; APTT, Activated Partial Thromboplastin Time; TT, Thrombin time; ALT, Alanine aminotransferase; ALP, Alkaline phosphatase; AST, Aspartate aminotransferase; AST/ALT, Aspartate aminotransferase/alanine aminotransferase Ratio; GGT, γ - Glutamyl transferase

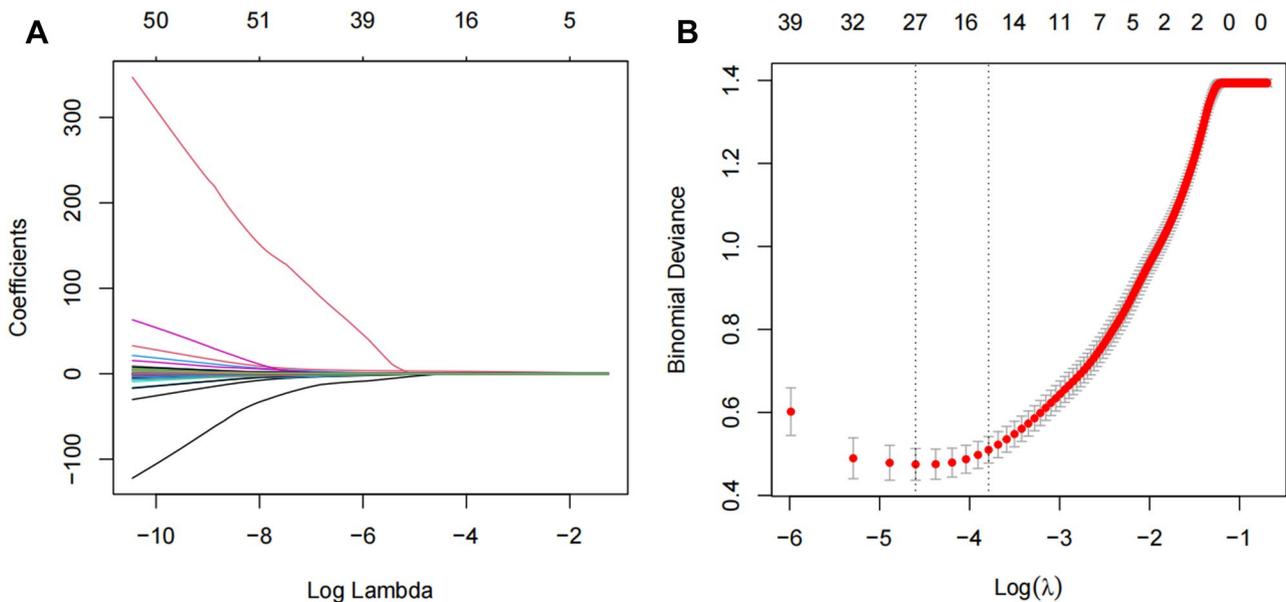


Fig. 2 LASSO regression model for clinical feature selection. **(A)** Plot of the model's coefficient distribution for logarithmic (lambda) sequences at different penalty levels. **(B)** For cross verification, the first column represents the minimum error, while the second column represents the cross-verification error of 1 standard deviation

showed that when RF model was used as reference, ANN ($Z=0.3941$, $p=0.6935$) had no statistical difference. LR model ($Z=3.7834$, $p=0.0002$), KNN model ($Z=6.4409$, $p<0.0001$) and SVM model ($Z=2.2513$, $p=0.02436$) had lower predictive power. Furthermore, NRI and IDI evaluated the model's ability to improve the classification effect in the training set, and the results showed that the RF model had positive improvement ability compared

with ANN, SVM, and KNN model ($NRI>0$, $IDI>0$). DCA showed that each model had good clinical practicality (Fig. 5A).

In the validation set, the RF model performed best in prediction and calibration with an AUC of 0.993 (95% CI: 0.986–1) and a Brier score of 0.043(95% CI: 0.028–0.054) (Figs. 3B and 4B). The DeLong test showed that the AUC of the RF model was higher than that of the LR

Table 2 Multivariate Logistic regression analysis of independent risk factors

| Characteristic | Estimate | SE | Wald χ^2 | p-value | OR(95%CI) | VIF |
|-----------------------------------|----------|-------|---------------|---------|--------------------------|-------|
| Age | 0.130 | 0.028 | 4.666 | <0.0001 | 1.138(1.083–1.210) | 1.649 |
| BLR | 1.331 | 0.374 | 3.561 | 0.0004 | 3.783(1.901–8.343) | 1.423 |
| D-Dimer | 0.770 | 0.227 | 3.399 | 0.0007 | 2.160(1.438–3.574) | 1.608 |
| AST/ALT | 1.358 | 0.534 | 2.542 | 0.0110 | 3.889(1.388–10.907) | 1.421 |
| GGT | 0.005 | 0.002 | 2.169 | 0.0301 | 1.005(1.001–1.010) | 1.320 |
| AFP(≥ 400 $\mu\text{g/L}$) | 4.729 | 0.975 | 4.848 | <0.0001 | 113.166(20.639–1027.837) | 1.355 |
| (Intercept) | -1.995 | 7.005 | -0.285 | 0.776 | 0.136(—) | — |

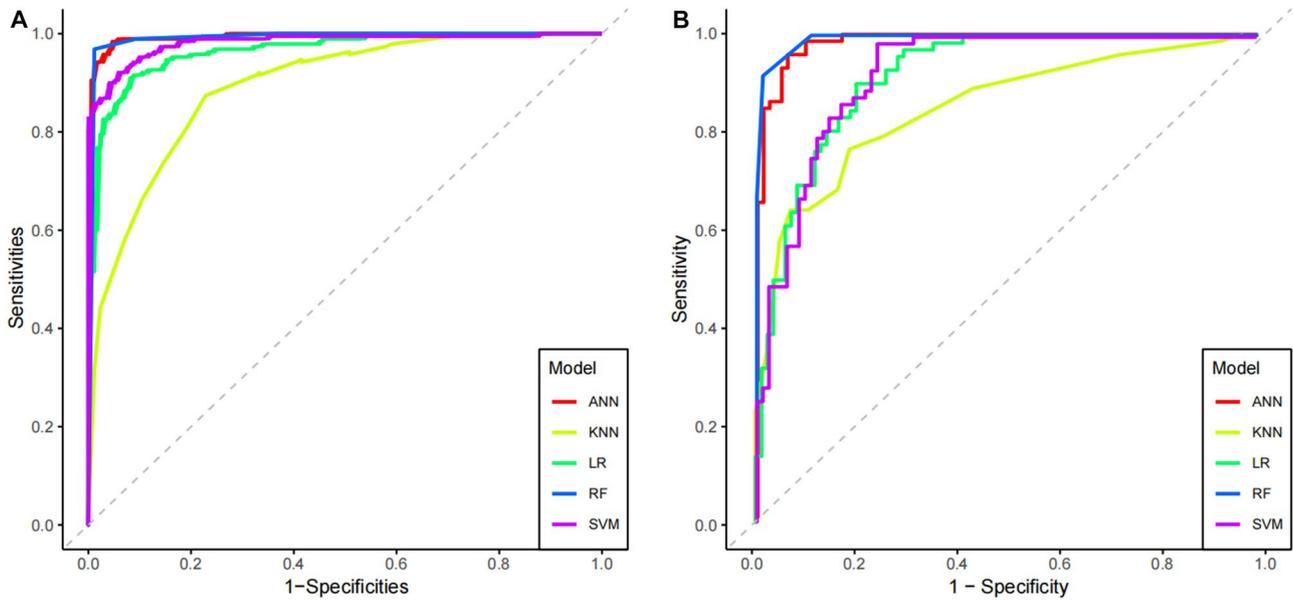


Fig. 3 Receiver operating characteristic analysis of the training set (A) and validation set (B)

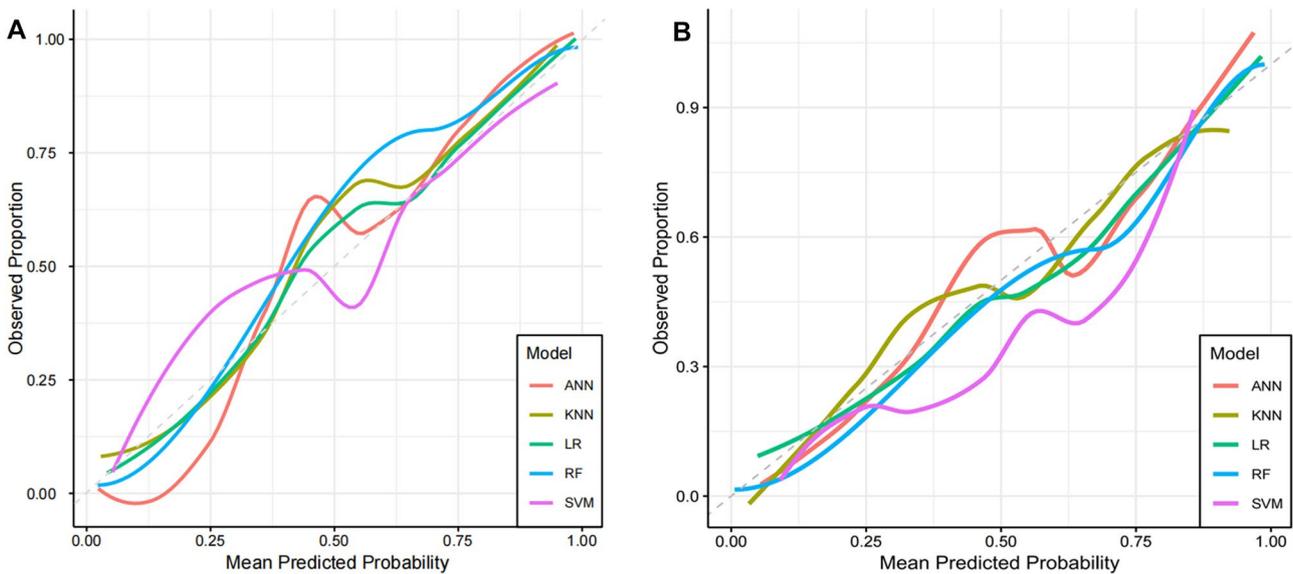


Fig. 4 Calibration curves of five machine learning models predicting HCC risk in patients with CHB in training set (A) and validation set (B)

Table 3 Performance of five machine learning methods in predicting HCC risk in CHB patients in training and validation sets

| Model | Accuracy(95% CI) | Precision(95% CI) | Recall(95% CI) | Specificity(95% CI) | F1 Score(95% CI) | AUC(95% CI) | Brier score(95% CI) |
|------------------------|--------------------|--------------------|--------------------|---------------------|--------------------|--------------------|---------------------|
| Train- ing set | | | | | | | |
| LR | 0.903(0.868–0.932) | 0.878(0.823–0.926) | 0.924(0.880–0.961) | 0.884(0.821–0.922) | 0.900(0.870–0.938) | 0.966(0.950–0.982) | 0.070(0.052–0.089) |
| KNN | 0.808(0.764–0.848) | 0.798(0.747–0.864) | 0.807(0.761–0.875) | 0.811(0.725–0.851) | 0.800(0.754–0.869) | 0.894(0.862–0.926) | 0.132(0.113–0.153) |
| SVM | 0.925(0.893–0.950) | 0.923(0.886–0.963) | 0.918(0.880–0.959) | 0.932(0.873–0.959) | 0.921(0.883–0.961) | 0.982(0.970–0.994) | 0.054(0.037–0.072) |
| RF | 0.978(0.957–0.990) | 0.966(0.933–0.988) | 0.988(0.962–0.999) | 0.968(0.927–0.987) | 0.977(0.947–0.993) | 0.996(0.991–0.999) | 0.025(0.013–0.039) |
| ANN | 0.964(0.939–0.981) | 0.965(0.959–0.968) | 0.959(0.926–0.985) | 0.968(0.925–0.987) | 0.962(0.929–0.987) | 0.994(0.989–0.999) | 0.027(0.017–0.038) |
| Valida- tion set | | | | | | | |
| LR | 0.832(0.764–0.887) | 0.811(0.733–0.905) | 0.833(0.756–0.921) | 0.831(0.703–0.893) | 0.822(0.744–0.913) | 0.915(0.873–0.958) | 0.115(0.087–0.148) |
| KNN | 0.761(0.686–0.826) | 0.778(0.733–0.904) | 0.681(0.649–0.834) | 0.831(0.656–0.872) | 0.726(0.688–0.868) | 0.846(0.784–0.908) | 0.155(0.126–0.185) |
| SVM | 0.839(0.771–0.892) | 0.841(0.775–0.932) | 0.806(0.742–0.908) | 0.867(0.733–0.918) | 0.823(0.758–0.920) | 0.913(0.884–0.964) | 0.121(0.089–0.153) |
| RF | 0.942(0.893–0.973) | 0.889(0.804–0.949) | 1.000(0.951–1.000) | 0.892(0.800–0.948) | 0.942(0.872–0.974) | 0.993(0.986–1.000) | 0.043(0.028–0.054) |
| ANN | 0.897(0.885–0.969) | 0.931(0.865–0.980) | 0.931(0.865–0.980) | 0.891(0.845–0.978) | 0.931(0.865–0.980) | 0.987(0.975–0.999) | 0.046(0.029–0.069) |

($Z = 3.7046$, $p = 0.000212$), KNN ($Z = 4.7276$, $p < 0.0001$), SVM ($Z = 3.5405$, $p < 0.0001$) and ANN ($Z = 1.1319$, $p = 0.2576$) models. However, there was no statistical significance between the RF and ANN models. The RF model had the highest accuracy (0.942), recall (1.000), specificity (0.892), and F1 score (0.942). NRI and IDI analysis showed that the LR model performed better than the KNN, RF, ANN and SVM models in terms of reclassification and overall discriminative ability (NRI > 0, IDI > 0). DCA showed that when the threshold of the model was set within a certain range, the decision curve lay above the None line and the All line, where the model had clinical practicability (Fig. 5B). In conclusion, the RF model performed well in the training and test sets and was thus the preferred model for predicting HCC risk in patients with CHB.

Interpretation of RF model

To make the predictive model more intuitive for clinicians to understand and accept, we adopted the SHAP approach. The results of the RF model were clearly interpreted by quantifying the contribution of each variable to the predicted outcome. As shown in the SHAP summary bar chart (Fig. 6A), the average SHAP value was used to evaluate the contribution of features to the model, ranked from highest to lowest as AST/ALT, AFP, D-Dimer, age, GGT, and BLR. The SHAP summary scatter plot (Fig. 6B) visually showed the direction and

intensity of each feature's influence on the model prediction. A higher feature value corresponded with a higher likelihood of HCC. The red and blue dots represented higher and lower eigenvalues, respectively. SHAP dependency plots are used to visualise the impact of individual features on model predictions. Specifically, the x-axis is the feature value and the y-axis is the SHAP value. These graphs can visually show whether a particular feature has a positive or negative effect on the model prediction. Figure 6D shows that six features have significant positive effects on the model prediction results.

Additionally, local interpretation helped understand the decision-making mechanism of the model by calculating and displaying the contribution of each feature to the predicted outcome of a single sample. The SHAP waterfall plot (Fig. 6C) shows the contribution of each feature to the model's prediction of HCC occurrence for a given patient. The specific value of each feature in the figure and its corresponding SHAP value represented the positive and negative influence of the feature on the prediction result. In this patient, AFP, D-Dimer, and AST/ALT had significant positive contributions to the prediction results of +0.17, +0.17, and +0.1 respectively. Conversely, BLR, age, and GGT had significant negative contributions to the prediction results of -0.12, -0.05, and -0.02, respectively. The SHAP waterfall plot visually demonstrated the formation of patient-specific

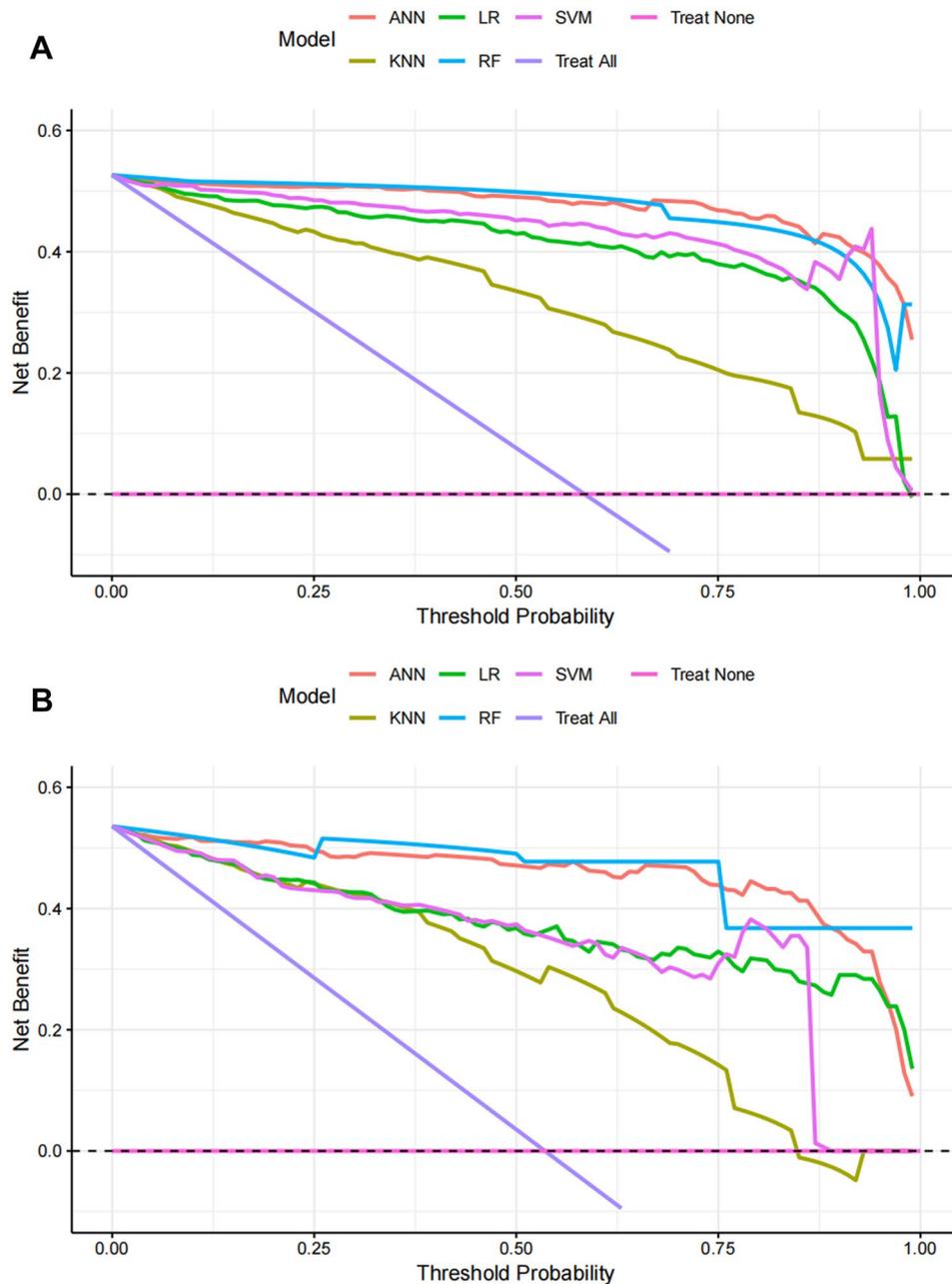


Fig. 5 DCA curves of five machine learning models predicting HCC risk in patients with CHB in training set (A) and validation set (B)

predictions by accumulating SHAP values, providing insight into the model’s decision-making mechanisms.

Discussion

HCC caused by HBV infection is a serious public health problem worldwide. Although antiviral therapy can effectively inhibit HBV replication and reduce the risk of HCC development, the risk of HCC after treatment remains a major concern. HCC is one of the most common malignant tumours, posing a serious threat to human life, health and safety. The treatment and prognosis of

patients are severely compromised due to occult presentation and low early diagnosis rate. Additionally, HCC is highly refractory to therapeutic intervention, with 70% of patients experiencing tumour recurrence within 5 years even after surgical resection or ablation [24]. Once the tumour has progressed to an advanced stage, currently available drug therapies offer only marginal survival benefits and are not cost effective [25]. Therefore, preventing the development and progression of HCC in high-risk patients rather than treating advanced disease with limited health benefits is prudent to consider.

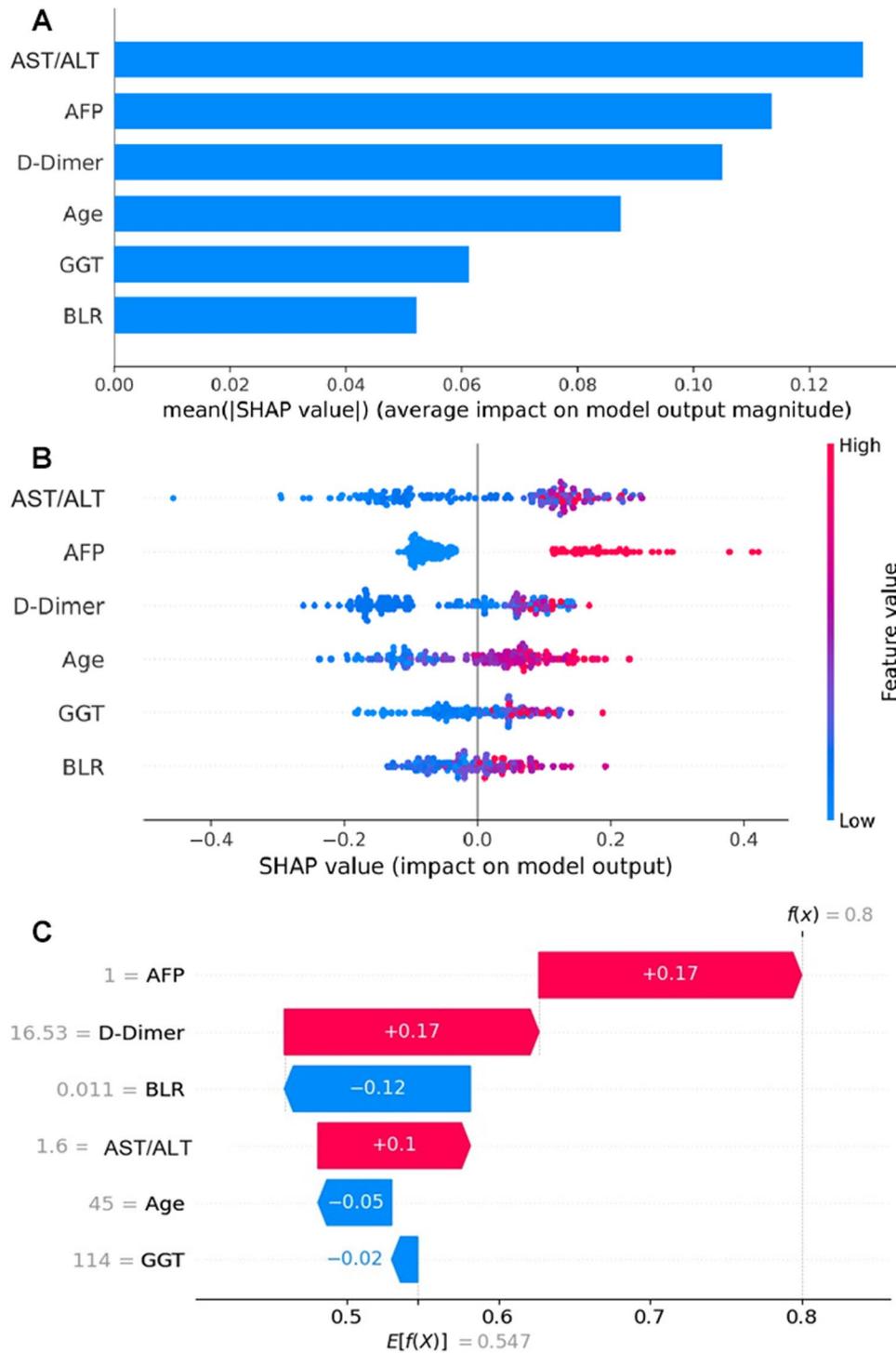


Fig. 6 Global and local model explanation by the SHapley Additive exPlanation (SHAP) method. **(A)** SHAP summary bar plot. **(B)** SHAP summary dot plot. **(C)** SHAP waterfall plot. **(D)** SHAP correlation plot for all predictors

Studies have been reported to predict the risk of HCC in CHB [26]. However, most of these studies use a single method (traditional logistic regression) for model construction, which may not be able to deal with complex problems, thereby affecting the predictive performance.

ML is described as a “marriage between Mathematics and Computer Science” and has proven to be a promising method for selecting biomarkers and building prognostic models [27]. Some studies have also reported some differences between traditional and ML-based methods

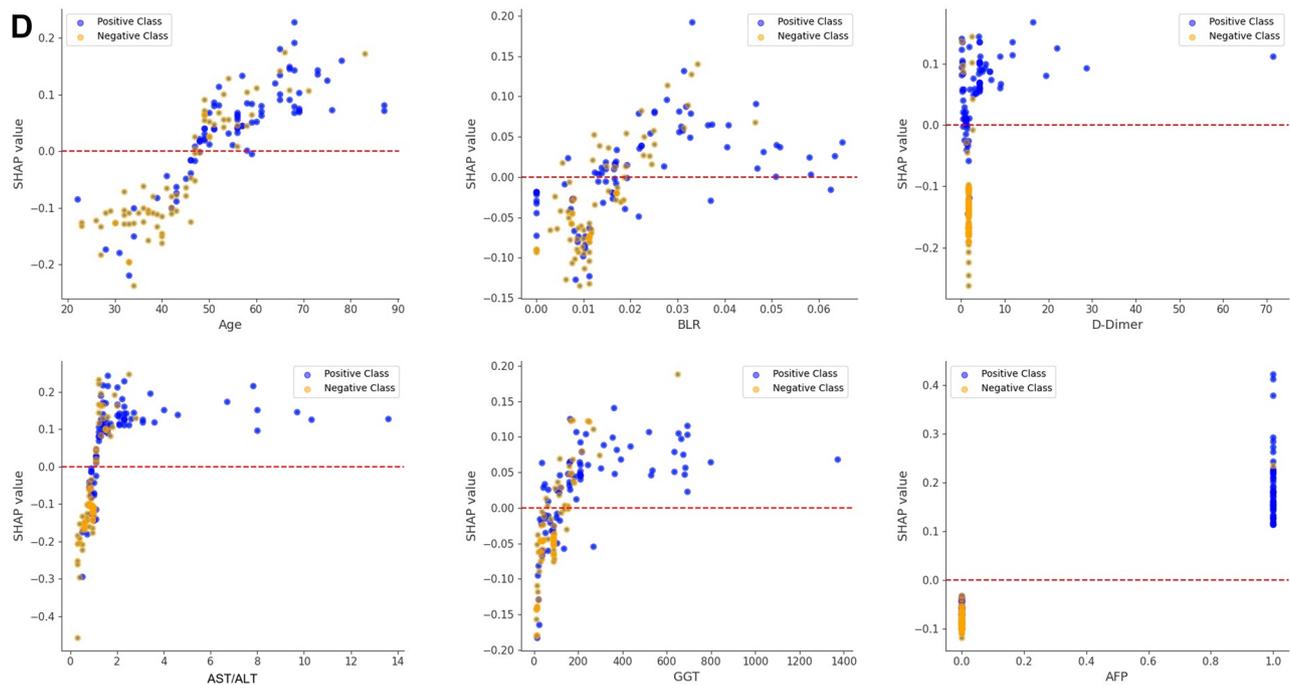


Fig. 6 (continued)

[28]. Data from clinical electronic health-record (EHR) systems are relatively objective, accurate, and easy to obtain. For clinicians and researchers, combining clinical EHR data with sophisticated ML algorithms can facilitate the development of clinical prediction models. Based on clinical and experimental data from the clinical EHR system, a set of predictors (age, BLR, D-Dimer, AST/ALT, GGT, and AFP) were identified by LASSO–logistic regression to predict HCC risk in patients with CHB. An interpretable ML model was constructed and validated using ML algorithms to assess HCC risk in CHB patients, which can be a valuable tool for the identification and early intervention of HCC risk in CHB patients.

Amongst the five ML models constructed in this study, the RF model has the highest AUC, with good accuracy and net benefit. The model also performs well in internal validation. This study shows that the RF-based prediction model has better discriminative ability than other ML algorithms. Compared to ANN, RF models are more robust when dealing with limited sample sizes and noisy data. Although ANN has significant advantages in modelling non-linear relationships, with its multi-layer structure capable of capturing complex feature interactions and relationships between high-dimensional data, its performance is hampered by sample size limitations and the complexity of data characteristics. Under small sample conditions, ANN is prone to overfitting and sensitive to the adjustment of hyper-parameters such as the number of layers and the number of nodes. In contrast, the RF algorithm is an ensemble learning method that solves

classification and regression problems by constructing multiple decision trees. It subsequently improves prediction accuracy, generalisation ability, and anti-overfitting ability by averaging (regression problem) or majority voting (classification problem) [29]. RF offers the advantages of high classification accuracy, identification of the significance of variables, multiple data analyses, and modelling of complex interactions between explanatory variables [30]. The results show that the RF model outperforms the ANN in terms of sensitivity, specificity and AUC, indicating that the RF model is more robust in processing limited sample size and high noise data. This result suggests that when designing predictive models, researchers should select appropriate algorithms according to data characteristics and research objectives, and explore model optimisation to fully exploit the potential of the data. At the same time, future research can consider integrating multiple models with a larger sample size or broader research background to improve the generalisation ability and anti-overfitting performance of the model. These measures will also help to improve the reliability and applicability of the model in clinical practice. Several studies have shown that RF algorithms are very valuable for predictive models in the medical field [31, 32]. Wang et al. [33] developed an RF model for macrosomia detection. The sensitivity, specificity, and AUC were 91.7%, 91.7%, and 0.953, respectively, which were significantly higher than logistic regression. Wang et al. [34] also developed five ML models to predict the risk of structural recurrence in papillary thyroid cancer patients.

Their RF performance was significantly higher than that of LR, SVM, extreme gradient boosting, and NN. In the current study, we developed and interpreted the final model with six features using the RF algorithm. These features can be easily obtained and evaluated during a patient's hospital stay, making the model a promising tool for effectively predicting the risk of developing HCC in CHB patients.

Currently, guidelines and consensus to guide the selection of features in predictive models are lacking. The number of features that should be included in the model remains unclear. More predictive features in the model usually mean better performance, but too many features may limit the clinical application of the model, and the addition of confounders also reduces the accuracy of prediction. Therefore, LASSO regression combined with multi-factor logistic method was used for feature selection in this study. Six indicators (age, BLR, D-Dimer, AST/ALT, GGT, and AFP) were finally identified as predictive indicators of the model, which are easy to obtain in daily clinical practice, making the model conducive to clinical decision making. AFP is an HCC-specific tumour marker and the most widely used biomarker for HCC worldwide [35, 36]. Determination of AFP levels has been used to monitor the onset and progression of HCC, as well as to assess the effectiveness of treatment and predict prognosis. High AFP levels are strong predictors of HCC risk. AFP can be used indirectly as an indicator of the stage of fibrosis in chronic hepatitis C virus infection [37]. AFP levels are associated with cirrhosis stage and HCC risk and also positively correlated with tumour size [38]. High AFP levels are associated with aggressive tumour biology and survival. AFP promotes carcinogenesis by promoting proliferation, immune escape, angiogenesis, invasion, metastasis, and recurrence of liver cancer cells. Consistent with previous studies, this study showed that high serum AFP levels (≥ 400 $\mu\text{g/L}$) predicted an increased risk of HCC. However, the increased AFP concentration in HCC patients and the mechanism by which AFP is associated with HCC pathogenesis are not fully understood. The underlying mechanism may be that AFP expression is associated with the expression of several proteins involved in angiogenesis and iron metabolism. Iron overload promotes the development of liver cancer through the production of oxygen-reactive substances and carcinogenic oxidative damage.

AST and ALT are liver enzymes that can indicate hepatocellular damage. The AST/ALT ratio is a non-invasive biomarker of liver function damage and can be used to assess the causes of liver diseases such as hepatic fibrosis, cirrhosis, and non-alcoholic fatty liver disease [39]. High levels of AST/ALT can be used as an indicator of HBV infection. AST/ALT may also serve as a biomarker for inflammation. In patients with cirrhosis, elevated AST/

ALT is significantly associated with progression of liver dysfunction and stage of cirrhosis and provides prognostic information similar to existing scoring systems such as the end-stage liver disease model or the Child–Pugh score [40]. Previous studies have shown that the AST/ALT ratio is a predictor of liver fibrosis and cirrhosis in patients with chronic HCV infection. A higher AST/ALT ratio is associated with an increased risk of death and relapse. High AST/ALT levels may reflect severe liver necrosis, leading to HCC invasion and recurrence. Increasing evidence indicates that the ratio usually exceeds 2.0 in patients with alcoholic liver disease and is less than 1.0 in patients with chronic hepatitis and chronic cholestatic syndrome [41]. In this study, AST/ALT levels in patients with CHB were significantly lower than those in patients with HCC (AST/ALT: 0.9 vs. 1.6, $p < 0.001$), and elevated AST/ALT levels also predicted an increased risk of HCC (OR: 3.889 (1.388–10.907), $p < 0.0001$). The underlying pathophysiological mechanism is unclear, but it may be related to the disruption of cancer-related changes in cell metabolism that is a hallmark of the disease in HCC, which may lead to the observed changes in AST and ALT levels [42]. Patients with high AST/ALT ratios may represent severe liver disease, so close monitoring is essential to control potential risks and optimise treatment behaviour. Careful management is essential when treating CHB patients with high AST/ALT ratios.

Serum GGT is an important enzyme in glutathione metabolism and cellular stress response and is attracting increased research attention [43]. GGT is a marker of liver damage. Alcohol consumption, acute and chronic liver disease, and oxidative stress can all contribute to elevated levels. Previous studies [44] have shown that elevated GGT levels are associated with an increased risk of cancer of the liver, colon, oesophagus, lung, and stomach. Hu et al. [45] indicated that high serum GGT levels, as measured by the highest quartile of serum GGT, increased the risk of primary liver cancer. Our results also suggest that elevated GGT levels predict an increased risk of HCC. Potential mechanisms may include cell damage, oxidative stress, inflammatory response, and bile acid metabolism [46]. Therefore, GGT may play an important role in assessing the risk of liver cancer in patients, and monitoring GGT levels may help in the early detection of liver cancer and its associated risks, especially in high-risk populations.

Additionally, systemic inflammation has been shown to be an important feature of malignant tumours, and the body's inflammation and immune response play a major role in the progression from chronic hepatitis to HCC. BLR, as a comprehensive indicator for predicting inflammation and immune response in blood routine, has the advantage of being simple and easy to obtain. Elevated

BLR is reportedly associated with improved overall survival in patients with non-small cell lung cancer [47]. Additionally, BLR is considered an independent prognostic factor for recurrence-free survival in distal cervical cancer [48]. Liao et al. [49] pointed out that PLR, NLR, MLR, BLR, and ELR all increase in patients with stable chronic obstructive pulmonary disease compared with healthy subjects. PLR, NLR, MLR, and BLR were further increased in the exacerbation phase. However, this study also examined the relationship between NLR, PLR, NMLR, SII, and other indicators and the risk of HCC, but no statistically significant difference was observed. This may be primarily due to the bias caused by the inclusion of samples, or the fact that CHB patients received antiviral treatment, resulting in changes in the relevant indicators.

D-Dimer is an indicator of coagulation and fibrinolysis that can indirectly determine thrombotic activity and is a reliable indicator of the degree of activation of the coagulation system [50]. Hypercoagulability plays an important role in tumour angiogenesis, invasion, and metastasis in patients with malignant tumours [51]. High plasma D-Dimer levels are strongly associated with poor prognosis in patients with colorectal cancer [52], but the screening value of D-Dimer has been neglected. The results of our study indicated that D-Dimer level was a valuable screening indicator for HBV-associated HCC. Thus, monitoring the D-Dimer should be considered in clinical practice to facilitate the early identification of high-risk patients and support timely intervention and treatment.

Finally, the results of this study showed that age was one of the most important risk factors for HCC. HCC development is known to be a gradual process that can take years or even decades. This may be related to the longer course of hepatitis B, host autoimmune resistance, and clinical antiviral therapy. The asymptomatic incubation period of HBV infection is long, and accurately determining the duration of HBV infection in real life is difficult. Age may also reflect the duration of HBV infection in China. Previous studies have shown that old age is a risk factor for liver cancer in patients with chronic liver disease [53]. Liu et al. [54] proposed similar results, suggesting that age is an independent predictor for the diagnosis of AFP-negative HCC. HCC reportedly occurs rarely before the age of 40 and peaks at around 60 years, consistent with the mean age of 56.99 years in the HCC group in this study.

Therefore, increased age, BLR, D-Dimer, AST/ALT, GGT, and AFP were important predictors of the risk of HCC in CHB patients. In clinical practice, the clinical monitoring of changes in the above indicators in HBV-infected patients can help to predict and diagnose the risk of HCC, as well as assess the liver function and

disease progression of patients. If the above indicators increase in clinical practice, we should be vigilant.

Strengths and limitations

ML techniques are often described as 'black boxes', making their predictive processes almost impossible to explain. This lack of transparency can make clinicians reluctant to use these technologies because they are reluctant to make medical decisions based on opaque information. However, a major strength of this study was that we use SHAP methods to meaningfully expose the 'black box' of ML models. The SHAP methods clarify the function of the model by providing global and local explanations that detail how personalised input data can be used to make specific predictions about individual patients. Additionally, another benefit of this study was the comparison of the predictive performance of different ML models for HCC risk in CHB. The performance evaluation of the model in the training and internal-validation sets and the model comparison show that the RF model had a better predictive value. Finally, the predictive factors included in the model in this study were all routine items during the hospitalisation of patients. These factors were accessible and affordable and can make full use of the available examination items for the risk assessment of CHB patients, thereby providing feasibility for the promotion and application of the model in clinical practice.

However, this study had several notable limitations. Firstly, it was a single-centre study with a small total sample size and inevitable selection bias. Secondly, this study focused only on a single centre, so only internal validation was performed without the support of external validation, proving that the stability of the predictive model performance was necessary. Therefore, future efforts are needed to conduct multi-centre prospective studies and provide more opportunities for multi-centre collaboration and better data-mining capabilities. Thirdly, geographical limitations may also affect the generalisability of the results. Because studies are based on patient data from specific regions, they may not reflect conditions in other regions or populations, which limits the general applicability of the results. Fourthly, CHB patients receiving treatment, but antiviral treatment was not included as part of the study. This limitation may affect the results of the study, as antiviral therapy may have an impact on serological markers (such as HBV-related indicators, liver function, etc.) and liver morphology. Considering the universality of the model, only routine laboratory test results easily obtained in clinical practice were included. We did not include some novel molecular markers associated with liver cancer risk, such as dehydroepiandrosterone sulphate [55] and miRNA [56]. Finally, many factors (such as imaging indicators, BCLC stage, physical

status, etc.) may be associated with HCC risk, but they were not fully included in this study, which may limit the comprehensiveness and predictive performance of the model. Therefore, more variables need to be added in the future to further optimise the prediction model and improve its clinical application value.

Conclusion

We identified age, BLR, D-Dimer, AST/ALT, GGT, and AFP as predictors of HCC risk in CHB patients and constructed and validated an interpretable ML model to assess the risk of HCC in CHB patients. It may provide a valuable tool for the identification and early intervention of HCC in patients with CHB.

Abbreviations

| | |
|-------|---|
| AFP | Alpha fetoprotein |
| ALT | Alanine aminotransferase |
| ANN | Artificial neural network |
| A/G | Albumin globule ratio |
| AST | Aspartate aminotransferase |
| AUC | Area under the curve |
| BLR | Basophil to lymphocyte ratio |
| BMI | Body mass index |
| CHB | Chronic hepatitis B |
| EHR | Electronic health-record |
| GGT | γ -Glutamyl transferase |
| HBV | Hepatitis B virus |
| HCC | Hepatocellular carcinoma |
| IDI | Integrated discrimination improvement |
| KNN | K-Nearest neighbour |
| LASSO | Least absolute shrinkage and selection operator |
| LR | Logistic regression |
| Lym | Lymphocyte |
| ML | Machine Learning |
| NE | Neutrophil |
| NLR | Neutrophil - lymphocyte ratio |
| NMLR | Neutrophil - monocyte - lymphocyte ratio |
| NRI | Net reclassification improvement |
| PLR | Platelet - lymphocyte ratio |
| RF | Random forest |
| ROC | Receiver operating characteristic |
| SHAP | SHapley Additive explanation |
| SVM | Support vector machine |
| VIF | Variance inflation factor |

Acknowledgements

The authors would like to thank all researchers for their contributions.

Author contributions

Conception and design: L Wu, T Huang, L Yang; Administrative support: C Fu, Y Lu, T Huang, L Yang; Provision of study materials or patients: L Wu, Z Liu, H Huang, D Pan, C Fu, Y Lu, M Zhou; Data collection and assembly: M Zhou, L Wu, H Huang, K Huang; Data analysis and interpretation: All authors; Drafting of the manuscript: All authors; Final approval of the manuscript: All authors.

Funding

Guangxi Zhuang Autonomous Region Health and Family Planning Commission self-funded research project (No. Z20210032).

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to data protection and ethical restrictions, but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and was reviewed by the Ethics Committee of the Fourth Affiliated Hospital of Guangxi Medical University (approval number: KY2021070). This study was retrospective and all data were anonymised, so informed consent was not required.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Occupational Health and Environmental Health, School of Public Health, Guangxi Medical University, Nanning, Guangxi 530021, China

²Medical Records Data Center, The Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, Guangxi 545000, China

³Medical Department, The Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, Guangxi 545000, China

⁴General Surgery, The Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, Guangxi 545000, China

⁵Department of Epidemiology and Biostatistics, School of Public Health, Guangxi Medical University, Nanning, Guangxi 530021, China

Received: 14 November 2023 / Accepted: 13 February 2025

Published online: 11 March 2025

References

- Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBO-CAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA Cancer J Clin.* 2024;74(3):229–63.
- Zheng RS, Chen R, Han BF, et al. Cancer incidence and mortality in China, 2022[J]. *Zhonghua Zhong Liu Za Zhi.* 2024;46(3):221–31.
- Xie D, Shi J, Zhou J, et al. Clinical practice guidelines and real-life practice in hepatocellular carcinoma: a Chinese perspective[J]. *Clin Mol Hepatol.* 2023;29(2):206–16.
- de Martel C, Maucort-Boulch D, Plummer M, et al. World-wide relative contribution of hepatitis B and C viruses in hepatocellular carcinoma[J]. *Hepatology.* 2015;62(4):1190–200.
- Chen S, Li J, Wang D, et al. The hepatitis B epidemic in China should receive more attention[J]. *Lancet.* 2018;391(10130):1572–1572.
- Su S, Wong WC, Zou Z, et al. Cost-effectiveness of universal screening for chronic hepatitis B virus infection in China: an economic evaluation[J]. *Lancet Global Health.* 2022;10(2):e278–87.
- Nishida N, Ohashi J, Suda G, et al. Prediction model with HLA-A*33:03 reveals number of days to develop Liver Cancer from Blood Test[J]. *Int J Mol Sci.* 2023;24(5):4761.
- Park JW, Chen M, Colombo M, et al. Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE Study[J]. *Liver Int.* 2015;35(9):2155–66.
- Wang C, Li K, Huang Z, et al. Repeat hepatectomy versus percutaneous ablation for recurrent hepatocellular carcinoma: emphasis on the impact of early or late recurrence[J]. *J Cancer Res Clin Oncol.* 2023;149(16):15113–25.
- Lloyd-Jones D. Cardiovascular risk prediction: basic concepts, current status, and future directions[J]. *Circulation.* 2010;121(15):1768–77.
- Khan SS, Coresh J, Pencina MJ, et al. Novel prediction equations for Absolute Risk Assessment of Total Cardiovascular Disease Incorporating Cardiovascular-Kidney-Metabolic Health: a Scientific Statement from the American Heart Association[J]. *Circulation.* 2023;148(24):1982–2004.
- Yu J, Cho S, Jin Y, et al. The best predictive model for hepatocellular carcinoma in patients with chronic hepatitis B infection[J]. *Clin Mol Hepatol.* 2022;28(3):351–61.
- Wong G, Chan H, Wong C, et al. Liver stiffness-based optimization of hepatocellular carcinoma risk score in patients with chronic hepatitis B[J]. *J Hepatol.* 2014;60(2):339–45.

14. Aghamaleki FS, Mollashahi B, Nosrati M, et al. Application of an artificial neural network in the diagnosis of chronic lymphocytic leukemia[J]. *Cureus*. 2019;11(2):e4004.
15. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery[J]. *Lancet Oncol*. 2019;20(5):e262–73.
16. Swanson K, Wu E, Zhang A, et al. From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment[J]. *Cell*. 2023;186(8):1772–91.
17. Wang C, Chang QX, Wang XM, et al. Prostate Cancer risk prediction and online calculation based on machine learning Algorithm[J]. *Chin Med Sci J*. 2022;37(3):210–7.
18. Hou F, Zhu Y, Zhao H, et al. Development and validation of an interpretable machine learning model for predicting the risk of distant metastasis in papillary thyroid cancer: a multicenter study[J]. *EclinicalMedicine*. 2024;77:102913.
19. Kim HY, Lampertico P, Nam JY, et al. An artificial intelligence model to predict hepatocellular carcinoma risk in Korean and caucasian patients with chronic hepatitis B[J]. *J Hepatol*. 2022;76(2):311–8.
20. National Health Commission of the People's Republic of China. Guidelines for diagnosis and treatment of primary liver cancer (2024 edition). *Chin J Hepatol*. 2024;32(7):581–630. (in Chinese).
21. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods[J]. *BMJ*. 2024;385:e078378.
22. Wang K, Tian J, Zheng C, et al. Improving risk identification of adverse outcomes in Chronic Heart failure using SMOTE + ENN and Machine Learning[J]. *Risk Manag Healthc Policy*. 2021;14:2453–63.
23. Angraal S, Mortazavi BJ, Gupta A, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection Fraction[J]. *JACC Heart Fail*. 2020;8(1):12–21.
24. Fujiwara N, Friedman SL, Goossens N, et al. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine[J]. *J Hepatol*. 2018;68(3):526–49.
25. Parikh ND, Singal AG, Hutton DW. Cost effectiveness of regorafenib as second-line therapy for patients with advanced hepatocellular carcinoma[J]. *Cancer*. 2017;123(19):3725–31.
26. Kim HS, Yu X, Kramer J, et al. Comparative performance of risk prediction models for hepatitis B-related hepatocellular carcinoma in the United States[J]. *J Hepatol*. 2022;76(2):294–301.
27. Deo RC. Machine learning in Medicine[J]. *Circulation*. 2015;132(20):1920–30.
28. Zhou LX, Zhou Q, Gao TM, et al. Machine learning predicts acute respiratory failure in pancreatitis patients: a retrospective study[J]. *Int J Med Inf*. 2024;192:105629.
29. Denisko D, Hoffman MM. Classification and interaction in random forests[J]. *Proc Natl Acad Sci U S A*. 2018;115(8):1690–2.
30. Cutler DR, Edwards TC Jr, Beard KH, et al. Random forests for classification in ecology[J]. *Ecology*. 2007;88(11):2783–92.
31. Moehring RW, Phelan M, Lofgren E, et al. Development of a Machine Learning Model Using Electronic Health Record Data to identify antibiotic use among hospitalized Patients[J]. *JAMA Netw Open*. 2021;4(3):e213460.
32. Zhang X, Yue P, Zhang J, et al. A novel machine learning model and a public online prediction platform for prediction of post-ERCP-cholecystitis (PEC)[J]. *EclinicalMedicine*. 2022;48:101431.
33. Wang F, Wang Y, Ji X, et al. Effective Macrosomia Prediction using Random Forest Algorithm[J]. *Int J Environ Res Public Health*. 2022;19(6):3245.
34. Wang H, Zhang C, Li Q, et al. Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches[J]. *BMC Cancer*. 2024;24(1):427.
35. Samban SS, Hari A, Nair B et al. An insight into the role of alpha-fetoprotein (AFP) in the Development and Progression of Hepatocellular Carcinoma[J]. *Mol Biotechnol*. 2023: 1–13.
36. Trevisani F, Garuti F, Neri A. Alpha-fetoprotein for diagnosis, prognosis, and transplant Selection[J]. *Semin Liver Dis*. 2019;39(2):163–77.
37. Liu YR, Lin BB, Zeng DW, et al. Alpha-fetoprotein level as a biomarker of liver fibrosis status: a cross-sectional study of 619 consecutive patients with chronic hepatitis B[J]. *BMC Gastroenterol*. 2014;14:145.
38. Zhao KX, Xu GW, Li N, et al. Establishment and validation of a predictive model for HBV-positive hepatocellular carcinoma[J]. *Chin J Lab Med*. 2022;45(5):516–21.
39. Huo RR, Pan LX, Wu PS et al. Prognostic value of aspartate aminotransferase/alanine aminotransferase ratio in hepatocellular carcinoma after hepatectomy[J]. *BJS Open*. 2024, 8(1).
40. Xu J, Xia Y, Li S, et al. A retrospective pilot study to examine the potential of aspartate aminotransferase to alanine aminotransferase ratio as a predictor of postoperative acute kidney injury in patients with hepatocellular carcinoma[J]. *Ann Clin Biochem*. 2019;56(3):357–66.
41. Eminler AT, Ayyildiz T, Irak K, et al. AST/ALT ratio is not useful in predicting the degree of fibrosis in chronic viral hepatitis patients[J]. *Eur J Gastroenterol Hepatol*. 2015;27(12):1361–6.
42. Caldez MJ, Van Hul N, Koh HWL, et al. Metabolic remodeling during liver Regeneration[J]. *Dev Cell*. 2018;47(4):425–e438425.
43. Anupam S, Goel S, Mehta DK, et al. Comprehending the role of serum GGT in colorectal carcinoma: cancer risk, prognostic and diagnostic significance[J]. *Clin Transl Oncol*; 2024.
44. Mok Y, Son DK, Yun YD, et al. γ -Glutamyltransferase and cancer risk: the Korean cancer prevention study[J]. *Int J Cancer*. 2016;138(2):311–9.
45. Hu G, Tuomilehto J, Pukkala E, et al. Joint effects of coffee consumption and serum gamma-glutamyltransferase on the risk of liver cancer[J]. *Hepatology*. 2008;48(1):129–36.
46. Pastore A, Federici G, Bertini E, et al. Analysis of glutathione: implication in redox and detoxification[J]. *Clin Chim Acta*. 2003;333(1):19–39.
47. Lang D, Ritzberger L, Rambousek V, et al. First-line Pembrolizumab Mono- or combination therapy of Non-small Cell Lung Cancer: baseline metabolic biomarkers predict Outcomes[J]. *Cancers (Basel)*. 2021;13(23):6096.
48. Lee JW, Jeon S, Mun ST, et al. Prognostic value of Fluorine-18 fluorodeoxyglucose uptake of bone marrow on Positron Emission Tomography/Computed Tomography for Prediction of Disease Progression in Cervical Cancer[J]. *Int J Gynecol Cancer*. 2017;27(4):776–83.
49. Liao QQ, Mo YJ, Zhu KW, et al. Platelet-to-lymphocyte ratio (PLR), neutrophil-to-lymphocyte ratio (NLR), monocyte-to-lymphocyte ratio (MLR), and Eosinophil-to-lymphocyte ratio (ELR) as biomarkers in patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease (AECOPD)[J]. *Int J Chron Obstruct Pulmon Dis*. 2024;19:501–18.
50. Ding R, Chen Z, He M et al. Application Value of Combined Detection of NLR, PNI, D-Dimer, CD3 + T Lymphocytes, and CEA in Colorectal Cancer Screening[J]. *Disease Markers*, 2022, 2022.
51. Falanga A, Marchetti M, Vignoli A. Coagulation and cancer: biological and clinical aspects[J]. *J Thromb Haemost*. 2013;11(2):223–33.
52. Ojima Y, Harano M, Sumitani D, et al. Prognostic value of preoperative carcinoembryonic antigen and D-dimer concentrations in patients undergoing curative resection for colorectal cancer[J]. *Surg Today*. 2021;51:1108–17.
53. Yao M, Zhao J, Lu F. Alpha-fetoprotein still is a valuable diagnostic and prognosis predicting biomarker in hepatitis B virus infection-related hepatocellular carcinoma[J]. *Oncotarget*. 2016;7(4):3702.
54. Liu L, Wang Q, Zhao X, et al. Establishment and validation of nomogram model for the diagnosis of AFP-negative hepatocellular carcinoma[J]. *Front Oncol*. 2023;13:1131892.
55. Hang D, Yang X, Lu J, et al. Untargeted plasma metabolomics for risk prediction of hepatocellular carcinoma: a prospective study in two Chinese cohorts[J]. *Int J Cancer*. 2022;151(12):2144–54.
56. Khare S, Khare T, Ramanathan R, et al. Hepatocellular Carcinoma: the role of MicroRNAs[J]. *Biomolecules*. 2022;12(5):645.

Publisher's note

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