Elevate Series

ANNALS OF MEDICINE
2025, VOL. 57, NO. 1, 2477294
https://doi.org/10.1080/07853890.2025.2477294

HEPATOLOGY

OPEN ACCESS Check for updates

Identifying liver cirrhosis in patients with chronic hepatitis B: an interpretable machine learning algorithm based on LSM

Xueting Bai[a‡], Chunwen Pu[b‡], Wenchong Zhen[a], Yushuang Huang[b], Qian Zhang[a], Zihan Li[a], Yixin Zhang[a], Rongxuan Xu[a], Zhihan Yao[a], Wei Wu[a], Mei Sun[b] and Xiaofeng Li[a]

[a]Department of Epidemiology and Health Statistics, Dalian Medical University, Dalian, China; [b]Dalian Public Health Clinical Center, Dalian, Liaoning province, China

ABSTRACT

Background: Chronic hepatitis B (CHB) is a common cause of liver cirrhosis (LC), a condition associated with an unfavourable prognosis. Therefore, timely diagnosis of LC in CHB patients is crucial.

Objective: This study aimed to enhance the diagnostic accuracy of LC in CHB patients by integrating liver stiffness measurement (LSM) with traditional indicators.

Methods: The study participants were randomly divided into training and internal validation sets. Employing the least absolute shrinkage and selection operator (LASSO) and random forest-recursive feature elimination (RF-RFE) for feature selection, we developed both traditional logistic regression and five machine learning models (k-nearest neighbors, random forest (RF), artificial neural network, support vector machine and eXtreme Gradient Boosting). Performance evaluation included receiver operating characteristic curves, calibration curves and decision curve analysis. Shapley additive explanations (SHAP) was employed to improve the interpretability of the optimal model.

Results: We retrospectively included 1609 patients with CHB, among whom 470 were diagnosed with cirrhosis. Cirrhosis was diagnosed based on histological confirmation or clinical assessment, supported by characteristic findings on abdominal ultrasound and corroborative evidence such as thrombocytopenia, varices or imaging from CT/MRI. In the internal validation, the RF model achieved an accuracy above 0.80 and an AUC above 0.80, with outstanding calibration ability and clinical net benefit. Additionally, the model exhibited excellent predictive performance in an independent external validation set. The SHAP analysis indicated that LSM contributed the most to the model. The model still showed strong discriminative power when using only LSM or traditional indicators alone.

Conclusions: Machine learning models, especially the RF model, can effectively identify LC in CHB patients. Integrating LSM with traditional indicators can enhance diagnostic performance.

KEY MESSAGES

- Liver cirrhosis (LC) is a common complication of chronic hepatitis B (CHB).
- The random forest (RF) model showed the best overall performance to identify LC in CHB patients in our study, which could assist in the clinical decision-making procedure.
- Integrating LSM with traditional indicators can enhance the diagnostic performance of LC in CHB patients. In the absence of LSM, other traditional indicators can also diagnose LC effectively.

ARTICLE HISTORY
Received 8 August 2024
Revised 17 January 2025
Accepted 13 February 2025

KEYWORDS
Chronic hepatitis B; liver cirrhosis; machine learning; liver stiffness measurement; diagnostic model

1. Introduction

Liver cirrhosis (LC) is a pathological condition marked by the gradual hardening of liver tissue, a consequence of prolonged inflammation [1]. Chronic hepatitis B (CHB) virus infection is recognized as a key contributor to the development of LC. Despite advances in CHB treatment, the prognosis for patients with severe cirrhosis remains a significant concern [2]. Cirrhosis is often incidentally diagnosed, typically at a stage when liver function has already suffered significant impairment. Fortunately, LC is a reversible condition [3,4]. Even patients in the decompensated stage can potentially achieve stabilization or even reversal through timely

CONTACT Mei Sun sunmei75@163.com Dalian Public Health Clinical Center, Dalian, Liaoning Province, China; Xiaofeng Li lxf_chen@dmu.edu.cn Department of Epidemiology and Health Statistics, Dalian Medical University, No. 9, West Section Lvshun Road, Dalian, Liaoning Province, China
‡Both authors contributed equally to this work.

management of underlying risk factors [5]. Undeniably, timely prevention and treatment of hepatitis B virus (HBV)-related cirrhosis are crucial.

Although liver biopsy is considered the gold standard for diagnosing liver fibrosis and cirrhosis in chronic liver disease, it is plagued by limitations such as sampling errors, high costs, patient discomfort and invasiveness [6]. These challenges have spurred the need for non-invasive serological diagnostic alternatives in clinical practice. Non-invasive methods like the γ-glutamyl transferase-to-platelet ratio (GPR), aspartate aminotransferase to platelet ratio (APRI) and fibrosis-4 (FIB-4) indices offer advantages in terms of simplicity, cost-effectiveness and reproducibility. However, these methods often produce conflicting results with relatively higher false negative and false positive rates [7–9].

Liver stiffness measurement (LSM) by transient elastography (TE) has emerged as a highly precise and non-invasive technique, particularly effective in identifying advanced fibrosis and cirrhosis. TE can swiftly evaluate liver tissue properties without causing discomfort or complications to patients, playing a crucial role in monitoring fibrosis progression [10]. LSM can serve as a valuable complementary tool for diagnosing cirrhosis in patients with CHB [11]. Nonetheless, it is essential to underscore that the LSM threshold for diagnosing cirrhosis should be adjusted based on whether the CHB patient's bilirubin or alanine aminotransferase levels are within the normal range [12]. Relying solely on a single threshold for diagnosing cirrhosis in CHB patients may lead to misdiagnosis. Combining LSM with other traditional indicators can provide a more comprehensive evaluation of cirrhosis, thereby enhancing diagnostic accuracy. However, the substantial expense associated with acquiring and maintaining FibroScan devices hinders their availability in low- and middle-income countries, thereby posing challenges in implementing LSM as a standard approach for cirrhosis surveillance in resource-limited settings or primary healthcare facilities [13]. Therefore, comparing the diagnostic performance of models with and without LSM can help determine the substitutability of traditional indicators when LSM is not available.

In recent years, there has been a notable increase in research on liver fibrosis among patients with chronic liver disease. However, these studies predominantly rely on traditional logistic regression (LR), with relatively few applications of machine learning techniques [14,15]. While machine learning has demonstrated tremendous potential in the medical field, its 'black box' nature still poses challenges for interpretation [16]. In this study, we construct multiple machine learning models that integrate LSM with traditional indicators and compare them with traditional LR model. After selecting the optimal model, we employ Shapley additive explanations (SHAP) to enhance the model's transparency and interpretability. Additionally, we delve into the clinical application value of combining LSM with traditional indicators, using LSM alone, and using traditional indicators alone based on the optimal model. Furthermore, we compare the diagnostic performance of LSM combined with traditional indicators to conventional serological markers. These investigations aim to provide scientific support to help healthcare professionals optimize monitoring and management strategies for CHB patients.

2. Materials and methods

2.1. Study participants

In the initial cohort, 5639 CHB patients treated at Dalian Public Health Clinical Center from October 2015 to April 2024 were included in the analysis. Patients who met the following inclusion criteria were included: (1) patients with a clear diagnosis of CHB; (2) patients who have had LSM; (3) patients who have a definitive histopathological examination or clinical diagnosis indicating the presence or absence of abnormalities in liver morphology and structure, including LC. The following exclusion criteria were used: (1) patients diagnosed with liver cancer or other malignancies; (2) patients with viral co-infections and systemic diseases impacting the liver (such as HIV infection, autoimmune liver disease, etc.); (3) patients diagnosed with fatty liver; (4) patients with missing data.

Based on the inclusion and exclusion criteria, 1609 CHB patients who underwent LSM were ultimately enrolled in the study. One thousand one hundred and twenty-two CHB patients from October 2015 to December 2021 were included in the training and internal validation sets, among whom 348 were diagnosed with LC. Data from CHB patients between January 2022 and April 2024 were used as an independent external validation set. The variables we extracted were all information that can be obtained from the hospital information system. Data extraction was performed independently by one of the authors and verified for accuracy by another author. The entire study process is shown in Figure 1.

2.2. Diagnostic criteria

The definition of CHB conforms to the 2022 guidelines for the prevention and treatment of CHB [17]. Chronic HBV infection is defined as the presence of HBsAg and/

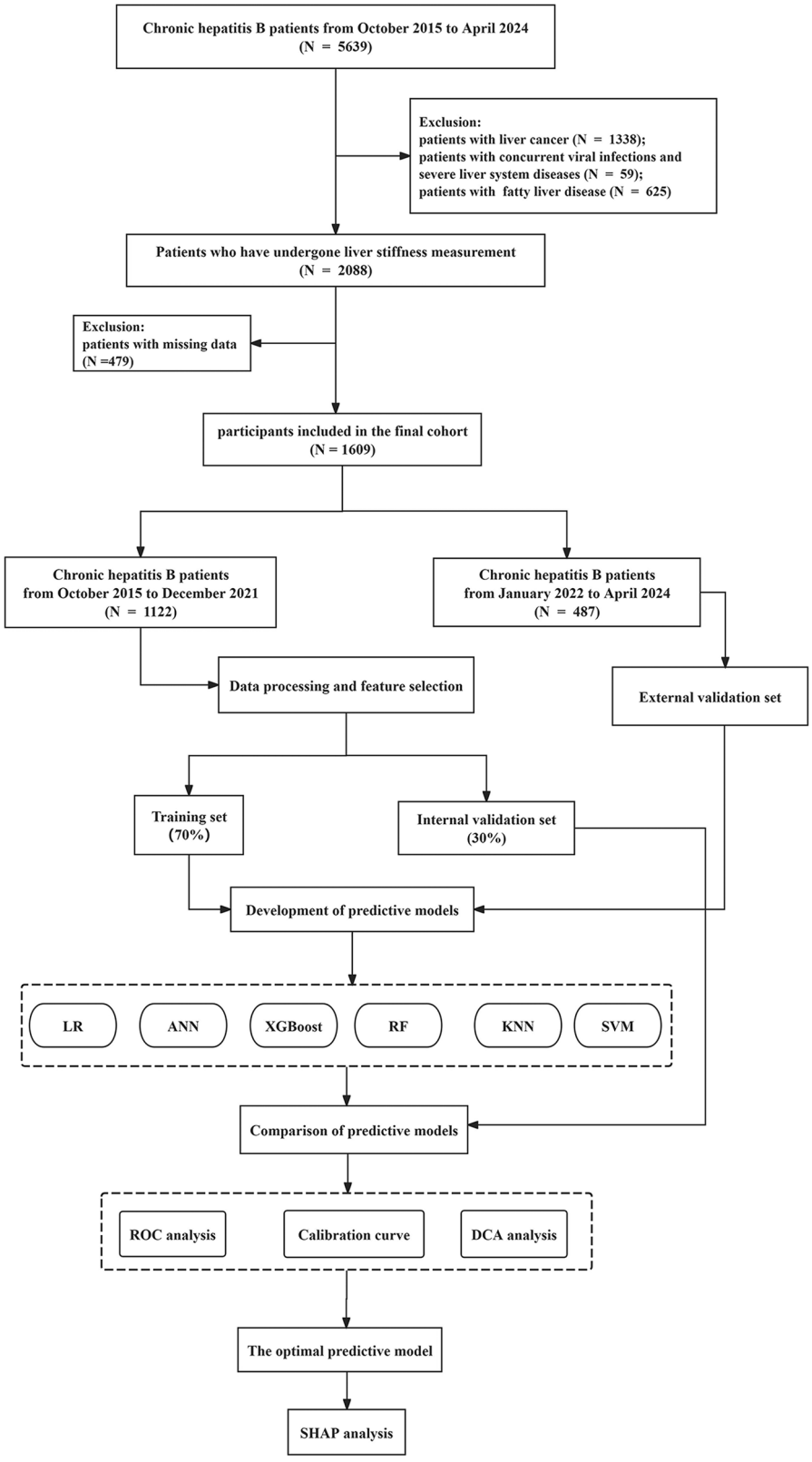

Figure 1. Flowchart. LR: logistic regression; XGBoost: eXtreme Gradient Boosting; ANN: artificial neural network; RF: random forest; KNN: k-nearest neighbors; SVM: support vector machine; ROC: receiver operating characteristic; DCA: decision curve analysis; SHAP: Shapley additive explanations.

or HBV DNA positive for more than 6 months. The presence of HBsAg for at least 6 months establishes the chronicity of infection. Chronic hepatitis B is defined as a chronic inflammatory liver disease caused by persistent HBV infection. These definitions are also consistent with the AASLD 2018 Hepatitis B Guidance [18].

The diagnosis of HBV-related cirrhosis should meet the following criteria: (1) the patient is currently HBsAg positive, or HBsAg negative and anti-HBc positive with a clear history of chronic HBV infection (with a history of being HBsAg positive for >6 months), with other aetiologies being ruled out. (2) The diagnosis of cirrhosis is established through either histological confirmation or clinical diagnosis, supported by repeated and consistent findings from abdominal ultrasound indicative of cirrhosis, along with corroborative evidence such as thrombocytopenia, oesophageal/gastric varices or additional imaging findings from CT or MRI suggestive of cirrhosis [19,20]. In Chinese patients, liver tissue pathology diagnosis is less common, with diagnosis mainly relying on clinical and imaging examinations (abdominal ultrasound, CT or MRI). Clinical manifestations, including splenomegaly, portal vein dilation, ascites and other signs, also play a significant role in the clinical diagnosis of cirrhosis [21].

2.3. Measurement of liver stiffness

LSM was performed using the FibroScan device, which assesses liver health using ultrasound technology [22]. Patients underwent the examination while fasting, lying on a bed as the physician positioned the device beneath the ribcage to emit sound waves for assessment. Following the examination, FibroScan produced results based on the collected data. Depending on the patient's physique, the device offered different models – the M type for most patients and the XL type for obese patients. All LSM values were obtained by experienced operators following the manufacturer's protocol. The LSM values were measured in kPa. To ensure reliable LSM values, the following conditions must be met: at least 10 valid measurements, a success rate exceeding 60%, and an interquartile range to median ratio not exceeding 30%.

2.4. Routine available serum algorithms

FIB-4, APRI and GPR are commonly used serological test indices. The specific calculation formulas are as follows:

$$\mathrm{FIB-4} = \left\{\mathrm{Age(years)} \times \mathrm{AST(U/L)}\right\} \Big/ \left\{\mathrm{PLT}\left(10^9/\mathrm{L}\right) \times \mathrm{ALT}^{1/2}\left(\mathrm{U/L}\right)\right\} \quad (1)$$

$$\mathrm{APRI} = \left\{\mathrm{AST}\left(\mathrm{U/L}\right)/\mathrm{ULN}\right\} / \left\{\mathrm{PLT}\left(10^9/\mathrm{L}\right) \times 100\right\} \quad (2)$$

$$\mathrm{GPR} = \left\{\mathrm{GGT}\left(\mathrm{U/L}\right)/\mathrm{ULN}\right\} / \left\{\mathrm{PLT}\left(10^9/\mathrm{L}\right) \times 100\right\} \quad (3)$$

The ULN for AST, ALT and GGT is defined as 40 U/L, 40 U/L and 60 U/L, respectively. The full names of AST, ALT, GGT, PLT and ULN are aspartate aminotransferase, alanine aminotransferase, γ-glutamyl transferase, platelet count and upper limit of normal, respectively.

2.5. Statistical analysis

2.5.1. Data processing

Statistical analysis was conducted using SPSS26.0 (SPSS Inc., Chicago, IL) and R4.3.3 (R Foundation for Statistical Computing, Vienna, Austria). The Kolmogorov–Smirnov test was used to check if the variables followed a normal distribution. Continuous variables were expressed as mean with standard deviation (SD), while categorical variables were presented as numbers (percentages). Student's *t*-test or Mann–Whitney's *U*-test was employed to compare continuous variables, and Pearson's Chi-squared test or Fisher's exact test was used to compare categorical variables. All tests were two-tailed, with a significance level set at $p < .05$. Variables with missing data exceeding 20% were excluded from the analysis, while those with missing data below 20% were imputed using the missForest method [23]. Multicollinearity among variables was assessed using the variance inflation factor (VIF). A VIF greater than 5 indicated the presence of multicollinearity, while a VIF greater than 10 indicated severe multicollinearity [24].

2.5.2. Feature selection

In this study, feature selection was performed by integrating random forest-recursive feature elimination (RF-RFE) with the least absolute shrinkage and selection operator (LASSO). In LASSO using the 'glmnet' package in R for feature selection, optimal regularization parameters are chosen via 10-fold cross-validation. This process entails randomly dividing the data into 10 subsets, where each subset serves as the validation set once while the remaining subsets form the training set in each iteration. Various regularization parameters (λ) are tested during training, and model performance is assessed on the validation set. The λ yielding the best performance across all folds is selected as the final parameter for the LASSO. Conversely, RF-RFE achieves a similar objective using the 'caret' package, systematically eliminating features by configuring recursive feature elimination control parameters. This method employs accuracy as the evaluation metric and continues iterating until it reaches a predefined number of features or achieves satisfactory performance.

2.5.3. Development and evaluation of predictive models

The study subjects were randomly allocated to training and internal validation sets at a 7:3 ratio. Machine learning models including LR, artificial neural network (ANN), support vector machine (SVM), random forest (RF), k-nearest neighbors (KNN) and eXtreme Gradient Boosting (XGBoost) were built. Ten-fold cross-validation and grid search techniques were used to optimize the model's parameters. After multiple iterations, refined parameters were identified as the optimal configuration for the current model. Receiver operating characteristic (ROC) curves were used to evaluate the models' diagnostic accuracy and discriminative power. The DeLong test was used to compare AUC values. Calibration curves and decision curve analysis were performed to assess the models' predictive capability and clinical applicability. The performance metrics for model evaluation included accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and *F*1-score. Additionally, SHAP was used to further reveal the impact and contributions of feature variables [25].

3. Results

3.1. Comparison of demographic and clinical characteristics between the LC and non-LC group

This study enrolled 1122 CHB patients for model training and internal validation. Among these patients, 348 were diagnosed with LC. The characteristics of patients with and without LC are presented in Table 1. The mean age of the patients was 46.54 (11.83) years, with males comprising 60.8%. The study revealed statistical differences between the two groups across various parameters

Table 1. Comparison of baseline characteristics between the LC and non-LC groups in the training and internal validation sets.

Characteristic	Overall	LC	Non-LC	*p*
N	1122	348	774	
Age, years, mean (SD)	46.54 (11.83)	51.88 (10.68)	44.14 (11.54)	<.001
Gender (male), n (%)	682 (60.8)	243 (69.8)	439 (56.7)	<.001
LSM, kPa, mean (SD)	10.20 (10.21)	16.36 (13.79)	7.42 (6.40)	<.001
CAP, dB/m, mean (SD)	215.19 (56.80)	212.21 (60.58)	216.53 (55.00)	.239
Platelet, 10^9/L, mean (SD)	174.42 (62.65)	126.21 (55.43)	196.09 (52.87)	<.001
Leukocyte, 10^9/L, mean (SD)	5.15 (1.60)	4.54 (1.73)	5.42 (1.45)	<.001
Erythrocyte, 10^{12}/L, mean (SD)	4.70 (0.58)	4.54 (0.70)	4.77 (0.50)	<.001
Haemoglobin, g/L, mean (SD)	145.29 (18.90)	140.89 (23.05)	147.28 (16.34)	<.001
NLR, 10^9/L, mean (SD)	1.91 (1.72)	2.31 (2.81)	1.73 (0.82)	<.001
PDW, fL, mean (SD)	15.47 (2.32)	15.90 (2.28)	15.28 (2.31)	<.001
LMR, 10^9/L, mean (SD)	4.78 (2.51)	4.37 (2.36)	4.97 (2.55)	<.001
Albumin, g/L, mean (SD)	44.08 (4.46)	42.09 (5.48)	44.97 (3.57)	<.001
AST/ALT, U/L, mean (SD)	0.97 (0.48)	1.06 (0.48)	0.93 (0.47)	<.001
ALP, U/L, mean (SD)	74.33 (73.59)	78.18 (31.98)	72.59 (85.93)	.239
LDH, U/L, mean (SD)	200.73 (57.51)	210.34 (69.33)	196.40 (50.78)	<.001
Total bilirubin, μmol/L, mean (SD)	15.91 (15.13)	18.57 (16.39)	14.72 (14.37)	<.001
Urea nitrogen, mmol/L, mean (SD)	5.40 (8.87)	5.39 (2.30)	5.40 (10.56)	.985
Prealbumin, mg/L, mean (SD)	209.59 (73.09)	183.24 (77.28)	221.44 (67.93)	<.001
Creatinine, μmol/L, mean (SD)	71.76 (60.31)	71.85 (54.41)	71.72 (62.81)	.975
Direct bilirubin, μmol/L, mean (SD)	7.01 (11.66)	8.53 (12.12)	6.32 (11.39)	.003
GGT, U/L, mean (SD)	52.89 (75.73)	60.02 (67.62)	49.68 (78.94)	.034
Total protein, g/L, mean (SD)	71.72 (6.23)	69.41 (7.52)	72.76 (5.24)	<.001
GFR, mL/min/1.73 m^2, mean (SD)	103.96 (16.30)	100.66 (16.87)	105.45 (15.82)	<.001
Glucose, mmol/L, mean (SD)	5.39 (1.49)	5.66 (1.96)	5.26 (1.20)	<.001
Cholesterol, mmol/L, mean (SD)	4.33 (0.90)	4.08 (0.92)	4.44 (0.87)	<.001
HDL-C, mmol/L, mean (SD)	1.28 (0.35)	1.23 (0.35)	1.30 (0.34)	.001
LDL-C, mmol/L, mean (SD)	2.78 (0.83)	2.60 (0.86)	2.86 (0.81)	<.001
Triglyceride, mmol/L, mean (SD)	1.23 (0.64)	1.12 (0.58)	1.27 (0.66)	<.001
Apolipoprotein A, g/L, mean (SD)	1.19 (0.26)	1.11 (0.24)	1.23 (0.26)	<.001
Apolipoprotein B, g/L, mean (SD)	0.86 (0.23)	0.80 (0.22)	0.89 (0.23)	<.001
Uric acid, μmol/L, mean (SD)	314.02 (76.81)	318.87 (90.18)	311.84 (69.93)	.156
Globulin, g/L, mean (SD)	27.62 (4.37)	27.42 (5.14)	27.72 (3.97)	.288
Bile acids, μmol/L, mean (SD)	11.89 (27.87)	20.34 (39.15)	8.09 (19.79)	<.001
Creatine kinase, U/L, mean (SD)	106.90 (173.94)	111.59 (155.95)	104.79 (181.51)	.545
cholinesterase, U/L, mean (SD)	8065.82 (2546.69)	6915.54 (2370.44)	8582.99 (2453.48)	<.001
HBsAg ≥250 IU/mL (yes), *n* (%)	974 (86.8)	273 (78.4)	701 (90.6)	<.001
HBV DNA ≥10^4 IU/mL (yes), *n* (%)	381 (34.0)	93 (26.7)	288 (37.2)	.001
FIB-4	2.00 (2.22)	3.34 (3.23)	1.39 (1.14)	<.001
APRI	8.40×10^{-5} (1.60×10^{-4})	1.13×10^{-4} (1.95×10^{-4})	7.10×10^{-5} (1.40×10^{-4})	<.001
GPR	6.56×10^{-5} (1.05×10^{-4})	1.01×10^{-4} (1.26×10^{-4})	4.98×10^{-5} (9.01×10^{-5})	<.001

LC: liver cirrhosis; LSM: liver stiffness measurement; CAP: controlled attenuation parameter; NLR: neutrophil to lymphocyte ratio; PDW: platelet distribution width; LMR: lymphocyte to monocyte ratio; AST: aspartate aminotransferase; ALT: alanine aminotransferase; ALP: alkaline phosphatase; LDH: lactate dehydrogenase; GGT: gamma-glutamyl transferase; GFR: glomerular filtration rate; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; HBsAg: hepatitis B surface antigen; FIB-4: fibrosis-4 index; APRI: aspartate aminotransferase to platelet ratio index; GPR: the γ-glutamyl transferase-to-platelet ratio.

Data were expressed as frequency (proportion) or mean (SD). The *p* value indicates statistical significance.

including age, gender, LSM, platelet, leukocyte, erythrocyte, haemoglobin, NLR, PDW, LMR, albumin, AST/ALT, LDH, total bilirubin, prealbumin, direct bilirubin, GGT, total protein, GFR, glucose, cholesterol, HDL-C, LDL-C, triglyceride, apolipoprotein A, apolipoprotein B, bile acids, cholinesterase, HBsAg, HBV DNA, FIB-4, APRI and GPR ($p < .05$). Compared to CHB patients without LC, those with LC were older ($p < .001$), had higher LSM ($p < .001$) and had lower platelet counts ($p < .001$). The detailed baseline characteristics of the external validation set, which includes 365 CHB patients without cirrhosis and 122 with cirrhosis, are presented in Table 2.

3.2. Feature selection

The LASSO algorithm constrains model complexity through regularization, enabling feature selection during the regression model fitting process [26]. The R package performs cross-validation over a specified range of λ values, yielding two key parameters: lambda_min and lambda_lse. Lambda_min is the λ value that minimizes the cross-validation error, while lambda_lse is the largest λ value within one standard error above the minimum cross-validation error. Opting for lambda_min offers the advantage of achieving the best model fit, thereby attaining superior predictive performance. Similarly, the RF-RFE method performs feature selection by employing a strategy based on a RF model, iteratively training the model to assess the importance of each feature and recursively eliminating irrelevant ones [27]. As shown in Figure 2, we ultimately selected the common feature variables of LASSO and RF-RFE algorithms to build the predictive model. These variables, listed in sequence, are LSM, platelet, age, leukocyte, glucose, cholinesterase, AST/ALT, apolipoprotein A, apolipoprotein B, HDL-C, GGT, PDW, triglyceride, uric acid, NLR, prealbumin, haemoglobin and erythrocyte.

Table 2. Comparison of baseline characteristics between the LC and non-LC groups in the external validation set.

Characteristic	Overall	LC	Non-LC	*p*
N	487	122	365	
Age, years, mean (SD)	48.27 (12.53)	55.91 (10.61)	45.72 (12.09)	<.001
Gender (male), *n* (%)	267 (54.8)	71 (58.2)	196 (53.7)	.448
LSM, kPa, mean (SD)	8.35 (7.74)	15.30 (11.61)	6.03 (3.69)	<.001
CAP, dB/m, mean (SD)	214.52 (51.45)	201.33 (54.20)	218.92 (49.80)	.001
Platelet, 10^9/L, mean (SD)	179.87 (67.54)	123.65 (62.40)	198.67 (58.17)	<.001
Leukocyte, 10^9/L, mean (SD)	5.32 (1.67)	4.37 (1.82)	5.64 (1.49)	<.001
Erythrocyte, 10^{12}/L, mean (SD)	4.68 (0.56)	4.42 (0.72)	4.77 (0.47)	<.001
Haemoglobin, g/L, mean (SD)	144.29 (19.07)	136.46 (25.04)	146.91 (15.81)	<.001
NLR, 10^9/L, mean (SD)	1.85 (1.00)	2.34 (1.49)	1.68 (0.71)	<.001
PDW, fL, mean (SD)	16.40 (1.82)	16.55 (2.09)	16.35 (1.71)	.301
LMR, 10^9/L, mean (SD)	4.67 (1.80)	3.83 (1.85)	4.95 (1.70)	<.001
Albumin, g/L, mean (SD)	43.99 (4.96)	39.73 (6.73)	45.42 (3.11)	<.001
AST/ALT, U/L, mean (SD)	1.12 (0.73)	1.29 (0.97)	1.07 (0.62)	.003
ALP, U/L, mean (SD)	66.34 (27.99)	76.07 (35.59)	63.08 (24.14)	<.001
LDH, U/L, mean (SD)	193.53 (47.69)	200.91 (43.87)	191.06 (48.71)	.048
Total bilirubin, μmol/L, mean (SD)	20.47 (17.64)	28.86 (29.11)	17.67 (10.11)	<.001
Urea nitrogen, mmol/L, mean (SD)	4.87 (1.62)	5.51 (1.93)	4.65 (1.44)	<.001
Prealbumin, mg/L, mean (SD)	226.05 (70.30)	172.46 (68.95)	243.97 (61.11)	<.001
Creatinine, μmol/L, mean (SD)	65.47 (16.43)	64.67 (17.48)	65.74 (16.08)	.534
Direct bilirubin, μmol/L, mean (SD)	7.11 (10.07)	11.43 (16.25)	5.67 (6.27)	<.001
GGT, U/L, mean (SD)	34.65 (52.32)	44.59 (61.64)	31.32 (48.45)	.015
Total protein, g/L, mean (SD)	72.40 (6.34)	67.67 (8.37)	73.98 (4.51)	<.001
GFR, mL/min/1.73 m^2, mean (SD)	104.04 (15.66)	99.53 (15.30)	105.55 (15.50)	<.001
Glucose, mmol/L, mean (SD)	5.55 (1.36)	5.89 (2.06)	5.44 (1.00)	.001
Cholesterol, mmol/L, mean (SD)	4.47 (1.03)	4.03 (1.15)	4.61 (0.95)	<.001
HDL-C, mmol/L, mean (SD)	1.20 (0.31)	1.12 (0.32)	1.23 (0.30)	.001
LDL-C, mmol/L, mean (SD)	2.65 (0.88)	2.30 (0.97)	2.77 (0.81)	<.001
Triglyceride, mmol/L, mean (SD)	1.11 (0.60)	0.99 (0.55)	1.15 (0.62)	.009
Apolipoprotein A, g/L, mean (SD)	1.40 (0.30)	1.28 (0.30)	1.44 (0.28)	<.001
Apolipoprotein B, g/L, mean (SD)	0.88 (0.23)	0.77 (0.24)	0.91 (0.22)	<.001
Uric acid, μmol/L, mean (SD)	308.18 (80.33)	307.93 (80.10)	308.26 (80.51)	.968
Globulin, g/L, mean (SD)	28.37 (4.12)	27.89 (4.81)	28.53 (3.85)	.136
Bile acids, μmol/L, mean (SD)	11.81 (33.94)	28.67 (59.54)	6.18 (15.25)	<.001
Creatine kinase, U/L, mean (SD)	109.00 (89.03)	121.06 (148.10)	104.96 (56.79)	.084
cholinesterase, U/L, mean (SD)	8099.04 (1993.02)	6683.56 (2236.25)	8572.16 (1656.83)	<.001
HBsAg ≥250 IU/mL, (Yes), *n* (%)	443 (91.0)	105 (86.1)	338 (92.6)	.046
HBV DNA ≥10^4 IU/mL (yes), *n* (%)	137 (28.1)	23 (18.9)	114 (31.2)	.012
FIB-4	2.25 (3.65)	4.88 (6.35)	1.37 (1.14)	<.001
APRI	7.58×10^{-5} (1.98×10^{-4})	1.41×10^{-4} (2.90×10^{-4})	5.39×10^{-5} (1.51×10^{-4})	.002
GPR	4.18×10^{-5} (7.30×10^{-5})	7.46×10^{-5} (8.09×10^{-5})	3.09×10^{-5} (6.68×10^{-5})	<.001

LC: liver cirrhosis; LSM: liver stiffness measurement; CAP: controlled attenuation parameter; NLR: neutrophil to lymphocyte ratio; PDW: platelet distribution width; LMR: lymphocyte to monocyte ratio; AST: aspartate aminotransferase; ALT: alanine aminotransferase; ALP: alkaline phosphatase; LDH: lactate dehydrogenase; GGT: gamma-glutamyl transferase; GFR: glomerular filtration rate; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; HBsAg: hepatitis B surface antigen; FIB-4: fibrosis-4 index; APRI: aspartate aminotransferase to platelet ratio index; GPR: the γ-glutamyl transferase-to-platelet ratio.

Data were expressed as frequency (proportion) or mean (SD). The *p* value indicates statistical significance.

3.3. Model performance and comparison

Based on the feature selection results, we constructed a traditional LR model and five machine learning models including ANN, RF, SVM, KNN and XGBoost. The ROC curves of these models are illustrated in Figure 3. In the training set, RF (AUC = 0.982) and XGBoost (AUC = 0.954) models exhibited superior discriminative ability, followed by SVM (AUC = 0.912), KNN (AUC = 0.898), LR (AUC = 0.875) and ANN (AUC = 0.867). In the internal validation set, XGBoost (AUC = 0.891) and RF (AUC = 0.889) models exhibited superior discriminative ability, followed by SVM (AUC = 0.872), LR (AUC = 0.861), ANN (AUC = 0.860) and KNN (AUC = 0.842). In the external validation set, the RF model had the highest AUC value (AUC = 0.890), followed by XGBoost (AUC = 0.889), ANN (AUC = 0.881), SVM (AUC = 0.872), LR (AUC = 0.866) and KNN (AUC = 0.853). A detailed comparison of specific performance metrics for each model is presented in Table 3. In the training set, the RF model ranks first in accuracy, sensitivity, specificity, PPV, NPV and *F*1-score. In the internal validation set, the RF model ranks first in accuracy, specificity, PPV and *F*1-score, while it ranks third in sensitivity and NPV. In the external validation set, the RF model also shows strong performance across all metrics. The confusion matrices for each model in the training and validation sets are provided in Figure 4.

We plotted calibration curves and DCA curves based on the training and validation sets. The former is utilized to assess the accuracy and reliability of the model predictions, while the latter is employed to evaluate the potential clinical utility of the model across different threshold ranges. The calibration curves for the six models are depicted in Figure 5. Except for the ANN model, the other models demonstrate good calibration performance. The DCA curves, as illustrated in Figure 6, show that the models achieve higher net benefits than the 'all-intervention' or 'no-intervention' strategies within a wide range of thresholds. Across different threshold probability ranges, the RF and XGBoost models demonstrate more pronounced clinical benefits compared to other models. Combining the results of the model performance evaluation, the RF model exhibits superior diagnostic capability and clinical utility, making it the optimal model for diagnosing LC in CHB patients.

3.4. Diagnostic efficacy comparison of different indicator combinations based on the RF model

3.4.1. LSM and 17 traditional indicators vs. serum biomarkers

The RF model emerged as the superior choice after evaluating the modelling performance of various machine learning methods. Building on this, we developed four RF models to evaluate their diagnostic efficacy for LC. These models included a comprehensive model combining LSM with 17 traditional indicators, as well as three models based on APRI, GPR and FIB-4 serological biomarkers, respectively. The results (Figure 7) showed that the comprehensive model combining LSM and 17 traditional indicators had the best diagnostic performance. Among the serological indicator models, the FIB-4 model performed the best, followed by GPR, while APRI had relatively lower diagnostic efficacy.

3.4.2. LSM and 17 traditional indicators vs. LSM-only vs. 17 traditional indicators

After determining the optimal RF model, we conducted modelling for three scenarios: (1) LSM alone, (2) 17 traditional indicators alone and (3) LSM combined with traditional indicators. The results of these models are presented in Figure 8. The incorporation of traditional indicators significantly enhanced model performance, with the combined model achieving the highest AUC compared to the LSM-only approach. Furthermore, in the absence of LSM, the diagnostic performance of the traditional 17-indicator model was comparable to that of the combined LSM and traditional indicators model. This indicates that traditional indicators remain highly effective for diagnosing LC when LSM is unavailable, outperforming the LSM-only model.

3.4.3. FIB-4 vs. LSM-only vs. 17 traditional indicators

To further evaluate the performance of FIB-4, we subsequently compared it with models using only LSM and only the 17 traditional indicators separately. In the training set, the FIB-4 model demonstrated superior performance compared to the LSM-only model. In the internal validation set, the LSM-only model exhibited comparable performance to the FIB-4 model. However, the LSM-only model surpassed the FIB-4 model in the external validation set. Notably, the model integrating 17 traditional indicators consistently outperformed the FIB-4 model in all sets (Figure 9). The evaluation of performance metrics for diverse indicator combinations is presented in Table 4.

3.5. Model interpretation

To gain a more intuitive understanding of the relationship between the model and feature factors, SHAP was employed to interpret the RF model. Figure 10(A)

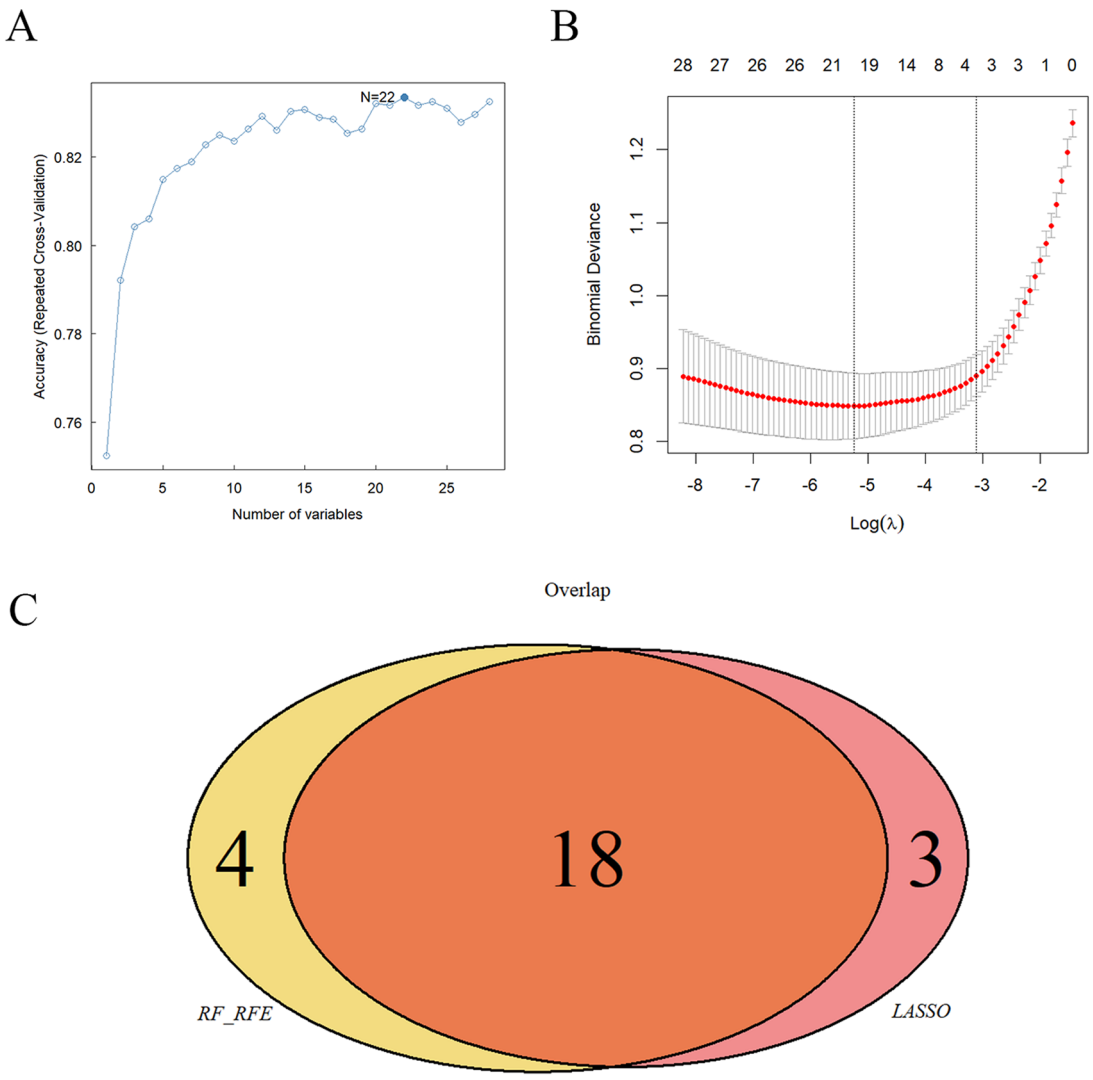

Figure 2. Screening of characteristic factors. (A) Feature variables screening based on RF-RFE. (B) Feature variables screening based on LASSO. (C) LASSO combined RF-RFE. LASSO: least absolute shrinkage and selection operator; RF: random forest; RFE: recursive feature elimination.

presents SHAP values for all features based on the RF model. Feature importance is ranked from top to bottom, with LSM scoring the highest importance, followed by platelet, age, leukocyte, glucose, cholinesterase, AST/ALT, apolipoprotein A, apolipoprotein B, HDL-C, GGT, PDW, triglyceride, uric acid, NLR, prealbumin, haemoglobin and erythrocyte. Figure 10(B) displays the distribution of SHAP values for each feature across different data points or samples. If a specific instance has relatively high variable values, it appears as a yellow point, whereas relatively low variable values appear as purple points. SHAP values illustrate the contribution of each feature to the target variable, whether positive or negative. Additionally, two individual SHAP force plots are provided in the study. Figure 10(C) illustrates the SHAP force plot for diagnosing LC in patients with CHB, showing that all features support the diagnosis of LC in this patient. The longer the yellow arrow of a certain feature, the greater its contribution to supporting the diagnosis of LC in CHB patients. Figure 10(D) displays the SHAP force plot for excluding LC in patients with CHB. Glucose supports the diagnosis of LC, the other 17 features do not support the diagnosis. Yellow arrows support the diagnosis of LC in CHB patients; purple arrows do not support the diagnosis of LC in CHB patients.

4. Discussion

Cirrhosis ranks among the top ten causes of mortality worldwide, placing a significant burden on public health. In China, HBV infection remains the primary cause of cirrhosis [28]. The 2030 Sustainable Development Agenda has set clear goals for

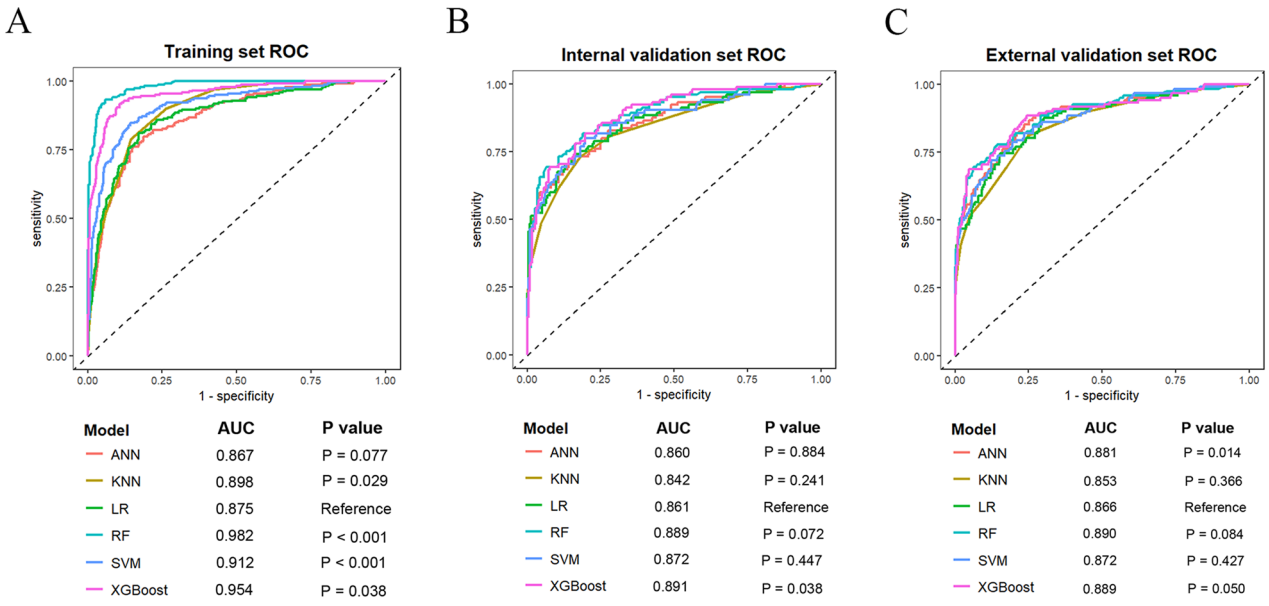

Figure 3. ROC curves for the prediction models. (A) ROC curve in the training set. (B) ROC curve in the internal validation set. (C) ROC curve in the external validation set. ROC: receiver operating characteristic; AUC: area under the curve; ANN: artificial neural network; KNN: k-nearest neighbors; LR: logistic regression; RF: random Forest; SVM: support vector machine; XGBoost: eXtreme Gradient Boosting.

Table 3. Performance parameters of the six prediction models.

Models	Accuracy	Sensitivity	Specificity	PPV	NPV	*F*1-score
Training set						
LR	0.823	0.811	0.828	0.679	0.907	0.739
XGBoost	0.901	0.918	0.893	0.794	0.960	0.851
RF	0.939	0.934	0.941	0.876	0.970	0.904
SVM	0.855	0.848	0.858	0.728	0.926	0.783
ANN	0.809	0.798	0.813	0.658	0.900	0.721
KNN	0.835	0.790	0.856	0.711	0.901	0.749
Internal validation set						
LR	0.787	0.771	0.794	0.628	0.885	0.692
XGBoost	0.805	0.790	0.811	0.654	0.896	0.716
RF	0.817	0.781	0.833	0.678	0.894	0.726
SVM	0.790	0.800	0.785	0.627	0.897	0.703
ANN	0.775	0.743	0.790	0.614	0.872	0.672
KNN	0.793	0.733	0.820	0.647	0.872	0.688
External validation set						
LR	0.774	0.770	0.775	0.534	0.910	0.631
XGBoost	0.813	0.770	0.827	0.599	0.915	0.674
RF	0.840	0.754	0.868	0.657	0.914	0.702
SVM	0.797	0.770	0.805	0.570	0.913	0.655
ANN	0.799	0.787	0.803	0.571	0.918	0.662
KNN	0.793	0.697	0.825	0.570	0.891	0.627

LR: logistic regression; ANN: artificial neural network; SVM: support vector machine; RF: random forest; KNN: k-nearest neighbors; XGBoost: eXtreme Gradient Boosting; PPV: positive predictive value; NPV: negative predictive value.

eradicating viral hepatitis. To achieve these objectives, the World Health Organization has devised strategies [29]. However, many CHB patients still lack adequate attention, which increases the risk of developing LC. Therefore, it is necessary to improve the diagnostic accuracy for cirrhosis among individuals with CHB.

LSM is a direct method for assessing liver stiffness, proving effective in diagnosing liver fibrosis and cirrhosis in chronic liver diseases. However, the accuracy of LSM can be affected by several factors, including obesity, respiratory status and the skill level of the operator [30]. These factors may introduce biases in LSM measurements, thereby compromising the reliability of the diagnosis. Research has demonstrated that both the stiffness of the surrounding liver tissue and the degree of liver inflammation significantly influence the diagnostic accuracy of LSM [31]. Traditional clinical indicators such as ALT, AST and platelet count provide insights into liver function and inflammatory status; however, their diagnostic efficacy is limited when used in

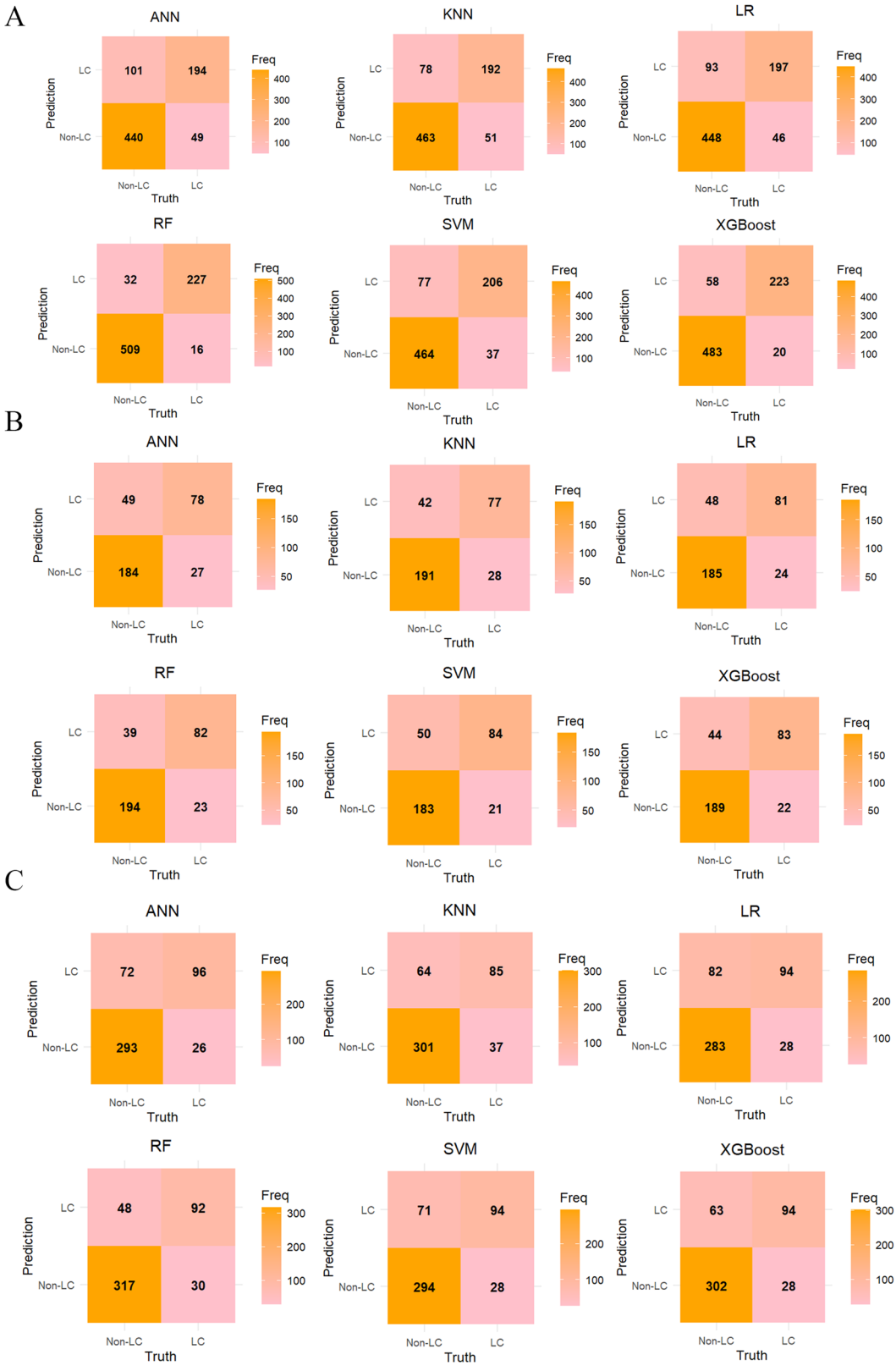

Figure 4. Confusion matrices of six models in the training and validation sets. (A) Confusion matrices in the training set. (B) Confusion matrices in the internal validation set. (C) Confusion matrices in the external validation set. LR: logistic regression; ANN: artificial neural network; SVM: support vector machine; RF: random Forest; KNN: k-nearest neighbors; XGBoost: eXtreme Gradient Boosting; LC: liver cirrhosis.

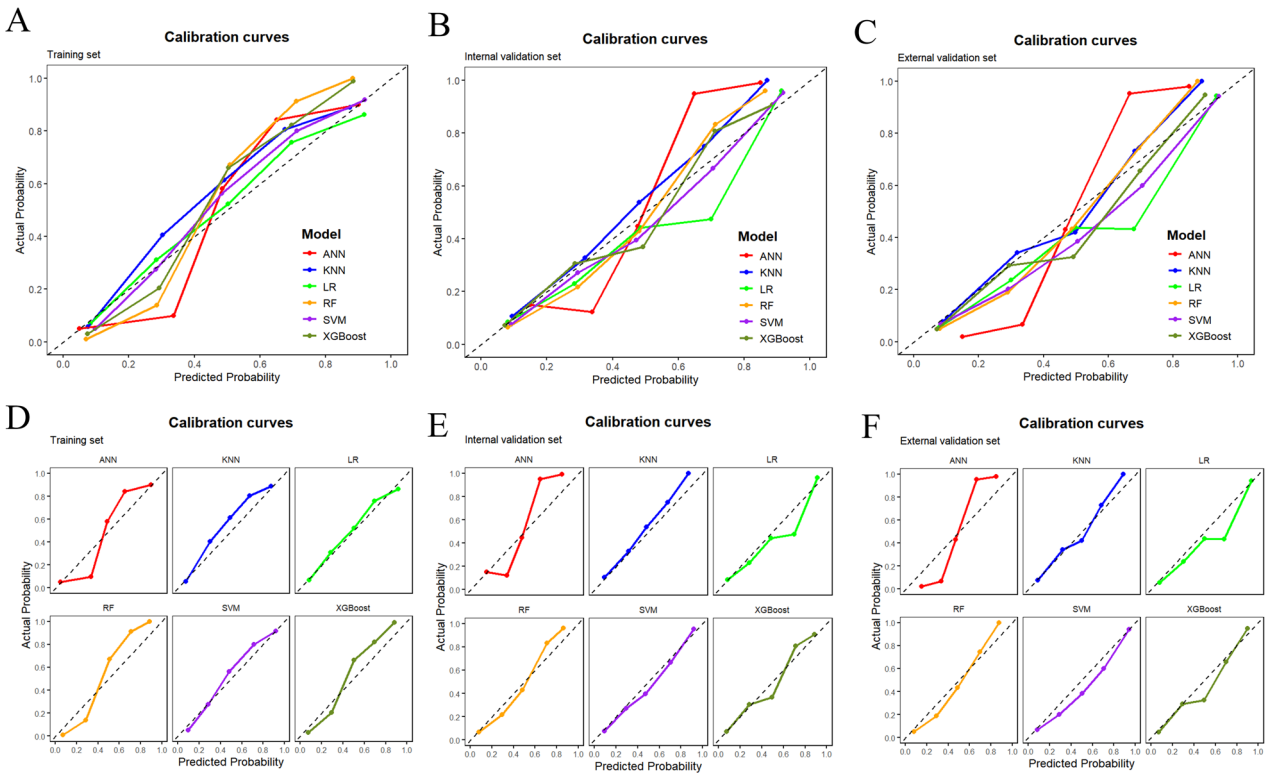

Figure 5. Calibrate curves for the prediction models. (A) Comprehensive summary figure of six models in the training set. (B) Comprehensive summary figure of six models in the internal validation set. (C) Comprehensive summary figure of six models in the external validation set. (D) Stratified plots of six models in the training set. (E) Stratified plots of six models in the internal validation set. (F) Stratified plots of six models in the external validation set. ANN: artificial neural network; KNN: k-nearest neighbors; LR: logistic regression; RF: random Forest; SVM: support vector machine; XGBoost: eXtreme Gradient Boosting.

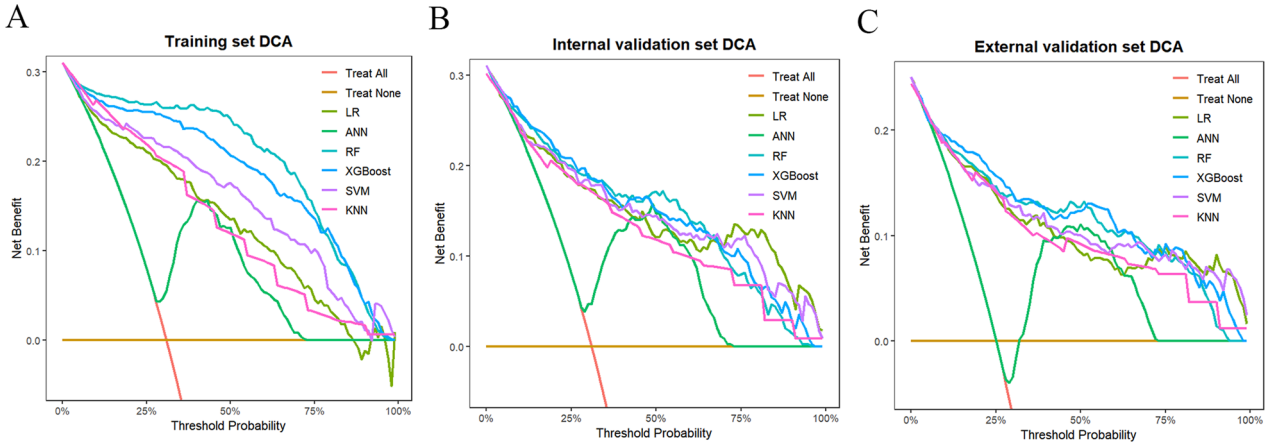

Figure 6. DCA curves for the prediction models. (A) DCA curve in the training set. (B) DCA curve in the internal validation set. (C) DCA curve in the external validation set. The treat all curve represents the benefit rates for all cases with intervention, while the treat none curve represents the benefit rates for all cases without intervention. The remaining curves denote various models. The threshold probability represents the probability cut-off used to make a decision, while the net benefit indicates the clinical utility gained from using the model compared to alternative strategies. ANN: artificial neural network; KNN: k-nearest neighbors; LR: logistic regression; RF: random forest; SVM: support vector machine; XGBoost: eXtreme Gradient Boosting.

isolation. Therefore, a comprehensive assessment using various indicators is a more reliable approach. Many studies have shown that combining LSM with various clinical indicators and scores significantly improves the diagnostic accuracy for assessing liver fibrosis and predicting related complications in patients with chronic liver diseases. Fan et al. [32] conducted a precise evaluation of liver fibrosis in patients with CHB by combining

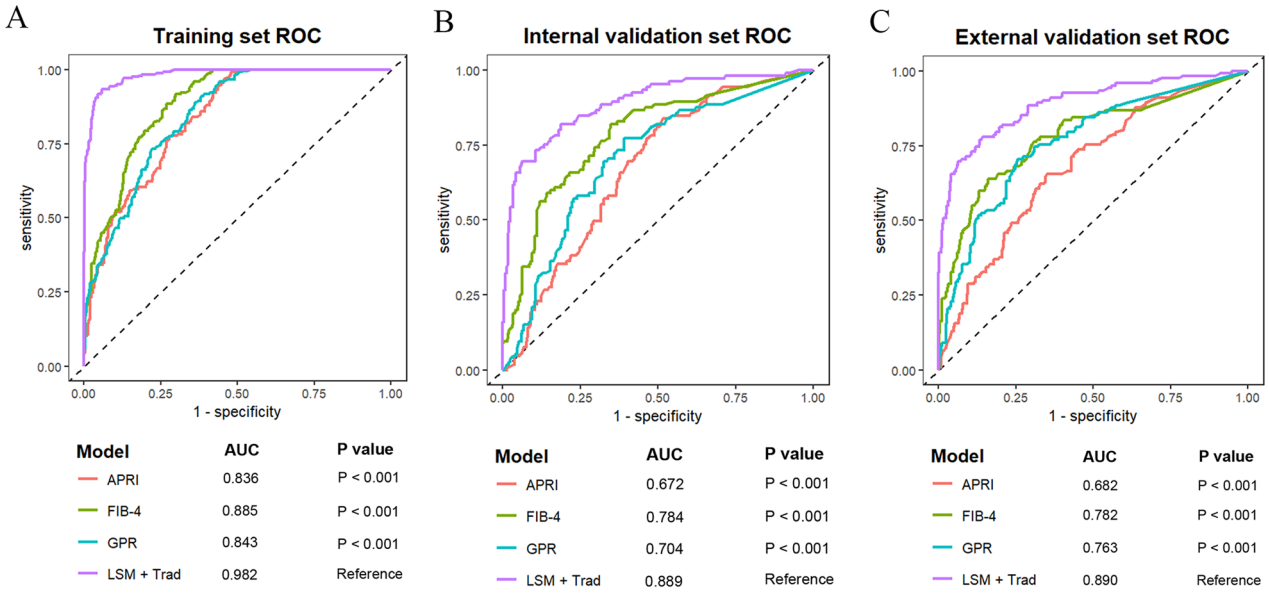

Figure 7. ROC curves for the RF model using LSM with traditional indicators, FIB-4, APRI and GPR. (A) ROC curve in the training set. (B) ROC curve in the internal validation set. (C) ROC curve in the external validation set. FIB-4: fibrosis-4 index; APRI: aspartate aminotransferase to platelet ratio index; GPR: the γ-glutamyl transferase-to-platelet ratio; LSM: liver stiffness measurement; Trad: 17 traditional indicators; ROC: receiver operating characteristic; AUC: area under the curve; RF: random forest.

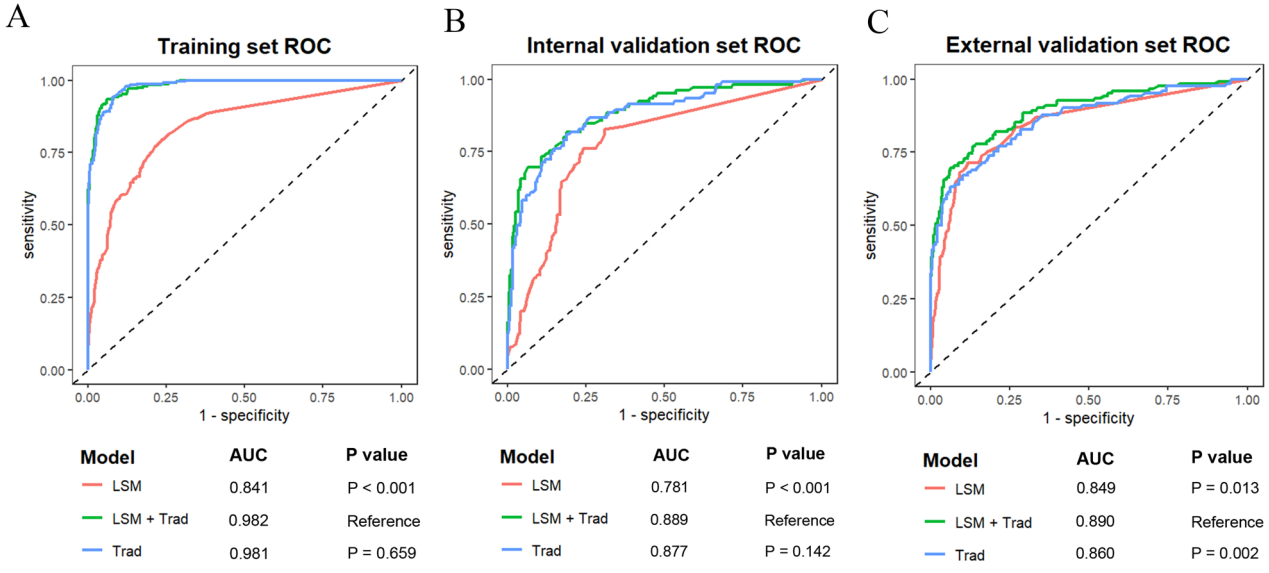

Figure 8. ROC curves for the RF model based on the traditional indicators, LSM and traditional indicators, and LSM-only. (A) ROC curve in the training set. (B) ROC curve in the internal validation set. (C) ROC curve in the external validation set. LSM: liver stiffness measurement; Trad: 17 traditional indicators; ROC: receiver operating characteristic; AUC: area under the curve; RF: random forest.

LSM with aMAP scores, which reflect the potential risk of hepatocellular carcinoma development. The joint utilization of aMAP and LSM has shown strong diagnostic efficacy in identifying liver fibrosis among CHB patients. Additionally, Fan et al. [33] developed a machine learning model that integrates clinical indicators with LSM to identify fibrosis associated with metabolic dysfunction-associated steatotic liver disease (MASLD, formerly known as NAFLD). Sanyal et al. [34] demonstrated that combining LSM with clinical biomarkers significantly enhances the accuracy of identifying atrial fibrillation or cirrhosis in patients with non-alcoholic fatty liver disease (NAFLD). However, when using only LSM or FIB-4, predictive accuracy may slightly decrease. Some studies have combined elements of FIB-4 with LSM to predict incident complications of portal

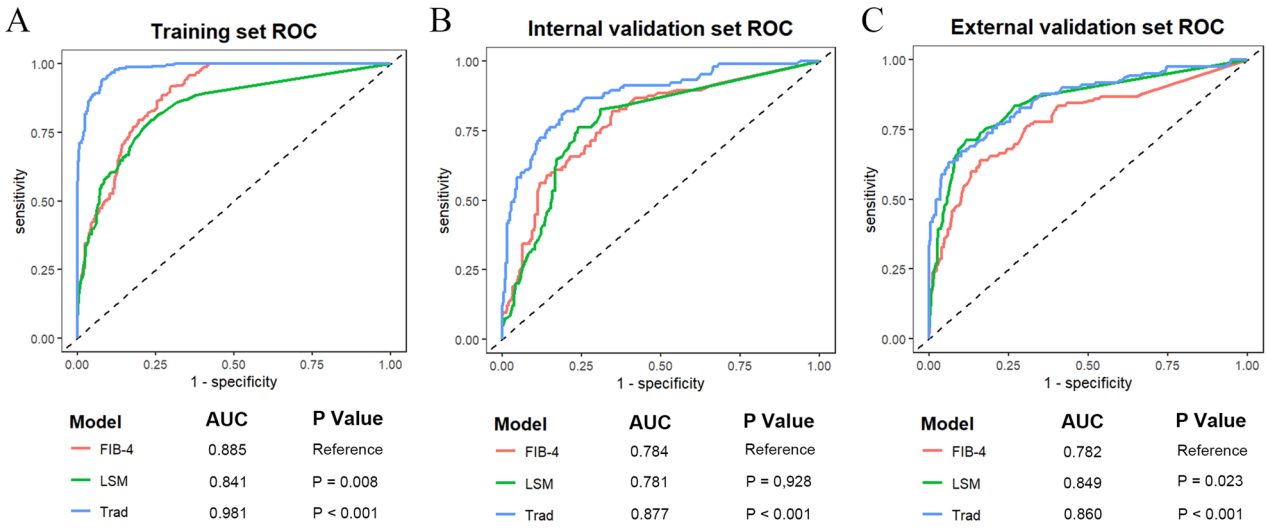

Figure 9. ROC curves for the RF model based on FIB-4, LSM-only and 17 traditional indicators. (A) ROC curve in the training set. (B) ROC curve in the internal validation set. (C) ROC curve in the external validation set. FIB-4: fibrosis-4 index; LSM: liver stiffness measurement; Trad: 17 traditional indicators; ROC: receiver operating characteristic; AUC: area under the curve; RF: random Forest.

Table 4. Evaluation of performance metrics for diverse indicator combinations.

Models	Accuracy	Sensitivity	Specificity	PPV	NPV	*F*1-score
Training set						
LSM + Trad (17)	0.939	0.934	0.941	0.876	0.970	0.904
LSM-only	0.774	0.794	0.765	0.603	0.892	0.686
Trad (17)	0.927	0.942	0.921	0.842	0.973	0.889
FIB-4	0.768	0.918	0.701	0.579	0.950	0.710
APRI	0.667	0.992	0.521	0.482	0.993	0.649
GPR	0.716	0.889	0.638	0.524	0.927	0.660
Internal validation set						
LSM + Trad (17)	0.817	0.781	0.833	0.678	0.894	0.726
LSM-only	0.760	0.762	0.760	0.588	0.876	0.664
Trad (17)	0.805	0.819	0.798	0.647	0.907	0.723
FIB-4	0.692	0.829	0.631	0.503	0.891	0.626
APRI	0.577	0.848	0.455	0.412	0.869	0.555
GPR	0.648	0.771	0.592	0.460	0.852	0.577
External validation set						
LSM + Trad (17)	0.840	0.754	0.868	0.657	0.914	0.702
LSM-only	0.807	0.746	0.827	0.591	0.907	0.659
Trad (17)	0.768	0.918	0.701	0.579	0.950	0.710
FIB-4	0.661	0.803	0.614	0.410	0.903	0.543
APRI	0.546	0.754	0.477	0.325	0.853	0.454
GPR	0.723	0.713	0.726	0.465	0.883	0.563

PPV: positive predictive value; NPV: negative predictive value; FIB-4: fibrosis-4 index; APRI: aspartate aminotransferase to platelet ratio index; GPR: the γ-glutamyl transferase-to-platelet ratio; LSM: liver stiffness measurement; Trad: 17 traditional indicators.

hypertension (PH) in individuals with compensated liver disease [35]. This approach has also demonstrated good predictive ability. In this study, we integrated LSM with 17 traditional indicators to identify LC in CHB patients. The results demonstrated that each model achieved an AUC–ROC value of over 0.80. In comparison, the diagnostic performance of LSM or traditional indicators alone was comparatively weaker. Notably, even in the absence of LSM, the selected 17 traditional clinical indicators were still effective in identifying LC in CHB patients. This finding has substantial clinical significance, especially in resource-constrained or equipment-limited environments, where clinicians can utilize these traditional indicators as a dependable approach for initial evaluation.

In the field of non-invasive LC diagnosis, serum-based indices such as FIB-4, APRI and GPR are widely used in patients with various types of hepatitis due to their simplicity and low cost. While these methods have proven useful in clinical practice, they have notable limitations in accuracy and sensitivity. Therefore, we aimed to develop a more accurate and sensitive diagnostic model by integrating LSM with 17 traditional indicators. Previous studies have robustly confirmed the high diagnostic value of FIB-4 in identifying cirrhosis [36]. Similarly, research has shown that GPR outperforms FIB-4 and APRI, particularly in staging liver fibrosis in CHB patients [7]. These findings form a crucial basis for assessing the diagnostic efficacy of integrating LSM with multiple indicators and traditional serum markers like FIB-4, APRI and GPR. Among the three serum-based indices, FIB-4 exhibited the best diagnostic performance for LC, followed by GPR, while APRI showed relatively weaker efficacy. To further evaluate the diagnostic performance of FIB-4, we compared it with LSM alone and the 17 traditional indicators. The results indicated that in the internal validation, the AUROC of FIB-4 was comparable to that of LSM alone; however, in the external validation, LSM alone significantly outperformed FIB-4. This finding aligns with previous studies, further underscoring the unique value of LSM in diagnosing LC. Notably, the diagnostic performance of the 17

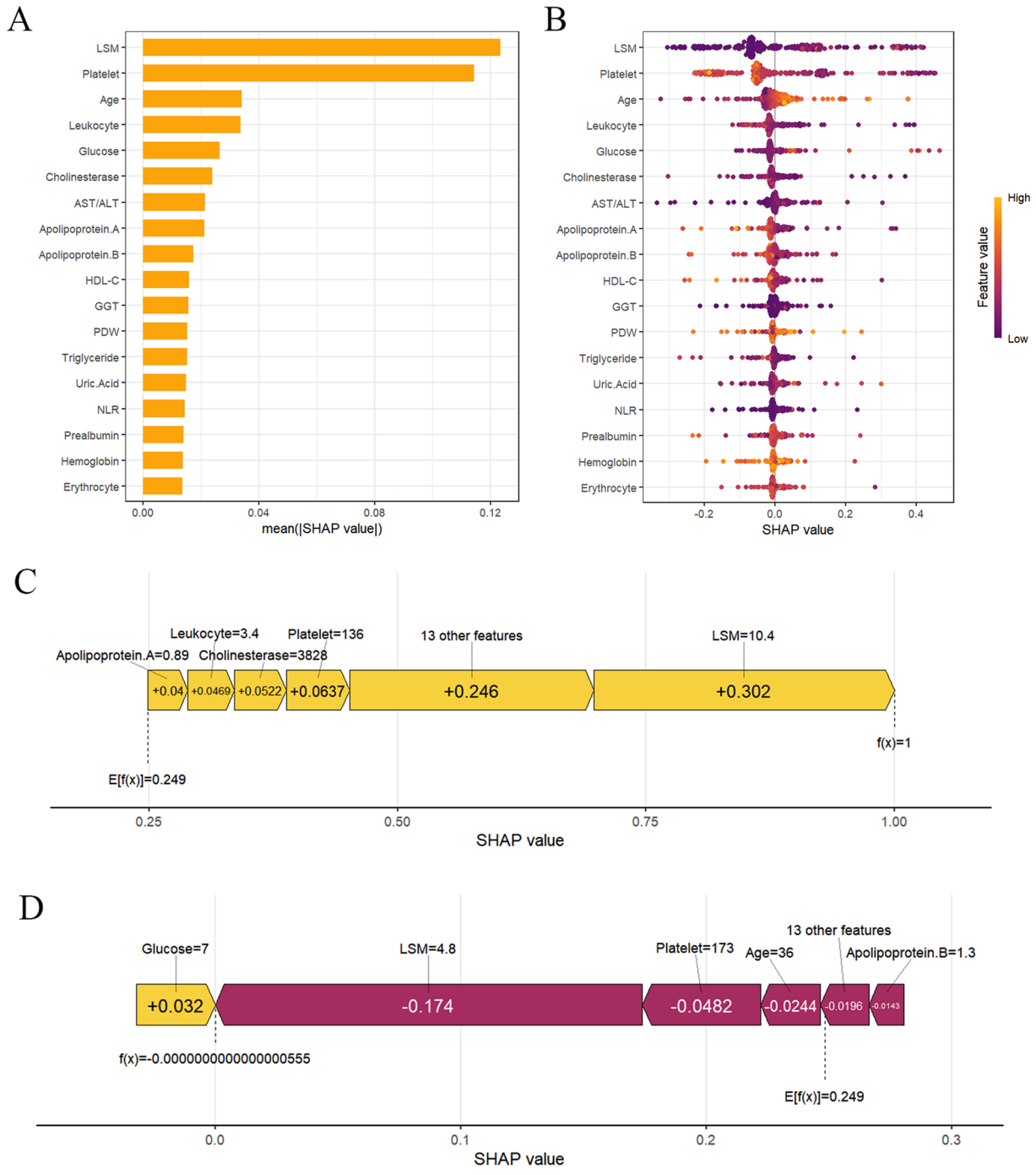

Figure 10. SHAP analysis based on the RF model. (A) Ranking of variable importance based on the mean SHAP value. (B) In the SHAP bee swarm plot, each row represents a feature, the *x*-axis represents the SHAP value, and each data point represents a sample. (C) SHAP analysis of liver cirrhosis in patients with chronic hepatitis B. (D) SHAP force plot of non-cirrhosis chronic hepatitis B patient.

traditional indicators yielded an AUROC significantly higher than that of LSM alone or FIB-4, demonstrating that the integration of multiple indicators can substantially enhance the accuracy of LC diagnosis.

In recent years, machine learning has been widely applied in the field of chronic liver disease. However, due to the 'black-box' nature of machine learning, its interpretability is relatively poor, making it difficult to explain why specific predictions are made for patients [37]. In this study, we utilized SHAP analysis to provide a detailed interpretation of the optimal model. SHAP analysis enables us to assess the contribution of different variables to the model's predictive outcomes. The results revealed that the top three most important variables are LSM, platelet and age. As CHB progresses, liver fibrosis gradually increases, which leads to the

progressive hardening of the liver. Higher LSM is associated with an increased risk of cirrhosis [38]. Patients with LC often have PH and hypersplenism, which increases platelet sequestration and destruction in the spleen, leading to thrombocytopenia. If a CHB patient exhibits abnormalities in platelet count or changes in platelet function, LC should be considered as a potential diagnosis [39]. Age is a major risk factor for chronic liver disease. Advanced liver disease is more common in older adults. Even in young patients after liver injury, the compensatory mechanism of liver cell activation may be impaired, leading to the development of serious liver diseases as age increases [40]. Additionally, changes in indicators such as leukocyte, glucose, cholinesterase, AST/ALT, apolipoprotein A, apolipoprotein B, HDL-C, GGT, PDW, triglyceride, uric acid, NLR, prealbumin, haemoglobin and erythrocyte can also help in identifying cirrhosis in CHB patients. Therefore, CHB patients should pay particular attention to fluctuations in these indicators.

Early diagnosis and intervention are crucial for CHB patients. CHB is a progressive liver disease, and once it progresses to cirrhosis, reversing the condition is challenging. The proposed model can enhance the accuracy of diagnosing cirrhosis in CHB patients, aiding clinicians in making informed decisions. The model, which is based on traditional indicators, demonstrates considerable effectiveness even in resource-limited environments where LSM is not accessible. However, this study also has its limitations. This study is a single-centre study conducted exclusively at a hospital in Dalian, which introduces potential selection bias. Therefore, external validation from other centres is necessary to improve the model's generalizability. Furthermore, the current model cannot capture the dynamic changes in key indicators, which may restrict its application in the dynamic monitoring and management of CHB patients. Future research should incorporate longitudinal data, explore the performance of new models at different time points, and analyse the impact of key indicators' changes over time on outcomes. Additionally, longitudinal data will facilitate the evaluation of new models' dynamic adjustment capabilities, potentially further enhancing their accuracy and practicality, especially in the long-term management of CHB patients.

5. Conclusions

This study utilized machine learning methods to select the optimal model from six candidates, the RF model, which integrates LSM with 17 conventional indicators. This approach significantly improved the diagnostic accuracy of LC in CHB patients. The integration of LSM and the 17 traditional indicators in the RF model demonstrated superior diagnostic performance compared to traditional serological markers such as FIB-4, APRI and GPR, as well as outperforming the use of LSM or the 17 traditional indicators alone. Even in the absence of LSM, the selected 17 traditional clinical indicators effectively identified LC in CHB patients, highlighting their potential utility in resource-limited settings. Despite these significant findings, future research should focus on validating this model in multicentre environments to enhance its generalizability. Additionally, incorporating longitudinal data could further explore the new models' application in dynamically assessing disease progression in CHB patients, thereby providing more comprehensive support for clinical decision-making. In summary, the combination of LSM and traditional indicators offers an efficient and reliable tool for diagnosing LC, holding substantial clinical value.

Acknowledgements

Appreciation is extended to the Dalian Public Health Clinical Center for supporting this study.

Author contributions

Conceptualization, data curation, methodology, software, writing – original draft: Xueting Bai; conceptualization, supervision, writing – review and editing: Chunwen Pu; data curation, writing – review and editing: Wenchong Zhen; software, writing – review and editing: Yushuang Huang, Qian Zhang, Zihan Li and Yixin Zhang; methodology, writing – review and editing: Rongxuan Xu, Zhihan Yao and Wei Wu; supervision, writing – review and editing: Mei Sun and Xiaofeng Li. All authors had read and approved the final version of the manuscript.

Ethical approval

The study is a retrospective analysis based on historical data, and all data involved were anonymized. This study was approved by the Ethics Committee of Dalian Public Health Clinical Center (2024-026KY-001). All patients provided written informed consent. All procedures conducted in this study adhered to the Declaration of Helsinki.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Data availability statement

The datasets used and analysed in the study are available from the corresponding author on reasonable request.

References

[1] Ginès P, Krag A, Abraldes JG, et al. Liver cirrhosis. Lancet. 2021;398(10308):1359–1376. doi: 10.1016/S0140-6736(21)01374-X.

[2] Shih C, Yang CC, Choijilsuren G, et al. Hepatitis B virus. Trends Microbiol. 2018;26(4):386–387. doi: 10.1016/j.tim.2018.01.009.

[3] Kisseleva T, Brenner D. Molecular and cellular mechanisms of liver fibrosis and its regression. Nat Rev Gastroenterol Hepatol. 2021;18(3):151–166. doi: 10.1038/s41575-020-00372-7.

[4] Jung YK, Yim HJ. Reversal of liver cirrhosis: current evidence and expectations. Korean J Intern Med. 2017;32(2):213–228. doi: 10.3904/kjim.2016.268.

[5] He ZY, Wang BQ, You H. Reversal of cirrhotic decompensation: re-compensation. Zhonghua Gan Zang Bing Za Zhi. 2019;27(12):915–918. doi: 10.3760/cma.j.issn.1007-3418.2019.12.002.

[6] Chowdhury AB, Mehta KJ. Liver biopsy for assessment of chronic liver diseases: a synopsis. Clin Exp Med. 2023;23(2):273–285. doi: 10.1007/s10238-022-00799-z.

[7] Hu YC, Liu H, Liu XY, et al. Value of gamma-glutamyltranspeptidase-to-platelet ratio in diagnosis of hepatic fibrosis in patients with chronic hepatitis B. World J Gastroenterol. 2017;23(41):7425–7432. doi: 10.3748/wjg.v23.i41.7425.

[8] Itakura J, Kurosaki M, Setoyama H, et al. Applicability of APRI and FIB-4 as a transition indicator of liver fibrosis in patients with chronic viral hepatitis. J Gastroenterol. 2021;56(5):470–478. doi: 10.1007/s00535-021-01782-3.

[9] Huang R, Wang G, Tian C, et al. Gamma-glutamyltranspeptidase to platelet ratio is not superior to APRI, FIB-4 and RPR for diagnosing liver fibrosis in CHB patients in China. Sci Rep. 2017;7(1):8543. doi: 10.1038/s41598-017-09234-w.

[10] Jung KS, Kim SU. Clinical applications of transient elastography. Clin Mol Hepatol. 2012;18(2):163–173. doi: 10.3350/cmh.2012.18.2.163.

[11] Kostallari E, Wei B, Sicard D, et al. Stiffness is associated with hepatic stellate cell heterogeneity during liver fibrosis. Am J Physiol Gastrointest Liver Physiol. 2022;322(2):G234–G246. doi: 10.1152/ajpgi.00254.2021.

[12] Consensus on clinical application of transient elastography detecting liver fibrosis: a 2018 update. Zhonghua Gan Zang Bing Za Zhi. 2019;27(3):182–191.

[13] Wilson ML, Fleming KA, Kuti MA, et al. Access to pathology and laboratory medicine services: a crucial gap. Lancet. 2018;391(10133):1927–1938. doi: 10.1016/S0140-6736(18)30458-6.

[14] Wei R, Wang J, Wang X, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. EBioMedicine. 2018;35:124–132. doi: 10.1016/j.ebiom.2018.07.041.

[15] Saboo K, Petrakov NV, Shamsaddini A, et al. Stool microbiota are superior to saliva in distinguishing cirrhosis and hepatic encephalopathy using machine learning. J Hepatol. 2022;76(3):600–607. doi: 10.1016/j.jhep.2021.11.011.

[16] Musolf AM, Holzinger ER, Malley JD, et al. What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. Hum Genet. 2022;141(9):1515–1528. doi: 10.1007/s00439-021-02402-z.

[17] Guidelines for the prevention and treatment of chronic hepatitis B (version 2022). Zhonghua Gan Zang Bing Za Zhi. 2022;30(12):1309–1331.

[18] Terrault NA, Lok ASF, McMahon BJ, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. Hepatology. 2018;67(4):1560–1599. doi: 10.1002/hep.29800.

[19] Chinese guidelines on the management of liver cirrhosis. Zhonghua Gan Zang Bing Za Zhi. 2019;27(11):846–865.

[20] Jeng WJ, Chien RN, Chen YC, et al. Hepatocellular carcinoma reduced, HBsAg loss increased, and survival improved after finite therapy in hepatitis B patients with cirrhosis. Hepatology. 2024;79(3):690–703. doi: 10.1097/HEP.0000000000000575.

[21] Smith A, Baumgartner K, Bositis C. Cirrhosis: diagnosis and management. Am Fam Physician. 2019;100(12):759–770.

[22] Wilder J, Patel K. The clinical utility of FibroScan(®) as a noninvasive diagnostic test for liver disease. Med Devices. 2014;7:107–114. doi: 10.2147/MDER.S46943.

[23] Tang F, Ishwaran H. Random forest missing data algorithms. Stat Anal Data Min. 2017;10(6):363–377. doi: 10.1002/sam.11348.

[24] Kim JH. Multicollinearity and misleading statistical results. Korean J Anesthesiol. 2019;72(6):558–569. doi: 10.4097/kja.19087.

[25] Han Y, Wang S. Disability risk prediction model based on machine learning among Chinese healthy older adults: results from the China Health and Retirement Longitudinal Study. Front Public Health. 2023;11:1271595.

[26] Wang Q, Qiao W, Zhang H, et al. Nomogram established on account of Lasso-Cox regression for predicting recurrence in patients with early-stage hepatocellular carcinoma. Front Immunol. 2022;13:1019638. doi: 10.3389/fimmu.2022.1019638.

[27] Chen Q, Meng Z, Liu X, et al. Decision variants for the automatic determination of optimal feature subset in RF-RFE. Genes. 2018;9(6):301. doi: 10.3390/genes9060301.

[28] Fortea JI, Crespo J, Puente Á. Cirrhosis, a global and challenging disease. J Clin Med. 2022;11(21):6512.

[29] Andriamandimby SF, Olive MM, Shimakawa Y, et al. Prevalence of chronic hepatitis B virus infection and infrastructure for its diagnosis in Madagascar: implication for the WHO's elimination strategy. BMC Public Health. 2017;17(1):636. doi: 10.1186/s12889-017-4630-z.

[30] Oeda S, Tanaka K, Oshima A, et al. Diagnostic accuracy of FibroScan and factors affecting measurements. Diagnostics. 2020;10(11):940. doi: 10.3390/diagnostics10110940.

[31] Ji D, Chen Y, Shang Q, et al. Unreliable estimation of fibrosis regression during treatment by liver stiffness measurement in patients with chronic hepatitis B. Am J Gastroenterol. 2021;116(8):1676–1685. doi: 10.14309/ajg.0000000000001239.

[32] Fan R, Li G, Yu N, et al. aMAP score and its combination with liver stiffness measurement accurately assess liver fibrosis in chronic hepatitis B patients. Clin Gastroenterol Hepatol. 2023;21(12):3070–3079.e13. doi: 10.1016/j.cgh.2023.03.005.

[33] Fan R, Yu N, Li G, et al. Machine-learning model comprising five clinical indices and liver stiffness measurement can accurately identify MASLD-related liver fibrosis. Liver Int. 2024;44(3):749–759. doi: 10.1111/liv.15818.

[34] Sanyal AJ, Foucquier J, Younossi ZM, et al. Enhanced diagnosis of advanced fibrosis and cirrhosis in individuals with NAFLD using FibroScan-based Agile scores. J Hepatol. 2023;78(2):247–259. doi: 10.1016/j.jhep.2022.10.034.

[35] Vutien P, Berry K, Feng Z, et al. Combining FIB-4 and liver stiffness into the FIB-5, a single model that accurately predicts complications of portal hypertension. Am J Gastroenterol. 2022;117(12):1999–2008. doi: 10.14309/ajg.0000000000001906.

[36] Kim BK, Kim DY, Park JY, et al. Validation of FIB-4 and comparison with other simple noninvasive indices for predicting liver fibrosis and cirrhosis in hepatitis B virus-infected patients. Liver Int. 2010;30(4):546–553. doi: 10.1111/j.1478-3231.2009.02192.x.

[37] McDonald L, Ramagopalan SV, Cox AP, et al. Unintended consequences of machine learning in medicine? F1000Res. 2017;6:1707. doi: 10.12688/f1000research.12693.1.

[38] Kim BK, Fung J, Yuen MF, et al. Clinical application of liver stiffness measurement using transient elastography in chronic liver disease from longitudinal perspectives. World J Gastroenterol. 2013;19(12):1890–1900. doi: 10.3748/wjg.v19.i12.1890.

[39] Chen SH, Tsai SC, Lu HC. Platelets as a gauge of liver disease kinetics? Int J Mol Sci. 2022;23(19):11460. doi: 10.3390/ijms231911460.

[40] Georgieva M, Xenodochidis C, Krasteva N. Old age as a risk factor for liver diseases: modern therapeutic approaches. Exp Gerontol. 2023;184:112334. doi: 10.1016/j.exger.2023.112334.