Original Article

Lasso Regression-based Model for Cross-Sectional Identification of Significant Hepatic Fibrosis in NAFLD LASSO-based identification of hepatic fibrosis in NAFLD

Wen Deng[1,2,&], Yaqin Zhang[1,2,&], Weihua Cao[1,2,&], Shuojie Wang[1,2,&], Shiyu Wang[1,2], Ziyu Zhang[1,2], Xinxin Li[1,2], Linmei Yao[1,2], Zixuan Gao[1,2], Xin Wei[1,2], Tianyu Ma[1,2], Dianya Qiu[1,2], Hongxiao Hao[1,2,#], Yao Xie[1,2,3,#], and Minghui Li[1,2,3,#]

1. Department of Hepatology Division 2, Beijing Ditan Hospital, Capital Medical University, Beijing 100015, China; 2. HBV Infection, Clinical Cure and Immunology Joint Laboratory, Capital Medical University, Beijing 100015, China; 3. Department of Hepatology Division 2, Peking University Ditan Teaching Hospital, Beijing 100015, China

Abstract

Objective Nonalcoholic fatty liver disease (NAFLD) is an increasing global health concern, with liver-related mortality increasing as fibrosis progresses. This study aimed to identify the key determinants and develop a noninvasive model to detect significant hepatic fibrosis.

Methods A total of 466 patients with biopsy-confirmed NAFLD were retrospectively analyzed at Beijing Ditan Hospital between 2008 and 2018. The patients were classified into non-significant (S0–1) and significant fibrosis (S2–4) groups. Relevant features were selected using least absolute shrinkage and selection operator (LASSO) regression, followed by multivariate logistic regression to construct a model for the cross-sectional identification of significant fibrosis. Model performance was assessed using receiver operating characteristic (ROC) curves, decision curve analysis (DCA), and bootstrap validation.

Results Of the 466 patients, 112 had significant fibrosis. LASSO regression identified 10 relevant features, and the model achieved an AUC of 0.919 (sensitivity, 83.9%; specificity, 85.3%) with a corrected AUC of 0.907 after bootstrap validation. It outperformed the APRI, FIB-4, and LSM ($P < 0.001$), and the DCA confirmed its clinical utility across probability thresholds.

Conclusion The noninvasive model, incorporating demographic, laboratory, and imaging parameters, accurately identified significant hepatic fibrosis in NAFLD and outperformed existing noninvasive scores. This may facilitate interventions and guide personalized management.

Key words: Hepatic fibrosis; LASSO; Noninvasive diagnostic model; Non-alcoholic fatty liver disease

Biomed Environ Sci, 2026; 39(7): 1-10 *doi: 10.3967/bes2026.065* *ISSN: 0895-3988*

www.besjournal.com (full text) *CN: 11-2816/Q*

[&]These authors contributed equally to this work.

[#]Correspondence should be addressed to Hongxiao Hao, MMed, Tel: 15001351665, Email: Haohongxiao1228@163.com; Yao Xie, MD, Tel: 13501093293, Email: xieyao00120184@sina.com; Minghui Li, MD, Tel: 13693259096, Email: wuhm2000@sina.com

Biographical notes of the first authors: Wen Deng, MMed, majoring in prevention, diagnosis, and treatment of viral hepatitis, E-mail: 1075233278@qq.com; Yaqin Zhang, MD, majoring in prevention, diagnosis, and treatment of viral hepatitis, E-mail: dr_zyq0209@163.com; Weihua Cao, MMed, majoring in prevention, diagnosis, and treatment of viral hepatitis, E-mail: weihuacaohappy@163.com; Shuojie Wang, MMed, majoring in prevention, diagnosis, and treatment of viral hepatitis, E-mail: 18947149477@163.com

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is a chronic liver disorder characterized by excessive hepatic fat accumulation and closely associated with metabolic dysfunction. It has become one of the most prevalent liver diseases worldwide, affecting an estimated 38% of adults, a figure that continues to increase alongside global obesity and diabetes epidemics[1-4]. By 2040, the global prevalence of NAFLD is projected to reach 55.4%[5]. Consequently, NAFLD has emerged as a leading chronic liver disease and a major contributor to liver-related morbidity and mortality[6].

The pathological spectrum of NAFLD ranges from simple steatosis to non-alcoholic steatohepatitis (NASH), liver fibrosis, cirrhosis, and hepatocellular carcinoma (HCC)[7]. The presence and severity of fibrosis are strongly associated with liver-related mortality, cardiovascular events, and overall survival, with prognosis worsening as fibrosis progresses[8,9]. Although liver biopsy remains the current "gold standard" for the assessment of fibrosis, its invasiveness, procedural risks, and limited patient acceptance restrict its widespread clinical use. Therefore, there is an urgent need for safe, convenient, cost-effective, and accurate noninvasive tools to evaluate liver injury, identify patients with NAFLD with significant fibrosis (SigFib), and guide clinical management.

Several noninvasive scoring systems, including the aspartate aminotransferase-to-platelet ratio index (APRI), fibrosis-4 index (FIB-4), and liver stiffness measurement (LSM), have been used to assess the likelihood of liver fibrosis in NAFLD. However, most of these tests were originally developed for viral hepatitis and have shown limited diagnostic accuracy in NAFLD populations. Therefore, the development of NAFLD-specific models is necessary. This study aimed to establish a practical clinical tool to identify patients with significant fibrosis, facilitate severity stratification, guide individualized management, and construct a noninvasive diagnostic model for SigFib in NAFLD.

MATERIALS AND METHODS

Data Sources and Study Population

This study was conducted at Beijing Ditan Hospital, Capital Medical University. Patients diagnosed with biopsy-confirmed NAFLD between October 2008 and December 2018 were included.

Patients with biopsy-confirmed NAFLD were enrolled for model development and internal validation. The inclusion criteria were as follows: (1) age 18–65 years; (2) complete clinical data, and (3) definite NAFLD diagnosis based on liver tissue. The exclusion criteria were as follows: (1) liver diseases, such as alcoholic hepatitis, autoimmune hepatitis, and drug-induced hepatitis; (2) liver diseases caused by viral infections, including various hepatitis viruses, Epstein-Barr virus (EBV), cytomegalovirus (CMV), and human immunodeficiency virus (HIV); (3) mental disorders; and (4) liver tumors.

Data Collection

All data were obtained from the Hospital Information System (HIS) and Laboratory Information System (LIS). Demographic data, including age, sex, height, weight, body mass index (BMI), and economic status were collected. Data on medical history, lifestyle (smoking, alcohol consumption, diabetes, hypertension, hyperlipidemia, and sedentary habits), and family history (diabetes, hypertension, hyperlipidemia, and hepatocellular carcinoma) were collected. Laboratory tests were conducted to assess liver function, including alanine aminotransferase (ALT), aspartate aminotransferase (AST), and total bilirubin (TBIL), albumin (ALB), alkaline phosphatase (ALP), and cholinesterase (CHE); renal function tests such as uric acid (UA); lipid profile, including total cholesterol (TC), triglycerides (TG), low-density lipoprotein cholesterol (LDL), high-density lipoprotein (HDL), apolipoprotein A1 (apoA1), apolipoprotein B (apoB), and homocysteine (HCY); blood glucose (GLU); blood routine tests such as white blood cell count (WBC), neutrophil count (NEUT), hemoglobin (HGB), and platelet count (PLT); coagulation function tests such as erythrocyte sedimentation rate (ESR) and international normalized ratio (INR); and liver fibrosis serum markers such as procollagen III (PCIII), procollagen type IV (PIVC), hyaluronic acid (HA), and laminin (LN) were collected. Imaging data such as the controlled attenuation parameter (CAP) and LSM were also obtained. Finally, histopathological data from the liver biopsies were assessed, including the degree of steatosis (S0–S3), inflammation grade (G0–G3), and fibrosis stage (F0–F4).

NAFLD Biopsy and Grouping

Under ultrasound guidance, liver tissue was obtained using a 16G biopsy needle, with a specimen

length ranging from 1.5 to 2.5 cm (minimum 1.0 cm). The specimens were sectioned and stained using routine hematoxylin-eosin, reticulin, and/or Masson's trichrome. Hepatocyte steatosis (S0–3), inflammation grade (G0–3), and fibrosis stage (F0–4) were evaluated according to the "Guidelines for the Diagnosis and Treatment of Nonalcoholic Fatty Liver Disease" (2006 edition). The patients were classified into non-significant fibrosis (non-SigFib, S0–1) and significant fibrosis (SigFib, S2–4) groups[10].

Statistical Analysis

Statistical analyses were performed using SPSS 26.0 and R 4.4.3. Categorical variables were compared using Pearson's chi-square or Fisher's exact tests. The normality of continuous variables was assessed using the Kolmogorov-Smirnov test. Non-normally distributed variables ($P < 0.05$) were summarized as median (interquartile range) and compared using the Mann–Whitney U test, whereas normally distributed variables ($P \geq 0.05$) were presented as mean ± standard deviation. The homogeneity of variance was evaluated using Levene's test; if variances were unequal, Welch's t-test was applied.

Least Absolute Shrinkage and Selection Operator (LASSO) regression was used to identify variables associated with SigFib in NAFLD, and a multivariate logistic regression model was subsequently constructed. A nomogram was developed based on the multivariate model. Receiver operating characteristic (ROC) curves were generated using the pROC package, and areas under the ROC curves (AUC) were compared using the DeLong test. Decision curve analysis (DCA) was performed using the rmda package and data visualization was conducted using ggplot2. All tests were two-sided, with a significance level of $\alpha = 0.05$.

Ethical Considerations

This study was approved by the Ethics Committee of Beijing Ditan Hospital (Ethical ID: Jing Lun Di Zi 2018-052-01). All data were used solely for academic purposes.

RESULTS

Clinical and Demographic Comparison between Significant and Non-significant Fibrosis

A total of 466 NAFLD patients were included, comprising 281 males (60.3%) and 185 females (39.7%). Among them, 354 patients were classified as non-SigFib and 112 as SigFib. Comparative analysis indicated that, relative to the non-SigFib group (F < 2), the SigFib group (F ≥ 2) exhibited the following distinguishing characteristics:

Demographic characteristics: The SigFib NAFLD group had a significantly higher proportion of female patients [54.46% (61/112) *vs.* 35.03% (124/354), χ^2 = 13.426, $P < 0.001$], a higher rate of positive family history of diabetes [35.71% (40/112) *vs.* 18.08% (64/354), χ^2 = 15.262, $P < 0.001$], and a higher prevalence of diabetes [16.07% (18/112) *vs.* 8.47% (30/354), χ^2 = 5.314, $P = 0.021$]; older age [median 46.50 (32.50, 54.00) *vs.* 38.00 (29.00, 48.25) years, $Z = 4.140$, $P < 0.001$] and higher BMI [29.72 (28.54, 30.67) *vs.* 27.76 (24.86, 29.40), $Z = 7.520$, $P < 0.001$]; and worse economic status [72.32% (81/112) *vs.* 51.41% (182/354), χ^2 = 15.129, $P < 0.001$]. No significant differences were observed in the rate of positive family history of hyperlipidemia or the prevalence of hyperlipidemia [47.32% (53/112) *vs.* 53.11% (188/354), χ^2 = 1.141, $P = 0.286$; 32.14% (36/112) *vs.* 40.11% (142/354), χ^2 = 2.289, $P = 0.130$].

Clinical laboratory indicators: The SigFib NAFLD group exhibited higher levels of ESR [median 8.00 (6.00, 15.00) *vs.* 7.00 (5.00, 12.00), $Z = 2.453$, $P = 0.014$], AST [58.50 (34.25, 94.75) *vs.* 42.00 (28.00, 64.25), $Z = 3.721$, $P < 0.001$], and GLU [6.27 (5.47, 7.97) *vs.* 5.74 (5.29, 6.32), $Z = 4.226$, $P < 0.001$]; as well as higher levels of INR [1.03 (0.97, 1.09) *vs.* 0.99 (0.93, 1.03), $Z = 5.147$, $P < 0.001$], PcIII [32.73 (14.59, 52.26) *vs.* 16.49 (12.08, 25.07), $Z = 5.775$, $P < 0.001$], and PIVC [76.46 (65.26, 88.65) *vs.* 67.34 (56.11, 78.36), $Z = 4.011$, $P < 0.001$]. Conversely, the SigFib NAFLD group had significantly lower levels of CHE [9,338.50 (7,863.00, 10,442.25) *vs.* 9,783.00 (8,592.75, 11,050.00), $Z = 2.625$, $P = 0.009$], peripheral neutrophil count [3.08 (2.44, 4.13) *vs.* 3.48 (2.91, 4.30), $Z = 2.837$, $P = 0.005$], HGB [144.00 (135.00, 158.00) *vs.* 152.00 (139.00, 162.00), $Z = 2.472$, $P = 0.013$], and PLT [184.50 (143.50, 234.75) *vs.* 225.00 (181.75, 262.00), $Z = 4.530$, $P < 0.001$].

Imaging indicators: The SigFib NAFLD group exhibited markedly higher LSM values [median 11.35 (8.33, 15.90) *vs.* 6.60 (5.60, 8.30), $Z = 9.968$, $P < 0.001$] and higher CAP values [275.00 (246.75, 310.00) *vs.* 262.50 (211.75, 289.00), $Z = 3.660$, $P < 0.001$]. The patient demographics and clinical characteristics are presented in Table 1.

Establishment of the Diagnostic Model

Variable selection was performed using LASSO regression, with the optimal regularization parameter determined by 10-fold cross-validation.

Table 1. Demographic and Clinical Characteristics of Patients

Values	All-patients (*n* = 466)	SigFib (*n* = 112)	Non-SigFib (*n* = 354)	*Statistic*	*P*
Male (%)	281 (60.30%)	51 (45.54%)	230 (64.97%)	χ^2 = 13.426	< 0.001
Age (years)	39.00 (29.00, 50.00)	46.50 (32.50, 54.00)	38.00 (29.00, 48.25)	*Z* = 4.140	< 0.001
BMI (kg/m^2)	28.31 (25.24, 29.85)	29.72 (28.54, 30.67)	27.76 (24.86, 29.40)	*Z* = 7.520	< 0.001
Smoking (%)	87 (18.67%)	22 (19.64%)	65 (18.36%)	χ^2 = 0.092	0.762
Alcohol consumption (%)	118 (25.32%)	31 (27.68%)	87 (24.58%)	χ^2 = 0.433	0.511
Bad economic status (%)	263 (56.44)	81 (72.32%)	182 (51.41%)	χ^2 = 15.129	< 0.001
Sedentary lifestyle (%)	444 (95.28%)	108 (96.43%)	336 (94.92%)	χ^2 = 0.433	0.510
Family history of diabetes (%)	104 (22.32%)	40 (35.71%)	64 (18.08%)	χ^2 = 15.262	< 0.001
Family history of hypertension (%)	96 (20.60%)	25 (22.32%)	71 (20.06%)	χ^2 = 0.267	0.605
Family history of dyslipidemia (%)	241 (51.72%)	53 (47.32%)	188 (53.11%)	χ^2 = 1.141	0.286
Hypertension (%)	54 (11.59%)	14 (12.50%)	40 (11.30%)	χ^2 = 0.120	0.729
Diabetes mellitus (%)	48 (10.30%)	18 (16.07%)	30 (8.47%)	χ^2 = 5.314	0.021
Dyslipidemia (%)	178 (38.20%)	36 (32.14%)	142 (40.11%)	χ^2 = 2.289	0.130
ESR (mm/h)	7.00 (5.000, 12.00)	8.00 (6.00, 15.00)	7.00 (5.00, 12.00)	*Z* = 2.453	0.014
ALT (U/L)	81.50 (47.15, 143.00)	85.50 (47.75, 144.50)	80.45 (47.15, 139.25)	*Z* = 0.187	0.852
AST (U/L)	45.00 (29.00, 73.00)	58.50 (34.25, 94.75)	42.00 (28.00, 64.25)	*Z* = 3.721	< 0.001
TBIL (μmol/L)	12.45 (9.78, 17.00)	13.00 (9.60, 16.75)	12.00 (9.88, 17.00)	*Z* = 0.692	0.489
ALB (g/L)	46.00 (43.00, 48.00)	45.00 (42.25, 48.00)	46.00 (44.00, 48.13)	*Z* = 1.917	0.055
ALP (U/L)	78.00 (65.00, 99.25)	81.00 (67.00, 103.00)	77.00 (64.00, 97.00)	*Z* = 1.217	0.224
CHE (U/L)	9591.00 (8317.75, 10978.85)	9338.50 (7863.00, 10442.25)	9783.00 (8592.75, 11050.00)	*Z* = 2.625	0.009
UA (μmol/L)	367.00 (299.00, 442.25)	362.50 (282.25, 424.00)	368.50 (304.00, 449.00)	*Z* = 1.551	0.121
TC (mmol/L)	4.81 (4.25, 5.61)	4.69 (4.12, 5.40)	4.84 (4.28, 5.64)	*Z* = 1.534	0.125
TG (mmol/L)	1.74 (1.27, 2.52)	1.52 (1.15, 2.29)	1.79 (1.29, 2.58)	*Z* = 1.819	0.069
LDL (mmol/L)	2.85 (2.27, 3.39)	2.77 (2.30, 3.23)	2.90 (2.26, 3.44)	*Z* = 1.169	0.243
HDL (mmol/L)	1.03 (0.89, 1.20)	1.02 (0.86, 1.14)	1.04 (0.90, 1.22)	*Z* = 1.833	0.067
ApoA1 (g/L)	1.32 (1.16, 1.50)	1.32 (1.17,1.52)	1.32 (1.15,1.50)	*Z* = 0.193	0.847
ApoB (g/L)	0.89 (0.73, 1.07)	0.88 (0.73, 1.06)	0.91 (0.74,1.07)	*Z* = 0.671	0.502
HCY (μmol/L)	12.35 (9.15, 16.09)	12.90 (9.50, 16.63)	12.23 (8.98, 16.04)	*Z* = 0.916	0.360
WBC (10^9/L)	6.13 (5.10, 7.18)	5.96 (4.61,7.05)	6.14 (5.19, 7.25)	*Z* = 1.716	0.086
N (10^9/L)	3.43 (2.79, 4.26)	3.08 (2.44, 4.13)	3.48 (2.91,4.30)	*Z* = 2.837	0.005
HGB (g/L)	150.00 (137.00, 161.00)	144.00 (135.00, 158.00)	152.00 (139.00,162.00)	*Z* = 2.472	0.013
PLT (10^9/L)	216.50 (172.00, 257.00)	184.50 (143.50, 234.75)	225.00 (181.75, 262.00)	*Z* = 4.530	< 0.001
INR	1.00 (0.94, 1.04)	1.03 (0.97, 1.09)	0.99 (0.93, 1.03)	*Z* = 5.147	< 0.001
PcIII (ng/mL)	18.37 (12.48, 32.45)	32.73 (14.59, 52.26)	16.49 (12.08,25.07)	*Z* = 5.775	< 0.001
PIVC (ng/mL)	68.35 (56.37, 80.53)	76.46 (65.26, 88.65)	67.34 (56.11,78.36)	*Z* = 4.011	< 0.001
HA (ng/mL)	75.73 (64.18, 89.27)	75.30 (64.48, 88.01)	76.24 (63.96,89.42)	*Z* = 0.422	0.673
LN (ng/mL)	87.00 (68.42, 110.36)	89.76 (75.24, 112.93)	86.22 (66.46,109.09)	*Z* = 1.867	0.062
GLU（mmol/L）	5.80 (5.33, 6.61)	6.27 (5.47, 7.97)	5.74 (5.29, 6.32)	*Z*=4.226	< 0.001
LSM (kPa)	7.20 (5.80, 10.50)	11.35 (8.33, 15.90)	6.60 (5.60, 8.30)	*Z* = 9.968	< 0.001
CAP (dB/m)	266.00 (216.00, 294.00)	275.00 (246.75,310.00)	262.50 (211.75,289.00)	*Z* = 3.660	< 0.001

Note. BMI, body mass index; ALT, alanine aminotransferase; AST, Aspartate Aminotransferase; TBIL, Total Bilirubin; ALB, Albumin; ALP, Alkaline Phosphatase; CHE, Cholinesterase; UA, Uric Acid; TC, Total Cholesterol; TG, Triglycerides; LDL, Low-Density Lipoprotein; HDL, High-Density Lipoprotein; apoA1, Apolipoprotein A1; ApoB, Apolipoprotein B; HCY, Homocysteine; WBC, White Blood Cell; N, Neutrophils; HGB, Hemoglobin; PLT, Platelets; INR, International Normalized Ratio; PcIII, Procollagen III N-terminal peptide; PIVC, Procollagen Type IV; HA, Hyaluronic Acid; LN, Laminin; GLU, Glucose; LSM, Liver Stiffness Measurement; CAP, Controlled Attenuation Parameter. SigFib: significant fibrosis, non-alcoholic fatty liver disease. Non-SigFib: Non-significant fibrosis, nonalcoholic fatty liver disease.

Based on the "one standard error" criterion (lambda.1se = 0.027), 10 variables with non-zero coefficients and significant discriminative value (sex, age, BMI, TBIL, HDL, PLT, INR, PcIII, LSM, GLU) were selected from an initial pool of 39 candidate variables. The LASSO variable selection results are shown in Figure 1.

Using stepwise regression, ten independent variables (sex, age, BMI, TBIL, HDL, PLT, INR, PcIII, LSM, GLU) were identified and incorporated into a multivariate logistic regression model. The final model equation is: *P(SigFib=1) = 1 / (1 + exp(-(−2.084*

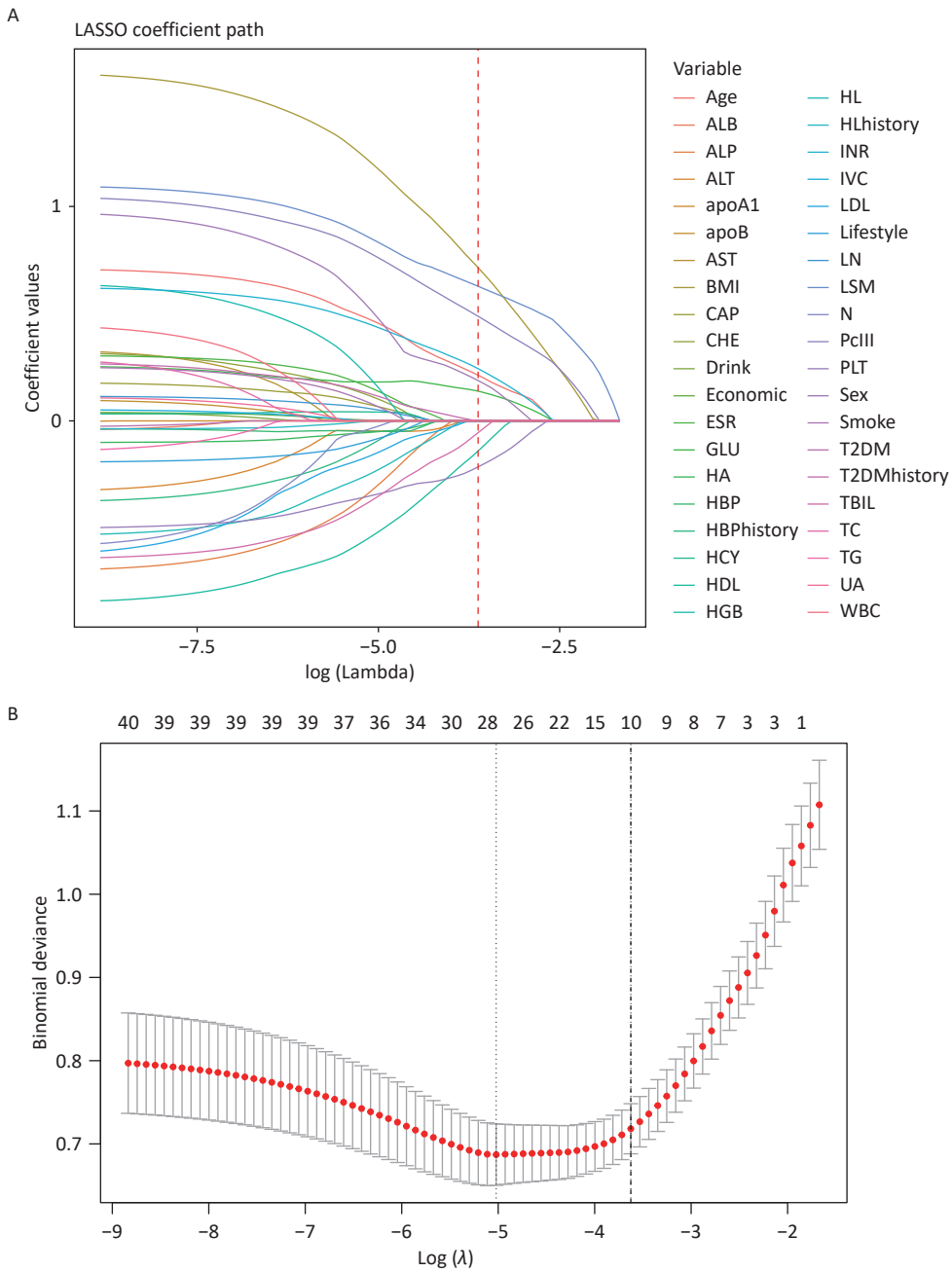

Figure 1. LASSO regression analysis diagram for variable selection in the model. (A) Regularization path plot: The trajectory of variable coefficients shrinks as the regularization parameter λ increases, ultimately retaining 10 variables with nonzero coefficients. (B) Determination of the optimal regularization parameter λ through 10-fold cross-validation. Based on the "one standard error" criterion (lambda.1se = 0.027), the model retained ten variables with nonzero coefficients.

+ 0.442Sex + 0.485Age + 1.327BMI - 0.668TBIL - 0.629HDL - 0.502PLT + 0.512INR + 0.857PcIII + 0.921LSM + 0.254GLU))).

Female sex [*adjusted odds ratio (aOR)* = 1.56, 95% *CI*: 1.07–2.29, Z = 2.29, *P* = 0.022], older age [a*OR* = 1.62, 95% *CI*: 1.10–2.39; Z = 2.42, *P* = 0.016], higher BMI [a*OR* = 3.77, 95% CI: 2.48–5.73; Z = 6.22, *P* < 0.001], higher INR [a*OR* = 1.67, 95% *CI*: 1.24–2.24; Z = 3.39, *P* < 0.001], higher PcIII [a*OR* = 2.36, 95% *CI*: 1.73–3.22; Z = 5.46, *P* < 0.001], higher LSM [a*OR* = 2.51, 95% *CI*: 1.74–3.62; Z = 4.87, *P* < 0.001], and higher GLU levels [a*OR* = 1.29, 95% *CI*: 0.96–1.74; Z = 1.69, *P* = 0.092] were associated with the presence of significant fibrosis in NAFLD. Conversely, higher levels of TBIL [a*OR* = 0.51, 95% *CI*: 0.30–0.87; Z = −2.48, *P* = 0.013], HDL [a*OR* = 0.53, 95% *CI*: 0.37–0.76; Z = −3.41, *P* < 0.001], and PLT [a*OR* = 0.61, 95% *CI*: 0.43–0.87; Z = −2.80, *P* = 0.005] were associated with a lower probability of existing significant fibrosis in NAFLD. The regression coefficients and significance values are presented in Table 2.

A nomogram was developed based on these ten variables to allow individualized assessment of the probability of significant fibrosis in patients with NAFLD (Figure 2).

Internal Validation of the Nomogram

The model was internally validated using bootstrap resampling (1,000 repetitions, random seed = 123). The original ROC-AUC was 0.919 (95% *CI*: 0.894–0.945), with an optimal cutoff value of 0.25, sensitivity of 83.9%, and specificity of 85.3% (Figure 3A). The corrected AUC was 0.907 (optimism = 0.024) and the calibration slope was 1.000 (95% CI: 0.813–1.215). Bootstrap confidence intervals closely followed the reference line, and the Hosmer-Lemeshow test yielded *P* = 0.845. These results indicate that the model demonstrated excellent discrimination and calibration with no significant deviations (Figure 3B).

The diagnostic performance of the SigFib model was compared with that of the APRI, LSM, and FIB-4 models for identifying significant liver fibrosis. The AUCs were 0.919 for the SigFib model, 0.681, 0.812, and 0.712 for FIB-4, respectively. DeLong's test confirmed that the SigFib model outperformed all traditional indicators (all *P* < 0.001), demonstrating its superior discriminative ability and potential clinical utility for noninvasive diagnosis of significant fibrosis in NAFLD (Figure 4).

The AUC values for the SigFib model were 0.919, 0.681, 0.812, and 0.712 for FIB-4, respectively. Statistically significant differences were observed between the SigFib model and APRI, LSM, and FIB-4 index (*P* < 0.001).

Decision Curve Analysis (DCA) of the Diagnostic Model

Across a wide range of threshold probabilities, the SigFib model consistently provided higher net clinical benefit than the reference strategies of "treat all" and "treat none." These results indicate that the clinical decisions guided by the SigFib model offer superior net benefits across diverse scenarios, highlighting its substantial clinical utility. The DCA for

Table 2. Regression Coefficients and Significance

Values	*β*	*SE*	*Wald Z*	*p*	*OR (95% CI)*
Sex-Female	0.442	0.193	2.29	0.022	1.56 (1.07–2.29)
Age	0.485	0.200	2.42	0.016	1.62 (1.10–2.39)
BMI	1.327	0.213	6.22	< 0.001	3.77 (2.48–5.73)
TBIL	−0.668	0.269	−2.48	0.013	0.51 (0.30–0.87)
HDL	−0.629	0.184	−3.41	< 0.001	0.53 (0.37–0.76)
PLT	−0.502	0.180	−2.80	0.005	0.61 (0.43–0.87)
INR	0.512	0.151	3.39	< 0.001	1.67 (1.24–2.24)
PcIII	0.857	0.157	5.46	< 0.001	2.36 (1.73–3.22)
LSM	0.921	0.189	4.87	< 0.001	2.51 (1.74–3.62)
GLU	0.255	0.151	1.69	0.092	1.29 (0.96–1.74)

Note. BMI, body mass index; TBIL, total bilirubin; HDL, high-density lipoprotein; PLT, platelets; INR, international normalized ratio; PcIII, pro-collagen III N-terminal peptide; LSM, Liver Stiffness Measurement; GLU, Glucose.

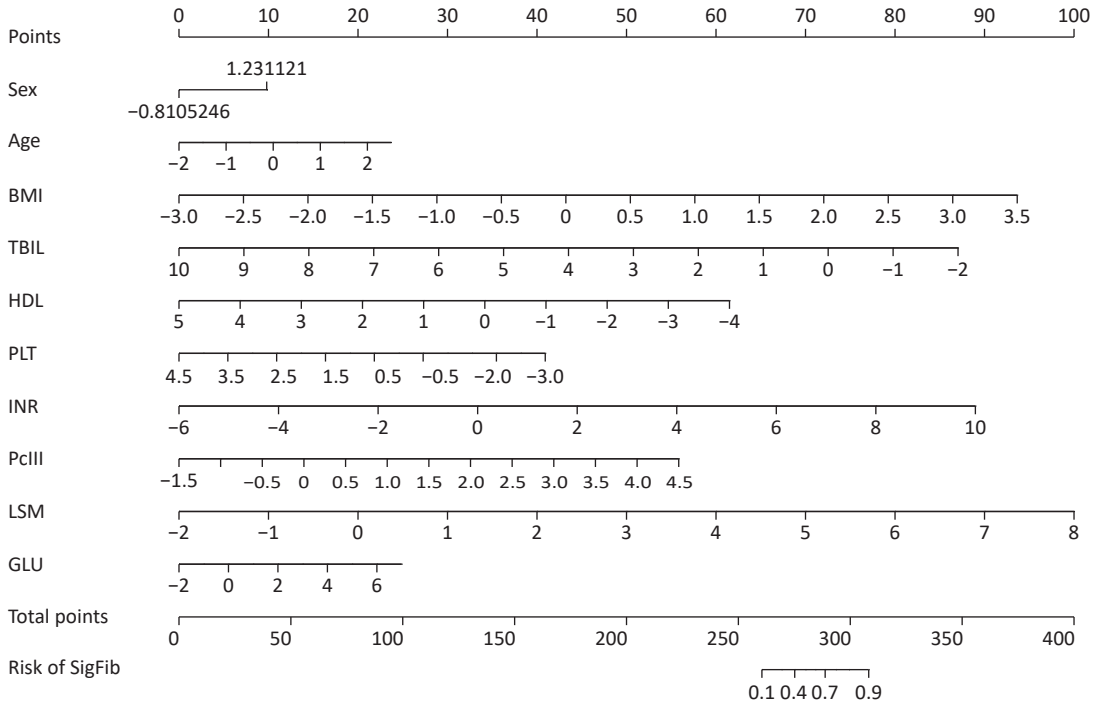

Figure 2. Nomogram for Detecting Significant Fibrosis in NAFLD. Note. BMI, body mass index; TBIL, total bilirubin; HDL, high-density lipoprotein; PLT, platelets; INR, international normalized ratio; PcIII, procollagen III N-terminal peptide; LSM, liver stiffness measurement; GLU, glucose; SigFib-NAFLD, significant fibrosis-non-alcoholic fatty liver disease. The patient's score for each indicator was marked on the corresponding scale, the scores of all indicators were summed to obtain the total score, and the estimated probability was read from the probability axis based on the total score. The final model included 10 variables: sex, age, BMI, TBIL, HDL, PLT, INR, PcIII, LSM, and GLU.

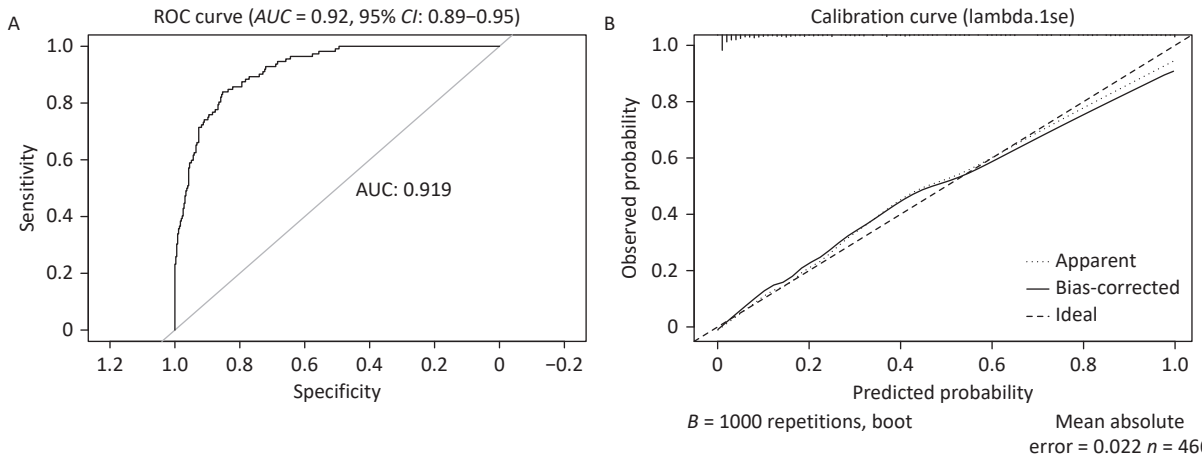

Figure 3. Discrimination and Calibration of Diagnostic Model. (A) Area under the receiver operating characteristic curve with a sensitivity of 83.9%, specificity of 85.3%, and an optimal cutoff value of 0.25. (B) Calibration curve visually compares the model's predicted probabilities with the observed outcomes, providing an intuitive assessment of the calibration performance. The calibration slope = 1.000 (95% *CI*: 0.813–1.215), and the Hosmer-Lemeshow test yielded χ^2 = 4.136, with *P* = 0.845.

the SigFib model is shown in Figure 5.

DISCUSSION

In this study, we developed a multivariate logistic regression model to identify significant fibrosis in NAFLD based on ten key variables using LASSO regression. The model demonstrated excellent discrimination (AUC = 0.919; bootstrap-corrected AUC = 0.907) and good calibration (Hosmer-Lemeshow test, *P* = 0.845). It significantly outperformed the APRI, FIB-4, and LSM alone. By integrating LSM with routine serological markers, the model can be applied to centers equipped with FibroScan. LASSO regression was employed to minimize overfitting and enhance model stability.

Our findings highlight the pivotal role of metabolic factors in the pathogenesis of NAFLD-related fibrosis. Specifically, female sex, older age, and higher BMI were identified as independent factors associated with significant fibrosis, consistent with previous studies highlighting the strong associations between obesity, insulin resistance, sex differences, and NAFLD progression[11-15].

The discriminative performance of the model was enhanced by incorporating key hematological, biochemical, and imaging markers. Reduced PLT levels and elevated INR reflect impaired liver functional reserves and progression of fibrosis[16,17]. Elevated PcIII levels indicate active hepatic collagen deposition and correlate significantly with the histological stage of fibrosis, serving as a marker of fibrogenesis[18,19]. Lower HDL levels may reflect dysregulated hepatic lipid metabolism and chronic inflammation[20-22]. The LSM, an imaging marker of liver stiffness, was identified as a significant factor. Integrating these variables allowed the model to outperform individual indicators in distinguishing between significant and nonsignificant fibrosis.

Compared with traditional scoring systems, the model developed in this study demonstrated clear advantages. APRI and FIB-4 were originally derived from populations with viral hepatitis and showed limited diagnostic accuracy for NAFLD[23]. Although LSM is widely used for noninvasive assessments, its performance can be affected by obesity or concomitant inflammation[24]. By integrating clinical characteristics, laboratory markers, and imaging parameters, the present model overcomes the limitations of individual indicators and improves discriminative ability.

DCA was performed to evaluate the clinical utility of this model. The results demonstrated that the model provides substantial net benefit across a wide range of decision thresholds, indicating that it is not only statistically robust but may also offer practical value in clinical decision-making, guiding stratified management, and intervention according to disease severity.

Previous studies largely relied on serological scores (APRI and FIB-4) or imaging-based elastography measures such as LSM alone. In contrast, the present study systematically integrated LSM with routine serological markers to identify significant hepatic fibrosis in patients with NAFLD. This model may serve as a noninvasive screening tool, potentially reducing unnecessary liver biopsies in low-risk patients.

However, this study had several limitations. First, this was a single-center retrospective study with a relatively small sample size, which may have

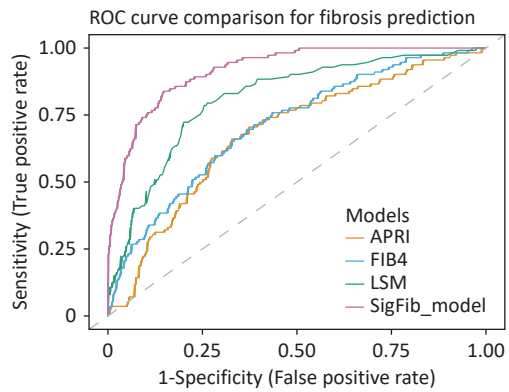

Figure 4. Comparison of different models. Note. APRI, aspartate aminotransferase-to-platelet ratio index; FIB-4, fibrosis-4 index; LSM, liver stiffness measurement; SigFib, significant fibrosis-nonalcoholic fatty liver disease.

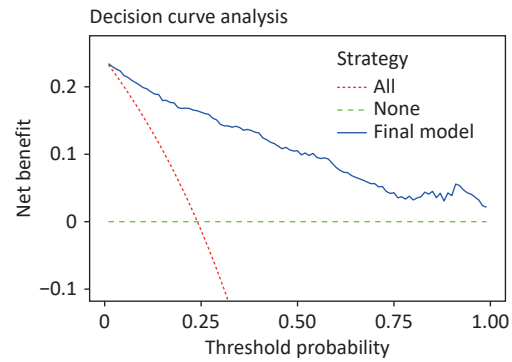

Figure 5. Decision curve of the SigFib NAFLD diagnostic model.

introduced selection bias. Second, the model was validated internally, and its generalizability requires confirmation in independent external cohorts. Third, some emerging noninvasive markers, such as metabolism-related biomarkers or radiomic features, were not included, which may further enhance the discriminative performance in future studies. Importantly, this model is not intended to replace FibroScan or expand its accessibility; rather, by integrating routinely available serological markers (e.g., those included in the APRI and FIB-4) with existing LSM values, it improves the ability to discriminate between significant and nonsignificant hepatic fibrosis.

In conclusion, the SigFib diagnostic model developed in this study using LASSO regression demonstrated robust performance in distinguishing significant fibrosis in NAFLD, surpassing traditional scoring systems and showing substantial clinical utility. Future studies should focus on external validation in multicenter, large-scale prospective cohorts and investigate the incorporation of emerging technologies, such as metabolomics, proteomics, and AI-based imaging analysis, to further improve the accuracy and generalizability of NAFLD fibrosis assessments.

Funding This study was supported by Beijing Municipal Health Commission high-level public health technical personnel construction project (discipline leader-03-26, discipline backbone-02-28); The Beijing Research Ward's Excellence Program (BRWEP2024W102170101); The National Key Research and Development Program (2022YFC2603500, 2022YFC2603505); Capital's Funds for Health Improvement and Research (2022-1-2172); Beijing Hospitals Authority Clinical Medicine Development of Special Funding Support (ZLRK202301); Beijing Hospitals Authority "Peak" Talent Training Program (DFL20241803); National Key Research and Development Program of China (2023YFC2306900); National Key Research and Development Program of the Ministry of Science and Technology (2023YFC2308105).

Competing Interests The authors declare no conflict of interest.

Ethics This study was approved by the Ethics Committee of Beijing Ditan Hospital (Ethical ID: Jing Lun Di Zi 2018-052-01). All data were used for academic research and not for other purposes.

Authors' Contributions Drafting of the manuscript: Wen Deng. Data analysis and manuscript revision: Wen Deng, Yaqin Zhang, Weihua Cao, and Shuojie Wang. Collection of clinical information: Shiyu Wang, Ziyu Zhang, Xinxin Li, Linmei Yao, Zixuan Gao, Xin Wei, Tianyu Ma, and Dianya Qiu. Conceptualization, design, result interpretation, critical review, and revision: Hongxiao Hao, Yao Xie, and Minghui Li.

Received: January 20, 2026;
Accepted: May 19, 2026

REFERENCES

1. Fan JG, Xu XY, Yang RX, et al. Guideline for the prevention and treatment of metabolic dysfunction-associated fatty liver disease (version 2024). J Clin Transl Hepatol, 2024; 12, 955−74.
2. Teng ML, Ng CH, Huang DQ, et al. Global incidence and prevalence of nonalcoholic fatty liver disease. Clin Mol Hepatol, 2023; 29, S32−42.
3. Younossi ZM, Razavi H, Sherman M et al. Addressing the high and rising global burden of metabolic dysfunction-associated steatotic liver disease (MASLD) and metabolic dysfunction-associated steatohepatitis (MASH): from the growing prevalence to payors' perspective. Aliment Pharmacol Ther, 2025; 61, 1467−78.
4. Stefan N, Cusi K. A global view of the interplay between non-alcoholic fatty liver disease and diabetes. Lancet Diabetes Endocrinol, 2022; 10, 284−96.
5. Le MH, Yeo YH, Zou BY, et al. Forecasted 2040 global prevalence of nonalcoholic fatty liver disease using hierarchical Bayesian approach. Clin Mol Hepatol, 2022; 28, 841−50.
6. Li WH, Alazawi W, Loomba R. Current and emerging therapeutic landscape for metabolic dysfunction-associated steatohepatitis. Lancet Gastroenterol Hepatol, 2026; 11, 150−62.
7. Fares R, Mohammed A, Khalil MAF, et al. miRNA-574-5p and miRNA-522 as potential biomarkers in nonalcoholic fatty liver disease. J Gene Med, 2025; 27, e70042.
8. Cerabino N, Di Chito M, Guido D, et al. Liver fibrosis is positively and independently associated with leptin circulating levels in individuals that are overweight and obese: a FibroScan-based cross-sectional study. Nutrients, 2025; 17, 1908.
9. Ng CH, Lim WH, Hui Lim GE, et al. Mortality outcomes by fibrosis stage in nonalcoholic fatty liver disease: a systematic review and meta-analysis. Clin Gastroenterol Hepatol, 2023; 21, 931-9. e5.
10. Chinese Society of Hepatology, Chinese Medical Association, Chinese Society of Gastroenterology, Chinese Medical Association, Chinese Society of Infectious Diseases, Chinese Medical Association. Consensus on the diagnosis and treatment of hepatic fibrosis (2019). Chin J Hepatol, 2019; 27, 657−67. (In Chinese)
11. Tilg H, Petta S, Stefan N, et al. Metabolic dysfunction-associated steatotic liver disease in adults: a review. JAMA, 2026; 335, 163−74.
12. Mehta M, Shah J, Joshi U. Understanding insulin resistance in NAFLD: a systematic review and meta-analysis focused on HOMA-IR in South Asians. Cureus, 2024; 16, e70768.
13. Kalavalapalli S, Leiva EG, Lomonaco R, et al. Adipose tissue insulin resistance predicts the severity of liver fibrosis in patients with type 2 diabetes and NAFLD. J Clin Endocrinol Metab, 2023; 108, 1192−201.
14. Li Q, Shen QL, Xu CN, et al. A self-test prediction model to

determine the probability risk of non-alcoholic fatty liver disease. Biomed Environ Sci, 2020; 33, 857−61.
15. Feng XX, Zhang RT, Yang ZY, et al. Mechanism of metabolic dysfunction-associated steatotic liver disease: important role of lipid metabolism. J Clin Transl Hepatol, 2024; 12, 815−26.
16. Zhang ZF, Wang J. Correlation between PLT, MPV, PDW and liver fibrosis and inflammatory activity in patients with NAFLD: a retrospective case-control study. Medicine, 2025; 104, e43815.
17. Ding RR, Zheng JM, Huang D, et al. INR-to-platelet ratio (INPR) as a novel noninvasive index for predicting liver fibrosis in chronic hepatitis B. Int J Med Sci, 2021; 18, 1159−66.
18. Tang X, Yang L, Zhang PL, et al. Occult hepatitis B virus infection and liver fibrosis in Chinese patients. J Infect Dis, 2023; 228, 1375−84.
19. Mosca A, Comparcola D, Romito I, et al. Plasma N-terminal propeptide of type III procollagen accurately predicts liver fibrosis severity in children with non-alcoholic fatty liver disease. Liver Int, 2019; 39, 2317−29.
20. Hoekstra M, Van Eck M. High-density lipoproteins and non-alcoholic fatty liver disease. Atheroscler Plus, 2023; 53, 33−41.
21. Liang B, Qiu X, Huang JS, et al. Nonlinear associations of the hs-CRP/HDL-C index with metabolic dysfunction-associated steatotic liver disease and advanced liver fibrosis in US adults: insights from NHANES 2017-2018. Sci Rep, 2025; 15, 4029.
22. Zhao H, Qiu X, Li HZ, et al. Association between serum uric acid to HDL-cholesterol ratio and nonalcoholic fatty liver disease risk among Chinese adults. Biomed Environ Sci, 2023; 36, 1−9.
23. Parikh ND, Mehta M, Tapper EB. FIB-4 and APRI for cirrhosis detection in a privately insured national cohort. JHEP Rep, 2024; 6, 100925.
24. Graupera I, Thiele M, Serra-Burriel M, et al. Low accuracy of FIB-4 and NAFLD fibrosis scores for screening for liver fibrosis in the population. Clin Gastroenterol Hepatol, 2022; 20, 2567-76. e6.