Construction of a community-based primary screening and hospital-based confirmatory screening pathway

Contents

Supplementary Methods

Definition of NAFLD and significant fibrosis

Children presenting with the following conditions may be considered for liver biopsy based on the doctor's recommendation: unexplained persistent liver function abnormalities, or the presence of severe liver disease manifestations such as jaundice, ascites, or hepatosplenomegaly. Liver tissue specimens were fixed in buffered formalin, embedded in paraffin and then subjected to hematoxylin and eosin (H&E) and Masson trichrome staining. The slides were reviewed by an experienced pathologist who was blinded to the patients' clinical information. Liver fibrosis in NAFLD patients was staged according to the Brunt staging system[15], with F1 defined as perisinusoidal or periportal fibrosis, F2 defined as perisinusoidal fibrosis with portal or periportal involvement, F3 defined as bridging fibrosis and F4 defined as cirrhosis. In this study, significant fibrosis was defined as a score of $\geq$ F2.

Other causes of fatty liver, such as viral hepatitis, drug-induced fatty liver, Wilson's disease, or autoimmune hepatitis, were excluded via serological examination or past medical history in the PLA 5th.

Statistical analyses

Initially, 33 factors were selected for inclusion in the screening model for tertiary hospitals (Text S1). Least absolute shrinkage and selection operator (LASSO) regression was performed combined with 10-fold cross-validation on the total sample, and stepwise logistic regression (LR) analysis was performed on 1000 bootstrap samples of the total sample. Second, the PLA 5th dataset was randomly divided into a training set and an internal validation set (8:2). A multivariable LR model was

constructed based on the selected factors in the training set; this model was named the full model. Due to the excessive inclusion of parameters in the model and the high cost, a simplified model was also constructed. A satisfactory simplified model should fulfill an AUC >0.80, and no significant difference was detected with the full model ($p \geq 0.01$, likelihood ratio test with Bonferroni correction). Third, the full model and the satisfactory simplified model were subsequently applied to the validation datasets, and their predictive performance was compared with that of previously published child indices, including the BMI-AST, M-APRI, M-FIB-4, PNFS, PNFI, and TyG indices (Text 2). Fourth, the screening performances of the ML models (LR, decision tree (DT), support vector machine (SVM), random forest (RF), artificial neural network (ANN), and extreme gradient boosting (XGBoost)) were compared using 10-fold cross-validation. The DT model was based on the "classification and regression trees" algorithm, and the SVM model was based on the radial basis function kernel. The model parameters were optimized using grid search. The RF model consisted of 50 decision trees and had no maximum tree depth; the ANN model included two hidden layers, utilized the S-shaped cross-entropy, and employed the logistic function as the activation function. The XGBoost model had a maximum tree depth set at 6. In addition, to interpret the predictions of the ML models, Shapley additive explanations (SHAP) were calculated to illustrate the contribution of each factor in the predictive models to the overall model output and to provide a ranking of important predictors. All cutoff values of the indices were the values with the maximum Youden index in the training set.

To enhance the practicality of the models in community health centers[20], 17 regularly

tested factors (sex, age, weight, BMI, SBP, DBP, FPG, HGB, PLT, ALT, creatinine, urea, insulin, UA, TG, TC, HDL, and LDL) were chosen for screening model construction. After excluding the missing data, 101 NAFLD children with complete data were included. All the procedures were similar to those used for the construction of screening models for tertiary hospitals.

In addition, we chose two cutoff values for each model, corresponding to a sensitivity of 90% and a specificity of 90% for predicting significant fibrosis. The diagnostic performance of the two cutoff values was determined by accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

To further evaluate the sequential use of screening models for community health centers and tertiary hospitals, based on the 101 subjects with complete data on all indices in the PLA 5th dataset, we first applied the BIU index, developed for community health center settings, followed by the ATS index, designed for tertiary hospital use to assess the integrated discriminative ability of the sequential model. Firstly, to enhance the NPV, we applied cut-off values corresponding to 90% sensitivity. If the two screening models' scores were below the cut-off values corresponding to each model, the result was classified as negative; otherwise, it was classified as positive. Subsequently, we applied cut-off values corresponding to 90% specificity. If the two screening models' scores were above the cut-off values of each model, the result was classified as positive; otherwise, it was classified as negative, aiming to improve the PPV of the sequential model.

Text 1: Thirty-three variables:

Including sex, age, weight, BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), FPG, international normalized ratio (INR), prothrombin activity, fibrinogen, lipoprotein a, apolipoprotein B (ApoB), apolipoprotein A1 (Apo A1), aspartate transaminase (AST), alanine transaminase (ALT), alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), total bile acid (TBA), cholinesterase, albumin, globulin, prealbumin, direct bilirubin, total bilirubin, triglycerides (TG), total cholesterol (TC), platelets (PLT), creatinine, high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), urea, uric acid (UA), and hemoglobin (HGB).

Text 2: Indices for predicting fibrosis in previous studies were calculated as follows:

B-AST=BMI z-score ×AST(U/l)；

APRI=$\frac{\frac{\text{AST(U/L)}}{\text{AST upper limit of normal (U/L)}}}{\text{platelets}(10^9/\text{L})}\times 100$； M-APRI=BMI z-score ×APRI；

FIB-4=$\frac{\text{age(years)}\times\text{AST(U/L)}}{\text{PLT}(10^9/\text{L})\times\sqrt{\text{ALT(U/L)}}}$； M-FIB-4=BMI z-score ×FIB-4；

Forns-Index=7.811-3.131×lnPLT(10^9/L) +0.781×ln GGT (U/L) +3.467 ×ln age (years)-0.014×cholesterol (mg/dL)；

HSI=8 ×ALT(U/L)/AST(U/L) + BMI(kg/m^2) (+2 if T2D, +2 if female)；

NFS = -1.675 + 0.037× age(years) +0.094 ×BMI + 1.13×IFG /T2D (yes=1; no=0)+ 0.99× AST/ALT - 0.013× PLT (10^9/L)- 0.66× albumin (g/dL);

PNFS=1/(1+e^{-z}) × 100 (z=1.1+(0.34×$\sqrt{\text{ALT(U/L)}}$)+0.002×ALP(U/L)-1.1×ln PLT (10^9/L)- 0.02× GGT (U/L));

PNFS=1/(1+e^{-lp}) × 10 (lp=-6.539× ln age(years) +0.207×waist(cm)+1.957× ln TG (mg/dL) -10.074；

TyG=ln [TG (mg/dL) × FPG (mg/dL)/2];

VAI-female=$\frac{\text{WC(cm)}}{39.68+(1.88\times\text{BMI})}\times\frac{\text{TG(mg/dL)}}{1.03}\times\frac{1.31}{\text{HDL(mg/dL)}}$;

VAI-male=$\frac{\text{WC(cm)}}{36.58+(1.89\times\text{BMI})}\times\frac{\text{TG(mg/dL)}}{0.81}\times\frac{1.52}{\text{HDL(mg/dL)}}$

Text 3: Indices for predicting fibrosis in the present study were calculated as follows:

ATC index=1.938 + 0.007×ALP (U/L)+ 0.079×TBA (μmol/L)- 0.048× creatinine (μmol/L)+ 0.007×AST (U/L)- 0.001× cholinesterase (U/L)+ 0.045× weight (kg) - 0.003× UA (μmol/L)-2.274× HDL (mmol/L) +0.768× fibrinogen (g/L) -0.026× DBP (mmHg) +1.402× INR +0.002× prealbumin (mg/L), the optimal cutoff value for screening significant fibrosis is 0.31.

HIU index = 5.063+ 0.072× insulin (mU/L)- 0.009× UA (μmol/L)- 0.054× HGB (g/L)- 0.644× FPG (mmol/L)- 0.057× creatinine (μmol/L) +0.271× age (year)+ 0.164× BMI (kg/m^2) + 0.022× SBP (mmHg)+ 2.030× HDL (mmol/L)- 0.409× urea (mmol/L). The cutoff value with the maximum Youden index was 0.91.

Supplementary Table 1 Number of missing variables

Variables	Missing	Variables	Missing
Sex	0	Prealbumin	7
Age	0	Direct bilirubin	6
Weight	0	Total bilirubin	3
BMI	10	ALT	1
SBP	6	AST	2
DBP	6	ALP	5
FPG	12	GGT	5
INR	5	TBA	6
Prothrombin activity	5	Creatinine	7
Fibrinogen	5	Cholinesterase	6
Lipoprotein a	17	Urea	10
ApoB	17	UA	10
ApoA1	17	TC	11
HGB	1	TG	10
PLT	1	HDL	18
Albumin	2	LDL	18
Globulin	13	Insulin	155

ALP=alkaline phosphatase, ALT=alanine aminotransferase, ApoB= apolipoprotein B, ApoA1= apolipoprotein A1, AST=aspartate aminotransferase, DBP=diastolic blood pressure, FPG=fasting plasma glucose, GGT=glutamyl transpeptidase, HDL=high-density lipoprotein cholesterol, HGB=hemoglobin, INR= international normalized ratio, LDL=low-density lipoprotein cholesterol, PLT=platelets, SBP=systolic blood pressure, TBA=total bile acid, TC=total cholesterol, TG=triglycerides, UA=uric acid.

Supplementary Table 2. Comparison of data characteristics for excluding missing values.

Variables [a]	*P* [b]	Variables	*P* [b]
ALP	0.349	Insulin	0.728
TBA	0.932	UA	0.092
Creatinine	0.529	HGB	0.713
AST	0.470	FPG	0.323
Cholinesterase	0.841	Creatinine	0.652
Weight	0.828	Age	0.377
UA	0.696	BMI	0.121
HDL	0.555	SBP	0.524
Fibrinogen	0.32	HDL	0.445
DBP	0.959	Urea	0.887
INR	0.91		
Prealbumin	0.782		

ALP=alkaline phosphatase, AST=aspartate aminotransferase, DBP=diastolic blood pressure, FPG=fasting plasma glucose, HDL=high-density lipoprotein cholesterol, HGB=hemoglobin, INR= international normalized ratio, SBP=systolic blood pressure, TBA=total bile acid, UA=uric acid.

a The left panel presents the comparison of data characteristics before and after the exclusion of missing values for variables included in the pediatric significant fibrosis screening model developed for tertiary hospitals. The right panel shows the corresponding comparison for the model developed for community health centers.

b T-test was used for normally distributed continuous variables, and Wilcoxon test for non-normally distributed continuous variables.

Supplementary Table 3 Selection of candidate predictors

Variables	Model 1	Model 2	Variables	Model 1	Model 2
Sex	335	--	Prealbumin	288	**1**
Age	549	--	Direct bilirubin	566	--
Weight	**759**	--	Total bilirubin	301	--
BMI	520	**1**	ALT	494	--
SBP	289	--	AST	**881**	**1**
DBP	**687**	--	ALP	**955**	**1**
FPG	605	--	GGT	540	--
INR	546	**1**	TBA	**930**	**1**
Prothrombin activity	438	--	Creatinine	**898**	**1**
Fibrinogen	**745**	**1**	Cholinesterase	**854**	--
Lipoprotein a	339	--	Urea	312	--
ApoB	503	--	UA	**755**	--
ApoA1	344	--	TC	463	--
HGB	406	--	TG	632	--
PLT	395	--	HDL	**752**	**1**
Albumin	333	--	LDL	510	--
Globulin	496	--			

Model 1 showed the number of times that candidate predictors were selected (1000 times) based on bootstrapped stepwise logistic regression, and the predictors included in the final model are marked in bold; Model 2 was LASSO regression analysis combined with 10-fold cross-validation, and the selected predictor was represented by "1".

ALP=alkaline phosphatase, ALT=alanine aminotransferase, ApoB= apolipoprotein B, ApoA1= apolipoprotein A1, AST=aspartate aminotransferase, DBP=diastolic blood pressure, FPG=fasting plasma glucose, GGT=glutamyl transpeptidase, HDL=high-density lipoprotein cholesterol, HGB=hemoglobin, INR= international normalized ratio, LDL=low-density lipoprotein cholesterol, PLT=platelets, SBP=systolic blood pressure, TBA=total bile acid, TC=total cholesterol, TG=triglycerides, UA=uric acid.

Supplementary Table 4 Variance inflation factor of variables based on ATC.

Variables	VIF
ALP	1.51
TBA	1.17
Creatinine	2.66
AST	1.19
Cholinesterase	1.46
Weight	2.66
UA	1.55
HDL	1.46
Fibrinogen	1.24
DBP	1.47
INR	1.19
Prealbumin	1.39

VIF = Variance inflation factor.

Supplementary Table 5 Screening performances of indices for the diagnosis of ≥F2 in the training and validation set

	AUC (95%CI)	*P* [a]	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
Training set								
ATC	0.84 [0.78, 0.89]	Ref	0.73	0.82	0.77	0.82	0.73	0.55
B-AST	0.65 [0.57, 0.73]	<0.001	0.45	0.88	0.65	0.81	0.58	0.32
APRI	0.64 [0.55, 0.72]	<0.001	0.51	0.80	0.64	0.74	0.59	0.30
M-APRI	0.63 [0.54, 0.70]	<0.001	0.44	0.84	0.63	0.76	0.57	0.27
FIB-4	0.58 [0.49, 0.66]	<0.001	0.56	0.61	0.59	0.62	0.55	0.18
M- FIB-4	0.60 [0.52, 0.69]	<0.001	0.46	0.82	0.63	0.74	0.57	0.27
Forns	0.54 [0.45, 0.62]	<0.001	0.88	0.30	0.61	0.59	0.69	0.19
HSI	0.55 [0.47, 0.64]	<0.001	0.59	0.57	0.58	0.60	0.55	0.15
NFS	0.54 [0.46, 0.63]	<0.001	0.17	0.94	0.53	0.76	0.50	0.11
PNFS	0.55 [0.47, 0.64]	<0.001	0.49	0.74	0.61	0.68	0.56	0.22
TyG	0.56 [0.48, 0.65]	<0.001	0.80	0.36	0.59	0.59	0.61	0.16
ATS	0.81 [0.75, 0.87]	/	0.64	0.88	0.75	0.86	0.68	0.51
Internal validation set								
ATC	0.80 [0.66, 0.91]	Ref	0.68	0.69	0.69	0.62	0.75	0.37
B-AST	0.65 [0.46, 0.81]	0.058	0.37	0.89	0.67	0.70	0.66	0.27
APRI	0.67 [0.50, 0.82]	0.188	0.47	0.92	0.73	0.82	0.71	0.42
M-APRI	0.62 [0.43, 0.79]	0.033	0.32	0.92	0.67	0.75	0.65	0.26
FIB-4	0.63 [0.45, 0.78]	0.114	0.53	0.54	0.53	0.46	0.61	0.06

M-FIB-4	0.62 [0.43, 0.79]	0.049	0.42	0.89	0.69	0.73	0.68	0.32
Forns	0.54 [0.36, 0.71]	<0.001	0.32	0.73	0.56	0.46	0.59	0.05
HSI	0.55 [0.37, 0.73]	0.002	0.42	0.69	0.58	0.50	0.62	0.12
NFS	0.60 [0.41, 0.77]	0.068	0.90	0.12	0.44	0.43	0.60	0.01
PNFS	0.60 [0.41, 0.77]	0.072	0.58	0.65	0.62	0.55	0.68	0.23
TyG	0.67 [0.51, 0.82]	0.205	0.26	0.85	0.60	0.56	0.61	0.12
ATS	0.70 [0.54, 0.85]	/	0.53	0.77	0.67	0.63	0.69	0.30

AUC=area under the curve, PPV=positive predictive value, NPV=negative predictive value. ATC was a logistic regression model developed in this study that included thirteen parameters: alkaline phosphatase, total bile acid, creatinine, aspartate aminotransferase, cholinesterase, weight, uric acid, high-density lipoprotein cholesterol, fibrinogen, diastolic blood pressure, BMI, international normalized ratio and prealbumin. ATS was a logistic regression model developed in this study that included six parameters: alkaline phosphatase, total bile acid, aspartate aminotransferase, cholinesterase, high-density lipoprotein cholesterol, and fibrinogen. The calculation methods for other indicators are shown in Text 2.

a. Delong test (two-sided 5% significance level)

Supplementary Table 6 Screening performances of ML models based on ATC for the diagnosis of ≥F2 using 10-fold cross-validation

	AUC (95%CI)	*P* [a]	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
LR	0.71 [0.66, 0.77]	Ref	0.69	0.73	0.71	0.73	0.70	0.42
DT	0.73 [0.67, 0.78]	0.722	0.73	0.72	0.73	0.73	0.72	0.45
SVM	0.69 [0.63, 0.75]	0.400	0.64	0.74	0.69	0.72	0.66	0.38
RF	0.83 [0.77, 0.87]	<0.001	0.81	0.84	0.82	0.84	0.81	0.65
ANN	0.71 [0.65, 0.77]	0.890	0.68	0.73	0.71	0.73	0.69	0.41
XGBoost	0.79 [0.74, 0.85]	0.016	0.81	0.78	0.79	0.78	0.79	0.56

LR= logistic regression, DT= decision tree, SVM= support vector machine, RF=random forest, ANN=artificial neural network, XGBoost=eXtreme gradient boosting, AUC=area under the curve, PPV=positive predictive, NPV=negative predictive value. ATC was a logistic regression model developed in this study that included thirteen parameters: alkaline phosphatase, total bile acid, creatinine, aspartate aminotransferase, cholinesterase, weight, uric acid, high-density lipoprotein cholesterol, fibrinogen, diastolic blood pressure, BMI, international normalized ratio and prealbumin.

a. Delong test (two-sided 5% significance level)

Supplementary Table 7 Simplification of ATC

	AUC (95%CI)	*P* [a]	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
TBA	0.68 [0.61, 0.76]	<0.001	0.83	0.43	0.64	0.62	0.69	0.27
TBA, AST	0.72 [0.64, 0.80]	<0.001	0.51	0.88	0.68	0.83	0.61	0.38
ALP, AST, HDL	0.75 [0.68, 0.82]	<0.001	0.72	0.75	0.73	0.76	0.71	0.47
ALP, TBA, cholinesterase, BMI	0.78 [0.71, 0.85]	<0.001	0.78	0.74	0.76	0.77	0.74	0.51
ALP, TBA, AST, HDL, fibrinogen	0.80 [0.73, 0.86]	0.004	0.68	0.83	0.75	0.82	0.70	0.51
ALP, TBA, AST, cholinesterase, HDL, fibrinogen	0.81 [0.75, 0.87]	0.081	0.64	0.88	0.75	0.86	0.68	0.51
ALP, TBA, AST, cholinesterase, weight, HDL, fibrinogen	0.82 [0.75, 0.88]	0.148	0.83	0.71	0.77	0.77	0.79	0.54
ALP, TBA, creatinine, AST, cholinesterase, weight, HDL, fibrinogen	0.83 [0.77, 0.89]	0.681	0.66	0.88	0.76	0.86	0.70	0.53
ALP, TBA, creatinine, AST, cholinesterase, weight, HDL, fibrinogen, DBP	0.83 [0.77, 0.89]	0.746	0.75	0.87	0.80	0.86	0.75	0.61

AST=aspartate aminotransferase, ALP=alkaline phosphatase, TBA=total bile acids, HDL=high-density lipoprotein cholesterol, DBP=diastolic blood pressure, AUC=area under the curve, PPV=positive predictive value, NPV=negative prediction value. ATC was a logistic regression model developed in this study that included thirteen parameters: alkaline phosphatase, total bile acid, creatinine, aspartate aminotransferase, cholinesterase, weight, uric acid, high-density lipoprotein cholesterol, fibrinogen, diastolic blood pressure, BMI, international normalized ratio and prealbumin.
a. Likelihood ratio test was conducted between simplified indictors and ATC and corrected with Bonferroni correction.

Supplementary Table 8 Screening performance of ATS and other indices for the diagnosis of ≥F2 in the training and validation set

Training set	*P* [a]	**Internal validation set**	*P* [a]
ATS	Ref	ATS	Ref
B-AST	<0.001	B-AST	0.594
APRI	<0.001	APRI	0.810
M-APRI	<0.001	M-APRI	0.445
FIB-4	<0.001	FIB-4	0.573
M- FIB-4	<0.001	M- FIB-4	0.498
Forns	<0.001	Forns	0.014
HSI	<0.001	HSI	0.070
NFS	<0.001	NFS	0.427
PNFS	<0.001	PNFS	0.408
TyG	<0.001	TyG	0.816

ATS was a logistic regression model developed in this study that included six parameters: alkaline phosphatase, total bile acid, aspartate aminotransferase, cholinesterase, high-density lipoprotein cholesterol, and fibrinogen. The calculation methods for other indicators are shown in Text 2.

a. Delong test (two-sided 5% significance level).

Supplementary Table 9 Screening performances of ML models based on ATS for the diagnosis of ≥F2 in the training and validation sets (90% Sensitivity and 90% Specificity)

	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
Training set						
90% Sensitivity						
LR	0.90	0.49	0.71	0.67	0.82	0.41
DT	0.89	0.73	0.82	0.79	0.86	0.63
SVM	0.89	0.61	0.76	0.72	0.84	0.52
RF	0.89	1.00	0.94	1.00	0.89	0.89
ANN	0.90	0.99	0.94	0.99	0.90	0.89
XGBoost	0.90	0.99	0.94	0.99	0.90	0.89
90% Specificity						
LR	0.44	0.90	0.65	0.84	0.59	0.33
DT	0.74	0.90	0.82	0.90	0.76	0.64
SVM	0.73	0.90	0.81	0.90	0.75	0.63
RF	1.00	0.90	0.95	0.92	1.00	0.91
ANN	0.99	0.90	0.95	0.92	0.99	0.90
XGBoost	0.98	0.90	0.94	0.92	0.97	0.89
Percentage of patients within indeterminate zone						
LR	44.07					
DT	15.82					

SVM	22.03					
RF	10.17					
ANN	8.47					
XGBoost	7.91					
Internal Validation set						
90% Sensitivity						
LR	0.95	0.46	0.67	0.56	0.92	0.37
DT	0.68	0.58	0.62	0.54	0.71	0.25
SVM	0.89	0.50	0.67	0.57	0.87	0.37
RF	0.53	0.92	0.76	0.83	0.73	0.47
ANN	0.58	0.69	0.64	0.58	0.69	0.27
XGBoost	0.68	0.81	0.76	0.72	0.78	0.50
90% Specificity						
LR	0.47	0.81	0.67	0.64	0.68	0.29
DT	0.53	0.73	0.64	0.59	0.68	0.26
SVM	0.68	0.73	0.71	0.65	0.76	0.41
RF	1.00	0.58	0.76	0.63	1.00	0.54
ANN	0.68	0.62	0.64	0.57	0.73	0.29
XGBoost	0.95	0.50	0.69	0.58	0.93	0.41
Percentage of patients within indeterminate zone						

LR	40.00
DT	15.56
SVM	22.22
RF	40.00
ANN	8.89
XGBoost	28.89

LR= logistic regression, DT= decision tree, SVM= support vector machine, RF=random forest, ANN=artificial neural network, XGBoost=eXtreme gradient boosting, AUC=area under the curve, PPV=positive predictive, NPV=negative predictive value. ATS was a logistic regression model developed in this study that included six parameters: alkaline phosphatase, total bile acid, aspartate aminotransferase, cholinesterase, high-density lipoprotein cholesterol, and fibrinogen.

Supplementary Table 10 Selection of candidate predictors

Variables	Model 1	Model 2
Sex	469	--
Age	**565**	--
Weight	449	--
BMI	**549**	--
SBP	**546**	--
DBP	344	--
Insulin	**996**	**1**
FPG	**592**	--
HGB	**776**	--
PLT	393	--
ALT	460	--
Creatinine	**566**	--
Urea	**521**	--
UA	**976**	**1**
TC	488	--
TG	502	--
HDL	**546**	--
LDL	425	--

Model 1 showed the number of times that candidate predictors were selected (1000 times) based on bootstrapped stepwise logistic regression, and the predictors included in the final model are marked in bold; Model 2 was LASSO regression analysis combined with 10-fold cross-validation, and the selected predictor was represented by "1".

SBP=systolic blood pressure, DBP=diastolic blood pressure, ALT=alanine aminotransferase, FPG=fasting plasma glucose, TC=total cholesterol,

TG=triglycerides, PLT=platelets, HDL=high-density lipoprotein cholesterol, LDL=low-density lipoprotein cholesterol, UA=uric acid, HGB=hemoglobin.

Supplementary Table 11 Variance inflation factor of variables based on HIU.

Variables	VIF
Insulin	1.52
UA	1.58
HGB	1.61
FPG	1.32
Creatinine	2.86
Age	2.64
BMI	2.12
SBP	2.31
HDL	2.07
Urea	1.85

VIF = Variance inflation factor.

Supplementary Table 12 Screening performance of indices for the diagnosis of ≥F2 in the training and validation sets

	AUC (95%CI)	*P* [a]	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
Training set								
HIU	0.86 [0.78, 0.94]	Ref	0.67	0.94	0.79	0.94	0.69	0.59
B-AST	0.65 [0.53, 0.77]	0.004	0.80	0.51	0.68	0.68	0.67	0.32
APRI	0.64 [0.51, 0.76]	0.006	0.56	0.80	0.66	0.78	0.58	0.34
M-APRI	0.63 [0.50, 0.75]	0.002	0.67	0.66	0.66	0.71	0.61	0.32
FIB-4	0.60 [0.47, 0.73]	0.001	0.62	0.63	0.63	0.68	0.56	0.25
M- FIB-4	0.63 [0.51, 0.75]	0.002	0.53	0.74	0.63	0.73	0.55	0.27
Forns	0.52 [0.39, 0.66]	<0.001	0.84	0.37	0.64	0.63	0.65	0.23
HSI	0.54 [0.41, 0.66]	<0.001	0.42	0.74	0.56	0.68	0.50	0.16
NFS	0.57 [0.44, 0.71]	<0.001	0.62	0.57	0.60	0.65	0.54	0.19
PNFS	0.62 [0.49, 0.74]	0.002	0.60	0.74	0.66	0.75	0.59	0.33
TyG	0.50 [0.37, 0.63]	<0.001	0.82	0.31	0.60	0.61	0.58	0.14
BIU	0.81 [0.71, 0.90]	/	0.84	0.74	0.80	0.81	0.79	0.59
Internal validation set								
HIU	0.91 [0.75, 1.00]	Ref	0.82	0.90	0.86	0.90	0.82	0.72
B-AST	0.61 [0.36, 0.85]	0.034	0.82	0.40	0.62	0.60	0.67	0.22
APRI	0.69 [0.46, 0.91]	0.125	0.46	0.90	0.67	0.83	0.60	0.35
M-APRI	0.66 [0.42, 0.91]	0.078	0.82	0.70	0.76	0.75	0.78	0.52

FIB-4	0.59 [0.34, 0.85]	0.028	0.55	0.60	0.57	0.60	0.55	0.15
M- FIB-4	0.60 [0.34, 0.86]	0.033	0.36	0.80	0.57	0.67	0.53	0.16
Forns	0.54 [0.26, 0.81]	0.005	1.00	0.40	0.71	0.65	1.00	0.41
HSI	0.56 [0.27, 0.82]	0.013	0.82	0.30	0.57	0.56	0.60	0.12
NFS	0.59 [0.32, 0.83]	0.009	0.70	0.60	0.65	0.64	0.67	0.30
PNFS	0.69 [0.42, 0.90]	0.101	0.55	0.80	0.67	0.75	0.62	0.34
TyG	0.64 [0.37, 0.88]	0.036	0.82	0.50	0.67	0.64	0.71	0.32
BIU	0.88 [0.70, 1.00]	/	0.82	0.70	0.76	0.75	0.78	0.52

AUC=area under the curve, PPV=positive predictive value, NPV=negative predictive value. HIU was a logistic regression model developed in this study that included ten parameters: insulin, uric acid, hemoglobin, fasting plasma glucose, creatinine, age, BMI, systolic blood pressure, high-density lipoprotein cholesterol, urea. BIU was a logistic regression model developed in this study that included three parameters: BMI, insulin, and uric acid. The calculation methods for other indicators are shown in Text 2.

a. Delong test (two-sided 5% significance level).

Supplementary Table 13 Screening performances of ML models based on HIU for the diagnosis of ≥F2 using 10-fold cross-validation

	AUC (95%CI)	*P* [a]	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
LR	0.75 [0.67, 0.84]	Ref	0.75	0.76	0.75	0.79	0.71	0.50
DT	0.71 [0.62, 0.80]	0.404	0.73	0.69	0.71	0.75	0.67	0.42
SVM	0.77 [0.69, 0.86]	0.616	0.77	0.78	0.77	0.81	0.73	0.54
RF	0.76 [0.68, 0.84]	0.853	0.77	0.76	0.76	0.80	0.72	0.52
ANN	0.68 [0.59, 0.77]	0.140	0.71	0.64	0.68	0.71	0.64	0.36
XGBoost	0.74 [0.65, 0.83]	0.833	0.75	0.73	0.74	0.78	0.70	0.48

LR= logistic regression, DT= decision tree, SVM= support vector machine, RF=random forest, ANN=artificial neural network, XGBoost=eXtreme gradient boosting, AUC=area under the curve, PPV=positive predictive, NPV=negative predictive value. HIU was a logistic regression model developed in this study that included ten parameters: insulin, uric acid, hemoglobin, fasting plasma glucose, creatinine, age, BMI, systolic blood pressure, high-density lipoprotein cholesterol, urea.

a. Delong test (two-sided 5% significance level)

Supplementary Table 14 Simplification the HIU

	AUC	*P* [a]	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
Insulin	0.74 [0.62, 0.85]	0.014	0.91	0.51	0.74	0.71	0.82	0.44
Insulin, UA	0.80 [0.69, 0.90]	0.140	0.71	0.86	0.78	0.87	0.70	0.55
insulin, UA, BMI	0.81 [0.71, 0.90]	0.196	0.84	0.74	0.80	0.81	0.79	0.59
insulin, UA, FPG, BMI	0.82 [0.73, 0.91]	0.223	0.84	0.74	0.80	0.81	0.79	0.59
insulin, UA, FPG, BMI, urea	0.84 [0.75, 0.92]	0.259	0.89	0.69	0.80	0.78	0.83	0.59
insulin, UA, FPG, HGB, BMI, HDL	0.84 [0.75, 0.92]	0.204	0.84	0.77	0.81	0.83	0.79	0.62

UA=uric acid, HDL=high-density lipoprotein cholesterol, FPG=fasting plasma glucose, HGB= hemoglobin, AUC=area under the curve, PPV=positive predictive value, NPV=negative prediction value. HIU was a logistic regression model developed in this study that included ten parameters: insulin, uric acid, hemoglobin, fasting plasma glucose, creatinine, age, BMI, systolic blood pressure, high-density lipoprotein cholesterol, urea.

a. Likelihood ratio test was conducted between simplified indictors and ATC and corrected with Bonferroni correction.

Supplementary Table 15 Screening performances of BIU and other indices for the diagnosis of ≥F2 in the training and validation set

Training set	*P* [a]	**Internal validation set**	*P* [a]
BIU	Ref	BIU	Ref
B-AST	0.034	B-AST	0.071
APRI	0.039	APRI	0.205
M-APRI	0.017	M-APRI	0.141
FIB-4	0.014	FIB-4	0.068
M- FIB-4	0.022	M- FIB-4	0.070
Forns	<0.001	Forns	0.020
HSI	<0.001	HSI	0.015
NFS	0.001	NFS	0.029
PNFS	0.016	PNFS	0.158
TyG	<0.001	TyG	0.077

BIU was a logistic regression model developed in this study that included three parameters: BMI, insulin, and uric acid. The calculation methods for other indicators are shown in Text 2.

a. Delong test (two-sided 5% significance level).

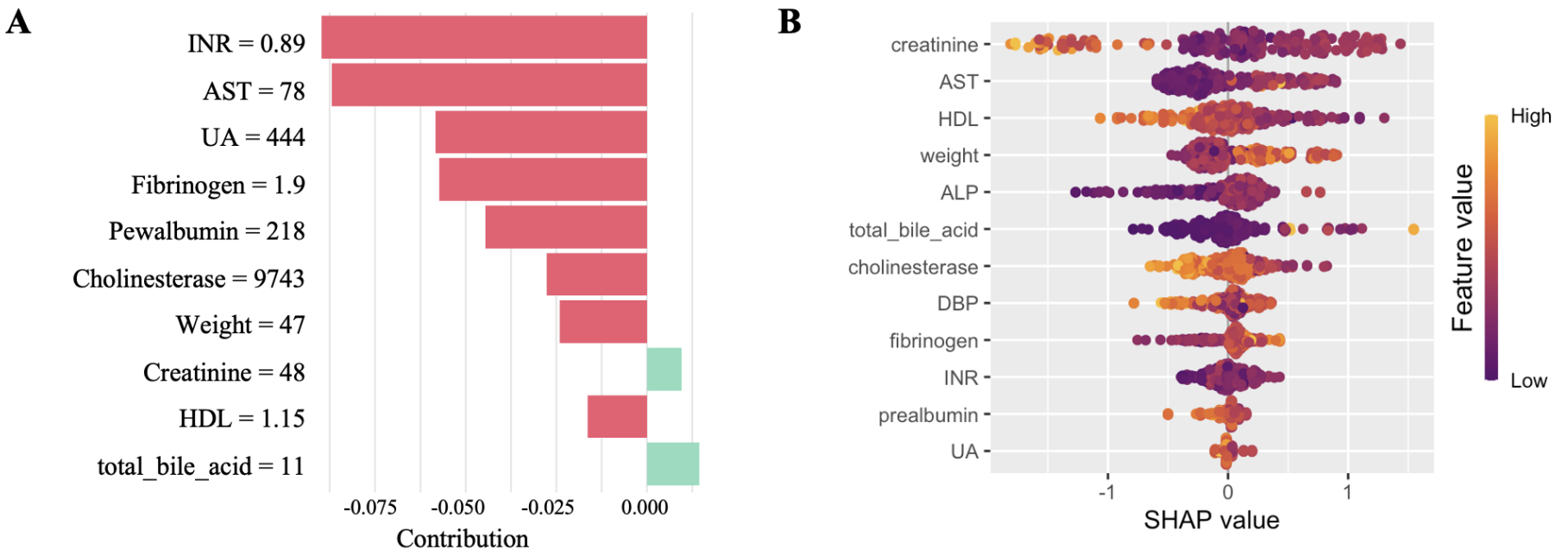
Supplementary Table 16 Screening performances of ML models based on BIU for the diagnosis of ≥F2 in the training and validation sets (90% Sensitivity and 90% Specificity)

	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
Training set						
90% Sensitivity						
LR	0.91	0.29	0.64	0.62	0.71	0.21
DT	0.98	0.40	0.72	0.68	0.93	0.40
SVM	0.91	0.46	0.71	0.68	0.80	0.39
RF	0.93	1.00	0.96	1.00	0.92	0.92
ANN	0.89	1.00	0.94	1.00	0.88	0.88
XGBoost	0.91	1.00	0.95	1.00	0.90	0.90
90% Specificity						
LR	0.60	0.91	0.74	0.90	0.64	0.49
DT	0.64	0.89	0.75	0.88	0.66	0.51
SVM	0.67	0.89	0.76	0.88	0.67	0.53
RF	1.00	0.91	0.96	0.94	1.00	0.92
ANN	1.00	0.91	0.96	0.94	1.00	0.92
XGBoost	0.98	0.91	0.95	0.94	0.97	0.90
Percentage of patients within indeterminate zone						
LR	45.00					
DT	40.00					

SVM	32.50					
RF	7.50					
ANN	10.00					
XGBoost	7.50					
Internal Validation set						
90% Sensitivity						
LR	0.91	0.40	0.67	0.62	0.80	0.32
DT	0.91	0.30	0.62	0.59	0.75	0.21
SVM	1.00	0.60	0.81	0.73	1.00	0.61
RF	0.73	0.80	0.76	0.80	0.73	0.52
ANN	0.91	0.90	0.90	0.91	0.90	0.81
XGBoost	0.73	0.60	0.67	0.67	0.67	0.33
90% Specificity						
LR	0.54	1.00	0.76	1.00	0.67	0.53
DT	0.55	1.00	0.76	1.00	0.67	0.53
SVM	0.45	0.90	0.67	0.83	0.60	0.35
RF	1.00	0.50	0.76	0.69	1.00	0.51
ANN	0.91	0.70	0.81	0.77	0.88	0.61
XGBoost	0.91	0.60	0.76	0.71	0.86	0.52
Percentage of patients within indeterminate zone						

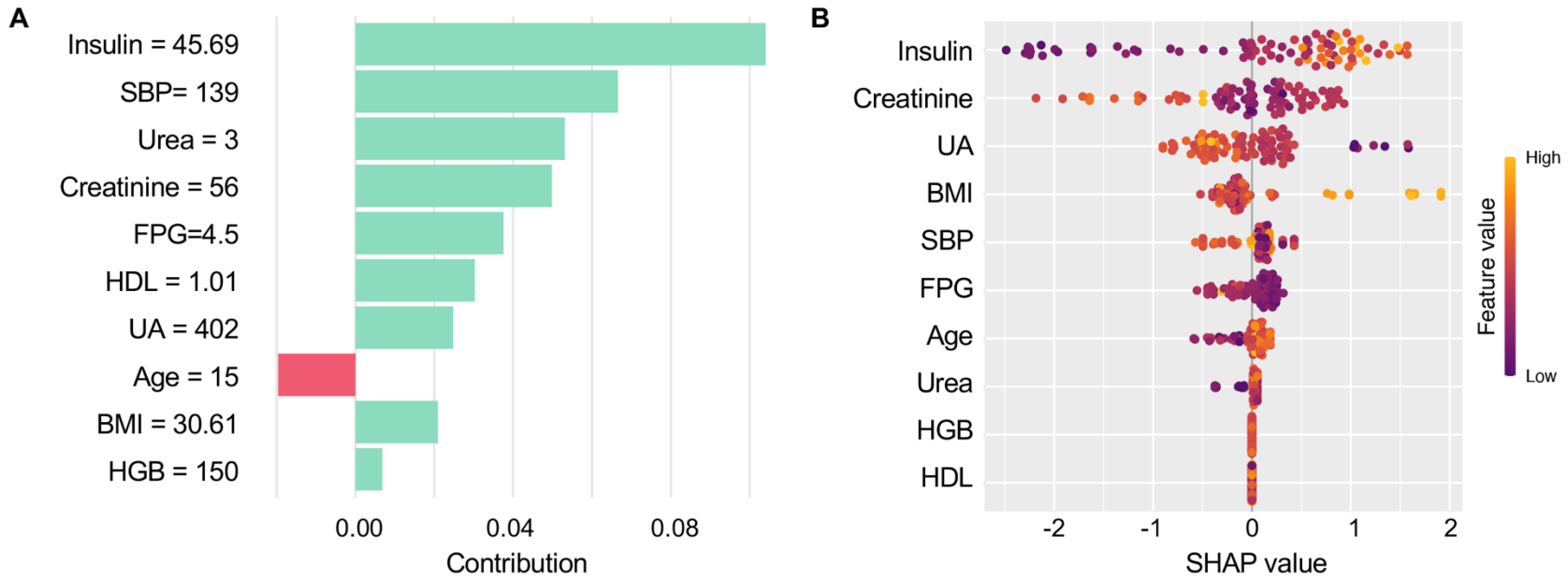
LR	47.62
DT	52.38
SVM	42.86
RF	28.57
ANN	9.52
XGBoost	9.52

LR= logistic regression, DT= decision tree, SVM= support vector machine, RF=random forest, ANN=artificial neural network, XGBoost=eXtreme gradient boosting, AUC=area under the curve, PPV=positive predictive value, NPV=negative predictive value. BIU was a logistic regression model developed in this study that included three parameters: BMI, insulin, and uric acid.



Supplementary Figure 1 SHAP summary plots of the 13 selected features of the RF model (A) and XGBoost model (B).

The coordinates on the x-axis represent SHAP values. The bar plots A depict the average SHAP values corresponding to each feature in the RF model. In plot B, each point corresponds to the SHAP value of an individual, where yellow indicates higher feature values, and purple indicates lower feature values. RF: random forest; XGBoost: eXtreme gradient boosting.



Supplementary Figure 2 SHAP summary plots of the 10 selected features of the RF model (A) and XGBoost model (B).

The coordinates on the x-axis represent SHAP values. The bar plots A depict the average SHAP values corresponding to each feature in the RF model. In plot B, each point corresponds to the SHAP value of an individual, where yellow indicates higher feature values, and purple indicates lower feature values. RF: random forest; XGBoost: eXtreme gradient boosting.