



Original article

An exploratory machine learning model for predicting advanced liver fibrosis in autoimmune hepatitis patients: A preliminary study

Qinglin Wei^{a,b,c,1}, Wen Li^{b,1}, Shubei He^d, Hongbo Wu^d, Qiaoling Xie^d, Ying Peng^{d,*}, Xingyue Zhang^{e,*}^a Department of Gastroenterology, The Seventh Medical Center of Chinese PLA General Hospital, Beijing 100700, China^b Department of Cadre Diagnosis and Treatment, The Seventh Medical Center of Chinese PLA General Hospital, Beijing 100700, China^c Department of Gastroenterology, Xinqiao Hospital of Third Military Medical University (Army Medical University), Chongqing 400038, China^d Department of Gastroenterology, Southwest Hospital of Third Military Medical University (Army Medical University), Chongqing 400038, China^e Department of Dermatology, The Seventh Medical Center of Chinese PLA General Hospital, Beijing 100700, China

ARTICLE INFO

Article History:

Received 18 May 2024

Accepted 23 October 2024

Available online 12 December 2024

Keywords:

Autoimmune hepatitis

Liver fibrosis

Machine learning

Random forest

Non-invasive testing

ABSTRACT

Introduction and Objectives: Advanced fibrosis is a crucial stage in the progression of autoimmune hepatitis (AIH), where fibrosis can either regress or advance. This study aims to leverage machine learning (ML) models for the assessment of advanced liver fibrosis in AIH patients using routine clinical features.

Patients and Methods: A total of 233 patients diagnosed with AIH and underwent liver biopsy were included in the discovery cohort. The dataset was randomly split into training and testing sets. Patients were categorized into groups with no/minimal/moderate fibrosis and advanced fibrosis. Six ML models were employed to identify the optimal model. Subsequently, the predictive capability of the best ML model was validated in an additional cohort ($n = 33$) and compared with conventional noninvasive fibrosis scores.

Results: Three key clinical features, including prothrombin time (PT), albumin (ALB), and ultrasound spleen thickness (UTST), were analyzed by least absolute shrinkage and selection operator (LASSO) regression. In the training set, the random forest (RF) model showed the highest diagnostic performance in predicting advanced fibrosis stage (AUC=0.951). In the testing cohort and validation cohort, the RF model maintained high accuracy (AUC = 0.863 and AUC = 0.843). Additionally, the random forest model outperformed the conventional noninvasive fibrosis scores.

Conclusions: ML models, particularly the RF model, can help improve the discrimination of advanced liver fibrosis in patients with AIH.

© 2024 Fundación Clínica Médica Sur, A.C. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Autoimmune hepatitis (AIH) is characterized by inflammation of the liver parenchyma resulting from autoimmune reactions. It manifests with elevated serum transaminase, positive autoantibodies, hyperglobulinemia, and interface hepatitis in liver histology, predominantly affecting females [1]. The prevalence of AIH has gradually increased in recent years [2]. The International Autoimmune Hepatitis Group (IAIHG) scoring system, widely used for AIH diagnosis, saw simplified diagnostic criteria in 2008, achieving high sensitivity (90 %) and specificity (95 %) [3]. Approximately one-third of AIH patients progress to advanced fibrosis and cirrhosis at diagnosis, increasing the risk of hepatocellular carcinoma (HCC) [4]. Advanced fibrosis significantly influences AIH prognosis, as patients in this

stage may either regress or progress to cirrhosis. Early identification and intervention at this stage are crucial for improving outcomes and the quality of life for AIH patients.

The diagnosis of AIH and the staging of fibrosis present significant challenges in clinical practice. Liver biopsy, the gold standard for evaluating liver fibrosis, faces limitations such as invasiveness, cost, associated risks (e.g., bleeding, infection), potential sampling errors, and variability in interpretation (both inter-observer and intra-observer), making it unsuitable for monitoring and long-term treatment response assessment [5–7]. As is well known, transient elastography (TE) is a special ultrasound scan that assesses liver stiffness as a surrogate marker for liver fibrosis [8,9]. However, the significance of TE in AIH patients is controversial, as the elevation of alanine aminotransferase (ALT) level and hepatic inflammation may affect the accuracy of TE in detecting liver fibrosis [10]. The high cost of TE devices and operator requirements limit its clinical use in resource-limited settings. Blood-based tests like fibrosis index based on the four

* Corresponding authors.

E-mail addresses: ying.peng@cldcsw.org (Y. Peng), 452935536@qq.com (X. Zhang).¹ These authors contributed equally to this work and share the first authorship.

factors (FIB-4), aspartate aminotransferase to alanine aminotransferase ratio (AAR), and aspartate aminotransferase to platelet ratio index (APRI) have shown limited capabilities in measuring advanced fibrosis in AIH [2,11,12]. Therefore, developing a noninvasive model to discriminate advanced liver fibrosis in AIH is crucial.

The advancement of electronic medical records and hospital information platforms has facilitated the easier acquisition of clinical data. In this era, machine learning (ML) models have demonstrated effectiveness in appraising diagnoses, implementing early warning systems, and predicting drug responses within the medical field [13–15]. Notably, supervised learning in ML has proven superior in predicting clinical outcomes compared to traditional statistical analyses [16]. In a groundbreaking approach, we propose the development of a noninvasive model using ML methods to predict the advanced fibrosis stage. This innovative model utilizes routine data readily available in clinical settings, marking the first instance of such an approach. By harnessing the power of ML, we aim to enhance the accuracy and efficiency of predicting advanced liver fibrosis, thereby providing a valuable tool for clinicians in their decision-making processes.

2. Patients and Methods

2.1. Patients' selection

This study included a total of 233 participants with AIH, aged 18 years or older, encompassing both type 1 and type 2 AIH, which were formed as discovery cohort. The participants were sourced from the First Affiliated Hospital of Army Medical University (Southwest Hospital) between January 2010 and April 2022. Additionally, nine AIH patients from the First Affiliated Hospital of Army Medical University (Southwest Hospital) between May 2022 and October 2023 and 24 AIH patients from the Second Affiliated Hospital of Army Medical University (Xinqiao Hospital) between January 2018 and October 2023 were recruited, which were formed as validation cohort.

All participants underwent a comprehensive assessment, including demographic data, medical history, clinical manifestations, and abdominal ultrasonography. Prior to receiving systematic therapy (including albumin infusion, liver-protecting drug, prednisone and azathioprine), all patients had undergone liver biopsy within a week of obtaining blood for laboratory tests. The liver biopsy confirmed clear pathological features of AIH, and fibrosis stages were also evaluated. The diagnosis of AIH was established when a simplified International Autoimmune Hepatitis Group (IAIHG) score reached at least 7.

Exclusion criteria were applied as follows: (1) patients with hepatocellular carcinoma (HCC) or other malignancies; (2) patients with concurrent liver diseases such as viral hepatitis, nonalcoholic fatty liver disease (NAFLD), alcohol-related liver disease, primary biliary cholangitis (PBC), primary sclerosing cholangitis (PSC), drug-induced liver disease (DILI), and other inherited or metabolic liver diseases; (3) patients suffering from other serious autoimmune diseases, heart, respiratory, or hematological diseases; (4) patients lacking a substantial amount of data.

Twenty-eight common clinical features were gathered for our study, including demographic information: age, sex, height, weight, body mass index (BMI); routine blood tests: red blood cells (RBC), white blood cells (WBC), platelet count (PLT), hemoglobin (HGB), red blood cell distribution width (RDW), mean platelet volume (MPV), platelet distribution width (PDW), hematocrit value (HCT), monocyte, lymphocyte, neutrophil; coagulation function: prothrombin time (PT), activated partial thromboplastin time (APTT); abdominal ultrasonography: ultrasonic spleen thickness (UTST); blood biochemical: serum globulin (GLB), serum albumin (ALB), total bilirubin (TBL), total bile acid (TBA), alanine aminotransferase (ALT), aspartate

aminotransferase (AST), alkaline phosphatase (ALP), gamma-glutamyl transpeptidase (GGT).

FIB-4, AAR, and APRI were calculated as follows:

$$\text{FIB-4} = (\text{Age years} \times \text{AST}) / (\text{PLT} \times \sqrt{\text{ALT}})$$

$$\text{AAR} = \text{AST} / \text{ALT}$$

$$\text{APRI} = (\text{AST} / \text{upper limit of normal AST}) \times 100 / \text{PLT}$$

Note: AST and ALT were measured in units per liter, and PLT was measured in 10^9 per liter.

Liver biopsy (LB) procedures were carried out using a 16-gauge disposable needle. All liver specimens obtained were independently scored by pathologists who were blinded to the patients' information. The Schauer grading system was employed to classify liver fibrosis into the following stages [17]: S0, no fibrosis; S1, portal fibrosis without septa; S2, portal fibrosis with rare septa; S3, numerous septa without cirrhosis, and S4, cirrhosis. Based on these fibrosis stages, patients were categorized into two groups: S0–S2, defining no/minimal/moderate liver fibrosis, and S3–S4, defining advanced liver fibrosis.

The research adhered to the principles of the Declaration of Helsinki and received approval from the Ethics Committee Boards of Southwest Hospital and Xinqiao Hospital. Informed consent was waived for this study.

2.2. Statistical analyses

K-nearest neighbor classification was employed to impute missing values in the dataset. Continuous variables were presented as mean \pm standard deviation (SD), while categorical variables were expressed as percentages. To compare normally distributed continuous variables, a two-sample independent *t*-test was utilized. For non-normally distributed continuous variables, the Mann-Whitney U test was applied. The comparison of categorical variables was conducted using the Chi-square test or Fisher test. A two-tailed *p*-value was used to indicate statistical significance. All statistical analyses were performed using IBM SPSS Statistics version 26.0.

2.3. Machine learning model

The Least Absolute Shrinkage and Selection Operator (LASSO) technique was employed to select variables, facilitating the creation of a simplified model by constructing a penalty function and mitigating overfitting. Independent predictive indicators identified through LASSO were then utilized to establish supervised machine learning (ML) models. SHapley Additive exPlanations (SHAP) were employed to explain the importance and role of each clinical feature included in the ML models.

Six ML methods were established, including random forest (RF), extreme gradient boosting (XGBoost), logistic regression (LR), multi-layer perceptron (MLP), decision tree (DT), and support vector machine (SVM), to determine the most fitting ML model. Ten-fold cross-validation was applied to create more reliable parameter combinations for the ML models. The predictive accuracies of the six ML models were assessed using the area under the curve (AUC). The models were subsequently validated in both the testing cohort and validation cohort.

Finally, the performance of the best ML model was compared with conventional noninvasive models based on serum indicators, including FIB-4, AAR, and APRI. This comprehensive approach aimed to identify the most effective and accurate model for predicting advanced liver fibrosis in patients with AIH.

Calibration curve and decision curve analyses (DCA) were employed for the assessment of clinical application. The calibration curve assessed whether the prediction probabilities of the model were in conformity with clinical practice. On the other hand, DCA calculated the benefit under different threshold probabilities to infer the clinical applicability of the model. To achieve a comprehensive

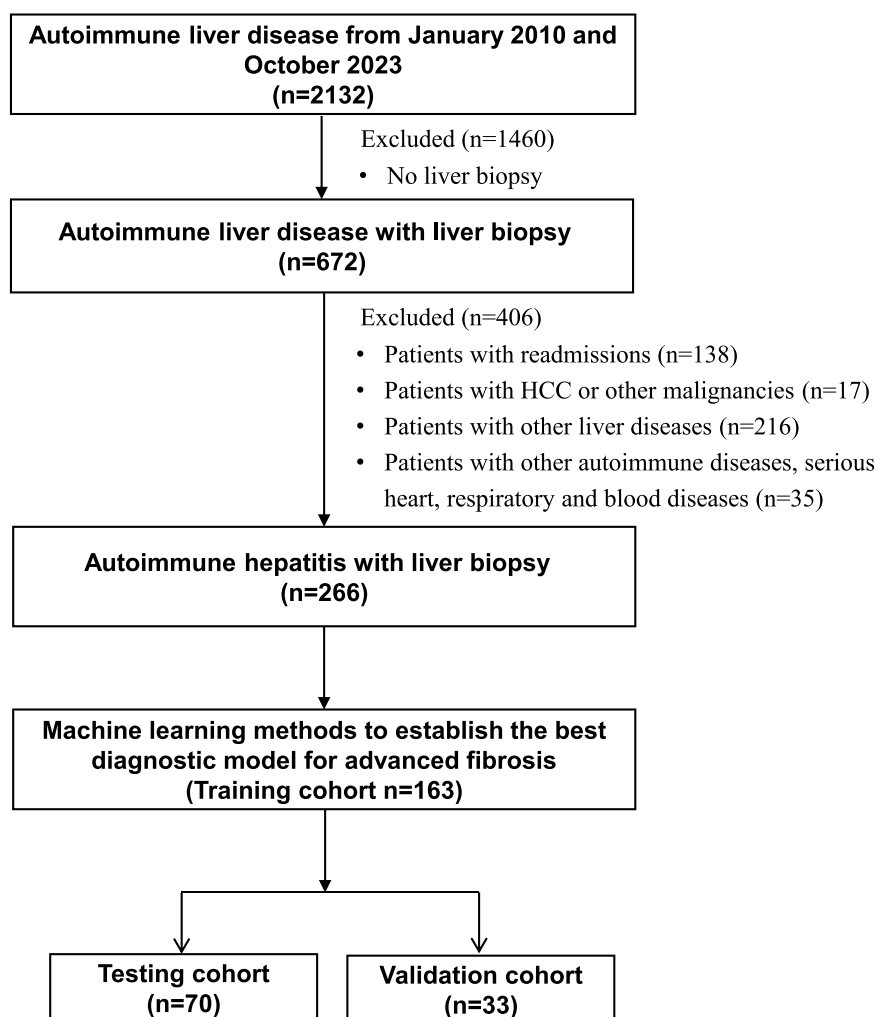


Fig. 1. Flow chart of study population. This flow chart illustrates the process of patient selection and cohort formation in the study, detailing the inclusion and exclusion criteria, and the final composition of the discovery and validation cohorts.

evaluation of the six ML models, sensitivity, specificity, accuracy, and recall were calculated. These metrics provide insights into the performance of the models in different aspects of prediction. Finally, the optimal ML model was transformed into an online calculator to facilitate its use in clinical practice. This step aims to enhance the practicality and accessibility of the model for healthcare professionals. The ML analyses were performed using R 4.1.2 software.

2.4. Machine learning model training, testing, and validation

The discovery dataset was randomly partitioned into training and testing cohorts. In the training cohort, ten-fold cross-validation was implemented, where the data was divided into ten groups. Nine groups were utilized for training the ML model, and one group was reserved for cross-validation to ensure robustness. Diagnostic indicators, including accuracy, AUC, sensitivity, specificity, and recall, were assessed for the ML model in both the training and testing sets. The optimal ML model was then validated in an external cohort to assess its performance on new and unseen data. Additionally, a comparison was conducted between the optimal ML model and noninvasive fibrosis models, which included FIB-4, AAR, and APRI. This comprehensive evaluation aimed to ascertain the superiority of the ML model in predicting advanced liver fibrosis compared to conventional diagnostic approaches.

2.5. Ethics statement

All subjects were waived the need for informed consent. In order to ensure confidentiality, the names of study participants were not included in the data. Information obtained from the data of the study participants is kept confidential. In addition, the Ethics Committee of the Southwest Hospital and Xinqiao Hospital Third Military Medical University (Army Medical University) approved the study. Informed patient consent was waived with the consent of the ethics committee of the local hospital.

3. Results

3.1. Clinical characteristics of all AIH patients

A total of 266 patients diagnosed with AIH were included in the study, with 233 patients forming the discovery cohort and 33 comprising the validation cohort (Fig. 1). Among all the participants, 236 were female, accounting for 88.7 % of the total. The gender frequency, body mass index (BMI), most liver function indicators, and noninvasive fibrosis tests showed comparable distributions between the discovery and validation cohorts (Table 1).

In the discovery cohort, the mean BMI was 21.89 kg/m², and the mean age was 49 years. The distribution of liver fibrosis stages was as follows: S0 in 54 patients (23.2 %), S1 in 76 patients (32.6 %), S2 in

Table 1
Comparative clinical features of AIH patients in the discovery versus validation cohort.

Characteristic	Discovery cohort (n = 233)	Validation cohort (n = 33)	p value
Age (year)	48.0 ± 11.2	52.6 ± 11.8	0.029
Sex (male/female)	27/206 (11.6 %)	3/30 (9.1 %)	0.896
Height (m)	1.6 ± 0.1	1.6 ± 0.1	0.632
Weight (kg)	55.4 ± 8.2	54.7 ± 7.4	0.938
BMI (kg/m ²)	22.3 ± 3.0	22.3 ± 3.4	0.923
WBC (× 10 ⁹ /L)	5.0 ± 1.6	4.9 ± 2.7	0.204
RBC (× 10 ⁹ /L)	3.9 ± 0.6	3.5 ± 0.6	<0.001
HGB (g/L)	119.1 ± 18.0	91.5 ± 45.0	0.002
PLT (× 10 ⁹ /L)	157.0 ± 71.4	152.0 ± 90.1	0.716
Monocyte (× 10 ⁹ /L)	0.4 ± 0.2	0.4 ± 0.3	0.461
Lymphocyte (× 10 ⁹ /L)	1.6 ± 0.7	1.4 ± 0.6	0.271
Neutrophil (× 10 ⁹ /L)	2.9 ± 1.2	3.0 ± 2.3	0.249
RDW-CV (%)	15.2 ± 2.9	15.0 ± 2.2	0.728
RDW-SD (%)	49.6 ± 8.7	51.8 ± 6.2	0.012
HCT (%)	36.2 ± 5.1	33.3 ± 6.5	0.004
MPV (fL)	11.5 ± 1.4	11.7 ± 1.3	0.296
PDW (fL)	15.7 ± 2.4	15.4 ± 2.8	0.803
ALT (U/L)	187.4 ± 287.7	186.2 ± 238.5	0.887
AST (U/L)	206.0 ± 276.8	195.1 ± 238.7	0.790
ALP (U/L)	165.2 ± 97.0	208.7 ± 155.0	0.122
GGT (U/L)	149.8 ± 107.8	230.1 ± 181.2	0.031
TBIL (umol/L)	68.7 ± 97.1	65.5 ± 87.5	0.243
TBA (umol/L)	72.9 ± 88.0	84.4 ± 108.9	0.572
ALB (g/L)	37.7 ± 6.6	34.8 ± 7.4	0.433
PT (sec)	12.0 ± 2.9	12.5 ± 3.0	0.415
APTT (sec)	32.7 ± 8.7	32.9 ± 8.8	0.866
UTST (mm)	39.0 ± 10.0	43.4 ± 15.5	0.546
FIB-4	6.3 ± 7.6	7.3 ± 7.4	0.259
AAR	1.5 ± 1.6	1.4 ± 1.0	0.855
APRI	4.1 ± 7.5	3.9 ± 4.5	0.668
IgG (g/L)	20.1 ± 3.5	21.3 ± 5.6	0.860
ANA (+)	184/233 (80.0 %)	26/33 (80.0 %)	0.981
SMA (+)	51/233 (21.9 %)	7/33 (21.2 %)	0.930
Anti-LKM (+)	11/233 (4.7 %)	1/33 (3.0 %)	0.661
Anti-SLA/LP (+)	4/233 (1.7 %)	1/33 (3.0 %)	0.603

AIH, autoimmune hepatitis; BMI, body mass index; WBC, white blood cells; RBC, red blood cells; HGB, hemoglobin; PLT, platelet count; RDW: red blood cell distribution width; HCT, hematocrit value; MPV, mean platelet volume; PDW, platelet distribution width; ALT: alanine aminotransferase; AST: aspartate aminotransferase; ALP: alkaline phosphatase; GGT: gamma-glutamyl transpeptidase; TBIL, total bilirubin; TBA, total bile acid; ALB, albumin; PT, prothrombin time; APTT, activated partial thromboplastin time; UTST, Ultrasonic spleen thickness; FIB4, fibrosis index based on the four factors; AAR, aspartate aminotransferase to alanine aminotransferase ratio; APRI, aspartate aminotransferase to platelet ratio index.

44 patients (18.9 %), S3 in 31 patients (13.3 %), and S4 in 28 patients (12.0 %). [Table 2](#) revealed statistically significant differences in most indicators and non-invasive fibrosis scores between the none-to-moderate fibrosis group (S0-S2) and the advanced fibrosis group (S3-S4).

3.2. Feature selection

A total of 28 candidate features were considered for analysis. Among these, two features, MPV and PDW, had missing values. The k-nearest neighbor algorithm was employed to fill in the missing values for these features. The process of feature selection was carried out using LASSO regression, as depicted in [Fig. 2](#). In accordance with clinical observations, it was found that decreased albumin (ALB), prolongation of prothrombin time (PT), and an increase in ultrasonic spleen thickness (UTST) were significantly associated with advanced fibrosis. These identified features are consistent with clinical expectations and align with established indicators of liver fibrosis in autoimmune hepatitis.

Table 2
Comparative clinical features of AIH patients in the discovery cohort.

Characteristic	Fibrosis S0-S2 (n = 174)	Fibrosis S3-S4 (n = 59)	p value
Age (year)	47.0 ± 11.1	50.8 ± 11.3	0.023
Sex (male/female)	22/152 (12.6 %)	5/54 (8.5 %)	0.387
Height (m)	1.6 ± 0.1	1.6 ± 0.7	0.296
Weight (kg)	55.5 ± 8.3	54.8 ± 8.0	0.640
BMI (kg/m ²)	22.3 ± 3.0	22.3 ± 2.9	1.000
WBC (× 10 ⁹ /L)	5.1 ± 1.5	4.6 ± 1.9	0.008
RBC (× 10 ⁹ /L)	4.0 ± 0.5	3.6 ± 0.6	<0.001
HGB (g/L)	122.1 ± 16.7	110.4 ± 19.2	<0.001
PLT (× 10 ⁹ /L)	173.2 ± 68.6	109.2 ± 56.8	<0.001
Monocyte (× 10 ⁹ /L)	0.4 ± 0.2	0.4 ± 0.2	0.466
Lymphocyte (× 10 ⁹ /L)	1.6 ± 0.7	2.3 ± 8.0	<0.001
Neutrophil (× 10 ⁹ /L)	2.9 ± 1.1	2.9 ± 1.6	0.204
RDW-CV (%)	14.7 ± 2.6	16.7 ± 3.2	<0.001
RDW-SD (%)	47.7 ± 7.2	55.3 ± 10.2	<0.001
HCT (%)	37.1 ± 4.7	33.4 ± 5.2	<0.001
MPV (fL)	11.5 ± 1.4	11.5 ± 1.4	0.851
PDW (fL)	15.5 ± 2.4	16.1 ± 2.5	0.151
ALT (U/L)	193.1 ± 305.3	170.7 ± 229.9	0.603
AST (U/L)	191.4 ± 269.6	248.8 ± 295.0	0.080
ALP (U/L)	161.9 ± 86.1	175.1 ± 124.1	0.726
GGT (U/L)	153.5 ± 111.7	138.9 ± 95.2	0.383
TBIL (μmol/L)	56.7 ± 90.1	104.1 ± 108.4	<0.001
TBA (μmol/L)	58.3 ± 81.0	115.7 ± 94.4	<0.001
ALB (g/L)	39.5 ± 5.9	32.7 ± 6.2	<0.001
PT (sec)	11.3 ± 2.0	14.0 ± 4.1	<0.001
APTT (sec)	31.2 ± 7.3	37.0 ± 10.7	<0.001
UTST (mm)	36.2 ± 7.8	47.2 ± 11.3	<0.001
FIB-4	4.7 ± 5.1	11.0 ± 11.2	<0.001
AAR	1.4 ± 1.4	1.9 ± 1.8	<0.001
APRI	3.3 ± 6.8	6.4 ± 8.9	<0.001

AIH, autoimmune hepatitis; BMI, body mass index; WBC, white blood cells; RBC, red blood cells; HGB, hemoglobin; PLT, platelet count; RDW: red blood cell distribution width; HCT, hematocrit value; MPV, mean platelet volume; PDW, platelet distribution width; ALT: alanine aminotransferase; AST: aspartate aminotransferase; ALP: alkaline phosphatase; GGT: gamma-glutamyl transpeptidase; TBIL, total bilirubin; TBA, total bile acid; ALB, albumin; PT, prothrombin time; APTT, activated partial thromboplastin time; UTST, Ultrasonic spleen thickness; FIB-4, fibrosis index based on the four factors; AAR, aspartate aminotransferase to alanine aminotransferase ratio; APRI, aspartate aminotransferase to platelet ratio index.

3.3. Machine learning model construction and evaluation

The discovery cohort, consisting of 233 cases, was randomly divided into a training set (163 cases) and a testing set (70 cases). The training set was utilized for model construction, while the testing set served to validate the model. To ensure the robustness of the classifiers across the training and testing data, a ten-fold cross-validation method was adopted to calculate the diagnostic value for advanced fibrosis (≥ S3).

The cross-validation results revealed that the RF model exhibited superior performance with an AUC of 0.893, outperforming other models such as XGB with an AUC of 0.886, MLP with an AUC of 0.883, LR and SVM both with an AUC of 0.867, and DT with an AUC of 0.779. The cross-validation receiver operating characteristic (ROC) curves of the six machine learning models were illustrated in [Figure S1](#).

3.4. Machine learning model testing and validation

In the training set, models for RF, LR, MLP, SVM, DT, and XGB were created. The ML model demonstrating the highest diagnostic value was then compared with classical noninvasive scores. [Table S1](#) summarizes the diagnostic indicators of the six ML models, including accuracy, AUC, sensitivity, specificity, and recall. Among the ML models, the RF model exhibited superior performance in both the training ([Figure S2a](#)) and testing sets ([Fig. 3a](#)), with AUCs of 0.951 and 0.869,

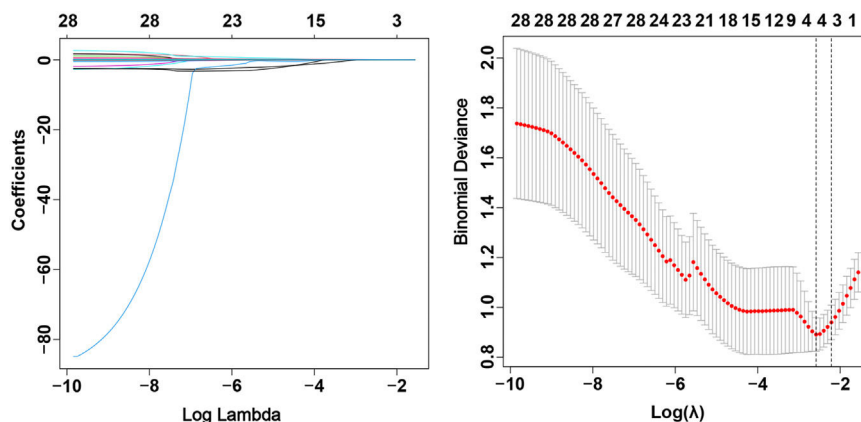


Fig. 2. Least absolute shrinkage and selection operator (LASSO) regression for candidate biomarker selection. LASSO regression analysis was utilized to identify key clinical features associated with advanced liver fibrosis in patients with autoimmune hepatitis (AIH). The plot displayed the regression coefficients, indicating the significance and direction of the association between each candidate biomarker and the fibrosis stage.

respectively. Calibration curve and decision curve analyses further supported the clinical utility of the RF model (Figure S3). As expected, the RF model also demonstrated better efficiency than traditional noninvasive predictive models. In the testing set, the AUCs of FIB-4, APRI, and AAR were 0.775, 0.669, and 0.765, respectively, although the differences did not reach statistical significance (Fig. 3b-d). The accuracy of the RF model was further validated in an external validation cohort, where the AUC of the model was 0.843 (Fig. 4), indicating its considerable and stable diagnostic accuracy across different datasets.

3.5. Machine learning model explanation and visualization

The SHapley Additive exPlanations (SHAP) analysis illustrated the performance of each indicator that constituted the RF model in predicting advanced fibrosis. UTST and PT were positively associated with advanced fibrosis, whereas ALB was negatively associated with advanced fibrosis (Figure S4). To assess the practical application of the RF model, two randomly selected patients were evaluated to determine if the model could accurately distinguish between those with or without advanced fibrosis (Figure S5a-b). One patient, with ALB of 30.39 g/L, PT of 10.6 s, and UTST of 60 mm, was diagnosed as positive by the RF model. Another patient, with ALB of 43.22 g/L, PT of 12.5 s, and UTST of 43 mm, was diagnosed as negative. In both cases, the RF model correctly diagnosed the patients. To enhance the accessibility of the RF model in clinical practice, it was transformed into a web calculator, allowing doctors to easily obtain the probability of advanced fibrosis by inputting numerical values for ALB, PT, and UTST. The web calculator can be accessed at <https://yingpeng.shinyapps.io/shiny/>.

4. Discussion

This study represents a groundbreaking endeavor as the first to employ ML methods in evaluating advanced fibrosis in patients with AIH. By exploring various ML methods for predicting advanced fibrosis in AIH, the study has showcased promising performance compared to established noninvasive predictors such as FIB-4, AAR, and APRI. Notably, the RF model emerged as the most effective in predicting advanced fibrosis. The diagnostic robustness of the RF model was validated in an additional cohort, reinforcing its reliability. To enhance practical applicability for clinicians, the RF model was transformed into a user-friendly web calculator, offering convenient access for healthcare professionals. This innovative approach not

only broadens our understanding of ML applications in AIH but also provides a potentially valuable tool for more accurate and efficient prediction of advanced liver fibrosis in AIH patients.

The RF model, incorporating UTST, ALB, and PT, exhibited the best performance in the prediction of advanced fibrosis. Originally designed to predict esophageal varices in cirrhosis, UTST has recently been recognized for its significant role in predicting liver fibrosis. Previous studies have supported the predictive value of spleen thickness for significant fibrosis, even in patients with persistently normal or slightly elevated levels of ALT [18]. Sheptulina et al. found UTST to be a significant prognosticator of advanced fibrosis in AIH, with a sensitivity of 72.7 % and a specificity of 80 % [19]. Advanced fibrosis and liver cirrhosis can result in metabolic and synthetic dysfunctions, potentially elevating bilirubin levels and decreasing coagulation factors and thrombopoietin. PT prolongation, dependent on coagulation factors synthesized by the liver, is positively related to liver function deterioration. Monitoring PT is often necessary in patients with liver dysfunction, providing a cost-effective and easily available test that generally reflects bleeding risk [20]. Boursier et al. established a model comprising age, gender, GGT, AST, PLT, and PT based on accessible clinical indicators, demonstrating preferable diagnostic ability for advanced fibrosis in chronic liver diseases [21]. In a recent study, ALB was identified as the only independent risk factor in predicting the severity of NAFLD [22]. Olteanu et al. also observed a correlation between decreased ALB levels and advanced fibrosis [23]. This underscores the significance of ALB as a potential biomarker for assessing liver disease severity and fibrosis progression in various liver conditions, including NAFLD and chronic liver diseases.

Indeed, the rapid advancement of ML has significantly broadened its applications in liver diseases, offering valuable insights and tools across various aspects of liver health. This includes predicting the severity of fibrosis in hepatitis B and C virus infections, as well as assessing complications and outcomes post-liver transplantation [24–26]. However, risk stratification of liver fibrosis in non-infectious liver diseases, particularly in cases of AIH, presents a formidable challenge [27]. The diagnostic process for AIH is intricate due to the absence of reliable biomarkers, necessitating alternative approaches for accurate assessment [28]. In this context, ML methods have emerged as valuable tools, although their application depends on factors like sample size and algorithm selection. The accuracy of ML methods is closely linked to the size of the dataset. Given this limitation, identifying the optimal ML model for a small sample size becomes particularly crucial. Recent studies have explored different ML methods to enhance liver disease diagnosis and create predictive

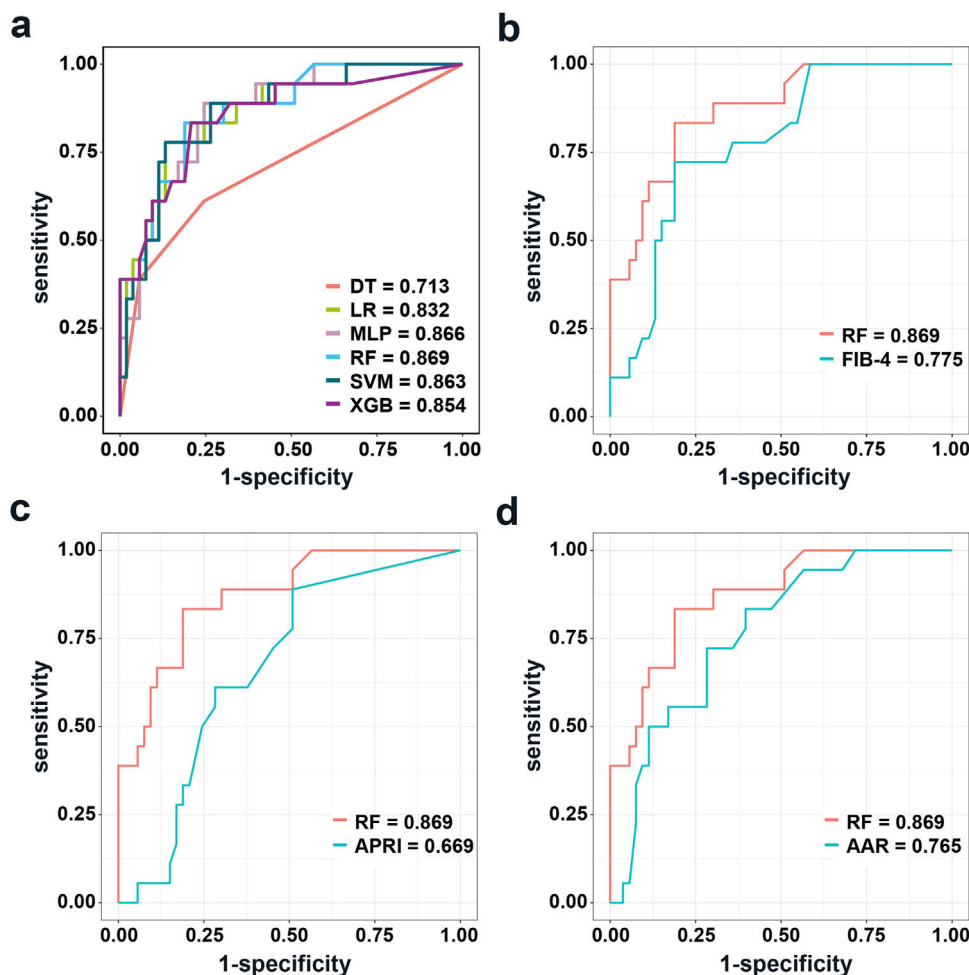


Fig. 3. The area under the receiver operating characteristic curve (AUC) of predictive models for advanced fibrosis in the testing cohort. The AUC values for various machine learning models were presented, demonstrating their predictive performance in identifying advanced liver fibrosis in the testing cohort of AIH patients. The random forest model showed superior diagnostic accuracy than the other machine learning models.

models for liver fibrosis stratification [29,30]. Prior research has highlighted the efficacy of two ML methods, RF and SVM, across various parameter categories. Consequently, we employed six models, including RF and SVM, to estimate the severity of liver fibrosis in AIH. Our study demonstrated that RF emerged as the optimal algorithm for small sample data. RF's strength lies in combining predictions from multiple weak classifiers, yielding a more accurate and stable prediction. Furthermore, its utilization of random samples and features ensures resilience to data noise, even with minimal tuning parameters and a small sample size [31,32]. This emphasizes the critical importance of selecting an algorithm tailored to the dataset's characteristics, particularly when dealing with limited samples in complex conditions like AIH.

In prior studies, RDW-CV and RDW-SD have been utilized to characterize the distribution of erythrocyte width, rather than RDW alone. AIH patients, especially those with advanced fibrosis, exhibit an increased susceptibility to hemolytic anemia [33]. The elevation of RDW-CV and RDW-SD in AIH patients may be attributed to the inhibition of erythrocyte maturation by pro-inflammatory cytokines. Furthermore, liver dysfunction and secondary malnutrition in AIH patients can result in a deficiency of hematopoietic substances such as iron, allowing many immature red blood cells to enter the peripheral circulation. Additionally, prior study observed a close relationship between RDW and the levels of inflammatory cytokines, which

were higher in AIH patients [34]. These findings suggest a complex interplay between liver function, erythrocyte indices, and inflammatory processes in AIH. In line with numerous prior studies, the current study also reports a significant association between a low platelet count (PLT) and advanced liver fibrosis in AIH patients [35,36].

Inevitably, there were limitations in this study. Firstly, it was a retrospective study, introducing potential selective bias. And there was a high dropout rate due to limited number of patients received liver biopsy. However, we have performed cross-validation to enhanced the analysis of model robustness. Future studies incorporated more patients may validate the reliability of our model. Secondly, the study did not differentiate between different subtypes of AIH, potentially introducing bias into the results. Thirdly, PT was consisted of our RF model instead of INR, which may limit its widely use across different centers. However, PT may directly reflect the coagulation dysfunction in patients with liver disease. Lastly, while TE is considered as a main noninvasive image method for staging fibrosis, it is not included in our study. The study population and environment in our research may not have had widespread access to TE, which is a significant reason for its exclusion from our analysis. Future iterations of our model may incorporate TE measurements to enhance its predictive performance, particularly in cases where there is discordance between clinical parameters and the suspected severity of fibrosis. Despite these

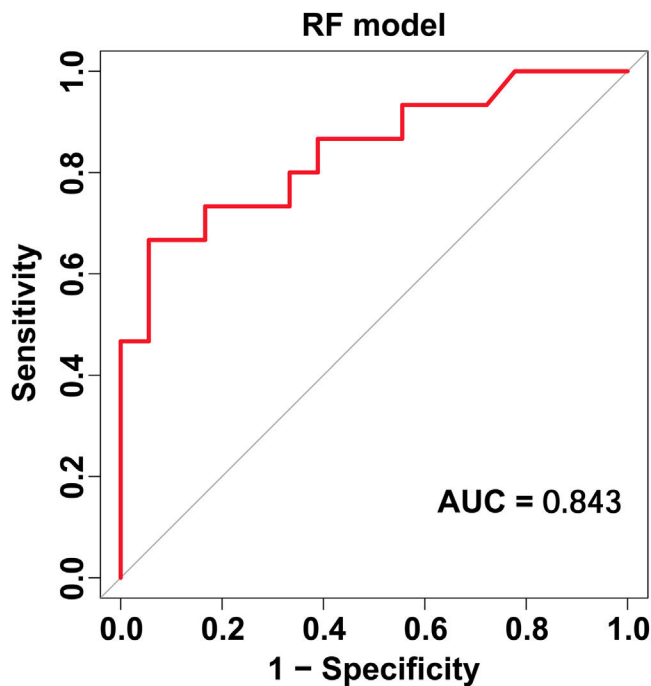


Fig. 4. AUC of random forest model for advanced fibrosis in the validation cohort. The predictive accuracy of the random forest model in an independent validation cohort was evaluated, with the AUC indicating its robustness in predicting advanced liver fibrosis in AIH patients.

limitations, our study constitutes a practical application of ML to address longstanding questions in the field, offering novel perspectives on the challenging issue of liver fibrosis stratification in AIH.

5. Conclusions

This study illustrates that ML methods, especially the RF model, have the potential to enhance the prediction of advanced fibrosis in AIH. The utilization of this model allows for individualized predictions of fibrosis in AIH patients while minimizing evaluation costs. It is advisable for risk stratification and the implementation of suitable preventive measures in high-risk AIH patients. Clinicians are encouraged to advocate for the adoption of this model, which holds promise for optimizing the noninvasive assessment of liver fibrosis on a larger scale.

Author contributions

QW: designed the study, analyzed and interpreted the data and drafted the manuscript; QX, WL and HW: collected the data, searched and selected the literature; YP and SH, analyzed the data; YP and XZ: revised the manuscript and supervised the study. All authors approved the submission.

Declaration of interests

None.

Data availability statement

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding authors.

Funding statement

This study is funded by the NSFC (82103710) and the Youth Independent Innovation Science Fund Project of PLA General Hospital (22QNFC065).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.aohep.2024.101754.

References

- [1] Mieli-Vergani G, Vergani D, Czaja AJ, et al. Autoimmune hepatitis. *Nat Rev Dis Primers* 2018;4:18017.
- [2] Wu S, Yang Z, Zhou J, et al. Systematic review: diagnostic accuracy of non-invasive tests for staging liver fibrosis in autoimmune hepatitis. *Hepatol Int* 2019;13:91–101.
- [3] Ohira H, Takahashi A, Zeniya M, et al. Clinical practice guidelines for autoimmune hepatitis. *Hepatol Res* 2022;52:571–85.
- [4] Chinese Society of Hepatology CMA. [Guidelines on the diagnosis and management of autoimmune hepatitis (2021)]. *Zhonghua Gan Zang Bing Za Zhi* 2022;30:482–92.
- [5] Soloway RD, Baggenstoss AH, Schoenfield LJ, Summerskill WH. Observer error and sampling variability tested in evaluation of hepatitis and cirrhosis by liver biopsy. *Am J Dig Dis* 1971;16:1082–6.
- [6] Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. American Association for the Study of Liver D. Liver biopsy. *Hepatology* 2009;49:1017–44.
- [7] Regev A, Berho M, Jeffers LJ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol* 2002;97:2614–8.
- [8] Sandrin L, Fourquet B, Hasquenoph JM, et al. Transient elastography: a new non-invasive method for assessment of hepatic fibrosis. *Ultrasound Med Biol* 2003;29:1705–13.
- [9] Boursier J, Zarski JP, de Ledinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology (Baltimore, Md)* 2013;57:1182–91.
- [10] Sagir A, Erhardt A, Schmitt M, Haussinger D. Transient elastography is unreliable for detection of cirrhosis in patients with acute liver damage. *Hepatology* 2008;47:592–5.
- [11] Yuan X, Duan SZ, Cao J, Gao N, Xu J, Zhang L. Noninvasive inflammatory markers for assessing liver fibrosis stage in autoimmune hepatitis patients. *Eur J Gastroenterol Hepatol* 2019;31:1467–74.
- [12] Abdollahi M, Pouri A, Ghojzadeh M, Estakhri R, Somi M. Non-invasive serum fibrosis markers: a study in chronic hepatitis. *Biol Impacts: BI* 2015;5:17–23.
- [13] Olisah CC, Smith L, Smith M. Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Comput Methods Programs Biomed* 2022;220:106773.
- [14] Kim HY, Lampertico P, Nam JY, et al. An artificial intelligence model to predict hepatocellular carcinoma risk in Korean and Caucasian patients with chronic hepatitis B. *J Hepatol* 2022;76:311–8.
- [15] Papatheodoridis GV, Sypsa V, Dalekos GN, et al. Hepatocellular carcinoma prediction beyond year 5 of oral therapy in a large cohort of Caucasian patients with chronic hepatitis B. *J Hepatol* 2020;72:1088–96.
- [16] Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 2022;23:40–55.
- [17] Scheuer PJ, Standish RA, Dhillon AP. Scoring of chronic hepatitis. *Clin Liver Dis* 2002;6:335–47 v-vi.
- [18] Zhang J, Du X, Zhou Z, Lv F, Yu Y. Spleen thickness can predict significant liver pathology in patients with chronic hepatitis B with persistently normal alanine aminotransferase or minimally raised alanine aminotransferase: a retrospective study. *J Int Med Res* 2019;47:122–32.
- [19] Sheptulina A, Shirokova E, Nekrasova T, Blum H, Ivashkin V. Platelet count to spleen diameter ratio non-invasively identifies severe fibrosis and cirrhosis in patients with autoimmune hepatitis. *J Gastroenterol Hepatol* 2016;31:1956–62.
- [20] O'Leary JG, Greenberg CS, Patton HM, Caldwell SH. AGA clinical practice update: coagulation in cirrhosis. *Gastroenterology* 2019;157:34–43 e1.
- [21] Boursier J, de Ledinghen V, Leroy V, et al. A stepwise algorithm using an at-a-glance first-line test for the non-invasive diagnosis of advanced liver fibrosis and cirrhosis. *J Hepatol* 2017;66:1158–65.
- [22] Kawaguchi K, Sakai Y, Terashima T, et al. Decline in serum albumin concentration is a predictor of serious events in nonalcoholic fatty liver disease. *Medicine* 2021;100:e26835.
- [23] Olteanu VA, Balan GG, Timofte O, et al. Risk Predictors of Advanced Fibrosis in Non-Alcoholic Fatty Liver Disease. *Diagnostics (Basel)* 2022;12.
- [24] Wang K, Lu X, Zhou H, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut* 2019;68:729–41.
- [25] Wei R, Wang J, Wang X, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. *EBioMedicine* 2018;35:124–32.

- [26] Bertsimas D, Kung J, Trichakis N, Wang Y, Hirose R, Vagefi PA. Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation. *Am J Transplant* 2019;19:1109–18.
- [27] Bossen L, Gerussi A, Lygoura V, Mells GF, Carbone M, Invernizzi P. Support of precision medicine through risk-stratification in autoimmune liver diseases - histology, scoring systems, and non-invasive markers. *Autoimmun Rev* 2018;17:854–65.
- [28] Sahebjam F, Vierling JM. Autoimmune hepatitis. *Front Med* 2015;9:187–219.
- [29] Wang K, Li Y, Pan J, et al. Noninvasive diagnosis of AIH/PBC overlap syndrome based on prediction models. *Open Med (Wars)* 2022;17:1550–8.
- [30] Eaton JE, Vesterhus M, McCauley BM, et al. Primary Sclerosing Cholangitis Risk Estimate Tool (PREsTo) Predicts Outcomes of the Disease: a Derivation and Validation Study Using Machine Learning. *Hepatology* 2020;71:214–24.
- [31] Agosta F, Ferraro PM, Canu E, et al. Differentiation between Subtypes of Primary Progressive Aphasia by Using Cortical Thickness and Diffusion-Tensor MR Imaging Measures. *Radiology* 2015;276:219–27.
- [32] Preziosa P, Rocca MA, Mesaros S, et al. Relationship between damage to the cerebellar peduncles and clinical disability in multiple sclerosis. *Radiology* 2014;271:822–30.
- [33] Liu L, Cao J, Zhong Z, et al. Noninvasive indicators predict advanced liver fibrosis in autoimmune hepatitis patients. *J Clin Lab Anal* 2019;33:e22922.
- [34] Allen LA, Felker GM, Mehra MR, et al. Validation and potential mechanisms of red cell distribution width as a prognostic marker in heart failure. *J Card Fail* 2010;16:230–8.
- [35] Hui AY, Chan HL, Wong VW, et al. Identification of chronic hepatitis B patients without significant liver fibrosis by a simple noninvasive predictive model. *Am J Gastroenterol* 2005;100:616–23.
- [36] Ding R, Zhou X, Huang D, et al. Nomogram for predicting advanced liver fibrosis and cirrhosis in patients with chronic liver disease. *BMC Gastroenterol* 2021;21:190.