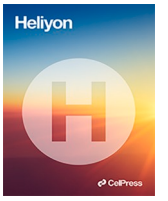
Heliyon 10 (2024) e38848

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

A machine learning-based predictive model discriminates nonalcoholic steatohepatitis from nonalcoholic fatty liver disease

Yuqi Yan [a,1], Danhui Gan [b,1], Ping Zhang [a], Haizhu Zou [a], MinMin Li [a,*]

[a] Department of Clinical Laboratory Medicine, The First Affiliated Hospital of Jinan University, Guangzhou, 510630, China
[b] Department of Clinical Pathology, The First Affiliated Hospital of Jinan University, Guangzhou, 510632, China

ARTICLE INFO

Keywords:
NAFLD
Machine learning
NASH
Non-invasive tests (NITs)

ABSTRACT

Background: Non-alcoholic fatty liver disease (NAFLD) is a leading cause of liver-related morbidity and mortality. The diagnosis of non-alcoholic steatohepatitis (NASH) plays a crucial role in the management of NAFLD patients.
Objective: The aim of our observational study was to build a machine learning model to identify NASH in NAFLD patients.
Methods: The clinical characteristics of 259 NAFLD patients and their initial laboratory data (Cohort 1) were collected to train the model and carry out internal validation. We compared the models built by five machine learning algorithms and screened out the best models. Receiver operating characteristic (ROC) curves, sensitivity, specificity, and accuracy were used to evaluate the performance of the model. In addition, the NAFLD patients in Cohort 2 (n = 181) were externally verified.
Results: We finally identified six independent risk factors for predicting NASH, including neutrophil percentage (NEU%), aspartate aminotransferase/alanine aminotransferase (AST/ALT), hematocrit (HCT), creatinine (CREA), uric acid (UA), and prealbumin (PA). The NASH-XGB6 model built using the XGBoost algorithm showed sufficient prediction accuracy, with ROC values of 0.95 (95 % CI, 0.91–0.98) and 0.90 (95 % CI, 0.88–0.93) in Cohort 1 and Cohort 2, respectively.
Conclusions: NASH-XGB6 can serve as an effective tool for distinguishing NASH patients from NAFLD patients.

1. Introduction

The prevalence of nonalcoholic fatty liver disease (NAFLD), a leading cause of chronic liver disease, is approximately 25 % globally, with an estimated 3.6 million new cases emerging every year [1]. NAFLD encompasses a range of chronic conditions associated with obesity, insulin resistance-related type 2 diabetes mellitus (T2DM), and/or metabolic syndromes, including nonalcoholic fatty liver (NAFL), nonalcoholic steatohepatitis (NASH), nonalcoholic fatty liver-associated fibrosis, cirrhosis, and liver cancer [2]. NASH, an active form of NAFLD, is associated with an increased incidence of adverse liver-related clinical outcomes and mortality [3]. Identifying individuals at risk of developing NASH is the cornerstone of clinical care and interventions [4].

* Corresponding author.
E-mail address: limm269@126.com (M. Li).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e38848
Received 14 March 2024; Received in revised form 31 August 2024; Accepted 1 October 2024
Available online 2 October 2024

Currently, liver biopsy remains the "gold standard" for diagnosing NASH. Although generally safe, liver biopsy is not risk-free, often unappealing to patients, and limited by sampling variations and inter- and intra-observer variability among pathologists [5]. Importantly, given the vast number of individuals at risk of NAFLD, performing liver biopsies on such a large population is impractical, making it an ineffective tool for identifying patients with NASH. There is an urgent need for noninvasive biomarkers to support clinical care and facilitate the evaluation of new therapies. Several noninvasive biomarkers have been proposed in previous studies. Many studies have attempted to identify new single variables to aid in the prediction of NASH, including cytokeratin-18 (CK18) [6], IL-8 [7], and miR122 [8]; however, their diagnostic efficacy is not ideal. Multi-index serum biomarker models have also been used in previous studies because a single factor is less accurate for assessing NASH. FIB-4 [9] and NFS [10] are the most widely used algorithms and both have high negative predictive values. However, the FIB-4 and NFS generally identify only patients with moderate-to-severe NASH-associated fibrosis, and both assays have large regions of uncertainty. Therefore, the development of simple, easily accessible, and proven noninvasive tests for early NASH diagnosis in the management of patients with NAFLD is essential.

Machine learning is an emerging method for data calculations and statistics that can be used to analyze medical data in multiple dimensions [11]. Compared with traditional statistical modeling, machine learning has the advantage of recognizing unique patterns and incorporating multiple factors to create predictive models. Given the heterogeneity, complexity, and overlapping confounding factors of liver diseases, machine learning is particularly suitable for applications in this field. Multiple studies have used ML technology in conjunction with liver imaging, histological imaging, or other omics data to predict NAFLD. Schawkat et al. achieved an impressive overall accuracy of 85.7 % using a support vector machine (SVM) to predict hepatic fibrosis in chronic liver disease based on MRI images of the liver [12]. Similarly, Teramoto et al. demonstrated an AUC of 0.95 with a convolutional neural network (CNN) to predict NASH from histopathology images [13]. Forlano et al. exhibited strong interclass correlation coefficients for steatosis (0.97), inflammation (0.96), ballooning (0.94), and fibrosis (0.92) between manual and software analysis, utilizing AI-based second-harmonic generation (SGH) technology for NASH prediction using histopathology images [14]. Oh et al. achieved an AUC of 0.91 with a random forest (RF) to predict NAFLD cirrhosis based on 19 distinct species from gut microbiome signatures [15]. However, most studies rely on costly examinations, such as MRI or microbiome sequencing, or fail to eliminate the need for liver biopsy, highlighting the need for more accessible and accurate diagnostic tools.

In this study, we developed an ML model to distinguish patients with NASH from those with NAFLD. We tested multiple ML models, including extreme gradient boosting (XGBoost), logistic regression (LR), light gradient boosting machine (LightGBM), random forest (RF), and adaptive boosting (AdaBoost), to screen the optimal algorithm for the effective diagnosis of NASH. Furthermore, we used the differences in the temporal dimension to collect the latest two-year clinical data as an externally validated dataset. We hypothesized that the ML model could partially replace liver biopsy for the diagnosis of patients with NASH.

2. Materials and methods

2.1. Study cohort

This study included two distinct cohorts of patients recruited from the First Affiliated Hospital of Jinan University. Cohort 1 was the primary cohort for model development and internal validation. This study included patients diagnosed with NAFLD between January 2019 and December 2021. These patients were further categorized into NAFL and NASH subgroups based on the liver biopsy results, facilitating the establishment of a machine learning-based diagnostic model for NASH. Furthermore, Cohort 2 represented the external validation cohort, comprising patients diagnosed with NAFLD from January 2022 to December 2023. This cohort underwent the same processing procedures as Cohort 1 to validate the diagnostic model externally. The study protocol was approved by the Medical Research Ethics Committee of the First Affiliated Hospital of Jinan University (KY-2023-366) and strictly adhered to the ethical standards outlined in the Declaration of Helsinki. All methodological approaches were implemented in accordance with approved guidelines.

NAFLD was diagnosed according to the 2017 Asia-Pacific Working Group guidelines [16]. The inclusion criteria were as follows: (1) no history of excessive alcohol consumption. Excessive alcohol consumption refers to drinking for more than 2 years before liver tissue examination and drinking >21 standard units per week in men and >14 standard units per week in women. One standard unit constitutes approximately 14 g of pure ethanol [17]. (2) Histological changes in the liver biopsy meet the pathological diagnostic criteria for NAFLD. Exclusion criteria: (1) all chronic liver diseases known to cause fatty liver diseases, such as viral hepatitis (especially genotype 3 hepatitis C), Wilson disease, hemochromatosis, and autoimmune hepatitis, and (2) all secondary causes of hepatic steatosis, such as drug factors (including tamoxifen, amiodarone, glucocorticoid, synthetic estrogen, methotrexate, and antiviral drugs), metabolic or genetic factors (including β-lipoprotein deficiency, lysosomal acid lipase deficiency, and fat metabolism disorders), nutritional factors (poor nutrition, malabsorption, total parenteral nutrition, rapid weight loss, and jejunoileal bypass), and other special circumstances (such as small intestinal diverticula, petrochemical product exposure, and organic solvent exposure).

2.2. Liver biopsy

The acquisition of histopathological specimens strictly followed the standard operating procedure (SOP). To minimize differences in histopathological evaluation, pathology reading involved central reading with a double-blind reading performed by two or more liver pathologists according to the fatty liver inhibition of progression algorithm (FLIP-SAF) [18]. This method separately assesses the steatosis, mobility, and fibrosis stages. Steatosis was assessed based on the percentage of hepatocytes containing large- and medium-sized cytoplasmic lipid droplets, ranging from grade 0 to 3: S0, <5 %; S1, 5%–33 %; S2, 34%–66 %; and S3, >67 %. Activity

included lobular inflammation and ballooning. Lobular inflammation was defined as a 20-fold microscopic count of hepatic inflammatory necrosis: 0—none, 1—1–2 foci of inflammatory necrosis in each liver lobule, and 2–2 or more in each liver lobule. The ballooning of hepatocytes varied from 0 to 2:0, normal hepatocytes were vertical with acute angles and pink eosinophilic cytoplasm; 1, there were aggregated clusters of hepatocytes with cytoplasmic cavitation (black and white), a grid shape, and differences in size, similar to hepatocytes; 2, on a one-point basis, there was at least one enlarged balloon-like cell in the field, and at least one cluster of the one-point balloon-like hepatocytes was twice the size of the other balloon-like cells. Liver fibrosis was graded from 0 to 4 as follows: 0, no fibrosis; 1, perisinus or periportal fibrosis; 2, perisinusoidal fibrosis with periportal fibrosis; 3, bridging fibrosis; 4, cirrhosis. NAFLD, according to histological changes in the liver, is divided into NAFL and NASH: (1) NAFL-hepatocyte fat change >5 %, without hepatocyte ballooning or lobular inflammation, with or without fibrosis. (2) NASH: hepatocyte fat change >5 % accompanied by ballooning and lobular inflammation with or without fibrosis.

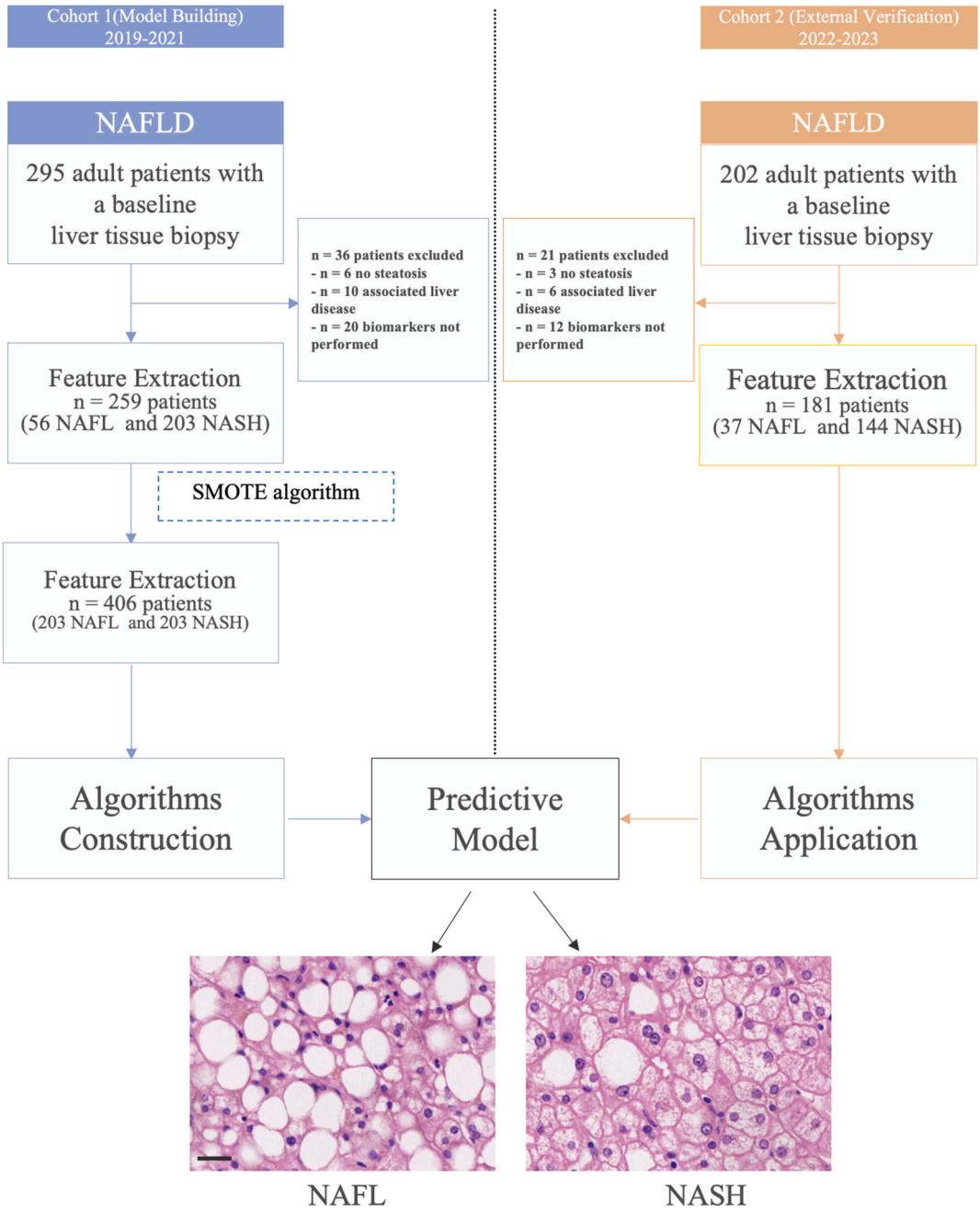


Fig. 1. Flow diagram illustrating the process of the machine learning algorithm employed in developing the predictive model. NAFLD: nonalcoholic fatty liver disease; NAFL, nonalcoholic fatty liver; NASH, nonalcoholic steatohepatitis. Scale bars = 25 μm.

2.3. Detection of clinical indicators and laboratory indicators

Age and sex were recorded for all patients, and basic measures of height, weight, systolic blood pressure (SBP), and diastolic blood pressure (DBP) were recorded following standard clinical procedures. Fasting blood samples were collected to analyze routine blood and biochemical variables. The indicators tested included 23 hematological measures: red blood cells (RBCs), the coefficient of variation of the width of the RBC distribution (RDW-CV), the standard deviation of the RBC distribution width (RDW-SD), hematocrit (HCT), hemoglobin (HGB), mean BC hemoglobin content (MCH), mean red blood cell hemoglobin concentration (MCHC), mean red blood cell volume (MCV), white blood cells (WBCs), neutrophils (NEU#), percentage of neutrophils (NEU%), lymphocytes (LYM#), lymphocytes (LYM%), monocytes (MONO#), percentage of monocytes (MONO%), basophiles (BASO#), percentage of basophiles (BASO%), eosinophils (EOS#), percentage of eosinophils (EOS%), platelet (PLT), mean platelet volume (MPV), NLR (neutrophil/lymphocyte), and dNLR (neutrophil/leukocyte–neutrophil), and 47 serological parameters: anionic gap (AG), albumin (ALB), globulins (GLB), A/G (albumin/globulin), prealbumin (PA), total protein (TP), alkaline phosphatase (ALP), aspartate aminotransferase (AST), alanine aminotransferase (ALT), AST/ALT (aspartate aminotransferase/alanine aminotransferase), amylase (AMY), apolipoprotein A (APOA), apolipoprotein B (APOB), APOA/APOB (apolipoprotein A/apolipoprotein B), ceruloplasmin cholinesterase (CHE), total cholesterol (TCHOL), triglyceride (TG), creatine kinase (CK), creatine kinase isozyme (CKMB), creatinine (CREA), cystatin C (Cys-C), combined with bilirubin (DBL), nonbound bilirubin (IBIL), total bilirubin (TBL), iron (Fe), γ-glutamyl transferase (GGT), glucose (GLU), glycated serum protein (GSP), α-hydroxybutyrate dehydrogenase (HBDH), high-density lipoprotein (HDL), high-density lipoprotein C (HDL-C), low-density lipoprotein C (LDL-C), lactate dehydrogenase (LDH), lipoprotein A (LPa), sodium, potassium, chlorine, magnesium, total carbon dioxide (CO_2), osmotic pressure (OSM), phosphorus, zinc, calcium, total bile acid (TBA), transferrin (TRF), uric acid (UA), and urea (UREA). All routine blood variables were measured using an UniCel DxH 800 (Beckman Coulter, California, USA). All biochemical variables were measured using a Hitachi 7600 autoanalyzer (Hitachi).

2.4. Statistical analysis

Statistical analyses were conducted using R software (version 3.6.2), SPSS 22.0, and the Beckman Coulter DxAI platform (w ww. xsm artanalysisbeckman/login/). Data with more than 30 % missing data were excluded from the analysis. Based on the characteristics of the data distribution, methods that could represent the central tendency of the variables, such as the median or average, were selected to fill in missing values. Continuous variables were represented either as the mean ± standard deviation (SD) or the median with the interquartile range (IQR). We employed Student's t-test, Mann–Whitney *U* test, ANOVA, or Kruskal–Wallis H-test to compare variable distributions across groups. A confusion matrix was used to calculate model evaluation indices. Categorical variables were expressed as composition ratios and were compared between groups using Pearson's χ^2 test or Fisher's exact test. Receiver operating characteristic (ROC) curves were used to assess the diagnostic value of the model. The areas under the ROC curves (AUC) were compared using DeLong's test. A calibration curve was used to assess the agreement between model predictions and sample probabilities. Decision curve analysis (DCA) was used to evaluate the clinical utility of the model. The interpretation of feature ranking was facilitated by Shapley additive explanation (SHAP) plots. Statistical significance was set at $P < 0.05$.

2.5. Model development and evaluation of machine learning

A total of 75 features per subject were included in the analysis, including nearly all common clinical indicators, including 5 demographic, 23 hematological, and 47 serological measures. Liver biopsy results were used as dependent variables. A detailed flowchart depicting the stepwise process of model development using a machine learning algorithm is presented in Fig. 1. The feature selection involved three steps. First, all differential features were compared with baseline data. Second, the Least Absolute Shrinkage and Selection Operator (LASSO) method was used to select the most useful differential features from the primary dataset [19]. Finally, the selected differential features were included in the model, and the final included indicators were constantly optimized according to the weight ranking of the features.

Cohort 1 employed a 5-fold cross-validation method, in which the entire dataset was partitioned into five folds. In each iteration, four folds served as the training dataset for model development, whereas the remaining fold functioned as an internal validation set for assessing model performance. This process was repeated five times to ensure a comprehensive evaluation. We compared the predictive models built using extreme XGBoost [20], LR [21], LightGBM [22], RF [23], and AdaBoost [24] and identified the most effective machine learning algorithms for predicting NASH patients within the validation set. To optimize the model performance, hyperparameters such as the number of estimators and criteria were carefully fine-tuned. The hyperparameters were tuned using a grid search. This involved systematically exploring a predefined set of hyperparameters, testing all possible combinations, and selecting the combination that yielded the best model performance according to the selected evaluation metric. The model generated a score that indicated the likelihood of each patient belonging to the NASH group. During the grid search process, we conducted extensive searches on multiple hyperparameters to determine the optimal model configuration. These parameters include the optimization objective function (binary:logistic or binary:logitraw), learning rate (ranging from 0 to 1), maximum tree depth (ranging from 0 to infinity), minimum split weight sum (ranging from 0 to infinity), and L2 regularization coefficient (ranging from 0 to infinity). The best estimator is the model instance that performs the best based on cross-validation results automatically selected after the grid search is completed. The specific parameter configuration of this model is as follows: objective (binary:logistic), learning rate (0.1), maximum tree depth (6), minimum split weight sum (1), and L2 regularization coefficient (1). These parameter values were found through grid search within the search range and enabled the model to achieve the best generalization performance on the validation set. Model

performance was comprehensively assessed using metrics such as the area under the curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV). Additionally, Cohort 2 served as an independent dataset for external validation of the model, ensuring its generalizability and reliability.

3. Results

3.1. Patient characteristics

Cohort 1 (for model building and internal validation) included 259 patients with NAFLD, including 56 with NAFL and 203 with NASH (Fig. 1). A total of 75 features per subject were included in the analysis, including 5 demographic measures, 23 hematological measures, and 47 serological measures. Because of the low event rate of NAFL, which leads to an imbalanced dataset, we used the SMOTE algorithm to balance the data from the study cohort [25]. In cohort 1, after testing for normality and homogeneity of variance, we chose the Mann–Whitney *U* test or Welch's *t*-test based on the results to compare the parameters of the two groups. Patients with NASH were typically older, female, and had higher values for BMI, DBP, AG, ALP, ALT, AST, AST/ALT, APOB, CYSC, GGT, GLU, CREA, GSP, NEU%, MCHC, MONO#, TG, UA, NLR, and dNLR, and lower values for A/G, AMY, APOA/APOB, DBIL, HCT, HGB, LYM#, PA, TBIL, and UREA than subjects without NASH. Table 1 presents a detailed overview of the baseline clinical characteristics of the study population.

3.2. Establishment and evaluation of the NASH diagnostic model based on machine learning

3.2.1. Feature selection

The original model encompassed 75 variables, which were subsequently narrowed down to 25 potential predictors through a LASSO regression analysis conducted in Cohort 1. These predictors encompass a diverse array of metrics, including UA, TG, RDWCV, PA, OSM, NEU%, MONO, LYM, LPA, HCT, GSP, FE, DBIL, CREA, CKM, APOB, APOA, AMY, AST/ALT, AST, ALP, ALB, DBP, age, and sex. The numerical coefficients of these variables are listed in Supplementary Table 1, and their corresponding profiles are presented in Fig. 2. Specifically, Fig. 2A shows a cross-validated error plot of the LASSO regression model, offering insight into its predictive accuracy and stability. Moreover, Fig. 2B highlights the most regularized and parsimonious model, which achieved a cross-validated error within one standard error of the minimum and retained 25 variables, thereby emphasizing its efficiency and effectiveness in capturing the underlying patterns in the data.

3.2.2. Comparison of machine learning algorithms and the identification of the optimal model

The fivefold cross-validation results for the five machine learning algorithms incorporating 25 variables are presented in Table 2 and Fig. 3. In the test cohort, the results of the five machine learning algorithms showed AUCs of 0.93 (0.89–0.96) for XGBoost, 0.83 (0.80–0.86) for logistic regression, 0.92 (0.88–0.95) for LightGBM, 0.92 (0.89–0.95) for random forest, and 0.87 (0.82–0.91) for AdaBoost. The performances of the algorithms were further visualized using ROC curves (Fig. 3A), calibration curves (Fig. 3B), and decision curve analyses (Fig. 3C). Notably, XGBoost emerged as a superior algorithm, leading to the establishment of a NASH diagnostic model named NASH-XGB25. The importance of each predictive variable in the prediction results of the XGBoost model is determined using the SHAP algorithm. All 25 variables are listed in descending order of importance in the variable importance plot shown in Supplementary Fig. 1. Subsequently, we conducted feature screening of this model by selecting the top six variables with the highest feature weights for model building, named NASH-XGB6. Statistical analysis using the DeLong test revealed no significant difference in the AUC between the first six indicators and all other variables (Table 3 and Fig. 4). Consequently, NASH-XGB6 was selected as the final diagnostic model for NASH, incorporating six key features: NEU%, AST/ALT, HCT, CREA, UA, and PA.

3.2.3. Analysis and evaluation of the machine learning model

The internally validated ROC curve results demonstrated the remarkable classification capabilities of the NASH-XGB6 model in predicting NASH risk, exhibiting an AUC of 0.95 (95 % CI; 0.91–0.98) along with a sensitivity of 0.85 and a specificity of 0.93 (Fig. 5A). The calibration curve exhibited a strong concordance between the probabilities predicted by the model and actual outcomes (Fig. 5B). Furthermore, decision curve analysis revealed the substantial clinical utility of the model (Fig. 5C). Fig. 5D illustrates the correlation between the observed and SHAP values of the six most pertinent features included in our final model. As shown in Fig. 5E, the feature-ranked interpretation of the NASH-XGB6 model using the SHAP algorithm identified NEU%, AST/ALT, HCT, CREA, UA, and PA as the most significant predictors for NASH patients. In addition, an online tool for the NASH-XGB6 model was established to provide convenient access to the predictions. (http://www.xsmartanalysis.com/model/list/predict/model/html?mid=12752&symbol=817090gKfG367dE0dq52).

3.3. External verification

Cohort 2, which was dedicated to external validation, included 181 patients with NAFLD, including 37 with NAFL and 144 with NASH. As presented in Table 4 and Fig. 6A, the NASH-XGB6 model exhibited a remarkable and consistent diagnostic accuracy for NASH, achieving an AUC of 0.90 (95 % CI; 0.88–0.93) along with a sensitivity of 0.78 and specificity of 0.90. Furthermore, Fig. 6B visually demonstrates the strong alignment between the risks predicted by the model and the actual observed risks, indicating calibration. The decision curve analysis (DCA) curves shown in Fig. 6C confirm the substantial clinical benefits of the model in the external

Table 1
Baseline parameters and characteristics of all patients in the training set and validation sets.

Variable	Cohort 1 (model building and internal validation)				Cohort 2 (external validation)			
	Total (n = 406)	NAFL (n = 203)	NASH (n = 203)	P-value	Total (n = 181)	NAFL (n = 37)	NASH (n = 144)	P-value
Age (year)	28 (23–33)	26 (22–32)	29 (24–34)	0.003[b]	27 (23–32)	26 (22–30)	29 (24–34)	<0.001[c]
Gender								
Female (n %)	321 (79.06)	176 (86.70)	145 (71.43)	<0.001[c]	227 (78.82)	124 (86.11)	103 (71.53)	0.002[b]
Male (n%)	85 (20.94)	27 (13.30)	58 (28.57)		61 (21.18)	20 (13.89)	41 (28.47)	
BMI (kg/m^2)	37.01 (33.65–41.11)	36.19 (33.38–39.57)	37.72 (33.83–42.28)	0.013[b]	36.93 (33.84–40.42)	36.45 (33.58–38.86)	37.53 (34.29–42.00)	0.011[a]
SBP (mmHg)	126 (120–135)	125 (120–134)	126 (118–138)	0.518	126 (120–137)	126 (120–135)	126 (120–135)	0.914
DBP (mmHg)	79 (72–87)	77 (71–84)	81 (74–91)	<0.001[c]	80 (74–86)	78.00 (73–83)	81 (74–89)	0.005[b]
A/G	1.52 (1.39–1.63)	1.55 (1.42–1.64)	1.48 (1.36–1.62)	0.032[a]	1.51 (1.38–1.61)	1.52 (1.41–1.60)	1.47 (1.36–1.62)	0.321[a]
AG	11.42 (10.20–12.60)	11.20 (10.13–12.20)	11.80 (10.20–13.31)	0.003[b]	11.70 (10.20–12.53)	11.60 (10.25–12.18)	11.80 (10.10–13.40)	0.06
ALP (U/L)	75 (64–88)	72 (62–80)	79 (66–95)	<0.001[c]	74 (64–88)	71 (61–79)	79 (66–94)	<0.001[c]
ALT (U/L)	36 (23–60)	32 (21–50)	43 (28–80)	<0.001[c]	38. (25–63)	32 (20–50)	46 (28–84)	<0.001[c]
AST (U/L)	22 (18–31)	21 (17–26)	25 (18–42)	<0.001[c]	22 (18–34)	20 (17–28)	27 (18–43)	<0.001[c]
AST/ALT	1.54 (1.21–1.93)	1.41 (1.15–1.87)	1.65 (1.36–2)	<0.001[c]	1.59 (1.28–1.92)	1.45 (1.20–1.78)	1.67 (1.42–2.06)	<0.001[c]
AMY (U/L)	41 (34–50)	44 (37–54)	39 (30–47)	<0.001[c]	40 (33–50)	44 (36–53)	37 (30–45)	<0.001[c]
APOA/APOB	1.47 (1.26–1.79)	1.58 (1.29–1.86)	1.42 (1.20–1.74)	<0.001[c]	1.44 (1.25–1.77)	1.52 (1.32–1.78)	1.40 (1.19–1.74)	0.01[b]
APOB (g/L)	0.93 (0.78–1.07)	0.90 (0.77–1.02)	0.96 (0.82–1.12)	<0.001[c]	0.95 (0.80–1.07)	0.94 (0.77–1.03)	0.97 (0.83–1.12)	0.017[a]
CREA (umol/L)	63 (55–71)	61 (51–70)	65 (58–74)	<0.001[c]	63 ± 13	60 ± 14	65 ± 11	0.003[b]
CYSC (mg/L)	0.87 (0.80–0.97)	0.85 (0.79–0.92)	0.90 (0.81–1.01)	<0.001[c]	0.87 (0.79–0.95)	0.86 (0.79–0.91)	0.89 (0.80–1.00)	0.017[a]
DBIL (umol/L)	3.00 (2.10–4.20)	3.14 (2.20–4.40)	2.70 (1.90–4.00)	0.02[a]	3.01 (2.15–4.30)	3.34 (2.39–4.52)	2.70 (2.00–4.00)	0.01[b]
GGT (U/L)	33 (23–50)	30 (21–45)	35 (24–59)	<0.001[c]	32 (23–43)	29 (20–39)	36 (25–57)	<0.001[c]
GLU (mmol/L)	5.22 (4.86–5.73)	5.10 (4.73–5.41)	5.47 (4.95–6.41)	<0.001[c]	5.19 (4.82–5.97)	5.10 (4.66–5.49)	5.55 (4.93–6.47)	<0.001[c]
GSP (umol/L)	146 (133–16)	141 (132–154)	152 (134–173)	<0.001[c]	145 (133–16)	134 (132–154)	153 (134–173)	<0.001[c]
HCT (%)	42.27 (39.85–45.1)	42.91 (40.59–45.31)	41.35 (39.38–45.09)	0.007[b]	41.68 (39.50–45.0)	42.26 (39.79–45.50)	41.37 (39.44–45.00)	0.354
HGB (g/L)	138 (131–14)	139 (133–148)	138 (129–148)	0.046[a]	138 (129–14)	139 (129–149)	138 (129–149)	0.459
LYM# (10^9/L)	2.39 (2.12–2.80)	2.47 (2.12–2.98)	2.35 (2.12–2.65)	0.011[a]	2.46 (2.19–2.85)	2.56 (2.15–3.00)	2.47 (2.20–2.73)	0.043[a]
MCHC (g/L)	327 (322–33)	326 (322–332)	329 (321–334)	0.043[a]	329 (323–33)	328 (324–332)	330 (322–334)	0.402
MONO# (10^9/L)	0.43 (0.37–0.56)	0.41 (0.36–0.53)	0.47 (0.37–0.58)	0.016[a]	0.48 (0.38–0.55)	0.49 (0.40–0.54)	0.47 (0.36–0.57)	0.178
NEU%	60.15 (55.40–64.1)	58.70 (53.90–64.70)	60.44 (57.35–63.54)	0.044[a]	59.69 (55.05–63.5)	59.00 (54.70–64.20)	60.80 (55.54–62.66)	0.048[a]
PA (g/L)	256 (227–28)	261 (236–291)	252 (217–283)	0.027[a]	255 (225–28)	258 (228–290)	251 (219–278)	0.039[a]
TBIL (mmol/L)	10.93 (8.50–13.80)	11.79 (8.96–13.80)	10.40 (8.10–14.00)	0.021[a]	10.90 (8.40–13.90)	11.60 (8.81–13.95)	10.00 (8.10–13.50)	0.079
TG (mmol/L)	1.46 (1.17–1.89)	1.36 (1.10–1.62)	1.63 (1.25–2.11)	<0.001[c]	1.53 (1.25–1.99)	1.43 (1.20–1.69)	1.69 (1.31–2.19)	<0.001[c]
UA (umol/L)	442 (382–51)	431 (382–497)	454 (378–528)	0.022[a]	442 (380–494)	430 (385–482)	455 (378–512)	0.037[a]
UREA (umol/L)	4.42 (3.79–5.13)	4.54 (3.88–5.20)	4.27 (3.75–5.04)	0.047[a]	4.58 ± 0.10	4.81 ± 0.93	4.36 ± 1.02	<0.001[c]
NLR	1.92 (1.59–2.36)	1.84 (1.49–2.38)	1.95 (1.70–2.34)	0.023[a]	1.89 (1.58–2.26)	1.89 (1.64–2.20)	1.96 (1.52–2.35)	0.039[a]
dNLR	1.51 (1.24–1.80)	1.42 (1.17–1.82)	1.53 (1.35–1.76)	0.023[a]	1.49 (1.23–1.74)	1.49 (1.26–1.68)	1.54 (1.21–1.80)	0.038[a]
Ultrasound				N/A				N/A
Fatty liver	406 (100)	203 (100)	203 (100)		181 (100)	37 (100)	144 (100)	
Fibrosis stage				0.084				0.2552
0–1	369 (90.89)	190 (93.60)	179 (88.18)		160 (88.40)	35 (94.59)	125 (86.81)	
2–4	37 (9.11)	13 (6.40)	24 (11.82)		21 (94.59)	2 (5.41)	2 (5.41)	

Normally or approximately normally distributed variables are expressed as the mean ± standard deviation, and skewed variables are expressed as the median (interquartile range). BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HCT, hematocrit; MCHC, mean red

blood cell hemoglobin concentration; NEU%, percentage of neutrophils; LYM#, lymphocytes; LYM%, lymphocytes; MONO#, monocytes; NLR, neutrophil/lymphocyte; dNLR, neutrophil/leukocyte–neutrophil; AG, anionic gap; A/G, albumin/globulin; PA, prealbumin; ALP, alkaline phosphatase; AST, aspartate aminotransferase; ALT, alanine aminotransferase; AST/ALT (aspartate aminotransferase/alanine aminotransferase); AMY, amylase; APOB, apolipoprotein B; APOA/APOB, apolipoprotein A/apolipoprotein B; TG, triglyceride; TBL, total bilirubin; GGT, γ-glutamyl transferase; GLU, glucose; GSP, glycated serum protein; UA, uric acid; UREA, urea; N/A, not applicable. Liver fibrosis grade from 0 to 4: 0—no fibrosis; 1—perisinus or periportal fibrosis; 2—perisinusoidal fibrosis with periportal fibrosis; 3—bridging fibrosis; 4—cirrhosis.

[a] $P < 0.05$.

[b] $P < 0.01$, and.

[c] $P < 0.001$.

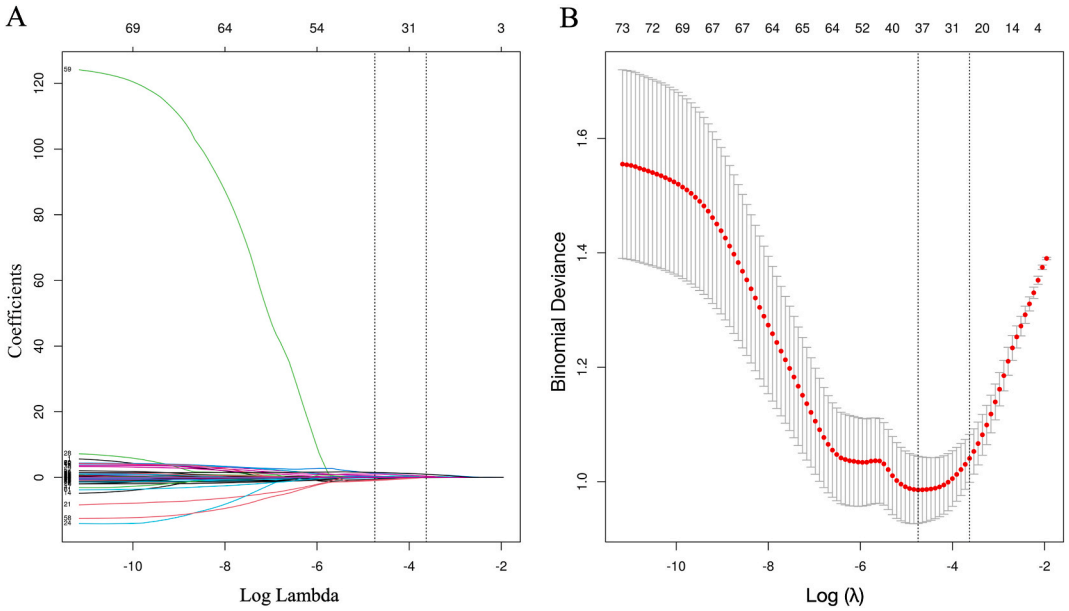


Fig. 2. Outcomes of the LASSO regression analysis. (A) Profiles of LASSO coefficients, which quantify the impact of each predictor variable on the response variable. These coefficients are obtained by fitting the LASSO model with varying regularization strengths. (B) Cross-validation error curve for the selection of the tuning parameter (λ). The curve plots the cross-validation error as a function of λ, indicating the λ value that minimizes the error and thus provides the best model fit.

Table 2
Diagnostic efficacy of the five machine learning algorithms in the internal validation set.

Algorithm	AUC±SD	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
XGBoost	0.93 ± 0.03	0.80 ± 0.03	0.92 ± 0.08	0.83 ± 0.06	0.93 ± 0.05	0.70 ± 0.04	0.89 ± 0.05
LR	0.83 ± 0.03	0.76 ± 0.05	0.72 ± 0.07	0.86 ± 0.03	0.80 ± 0.06	0.75 ± 0.03	0.75 ± 0.06
LightGBM	0.92 ± 0.04	0.73 ± 0.04	0.90 ± 0.07	0.85 ± 0.04	0.91 ± 0.06	0.66 ± 0.05	0.90 ± 0.04
RF	0.92 ± 0.03	0.80 ± 0.03	0.90 ± 0.06	0.80 ± 0.07	0.90 ± 0.07	0.78 ± 0.03	0.89 ± 0.04
AdaBoost	0.87 ± 0.04	0.80 ± 0.05	0.84 ± 0.08	0.81 ± 0.12	0.81 ± 0.07	0.80 ± 0.06	0.82 ± 0.03

XGBoost, extreme gradient boosting; LR, logistic regression; LightGBM, light gradient boosting machine; RF, random forest; AdaBoost, adaptive boosting. AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

validation cohort.

4. Discussion

Despite these limitations, the initial diagnosis of NAFLD is generally feasible owing to the development of ultrasonographic techniques. However, ultrasonography cannot be used to determine the severity of NAFLD. Distinguishing patients with NASH from those with NAFLD is an important challenge for clinicians. A liver biopsy is the gold standard for diagnosing NASH. Nevertheless, performing a liver biopsy for each NAFLD patient is not feasible, and the risks and cost-effectiveness of invasive procedures must be considered [26]. It is difficult for clinicians to assess the status of patients with NAFLD when clinical symptoms are inconspicuous and pathological findings are lacking. This study aimed to solve practical clinical problems to establish a diagnostic model that could effectively identify patients with NASH.

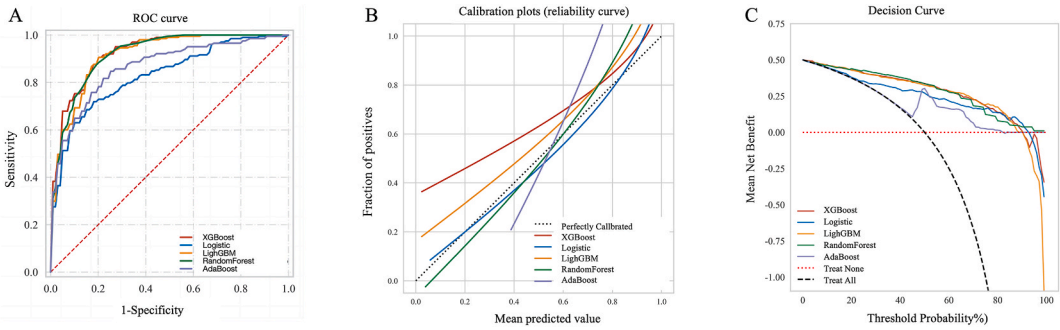


Fig. 3. Performance evaluation of five machine learning algorithms in the internal validation set. (A) Receiver operating characteristic curve analysis assessing the discriminatory capacity of the model based on true positive rate and false positive rate. (B) Calibration curve analysis evaluating how well the predicted probabilities of the algorithms align with the actual outcomes. (C) Decision curve analysis comparing the net benefit of using the model predictions versus making decisions solely based on the majority class.

Table 3
DeLong test of NASH-XGB6 and NASH-XGB25.

Algorithm	AUC (95 % CI)	z	p
NASH-XGB25	0.93 (0.89–0.96)	−1.437	0.151
NASH-XGB6	0.95 (0.91–0.98)		

AUC, area under the curve; CI, confidence interval. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.

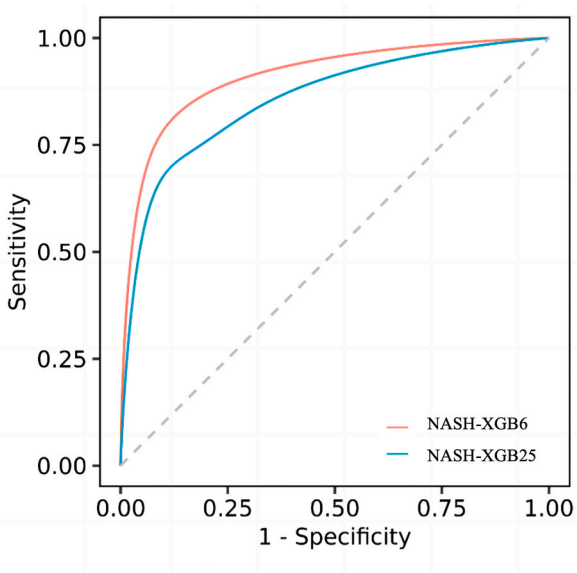


Fig. 4. Receiver Operating Characteristic curve analysis in the NASH-XGB6 and NASH-XGB25 models.

Traditional statistical techniques are hypothesis-driven and have inherent limitations. Specifically, they often restrict the analysis to only potential risk factors that are presumed relevant, potentially leading to the neglect of crucial information pertinent to outcome prediction. Alternatively, machine learning methods exhibit a remarkable capacity to handle multiple variables, capture intricate nonlinear interactions, and generate highly accurate predictive models [27]. We developed an ML-based NIT to discriminate between patients with NASH and NAFLD. It showed excellent performance on both internal and external validation datasets. The challenge in developing accurate predictive models for NASH is to achieve a good balance between specificity and sensitivity. Before applying the SMOTE algorithm, 56 and 203 patients were included in the NAFL and NASH groups, respectively. Misclassification errors may occur if the outcome classes are not equally distributed. The performance of machine learning algorithms is typically evaluated using predictive accuracy, which is not appropriate when the data are imbalanced or the costs of different errors vary significantly. The SMOTE algorithm was used to address the issue of imbalanced datasets and improve the accuracy of ML classifiers [28]. Compared with other methods such as random oversampling, SMOTE not only helps the model learn more nuanced patterns of the minority class but also

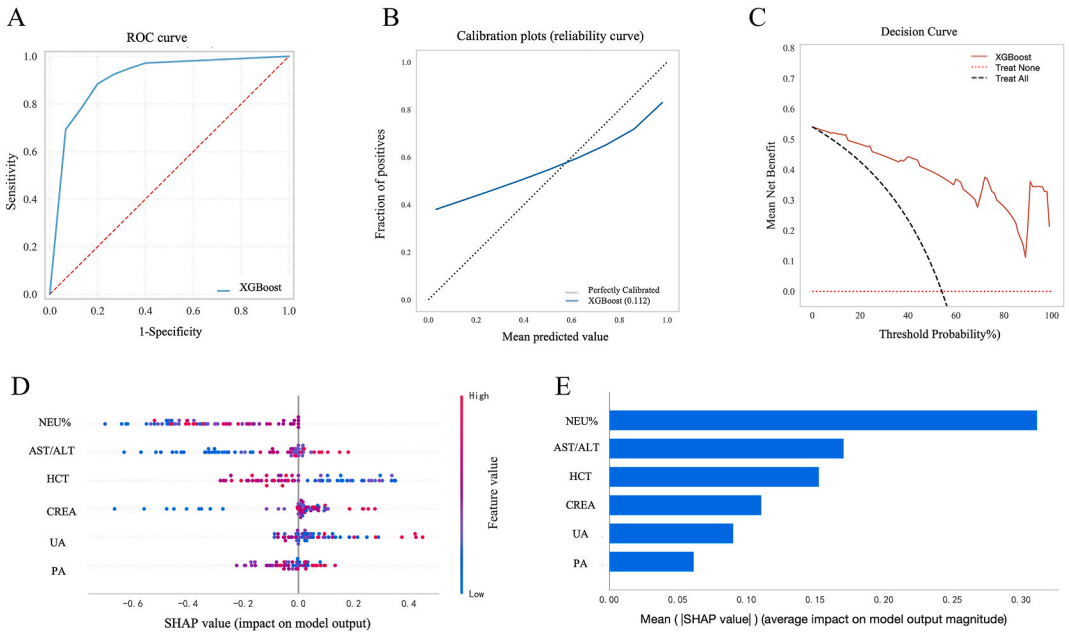


Fig. 5. Performance evaluation of the NASH-XGB6 model within the internal validation dataset. (A) Receiver operating characteristic curve analysis. (B) Calibration curve analysis. (C) Decision curve analysis. (D) SHAP value indicating the impact of each feature on the model predictions along the abscissa. (E) Ranking of feature importance based on the SHAP analysis providing insights into the relative contribution of each feature to the overall performance of the model. SHAP, SHapley Additive exPlanations.

Table 4
Diagnostic efficacy of XGBoost in the external validation set.

Algorithm	AUC (95 % CI)	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
XGBoost	0.90 (0.88–0.93)	0.81	0.78	0.90	0.92	0.75	0.84

AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

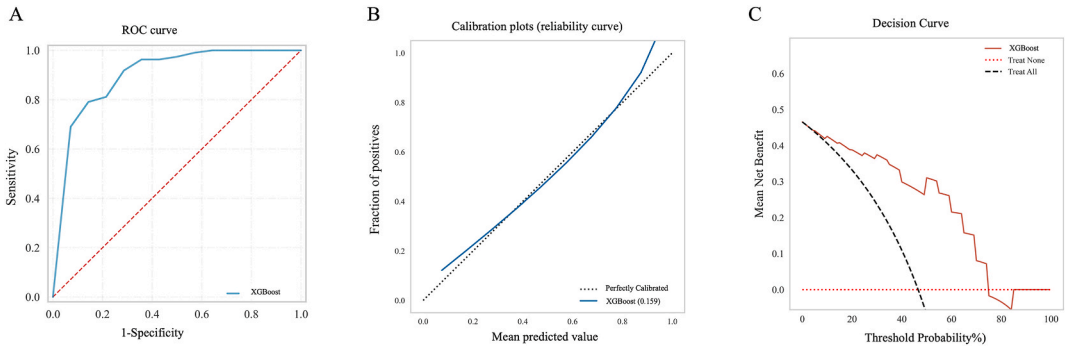


Fig. 6. Performance evaluation of the NASH-XGB6 model in the external validation set. (A) Receiver operating characteristic curve analysis. (B) Calibration curve analysis. (C) Decision curve analysis.

avoids the problem of simply duplicating existing minority class examples. Random undersampling is not suitable for our study because it reduces the data of the majority of NASH patients and decreases the sample size, which could lead to the loss of potentially valuable information. To construct the NASH prediction model, the predictor–outcome association was tested by narrowing the regression coefficient using the LASSO method, reducing 75 candidate features to 25 potential predictors. This method not only exceeds the method of selecting predictors based on the univariate association strength of predictors and outcomes [29] but also enables the panel of selected features to be combined into a NASH prediction model. In the internal validation dataset, the XGBoost algorithm, which included 25 variables to establish the model (NASH-XGB25), exhibited the best diagnostic efficacy (Table 2 and Fig. 3). Through further screening, we reduced the number of features to six to establish the NASH-XGB6 model (Fig. 5). To validate the generalization

capability of NASH-XGB6, we included a study cohort based on latitude for external validation. External validation showed that NASH-XGB6 had excellent diagnostic ability for identifying patients with NASH (Table 4 and Fig. 6), with the model's AUC, specificity, and PPV performing well. Based on these results, the NASH-XGB6 model was selected as the final algorithm. NASH-XGB6 is a noninvasive, easy-to-perform serum marker diagnostic model that can effectively support the management of patients with NAFLD.

Based on NAFLD cases diagnosed through liver tissue biopsy, we constructed a noninvasive diagnostic model characterized by six independent indicators (NEU%, AST/ALT, HCT, CREA, UA, and PA). The six features used in the ML model were closely related to the pathophysiology of NAFLD and its disease progression, which is similar to the findings on NAFLD reported in recent years. The SHAP value explained the contribution of a single sample to the prediction results and the percentage of neutrophils had the greatest impact on the performance of the NASH-XGB6 model (Fig. 5E). Neutrophils are the most abundant type of leukocytes in humans. In addition to acute inflammation, they play an important role in chronic inflammation such as that observed in nonalcoholic fatty liver disease [30]. NASH is an important stage in the development of NAFLD, and one of its characteristics is the large amount of neutrophil infiltration around lipotoxic hepatocytes, which promotes liver injury [31]. Our results also showed a significant increase in the number of neutrophils in the circulation of patients with NASH (Table 1), which is in line with previous findings [32]. In our study cohort, AST and ALT levels and AST/ALT ratios were significantly higher in the NASH group than in the NAFL group (Table 1), as expected. AST/ALT ratio, also known as the De Ritis ratio, is clinically significant [33] and was included in the model in this study. Some researchers believe that the prevalence of NAFLD increases with an increase in uric acid levels and that hyperuricemia is a risk factor for NAFLD [34]. Interestingly, PA levels significantly decreased in patients with NASH. When liver damage leads to the synthesis of hepatocyte proteins, the short half-life of PA in serum is more sensitive and can quickly reflect the presence of hepatocyte damage in the early stages [35].

The current study had some limitations. This was a retrospective, single-center study with a sample size limited to Chinese patients. Our model showed excellent performance in the current research environment, but we recognized that it may not be applicable to all populations. Specifically, our study may not have fully considered the characteristics of certain specific populations, such as the elderly, children, or certain ethnic minority groups. Therefore, when applying these models to a broader clinical setting, further validation and adjustment may have been necessary. Differences in resource availability, patient management, and clinical practices across different clinical environments may also have affected the performance of the model. Subsequent studies in different populations and centers, including various clinical settings, are needed to further validate our findings. In addition, there was a bias in the proportion of patients with NAFL and NASH. This is because clinicians decide which patients with NAFLD will undergo liver biopsy and select those who are in the more advanced stages of NASH through experience; thus, there is an inherent selection bias. However, this also makes the proposed ML model fit better in real-world situations. However, this study has several strengths. Firstly, by utilizing advanced algorithms and deep learning techniques, our model was able to analyze data more accurately, providing results that were more precise than those obtained from traditional non-invasive tests. By leveraging machine learning, our model analyzed a wide range of clinical and demographic features, enabling it to capture complex patterns and relationships that may not be considered in conventional tests. This comprehensive analysis contributed to its improved diagnostic performance. Secondly, by providing accurate predictions without requiring invasive procedures, our model offers a more effective and cost-efficient solution for diagnosing NASH and NAFLD. This can help reduce medical costs and improve patient outcomes. Thirdly, the liver biopsy results were obtained through a central review with a double-blind reading by two or more liver pathologists. This approach ensured the accuracy of pathological outcomes in our patients and prevented observer bias. Additionally, we have created a web-based version of the NASH-XGB6 model to provide greater convenience and accessibility to medical professionals in their daily practice. This web version allows healthcare providers to easily access and utilize the model's predictive capabilities without the need for specialized software or hardware, enabling them to make more informed decisions regarding the diagnosis and management of non-alcoholic steatohepatitis (NASH) and related liver conditions.

5. Conclusions

After evaluating the performance of five machine learning models built using different algorithms, we identified the NASH-XGB6 model, which was developed using the XGBoost algorithm, as the most effective. This model incorporates six key features: NEU%, AST/ALT, HCT, CREA, UA, and PA. Its excellent diagnostic capabilities for NASH enable timely intervention, ultimately improving the clinical outcomes for NAFLD patients with NAFLD. Our findings establish a solid foundation for the development of future machine learning models for disease prediction and provide valuable insights for clinical decision-making.

Ethics declarations

This study was reviewed and approved by the Medical Ethics Committee of the First Affiliated Hospital of Jinan University with the approval number: KY-2023-366, August 3, 2023. The written informed consent of the patient was waived as the medical records were obtained from previous clinical diagnosis and treatment, which had been approved by the ethics committee.

Funding

This research was funded by the National Natural Science Foundation of China (grant no.: 82172346), Science and Technology Projects in Guangzhou (grant no.: 2023A03J1041), the Guangdong Basic and Applied Basic Research Foundation (grant no.: 2020A1515110484), the Guangdong Bureau of Traditional Chinese Medicine (grant no.: 20212046), and the Fundamental Research

Funds for the Central Universities (grant no.: 21620307).

Data availability statement

The data related to this study is not stored in a publicly available repository. The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy and ethical reasons.

CRediT authorship contribution statement

Yuqi Yan: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Danhui Gan:** Validation, Software, Data curation. **Ping Zhang:** Visualization, Formal analysis. **Haizhu Zou:** Validation, Data curation. **MinMin Li:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38848.

References

[1] Z. Younossi, Q.M. Anstee, M. Marietti, T. Hardy, L. Henry, M. Eslam, J. George, E. Bugianesi, Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention, Nat. Rev. Gastroenterol. Hepatol. 15 (1) (2018) 11–20.
[2] M.A. Tincopa, R. Loomba, Non-invasive diagnosis and monitoring of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis, Lancet Gastroenterol Hepatol 8 (7) (2023) 660–670.
[3] C. Estes, H. Razavi, R. Loomba, Z. Younossi, A.J. Sanyal, Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease, Hepatology 67 (1) (2018) 123–133.
[4] A.J. Sanyal, M.L. Van Natta, J. Clark, B.A. Neuschwander-Tetri, A. Diehl, S. Dasarathy, R. Loomba, N. Chalasani, K. Kowdley, B. Hameed, L.A. Wilson, K.P. Yates, P. Belt, M. Lazo, D.E. Kleiner, C. Behling, J. Tonascia, Prospective study of outcomes in adults with nonalcoholic fatty liver disease, N. Engl. J. Med. 385 (17) (2021) 1559–1569.
[5] A.J. Sanyal, Past, present and future perspectives in nonalcoholic fatty liver disease, Nat. Rev. Gastroenterol. Hepatol. 16 (6) (2019) 377–386.
[6] L. He, L. Deng, Q. Zhang, J. Guo, J. Zhou, W. Song, F. Yuan, Diagnostic value of CK-18, FGF-21, and related biomarker panel in nonalcoholic fatty liver disease: a systematic review and meta-analysis, BioMed Res. Int. 2017 (2017) 9729107.
[7] S. Qi, D. Xu, Q. Li, N. Xie, J. Xia, Q. Huo, P. Li, Q. Chen, S. Huang, Metabonomics screening of serum identifies pyroglutamate as a diagnostic biomarker for nonalcoholic steatohepatitis, Clin. Chim. Acta 473 (2017) 89–95.
[8] A. Braza-Boïls, J. Marí-Alexandre, P. Molina, M.A. Arnau, M. Barceló-Molina, D. Domingo, J. Girbes, J. Giner, L. Martínez-Dolz, E. Zorio, Deregulated hepatic microRNAs underlie the association between non-alcoholic fatty liver disease and coronary artery disease, Liver Int. 36 (8) (2016) 1221–1229.
[9] P. Angulo, J.M. Hui, G. Marchesini, E. Bugianesi, J. George, G.C. Farrell, F. Enders, S. Saksena, A.D. Burt, J.P. Bida, K. Lindor, S.O. Sanderson, M. Lenzi, L. A. Adams, J. Kench, T.M. Therneau, C.P. Day, The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD, Hepatology 45 (4) (2007) 846–854.
[10] A.G. Shah, A. Lydecker, K. Murray, B.N. Tetri, M.J. Contos, A.J. Sanyal, Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease, Clin. Gastroenterol. Hepatol. 7 (10) (2009) 1104–1112.
[11] S. De Bruyne, M.M. Speeckaert, W. Van Biesen, J.R. Delanghe, Recent evolutions of machine learning applications in clinical laboratory medicine, Crit. Rev. Clin. Lab Sci. 58 (2) (2021) 131–152.
[12] K. Schawkat, A. Ciritsis, S. von Ulmenstein, H. Honcharova-Biletska, C. Jüngst, A. Weber, C. Gubler, J. Mertens, C.S. Reiner, Diagnostic accuracy of texture analysis and machine learning for quantification of liver fibrosis in MRI: correlation with MR elastography and histopathology, Eur. Radiol. 30 (8) (2020) 4675–4685.
[13] T. Teramoto, T. Shinohara, A. Takiyama, Computer-aided classification of hepatocellular ballooning in liver biopsies from patients with NASH using persistent homology, Comput Methods Programs Biomed 195 (2020) 105614.
[14] R. Forlano, B.H. Mullish, N. Giannakeas, J.B. Maurice, N. Angkathunyakul, J. Lloyd, A.T. Tzallas, M. Tsipouras, M. Yee, M.R. Thursz, R.D. Goldin, P. Manousou, High-throughput, machine learning-based quantification of steatosis, inflammation, ballooning, and fibrosis in biopsies from patients with nonalcoholic fatty liver disease, Clin. Gastroenterol. Hepatol. 18 (9) (2020) 2081–2090.e9.
[15] T.G. Oh, S.M. Kim, C. Caussy, T. Fu, J. Guo, S. Bassirian, S. Singh, E.V. Madamba, R. Bettencourt, L. Richards, R.T. Yu, A.R. Atkins, T. Huan, D.A. Brenner, C. B. Sirlin, M. Downes, R.M. Evans, R. Loomba, A universal gut-microbiome-derived signature predicts cirrhosis, Cell Metabol. 32 (5) (2020) 878–888.e6.
[16] V.W. Wong, W.K. Chan, S. Chitturi, Y. Chawla, Y.Y. Dan, A. Duseja, J. Fan, K.L. Goh, M. Hamaguchi, E. Hashimoto, S.U. Kim, L.A. Lesmana, Y.C. Lin, C.J. Liu, Y. H. Ni, J. Sollano, S.K. Wong, G.L. Wong, H.L. Chan, G. Farrell, Asia-pacific working party on non-alcoholic fatty liver disease guidelines 2017-Part 1: definition, risk factors and assessment, J. Gastroenterol. Hepatol. 33 (1) (2018) 70–85.
[17] A.J. Sanyal, E.M. Brunt, D.E. Kleiner, K.V. Kowdley, N. Chalasani, J.E. Lavine, V. Ratziu, A. McCullough, Endpoints and clinical trial design for nonalcoholic steatohepatitis, Hepatology 54 (1) (2011) 344–353.

[18] P. Bedossa, Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease, Hepatology 60 (2) (2014) 565–575.
[19] W. Sauerbrei, P. Royston, H. Binder, Selection of important variables and determination of functional form for continuous predictors in multivariable model building, Stat. Med. 26 (30) (2007) 5512–5528.
[20] J. Zhang, X. Ma, J. Zhang, D. Sun, X. Zhou, C. Mi, H. Wen, Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model, J Environ Manage 332 (2023) 117357.
[21] T.G. Nick, K.M. Campbell, Logistic regression, Methods Mol. Biol. 404 (2007) 273–301.
[22] T. Kuno, Y. Sahashi, S. Kawahito, M. Takahashi, M. Iwagami, N.N. Egorova, Prediction of in-hospital mortality with machine learning for COVID-19 patients treated with steroid and remdesivir, J. Med. Virol. 94 (3) (2022) 958–964.
[23] Y.V. Sun, Multigenic modeling of complex disease by random forests, Adv. Genet. 72 (2010) 73–99.
[24] Z. Zheng, Y. Yang, Adaptive boosting for domain adaptation: toward robust predictions in scene segmentation, IEEE Trans. Image Process. 31 (2022) 5371–5382.
[25] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (1) (2002) 321–357.
[26] Z.D. Goodman, Role of liver biopsy in clinical trials and clinical management of nonalcoholic fatty liver disease, Clin. Liver Dis. 27 (2) (2023) 353–362.
[27] R.Y. Choi, A.S. Coyner, J. Kalpathy-Cramer, M.F. Chiang, J.P. Campbell, Introduction to machine learning, neural networks, and deep learning, Transl Vis Sci Technol 9 (2) (2020) 14.
[28] R. Blagus, L. Lusa, SMOTE for high-dimensional class-imbalanced data, BMC Bioinf. 14 (2013) 106.
[29] R. Tibshirani, The lasso method for variable selection in the Cox model, Stat. Med. 16 (4) (1997) 385–395.
[30] A. Herrero-Cervera, O. Soehnlein, E. Kenne, Neutrophils in chronic inflammatory diseases, Cell. Mol. Immunol. 19 (2) (2022) 177–191.
[31] C. Luci, M. Bourinet, P.S. Leclère, R. Anty, P. Gual, Chronic inflammation in non-alcoholic steatohepatitis: molecular mechanisms and therapeutic strategies, Front. Endocrinol. 11 (2020) 597648.
[32] L. Antonucci, C. Porcu, E. Timperi, S.J. Santini, G. Iannucci, C. Balsano, Circulating neutrophils of nonalcoholic steatohepatitis patients show an activated phenotype and suppress T lymphocytes activity, J Immunol Res 2020 (2020) 4570219.
[33] M. Botros, K.A. Sikaris, The de ritis ratio: the test of time, Clin. Biochem. Rev. 34 (3) (2013) 117–130.
[34] X. Zheng, L. Gong, R. Luo, H. Chen, B. Peng, W. Ren, Y. Wang, Serum uric acid and non-alcoholic fatty liver disease in non-obesity Chinese adults, Lipids Health Dis. 16 (1) (2017) 202.
[35] M.Y. Yasmin, B. Aziz, M. Nazim, R.K. Madhavan, Prealbumin rather than albumin is a more sensitive indicator of acute liver disease, Malays. J. Pathol. 15 (2) (1993) 147–150.