Cardiovascular Diabetology

RESEARCH

Open Access

Interpretable machine learning analysis of immunoinflammatory biomarkers for predicting CHD among NAFLD patients

Wenyuan Dong[1,2,3], Hongcheng Jiang[1,2], Yu Li[1,2], Luo Lv[3], Yuxin Gong[4], Bao Li[3], Hongjie Wang[1,2,5*] and Hesong Zeng[1,2*]

Abstract

Background Coronary Heart Disease (CHD) and Non-Alcoholic Fatty Liver Disease (NAFLD) share overlapping pathogenic mechanisms including adipose tissue dysfunction, insulin resistance, and systemic inflammation mediated by adipokines. However, the specific impact of inflammation and immune responses on CHD risk in NAFLD patients remains poorly understood. This study evaluated the predictive value of ten immunoinflammatory indexes for CHD risk in NAFLD patients using an interpretable machine learning framework.

Methods We retrospectively analyzed 407 NAFLD patients undergoing coronary angiography, and stratifying them into NAFLD + CHD (n = 250) and NAFLD (n = 157) groups. Ten immunoinflammatory indexes were derived from the blood laboratory results. Lasso regression analysis and propensity score matching (PSM) were employed to mitigate confounding effects. Subsequently, univariate and multivariate logistic regression analyses were used to identify independent risk factors for CHD occurrence among NAFLD patients. While restricted cubic splines (RCS) and Receiver operating characteristic (ROC) curve evaluated the relationship between each immunoinflammatory indexes and CHD risk. Linear correlation methods were employed to evaluate the relationship between Gensini score and immunoinflammatory indexes. Finally, three machine learning algorithms (RF, SVM and GLM) were used to identify significant risk factors. To interpret the diagnostic model built by Random Forest, the SHapley Additive exPlanations (SHAP) method was employed, and features were ranked according to their SHAP values. Based on these rankings, a diagnostic nomogram was further constructed and the accuracy of the diagnostic model was evaluated using ROC curves.

Result After PSM, among the 282 included patients with NAFLD, 141 cases (50%) were complicated with CHD. Multivariate logistic regression analysis revealed that after adjusting for age, sex, hypertension, and smoking history, the NHR index was identified as the most significant risk factor for CHD in NAFLD patients (OR, 1.375; 95% CI, 1.021–

[†]Wenyuan Dong, Hongcheng Jiang and Yu Li contributed equally to this manuscript.

*Correspondence:
Hongjie Wang
hongjie.wang@tjh.tjmu.edu.cn
Hesong Zeng
zenghs@tjh.tjmu.edu.cn

Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

1.852; $P < 0.001$). Additionally, NLR, SII, SIRI and NMR were also identified as risk factors. PNR was a protective factor for CHD events in patients with NAFLD. RCS analysis demonstrated linear relationships between the NHR, NLR, and PNR index with CHD occurrence, whereas the SII index exhibited a non-linear J-shaped relationship with CHD risk (non-linear $P = 0.025$). Correlation analysis with Gensini score showed that the NHR index had the highest correlation with the severity of CHD (R = 0.256, $P < 0.001$). ROC curves indicated that the NHR index had good predictive and diagnostic performance (AUC = 0.703,95% CI, 0.652–0.754). Finally, the diagnostic nomogram constructed based on SHAP values demonstrated good accuracy and predictive performance (AUC = 0.834,95% CI, 0.795–0.873; $P < 0.001$).

Conclusion Six immunoinflammatory markers demonstrated significant associations with CHD risk in NAFLD populations, among which the NHR index exhibited particularly promising predictive potential.

Introduction

The severity of NAFLD is predominantly assessed through the evaluation of hepatic fat accumulation, inflammation activity, and the progression of fibrosis [1]. As a prevalent hepatic disorder, NAFLD encompasses a spectrum of pathological conditions, ranging from benign steatosis to nonalcoholic steatohepatitis (NASH), which is characterized by significant inflammation and fibrotic changes [2]. The pathogenesis and progression of NAFLD are intrinsically linked to inflammatory processes, with inflammatory mediators playing a crucial role in aberrant lipid metabolism and hepatocellular injury [3]. These mediators not only facilitate the transition from NAFLD to NASH but also accelerate fibrogenesis, potentially culminating in hepatocellular carcinoma [4, 5].

Similarly, the severity of CHD is generally gauged by the degree of coronary artery stenosis [6], the incidence of myocardial infarction [7], and the extend of cardiac functional impairment [8]. CHD and inflammation exhibit a profound interconnection, as the inflammatory mechanisms are central to both the initiation and progression of CHD [9]. Chronic inflammation within the arterial intima contributes to the formation of atherosclerotic plaques as well as the destabilization of plaques, which is the pathological hallmark of CHD [10, 11]. Moreover, systemic inflammation, as observed in obesity, diabetes mellitus, and metabolic syndrome, can also potentiate the inflammatory cascade in CHD [12, 13]. These metabolic disorders are frequently associated with elevated circulating inflammatory markers, which further exacerbate endothelial dysfunction, promote atherogenesis, and accelerate CHD progression [14].

The association between NAFLD and CHD is primarily mediated through shared metabolic risk factors, inflammation and immune responses, dysfunctional adipose tissue, and potential genetic and environmental influences [15]. Patients with NAFLD demonstrate an increased propensity for developing CHD, suggesting that effective management of NAFLD may potentially mitigate CHD risk [16].

The immunoinflammatory indexes plays a crucial role in elucidating the intricate relationship between CHD and NAFLD, underscoring the profound impact of inflammation on the pathogenesis of both diseases [17, 18]. Both CHD and NAFLD are often associated with systemic low-grade inflammation, which not only exacerbates disease progression but also serves as a potential mechanistic link between these conditions [19]. Although the calculation of these indexes may vary depending on specific research protocols or clinical contexts, they generally serve to quantify the overall inflammatory burden, which correlates with disease activity, prognostic outcomes, and therapeutic responses. Consequently, these indexes have emerged as valuable clinical tools for disease monitoring and management. Leveraging PSM to balance covariates and explainable machine learning (SHAP), we aim to identify the most robust inflammatory predictors of CHD in NAFLD and unravel their non-linear associations with disease severity. Our approach advances the field by providing a transparent, data-driven framework for personalized risk stratification, potentially guiding early therapeutic interventions in this high-risk population.

Methods

Study population and data collection

This cross-sectional study included 407 patients diagnosed with NAFLD by liver ultrasound upon admission. They underwent coronary angiography for chest pain at the Second Affiliated Hospital of Shanxi Medical University from January 2021 to December 2024. Based on the results of coronary angiography, patients were categorized into NAFLD + CHD group and NAFLD group. The Gensini score is a method used to quantify the severity and extent of coronary artery disease (CAD) based on the degree of luminal narrowing and the anatomical location of stenosis observed during coronary angiography [20]. It assigns different scores to various segments of the coronary arteries depending on the severity of narrowing: higher scores indicate more severe disease. The diagnostic criteria for NAFLD include liver ultrasound indicating hepatic steatosis, exclusion of excessive alcohol consumption (more than 140 g per week for males, more than 70 g per week for females), history of viral hepatitis, and use of hepatotoxic drugs. The ultrasound assessment

is conducted by skilled sonographers to evaluate hepatic fat deposition [21].

This study excluded participants with a history of prior use of statin or hepatotoxic drugs, coronary heart disease and other cardiac conditions such as rheumatic heart disease, valvular heart disease, congenital heart disease, heart failure, and cardiomyopathy. We also excluded participants with a history of malignancy, autoimmune diseases, acute or chronic infectious diseases, or severe cerebrovascular accidents. This study was approved by the Ethics Committee of the Second Affiliated Hospital of Shanxi Medical University. We obtained informed consent from all participants involved.

Through electronic medical record review, general clinical data such as gender, age, height, weight, systolic blood pressure, diastolic blood pressure, history of diabetes, and smoking were collected. Biochemical parameters including platelet count, neutrophil count, lymphocyte count, monocyte count, total cholesterol, triglycerides, high-density lipoprotein, and low-density lipoprotein were obtained through blood laboratory tests. Based on the biochemical parameters, we calculated BMI and ten immune-inflammatory surrogate indexes. $BMI = W/H^2$, $SII = (Plt \times N/L)$, $SIRI = (N \times M/L)$, $NLR = N/L$, $PLR = P/L$, $PNR = P/N$, $NHR = N/HDL\text{-}C$, $MHR = M/HDL\text{-}C$, $PHR = P/HDL\text{-}C$, $LHR = L/HDL\text{-}C$, $NMR = N/M$.

Statistical analysis

All data analyses were conducted using IBM SPSS Statistics version 23.0 and R Studio version 4.2.0. The normality of distribution for continuous variables was assessed using the Shapiro–Wilk test. Normally distributed data were described using means and standard deviations, while non-normally distributed data were described using medians and interquartile ranges. Categorical variables were presented as frequencies and percentages (%). Continuous variables were compared using independent sample t-tests or Mann–Whitney tests, and categorical variables were compared using chi-square tests or Fisher's exact tests. Initially, single-factor logistic regression analysis was used to identify potential predictors of CHD, and variables with p values < 0.05 from the univariate logistic regression analysis were included in the multivariate logistic regression analysis. Multivariate logistic regression analysis was then employed to determine independent predictors. The relationship between inflammation-immune indexes and CHD was reflected using restricted cubic splines (RCS). A p value < 0.05 (two-tailed) was considered statistically significant. Receiver operating characteristic (ROC) curve analysis and area under the curve (AUC) values were used to assess the predictive ability of different immunoinflammatory indexes for CHD. Lastly, linear correlation analysis was used to evaluate the relationship between Gensini score and immunoinflammatory indexes.

Lasso regression analysis

Lasso (Least Absolute Shrinkage and Selection Operator) regression analysis is a statistical method used for variable selection and modeling high-dimensional data. It applies an L1 regularization penalty to regression coefficients, which encourages some coefficients to shrink to zero, thereby achieving variable selection and model simplification. In this study, all variables between two groups (including gender, age, BMI, hypertension, diabetes, smoking history, platelets, monocytes, neutrophils, TC, TG, LDL-C, and HDL-C) were included in the Lasso regression analysis, using CHD as the dependent variable, the optimal penalty coefficient (λ) was determined through tenfold cross-validation. The λ value yielding the minimum mean squared error (MSE) (lambda.min) and the simplified model λ within one standard error (lambda.1se) were selected. Based on the lambda.1se threshold, 8 variables with non-zero coefficients were ultimately retained as significant predictors.

Propensity score matching analysis

We used propensity score matching (PSM) to mitigate the influence of potential confounding variables. Confounding factors identified through Lasso regression analysis, including gender, BMI, platelets, lymphocytes, monocytes, total cholesterol, and triglycerides, were adjusted to eliminate their impact on CHD. Propensity scores were calculated for each study subject with a caliper of 0.02. Matching adequacy between groups was assessed using absolute standardized differences and quartiles. Following PSM (1:1), the results demonstrated good balance between the two groups.

Machine learning algorithms

The dataset was randomly split into training and testing sets at a 7:3 ratio using stratified sampling, with the training set used for model construction and the testing set for model validation. Three machine learning models (RF, GLM, and SVM) were developed using the caret package (v7.1.0), with CHD as the dependent variable and other clinical indicators as independent variables. Model performance was evaluated using the DALEX package (v2.4.3), and based on residual plots analysis, the RF model was selected as the optimal model for subsequent analyses.

Model interpretation

To investigate the contribution of each individual feature to the prediction of CHD risk, we employed Tree SHAP to compute SHAP values from the RF model [22, 23]. The SHAP values, representing the logarithmic probability

of individual contributions, were visualized using the R package shapviz (v.0.9.6). SHAP facilitates both local and global interpretability, and we leveraged SHAP to provide an explanation for our predictive model, which encompasses relevant risk factors for the occurrence of CHD in patients with NAFLD. To identify the primary predictors of all-cause mortality within the patient cohort, we calculated the importance of feature rankings from the final model.

Results

Refinement of clinical feature selection using Lasso regression for CHD analysis

Through the application of the Lasso regression analysis, which is renowned for its ability to perform variable selection and regularization, we have effectively refined our baseline dataset. In our study, based on Lasso regression analysis, we successfully screened out seven confounding factors (sex, BMI, platelets, lymphocytes, monocytes, total cholesterol, and triglycerides) unrelated to CHD from a baseline dataset comprising 14 clinical features (Fig. 1). Lasso regression enabled us to eliminate the interference of confounding factors through the PSM method and select clinical features highly associated with CHD to streamline the analysis, thereby enhancing the robustness and interpretability of the results.

Clinical and biochemical characteristics in NAFLD and NAFLD-CHD groups

Through the Lasso algorithm, seven clinical features unrelated to CHD were identified and subsequently neutralized for their impact on outcomes using the PSM method. PSM is a statistical technique that balances covariates between treatment and control groups, effectively reducing bias and confounding variables in observational studies. After applying the PSM (1:1) method, this study included a total of 282 participants, consisting of 141 NAFLD patients and 141 NAFLD patients with CHD, with average ages of 54.52 and 58.89 years, respectively. Compared to patients without CHD, those with CHD had a higher proportion of smokers, hypertension and diabetes (Table 1). In contrast, sex, BMI, platelet count, lymphocyte count, monocyte count, total cholesterol, and triglycerides were balanced between the groups, with no significant differences observed ($P > 0.05$).

Univariate and multivariate analyses of factors associated with CHD in NAFLD

The univariate logistic regression analysis showed that neutrophil(N)(OR, 1.407; 95% CI, 1.194—1.657; $P < 0.001$), Low-Density Lipoprotein Cholesterol (LDL-C) (OR, 1.754; 95% CI, 1.225—2.513; $P = 0.002$), Systemic Immune-Inflammation Index (SII)(OR, 1.002; 95% CI, 1.001—1.002; $P < 0.001$), Systemic Inflammation Response Index (SIRI)(OR, 1.532; 95% CI, 1.115—2.107; $P = 0.009$), Neutrophil-to-Lymphocyte Ratio (NLR)(OR, 1.401; 95% CI, 1.144—1.717; $P = 0.001$), Platelet-to-Neutrophil Ratio (PNR)(OR, 0.984; 95% CI, 0.975—0.994; $P < 0.001$), Neutrophil-to-High-density lipoprotein Cholesterol Ratio (NHR)(OR, 1.301; 95% CI, 1.132—1.496; $P < 0.001$), and Neutrophil-to-Monocyte Ratio (NMR) (OR, 1.155; 95% CI, 1.079—1.236; $P < 0.001$) were correlated with the occurrence of CHD in NAFLD patients. The multivariate logistic regression analysis showed that N(OR, 1.420; 95% CI, 1.193—1.690; $P < 0.001$), LDL-C(OR, 1.948; 95% CI, 1.326—2.861; $P < 0.001$), SII(OR, 1.002; 95% CI, 1.001—1.003; $P = 0.001$), SIRI(OR, 1.527;

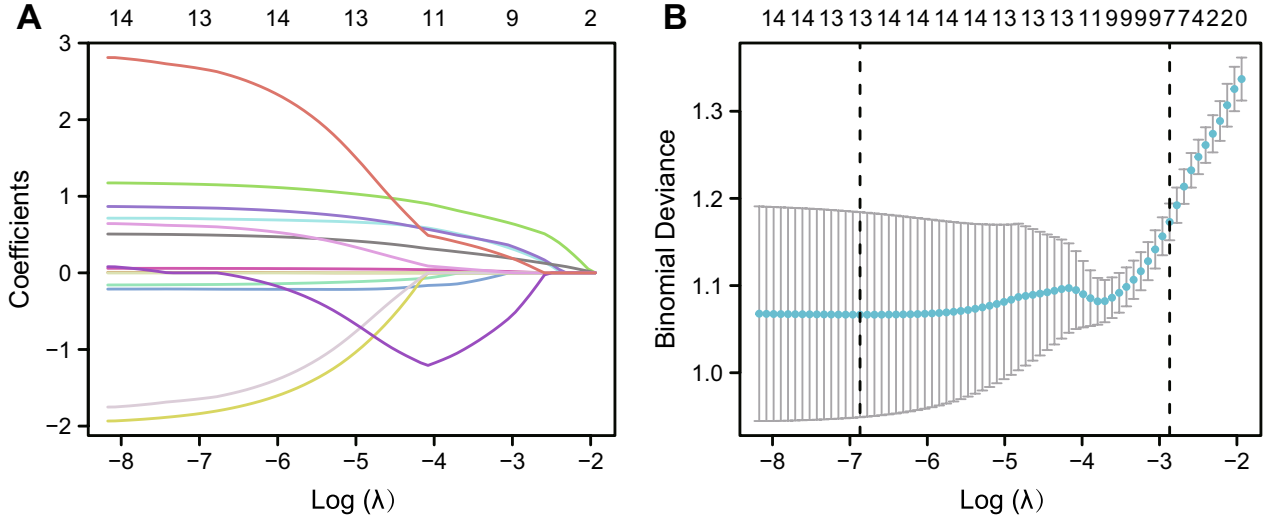

Fig. 1 LASSO regression screening results. **A** Co-variation of the 14 variables when the regularisation parameter λ is varied. **B** The regularisation parameter λ screening process

Table 1 Baseline characteristics of patients with NAFLD (control) and with NAFLD and CHD

Characteristics	Before PSM			After PSM (1: 1)		
	CHD + NAFLD(250)	NAFLD(157)	*P* value	CHD + NAFLD(141)	NAFLD(141)	*P* value
Sex (male)	176 (43.2%)	81 (19.9%)	< 0.001	82 (29.1%)	80 (28.4%)	0.81
Age (years)	59 (51, 67)	55 (48, 62)	< 0.001	58.887 ± 11.145	54.518 ± 10.04	< 0.001
BMI (kg/m^2)	25.96 (24.77, 27.775)	25.82 (24.16, 27.72)	0.429	26.04 (24.22, 28.01)	25.91 (24.16, 27.72)	0.559
HT (%)	169 (41.5%)	73 (17.9%)	< 0.001	101 (35.8%)	65 (23%)	< 0.001
DM (%)	96 (23.6%)	19 (4.7%)	< 0.001	52 (18.4%)	14 (5%)	< 0.001
Smoke (%)	134 (32.9%)	49 (12%)	< 0.001	59 (20.9%)	48 (17%)	0.177
Plt (10^9/L)	216.5 (177.25, 258.25)	225 (195, 258)	0.117	226 (185, 261)	224 (198, 257)	0.966
N, median (10^9/L)	4.335 (3.4325, 5.9375)	3.48 (2.73, 4.61)	< 0.001	4.07 (3.32, 5.67)	3.54 (2.73, 4.77)	< 0.001
L, median (10^9/L)	1.84 (1.4525, 2.33)	1.94 (1.58, 2.37)	0.105	2.004 ± 0.71162	1.9501 ± 0.53722	0.473
M, median (10^9/L)	0.475 (0.37, 0.6175)	0.47 (0.34, 0.54)	0.048	0.45 (0.35, 0.53)	0.48 (0.37, 0.55)	0.391
TC (mmol/L)	4.4689 ± 1.1113	4.3691 ± 0.98557	0.358	4.4222 ± 1.0434	4.3215 ± 0.93336	0.394
TG (mmol/L)	1.805 (1.335, 2.6)	1.61 (1.19, 2.2)	0.001	1.65 (1.27, 2.22)	1.59 (1.19, 2.22)	0.305
HDL-C (mmol/L)	1.05 (0.91, 1.23)	1.17 (1, 1.34)	< 0.001	1.13 (0.94, 1.3)	1.16 (1, 1.32)	0.214
LDL-C (mmol/L)	2.485 (2.075, 2.975)	2.27 (1.78, 2.71)	< 0.001	2.46 (2.09, 2.95)	2.26 (1.78, 2.66)	0.002

Data are shown as means ± SD or medians with interquartile ranges (IQRs). *CHD* Coronary heart disease, *BMI* body mass index, *HT* hypertension, *DM* diabetes mellitus, *NAFLD* Non-alcoholic fatty liver disease, *PSM* propensity score matching, *N* Neutrophil, *L* Lymphocyte, *M* Monocyte, *Plt* Platelet, *TC* Total cholesterol, *TG* Triglycerides

Table 2 Univariate and multivariate analyses of factors associated with NAFLD-CHD

Characteristics	Univariate analysis		Multivariate analysis	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
N	1.407 (1.194–1.657)	< 0.001	1.420 (1.193–1.690)	< 0.001
HDL-C (mmolL)	0.570 (0.226–1.434)	0.232		
LDL-C (mmolL)	1.754 (1.225–2.513)	0.002	1.948 (1.326–2.861)	< 0.001
SII	1.002 (1.001–1.002)	< 0.001	1.002 (1.001–1.003)	0.001
SIRI	1.532 (1.115–2.107)	0.009	1.527 (1.072–2.175)	0.019
NLR	1.401 (1.144–1.717)	0.001	1.418 (1.129–1.781)	0.003
PLR	1.003 (0.998–1.007)	0.209		
PNR	0.984 (0.975–0.994)	< 0.001	0.984 (0.974–0.994)	0.002
NHR	1.301 (1.132–1.496)	< 0.001	1.375 (1.021–1.852)	< 0.001
MHR	1.074 (0.334–3.453)	0.905		
PHR	1.001 (0.998–1.004)	0.442		
LHR	1.172 (0.854–1.606)	0.326		
NMR	1.155 (1.079–1.236)	< 0.001	1.161 (1.080–1.248)	< 0.001

Univariate analyse, non-adjusted model. Multivariate analyse was adjusted for age, hypertension, diabetes mellitus and smoking history

95% CI, 1.072—2.175; $P = 0.019$), NLR(OR, 1.418; 95% CI, 1.129—1.781; $P = 0.003$), NHR(OR, 1.375; 95% CI, 1.021—1.852; $P < 0.001$) and NMR(OR, 1.161; 95% CI, 1.080—1.248; $P < 0.001$) were still risk factors for CHD in NAFLD patients. PNR was a protective factor for CHD (Table 2; Fig. 2).

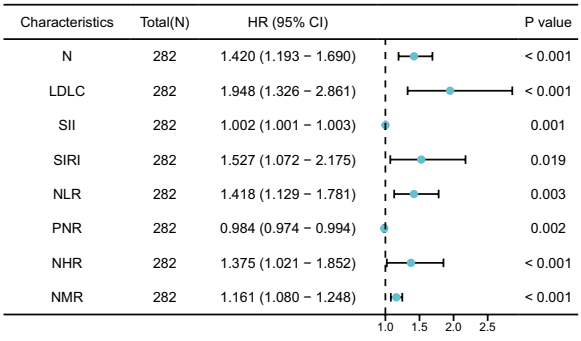

Fig. 2 Forest plots of independent factors associated with CHD in NAFLD

Using RCS analysis to explore the relationship between immunoinflammatory indexes and CHD in NAFLD populations

We used RCS to analyze the dose–response relationship between six immunoinflammatory indexes and the risk of CHD (Fig. 3). After adjusting for all covariates in the analytical model, a linear correlation was observed between NHR, NLR, NMR, and PNR with CHD (P for overall < 0.05, P for nonlinear > 0.05). In contrast, SII showed a nonlinear correlation with CHD (P for overall = 0.001, P for nonlinear < 0.05). Moreover, there was a negative correlation between PNR and CHD, with the incidence of CHD decreasing as PNR increased. It is noteworthy that in the NAFLD population, SIRI does not show a significant correlation with the occurrence of CHD (P for overall = 0.155, P for nonlinear = 0.254).

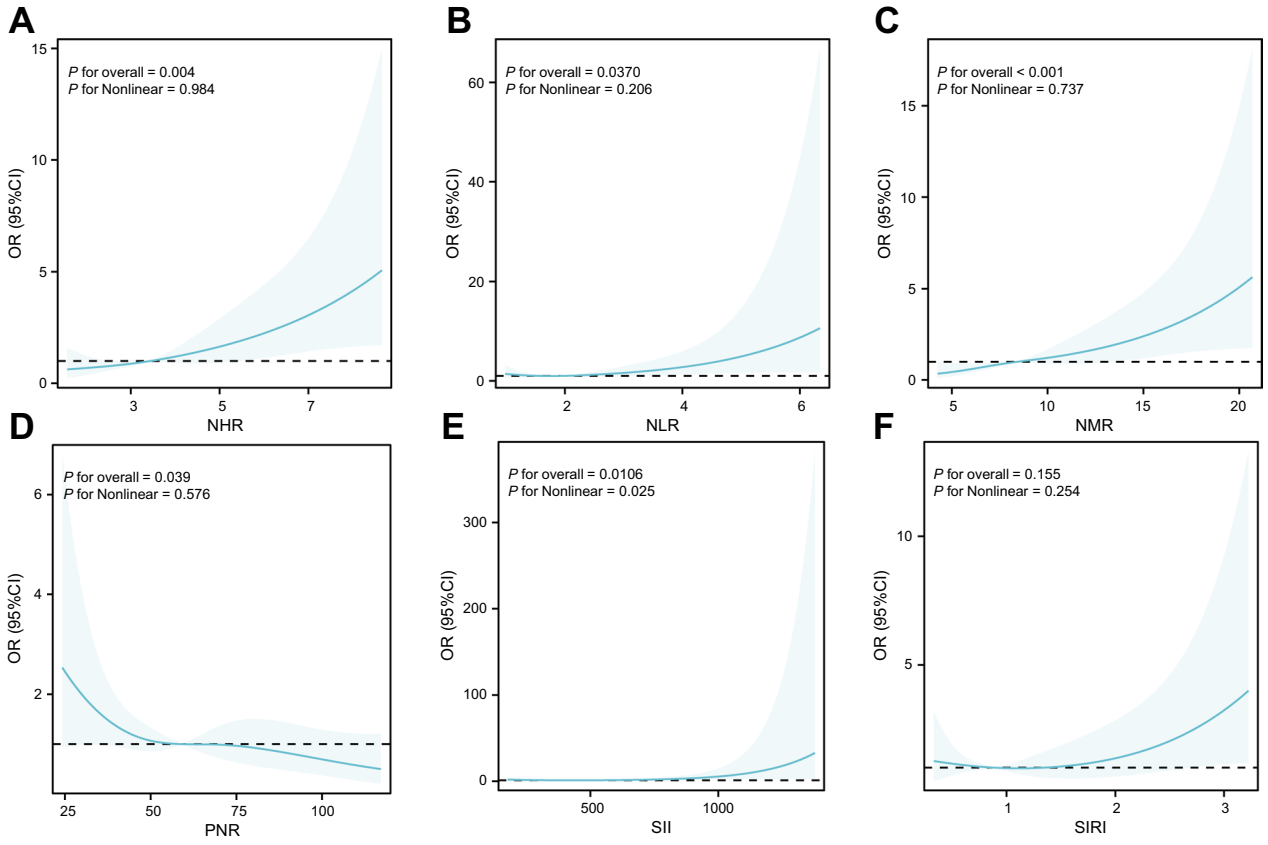

Fig. 3 The association between 6 variables and CHD was shown using restricted cubic splines (RCS)

Association between immunoinflammatory indexes and coronary Gensini score

Spearman's correlation analysis between Gensini score and six immunoinflammatory indexes showed that Gensini score was not significantly correlated with NLR (r = 0.169, $P = 0.008$), SII (r = 0.116, $P = 0.072$), SIRI (r = 0.176, $P = 0.005$) and NMR (r = 0.085, $P = 0.183$). Weak relationships were detected between Gensini score and NHR (r = 0.256, $P < 0.001$) and PNR (r = -0.200, $P = 0.001$) (Fig. 4).

Population risk biomarkers predictive of CHD risk using machine learning models

We established a disease prediction model based on univariate and multivariate logistic regression analysis to identify risk factors for CHD in NAFLD. Using Generalized Linear Models (GLM), Support Vector Machines (SVM), and Random Forest (RF) algorithms, we compared residual values with residual plot to determine the most effective predictive model. RF exhibited the smallest residuals, indicating superior predictive capability. Therefore, we proceeded with the RF model for further analysis (Fig. 5A). To explain which risk biomarkers are most important for predicting the diagnosis of CHD, we calculated the Shapley Additive Explanations (SHAP) values. SHAP values provide a transparent and interpretable method for evaluating the importance of individual features in a model. By calculating SHAP values, we can better identify which risk biomarkers contribute the most to predicting the diagnosis of CHD. The top three predictive factors for increased CHD risk include higher NHR levels, older age, and the presence of type 2 diabetes. In addition, higher LDLC levels and lower PNR levels are also important predictive factors for the occurrence of CHD. The levels of NMR, SII, and SIRI have a weak impact on the occurrence of CHD(Fig. 5 B,C). Moreover, male gender, smoking, hypertension, and higher triglyceride levels are risk factors for the occurrence of CHD, while higher HDLC levels serve as a protective factor against CHD. We extracted and visualized individual CHD risk profiles, using waterfall plots to show the positive and negative impacts of personal CHD risk biomarkers on the prediction results (Fig. 5D). In conclusion, for patients suffered CHD and NAFLD, the predicted risk was primarily influenced by nine factors: diabetes, age, hypertension, smoking, TG, NLR, NHR, LDL-C, and HDL-C. Furthermore, to evaluate the predictive performance of NHR in patients with different myocardial injury statuses, we conducted stratified subgroup analyses based on cardiac biomarkers (Creatine

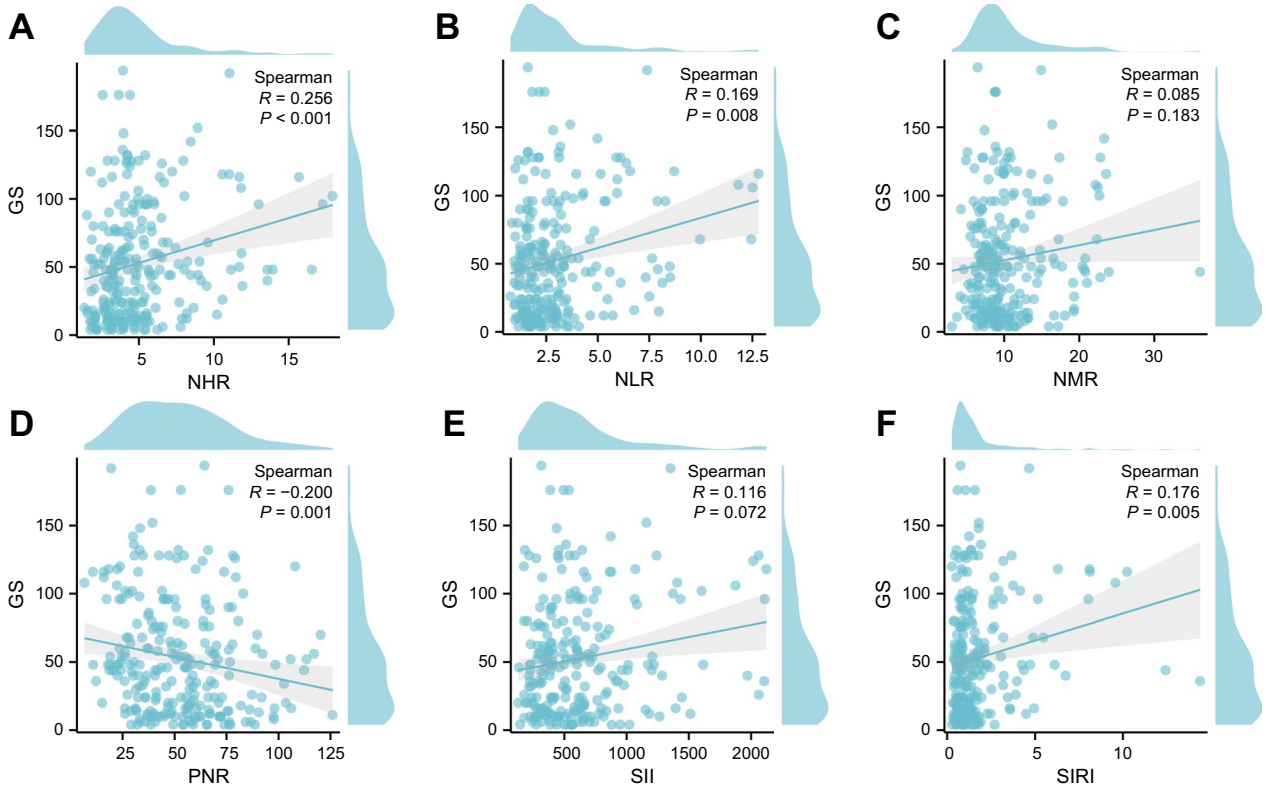

Fig. 4 Analysis of the correlation between six immunoinflammatory indexes and the severitly of CHD

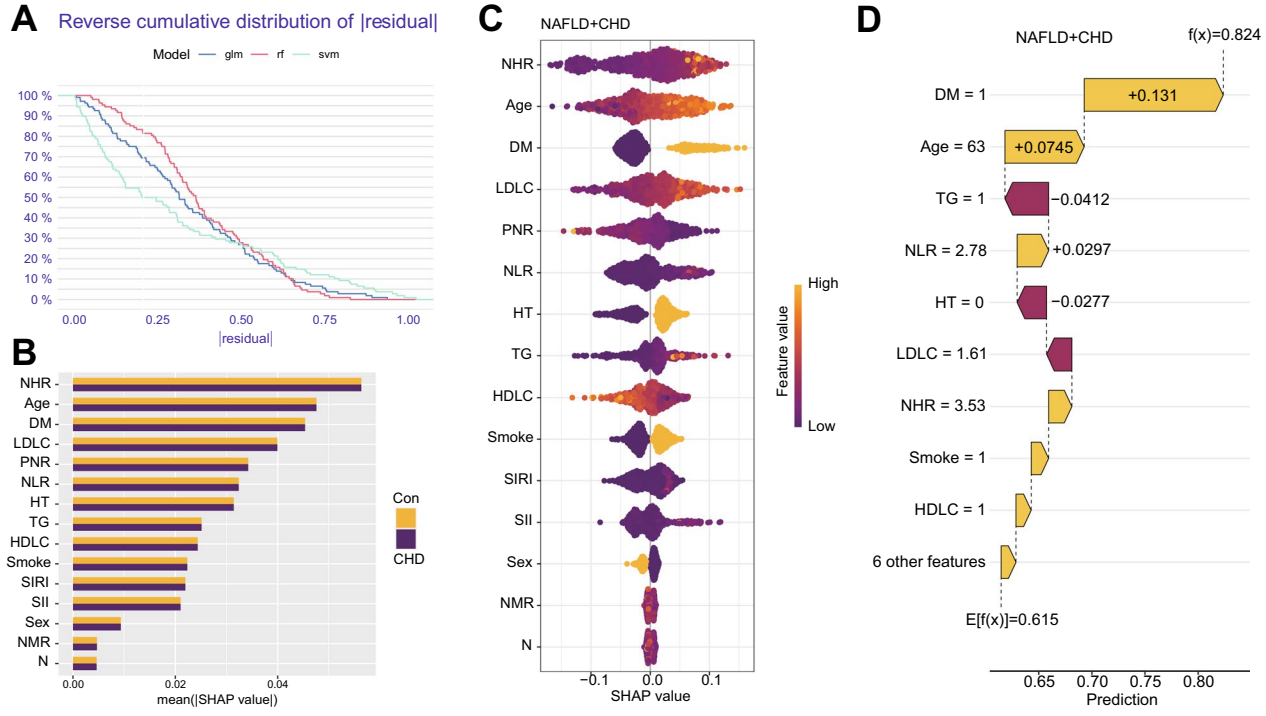

Fig. 5 Machine Learning Screening Variables. **A** Residual plots assist in evaluating and comparing the fitting effectiveness of different machine learning models. **B** Summary Plot visually ranks features by their average impact on model predictions using SHAP values. **C** Beeswarm plot of SHAP values for different variables in the CHD group. **D** The Waterfall plot displays the SHAP value contribution of each feature to an individual prediction outcome

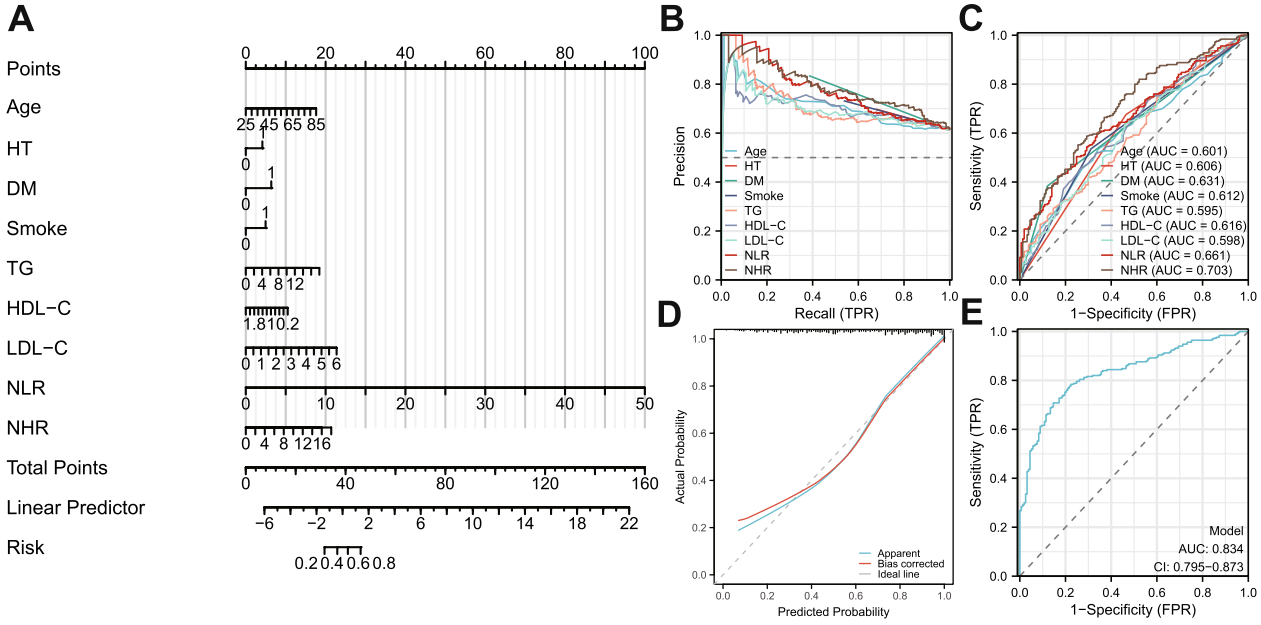

Fig. 6 Establishing a diagnostic nomogram for CHD in patients with NAFLD. **A** Nomogram for the diagnosis of CHD. **B** PR curve for diagnosis of CHD. **C** Calibration curve for prediction accuracy. **D** ROC curves of nine indexes in the overall population. **E** The ROC curve of the combination of nine indexes in the overall population

Kinase-Myocardial Band, Myoglobin and Cardiac Troponin T) in the NAFLD population (Supplementary 1). The results demonstrated that NHR consistently exhibited significant predictive value for CHD in biomarker-negative subgroups, including CKMB-negative (OR = 2.22, 95% CI: 1.36–3.62), Mb-negative (OR = 2.32, 95% CI: 1.40–3.83), and cTnT-negative (OR = 1.93, 95% CI: 1.00–3.73) patients (all $P < 0.05$), indicating that in the NAFLD population, even when myocardial injury markers are not elevated, NHR can serve as a predictive biomarker for the occurrence of CHD.

Construction of a diagnostic nomogram to predict CHD in the NAFLD

We employed machine learning algorithms to identify nine biomarkers associated with CHD. Building upon this, we aimed to construct a nomogram model using these biomarkers to predict CHD risk specifically in NAFLD populations (Fig. 6A). Evaluation using PR curves confirmed the model's robust diagnostic accuracy and reliability, with all predictive factors surpassing baseline measures and converging towards optimal performance in the upper-right corner (Fig. 6B). Further validation through Diagnostic Calibration Curve showed minimal deviation from the ideal line, affirming the model's precision and utility in clinical prediction of CHD risk (Fig. 6C). We evaluated the predictive capability of nine biomarkers for CHD risk in NAFLD patients using ROC curves, with discrimination assessed by the Area Under the Curve (AUC). LDL-C (AUC = 0.598), SII (AUC = 0.631), SIRI (AUC = 0.656), NLR (AUC = 0.661), PNR (AUC = 0.675), and NHR (AUC = 0.703,95% CI, 0.652–0.754) exhibited varying accuracies (Fig. 6D). Then, we investigated their combined predictive accuracy, revealing a significant improvement with an AUC of 0.834 when these nine biomarkers were used together (Fig. 6E).

Discussion

CHD, occupying a pivotal position in global health burdens as the most prevalent cardiovascular disorder, demands urgent attention in both epidemiological surveillance and clinical intervention paradigms [24]. Besides, emerging evidence positions NAFLD as a critical yet underrecognized risk factor for CHD development, sharing common pathophysiological determinants through systemic inflammatory pathways. Our pioneering investigation systematically evaluated and contrasted the prognostic capacity of ten distinct systemic inflammation biomarkers in predicting CHD susceptibility within NAFLD populations. Following comprehensive adjustment for confounding variables, six immune-inflammatory indexes demonstrated significant associations with CHD risk stratification, among which the Neutrophil-to-HDL Ratio (NHR) was regarded as a particularly potent predictive biomarker. This convergence of findings substantiates the hypothesis that dysregulated immune homeostasis constitutes a shared etiological axis in NAFLD-CHD comorbidity [25, 26].

Neutrophils play a significant role in the development of atherosclerosis and plaque instability [27], while also exacerbating tissue damage during the pathogenesis of NAFLD through the release of cytokines and neutrophil extracellular traps (NETs) [28]. In alignment with previous findings, our multivariate analyses corroborate neutrophils as an independent CHD risk determinant in NAFLD patients, with derived indexes including SII, SIRI, NLR, NMR, PNR, and NHR all demonstrating significant predictive validity. This compelling evidence positions neutrophil-mediated pathways as promising therapeutic targets for dual metabolic-cardiovascular risk mitigation.

Emerging evidence highlights distinct inflammatory signatures in CHD and NAFLD, yet their interplay remains underexplored. While prior investigations have independently characterized immunoinflammatory indices in these conditions, including the Dongfeng-Tongji cohort linking elevated SII to stroke risk [29], a retrospective analyses associating SIRI/HDL-C and MHR with CHD severity (n = 3,201) [30], and NHANES-based study identifing NLR as an independent risk factor for CHD development [31]. All these studies predominantly focus on singular disease entities. Notably, Sun et al. identified U-shaped SII/SIRI-NAFLD associations in cross-sectional analyses [32], and our study also demonstrates that in the NAFLD population, the SII index exhibits a J-shaped relationship with the incidence of CHD, a pattern corroborated by mortality trends in US NAFLD populations [33]. Critically, despite NAFLD's established role as a CHD risk amplifier [34], mechanistic insights into inflammation-driven cardiovascular disease within this vulnerable population remain unclear. Our study uniquely bridges this knowledge gap through dual innovations: Firstly, we systematically evaluate ten immunoinflammatory indices, confirming SII, SIRI, and NLR as independent CHD predictors in NAFLD patients, which extended prior single-disease observations. Secondly, leveraging machine learning, we establish neutrophil-derived ratios (NHR, NMR, PNR) as superior discriminators of CHD risk, surpassing conventional biomarkers. This comprehensive approach not only validates chronic inflammation as a shared NAFLD-CHD axis but pioneers precision CHD risk stratification tools for NAFLD populations.

To ensure analytical robustness, we implemented a multi-methodological framework incorporating 407 angiographically confirmed CHD cases with Gensini score-quantified lesion severity. Following rigorous inclusion criteria, Lasso regression with PSM eliminated seven key confounders, establishing a purified cohort for multivariable-adjusted models. Our integrated analytical approach revealed SII, SIRI, NLR, NHR, and NMR as independent CHD risk determinants, among which NHR demonstrated the strongest association, which extending prior evidence on neutrophil-derived ratios' predictive capacity [35]. RCS analyses uncovered nonlinear SII-CHD correlations mirroring mortality patterns in CVD populations [36], suggesting threshold-based CHD risk at elevated SII levels. While NLR exhibited linear CVD mortality associations [37], our Spearman correlations uniquely identified the linkageof NHR and Gensini score. Our methodological synergy, spanning machine learning-driven confounder control, nonparametric dose–response modeling, and comparative epidemiology, confers superior validity over single-method studies. By reconciling divergent findings from prior CHD-focused and NAFLD-centric research, our multi-modal paradigm advances precision in inflammation-associated CHD risk stratification in NAFLD populations.

We conducted a cross-sectional comparison of 10 immune inflammation indices and found that six of them are independent risk factors for CHD. Although research on prognostic risk factors for CHD and NAFLD has become increasingly in-depth [38], there is currently no systematic model that integrates these risk factors to accurately predict the occurrence of CHD. In this study, we established an RF-based model using 15 clinical factors, demonstrating that our model can better predict the occurrence of CHD in NAFLD populations. Three machine learning algorithms were employed to compare the importance of these factors in predicting CHD. One advantage of our study is the use of interpretable AI approaches (SHAP estimation) to help explain the predictions of complex machine learning models. By displaying the contribution of each feature to the model's output, SHAP makes the model more transparent, facilitating understanding and reliable. Furthermore, through SHAP values, the relative importance of each feature in the prediction can be determined. It provides a more precise and consistent evaluation compared to traditional feature importance methods. In addition, we also used SHAP to provide explanations for individual predictions. This is particularly useful for predicting the occurrence of CHD in specific individuals, allowing us to gain a more intuitive understanding of the diagnostic basis of the RF model. Based on the importance ranking from the SHAP analysis, the top three predictive factors for increased CHD risk include higher NHR levels, older age, and the presence of type 2 diabetes. The analysis results from the SHAP model are consistent with the results of the multivariate logistic regression, which suggests that in the NAFLD population, NHR is the best indicator for predicting the occurrence of CHD, and patients with comorbid diabetes and advanced age should receive more attention [39]. Finally, we constructed a diagnostic model based on the importance ranking from the SHAP analysis. which showed a higher AUC than single indexes. The

combined diagnostic model demonstrated good predictive performance. PR curve and diagnostic calibration curve analyses also indicated that this nomogram had good predictive ability.

This study employed a multi-phase analytical approach to ensure the reliability of conclusions: First, potential confounding variables were screened using LASSO regression and balanced through propensity score matching (PSM). Subsequently, univariate and multivariate logistic regression analyses identified six immunoinflammatory markers significantly associated with CHD. Further analysis utilized restricted cubic spline (RCS) models to examine nonlinear relationships, combined with Spearman rank correlation to evaluate associations between these indices and CHD severity. To enhance predictive performance, three machine learning algorithms (RF, SVM and GLM) were constructed and compared. Finally, based on the SHAP (SHapley Additive exPlanations) interpretability framework, key predictive factors were selected to construct a diagnostic nomogram.

However, this study has several noteworthy limitations: First, despite rigorous adjustment for known confounders including demographic characteristics and clinical indicators, unmeasured variables such as genetic background and environmental exposures may still influence results. Second, the exclusion of patients with non-alcoholic steatohepatitis (NASH) and significant liver fibrosis may limit the generalizability of conclusions across the full NAFLD spectrum. Third, the lack of time-to-event analysis means that while findings demonstrate a positive correlation between NHR index and CHD occurrence in NAFLD patients, predictive value cannot be established. Fourth, as the study population was limited to Chinese individuals, future large-scale multicenter prospective studies are needed to further validate the predictive capability of NHR index for CHD risk in NAFLD patients.

Conclusions

In summary, our findings demonstrate that immunoinflammatory markers are associated with CHD occurrence in NAFLD populations. Among these markers, NHR index exhibits optimal performance in assessing CHD risk in NAFLD patients, and NHR level showing a positive correlation with CHD severity. However, this study has certain limitations, including a relatively small sample size. Therefore, future multicenter studies involving diverse ethnic populations are warranted to further explore and validate these findings.

Abbreviations

CHD	Coronary heart disease
NAFLD	Nonalcoholic fatty liver disease
DM	Diabetes mellitus
BMI	Body mass index
HT	Hypertension
LDL-C	Low-density lipoprotein cholesterol
HDL-C	High density lipoprotein cholesterol
TC	Total cholesterol
TG	Triglyceride
Plt	Platelet
N	Neutrophil
L	Lymphocyte
M	Monocyte
SII index	Systemic Immune-Inflammation Index
SIRI index	Systemic Inflammatory Response Index
NLR	Neutrophil-to-Lymphocyte Ratio
PLR	Platelet-to-Lymphocyte Ratio
PNR	Platelet-to-Neutrophil Ratio
NHR	Neutrophil-to-high density lipoprotein ratio
MHR	Monocyte-to-high density lipoprotein ratio
PHR	Platelet-to-high density lipoprotein ratio
LHR	Lymphocyte-to-high density lipoprotein ratio
NMR	Neutrophil-to-Monocyte Ratio
Lasso	Least absolute shrinkage and selection operator
GLM	Generalized linear model
RF	Random forest
SVM	Support vector machine
RCS	Restricted cubic splines
ROC	Receiver operating characteristic
GS	Gensini score

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12933-025-02818-1.

Supplementary Material 1.

Author contributions

DWY, JHC and LY conceived the idea for the study and wrote the manuscript. LL and GYX helped analyze the study data. LB revised the manuscript. WHJ and ZHS directed the writing and analysis of this article. All authors read and approved the final manuscript. All authors contributed to this article and approved the version submitted.

Funding

This work was supported by the National Natural Science Foundation of China (NO. 82070490 to HZ) and the Hubei Provincial Engineering Research Center of Vascular Interventional Therapy.

Availability of data and materials

The original contributions of this study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate

The study was approved by the Medical Ethics Committee of Shanxi Medical University, which waived the need for written informed consent due to the retrospective study design.

Consent for publication

Not applicable.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Author details

[1]Division of Cardiology, Department of Internal Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, No. 1095, Jiefang Avenue, Wuhan, Hubei, China
[2]Hubei Provincial Engineering Research Centre of Vascular Interventional Therapy, Wuhan, China

[3]Department of Cardiology, The Second Hospital of Shanxi Medical University, 382 Wuyi Road, Taiyuan 030001, Shanxi, China
[4]Department of Pathology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, No. 1095, Jiefang Avenue, Wuhan 430030, Hubei, China
[5]Department of Cardiology, Tongji Xianning Hospital, Xianning 437011, Hubei, China

Received: 12 March 2025 / Accepted: 6 June 2025
Published online: 03 July 2025

References

1. Kwon Y, Gottmann P, Wang S, et al. Induction of steatosis in primary human hepatocytes recapitulates key pathophysiological aspects of metabolic dysfunction-associated steatotic liver disease. J Hepatol. 2025;82(1):18–27.
2. Yang M, Wang D, Wang X, Mei J, Gong Q. Role of folate in liver diseases. Nutrients. 2024;16(12):1872.
3. Friuli M, Sepe C, Panza E, Travelli C, Paterniti I, Romano A. Autophagy and inflammation an intricate affair in the management of obesity and metabolic disorders: evidence for novel pharmacological strategies? Front Pharmacol. 2024;4(15):1407336.
4. Pan Y, Li Y, Fan H, Cui H, Chen Z, Wang Y, Jiang M, Wang G. Roles of the peroxisome proliferator-activated receptors (PPARs) in the pathogenesis of hepatocellular carcinoma (HCC). Biomed Pharmacotherapy. 2024;1(177):117089.
5. Taru V, Szabo G, Mehal W, et al. Inflammasomes in chronic liver disease: hepatic injury, fibrosis progression and systemic inflammation. J Hepatol. 2024;81(5):895–910.
6. Jouda H, Larrea Murillo L, Wang T. Current progress in vascular engineering and its clinical applications. Cells. 2022;11(3):493.
7. Song L, Jia K, Yang F, Wang J. Advanced nanomedicine approaches for myocardial infarction treatment. Int J Nanomed. 2024;31:6399–425.
8. Griffiths K, Lee JJ, Frenneaux MP, Feelisch M, Madhani M. Nitrite and myocardial ischaemia reperfusion injury Where are we now? Pharmacol Therapeutics. 2021;1(223):107819.
9. Shaya GE, Leucker TM, Jones SR. Coronary heart disease risk lowdensity lipoprotein and beyond. Trends Cardiovasc Med. 2022;32(4):181–94.
10. Porsch F, Binder CJ. Autoimmune diseases and atherosclerotic cardiovascular disease. Nature Rev Cardiol. 2024;21(11):780–807.
11. Xing Y, Lin X. Challenges and advances in the management of inflammation in atherosclerosis. J adv res. 2024;71:317–35.
12. Katta N, Loethen T, Lavie CJ, Alpert MA. Obesity and coronary heart disease: epidemiology, pathology, and coronary artery imaging. Curr Probl Cardiol. 2021;46(3):100655.
13. Amiri MP, Khoshkam M, Salek RM, Madadi R, Ganji GF, Ramazani A. Metabolomics in early detection and prognosis of acute coronary syndrome. Clin Chimica acta. 2019;1(495):43–53.
14. Totoń-Żurańska J, Mikolajczyk TP, Saju B, Guzik TJ. Vascular remodelling in cardiovascular diseases: hypertension, oxidation, and inflammation. Clin Sci. 2024;138(13):817–50.
15. Toh JZ, Pan XH, Tay PW, Ng CH, Yong JN, Xiao J, Koh JH, Tan EY, Tan EX, Dan YY, Loh PH. A meta-analysis on the global prevalence, risk factors and screening of coronary heart disease in nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol. 2022;20(11):2462–73.
16. Vega L, Simian D, Gajardo AI, Salinas M, Urra A, Cattaneo M, Pino R, Roblero JP, Urzúa Á, Rojas K, Poniachik J. Coronary artery disease as a risk factor for metabolic dysfunction-associated steatotic liver disease and liver fibrosis. Ann Hepatol. 2024;29(4):101511.
17. Dugani SB, Moorthy MV, Li C, Demler OV, Alsheikh-Ali AA, Ridker PM, Glynn RJ, Mora S. Association of lipid, inflammatory, and metabolic biomarkers with age at onset for incident coronary heart disease in women. JAMA Cardiol. 2021;6(4):437–47.
18. Sheka AC, Adeyi O, Thompson J, Hameed B, Crawford PA, Ikramuddin S. Nonalcoholic steatohepatitis: a review. Jama. 2020;323(12):1175–83.
19. Dong W, Gong Y, Zhao J, Wang Y, Li B, Yang Y. A combined analysis of TyG index, SII index, and SIRI index: positive association with CHD risk and coronary atherosclerosis severity in patients with NAFLD. Front Endocrinol. 2024;8(14):1281839.
20. Rampidis GP, Benetos G, Benz DC, Giannopoulos AA, Buechel RR. A guide for Gensini Score calculation. Atherosclerosis. 2019;1(287):181–3.
21. Campos-Murguía A, Ruiz-Margáin A, González-Regueiro JA, Macías-Rodríguez RU. Clinical assessment and management of liver fibrosis in non-alcoholic fatty liver disease. World J Gastroenterol. 2020;26(39):5919.
22. Lundberg S M, Lee S-I J A P A. Consistent feature attribution for tree ensembles [J]. 2017.
23. Nielsen RL, Monfeuga T, Kitchen RR, Egerod L, Leal LG, Schreyer AT, Gade FS, Sun C, Helenius M, Simonsen L, Willert M. Data-driven identification of predictive risk biomarkers for subgroups of osteoarthritis using interpretable machine learning. Nature Commun. 2024;15(1):2817.
24. Stone PH, Libby P, Boden WE. Fundamental pathobiology of coronary atherosclerosis and clinical implications for chronic ischemic heart disease management—the plaque hypothesis: a narrative review. JAMA Cardiol. 2023;8(2):192–201.
25. Krenkel O, Tacke F. Liver macrophages in tissue homeostasis and disease. Nature Rev Immunol. 2017;17(5):306–21.
26. Krenkel O, Puengel T, Govaere O, Abdallah AT, Mossanen JC, Kohlhepp M, Liepelt A, Lefebvre E, Luedde T, Hellerbrand C, Weiskirchen R. Therapeutic inhibition of inflammatory monocyte recruitment reduces steatohepatitis and liver fibrosis. Hepatology. 2018;67(4):1270–83.
27. Afanasieva OI, Tyurina AV, Klesareva EA, Arefieva TI, Ezhov MV, Pokrovsky SN. Lipoprotein (a), immune cells and cardiovascular outcomes in patients with premature coronary heart disease. J Personal Med. 2022;12(2):269.
28. Jorch SK, Kubes P. An emerging role for neutrophil extracellular traps in noninfectious disease. Nature Med. 2017;23(3):279–87.
29. Xu M, Chen R, Liu L, Liu X, Hou J, Liao J, Zhang P, Huang J, Lu L, Chen L, Fan M. Systemic immune-inflammation index and incident cardiovascular diseases among middle-aged and elderly Chinese adults: the Dongfeng-Tongji cohort study. Atherosclerosis. 2021;1(323):20–9.
30. Peng A, Zhang B, Wang S, Feng Y, Liu S, Liu C, Li S, Li F, Peng Y, Wan J. Comparison of the value of various complex indexes of blood cell types and lipid levels in coronary heart disease. Front Cardiovasc Med. 2023;15(10):1284491.
31. Shah N, Parikh V, Patel N, Patel N, Badheka A, Deshmukh A, Rathod A, Lafferty J. Neutrophil lymphocyte ratio significantly improves the Framingham risk score in prediction of coronary heart disease mortality: insights from the national health and nutrition examination survey-III. Int J Cardiol. 2014;171(3):390–7.
32. Sun W, Fang Y, Zhou B, Mao G, Cheng J, Zhang X, Liu Y, Chen H. The association of systemic inflammatory biomarkers with non-alcoholic fatty liver disease: a large population-based cross-sectional study. Prevent Med Rep. 2024;1(37):102536.
33. Zhao E, Cheng Y, Yu C, et al. The systemic immune-inflammation index was non-linear associated with all-cause mortality in individuals with nonalcoholic fatty liver disease. Ann Med. 2023;55(1):2197652.
34. Zhao J, Fan H, Wang T, et al. TyG index is positively associated with risk of CHD and coronary atherosclerosis severity among NAFLD patients. Cardiovasc Diabetol. 2022;21(1):123.
35. He L, Chen S, Zhu X, et al. The change of inflammatory markers may predict long-term major adverse cardiovascular events in elderly patients with coronary heart disease: a retrospective cohort study. Front Med. 2025;11:1523581.
36. Cao Y, Li P, Zhang Y, et al. Association of systemic immune inflammatory index with all-cause and cause-specific mortality in hypertensive individuals: results from NHANES. Front Immunol. 2023;14:1087345.
37. Song S, Chen L, Yu R, et al. Neutrophil-to-lymphocyte ratio as a predictor of all-cause and cardiovascular mortality in coronary heart disease and hypertensive patients: a retrospective cohort study. Front Endocrinol. 2024;15:1442165.
38. Qiao Y, Wang Y, Chen C, et al. Association between triglyceride-glucose (TyG) related indices and cardiovascular diseases and mortality among individuals with metabolic dysfunction-associated steatotic liver disease: a cohort study of UK Biobank. Cardiovasc Diabetol. 2025;24(1):12.
39. Ye J, Zhuang X, Li X, et al. Novel metabolic classification for extrahepatic complication of metabolic associated fatty liver disease: a data-driven cluster analysis with international validation. Metabolism. 2022;136:155294.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.