·临床与预防·

基于七种机器学习算法的非肥胖人群非酒精性脂肪性肝病预测模型的构建与验证

邱林杰，任海燕，任燕，李美洁，邹姹姹，吴梓敬，张晋
中国中医科学院西苑医院治未病中心，北京 100091

摘要：目的　探究机器学习方法对非肥胖人群非酒精性脂肪性肝病（non-alcoholic fatty liver disease，NAFLD）的预测价值，并从中分析非肥胖人群 NAFLD 的危险因素。**方法**　回顾性收集并筛选出 2021 年 1 月—2024 年 12 月来自中国中医科学院西苑医院体检中心的 1 517 名非肥胖研究对象的人体成分、既往史、检验等 43 个变量信息，结局指标为腹部超声诊断是否为 NAFLD，并据此分为 NAFLD 组和非 NAFLD 组。通过单因素、LASSO 回归和多因素 logistic 回归分析筛选预测因子，通过 logistic 回归、决策树模型、随机森林、极端梯度提升、轻量级梯度提升机、支持向量机和人工神经网络七种机器学习方法构建并验证非肥胖 NAFLD 预测模型，通过采用沙普利加性解释（Shapley additive explanations, SHAP）法解析变量贡献度。**结果**　纳入的 1 517 例研究对象中 NAFLD 为 215 例（14.2%）。共筛选出 10 个独立预测因子，七种预测模型验证集曲线下面积（area under the curve，AUC）值范围为 0.876～0.924，其中基于 XGBoost 构建的模型区分度最佳（AUC=0.924，95% *CI*：0.895～0.949）。基于 SHAP 值的变量重要性排序结果显示，相位角、ZJU 指数及 UHR 是前三位重要预测因子。**结论**　基于常规体检数据构建的 XGBoost 模型对非肥胖人群 NAFLD 的发病具有较高预测效能，其可解释性特征有助于临床优先识别高危个体。

关键词：非肥胖 NAFLD；机器学习；预测模型；相位角；XGBoost
中图分类号：R575.5　　**文献标志码**：A　　**文章编号**：1003-8507（2026）03-539-07
DOI：10.20043/j.cnki.MPM.202508081

Development and validation of a prediction model for non-alcoholic fatty liver disease in non-obese population using seven machine learning algorithms

QIU Lin-jie, REN Hai-yan, REN Yan, LI Mei-jie, ZOU Cha-cha, WU Zi-jin, ZHANG Jin
Department of Physical Examination, Future Disease Center, Xiyuan Hospital, China Academy of Traditional Chinese Medicine, Beijing 100091, China

Abstract: Objective To investigate the predictive value of machine learning methods for non-alcoholic fatty liver disease (NAFLD) in non-obese populations and to analyze the risk factors for NAFLD within this group. **Methods** Data on 43 variables, including body composition, medical history, and laboratory test results, were retrospectively collected and screened from 1 517 non-obese subjects at the Physical Examination Center of Xiyuan Hospital, China Academy of Chinese Medical Sciences, between January 2021 and December 2024. The outcome indicator was the diagnosis of NAFLD by abdominal ultrasonography, based on which subjects were stratified into NAFLD and non-NAFLD groups. Predictors were screened using univariate analysis, lasso regression, and multivariable logistic regression analysis. Seven machine learning methods including logistic regression, decision tree, random forest, extreme gradient boosting (XGBoost), light hradient boosting machine (LightGBM), support vector machine (SVM), and artificial neural network (ANN), were employed to develop and validate prediction models for NAFLD in non-obese individuals. The shapley additive explanations (SHAP) method was used to interpret variable contributions. **Results** Among the 1 517 included subjects, 215 (14.2%) were diagnosed with NAFLD. Ten independent predictors were identified. The area under the curve (AUC) values for the seven prediction models on the validation set ranged from 0.876 to 0.924. The model based on XGBoost demonstrated the best discriminative ability (AUC=0.924, 95% *CI*: 0.895-0.949). The SHAP value-based variable importance ranking identified phase angle, the ZJU index, and UHR as the top three significant predictors. **Conclusion** The XGBoost model, constructed using routine physical

基金项目：国家重点研发计划资助项目（2018YFC2000600）；北京中医药“薪火传承 3+3 工程”（2023-SZ-A-51）；中国中医科学院科技创新工程（CI2021A03005）
作者简介：邱林杰（1995—），男，博士在读，研究方向：中医治未病、中医药防治代谢性疾病
通信作者：张晋，E-mail: jin_zhang2000@hotmail.com

examination data, exhibits high predictive performance for NAFLD incidence in non-obese populations. Its interpretable features facilitate the clinical prioritization of high-risk individuals.

Keywords: Non-obese NAFLD; Machine learning; Prediction model; Phase angle; XGBoost

非酒精性脂肪性肝病（non-alcoholic fatty liver disease，NAFLD）已经成为一种常见的肝病。它是指一种以肝脏脂肪变性为特征的临床病理综合征，不是由过量饮酒、病毒或药物引起的[1-2]。该综合征包括非酒精性脂肪性肝炎，常伴有炎症、肝细胞损伤、肝硬化，甚至增生性癌和死亡。据估计，全球约有四分之一的人口患有NAFLD，NAFLD 常伴有 2 型糖尿病、肥胖或心血管疾病。值得注意的是，非肥胖 NAFLD 患病率约为 10%～20%，且患病率逐年增加[3]。此外，研究发现与肥胖 NAFLD 相比，非肥胖 NAFLD 疾病进展更快，死亡率更高[4]。因此，早期识别非肥胖 NAFLD 高危人群，对疾病发生发展的预防具有重要意义。在信息技术和多学科交叉发展的当下，机器学习算法非肥胖 NAFLD 预测提供了可靠的有效方法。本研究基于中国中医科学院西苑医院体检中心的体检数据，系统比较七种常用机器学习算法预测非肥胖人群 NAFLD 效能，通过可解释性分析揭示关键风险因子，旨在为非肥胖 NAFLD 防治提供依据。

1 对象与方法

1.1 研究对象 本研究是基于中国中医科学院西苑医院体检中心开展的一项回顾性研究，以 2021 年 1 月—2024 年 12 月于该体检中心体检的 1 517 例健康体检者为调查对象。

纳入标准：年龄大于等于 18 岁且身体质量指数（body mass index,BMI）小于 28 kg/m^2。排除标准：（1）饮酒史；（2）自身免疫性肝病、病毒性、肝炎、肝硬化、肝癌；（3）使用糖皮质激素等引发的药物性肝病；（4）妊娠或哺乳期妇女；（5）健康体检中缺少腹部超声检查者；（6）健康体检中缺少人体成分检查者。本研究已通过中国中医科学院西苑医院伦理委员会审查批准（2025XLA031）。

1.2 相关定义 （1）NAFLD 的诊断：以《非酒精性脂肪肝诊疗指南（2018 年更新版）》为依据，根据腹部 B 超肝前场回声增强（“明亮肝”）、远场回声衰减，以及肝内管道结构显示不清楚等特征进行判别；（2）全身免疫炎症指数（systemic immune inflammation,SII）的定义：SII=P × N/L，其中 P、N 和 L 分别对应外周血中血小板、中性粒细胞和淋巴细胞计数；（3）系统炎症反应指数（systemic inflammatory response index ,SIRI）的定义：SIRI=M × N/L，其中 M 为外周血中单核细胞计数；（4）全身性炎症聚集指数（systemic inflammatory score index,AISI）的定义：AISI=P × M × N/L；（5）甘油三酯－葡萄糖指数（triglyceride-glucose index,TyG）的定义：TyG=Ln[TG（mg/dl）× FPG（mg/dl）/2]，其中 TG 为甘油三酯，FPG 为空腹血糖；（6）甘油三酯葡萄糖 -BMI（triglyceride glucose index-BMI,TyG-BMI）的定义：TyG-BMI=Ln［TG（mg/dl）× FPG（mg/dl）/2］× BMI（kg/m^2）；（7）胰岛素抵抗代谢评分（metabolic score for insulin resistance,METS-IR）的定义：METS-IR=Ln[2 × FPG（mg/dl）+TG（mg/dl）× BMI（kg/m^2）]/Ln[HDL-C（mg/dl）]，其中 HDL-C为高密度脂蛋白胆固醇；（8）血清尿酸与高密度脂蛋白胆固醇比值（serum uric acid to high-density lipoprotein cholesterol ratio,UHR）的定义：UHR=UA（mg/dl）/HDL-C（mg/dl），其中 UA 为血清尿酸；（9）非高密度与高密度脂蛋白胆固醇比值（non-high-density lipoprotein cholesterol to high-density lipoprotein cholesterol ratio，NHHR）的定义：NHHR=[TC（mg/dl）-HDL-C（mg/dl）]/HDL-C（mg/dl），其中 TC 为总胆固醇；（10）浙江大学指数（Zhejiang University index,ZJU）的定义：ZJU =FPG (mmol/L)+BMI (kg/m^2) +3 × ALT(U/L)/AST(U/L) (+2，如果是女性)+TG (mmol/L)；其中 ALT 为丙氨酸氨基转移酶，AST 为天门冬氨酸氨基转移酶；（11）肌脂比的定义：肌肉与脂肪的比值；（12）体脂肪、肌肉量、骨骼肌、体脂百分比、基础代谢率、内脏脂肪面积、相位角及骨骼肌指数是通过 Inbody 770 身体成分分析仪检测获得的。

1.3 缺失值处理 缺失值分两种类型处理：（1）缺失值超过 15%的变量予以删除；（2）缺失值低于 15%的变量采用多重插补进行填补。

1.4 统计分析 使用 R 4.5.1 和 Python 3.10 统计软件进行数据分析。不遵循正态分布的连续变量使用［$M(P_{25},P_{75})$］进行描述，并使用非参数 Mann-Whitney 秩和检验比较组间差异。分类变量用例（%）表示，使用 χ^2 检验分析组间差异。

将研究数据按照 7∶3 随机划分为训练集和验证集。本研究采用分步式的变量处理策略。首先，基于单因素分析、LASSO 回归和多因素 logistic 回归进行特征初筛。该策略旨在从大量临床变量中，筛选出具有明确统计学意义和临床解释性的核心预测因子集合，以增强最终模型的透明度和落地潜力。随后，将此核心变量集通过 logistic 回归、决策树模型（decision tree, DT）、随机森林（random forest）、极端梯度提升（extreme gradient boosting, XGBoost）、轻量级梯度提升机（light gradient boosting machine, LightGBM）、支持向量机（support vector machine, SVM）、人工神经网络

(artificial neural network, ANN) 七种机器学习方法构建非肥胖 NAFLD 预测模型,旨在充分利用复杂模型(如树模型、神经网络)的优势,深入挖掘这些核心变量之间潜在的非线性关系与交互效应,以进一步提升预测性能。然后通过绘制受试者工作曲线(receiver operating characteristic curve,ROC),计算各模型 ROC 的曲线下面积(area under the curve,AUC)、灵敏度、特异度、约登指数等来评价模型性能。将区分度最优模型通过沙普利加性解释(Shapley additive explanations, SHAP) 法进行模型解释。使用合成少数过采样技术(synthetic minority over-sampling technique,SMOTE)处理训练集,使发病患者与未发病患者比例接近 1∶1 平衡,以提高模型的效能,并使用十折交叉验证法(cross-validation,CV)和网格搜索法用于寻找模型的最优参数。双侧检验,检验水准 α=0.05。

2 结　果

2.1 基线特征比较　本研究共纳入 1 517 例患者,其中男性 546 例(36.0%),女性 971 例(64.0%)。根据非肥胖人群是否发生 NAFLD 将研究对象分为非 NAFLD 组和 NAFLD 组,其中 NAFLD 组 215 例(14.2%)。不同性别、吸烟、BMI、腰围、臀围、腰臀比、收缩压、舒张压、中性粒细胞计数、血小板计数、单核细胞计数、淋巴细胞计数、ALT、AST、ALT/AST、TC、TG、HDL-C、LDL-C、肌酐、UA、NHHR、UHR、TyG、TyG-BMI、METS-IR、ZJU、体脂肪、相位角、是否高血压、是否糖尿病、是否胆结石等风险因素在非肥胖非 NAFLD 组和非肥胖 NAFLD 组患者中,差异具有统计学意义(P 均<0.05)。见表 1。

表 1　非肥胖非 NAFLD 组和非肥胖 NAFLD 组基线特征的比较[n(%),M(P_{25},P_{75})]

Table 1　Comparison of baseline characteristics between the non-obese non-NAFLD group and the non-obese NAFLD group [n(%), M (P_{25}, P_{75})]

变量	非肥胖非 NAFLD 组(n=1 302)	非肥胖 NAFLD 组(n=215)	统计值(χ^2/Z 值)	P 值
性别			9.052	0.003
女性	853(65.51)	118(54.88)		
男性	449(34.49)	97(45.12)		
吸烟			11.814	0.002
否	1 285(98.69)	205(95.35)		
是	17(1.31)	10(4.65)		
年龄(岁)	47.00 (33.00, 62.00)	49.00 (36.50, 63.00)	−1.620	0.106
BMI(kg/m^2)	23.80(21.60, 25.67)	23.20 (20.60, 25.55)	−2.191	0.028
腰围(cm)	83.50(77.40, 89.35)	81.50 (75.65, 88.35)	−2.723	0.006
臀围(cm)	94.90 (90.82, 98.20)	94.40(90.00, 97.20)	−2.344	0.020
腰臀比	0.88(0.84, 0.92)	0.87 (0.83, 0.91)	−2.591	0.010
收缩压(mm Hg)	122.00 (111.00, 134.00)	129.00(119.50, 140.00)	−5.887	<0.001
舒张压(mm Hg)	70.00 (63.00, 78.00)	75.00(68.50, 82.00)	−6.072	<0.001
中性粒细胞计数($\times 10^9$/L)	3.19(2.59, 3.93)	3.66(3.01, 4.59)	−5.722	<0.001
血小板计数($\times 10^9$/L)	239.50 (204.00, 279.00)	254.00 (224.00, 295.00)	−4.220	<0.001
单核细胞计数($\times 10^9$/L)	0.29(0.24, 0.35)	0.34(0.28, 0.42)	−6.080	<0.001
淋巴细胞计数($\times 10^9$/L)	1.84 (1.50, 2.21)	2.12(1.77, 2.49)	−6.352	<0.001
ALT(U/L)	14.10(10.50, 21.30)	21.70(15.15, 32.20)	−10.543	<0.001
AST(U/L)	18.10(15.50, 21.90)	20.20 (17.05, 25.50)	−5.607	<0.001
ALT/AST	0.79 (0.64, 1.02)	1.05(0.87, 1.40)	−10.991	<0.001
总胆红素(μmol/L)	10.75 (8.40, 14.00)	11.10 (8.05, 14.20)	−0.093	0.931
TC(mmol/L)	4.74 (4.12, 5.42)	4.93 (4.20, 5.53)	−2.113	0.035
TG(mmol/L)	1.04 (0.74, 1.46)	1.37(1.04, 1.91)	−7.934	<0.001
HDL-C(mmol/L)	1.34 (1.13, 1.55)	1.22 (1.02, 1.45)	−5.031	<0.001
LDL-C(mmol/L)	2.92 (2.39, 3.56)	3.16(2.59, 3.76)	−3.090	0.002
肌酐(mmol/L)	67.00 (59.00, 78.00)	71.00 (61.00, 83.00)	−2.912	0.004
FPG(mmol/L)	5.30(5.03, 5.73)	5.57(5.18, 6.19)	−5.584	<0.001
UA(mmol/L)	300.00 (253.00, 357.00)	380.00(338.50, 423.50)	−12.765	<0.001
NHHR	2.53(1.89, 3.28)	2.94(2.38, 3.63)	−5.724	<0.001
UHR	0.10(0.07, 0.13)	0.13(0.09, 0.16)	−8.173	<0.001
TyG	6.81(6.43, 7.22)	7.12 (6.86, 7.54)	−8.572	<0.001
TyG-BMI	160.59 (143.52, 176.84)	166.63(149.58, 181.64)	−2.753	0.006
METS-IR	2.02(1.88, 2.19)	2.17(2.02, 2.34)	−7.232	<0.001
ZJU	34.32 (31.88, 36.48)	36.27 (33.87, 38.97)	−8.441	<0.001
体脂肪(kg)	19.60(15.90, 23.10)	18.40 (14.50, 22.40)	−2.676	0.007
肌肉量(kg)	39.30(35.60, 47.08)	39.40(34.95, 45.65)	−0.891	0.373
肌脂比	2.07(1.72, 2.71)	2.21(1.77, 2.91)	−1.932	0.054
骨骼肌(kg)	22.60(20.20, 27.58)	22.70 (19.75, 26.60)	−1.020	0.309
体脂百分比(%)	31.25(25.80, 35.38)	29.80 (24.45, 34.60)	−1.944	0.052
基础代谢率	1 272.00 (1 188.00, 1 445.75)	1 274.00 (1 173.00, 1 412.00)	−0.881	0.378
内脏脂肪面积(cm^2)	89.50(68.90, 112.30)	89.00(76.00, 107.45)	−0.762	0.448
相位角(°)	5.05 (4.60, 5.50)	4.50(4.50, 4.50)	−13.281	<0.001
骨骼肌指数	6.60 (6.00, 7.40)	6.60 (6.00, 7.20)	−0.942	0.347
高血压			26.812	<0.001
否	973 (74.73)	124 (57.67)		
是	329 (25.27)	91 (42.33)		
糖尿病			10.802	0.002
否	1 157 (88.86)	174 (80.93)		
是	145 (11.14)	41 (19.07)		
胆结石			75.878	<0.001
否	1 286 (98.77)	190 (88.37)		
是	16 (1.23)	25 (11.63)		

2.2　非肥胖 NAFLD 风险的 LASSO回归分析　将单因素筛选出来的风险因子纳入 LASSO回归分析，通过交叉验证确定最优 λ 值以筛选特征变量，结果共筛选出 19 个特征变量，包括性别、血小板计数、中性粒细胞计数、淋巴细胞计数、相位角、腰臀比、体脂肪、ALT、AST、ALT/AST、HDL-C、TG、FPG、UA、UHR、ZJU、高血压、糖尿病、胆结石。筛选过程见图 1、2。

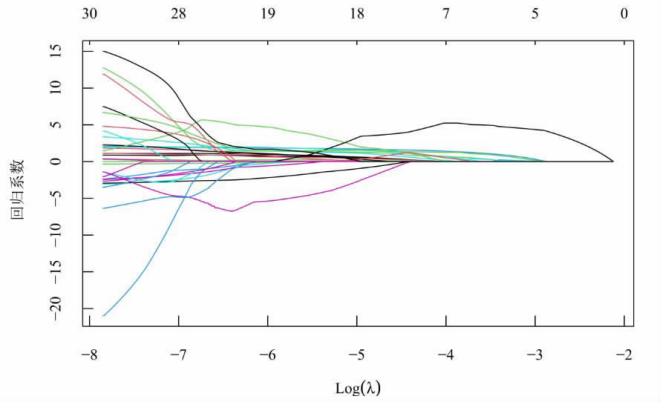


图 1　基于 LASSO 回归分析的非肥胖 NAFLD 发生风险变量选择路径图

Figure 1　Path diagram of risk factor selection for non-obese NAFLD based on lasso regression analysis

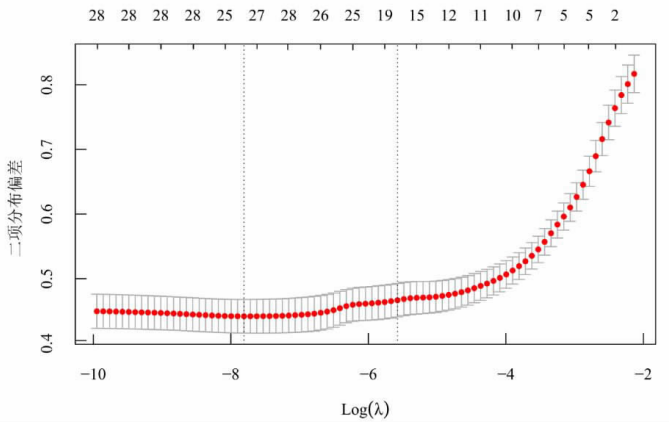


图 2　LASSO 回归分析筛选影响非肥胖 NAFLD 发生的关键特征的交叉验证结果

Figure 2　Cross-validation results of lasso regression analysis screening key characteristics influencing the occurrence of non-obese NAFLD

2.3　非肥胖 NAFLD 风险的多因素 logistic 回归分析　将 LASSO回归筛选的 19 个特征变量纳入多因素 logistic 回归分析中，结果表明血小板计数、相位角、体脂肪、ALT、AST、HDL-C、TG、UHR、ZJU、胆结石是非肥胖 NAFLD 风险的独立危险因素（$P<0.05$）。见表 2。

2.4　模型构建与验证　经 SMOTE 处理后的训练集中，基于多因素 logistic 回归分析筛选出的 10 个预测因子，通过 logistic 回归、DT、随机森林、XGBoost、LightGBM、SVM、ANN 七种机器学习方法构建非肥胖 NAFLD 预测模型，并在验证集中进行内部验证。结果显示，七种模型在验证集中 AUC 取值范围为 0.876～0.924，模型具有较好的区分度，其中基于 XGBoost 构建的模型区分度最佳（AUC=0.924，95% *CI*：0.895～0.949）。见图 3、表 3。

表 2　非肥胖患者 NAFLD 风险影响因素的多因素 logistic 回归分析

Table 2　Multivariable logistic regression analysis of risk factors for NAFLD in non-obese patients

变量	偏回归系数	标准误	Wald χ^2 值	*P* 值
血小板计数(×10⁹/L)	0.006	0.003	2.418	0.016
相位角(°)	−2.277	0.246	−9.266	<0.001
体脂肪(kg)	−0.192	0.045	−4.294	<0.001
ALT(U/L)	−0.046	0.023	−1.985	0.047
AST(U/L)	0.0789	0.030	2.613	0.009
HDL-C(mmol/L)	−3.398	1.042	−3.262	0.001
TG(mmol/L)	−0.339	0.130	−2.608	0.009
UHR	−38.183	10.100	−3.780	<0.001
ZJU	0.598	0.081	7.363	<0.001
胆结石	2.491	0.587	4.243	<0.001

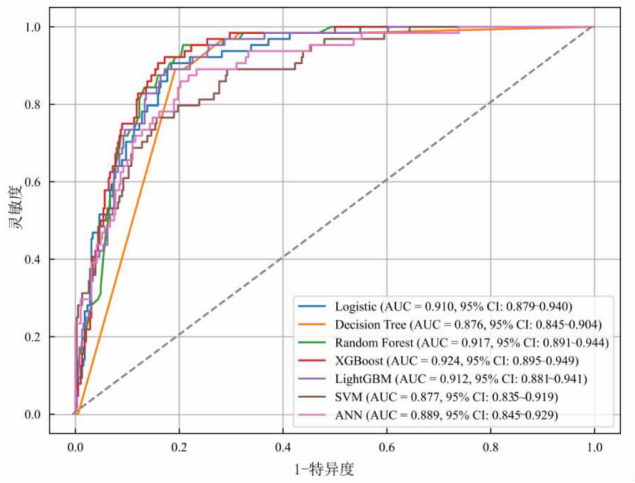


图 3　基于机器学习模型的测试集 ROC

Figure 3　Test set operating characteristic curve based on machine learning model

2.5　SHAP 模型解释　对 AUC 数值最高的 XGBoost 模型进行 SHAP 解释。SHAP 特征性重要性排序结果显示，相位角、ZJU 指数及 UHR 是非肥胖人群 NAFLD 发生风险的前三个重要预测因子，见图 4。SHAP 蜂群图直观地展示了各特征值与模型预测的关系。例如，在前三个重要预测因子中，相位角数值越大对模型有负向影响，是非肥胖人群 NAFLD 发生风险的保护因素；而 ZJU 指数和 UHR 数值越大对模型有正向影响，增加非肥胖人群 NAFLD 发生风险，见图 5。

表 3　七种机器模型测试集筛检结果

Table 3　Screening results for seven machine model test sets

机器学习方法	AUC (95%*CI*)	灵敏度(%)	特异度(%)	约登指数(%)
logistic	91.01(87.87 ~ 93.97)	51.56	94.10	45.67
DT	87.6(84.51 ~ 90.43)	96.88	71.03	67.90
随机森林	91.47(89.90 ~ 94.36)	90.63	81.79	72.42
XGBoost	92.36(89.47 ~ 94.91)	92.19	82.56	74.75
LightGBM	91.17(88.10 ~ 94.10)	90.63	77.95	68.57
SVM	87.75(83.48 ~ 91.87)	76.56	81.54	58.10
ANN	88.89(84.51 ~ 92.85)	78.13	81.79	59.92

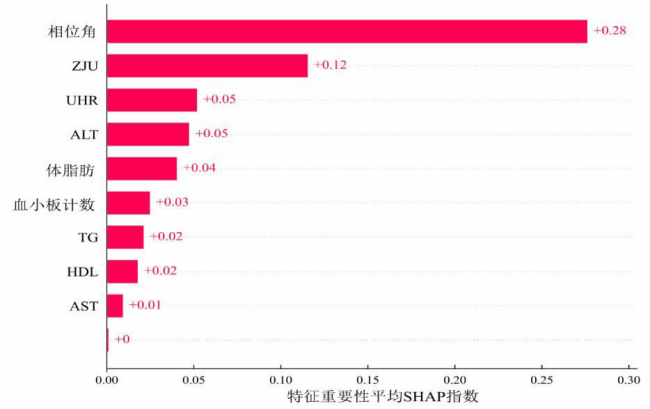


图 4　特征重要性排序图

Figure 4　Feature importance ranking chart

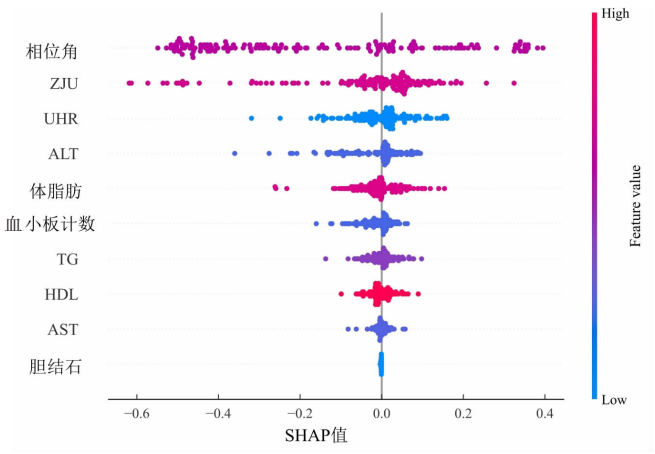


图 5　SHAP 蜂群图

Figure 5　SHAP swarm diagram

3　讨　论

肥胖是发生 NAFLD 的一个众所周知的风险因素[5]，然而越来越多的证据表明，非肥胖人群中 NAFLD 的患病率正在逐渐增加，尤其是亚洲国家[6]。预计到 2040 年，日本的非肥胖 NAFLD 的患病率将增加至 20.7%[7]。在中国的一项前瞻性研究表明，在五年随访期间约有 14.4%的非肥胖人群发生了 NAFLD[8]。在亚洲的一项国际注册研究发现 NAFLD 患者中 21.6%为非肥胖人群，且非肥胖 NAFLD 患者中肝炎和晚期肝纤维化患者的比例分别为 50.5%和 14%[9]。这些研究都表明了 NAFLD 在非肥胖人群的普遍性。目前对非肥胖 NAFLD 的研究相对较少，但非肥胖 NAFLD 通常是无症状的，但其是隐源性肝病的常见病因。因此迫切需要确定非肥胖 NAFLD 的早期和有效的危险因素，建立非肥胖人群 NAFLD 发病风险的预测模型。

机器学习能够自动挖掘海量数据中的复杂非线性关系和模式，处理高维特征，无需预设固定方程形式，相比依赖人工经验和简化假设的传统模型，其预测精度更高、适应性更强[10]。本研究回顾性分析了 1 517 例非肥胖健康体检者的基本信息、人体成分分析、既往疾病史、临床血液检验等指标，使用 logistic 回归、DT、随机森林、XGBoost、LightGBM、SVM、ANN 七种机器学习算法建立并验证了非肥胖人群 NAFLD 发病风险的预测模型，旨在为临床医生提供可靠的病情评估工具，以支持其制定更合理的治疗决策。

尽管本研究中构建的七种机器学习模型均表现出色且区分度差异无统计学意义，我们仍推荐 XGBoost 模型为最优选择。这一推荐不仅基于其在验证集中取得了最高的 AUC 值（0.924）和约登指数（74.75），更考虑到其 92.19%的高灵敏度在疾病筛查中的重要临床价值，以及该模型出色的鲁棒性。尤为关键的是，XGBoost 模型与 SHAP 框架的结合提供了卓越的可解释性，能够清晰量化各特征对预测结果的贡献，这极大地增强了模型的透明度和临床实践的转化潜力。XGBoost 模型在处理复杂、高维的监测数据时展现出卓越的预测精度和泛化能力，能够更精准地识别关键特征[11]。对 XGBoost 模型进行 SHAP 解释，SHAP 特征重要性排序结果显示，相位角、ZJU 指数和 UHR 是非肥胖人群 NAFLD 发病风险的三大重要危险因素。相位角是电流穿透细胞膜时产生的电抗值与电阻值的夹角，电阻是电流流动的限制，主要与组织中的水有关；电抗是组织界面和细胞膜产生的电阻效应，可以代表细胞膜的完整性。相位角是分析评估细胞膜完整性和代谢活性的核心指标，在临床中可作为营养状态评估、炎症监测及慢性疾病预后判断的重要生物标志物[12-13]。来自武汉的一项涉及 953 名超重受试者的研究发现相位角与 NAFLD 之间存在显著的负相关，即超重受试者相位角越高，NAFLD 发病风险越低，并且这种关联在 BMI＜30 kg/m^2 的受试者中显著，但在 BMI≥30 kg/m^2 的受试者中不显著[14]。这

与我们的研究结果具有一致性。本研究发现相位角是非肥胖人群 NAFLD 发病风险的重要因素，这可能是因为 NAFLD 状态下，会导致细胞质量的减少，和细胞膜表面积的损失，从而导致相位角的降低[15]。ZJU 指数是识别中国普通人群 NAFLD 的可靠指标[16]。一项针对中国≥ 60 岁老年人的 NAFLD 风险预测模型显示 ZJU 指数是 NAFLD 预测模型贡献度高的特征[17]。另一项针对非肥胖受试者 ZJU 指数与新发 NAFLD 的前瞻性队列研究也证实随着 ZJU 指数的增加，新发 NAFLD 的累积发病率逐渐增加[18]。我们的研究也表明 ZJU 指数是非肥胖人群 NAFLD 发病风险不可忽视的重要危险因素。NAFLD 的发展与脂质代谢中断、葡萄糖加工异常和脂肪组织功能受损有关[19]。ZJU 指数的组成部分，包括 ALT、AST、TG 和 FPG，可作为肝脏代谢异常的代表性标志物。此外，ZJU 指数还显示了与胰岛素抵抗的深刻联系。胰岛素抵抗是 NAFLD 的重要危险因素，其通过减少脂解抑制而促进游离脂肪酸流向肝脏，进而导致 NAFLD[20]。本研究发现 UHR 是非肥胖人群 NAFLD 发病风险的第三大重要危险因素。来自美国和中国的两项研究均证实 UHR 水平升高与 NAFLD 风险增加显著相关[21-22]。UHR 是一种新的炎症和代谢指标，UA 水平与组织学肝损伤呈正相关，高尿酸血症通过各种机制促进代谢综合征和 NAFLD 的发生，包括胰岛素抵抗、氧化应激和果糖代谢紊乱[23]。另有研究表明 HDL-C 也参与代谢综合征，NAFLD 患者 HDL-C 水平较低[24]。

本研究也存在一些局限性。首先，本研究为单中心研究，缺乏外部数据对模型进行外部验证，未来可能进行大样本多中心的队列数据来进一步拟合并验证模型；其次，本研究依托体检中心的体检数据，由于体检数据的局限性，未对参与者的饮食、运动情况等进行调查，可能导致影响因素的遗漏。最后，本研究人群主要为中国北方人群，需要进一步收集各地区的数据来证实结果的稳健性。

综上所述，本研究建立的 XGboost 模型对非肥胖 NAFLD 发病风险的预测具有良好的效果，可为医生提供良好的临床辅助决策，进而尽早防治 NAFLD 的发生。

利益冲突声明 本研究不存在任何利益冲突

参考文献

[1] Yip TCF, Lee HW, Chan WK, et al. Asian perspective on NAFLD-associated HCC[J]. Journal of Hepatology, 2022, 76(3): 726-734.

[2] Ye Q, Zou BY, Yeo YH, et al. Global prevalence, incidence, and outcomes of non-obese or lean non-alcoholic fatty liver disease: a systematic review and meta-analysis[J]. The Lancet Gastroenterology & Hepatology, 2020, 5(8): 739-752.

[3] Friedman SL, Neuschwander-Tetri BA, Rinella M, et al.Mechanisms of NAFLD development and therapeutic strategies [J]. Nature Medicine, 2018, 24(7): 908-922.

[4] Dai HJ, Xiang JL, Hou YN, et al. Fat mass to fat-free mass ratio and the risk of non-alcoholic fatty liver disease and fibrosis in non-obese and obese individuals[J]. Nutrition & Metabolism, 2021, 18(1): 21.

[5] 邱林杰，任燕，李美洁，等. 136 例不同证型肥胖症患者红外热成像特征研究[J]. 江苏中医药，2022，54(8)：41-44.
Qiu LJ, Ren Y, Li MJ, et al. Study on Infrared Thermal Imaging Characteristics of 136 Patients with Obesity of Different Syndromes [J]. Jiangsu Journal of Traditional Chinese Medicine, 2022, 54 (8): 41-44.(In Chinese)

[6] Han XX, Kong JW, Zhang HM, et al. Triglycerides mediate the influence of body mass index on Non-Alcoholic fatty liver disease in a Non-Obese Chinese population with normal Low-Density lipoprotein cholesterol levels [J]. Obesity Facts, 2024, 17 (2): 191-200.

[7] Yabe Y, Chihara K, Oshida N, et al. Survey of dietary habits and physical activity in Japanese patients with Non-Obese Non-Alcoholic fatty liver disease[J]. Nutrients, 2023, 15(12): 2764.

[8] Sun DQ, Wu SJ, Liu WY, et al.Association of low-density lipoprotein cholesterol within the normal range and NAFLD in the non-obese Chinese population: a cross-sectional and longitudinal study[J]. BMJ Open, 2016, 6(12): e013781.

[9] Tan EXX, Lee JWJ, Jumat NH, et al. Non-obese non-alcoholic fatty liver disease (NAFLD) in Asia: an international registry study [J]. Metabolism, 2022, 126: 154911.

[10] Greener JG, Kandathil SM, Moffat L, et al. A guide to machine learning for biologists [J]. Nature Reviews Molecular Cell Biology, 2022, 23(1): 40-55.

[11] He YX, Chen ZK, Tang ZH, et al. Machine learning with the body roundness index and associated indicators: a new approach to predicting metabolic syndrome[J]. BMC Public Health, 2025, 25(1): 2608.

[12] Bellido D, García-García C, Talluri A, et al. Future lines of research on phase angle: Strengths and limitations [J]. Reviews in Endocrine and Metabolic Disorders, 2023, 24(3): 563-583.

[13] 邱林杰，任燕，李纪新，等. 温阳化饮方治疗脾肾阳虚型肥胖症患者的临床研究[J]. 中医药学报，2025，53(5)：84-90.
Qiu LJ, Ren Y, Li JX, et al. Clinical study on the treatment of Spleen-Kidney yang deficiency type obesity with the wenyang huayin formula [J]. Acta Chinese Medicine and Pharmacology, 2025, 53(5): 84-90.(In Chinese)

[14] Chen G, Lv Y, Ni W, et al. Associations between phase angle values obtained by bioelectrical impedance analysis and nonalcoholic fatty liver disease in an overweight population [J]. Can J Gastroenterol Hepatol, 2020, 8(4): 1-7.

[15] Pagano AP, Sicchieri JMF, Schiavoni IL, et al. Phase angle as a severity indicator for liver diseases[J]. Nutrition, 2020, 70: 110607.

[16] Li LM, You WY, Ren W. Correction to: the ZJU index is a powerful index for identifying NAFLD in the general Chinese population [J]. Acta Diabetologica, 2020, 57(2): 251-251.

[17] Pan H, Liu BC, Luo X, et al. Non-alcoholic fatty liver disease risk prediction model and health management strategies for older Chinese adults: a cross-sectional study [J]. Lipids in Health and Disease, 2023, 22(1): 205.

[18] Zheng KY, Yin YZ, Guo H, et al. Association between the ZJU index and risk of new-onset non-alcoholic fatty liver disease in non-obese

participants: a Chinese longitudinal prospective cohort study [J]. Frontiers in Endocrinology, 2024, 15: 1340644.

[19] Gaggini M, Morelli M, Buzzigoli E, et al. Non-alcoholic fatty liver disease (NAFLD) and its connection with insulin resistance, dyslipidemia, atherosclerosis and coronary heart disease [J]. Nutrients, 2013, 5(5): 1544-1560.

[20] Boushehri YG, Meymanatabadi Z, Tanha AE, et al. Association of triglyceride glucose-body mass index (TyG-BMI)with metabolic dysfunction-associated steatotic liver disease: A systematic review and meta-analysis[J]. PLOS One, 2025, 20(8): e324483.

[21] Xie Y, Huang K, Zhang X, et al. Association of serum uric acid-to-high-density lipoprotein cholesterol ratio with non-alcoholic fatty liver disease in American adults: a population-based analysis [J]. Frontiers of Medicine, 2023, 10: 1164096.

[22] Zhu WT, Liang A, Shi P, et al. Higher serum uric acid to HDL-cholesterol ratio is associated with onset of non-alcoholic fatty liver disease in a non-obese Chinese population with normal blood lipid levels[J]. BMC Gastroenterology, 2022, 22(1): 196.

[23] Chen XY, Dong H, Shi FY, et al. Association of uric acid to high-density cholesterol ratio (UHR) with biological age acceleration: evidence from NHANES 2009-2018[J]. Scientific Reports, 2025, 15 (1): 41847.

[24] Sumner AE, Zhou J, Doumatey A, et al. Low HDL-Cholesterol with normal triglyceride levels is the most common lipid pattern in West Africans and African Americans with metabolic syndrome: implications for cardiovascular disease prevention [J]. CVD Prevention and Control, 2010, 5(3): 75-80.

收稿日期:2025-08-08

（上接第 501 页）

data—— analysis of Cox's proportional hazards regression model of survival data [J]. Sichuan Mental Health, 2020, 33 (1): 27-32.(In Chinese)

[8] 何福逵，周钰山，王欣，等. 休闲活动,膳食多样性及体脂水平对老年人身心脑共病的影响研究[J]. 现代预防医学，2025，52（4）：642-647, 684.
He FK, Zhou YS, Wang X, et al. Research on the effects of leisure activities, dietary diversity and body fat levels on physical, mental and brain comorbidity in the older adults [J]. Modern Preventive Medicine, 2025, 52(4): 642-647, 684.(In Chinese)

[9] 曹淑丽. 文体活动优化老年心理健康的策略与实践分析[J]. 老龄化研究，2024，11（6）：2691-2697.
Cao SL. Strategies and practices for optimizing the mental health of the elderly through cultural and physical activities [J]. Aging Research, 2024, 11(6): 2691-2697.(In Chinese)

[10] 曹红梅，何新羊. 积极老龄化视域下社会活动参与对老年人健康的影响[J]. 江苏社会科学，2022,（2）：166-175.
Cao HM, He XY. The impact of social activity participation on the health of the older adults from the perspective of active aging [J]. Jiangsu Social Sciences, 2022, (2): 166-175.(In Chinese)

[11] 陈弘，秦栎. 健康老龄化背景下社会支持对农村老年人健康的影响研究——基于 2020 年 CHARLS 数据的实证分析[J]. 中国卫生事业管理，2024，41（7）：823-829.
Chen H, Qin L. Empirical analysis of the impact of social support on the health of rural older People in the context of healthy ageing based on 2020 CHARLS data [J]. Chinese Health Service Management, 2024, 41(7): 823-829.(In Chinese)

[12] 刘玟，温勇. 社会支持和代际支持对老年人健康的影响[J]. 生产力研究，2024,（1）：44-50.
Liu M, Wen Y. The impact of social support and inter-generational support on the health of the older adults [J]. Productivity Research, 2024, (1): 44-50.(In Chinese)

[13] 李阳，王振，曾智. 丧偶对我国老年人心理健康的影响:基于精神虚弱指数视角的研究[J]. 中国全科医学，2024，27（6）：663-669.
Li Y, Wang Z, Zeng Z. The impact of widowhood on the mental health of older adults in China: analysis based on the perspective of mental frailty index [J]. Chinese General Practice, 2024, 27 (6): 663-669.(In Chinese)

收稿日期:2025-07-24

（上接第 524 页）

monosodium glutamate on testicular morphology: insights into pyroptosis and therapeutic potential of resveratrol [J]. Folia Morphologica, 2025, 84(1): 151-166.

[25] Mohammadi Z, Alaee S, Namavar MR, et al. The antioxidant properties of resveratrol on sperm parameters, testicular tissue, antioxidant capacity, and lipid peroxidation in isoflurane-induced toxicity in mice[J]. Human & Experimental Toxicology, 2023, 42(1): 36.

[26] 刘贵珊，杨博，张泽生，等. 白藜芦醇对 D- 半乳糖致衰老小鼠学习记忆能力和脑组织抗氧化能力的影响 [J]. 食品科学，2014，35（5）：204-207.
Liu GS, Yang B, Zhang ZS, et al. Effects of resveratrol on learning, memory and brain antioxidant abilities in D-Galactose-Induced aging mice[J]. Food Science, 2014, 35(5): 204-207.(In Chinese)

[27] 朱根生，夏苏干，佘进进，等. 白藜芦醇对玉米赤霉烯酮致小鼠肝脏氧化损伤及炎症的保护作用[J]. 中国畜牧兽医，2021，48（11）：4254-4261.
Zhu GS, Xia SG, She JJ, et al. Protective effects of resveratrol on zearalenone-induced liver oxidative damage and inflammation in mice [J]. China Animal Husbandry & Veterinary Medicine, 2021, 48 (11): 4254-4261.(In Chinese)

[28] Ferramosca A, Lorenzetti S, Di Giacomo M, et al. Modulation of human sperm mitochondrial respiration efficiency by plant polyphenols[J]. Antioxidants, 2021, 10(2): 217.

[29] 石拴霞，王纪田，魏璐晓，等. 白藜芦醇治疗精子线粒体损伤导致的男性不育的作用机制研究进展[J]. 中国性科学，2023，32（9）：5-9.
Shi SX, Wang JT, Wei LX, et al. Research progress on the mechanism of resveratrol in the treatment of male infertility caused by sperm mitochondrial damage [J]. Chinese Journal of Human Sexuality, 2023, 32(9): 5-9.(In Chinese)

收稿日期:2025-09-27