

基于生物信息学技术构建代谢功能障碍相关脂肪性肝病风险预测模型的研究进展

卢 琛, 钟东升

赣南医科大学第一临床医学院, 江西 赣州

收稿日期: 2026年8月2日; 录用日期: 2026年8月26日; 发布日期: 2026年9月4日

摘 要

代谢功能障碍相关脂肪性肝病(Metabolic Associated Steatotic Liver Disease, MASLD)是全球发病率最高的慢性肝脏疾病之一, 其特征是肝脏脂肪过度积累, 与胰岛素抵抗相关, 并定义为肝细胞中脂肪变性>5%的组织学存在。MASLD疾病谱广泛, 早期阶段往往没有明显症状, 导致很多患者错过了最佳干预时机。因此, 通过有效手段早期识别高危人群并进行风险分层, 对于阻止或延缓疾病进展具有重要意义。随着生物信息学技术的飞速发展, 基于多组学数据的MASLD风险预测模型研究取得了显著进展, 本综述总结了风险预测模型最新研究进展及其特点。

关键词

代谢功能障碍相关脂肪性肝病, 生物信息学, 风险预测模型

Research Progress on Constructing Risk Prediction Models for Metabolic Associated Steatotic Liver Disease Based on Bioinformatics Technology

Chen Lu, Dongsheng Zhong

The First Clinical Medical School of Gannan Medical University, Ganzhou Jiangxi

Received: August 2, 2026; accepted: August 26, 2026; published: September 4, 2026

Abstract

Metabolic Associated Steatotic Liver Disease (MASLD) is one of the most prevalent chronic liver

文章引用: 卢琛, 钟东升. 基于生物信息学技术构建代谢功能障碍相关脂肪性肝病风险预测模型的研究进展[J]. 临床医学进展, 2026, 16(9): 529-533. DOI: 10.12677/acm.2026.1693076

diseases in the world. MASLD is characterized by excessive accumulation of fat in the liver and is associated with insulin resistance. And defined as the histological presence of steatosis >5% in hepatocytes. MASLD has a wide spectrum of diseases, and there are often no obvious symptoms in the early stage, resulting in many patients missing the best opportunity for intervention. Therefore, early identification and risk stratification of high-risk groups by effective means is of great significance for preventing or delaying disease progression. With the rapid development of bioinformatics technology, significant progress has been made in the research of MASLD risk prediction models based on multi-omics data. This review summarizes the latest research progress and characteristics of risk prediction models.

Keywords

Metabolic Associated Steatotic Liver Disease, Bioinformatics, Risk Prediction Model

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

代谢功能障碍相关脂肪性肝病(MASLD)已经成为全球最常见的慢性肝病之一,影响全球超过 30%的成年人口,并且是导致肝病相关发病率和死亡率最普遍的原因[1][2]。其特征是肝脏脂肪过度积累,与胰岛素抵抗相关,并定义为肝细胞中脂肪变性>5%的组织学存在[3]。MASLD 疾病谱广泛,可从简单的肝脏脂肪堆积(单纯性脂肪肝)逐步发展为代谢功能障碍相关脂肪性肝炎(MASH)、肝纤维化、肝硬化,最终可能进展为肝细胞癌(HCC) [4]-[6]。这一自然病程漫长且复杂,但早期阶段往往没有明显症状,导致很多患者错过了最佳干预时机。因此,通过有效手段早期识别高危人群并进行风险分层,对于阻止或延缓疾病进展具有重要意义。随着生物信息学技术的飞速发展,基于多组学数据的 MASLD 风险预测模型研究取得了显著进展。这些技术使研究人员能够从遗传学、分子生物学和临床医学等多个维度深入探究 MASLD 的发病机制,并构建更为精准的风险预测工具。本综述系统梳理了基于生物信息学技术构建 MASLD 风险预测模型的最新研究进展,重点围绕遗传特征解析、多维度数据整合及机器学习算法应用三大方向,以期 MASLD 的风险预测模型构建提供新思路。

2. 基于遗传特征解析的风险预测模型

遗传因素在 MASLD 的发病与进展中发挥关键作用,基于全基因组关联分析(GWAS)及多基因风险评分(PRS)的模型可精准捕捉个体遗传易感性差异,为疾病分层与预后预测提供依据。

2.1. 多基因风险评分的分型与预测价值

既往 PRS 模型多聚焦单一疾病表型,最新研究通过细化遗传位点功能特征,实现了 MASLD 的精准分型预测。Jamialahmadi, O 等人[7]的研究创新性地利用内脏脂肪体积等精准肥胖指标校正 GWAS 分析,从 36,394 例样本中识别出 27 个新的 MASLD 相关遗传位点,其中 6 个在独立队列中得到验证。基于肝脏脂蛋白滞留机制,研究构建了两种分隔多基因风险评分(pPRS): 不一致 pPRS 包含 10 个与肝脏甘油三酯含量和循环甘油三酯关联不一致的位点,主要反映肝脏局限性病变,与 HCC 等严重肝病风险强相关;一致 pPRS 包含 13 个关联一致的位点,对应全身性心代谢综合征相关的 MASLD 表型。该分型策略揭示了 MASLD 的遗传异质性,为精准医疗干预提供了分子靶点。

2.2. 遗传评分与临床指标的整体优化

将遗传特征与临床指标结合可显著提升预测效能。Liu Z 等人[8]利用英国生物库中 12,838 例磁共振成像肝脏质子密度脂肪分数数据, 构建了遗传预测的肝脏脂肪含量(gHFC)评分, 发现其与严重肝病风险呈显著剂量反应关系。通过 gHFC 五等分分层后, 最高遗传风险组中 FIB-4 对 HCC 的预测 AUROC 达 0.819, Forns 评分 AUROC 更是提升至 0.877, 大幅优化了临床纤维化评分的预测价值。这种“遗传分层 + 临床指标”的整合模式, 解决了传统临床评分在不同遗传背景人群中效能差异大的问题。

3. 基于多维度数据整合的风险预测模型

整合基因组学、转录组学、蛋白质组学及临床数据的多组学策略, 可全面捕捉 MASLD 的病理生理特征, 显著提升模型的诊断与预警能力。

3.1. 血浆蛋白质组学驱动的超早期预警

血浆蛋白质作为疾病状态的直接生物标志物, 在 MASLD 超早期预测中展现巨大潜力。Hao Chen 等人构建五种蛋白超早期预测模型[9], 分析 5 万余例未患 MASLD 参与者的血浆蛋白质组学数据, 通过多变量 Cox 回归筛选出 5 种关键蛋白质, 构建的模型可提前 16 年识别 MASLD 高风险个体, 在不同验证队列中 AUC 均高达 0.90, 为疾病一级预防提供了可行工具。

3.2. 多组学整合的诊断模型优化

Zhang J 等人[10]通过“蛋白质组学筛选 - 转录组学定位 - 单细胞测序验证”的多组学流程, 确定了库普弗细胞中特异性表达的差异蛋白, 结合临床参数构建 ABD-LTyG 预测模型。该模型在诊断高风险代谢相关脂肪性肝炎(MASH)方面表现优异, 发现队列与验证队列 AUC 分别达 0.832 和 0.807, 显著优于传统非侵入性方法, 实现了从分子机制到临床应用的转化。

3.3. 基因表达模式的疾病进展预测

基于公共数据库的基因表达分析为模型构建提供了丰富数据源。Lan Y 等人[11]整合 GEO 数据库三个数据集, 通过差异表达分析筛选出 69 个在各疾病阶段均异常调节的基因, 结合加权基因共表达网络分析识别出与 BMI、MASH 及 HCC 显著相关的核心模块。基于 62 个关键基因构建的 LASSO 回归模型, 在不同队列中均保持良好预测性能, 为追踪疾病进展提供了分子层面的预测工具。

4. 基于机器学习算法的风险预测模型

机器学习算法可高效挖掘复杂数据中的潜在关联, 通过特征筛选与模型优化, 实现预测性能与临床实用性的平衡, 尤其适用于大规模人群筛查。

4.1. 常见临床指标的简化模型构建

针对基层医疗机构需求, 基于常规检测指标的简化模型成为研究热点。一项针对中国汉族老年人群的大规模研究[12], 从 14,922 例健康数据中, 通过 Spearman 相关性分析、贪心算法特征筛选, 最终确定 BMI、白蛋白、ALT、葡萄糖、HDL、甘油三酯、肌酐 7 个核心指标, 利用 KNN、SVM、ANN 等算法构建模型。其中 ANN 模型在独立验证集中 AUC 达 0.848, 且指标易获取、检测成本低, 为社区基层筛查提供了实用工具。

4.2. 核心预测因子的精准识别

明确核心预测因子是优化模型性能的关键。一项发表于 PubMed 的研究[13]采用 7 种机器学习算法,

分别基于全变量、胰岛素相关变量等数据集构建模型, 通过 SHAP 值与部分依赖图分析发现, 胰岛素抵抗指数(HOMA-IR)和甘油三酯 - 葡萄糖 - 腰围指数(TyG-WC)是最关键的两个预测因子。基于 HOMA-IR、TyG-WC、年龄、AST、种族 5 个指标构建的最优模型, 平均 AUC 高达 0.960, 为精准风险评估提供了高效方案。

4.3. 特定人群的定制化模型开发

针对特定人群的生理特征开发定制化模型可提升预测准确性。基于中国北京和宁波的儿童数据一项研究[14], 开发了 6~14 岁儿童 MASLD 无创预测模型, 发现腰围身高比(WHtR)是最具成本效益的无创预测指标, 解决了儿童群体 MASLD 筛查缺乏专属工具的问题。针对炎症性肠病(IBD)合并 MASLD 的特殊人群, 研究者识别出吸烟、血脂异常、高血压等独立预测因素, 开发的加法临床评分和逻辑回归评分模型, 预测效能显著优于传统指标。

4.4. 基于 NHANES 的 ML 模型

Zhu, G 等[15]的研究基于 NHANES 数据库, 利用机器学习对基本人口与临床特征进行加权分析, 构建了针对美国人群的 MASLD 患病风险预测模型。该研究采用 10 种机器学习算法构建基于人口学特征的 MASLD 诊断模型。结果表明, 逻辑回归(LR)模型性能最优, 在内部验证队列中准确率、灵敏度和 AUC 分别达到 0.728、0.749 和 0.806, 且在测试队列中表现稳健。决策曲线分析与校准曲线进一步显示 LR 模型具有显著的临床实用性和良好的校准度, 其精确召回曲线也证实其在保持高灵敏度的同时能有效控制误报率, 凸显了 LR 模型在 MASLD 风险预测中的可靠性与临床应用潜力。

5. 总结与展望

当前研究虽取得显著进展, 但仍存在不足: 部分模型基于单中心数据构建, 缺乏多中心、跨人群外部验证; 多组学数据整合多停留在初步层面, 不同组学数据间的因果关联尚未完全明确; 研究链条多止步于模型构建或机制探讨, 未能形成“机制发现 - 模型验证 - 治疗提示”的闭环; 超早期预测模型的生物标志物临床检测试剂盒开发滞后, 转化应用不足。

动态预测模型是当前预后模型领域的研究热点, 其可依托患者最新随访数据实时更新风险水平, 能够更真实地还原疾病演变过程, 助力临床实现个体化风险评估与早期精准干预, 相较于传统静态预测模型具备明显优势。Moreau C 等[16]构建联合潜在类别模型(JLCM), 整合肝纤维化非侵入性指标的纵向动力学特征, 针对 MASLD 人群开展肝脏相关并发症(LRE, 含肝硬化失代偿、肝细胞癌)的动态预后预测。该国际多中心队列共纳入 13 627 例 MASLD 患者, 中位随访 4.0 年共发生 238 例 LRE 事件; 结果显示 JLCM 预测风险与实际事件发生率一致性良好, 不同时间点 AUROC 达 0.872~0.888, 预测效能优于单纯肝脏硬度(LSM)及 FIB-4。研究依据 5 年 LRE 风险划分低、中、高风险层级, 基线及随访中 85%~89%患者归为低风险组; 发生 LRE 的病例中 67%~72%被 JLCM 归入高风险组, 显著高于 LSM (54%~56%)与 FIB-4 (46%~50%), 证实该模型风险分层性能更突出。这类动态建模思路为 MASLD 患者的个体化随访管理提供了新的技术方向。但动态预测模型应用复杂, 需要患者多次连续无创检查结果, 在临床随访中较难完美实现。

未来研究可聚焦三方面: 一是深化多维度数据整合, 结合单细胞测序、空间转录组等新技术, 揭示遗传 - 环境 - 表型间的调控网络; 二是推动模型的临床转化, 开发便携式检测设备与快速分析工具, 适配基层医疗场景; 三是构建动态预测模型, 结合纵向随访数据, 实现个体风险的实时更新与干预效果的精准评估。随着生物信息学技术与临床研究的深度融合, MASLD 风险预测模型将在疾病精准防控中发挥愈发重要的作用。

参考文献

- [1] Miao, L., Targher, G., Byrne, C.D., Cao, Y.Y. and Zheng, M.H. (2024) Current Status and Future Trends of the Global Burden of MASLD. *Trends in Endocrinology & Metabolism*, **35**, 697-707. <https://doi.org/10.1016/j.tem.2024.02.007>
- [2] Chan, W.K., Chuah, K.H., Rajaram, R.B., Lim, L.L., et al. (2023) Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD): A State-of-the-Art Review. *Journal of Obesity & Metabolic Syndrome*, **32**, 197-213. <https://doi.org/10.7570/jomes23052>
- [3] The Lancet Gastroenterology & Hepatology, (2020) Redefining Non-Alcoholic Fatty Liver Disease: What's in a Name? *The Lancet Gastroenterology & Hepatology*, **5**, 419. [https://doi.org/10.1016/s2468-1253\(20\)30091-1](https://doi.org/10.1016/s2468-1253(20)30091-1)
- [4] Powell, E.E., Wong, V.W. and Rinella, M. (2021) Non-Alcoholic Fatty Liver Disease. *The Lancet*, **397**, 2212-2224. [https://doi.org/10.1016/s0140-6736\(20\)32511-3](https://doi.org/10.1016/s0140-6736(20)32511-3)
- [5] Ioannou, G.N. (2021) Epidemiology and Risk-Stratification of NAFLD-Associated HCC. *Journal of Hepatology*, **75**, 1476-1484. <https://doi.org/10.1016/j.jhep.2021.08.012>
- [6] Pouwels, S., Sakran, N., Graham, Y., Leal, A., Pintar, T., Yang, W., et al. (2022) Non-Alcoholic Fatty Liver Disease (NAFLD): A Review of Pathophysiology, Clinical Management and Effects of Weight Loss. *BMC Endocrine Disorders*, **22**, Article No. 63. <https://doi.org/10.1186/s12902-022-00980-1>
- [7] Jamialahmadi, O., De Vincentis, A., Tavaglione, F., Malvestiti, F., et al. (2024) Partitioned Polygenic Risk Scores Identify Distinct Types of Metabolic Dysfunction-Associated Steatotic Liver Disease. *Nature Medicine*, **30**, 3614-3623. <https://doi.org/10.1038/s41591-024-03284-0>
- [8] Liu, Z., Yuan, H., Suo, C., Jin, L., Zhang, T. and Chen, X. (2025) Population Stratification Using MASLD Polygenic Risk Score Improves Severe Liver Disease Prediction by Clinical Fibrosis Scores. *Digestive and Liver Disease*, **57**, 1819-1825. <https://doi.org/10.1016/j.dld.2025.06.016>
- [9] Chen, H., Yu, S.Y., Chen, C.L., et al. (2025) Plasma Proteomic Profiles in Diverse Populations Identify Biomarkers Predicting Metabolic Dysfunction-Associated Steatotic Liver Disease up to 16 Years before Onset. Research Square. <https://doi.org/10.21203/rs.3.rs-6405995/v1>
- [10] Zhang, J., Wang, W., Wang, X.Q., et al. (2025) A Robust Diagnostic Model for High-Risk MASH: Integrating Clinical Parameters and Circulating Biomarkers through a Multi-Omics Approach. *Hepatology International*, **19**, 820-835.
- [11] Lan, Y., Song, R., Feng, D. and He, J. (2025) Bioinformatic Analysis of Molecular Expression Patterns during the Development and Progression of Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD). *Scientific Reports*, **15**, Article No. 7294. <https://doi.org/10.1038/s41598-025-90744-3>
- [12] Wu, N., Feng M.F., Zhao, H.H., et al. (2025) Prediction of Metabolic Dysfunction-Associated Steatotic Liver Disease via Advanced Machine Learning among Chinese Han Population. *Obesity Surgery*, **35**, 4192-4204.
- [13] Chen, H., Zhang, J., Chen, X., Luo, L., Dong, W., Wang, Y., et al. (2025) Development and Validation of Machine Learning Models for MASLD: Based on Multiple Potential Screening Indicators. *Frontiers in Endocrinology*, **15**, Article 1449064. <https://doi.org/10.3389/fendo.2024.1449064>
- [14] Liu, Y., Wang, Y., Xing, Y., Wolters, M., Shi, D., Zhang, P., et al. (2024) Establish a Noninvasive Model to Screen Metabolic Dysfunction-Associated Steatotic Liver Disease in Children Aged 6-14 Years in China and Its Applications in High-Obesity-Risk Countries and Regions. *The Lancet Regional Health—Western Pacific*, **49**, Article 101150. <https://doi.org/10.1016/j.lanwpc.2024.101150>
- [15] Zhu, G., Song, Y., Lu, Z., Yi, Q., Xu, R., Xie, Y., et al. (2025) Machine Learning Models for Predicting Metabolic Dysfunction-Associated Steatotic Liver Disease Prevalence Using Basic Demographic and Clinical Characteristics. *Journal of Translational Medicine*, **23**, Article No. 381. <https://doi.org/10.1186/s12967-025-06387-5>
- [16] Moreau, C., Roux, M., Lin, H., Lee, H.W., Yip, T.C., Tsochatzis, E., et al. (2025) Personalised Prediction of the Individual Risk of Liver-Related Complications in MASLD Using the Dynamics of Non-Invasive Tests of Liver Fibrosis. *Journal of Hepatology*, **82**, S67. [https://doi.org/10.1016/s0168-8278\(25\)00414-3](https://doi.org/10.1016/s0168-8278(25)00414-3)