

· 综述 ·

基于机器学习的慢性肝脏疾病预测模型研究进展

吕阿兴¹ 丛雪¹ 孙点剑^{1,2,3} 余灿清^{1,2,3} 吕筠^{1,2,3,4} 李立明^{1,2,3} 庞元捷^{1,2,3}

1. 北京大学公共卫生学院流行病与卫生统计学系, 北京 100191; 2. 北京大学公众健康与重大疫情防控战略研究中心, 北京 100191; 3. 重大疾病流行病学教育部重点实验室(北京大学), 北京 100191; 4. 血管稳态与重构全国重点实验室

摘要: 慢性肝脏疾病在中国乃至全球范围内造成了沉重的疾病负担。近年来, 在健康医疗大数据的背景下, 机器学习凭借其自动化数据学习模式, 能够应对传统模型难以处理的复杂数据, 更准确地预测慢性肝脏疾病发病风险, 并为高危人群风险分层管理提供理论支持, 在医学研究领域内广泛应用。本综述通过系统检索和归纳慢性肝脏疾病相关的机器学习模型研究, 梳理了这些模型较传统模型在预测能力方面的差异, 旨在探讨机器学习在肝脏疾病风险评估中的应用价值。同时, 本文针对当前研究的目标疾病类型集中、信息报告不完整和缺乏严谨的外部验证等局限性, 提出了未来研究应拓展研究范围、遵循严格的科研规范的建议, 以期促进机器学习模型在慢性肝脏疾病早期预警和分层管理方面的有效应用。

关键词: 机器学习; 慢性肝脏疾病; 预测模型

中图分类号: R575 文献标识码: A 文章编号: 1004-6194(2026)02-0139-08

Machine learning for chronic liver disease prediction: a literature review

LYU Axing*, CONG Xue, SUN Dianjianyi, YU Canqing, LYU Jun, LI Liming, PANG Yuanjie

*Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China

Corresponding author: PANG Yuanjie, E-mail: ypang@bjmu.edu.cn

Abstract: Chronic liver disease (CLD) has imposed significant disease burdens in China and globally. Against the backdrop of healthcare big data, machine learning, with its automated data learning model, is capable of handling complex data that traditional models struggle to deal with, predicts the onset risk of CLD more accurately, and provides theoretical support for the stratification management of high-risk populations, thus being widely employed in medical research. The review systematically searched and summarized the machine learning models related to CLD, sorted out the differences in prediction performance between these models and traditional models, with the intention of exploring the application value of machine learning in CLD risk assessment. At the same time, the review pointed out the limitations of the current research, such as a narrow focus on specific diseases, incomplete reporting of key details, and a lack of rigorous external validation, and proposed that future research should expand the research scope, follow strict scientific norms, in order to promote the extensive application of machine learning models in early warning and stratification management for CLD.

Keywords: Machine learning; Chronic liver diseases; Prediction model

慢性肝脏疾病通常包括非酒精性脂肪性肝病(non-alcoholic fatty liver disease, NAFLD)^[1]、肝硬化^[2]和肝癌^[3]等, 在全球范围内造成了严重的疾病负担。2000—2021

年中国 NAFLD 的发病密度为 6 084.9/10 万人年, 其中 2015—2021 年的发病密度为 8 259.1/10 万人年^[4], 位居全球第一。2021 年中国肝硬化患者为 3.9 亿人, 肝癌患者为 26.6 万人, 分别占全球病例总数的 23.2%、35.9%^[5]。慢性肝脏疾病早期的临床表现隐匿, 往往在病情进展至晚期并出现显著症状后患者才就医, 错过最佳治疗时机^[6]。肝活组织检查是慢性肝脏疾病诊断的金标准, 却因侵入性难以广泛实施^[7-9]。影像学检查则是临床常

DOI: 10.16386/j.cjpcd.issn.1004-6194.20241205.0965

基金项目: 国家重点研发计划(2023YFC3606300), 国家自然科学基金(82304223, 82192901, 82192904, 82192900)

作者简介: 吕阿兴, 硕士研究生在读, 研究方向: 慢性病流行病学, E-mail: lvaxing@hsc.pku.edu.cn

通信作者: 庞元捷, E-mail: ypang@bjmu.edu.cn

居民主要慢性代谢性疾病现状及流行趋势[J]. 卫生研究, 2024, 53(6): 880-886, 913.

[14] 邵继红, 莫宝庆, 喻荣彬, 等. 南京市社区人群高尿酸血症与痛风的流行病学调查[J]. 疾病控制杂志, 2003, 7(4): 305-308.

[15] 洪佳慧, 田瑞平, 魏姝, 等. 天津市体检人群高尿酸血症影响因素及交互作用分析[J]. 社区医学杂志, 2021, 19(13): 788-791.

[16] 李瑛, 郭宏纪. 舟山海岛居民高尿酸血症与痛风的流行病学调查[J]. 全科医学临床与教育, 2025, 23(2): 113-116.

[17] 赵雅茜, 陈桃, 孙蕊, 等. 无症状高尿酸血症的诊治争议[J]. 中国临床新医学, 2021, 14(11): 1068-1071.

[18] 冀盈盈, 高宁宁, 许丽洁, 等. 高尿酸血症肾病的发病机制[J]. 中

华风湿病学杂志, 2025, 29(1): 62-69.

[19] 胡彩红, 程云凤, 王美兰, 等. 南京市 60 岁及以上人群肥胖与高尿酸血症的关联研究[J]. 中国慢性病预防与控制, 2024, 32(11): 815-820.

[20] 毛丹卉. 中国老年人饮食模式与血清尿酸的关系[J]. 中国慢性病预防与控制, 2022, 30(9): 712-716.

[21] 崔巍, 覃珊珊, 朱梦麟. 高尿酸血症运动疗法的研究进展[J]. 新疆中医药, 2024, 42(2): 99-103.

收稿日期: 2024-10-31 修回日期: 2024-12-26 本文编辑: 李翔

用的无创检查手段,但 B 型超声易受操作者主观判断影响,CT、核磁共振成像(MRI)、瞬时弹性成像(transient elastography, TE)以及磁共振瞬时成像(magnetic resonance elastography, MRE)等手段又因成本较高难以普及^[7-9]。因此,开发准确、可靠且易于使用的慢性肝脏疾病风险预测模型,对高危人群实现精准分层与管理,为疾病早期预警和医疗资源优化配置提供证据,对我国慢性肝脏疾病防控至关重要。目前,临床实践中应用较多的慢性肝脏疾病预测模型主要基于社会人口学、体格检查指标和血液生物标志物构建,如肝癌预测模型 GALAD、ASAP 和 aMAP 等^[10-12],肝脏纤维化预测模型 FIB-4、APRI、NFS、BARD 评分和 Forns 指数等^[13],脂肪肝预测模型 FLI、NLFS 和 HSI 等^[13]。此类预测模型优点在于预测因子简单,易于推广使用,但无法考虑复杂预测因子作用和预测因子间的交互作用,往往特异度较高,而灵敏度不足。机器学习通过计算机自动学习能发现数据特征间的未知模式和未知联系,已广泛用于疾病预测^[14]。已有综述总结了人工智能与机器学习方法在诊断肝脏疾病、辅助疾病分期管理和预测临床预后结局等领域的良好表现^[15-17],但存在以下局限性:(1)对慢性肝脏疾病发病预测相关研究举证不足;(2)未区分研究资料类型,其中病理学、影像学资料难以在一般人群中推广使用。因此,本文将聚焦肝癌、肝硬化和 NAFLD 3 种慢性肝脏疾病,对现有基于相对易获得资料开发的机器学习模型预测表现展开综述,并与传统预测模型相比较,探讨机器学习在慢性肝脏疾病风险预测方面的价值,推动未来基于机器学习的慢性肝脏疾病风险预测模型的推广使用。

1 资料与方法

1.1 检索策略 本研究同时检索中文数据库(中国知网、万方数据知识服务平台和维普期刊中文期刊服务平台)和英文数据库(PubMed、Scopus、Web of Science 和 Embase),以肝脏疾病(肝癌、肝硬化和 NAFLD 等)、机器学习、风险评估和预测模型为主题词或关键词进行检索,并将日期限定为 2024 年 7 月 15 日及之前。

1.2 纳入和排除标准 纳入标准:基于队列研究预测慢性肝脏疾病发病风险的机器学习模型文章。排除标准:(1)非人群队列研究;(2)儿童、青少年以及孕产妇等特殊人群研究;(3)针对接受临床处理人群的预后研究;(4)肝脏罕见病的研究(如肝豆状核变性、先天性胆汁酸合成障碍等);(5)预测因子包括组学数据(如基因组学、代谢组学、影像组学)或病理学数据的研究。

1.3 文献筛选与数据提取 首先对所检索文献剔除重复,后续通过标题与摘要初步筛选,最终通过阅读全文确定入选文献。本研究参考“个体预后与诊断多变量预测模型报告规范(transparent reporting of a multivariable

prediction model for individual prognosis or diagnosis, TRIPOD)+人工智能(artificial intelligence, AI)声明”^[18]对纳入文献进行信息提取,提取内容包括文献基本情况(作者、发表年份、国家或地区)、研究设计(研究设计类型、纳入起止时间、随访时长)、人群特征(纳入人群类型、人数、年龄、国家或地区)、结局特征(目标疾病、发病时间、疾病定义)和模型特征(模型类型、预测因子、预测指标、性能表现)。本研究将统计模型,包括 logistic 回归(logistic regression, LR)、Cox 比例风险模型(Cox proportional hazards model, Cox)、朴素贝叶斯(naive Bayes, NB)等以及常用的临床风险评分(如 FIB-4、APRI 和 GALAD 等)视为传统模型。经典机器学习模型如随机森林(random forest, RF)、支持向量机(support vector machine, SVM)等以及深度学习方法,则被归类为机器学习模型。文献筛选与数据提取由两名研究者独立完成并交叉核对,若存在分歧,则征询第 3 名研究者的意见。

1.4 统计学分析 所有统计学分析均使用 R 4.4.0 软件完成。在纳入的 20 项研究中,仅有 5 项研究进行了外部验证,而全部研究进行了内部验证或训练集的交叉验证。因此,本研究主要研究指标为内部验证的受试者工作特征曲线下面积(area under the receiver operator characteristic curve, AUROC),未进行内部验证的研究采用训练集的交叉验证结果代替,同时选择结果最好的传统模型和机器学习模型在森林图中展示。对于未报告置信区间的研究,通过以下公式使用 AUROC 值、发病人数与总人数估计其标准误^[19]。

$$SE(\theta) = \sqrt{\frac{(\theta(1-\theta) + (n_A - 1)(Q_1 - \theta)^2) + (N - n_A - 1)(Q_2 - \theta)^2}{n_A(N - n_A)}}$$

其中 θ 为 AUROC 值, n_A 为研究中的病例数, N 为研究中的总人数, $Q_1 = \theta \div (2 - \theta)$, 表示在随机选取的一个健康个体和两个发病个体中,模型对这两个发病个体的预测风险值均高于健康个体的概率; $Q_2 = 2\theta \div (1 + \theta)$, 表示在随机选取的两个健康个体和一个发病个体中,模型对该发病个体的预测风险值高于这两个健康个体中至少一个的概率。

2 结果

2.1 纳入文献情况 在剔除重复文献后,本研究共检索出 1 183 篇英文文献和 110 篇中文文献。根据纳入排除标准,最终纳入 20 篇,均为英文文献^[20-39]。根据研究对象可以进一步分为针对一般人群或高危人群的研究,后者主要为病毒性肝炎、脂肪肝等肝脏疾病患者。其中,针对一般人群的研究共 8 项,包括结局为肝癌的研究 1 项、肝硬化及纤维化 3 项、NAFLD 4 项(Goldman 的研究包括肝纤维化和 NAFLD 2 种结局^[25]),见表 1;针对高危人群的研究共 13 项,包括肝癌 4 项、肝硬化

及纤维化 9 项,见表 2。在本研究中,共有 14 项研究同时汇报了机器学习模型与传统模型的 AUROC,其中多数以肝硬化及纤维化为结局 ($n=9$)。在机器学习模型中,RF 模型[包括基于 RF 的其他模型,如随机生存森林(random survival forest,RSF)和脂肪肝纤维化随机森林评分(MAFLD fibrosis score,MFS)等]被采用得最多($n=8$)。而传统模型中,LR 模型[包括基于 LR 的其他模型,如纵向 LR 模型和脂肪肝纤维化评分(steatosis-associated fibrosis estimator score,SAFE-score),但不包括基于 LR 的、已广泛作为风险评分截断值的模型,如 GALAD、FIB-4 等]的使用频率最高($n=9$)。

2.2 以肝癌为结局的预测模型 共 5 项研究利用机器学习模型预测肝癌风险,主要针对欧美和东亚人群,结局均为肝细胞癌,预测因子数目至少为 8 个,中位数为 18 个。这些研究基于体检数据^[32]、疾病监测数据^[34]或电子病历信息^[33,35-36],汇报的 AUROC 值范围为 0.759~0.966,见图 1。其中,具有代表性的是逆时注意力模型(reverse time attention model,RETAIN),这是一种基于循环神经网络(recurrent neural network,RNN)的分类方法,特点是能通过内部门控循环单位有效利用电子病历信息中的纵向医学数据,将电子病历中大量的人口学特征、疾病诊断代码、药物通用名称、实验室检查代码和临床事件代码等加权汇总,计算疾病进展概率,从而预测 NAFLD 人群肝癌发病风险,AUROC 高达

0.966(95%CI 0.947~0.985)^[35]。

对于一般人群的肝癌预测模型,An 等^[32]开发的“9 年-肝细胞癌”预测模型表现出优秀的预测能力。该模型利用韩国 2004—2007 年进行健康体检的 417 346 名健康人群及其后续 9 年的随访资料开发,模型通过 Cox 模型筛选出 13 个模型预测因子,建立随机生存森林、极限梯度提升机(extreme gradient boosting,XGBoost)模型以及二者结合的集成模型(RSF+XGBoost)。结果显示,RSF+XGBoost 在训练过程与内部验证过程的 AUROC 分别为 0.871 (95%CI 0.852~0.890)与 0.873(95%CI 0.860~0.885),表现出最强的预测能力和优秀的稳健性。

对于高危人群,Johnson 等^[34]同时建立了 2 种 LR 模型[GALAD、GALAD-Ogaki(基于大阪市立医院数据的 GALAD 评分)]和 2 种 RF 模型[GALAD-RF(与 GALAD 预测因子一致)、RF-Practical(基于 8 种常规实验室指标)]探索机器学习模型用于肝癌预测的应用价值。这些模型利用日本一家医院的监测项目,纳入 1998—2014 年 3 473 例基线无肝癌因慢性肝脏疾病在该医院就诊的患者,用以预测肝细胞癌发病风险,结果显示,RF-Practical 的预测表现最好,AUROC 为 0.911 (95%CI : 0.896~0.926)。此外,Joannou 等^[33]基于美国退伍军人健康管理局(Veterans Health Administration,VHA)中 48 151 例丙型肝炎相关肝硬化患者的数据,利用 31 个预测因

表 1 一般人群慢性肝脏疾病预测模型信息

作者和年份	预测年限-疾病	机器学习模型(传统模型)	预测因子数	数据集人数	评价指标
An 2021 ^[32]	9 年-肝细胞癌	RSF,XGBoost,RSF+XGBoost	13	训练集 331 694,内部验证集: 85 652	AUROC
Goldman 2021 ^[25]	肝纤维化	CHAID	140	训练集 4 438(5 折交叉验证)	AUROC,准确率
Ben-Assuli 2022 ^[20]	肝纤维化	HMM,GBTM,CHAID,XGBoost	84	训练集 5 579(5 折交叉验证)	AUROC,错分率,准确率,精确度
Blanes Vidal 2022 ^[21]	肝硬度	LiverAID,XXS,LiverAID,XS,LiverAID,S,LiverAID,M,LiverAID,L,LiverAID,4XL(LR,FIB-4,Forns,APRI,FIB-4+Forns+APRI,TE)	233	训练集 1 810,内部验证集: 335	AUROC,灵敏度,特异度,准确率,阳性预测值,阴性预测值
Goldman 2021 ^[25]	NAFLD	CHAID	140	训练集 7 581(5 折交叉验证)	AUROC,准确率
Huang 2023 ^[38]	5 年-NAFLD	DT,SVM,RF,CatBoost,XGBoost(LR)	11	训练集 4 337,内部验证集: 1 859,外部验证集 6 196	AUROC,准确率,精确度,召回率,F1 分数
Cao 2024 ^[37]	5 年-NAFLD	XGBoost,DT(LR,NB)	16	训练集 15 498,内部验证集: 6 642	AUROC,准确率,精确度,召回率,F1 分数
Zeng 2024 ^[39]	5 年-NAFLD	LightGBM,SVM,RF(LR)	272	训练集 1 546,内部验证集: 337,外部验证集 749	AUROC

注:APRI.天冬氨酸氨基转移酶/血小板比值指数;AUROC.受试者工作特征曲线下面积;CatBoost.梯度提升树;CHAID.卡方自动交互检测模型;Cox.Cox 比例风险模型;DT.决策树;FIB-4.纤维化-4 指数;Forns.Forns 指数;GBTM.群体轨迹模型;HMM.隐马尔可夫模型;LightGBM.轻量级梯度提升机;LiverAID.肝硬度机器学习集成模型(编号分别为 XXS,XS,S,M,L,4XL);LR.logistic 回归;NAFLD.非酒精性脂肪性肝病;NB.朴素贝叶斯;RF.随机森林;RSF.随机生存森林;SVM.支持向量机;TE.瞬时弹性成像;XGBoost.极限梯度提升机。

表 2 高危人群预测模型信息

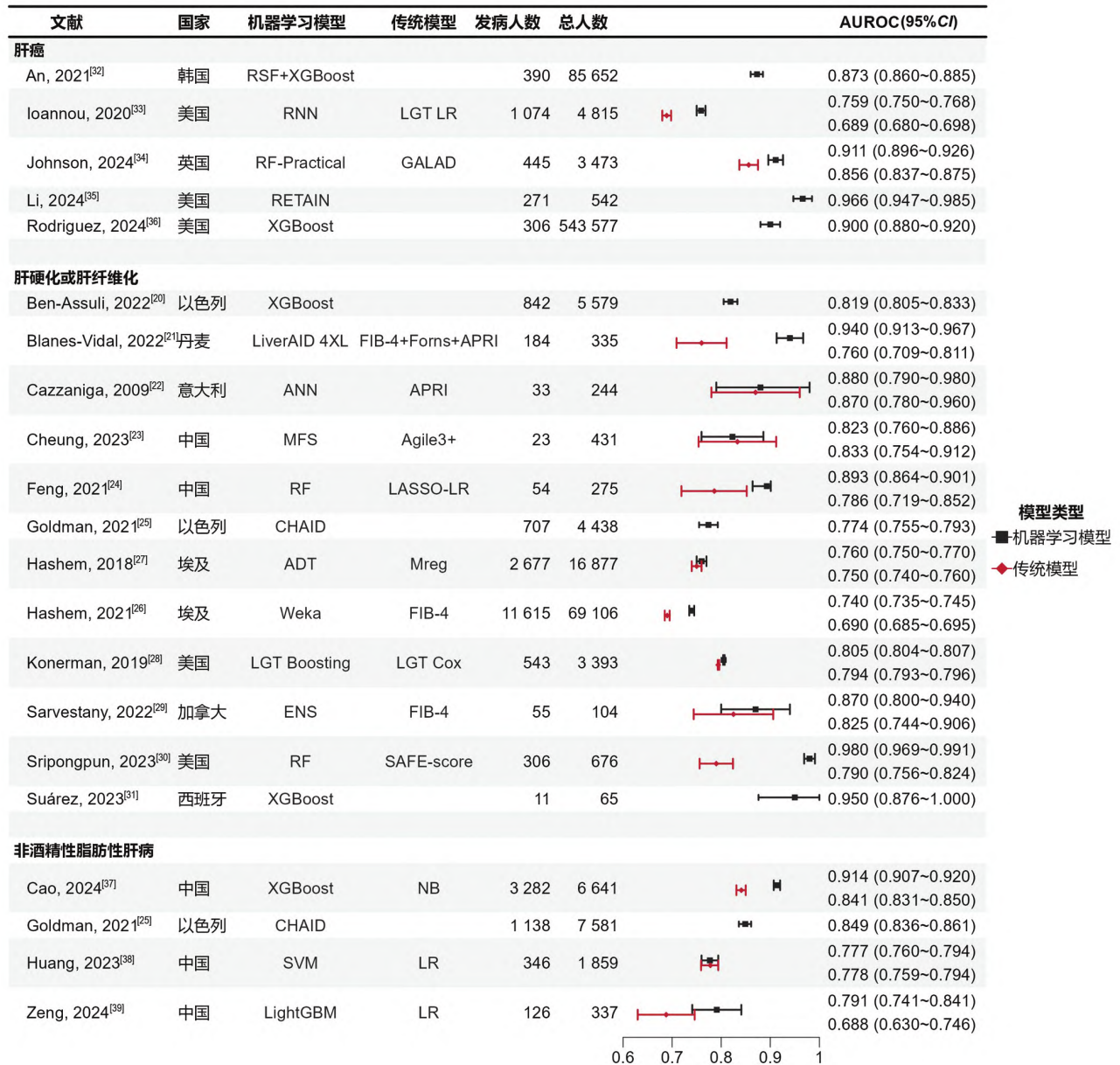
作者和年份	人群	预测年限-疾病	机器学习模型 (传统模型)	预测 因子数	数据集人数	评价指标
Ioannou 2020 ^[33]	丙型肝炎相关 肝硬化	3 年-肝细胞癌	RNN(CS LR ,LGT LR)	31	训练集 :43 336 ,内部验证集 :4 815	AUROC
Johnson 2024 ^[34]	非肝癌的慢性 肝脏疾病	肝细胞癌	GALAD -RF ,RF -Practical (GALAD ,GALAD-Ogaki)	8	训练集 :3 473(5 折交叉验证)	AUROC
Li 2024 ^[35]	NAFLD	0.5 年-肝细胞癌	RETAIN ,DeepHit	-	训练集 :57 691 ,内部验证集(5 折交叉验证) :542	AUROC
Rodriguez 2024 ^[36]	MASLD	肝细胞癌	XGBoost	18	训练集 :1 267 884 ,内部验证集 :543 577	AUROC
Hashem 2018 ^[27]	丙型肝炎	肝纤维化	ADT ,GA ,PSO(MReg)	4	训练集 :22 690 ,内部验证集 :16 877	AUROC ,灵敏度 ,特异度 ,准确率 ,阳性预测值 ,阴性预测值
Hashem 2021 ^[26]	丙型肝炎	肝纤维化	Weka(FIB-4 ,APRI)	15	训练集 :69 106(10 折交叉验证)	AUROC ,灵敏度 ,特异度 ,阳性预测值 ,阴性预测值
Feng 2021 ^[24]	NAFLD	肝纤维化	RF (LASSO -LR ,FIB -4 ,NAS ,APRI)	5	训练集 :278 ,内部验证集 :275	AUROC ,灵敏度 ,特异度 ,阳性预测值 ,阴性预测值
Sarvestany 2022 ^[29]	非失代偿期 肝硬化	肝纤维化	ENS ,SVM ,RF ,GB ,ANN (LR ,FIB-4 ,APRI)	12	训练集 :1 703 ,内部验证集 :104	AUROC ,灵敏度 ,特异度 ,准确率 ,阳性预测值 ,阴性预测值
Cheung 2023 ^[23]	MAFLD	肝纤维化	MFS (FIB -4 ,NFS ,APRI ,VCTE ,Agile3+)	7	训练集 :276 ,内部验证集 :431	AUROC
Sripongpun 2023 ^[30]	NAFLD	5 年-肝纤维化	GAM ,RF ,GB(SAFE-score ,FIB-4 ,NFS)	20	训练集 (10 折交叉验证) :676 ,外部验证集 1 :280 ,外部验证集 2 :130 ,外部验证集 3 :7 647	AUROC ,灵敏度 ,特异度 ,阳性预测值 ,阴性预测值
Suárez 2023 ^[31]	NASH	肝纤维化	SVM ,DT ,GNB ,KNN ,XGBoost	22	训练集 :150 ,内部验证集 :65	AUROC ,准确率 ,精确度 ,召回率 ,F1 分数 ,约登指数
Cazzaniga 2009 ^[22]	丙型肝炎	肝硬化	ANN(LR ,APRI)	5	训练集 :144 ,内部验证集 :244 ,外部验证集 :220	AUROC ,灵敏度 ,特异度 ,阳性预测值 ,阴性预测值
Konerman 2019 ^[28]	丙型肝炎	5 年-肝硬化	CS Boosting ,LGT Boosting (CS Cox ,LGT Cox)	21	训练集 :7 934 ,内部验证集 :3 393	AUROC ,灵敏度 ,特异度 ,阳性预测值 ,阴性预测值

注 :ADT. 替代决策树 ;Agile3+. 新型无创肝纤维化评分 ;ANN. 人工神经网络 ;APRI. 天冬氨酸氨基转移酶/血小板比值指数 ;AUROC.受试者工作特征曲线下面积 ;Boosting.Boosting 算法 ;Cox.Cox 比例风险模型 ;CS.横断面数据 ;DeepHit.深度生存模型 ;DT.决策树 ;ENS. 机器学习集成模型 ;FIB-4. 纤维化-4 指数 ;GA. 遗传算法 ;GALAD.GALAD 评分 ;GALAD-Ogaki. 基于大阪市立医院数据的 GALAD 评分 ;GAM.广义加性模型 ;GB.梯度提升机 ;GNB.高斯朴素贝叶斯 ;KNN.K 近邻 ;LASSO.最小绝对收缩选择算法 ;LGT.纵向数据 ;LR.logistic 回归 ;MAFLD.代谢相关脂肪性肝病 ;MASLD.代谢功能障碍相关脂肪性肝病 ;MFS.脂肪肝纤维化随机森林评分 ;MReg.多重线性回归 ;NAFLD.非酒精性脂肪性肝病 ;NAS.NAFLD 活动度评分 ;NASH.非酒精性脂肪性肝病 ;NFS.非侵入性肝纤维化评分 ;PSO.粒子群优化算法 ;RETAIN.逆时注意力模型 ;RF.随机森林 ;RF-Practical.简单资料随机森林模型 ;RNN.循环神经网络 ;SAFE-score.脂肪肝纤维化评分 ;SVM.支持向量机 ;VCTE.振动控制瞬态弹性成像 ;Weka.怀卡托分析环境 ;XGBoost.极限梯度提升机。

子开发 3 种“3 年-肝细胞癌”预测模型 :横断面 LR 模型(cross sectional logistic regression model ,CS LR)、纵向 LR 模型(longitudinal logistic regression model ,LGT LR)和纵向 RNN。RNN 模型通过输入的 31 个预测因子 ,在模型内部形成 118 个隐藏预测因子 ,进一步预测 3 年内肝细胞癌发病风险 ,AUROC 为 0.759(95%CI 0.750~0.768) ,高于 CS LR 模型[AUROC=0.682(95%CI 0.675~0.689)]和 LGT LR 模型[AUROC=0.689(95%CI 0.680~0.698)]。同时 ,根据 RNN 模型评估 ,在全部丙型肝炎相关肝硬化患者中 ,风险评分最高的前 51%和前 66%人群 ,分别覆盖了随访 3 年后发展为肝细胞癌患者的 80%

和 90%。相较之下 ,CS LR 模型和 LGT LR 模型需要筛查更多的人群(80%肝细胞癌 :60%和 59%风险评分最高患者 ;90%肝细胞癌 :75%和 74%风险评分最高患者)。这一研究发现说明机器学习预测模型在优化慢性肝脏疾病患者风险分层管理 ,实现有限医疗资源合理规划方面具有重要价值。

2.3 以肝硬化与肝纤维化为结局的预测模型 共 12 项研究使用机器学习模型进行肝硬化与肝纤维化(包括肝硬度)的风险预测(表 1、表 2) ,9 项研究关注肝纤维化^[20 23-27 29-31] ,2 项研究关注肝硬化^[22 28] ,1 项研究关注肝硬度^[21] ,预测因子数目范围为 4~233(中位数 17.5)。



注:ADT. 替代决策树 ;Agile3+. 新型无创肝纤维化评分 ;ANN. 人工神经网络 ;APRI. 天冬氨酸氨基转移酶/血小板比值指数 ;AUROC. 受试者工作特征曲线下面积 ;Boosting. Boosting 算法 ;CHAID. 卡方自动交互检测模型 ;ENS. 机器学习集成模型 ;FIB-4. 纤维化-4 指数 ;Forns. Forns 指数 ;GALAD. GALAD 评分 ;LASSO. 最小绝对收缩选择算法 ;LGT. 纵向数据 ;LightGBM. 轻量级梯度提升机 ;LiverAID. 肝硬度机器学习集成模型 ;LR. logistic 回归 ;MFS. 脂肪肝纤维化随机森林评分 ;Mreg. 多重线性回归 ;NB. 朴素贝叶斯 ;RETAIN. 逆时注意力模型 ;RF. 随机森林 ;RF-Practical. 简单资料随机森林模型 ;RSF. 随机生存森林 ;RNN. 循环神经网络 ;SAFE-score. 脂肪肝纤维化评分 ;SVM. 支持向量机 ;Weka. 怀卡托分析环境 ;XGBoost. 极限梯度提升机。

图 1 慢性肝脏疾病预测模型表现森林图

预测肝纤维化疾病进展风险的模型 AUROC 在 0.740~0.980 之间,见图 1,其中,RF 模型对 NAFLD 人群发展为进展期肝纤维化的风险预测经训练集 10 折交叉验证后,AUROC 为 0.980(95%CI 0.969~0.991) 远高于传统 LR 模型的 0.790(95%CI 0.756~0.824)^[30]。然而,这种优势在外部验证集的检验下并未得到证实。在外部验证集中,传统 LR 模型外部验证 AUROC 为 0.800,高于其余 3 种机器学习预测模型,包括训练集中表现最优的 RF 模型^[30]。预测肝硬化发病风险的模型 AUROC 为

0.805~0.880,见图 1,其中的人工神经网络(artificial neural network, ANN)模型针对 244 位慢性丙型肝炎患者肝硬化风险预测的 AUROC 为 0.880(95%CI 0.790~0.980)^[22],高于 APRI 的 0.870(95%CI 0.780~0.960)。

对于一般人群的肝纤维化预测,Ben-Assuli 等^[20]开发的模型使用 2007—2017 年在一家以色列医院进行多次健康体检的“特拉维夫医学中心炎症调查”(Tel-Aviv Medical Center Inflammation Survey, TAMCIS)研究人群的多时点资料,发现时间序列机器学习模型能够

有效提高初始模型预测能力。研究基于 5 579 位健康体检者多时点 FIB-4 指标建立了隐马尔可夫模型(hidden Markov model, HMM) 和群体轨迹模型(group-based trajectory modelling, GBTM) 两种时间序列机器学习模型, 估计个体潜在的肝纤维化发病概率, 聚类分析获得代表肝纤维化结局的风险状态值。在以下 3 种预测因子方案中分别利用卡方自动交互检测模型(chi-square automatic interaction detector, CHAID) 和 XGBoost 两种算法进行模型开发: (1) 83 种预测因子+HMM 风险值; (2) 83 种预测因子+GBTM 风险值; (3) 仅 83 种预测因子。由于 GBTM 模型可以有效识别个体肝纤维化风险演变轨迹并聚类获得代表肝纤维化风险的状态值, 第 2 种预测因子方案的 XGBoost 模型 AUROC 达到 0.819(95%CI 0.805~0.833), 疾病风险预测能力最佳。

对于高危人群, Konecny 等^[28]利用 VHA 2000—2016 年 72 683 例丙型肝炎患者数据, 基于 3 个人口学预测因子和 18 个实验室指标预测因子, 其中纵向预测模型的实验室指标均通过最大值、最小值、最大斜率、最小斜率和绝对变化量 5 个方面描述。研究比较横断面 Cox 模型(cross sectional Cox model, CS Cox)、横断面 Boosting 模型(cross sectional boosting model, CS Boosting)、纵向 Cox 模型(longitudinal Cox model, LGT Cox) 和纵向 Boosting 模型(longitudinal boosting model, LGT Boosting) 4 种模型对 1 年、3 年和 5 年肝硬化发病风险的预测能力。结果显示, LGT Boosting 模型对 3 年肝硬化发病风险预测能力最强, AUROC 为 0.805(95%CI 0.804~0.807)。而对于非失代偿期肝硬化的肝病医院就诊患者, Sarvestany 等^[29]则使用 SVM、RF、GB、LR 和 ANN 5 种模型的集成机器学习模型(ensemble algorithm, ENS) 预测, 研究纳入 2000 年 1 月—2014 年 12 月在加拿大多伦多肝病所进行肝活组织检查的 1 703 例患者, 内部验证的 AUROC 为 0.870(95%CI 0.800~0.940), 优于常用的 FIB-4 评分(AUROC 0.825, 95%CI 0.744~0.906)。

2.4 以脂肪肝为结局的预测模型 4 项研究^[25, 37-39]针对脂肪肝使用机器学习模型预测疾病风险(表 1) 结局均为 NAFLD, 人群均为一般人群, 预测因子数目为 11~272 不等(中位数为 78)。这些研究利用现有队列, 训练了 RF、SVM 和 XGBoost 等常见机器学习模型, 机器学习模型的 AUROC 在 0.777~0.914 之间。其中, 对于北京健康管理队列的 22 140 名健康参与者, XGBoost 展现出最强的预测 NAFLD 风险能力, AUROC 为 0.914(95%CI 0.907~0.920)^[37], 高出传统模型中表现最佳的 NB 模型的 AUROC 值(差值=0.073)。

针对脂肪肝的研究主要在人群中展开。Zeng 等^[39]在 2008—2013 年从广州营养与健康研究中招募 1 546 名 40~80 岁参与者, 每 3 年随访一次, 纳入 272

个元数据进行 4 种模型[RF、SVM、LR 和 LightGBM] 训练。在 NAFLD 风险预测能力比较中, LightGBM (AUROC=0.791) 在机器学习模型中表现最佳, 同时也高于 LR 模型(AUROC=0.688)。与此相似, Cao 等^[37]使用北京健康管理队列的 22 140 名参与者的纵向数据, 引入最小绝对收缩选择算法 (least absolute shrinkage and selection operator, LASSO) 惩罚机制后进行 Cox 回归, 最终保留 16 个系数非零的预测因子, 比较基于这 16 个预测因子的 XGBoost 与另外 3 个预测模型(DT、LR、NB) 对 NAFLD 发病风险的预测能力, 结果显示, XGBoost 模型的预测能力最佳, AUROC 为 0.914(95%CI 0.907~0.920)。尽管机器学习模型在大多研究中优于传统模型, Huang 等^[38]在使用 6 196 位宁波健康体检人群数据预测 5 年 NAFLD 发病风险时, 基于 LASSO 回归筛选后的 11 个预测因子构建 SVM 模型, 出现了传统预测模型 LR 的 AUROC(0.778) 略高于机器学习模型 AUROC(0.777) 的现象。

3 讨论

本综述系统检索和梳理了慢性肝脏疾病的机器学习模型, 并对比了其与传统模型的预测表现差距, 总结以下两点优势。第一, 基于机器学习的预测模型能更好地预测慢性肝脏疾病发病风险, 优化高危人群分层管理, 实现个性化疾病监测, 促进有限医疗资源充分利用。在纳入的 14 项比较了机器学习模型和传统模型预测能力的研究中, 共有 12 项研究^[21-22, 24, 26-30, 33-34, 37, 39]在训练集的交叉验证或内部验证集上, 最优机器学习模型表现优于传统模型。具体而言, 在同时构建机器学习模型与传统模型并开展对比分析的各类研究中, 结论为“机器学习模型表现优于传统模型”的研究的两类模型 AUROC 差值范围为 0.01~0.19(中位数 0.063)。仅有 2 项研究^[23, 38]结果表明传统模型表现更好, AUROC 值分别高出机器学习模型 0.01 和 0.001。Ioannou 等^[33]基于机器学习的丙型肝炎相关肝硬化患者肝细胞癌发病预测模型以相较传统模型更少的筛检所需人数成功预测 80% 或 90% 肝细胞癌新发病例。机器学习模型强大的预测能力提供了优化慢性肝脏疾病患者分层管理、实现个性化监测肝细胞癌发病风险的新策略, 也为其他慢性肝脏疾病的分层管理提供了新的发展方向。第二, 机器学习能处理复杂的医疗数据, 使用影像学、组织病理学、电子病历信息等多模态数据, 丰富预测模型数据来源, 提高慢性肝脏疾病风险预测能力。对于医疗资源相当有限的国家或地区, 现实中难以对所有慢性肝脏疾病高危人群进行定期监测随访, 可用于机器学习模型开发的高质量数据不足。电子病历资料作为临床上最基本的要素, 记载了详细的患者信息, 提供了追踪患者健康信息的可行机制, 是推动基于数据做出更好决

策的强大资源。但电子病历内含大量结构性(如药品代码、疾病诊断代码等)和非结构性(如描述症状、体征等的整段文字说明)数据,应用过程中存在数据量大、维度高、数据模式多、缺失值多、质量差等问题^[40]。机器学习,尤其是深度学习方法,能直接自动从数据本身学习最优特征,发现潜在数据关系,更快地实现电子病历数据预处理和特征工程^[41]。因此,深度学习有能力解决电子病历数据带来的挑战,基于更全面、更易得的电子病历信息开发出预测能力更强的慢性肝脏疾病预测模型,而这一点对于传统模型而言难以实现。

尽管当前研究已在机器学习领域取得进展,但仍存在若干局限性。第一,多数研究集中在肝脏纤维化的预测上,而关于肝癌、肝硬化和NAFLD的研究相对较少,药物性肝损伤(drug-induced liver injury, DILI)等少见肝脏疾病的研究更是近乎于无。相关机器学习模型的研究证据数量不足,且多为疾病发病的预测,未考虑更具体的疾病分期。未来研究亟须更多、更优质的证据补充,以充分评价慢性肝脏疾病的机器学习预测模型推广应用的前景。第二,大部分研究仅使用病历报告的疾病诊断信息用于结局判断而非肝活组织检查进行确认,而且肝活组织检查的判断标准以及相关检验专家的水平在不同研究中也存在差异,因此,已有研究对相关结局的客观判断并不能确保结局的真实性。然而,肝活组织检查的有创性限制了其大规模应用于人群的可能性,尤其是无疾病症状人群,这种局限性的弥补有赖于更严谨的研究设计,如尽可能使用肝活组织检查或者影像学资料保证模型结果真实可靠。第三,已发表研究普遍存在汇报信息不全的问题,如未提供置信区间或标准差,这限制了对数据的深入分析和有效利用。正因汇报信息不完整,本研究仅选取了既往综述最常用的AUROC^[42-44]作为评估模型性能的指标进行汇报,尽管它在处理不平衡数据集时可能缺乏精确度^[45]。因此,推广并遵循“TRIPOD+AI声明”^[18]的相关要求,规范后续研究实施过程与报告内容显得尤为重要。第四,目前绝大多数研究缺乏有效的外部验证,这也是机器学习领域长期存在的方法学挑战和缺陷^[46-48]。有效的外部验证是检验预测模型实用性与推广可能性的重要组成部分,保证后续研究重视并实施外部验证是提高研究质量和可靠性的关键。

不同机器学习模型具有不同优势,需根据具体应用场景选择最适宜的模型。然而,当前研究仍存在不足,有待在未来研究中得到发展。第一,需要建立科学、严格、规范的研究流程,根据目标人群、目标疾病的特点开发和推广最适合的模型。比如,在原始医疗记录的挖掘过程中,运用RNN算法构建个体的纵向健康轨迹数据^[33],在动态数据的分析中,运用GBTM等时间序列机器学习模型,识别个体疾病发展过程,预测一定时

间内发病风险^[20]。第二,机器学习预测模型与真实环境的整合是未来的重要发展方向。(1)建立行业范围内的机器学习预测模型评价标准,充分验证模型,发挥模型在初级卫生保健的应有价值。(2)与计算机、软件开发专业人员合作,对开发模型进行可视化处理,从而增强实用性,以指导后续干预或治疗。在促进卫生保健工作高效、精准开展的同时,也能加强对全人群或高危人群的健康教育。

综上所述,机器学习预测模型,能够解决传统模型难以应对的电子病历信息等复杂数据难题,更准确地量化慢性肝脏疾病发病风险,实现高危人群的识别与分层管理,为卫生资源的合理配置和后续临床决策提供指导。同时,未来的研究需要遵循严格规范,提高研究和汇报的质量。研究者应在后续工作中开发更准确可靠、适用性更广的机器学习预测模型并进行有效的外部验证,以推进机器学习在慢性肝脏疾病防治实践中的应用。

参考文献

- [1] Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease [J]. *Lancet*, 2021, 397(10290): 2212-2224.
- [2] Ginès P, Krag A, Abraldes JG, et al. Liver cirrhosis [J]. *Lancet*, 2021, 398(10308): 1359-1376.
- [3] 缪伟刚, 周金意, 韩仁强. 全球肝癌流行数据解析 [J]. *中华流行病学杂志*, 2024, 45(6): 865-869.
- [4] Le MH, Le DM, Baez TC, et al. Global incidence of non-alcoholic fatty liver disease: a systematic review and meta-analysis of 63 studies and 1,201,807 persons [J]. *J Hepatol*, 2023, 79(2): 287-295.
- [5] GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021 [J]. *Lancet*, 2024, 403(10440): 2133-2161.
- [6] Wang FS, Fan JG, Zhang Z, et al. The global burden of liver disease: the major impact of China [J]. *Hepatology*, 2014, 60(6): 2099-2108.
- [7] 中华人民共和国国家卫生健康委员会医政司. 原发性肝癌诊疗指南(2024年版) [J]. *中华肝脏病杂志*, 2024, 32(7): 581-630.
- [8] 中华医学会肝病学分会. 肝硬化诊治指南 [J]. *中华肝脏病杂志*, 2019, 27(11): 846-865.
- [9] 中华医学会肝病学分会. 非酒精性脂肪性肝病防治指南(2018更新版) [J]. *中华肝脏病杂志*, 2018, 26(3): 195-203.
- [10] Chan YT, Zhang C, Wu J, et al. Biomarkers for diagnosis and therapeutic options in hepatocellular carcinoma [J]. *Mol Cancer*, 2024, 23(1): 189. DOI: 10.1186/s12943-024-02101-z.
- [11] Zhou J, Sun H, Wang Z, et al. Guidelines for the diagnosis and treatment of primary liver cancer (2022 edition) [J]. *Liver Cancer*, 2023, 12(5): 405-444.
- [12] European Association for the Study of the Liver. EASL Clinical Practice Guidelines on the management of hepatocellular carcinoma [J]. *J Hepatol*, 2025, 82(2): 315-374.
- [13] Kim MN, Han JW, An J, et al. KASL clinical practice guidelines for noninvasive tests to assess liver fibrosis in chronic liver disease [J]. *Clin Mol Hepatol*, 2024, 30(Suppl): S5-S105. DOI: 10.3350/cmh.2024.0506.
- [14] Deo RC. Machine learning in medicine [J]. *Circulation*, 2015, 132(20): 1920-1930.
- [15] Ahn JC, Connell A, Simonetto DA, et al. Application of artificial

- intelligence for the diagnosis and treatment of liver diseases [J]. *Hepatology* 2021, 73(6) :2546–2563.
- [16] Calderaro J, Seraphin TP, Luedde T et al. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma [J]. *J Hepatol* 2022, 76(6) :1348–1361.
- [17] Schattenberg JM, Chalasani N, Alkhoury N. Artificial intelligence applications in hepatology[J]. *Clin Gastroenterol Hepatol* 2023, 21(8) :2015–2025.
- [18] Collins GS, Moons KGM, Dhiman P et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods[J]. *BMJ* 2024, 385 :e078378. DOI :10.1136/bmj-2023-078378.
- [19] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. *Radiology* 1982, 143(1) :29–36.
- [20] Ben-Assuli O, Jacobi A, Goldman O et al. Stratifying individuals into non-alcoholic fatty liver disease risk levels using time series machine learning models[J]. *J Biomed Inform* 2022, 126 :103986. DOI :10.1016/j.jbi.2022.103986.
- [21] Blanes-Vidal V, Lindvig KP, Thiele M et al. Artificial intelligence outperforms standard blood-based scores in identifying liver fibrosis patients in primary care[J]. *Sci Rep* 2022, 12(1) :2914. DOI :10.1038/s41598-022-06998-8.
- [22] Cazzaniga M, Salerno F, Borroni G, et al. Prediction of asymptomatic cirrhosis in chronic hepatitis C patients: accuracy of artificial neural networks compared with logistic regression models [J]. *Eur J Gastroenterol Hepatol*, 2009, 21(6) :681–687.
- [23] Cheung JTK, Zhang X, Wong GL et al. MAFLD fibrosis score using routine measures to identify advanced fibrosis in metabolic-associated fatty liver disease[J]. *Aliment Pharmacol Ther* 2023, 58(11/12) :1194–1204.
- [24] Feng G, Zheng KI, Li YY et al. Machine learning algorithm outperforms fibrosis markers in predicting significant fibrosis in biopsy-confirmed NAFLD[J]. *J Hepatobiliary Pancreat Sci* 2021, 28(7) :593–603.
- [25] Goldman O, Ben-Assuli O, Rogowski O et al. Non-alcoholic fatty liver and liver fibrosis predictive analytics: risk prediction and machine learning techniques for improved preventive medicine[J]. *J Med Syst* 2021, 45(2) :22. DOI :10.1007/s10916-020-01693-5.
- [26] Hashem A, Awad A, Shousha H et al. Validation of a machine learning approach using FIB-4 and APRI scores assessed by the metavir scoring system: a cohort study[J]. *Arab J Gastroenterol* 2021, 22(2) :88–92.
- [27] Hashem S, Esmat G, Elakel W et al. Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients[J]. *IEEE/ACM Trans Comput Biol Bioinform* 2018, 15(3) :861–868.
- [28] Konerman MA, Beste LA, Van T et al. Machine learning models to predict disease progression among veterans with hepatitis C virus[J]. *PLoS One* 2019, 14(1) :e0208141. DOI :10.1371/journal.pone.0208141.
- [29] Sarvestany SS, Kwong JC, Azhie A et al. Development and validation of an ensemble machine learning framework for detection of all-cause advanced hepatic fibrosis: a retrospective cohort study[J]. *Lancet Digit Health* 2022, 4(3) :e188–e199. DOI :10.1016/S2589-7500(21)00270-3.
- [30] Sripongpun P, Kim WR, Mannalithara A et al. The steatosis-associated fibrosis estimator (SAFE) score: a tool to detect low-risk NAFLD in primary care[J]. *Hepatology* 2023, 77(1) :256–267.
- [31] Suárez M, Martínez R, Torres AM et al. A machine learning method to identify the risk factors for liver fibrosis progression in nonalcoholic steatohepatitis[J]. *Dig Dis Sci* 2023, 68(9) :3801–3809.
- [32] An C, Choi JW, Lee HS et al. Prediction of the risk of developing hepatocellular carcinoma in health screening examinees: a Korean cohort study[J]. *BMC Cancer* 2021, 21(1) :755. DOI :10.1186/s12885-021-08498-w.
- [33] Ioannou GN, Tang W, Beste LA et al. Assessment of a deep learning model to predict hepatocellular carcinoma in patients with hepatitis C cirrhosis[J]. *JAMA Netw Open* 2020, 3(9) :e2015626. DOI :10.1001/jamanetworkopen.2020.15626.
- [34] Johnson PJ, Bhatti E, Toyoda H et al. Serologic detection of hepatocellular carcinoma: application of machine learning and implications for diagnostic models[J]. *JCO Clin Cancer Inform* 2024, 8 :e2300199. DOI :10.1200/CCI.23.00199.
- [35] Li Z, Lan L, Zhou Y et al. Developing deep learning-based strategies to predict the risk of hepatocellular carcinoma among patients with nonalcoholic fatty liver disease from electronic health records [J]. *J Biomed Inform* 2024, 152 :104626. DOI :10.1016/j.jbi.2024.104626.
- [36] Rodriguez LA, Schmittiel JA, Liu L et al. Hepatocellular carcinoma in metabolic dysfunction-associated steatotic liver disease [J]. *JAMA Netw Open* 2024, 7(7) :e2421019. DOI :10.1001/jamanetworkopen.2024.21019.
- [37] Cao T, Zhu Q, Tong C et al. Establishment of a machine learning predictive model for non-alcoholic fatty liver disease: a longitudinal cohort study[J]. *Nutr Metab Cardiovasc Dis* 2024, 34(6) :1456–1466.
- [38] Huang G, Jin Q, Mao Y. Predicting the 5-year risk of nonalcoholic fatty liver disease using machine learning models: Prospective cohort study[J]. *J Med Internet Res* 2023, 25 :e46891. DOI :10.2196/46891.
- [39] Zeng F, Su X, Liang X et al. Gut microbiome features and metabolites in non-alcoholic fatty liver disease among community-dwelling middle-aged and older adults[J]. *BMC Med* 2024, 22(1) :104. DOI :10.1186/s12916-024-03317-y.
- [40] Si Y, Du J, Li Z et al. Deep representation learning of patient data from Electronic Health Records (EHR): a systematic review[J]. *J Biomed Inform* 2021, 115 :103671. DOI :10.1016/j.jbi.2020.103671.
- [41] Shickel B, Tighe PJ, Bihorac A et al. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis [J]. *IEEE J Biomed Health Inform* 2018, 22(5) :1589–1604.
- [42] Fleuren LM, Klausch TLT, Zwager CL et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy[J]. *Intensive Care Med* 2020, 46(3) :383–400.
- [43] Smith LA, Oakden-Rayner L, Bird A et al. Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis[J]. *Lancet Digit Health* 2023, 5(12) :e872–e881.
- [44] Nadarajah R, Younsi T, Romer E et al. Prediction models for heart failure in the community: a systematic review and meta-analysis[J]. *Eur J Heart Fail* 2023, 25(10) :1724–1738.
- [45] He H, Garcia EA. Learning from imbalanced data [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(9) :1263–1284. DOI :10.1109/TKDE.2008.239
- [46] Bellou V, Belbasis L, Konstantinidis AK et al. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal[J]. *BMJ* 2019, 367 :15358. DOI :10.1136/bmj.15358.
- [47] Siontis GCM, Tzoulaki I, Castaldi PJ et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination[J]. *J Clin Epidemiol* 2015, 68(1) :25–34.
- [48] Steyerberg EW, Moons KGM, van der Windt DA et al. Prognosis research strategy (PROGRESS) 3: prognostic model research[J]. *PLoS Med* 2013, 10(2) :e1001381. DOI :10.1371/journal.pmed.1001381.

收稿日期 2024-12-05 修回日期 2025-01-21 本文编辑 李翔