

· 数智医疗与信息化研究 ·

初级保健领域基于机器学习预测模型研究的设计特征与方法学质量：范围综述

钟锦佳, 李文涛, 黄亚芳*, 吴浩

100069 北京市, 首都医科大学全科医学与继续教育学院

*通信作者: 黄亚芳, 副教授; E-mail: yafang@ccmu.edu.cn

扫描二维码
查看原文

【摘要】 **背景** 近年来初级保健领域基于机器学习预测模型研究发展迅速, 但关于其设计特征与方法学质量的研究报道较少。**目的** 系统总结、分析初级保健领域基于机器学习预测模型研究的设计特征与方法学质量。**方法** 采用计算机检索 PubMed、Embase、中国知网、万方数据知识服务平台建库至 2023-02-21 发布的初级保健领域基于机器学习预测模型研究, 采用叙述性总结和描述方法分析纳入文献的基本特征、预测模型类型、样本量、缺失值处理方法、机器学习算法类型、模型性能评价指标及预测效能、模型验证方法等。**结果** 最终纳入 30 篇文献, 涉及 106 个预测模型, 其中发表时间为 2021~2023 年 17 篇; 研究主题涉及呼吸系统疾病 6 篇, 肿瘤 4 篇, 门诊预约 3 篇; 26 篇文献样本量 >1 000 (占 86.67%, 95%CI=68.36%~95.64%); 使用机器学习方法处理缺失值者 7 篇; 65 个预测模型使用基于树的机器学习算法, 其中随机森林使用频率最高 (占 32.08%, 95%CI=23.53%~41.95%); 61 个预测模型使用受试者工作特征 (ROC) 曲线下面积 (AUC) 或一致性 (C 统计量) 作为区分度评价指标 (占 57.55%, 95%CI=47.57%~66.97%), 但仅 14 个预测模型报告了校准度指标 (占 13.21%, 95%CI=7.67%~21.50%); 106 个预测模型多数区分度良好, 但 92 个预测模型偏倚风险评估结果为高风险 (占 86.79%, 95%CI=78.50%~92.33%); 仅 7 篇文献所涉预测模型进行了外部验证。**结论** 近 3 年来初级保健领域基于机器学习预测模型研究逐渐增多, 研究主题主要涉及呼吸系统疾病、肿瘤、门诊预约等; 预测模型在样本量、缺失值处理方法等方面存在较大差异, 多数预测模型区分度良好, 但大部分预测模型未进行外部验证, 总体偏倚风险较高。

【关键词】 初级保健; 机器学习; 研究设计; 预测模型; 方法学评价; 范围综述**【中图分类号】** R 161 R 319 **【文献标识码】** A DOI: 10.12114/j.issn.1007-9572.2023.0561**Design Features and Methodological Quality of Researches about Prediction Models Based on Machine Learning in Primary Care: a Scoping Review**

ZHONG Jinjia, LI Wentao, HUANG Yafang*, WU Hao

School of General Practice and Continuing Education, Capital Medical University, Beijing 100069, China

*Corresponding author: HUANG Yafang, Associate professor; E-mail: yafang@ccmu.edu.cn

【Abstract】 **Background** Researches about prediction models based on machine learning in primary care developed rapidly in recent years, but there are few researches about the design features and methodological quality. **Objective** To systematically summarize and analyze the design features and methodological quality of researches about prediction models based on machine learning in primary care. **Methods** Researches about prediction models based on machine learning in primary care was searched in PubMed, Embase, CNKI, Wanfang Data published from base-building to 2023-02-21, descriptive summary and description methods were used to analyze the basic characteristics of the included literature, types of prediction models, sample size, handling method of missing value, types of machine learning algorithms, model performance evaluation index and prediction efficiency, and model verification method. **Results** Totally 30 literature were enrolled, involving 106 prediction models, thereinto 17 literature were published between 2021 and 2023; research topics: respiratory disease in 6 literature, tumour in 4 literature, outpatient appointment in 3 literature; sample size over 1 000 in 26 literature (accounting for 86.67%,

基金项目: 国家自然科学基金资助项目 (82104133); 北京市教育委员会科学研究计划项目 (SM202110025003)**引用本文:** 钟锦佳, 李文涛, 黄亚芳, 等. 初级保健领域基于机器学习预测模型研究的设计特征与方法学质量: 范围综述 [J]. 中国全科医学, 2024, 27 (10): 1271-1276. DOI: 10.12114/j.issn.1007-9572.2023.0561. [www.chinagp.net]

ZHONG J J, LI W T, HUANG Y F, et al. Design features and methodological quality of researches about prediction models based on machine learning in primary care: a scoping review [J]. Chinese General Practice, 2024, 27 (10): 1271-1276.

© Chinese General Practice Publishing House Co., Ltd. This is an open access article under the CC BY-NC-ND 4.0 license.

95%CI=68.36%–95.64%) ; using machine learning methods to hand missing value in 7 literature; 65 prediction models used tree-based machine learning algorithm, in which random forest was the most frequently used (accounting for 32.08%, 95%CI=23.53%–41.95%) ; 61 prediction models used AUC of ROC or consistency (C statistic) as the differentiation evaluation index (accounting for 57.55%, 95%CI=47.57%–66.97%), but only 14 prediction models reported prediction models (accounting for 13.21%, 95%CI=7.67%–21.50%) ; the differentiation of most of the 106 prediction models was good, but bias risk assessment results of 92 prediction models were high-risk (accounting for 86.79%, 95%CI=78.50%–92.33%) ; only 7 literature involved prediction models conducted the external validation. **Conclusion** Researches about prediction models based on machine learning in primary care increase gradually in the past three years, in which the topics mainly involve respiratory disease, tumour, outpatient appointment and so on; there are significant difference in sample size and handling method of missing value in the 106 prediction models, most of the 106 prediction models are with good differentiation, but most of them did not conducted the external validation, and the overall risk of bias is relatively high.

【Key words】 Primary care; Machine learning; Research design; Prediction model; Methodological evaluation; Scoping review

近年来,随着以信息技术和人工智能技术为代表的新兴技术的迅速发展,初级保健服务和基本医疗在数字化、智能化、协同化等方面不断发展、完善^[1-3]。与专科医疗机构相比,初级保健服务机构常规登记数据、区域健康医疗数据、健康档案数据、患者自行报告的数据及移动设备采集的数据体量更为庞大,基于这些数据的临床研究报告及文献数量呈现逐渐升高趋势。在初级保健研究领域,学者开始尝试借助大数据和人工智能技术,通过机器学习构建预测模型,以辅助医疗决策^[4-6]。

广义上的机器学习是指使用计算机系统对非线性关联性和复杂交互作用的假设进行数学模型拟合过程^[7]。近年来,随着初级保健领域医疗数据的不断增多、数据采集系统的逐步规范及算法的优化,通过机器学习开发预测模型相关研究迅速发展,并已有经过验证的预测模型应用于基层医疗和管理实践中^[8-9]。基于机器学习预测模型研究的设计与方法学质量对医疗实践具有重要的指导意义^[10-11],但目前初级保健领域尚缺少基于机器学习预测模型研究相关设计与方法学质量评价研究。本研究旨在系统总结、分析初级保健领域基于机器学习预测模型研究的设计特征与方法学质量,以期人工智能技术在基层的科学、合理应用提供参考。

1 资料与方法

1.1 研究注册与报告

本研究已在国际前瞻性系统评价注册数据库(International Prospective Register of Systematic Reviews, PROSPERO)注册,注册号:CRD42022358128,遵循范围综述报告的PRISMA扩展版清单(PRISMA extension for scoping reviews, PRISMA-ScR)进行报告^[12]。

1.2 初级保健研究的界定

1996年,STARFIELD^[13]对初级保健研究做出定义:在基本保健的环境背景中开展的研究即为初级保健研究,本研究参照该定义对拟纳入研究进一步规定如下:

数据来源于初级保健/基层医疗卫生机构,或构建的预测模型将应用于初级保健/基本医疗领域。

1.3 文献纳入与排除标准

文献纳入标准:(1)应用机器学习方法进行数据处理或模型构建;(2)有开发、验证或更新预测模型的过程;(3)数据来源于初级保健/基本医疗机构,或构建的预测模型将应用于初级保健/基本医疗领域;(4)语种为中文或英文。文献排除标准:(1)综述、系统评价、会议摘要、研究方案;(2)数据来源于急诊室、ICU或普通住院病房;(3)未报告机器学习方法;(4)动物研究。

1.4 文献检索策略

采用主题词与自由词结合方式进行检索,通过检索逻辑词“OR”和“AND”进行自由组合,并针对不同数据库进行调整(不同数据库文献检索策略扫描本文首页二维码查看附录1);采用计算机检索PubMed、Embase、中国知网、万方数据知识服务平台建库至2023-02-21发布的相关文献,同时手工检索纳入文献的参考文献。英文检索词包括:primary health care, general practice, family practice, machine learning, deep learning, supervised machine learning, unsupervised machine learning, support vector machine, prediction model, risk model;中文检索词包括:基本医疗、初级保健、全科医生、全科医学、机器学习、随机森林、支持向量机、预测模型、风险评分、风险模型。

1.5 文献筛选、数据提取及偏倚风险评估

文献筛选:将检索到的文献导入EndNote 20软件并使用重复识别工具剔除重复文献。由2名研究人员严格按照文献纳入与排除标准先对文题、摘要进行初步筛选,再根据初步筛选结果进行全文阅读并最终决定是否纳入,如遇分歧则讨论解决或征询第3名研究人员意见。

数据提取:由2名研究人员共同参与设计文献资料提取表并独立提取相关数据,主要包括纳入文献的基本

特征、预测模型类型、样本量、缺失值处理方法、机器学习算法类型、模型性能评价指标及预测效能、模型验证方法等。

偏倚风险评估：由2名经过培训的研究人员根据预测模型偏倚风险评估工具（PROBAST）进行偏倚风险评估，如遇分歧则讨论解决或征询第3名研究人员意见。PROBAST共涵盖4个领域（分析、结果、预测因子、研究对象）20个问题，各领域均以“低风险”“高风险”“不清楚”进行评价，如至少1个领域被判定为高风险，则总体评价为高风险；如至少1个领域被判定为不清楚，而其他领域为低风险，则总体判定为不清楚；如全部领域被判定为低风险，但预测模型建立后却没有进行外部验证，则总体仍判定为高风险，而若该预测模型是基于大样本数据建立的并进行了某种内部验证则总体判定为低风险^[14]。

1.6 统计学方法

本研究采用叙述性总结和描述方法，无定量综合分析，结果以频率、百分比、中位数及四分位数间距（interquartile range, IQR）表示，并通过Wilson评分连续矫正区间计算95%CI。

2 结果

2.1 文献检索结果

计算机检索四大数据库共获得6070篇文献，手工检索纳入文献的参考文献共获得34篇文献，经EndNote 20软件及手工去重共剔除文献732篇，经阅读文题、摘要后排除明显不符合研究主题的文献4969篇，经阅

读全文并结合文献纳入与排除标准排除文献373篇，最终纳入30篇文献（纳入文献扫描本文首页二维码查看附录2）进行范围综述，涉及106个预测模型。本研究文献筛选流程见图1。

2.2 纳入文献的基本特征

纳入的30篇文献发表时间：2021—2023年17篇，2010—2020年13篇，无2010年以前发表文献；地区分布：美国8篇，英国7篇，瑞典、德国各3篇，荷兰2篇，巴基斯坦、比利时、芬兰、韩国、加拿大、罗马尼亚、以色列各1篇；研究主题：呼吸系统疾病6篇，肿瘤4篇，门诊预约3篇，其他17篇（精神疾病、糖尿病、脑卒中、心血管疾病各2篇，退伍军人10年全因死亡率、传染病、儿科疾病、感染性疾病、骨质疏松、声带疾病、脂肪肝、血液系统疾病、运动系统疾病各1篇）；预测模型类型（开发和/或验证）：开发和内部验证20篇，开发和内、外部验证5篇，仅开发3篇，仅外部验证2篇，详见表1。

2.3 范围综述结果

样本量：纳入的30篇文献样本量：<1000者4篇，1000~2999者6篇，3000~9999者4篇，10000~29999者5篇，30000~300000者5篇，>300000例6篇，共26篇文献样本量>1000（占86.67%，95%CI=68.36%~95.64%）。16篇样本量>10000的文献中有11篇数据来源于电子健康记录（electronic health record, EHR）或电子医疗记录（electronic medical record, EMR）。

缺失值处理方法：纳入的30篇文献中使用机器学习方法处理缺失值者7篇，其中使用梯度提升机（gradient

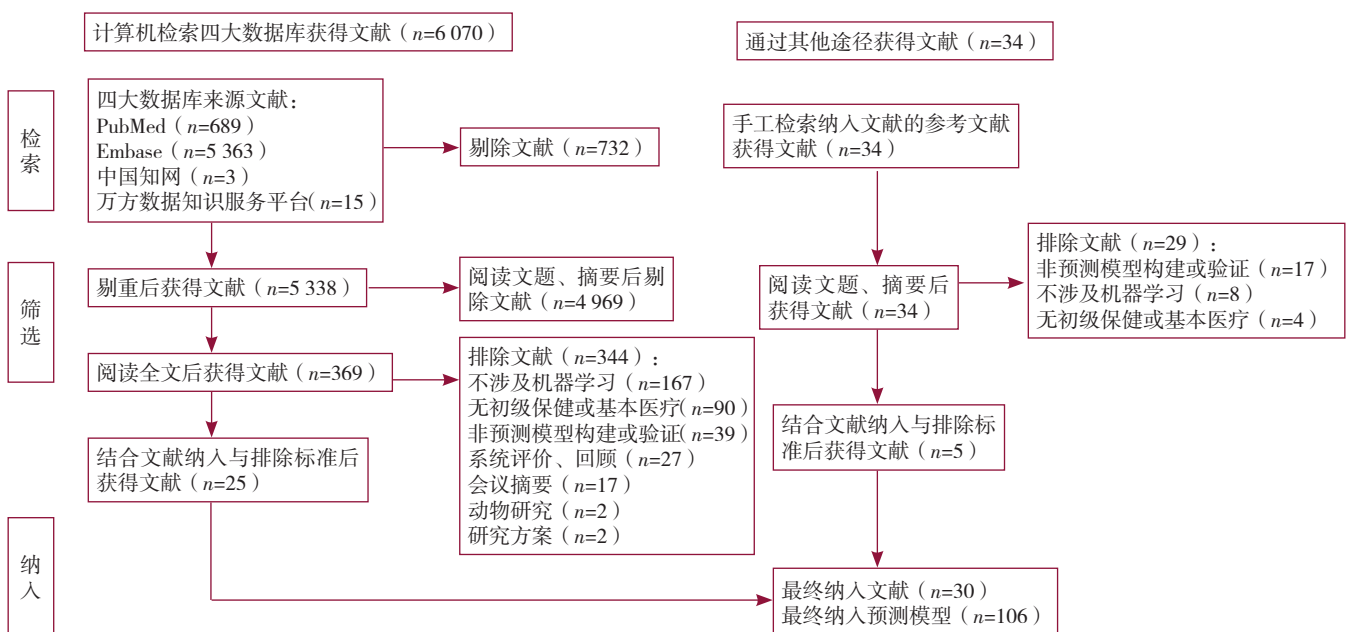


图1 文献筛选流程

Figure 1 Flowchart of literature screening

表1 纳入文献的基本特征 (n=30)

Table 1 Basic characteristics of the included literature

文献基本特征	篇数	百分比 (%)	95%CI (%)
发表时间			
2010—2020年	13	43.33	(25.97~62.34)
2021—2023年	17	56.67	(37.66~74.03)
地区分布			
美国	8	26.67	(12.98~46.18)
英国	7	23.33	(10.63~42.70)
瑞典	3	10.00	(2.62~27.68)
德国	3	10.00	(2.62~27.68)
其他	9	30.00	(15.41~49.56)
研究主题			
呼吸系统疾病	6	20.00	(8.40~39.13)
肿瘤	4	13.33	(4.36~31.64)
门诊预约	3	10.00	(2.62~27.68)
其他	17	56.67	(37.66~74.03)
预测模型类型			
开发和内部验证	20	66.67	(47.14~82.06)
开发和内、外部验证	5	16.67	(6.31~35.45)
仅开发	3	10.00	(2.62~27.68)
仅外部验证	2	6.67	(1.16~23.51)

表2 纳入文献所涉预测模型的机器学习算法类型 (n=106)

Table 2 Types of machine learning algorithms of the included prediction models

机器学习算法类型	预测模型数	百分比 (%)	95%CI (%)
基于树的机器学习算法	65	61.32	(51.33~70.48)
随机森林	34	32.08	(23.53~41.95)
梯度提升机	17	16.02	(9.89~24.72)
决策树	10	9.43	(4.86~17.06)
极端梯度提升树	4	3.77	(1.21~9.94)
回归模型	20	18.87	(12.17~27.88)
最大似然逻辑回归	14	13.21	(7.67~21.50)
Lasso 回归	2	1.89	(0.33~7.32)
最小二乘法回归	1	0.94	(0.05~5.90)
Cox 回归	1	0.94	(0.05~5.90)
弹性网络回归	1	0.94	(0.05~5.90)
岭回归	1	0.94	(0.05~5.90)
神经网络	9	8.49	(4.20~15.93)
支持向量机	6	5.66	(2.32~12.41)
贝叶斯网络	2	1.89	(0.33~7.32)
k-邻近算法	2	1.89	(0.33~7.32)
朴素贝叶斯	1	0.94	(0.05~5.90)
超级学习者集成	1	0.94	(0.05~5.90)

boosting machine, GBM)、极端梯度提升 (eXtreme gradient boosting, XGBoost) 者各 2 篇, 使用贝叶斯网络、k-近邻算法 (K-nearest neighbor, KNN) 者各 1 篇, 同时使用 XGBoost 和轻量梯度提升机 (light GMB) 者 1 篇; 使用多重插补方法处理缺失值者 3 篇; 2 篇文献中涉及连续变量使用均值表示, 分类变量使用“未报告或拒绝报告”替代; 2 篇文献未报告缺失值处理方法; 使用简单插补方法、不处理缺失值、不纳入不完整记录者各 1 篇; 其余 13 篇文献未报告缺失值。

机器学习算法类型: 纳入的 30 篇文献所涉 106 个预测模型中 65 个使用基于树的机器学习算法, 其中随机森林使用频率最高、GBM 次之; 20 个使用回归模型, 其中最大似然逻辑回归使用频率最高; 9 个使用神经网络, 6 个使用支持向量机, 使用贝叶斯网络、k-邻近算法者各 2 个, 使用朴素贝叶斯、超级学习者集成者各 1 个, 详见表 2。

模型评价指标: 在区分度方面, 纳入的 30 篇文献所涉 106 个预测模型中 61 个使用受试者工作特征 (ROC) 曲线下面积 (AUC) 或一致性 (C 统计量) 作为评价指标 (占 57.55%, 95%CI=47.57%~66.97%); 在校准度方面, 仅 14 个报告了校准度指标 (占 13.21%, 95%CI=7.67%~21.50%), 其中校准图 14 个 (占 13.21%, 95%CI=7.67%~21.50%), 校准斜率 6 个 (占 5.66%, 95%CI=2.32%~12.41%), 校准截距 6 个 (占 5.66%, 95%CI=2.32%~12.41%); 在综合评价方面,

47 个使用准确度 (占 44.33%, 95%CI=34.80%~54.29%), 36 个使用 F1 分数 (占 33.96%, 95%CI=25.22%~43.88%), 7 个使用 Brier 分数 (占 6.60%, 95%CI=2.92%~13.60%)。

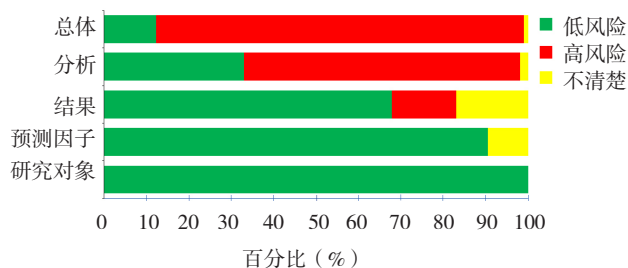
模型预测效能: 纳入的 30 篇文献所涉 106 个预测模型多数区分度良好, AUC 中位数为 0.79 (IQR 介于 0.74~0.83, 全距介于 0.63~0.92), 其中内部验证的 AUC 中位数为 0.79 (IQR 介于 0.73~0.82, 全距介于 0.63~0.92), 外部验证的 AUC 中位数为 0.83 (IQR 介于 0.78~0.84, 全距介于 0.71~0.92); 在校准度方面, 校准斜率的中位数为 1.03 (IQR 介于 1.00~1.06, 全距介于 0.97~1.18), 校准截距中位数为 0.02 (IQR 介于 0.00~0.03, 全距介于 -0.01~0.09)。

模型验证方法: 在内部验证方面, 纳入的 30 篇文献中 19 篇 (占 63.33%, 95%CI=43.90%~79.45%) 使用了交叉验证, 是最常用的模型验证方法; 6 篇使用了简单随机拆分验证 (占 20.00%, 95%CI=8.40%~39.13%); 3 篇仅进行模型开发而无内部验证 (占 10.00%, 95%CI=2.62%~27.68%)。在外部验证方面, 纳入的 30 篇文献中仅 7 篇 (占 23.33%, 95%CI=10.63%~42.70%) 使用独立数据库进行了外部验证。

2.4 预测模型的偏倚风险评估结果

纳入的 30 篇文献所涉 106 个预测模型中总体评估为低风险者 13 个 (占 12.26%, 95%CI=6.95%~20.40%), 高风险者 92 个 (占 86.79%, 95%CI=78.50%~

92.33%)，不清楚者1个(占0.94%，95%CI=0.05%~5.90%)，见图2(纳入文献所涉预测模型的偏倚风险评估结果扫描本文首页二维码查看附录3)。预测模型的高偏倚风险主要来自分析领域，常见于连续型预测因子和分类型预测因子是否被恰当地处理、出现缺失值的研究对象是否被恰当地处理、是否恰当地评估了相应的预测模型效能测量指标、预测模型过度拟合和预测模型效能上的乐观偏差是否被恰当地考量和处理^[14]。



注：PROBAST= 预测模型偏倚风险评估工具。

图2 基于PROBAST的预测模型偏倚风险评估结果

Figure 2 Risk of bias assessment results of prediction models based on PROBAST

3 讨论

本研究通过范围综述与文献计量方法系统分析了初级保健领域基于机器学习预测模型研究的设计特征与方法学质量，结果显示：(1)近3年来初级保健领域基于机器学习预测模型研究逐渐增多，研究主题主要涉及呼吸系统疾病、肿瘤、门诊预约等；(2)106个预测模型在样本量、缺失值处理方法等方面存在较大差异；(3)基于树的机器学习算法在算法类型使用最多，其中随机森林使用频率最高、GBM次之，二者约占全部机器学习算法类型的1/2；(4)多数预测模型区分度良好，对区分度的报告率较高，但对校准度的报告率较低；(4)大部分预测模型未进行外部验证，总体偏倚风险较高。

本研究结果显示，纳入的30篇文献中有26篇文献样本量>1 000，有16篇样本量>10 000，其中11篇数据来源于EHR或EMR，总体上样本量偏大，分析其原因如下：(1)初级保健服务机构常规登记数据、区域健康医疗数据、健康档案数据等体量庞大；(2)基于机器学习预测模型研究需要大量的训练集数据进行拟合等。分析本研究纳入的30篇文献所涉106个预测模型总体偏倚风险较高的原因为：与回归模型类似，基于机器学习预测模型研究在方法学、报告方面也存在缺失值处理方法、模型性能评价指标及预测效能、模型过度拟合报告不充分及未使用独立数据库进行外部验证等问题^[15-16]。回归模型常使用均值、众数进行插补以处理缺失值^[17-18]，基于机器学习预测模型研究缺失值处理

方法更为丰富，可以通过GBM、贝叶斯网络、KNN等机器学习算法直接处理缺失值^[19-21]。SAVAGE等^[22]研究指出，GBM包括一种内置输入缺失数据的方法，可从既有数据中推断缺失数据，并通过学习集成中的每个决策树节点与既有数据推断如何处理缺失数据。

本研究纳入的多数文献数据来源于EHR或EMR，也有一些文献数据来源于区域健康医疗数据库，虽然部分文献来源数据可能并不是为了研究目的而采集或测量的，但本研究设置了较严格的文献纳入与排除标准，因此，其在“研究对象”领域的偏倚风险评估结果为低风险。需要指出的是，在初级保健和社区卫生服务领域，大量真实场景下基于常规医疗服务而获得的数据对于基本医疗和初级保健研究而言既是机遇也是挑战：学者虽然可以较方便地获得大样本量数据，但可能会遭遇数据采集不规范、数据缺失、患者失访、各个数据收集系统协调性能与融合不足等问题^[23-24]。机器学习方法利用数据和经验使计算机系统自动学习和改进，可基于庞大的训练集生成越来越可靠的预测结果，且预测结果的可靠性仍主要取决于训练集的数据质量和样本量^[25]。随着大数据技术、EMR的迅速发展，医疗数据的记录越来越规范，机器学习方法在处理EMR及提取数据方面潜力巨大，未来初级保健领域可能会出现更多的基于机器学习预测模型研究。因此，今后基于社区卫生服务机构收集医疗数据并开展研究时可进一步完善采集数据的真实性与规范性，并规范数据纳入与排除标准、数据清洗等，而在模型开发及验证过程中需注重样本量、缺失值处理方法、模型性能评价指标及预测效能、模型验证方法等，以进一步降低初级保健领域基于机器学习预测模型研究报告的偏倚风险，促进今后以机器学习为代表的人工智能技术在初级保健的高质量应用。

本研究局限性：(1)提取的数据限于纳入文献的报告，若纳入文献未报告研究所需信息则可能会在一定程度上产生信息偏倚；(2)PROBAST是基于回归模型研究而开发的，而基于机器学习预测模型研究尚缺乏专门的报告条目指引，不利于进行数据提取及方法学质量评价^[26]；(3)纳入文献语种仅限中文与英文，纳入文献数量有限，并可能会在一定程度上产生选择偏倚。

综上所述，本研究通过范围综述与文献计量方法系统分析了初级保健领域基于机器学习预测模型研究的设计特征与方法学质量，并通过对初级保健研究进行界定、明确研究对象(是基于机器学习预测模型，而非针对人工智能这一较宽泛的概念)等分析了目前初级保健领域基于机器学习预测模型研究的设计特征与方法学质量，结果发现近3年来初级保健领域基于机器学习预测模型研究逐渐增多，研究主题主要涉及呼吸系统疾病、肿瘤、门诊预约等；预测模型在样本量、缺失值处理方法等方

面存在较大差异,多数预测模型区分度良好,但大部分预测模型未进行外部验证,总体偏倚风险较高。

作者贡献:钟锦佳负责研究的实施与可行性分析、数据收集与整理、结果分析与解释、撰写论文;李文涛负责研究的构思与设计、数据收集与整理;黄亚芳负责研究的构思与设计、修订论文、文章质量控制及审校,对文章整体负责;吴浩负责研究的构思与设计、研究实施的监督与管理。

本文无利益冲突。

参考文献

- [1] LIAW W, KAKADIARIS I A. Artificial intelligence and family medicine: better together [J]. *Fam Med*, 2020, 52 (1): 8–10.
- [2] KONG X, AI B, KONG Y, et al. Artificial intelligence: a key to relieve China's insufficient and unequally-distributed medical resources [J]. *Am J Transl Res*, 2019, 11 (5): 2632–2640.
- [3] AGRAWAL R, PRABAKARAN S. Big data in digital healthcare: lessons learnt and recommendations for general practice [J]. *Heredity (Edinb)*, 2020, 124 (4): 525–534.
- [4] SØRENSEN N, BEMMAN B, JENSEN M B, et al. Machine learning in general practice: scoping review of administrative task support and automation [J]. *BMC Prim Care*, 2023, 24 (1): 14.
- [5] ABBASGHOLIZADEH RAHIMI S, LÉGARÉ F, SHARMA G, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal [J]. *J Med Internet Res*, 2021, 23 (9): e29839. DOI: 10.2196/29839.
- [6] KUEPER J K, TERRY A L, ZWARENSTEIN M, et al. Artificial intelligence and primary care research: a scoping review [J]. *Ann Fam Med*, 2020, 18 (3): 250–258. DOI: 10.1370/afm.2518.
- [7] GREENER J G, KANDATHIL S M, MOFFAT L, et al. A guide to machine learning for biologists [J]. *Nat Rev Mol Cell Biol*, 2022, 23 (1): 40–55. DOI: 10.1038/s41580-021-00407-0.
- [8] SHICKEL B, TIGHE P J, 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. DOI: 10.1109/JBHI.2017.2767063.
- [9] HOBENSACK M, SONG J, SCHARP D, et al. Machine learning applied to electronic health record data in home healthcare: a scoping review [J]. *Int J Med Inform*, 2023, 170: 104978.
- [10] BOUWMEESTER W, ZUITHOFF N P A, MALLETT S, et al. Reporting and methods in clinical prediction research: a systematic review [J]. *PLoS Med*, 2012, 9 (5): 1–12.
- [11] CHOWDHURY M Z I, TURIN T C. Variable selection strategies and its importance in clinical prediction modelling [J]. *Fam Med Community Health*, 2020, 8 (1): e000262.
- [12] TRICCO A, LILLIE E, ZARIN W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation [J]. *Ann Intern Med*, 2018, 169 (7): 467–473.
- [13] STARFIELD B. A framework for primary care research [J]. *J Fam Pract*, 1996, 42 (2): 181–185.
- [14] MOONS K G M, WOLFF R F, RILEY R D, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration [J]. *Ann Intern Med*, 2019, 170 (1): W1–33. DOI: 10.7326/M18–1377.
- [15] CHRISTODOULOU E, MA J, COLLINS G S, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models [J]. *J Clin Epidemiol*, 2019, 110: 12–22. DOI: 10.1016/j.jclinepi.2019.02.004.
- [16] KABORÉ R, HALLER M C, HARAMBAT J, et al. Risk prediction models for graft failure in kidney transplantation: a systematic review [J]. *Nephrol Dial Transplant*, 2017, 32 (suppl_2): ii68–76. DOI: 10.1093/ndt/gfw405.
- [17] MASCONI K L, MATSHA T E, ECHOUFFO-TCHEUGUI J B, et al. Reporting and handling of missing data in predictive research for prevalent undiagnosed type 2 diabetes mellitus: a systematic review [J]. *EPMA J*, 2015, 6 (1): 7.
- [18] MASCONI K L, MATSHA T E, ERASMUS R T, et al. Effects of different missing data imputation techniques on the performance of undiagnosed diabetes risk prediction models in a mixed-ancestry population of South Africa [J]. *PLoS One*, 2015, 10 (9): e0139210. DOI: 10.1371/journal.pone.0139210.
- [19] TIERNEY N J, HARDEN F A, HARDEN M J, et al. Using decision trees to understand structure in missing data [J]. *BMJ Open*, 2015, 5 (6): e007450.
- [20] HOWEY R, CLARK A D, NAAMANE N, et al. A Bayesian network approach incorporating imputation of missing data enables exploratory analysis of complex causal biological relationships [J]. *PLoS Genet*, 2021, 17 (9): e1009811.
- [21] HAUN M W, SIMON L, SKLENAROVA H, et al. Predicting anxiety in cancer survivors presenting to primary care—a machine learning approach accounting for physical comorbidity [J]. *Cancer Med*, 2021, 10 (14): 5001–5016. DOI: 10.1002/cam4.4048.
- [22] SAVAGE R, MESSENGER M, NEAL R D, et al. Development and validation of multivariable machine learning algorithms to predict risk of cancer in symptomatic patients referred urgently from primary care: a diagnostic accuracy study [J]. *BMJ Open*, 2022, 12 (4): e053590. DOI: 10.1136/bmjopen-2021-053590.
- [23] GOLDSTEIN B A, NAVAR A M, PENCINA M J, et al. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review [J]. *J Am Med Inform Assoc*, 2017, 24 (1): 198–208.
- [24] KHARRAZI H, CHI W, CHANG H Y, et al. Comparing population-based risk-stratification model performance using demographic, diagnosis and medication data extracted from outpatient electronic health records versus administrative claims [J]. *Med Care*, 2017, 55 (8): 789–796.
- [25] AZMI J, ARIF M, NAFIS M T, et al. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data [J]. *Med Eng Phys*, 2022, 105: 103825.
- [26] COLLINS G S, DHIMAN P, ANDAUR NAVARRO C L, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence [J]. *BMJ Open*, 2021, 11 (7): e048008.

(收稿日期: 2023-05-11; 修回日期: 2023-12-27)

(本文编辑: 鹿飞飞)