

基于机器学习算法的初治结核病患者 复发预测模型研究

买日哈巴·卡米力¹, 买吾拉江·依马木², 王艳杰¹, 王雨薇¹, 阿丽米热·阿不力米提¹, 麦迪努尔·卡米力³, 向阳¹

1. 新疆医科大学公共卫生学院, 新疆 乌鲁木齐 830017; 2. 喀什地区疾病预防控制中心; 3. 新疆医科大学临床医学部

摘要: **目的** 基于新疆喀什地区结核病常规监测数据, 系统比较 7 种机器学习算法构建的初治结核病患者复发预测模型的效能, 为高负担地区优化复发干预策略提供数据支持。**方法** 以喀什地区 2016—2022 年成功治疗的初治结核病患者为研究对象, 分析其截至 2023 年末的复发情况。通过多因素 logistic 回归筛选独立预测因子, 并基于 logistic 回归、决策树、随机森林、多层感知器、极限梯度提升树、轻量级梯度提升机算法和弹性网络构建并验证复发预测模型, 对最优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 69 476 例研究对象截至 2023 年末复发 9 444 例 (13.59%)。通过 logistic 回归确定出 14 个独立预测因子, 7 种预测模型训练集 AUC 值范围为 0.705 ~ 0.762, 其中决策树模型表现最优 (AUC=0.762, 95%CI: 0.758 ~ 0.766), 且校准度良好。基于 SHAP 值的变量重要性排序结果显示, 初诊痰培养结果、现住地结核病负担及就诊方式是前三位重要预测因子。**结论** 基于常规监测数据构建的决策树模型对初治结核复发具有较高预测效能, 其可解释性特征有助于临床优先识别高危个体。

关键词: 结核病; 复发; 机器学习; 治疗成功

中图分类号: R521 文献标志码: A 文章编号: 1003-8507(2025)13-2310-08

DOI: 10.20043/j.cnki.MPM.202503139

Prediction model for tuberculosis recurrence in newly treated patients based on machine learning algorithms

KAMILI Mai-ri-ha-ba*, YIMAMU Mai-wu-la-jiang, WANG Yan-jie, WANG Yu-wei, ABULIMITI A-li-mi-re,

KAMILI Mai-di-nu-er, XIANG Yang

*School of Public Health, Xinjiang Medical University, Urumqi, Xinjiang 830017, China

Abstract: **Objective** To systematically compare the performance of seven machine learning algorithms in constructing prediction models for tuberculosis (TB) recurrence among newly treated patients in Kashgar, Xinjiang, providing data support for optimizing recurrence intervention strategies in high-burden areas. **Methods** We analyzed 69 476 successfully treated new TB patients from 2016 to 2022 in Kashgar, with follow-up through 2023. Independent predictors were selected through multivariate logistic regression. Seven models (logistic regression, decision tree, random forest, multilayer perceptron, XGBoost, LightGBM, and elastic net) were developed and validated. The optimal model was interpreted using SHapley Additive exPlanations (SHAP). **Results** Among 69 476 cases, 9 444 (13.59%) experienced recurrence by 2023. Fourteen independent predictors were identified. The seven models showed AUC values ranging from 0.705 to 0.762 in the training set, with the decision tree model performing best (AUC=0.762, 95%CI: 0.758-0.766) and demonstrating good calibration. SHAP analysis revealed sputum culture results at diagnosis, local TB burden, and treatment modality as the top three predictive factors. **Conclusion** The decision tree model based on routine surveillance data shows high predictive performance for TB recurrence, with interpretable features that can facilitate early identification of high-risk individuals in clinical practice.

Keywords: Tuberculosis; Recurrence; Machine learning; Treatment success

结核病复发是阻碍全球终止结核病进程的重要挑战, 高负担地区尤为显著。多项研究报告显示, 初治结核病患者中 2% ~ 10% 在治疗成功后约五年内出

基金项目: 国家自然科学基金(81860589); 自治区自然科学基金(2022 D01C203); 新疆维吾尔自治区“十四五”高等学校特色学科—公共卫生与预防医学

作者简介: 买日哈巴·卡米力 (1998—), 女, 硕士在读, 研究方向: 传染病流行病学

通信作者: 向阳, E-mail: 893664450@qq.com

现复发, 在新疆喀什地区这类高负担地区该比例可高达约 15%^[1]。复发患者耐药率增高、病情变复杂, 易造成疾病的迁延难愈和进一步扩散^[2]。精准预测复发、优化干预策略刻不容缓。在大数据与人工智能技术快速发展的当下, 机器学习算法为结核病复发预测提供了有效路径。本研究基于喀什地区近七年结核病监测数据, 系统比较七种常用机器学习算法预测复发效能, 通过可解释性分析揭示关键风险因子, 旨在构建

适配资源受限场景的复发预测工具,为优化高危人群靶向干预提供依据。

1 资料与方法

1.1 资料来源 本研究数据从《结核病管理信息系统》系统中导出,是经匿名化处理的常规监测数据,不涉及伦理审查和知情同意。

1.2 研究对象 喀什地区 2016—2022 年治疗成功的初治结核病患者(排除信息错误或缺失、现驻地非喀什地区的患者)。

1.3 相关定义 “复发”是指曾接受过抗结核治疗,且在疗程结束后被判定为“成功治疗”(包括“治愈”或“完成疗程”),但后期又被重新登记为结核病^[3]。本研究为了排除治疗不成功患者被误判的可能性,将“成功治疗”后六个月内就出现再次复发的患者视为治疗不成功患者^[4]。“诊疗延误时长”是指从患者出现结核病相关症状或体征开始,到最终获得适当治疗之间所经历的时间间隔^[5]。现驻地结核病负担以 2016—2022 年喀什地区各县市结核病年均发病率划分:<200/10 万为“低”;200/10 万~300/10 万为“中”;300/10 万~400/10 万为“较高”;>400/10 万为“严重”。治疗方案中“2HRZE”表示强化期药物组合,包括异烟肼、利福平、吡嗪酰胺和乙胺丁醇,每日一次,连续使用 2 个月,“4HR”表示巩固期药物组合,包括异烟肼和利福平,每日一次,连续使用 4 个月,“7-10HRE”表示巩固期药物组合,包括异烟肼、利福平和乙胺丁醇,每日一次,连续使用 7~10 个月。

1.4 预测模型

1.4.1 预测模型的构建 将研究对象按 7:3 随机分为训练集和验证集。在训练集中,以“是否复发”为因变量,纳入病案信息中性别、年龄等 14 个因素为自变量,采用单因素、多因素 logistic 回归筛选复发的独立影响因素,以此作为预测因子,构建基于 logistic 回归、决策树(decision tree)、随机森林(random forest)、多层感知器(multilayer perceptron, MLP)、极限梯度提升树(eXtreme gradient boosting, XGboost)、轻量级梯度提升机算法(light gradient boosting machine, light GBM)、弹性网络(elastic net, Enet)7 种机器学习算法的复发预测模型。随机过采样实例法(random over-sampling examples, ROSE)用于使训练集中复发患者与未复发患者比例接近 1:1 平衡,以提高模型预测效能,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超参数。

1.4.2 预测模型的验证与评价 以训练集上预测效果最佳的超参数作为模型的初始化超参数对验证集进行验证。受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under the curve,

AUC)、校准曲线用于评价模型的区分度、准确性。

1.4.3 SHAP 模型解释 对区分度最优模型采用 SHAP 分析法进行模型解释,绘制 SHAP 特征重要性排序图、SHAP 蜂群图及 SHAP 值特征依赖图,深入理解模型的决策过程和各特征对预测结果的具体贡献。

1.5 统计学方法 分别采用 Excel 2016 和 R 4.4.1 进行数据整理与分析。通过计算自变量间的容差和方差膨胀因子判断变量共线性。定性资料采用频数和率进行统计描述,采用 χ^2 检验分析组间差异性。使用 sample 函数对数据集进行随机划分。Rose 包中的 ROSE 函数用于数据不平衡处理。采用 logistic 回归对各变量进行单因素分析,对其 P 值小于 0.05 的变量纳入多因素 logistic 回归分析中,以 P 值小于 0.05 的变量作为预测因子纳入到模型构建过程。Tidymodels 包用于模型构建和模型评估,fastshap 包用于 SHAP 模型解释,iml 包用于 H 统计量交互作用强度分析。双侧检验,检验水准 $\alpha=0.05$ 。

2 结果

2.1 复发概况 2016—2022 年治疗成功的 69 476 例初治结核病患者中,截至 2023 年末复发 9 444 例(13.59%)。男性复发率为 14.09%,高于女性的 13.09%;农民复发率为 14.20%,高于非农民的 10.74%; ≥ 60 岁者复发率为 18.61%,高于其他年龄段者的 11.93%。

2.2 预测因子筛选 以研究对象性别、年龄段、职业、就诊方式、是否单纯结核性胸膜炎、治疗结局、初诊痰检结果、初诊痰培养结果、初治两个月末痰检结果、诊疗延误时长、治疗方案、固定剂量复合制剂(fixed-dose combination, FDC)使用情况、管理单位类型、现驻地结核病负担等 14 个因素作为研究变量,各变量间容差均小于 0.2、方差膨胀因子均大于 5,不存在显著共线性。将研究对象按 7:3 随机划分为训练集(48 632 例)和验证集(20 844 例),两组患者各研究变量间差异均无统计学意义(P 值均 >0.05),见表 1。变量赋值情况见表 2。在训练集,对单因素 logistic 回归分析有统计学意义的变量进一步纳入多因素 logistic 回归分析,多因素分析结果显示,14 个研究变量均为复发的独立影响因素。见表 3。

表 1 训练集和验证集比较

特征	训练集	验证集	χ^2 值	P 值
性别			0.317	0.574
女性	24 126	10 292		
男性	24 506	10 552		
年龄段(岁)			2.873	0.412
0 ~ <30	4 515	1 914		

(续表)

特征	训练集	验证集	χ^2 值	P 值
30 ~ <45	5 049	2 141		
45 ~ <60	10 798	4 532		
≥60	28 270	12 257		
职业			3.708	0.295
学生和教师	1 263	530		
居家人员	4 360	1 781		
农民	42 018	18 118		
其他	991	415		
就诊方式			0.537	0.464
被动方式	25 767	11 107		
主动方式	22 865	9 737		
是否单纯结核性胸膜炎			0.003	0.958
否	47 995	20 572		
是	637	272		
治疗结局			0.046	0.830
完成疗程	37 630	16 113		
治愈	11 002	4 731		
是否使用 FDC			0.039	0.844
否	5 889	2 513		
是	42 743	18 331		
初诊痰检结果			0.167	0.683
阴性	43 345	18 556		
阳性	5 287	2 288		
初治两个月末痰检结果			0.054	0.817
阴性	48 246	20 675		
阳性	386	169		
初诊痰培养结果			0.410	0.815
阴性	12 778	5 516		
阳性	6 541	2 818		
无结果	29 313	12 510		
诊疗延误时长(d)			0.007	0.934
<30	23 737	10 181		
≥30	24 895	10 663		
治疗方案			4.242	0.120
2HRZE/4HR	47 034	20 149		
2HRZE/7-10HRE	1 156	531		
其他	442	164		
管理单位类型			0.523	0.469
医院	20 959	9 045		
疾病预防控制中心	27 673	11 799		
现驻地结核病负担			0.770	0.857
低	4 701	1 999		
中	8 346	3 533		
高	17 121	7 340		
严重	18 464	7 972		

2.3 模型构建与验证 在训练集中, 基于多因素 logistic 回归分析筛选出的 14 个预测因子, 构建 logistic、决策树、随机森林、MLP、XGBoost、Light GBM 和 ENet 等 7 种复发预测模型, 并在验证集中进行内部验证。结果显示, 7 种模型在训练集中的 AUC 值范围为 0.705 ~ 0.762, 在验证集中的 AUC 值范围为

表 2 研究变量赋值表

变量	赋值
复发状态	未复发 =0, 复发 =1
性别	女 =0, 男 =1
年龄(岁)	0 ~ <30 =1, 30 ~ <45 =2, 45 ~ <60 =3, ≥60 =4
职业	学生和教师 =1, 居家人员 =2, 农民 =3, 其他 =4
就诊方式	主动方式 =0, 被动方式 =1
是否单纯结核性胸膜炎	否 =0, 是 =1
治疗结局	治愈 =0, 完成疗程 =1
是否使用 FDC	否 =0, 是 =1
初诊痰检结果	阴性 =0, 阳性 =1
初治两个月末痰检结果	阴性 =0, 阳性 =1
初诊痰培养结果	阴性 =0, 阳性 =1, 无结果 =2
诊疗延误时长(d)	<30 =0, ≥30 =1
治疗方案	2HRZE/4HR =1, 2HRZE/7-10HRE =2, 其他 =3
管理单位类型	医院 =0, 疾病预防控制中心 =1
现驻地结核病负担	低 =1, 中 =2, 高 =3, 严重 =4

0.691 ~ 0.712, 模型均具有较好的区分度, 其中基于决策树的复发预测模型区分度最佳 (AUC=0.762, 95% CI: 0.758 ~ 0.766), 见图 1。校准曲线图表明, logistic、决策树、ENet、Light GBM 和 XGBoost 模型的校准曲线在训练集和验证集中均接近理想对角线, 显示出良好的准确性和可靠性, 见图 2。

2.4 SHAP 模型解释 对 AUC 值最高的决策树模型进行 SHAP 解释。SHAP 特征重要性排序结果显示, 初诊痰培养结果、现驻地结核病负担和就诊方式是初治结核病复发的前三个重要预测因子, 对模型输出影响显著, 见图 3A。这些变量与其他变量间交互作用均较弱。SHAP 蜂群图和特征依赖图直观展示了各特征值与模型预测的关系, 揭示了特征对模型输出的具体贡献, 见图 3B、C。在 14 个预测因子中, 初诊痰培养结果未知、现驻地结核病负担严重、年龄 ≥60 岁、管理单位为疾控中心、诊疗延误时长 ≥30 d、初诊痰检阳性、男性、初治两个月末痰检阳性等因素对模型有正向影响, 增加复发风险; 而主动就诊、治疗结局为治愈、治疗方案为 2HRZE/7-10HRE 或其他方式等因素对模型有负向影响, 是复发的保护因素。

3 讨论

结核病复发问题是我国结核病防控工作中亟待解决的重要难题, 实现复发精准预测并优化干预策略具有重要的现实意义。随着大数据和人工智能技术的迅猛发展, 各类机器学习算法已被广泛应用于结核病各个领域, 包括结核分枝杆菌潜伏感染与活动性结核病的鉴别诊断^[5]、耐药结核病的诊断与预测^[6]、抗结

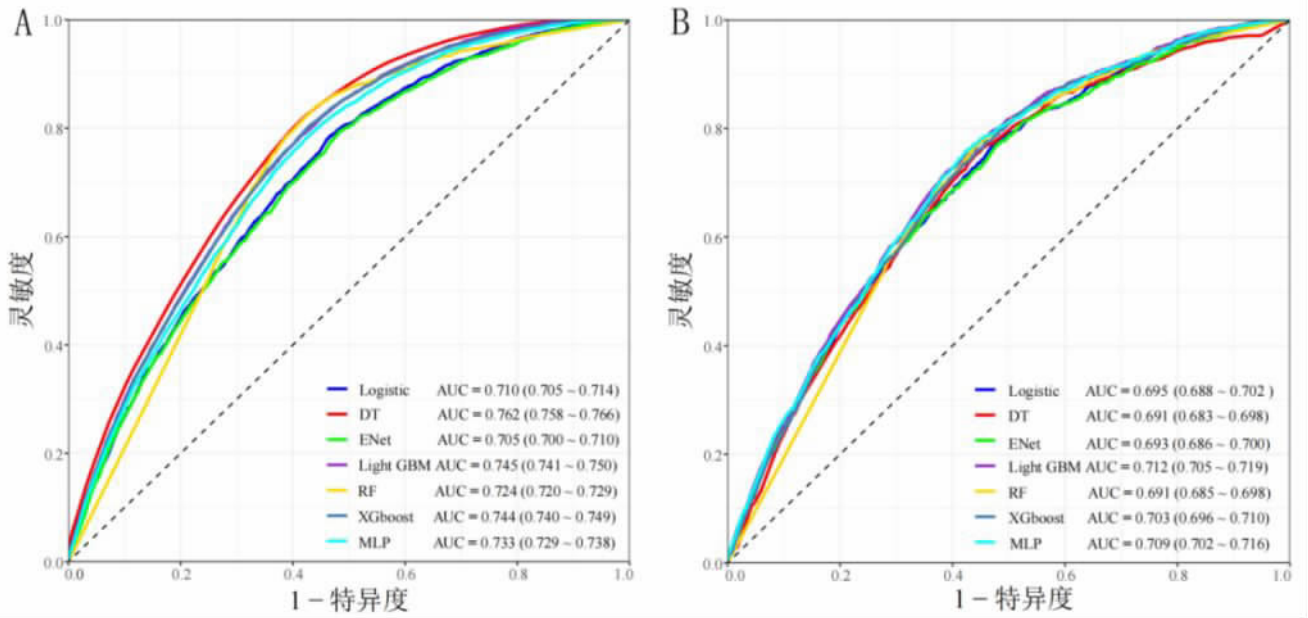
核治疗不良反应预测^[7]、结核分枝杆菌和非结核分枝杆菌感染的分类^[8]以及结核病治疗转归结果的预测等方面^[9]。这些研究通过整合机器学习技术与医学影像、生物标志物及临床数据,显著提升了预测准确性和效率。然而,几乎很少有研究将机器学习算法运用到结

核病复发领域。本研究利用喀什地区常规监测数据,构建了基于 logistic、决策树、随机森林、MLP、XGboost、Light GBM、ENet 等 7 种机器学习算法的初治结核病患者复发预测模型。

表 3 复发影响因素的单因素、多因素 logistic 回归分析

Table 3 Univariate and multivariate logistic regression analysis of factors influencing recurrence

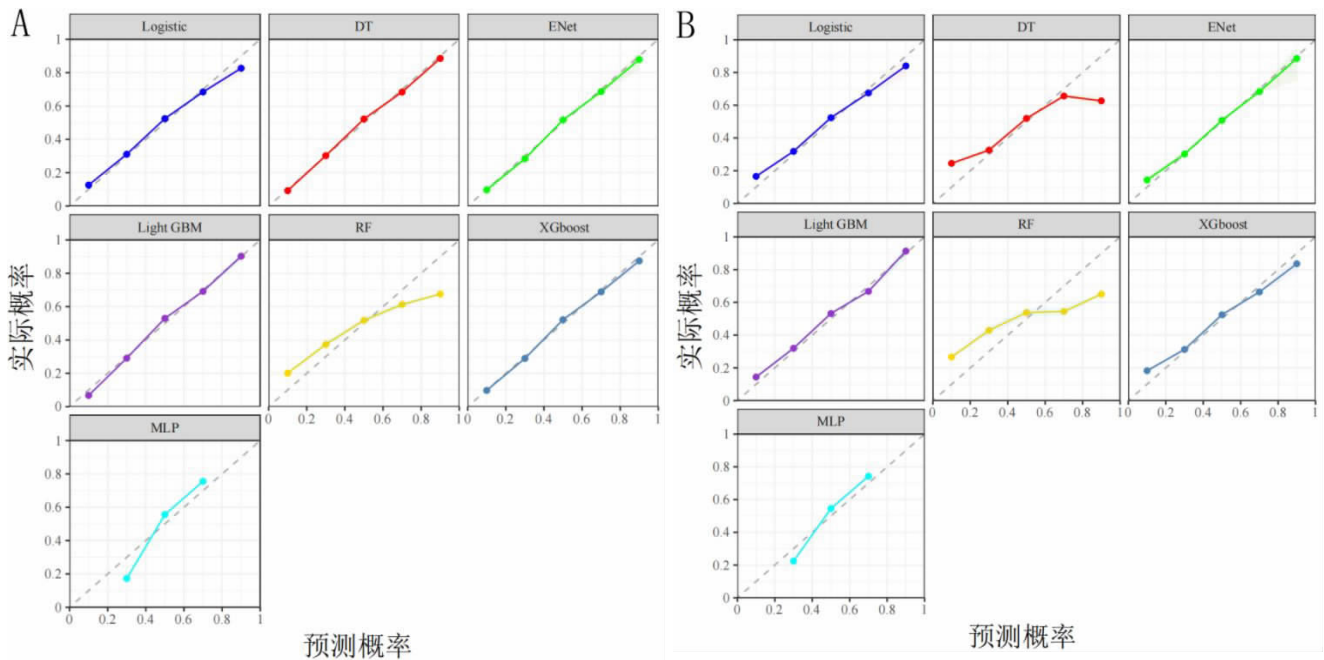
特征	复发	未复发	单因素 logistic 回归分析		多因素 logistic 回归分析	
			OR(95%CI)	P 值	aOR(95%CI)	P 值
性别						
女性	3 142	20 984	1		1	
男性	3 468	21 038	1.101 (1.045 ~ 1.160)	<0.001	1.175 (1.113 ~ 1.241)	<0.001
年龄段(岁)						
0 ~ <30	306	4 209	1		1	
30 ~ <45	400	4 649	1.183 (1.014 ~ 1.381)	0.033	1.092 (0.928 ~ 1.284)	0.288
45 ~ <60	1 462	9 336	2.154 (1.894 ~ 2.449)	<0.001	1.864 (1.623 ~ 2.140)	<0.001
≥60	4 442	23 828	2.564 (2.273 ~ 2.892)	<0.001	2.200 (1.930 ~ 2.508)	<0.001
职业						
学生和教师	51	1 212	1		1	
居家人员	531	3 829	3.296 (2.455 ~ 4.424)	<0.001	1.712 (1.251 ~ 2.343)	0.001
农民	5 954	36 064	3.923 (2.961 ~ 5.199)	<0.001	2.227 (1.651 ~ 3.004)	<0.001
其他	74	917	1.918 (1.329 ~ 2.768)	0.001	1.255 (0.857 ~ 1.838)	0.243
就诊方式						
被动方式	4 678	21 089	1		1	
主动方式	1 932	20 933	0.416 (0.393 ~ 0.440)	<0.001	0.515 (0.484 ~ 0.548)	<0.001
是否单纯结核性胸膜炎						
否	6 597	41 398	1		1	
是	13	624	0.131 (0.075 ~ 0.227)	<0.001	0.333 (0.182 ~ 0.611)	<0.001
治疗结局						
完成疗程	5 491	32 139	1		1	
治愈	1 119	9 883	0.663 (0.619 ~ 0.709)	<0.001	0.543 (0.483 ~ 0.612)	<0.001
是否使用 FDC						
否	651	5 238	1		1	
是	5 959	36 784	1.303 (1.196 ~ 1.420)	<0.001	1.190 (1.072 ~ 1.321)	0.001
初诊痰检结果						
阴性	5 677	37 668	1		1	
阳性	933	4 354	1.422 (1.318 ~ 1.534)	<0.001	2.523 (2.240 ~ 2.843)	<0.001
初治两个月末痰检结果						
阴性	6 523	41 723	1		1	
阳性	87	299	1.861 (1.464 ~ 2.366)	<0.001	2.027 (1.567 ~ 2.622)	<0.001
初诊痰培养结果						
阴性	711	12 067	1		1	
阳性	614	5 900	1.844 (1.650 ~ 2.061)	<0.001	1.728 (1.520 ~ 1.965)	<0.001
无结果	5 258	24 055	3.710 (3.420 ~ 4.024)	<0.001	2.559 (2.347 ~ 2.790)	<0.001
诊疗延误时长(d)						
<30	2 693	21 044	1		1	
≥30	3 917	20 978	1.459 (1.384 ~ 1.538)	<0.001	1.238 (1.158 ~ 1.324)	<0.001
治疗方案						
2HRZE/4HR	6 554	40 480	1		1	
2HRZE/7-10HRE	34	1 122	0.187 (0.133 ~ 0.264)	<0.001	0.392 (0.269 ~ 0.571)	<0.001
其他	22	420	0.324 (0.211 ~ 0.497)	<0.001	0.572 (0.366 ~ 0.896)	0.015
管理单位类型						
医院	2 706	18 253	1		1	
疾病预防控制中心	3 904	23 769	1.108 (1.051 ~ 1.168)	<0.001	1.144 (1.073 ~ 1.218)	<0.001
现住地结核病负担						
低	589	4 112	1		1	
中	1 052	7 294	1.007 (0.904 ~ 1.122)	0.901	1.068 (0.944 ~ 1.210)	0.296
高	1 889	15 232	0.866 (0.784 ~ 0.956)	0.004	0.888 (0.796 ~ 0.992)	0.035
严重	3 080	15 384	1.398 (1.272 ~ 1.536)	<0.001	1.351 (1.207 ~ 1.511)	<0.001



注:A 为训练集;B 为验证集。

图 1 复发预测模型训练集和验证集的 ROC 图

Figure 1 ROC curves of the recurrence prediction model for training (A) and validation set (B)



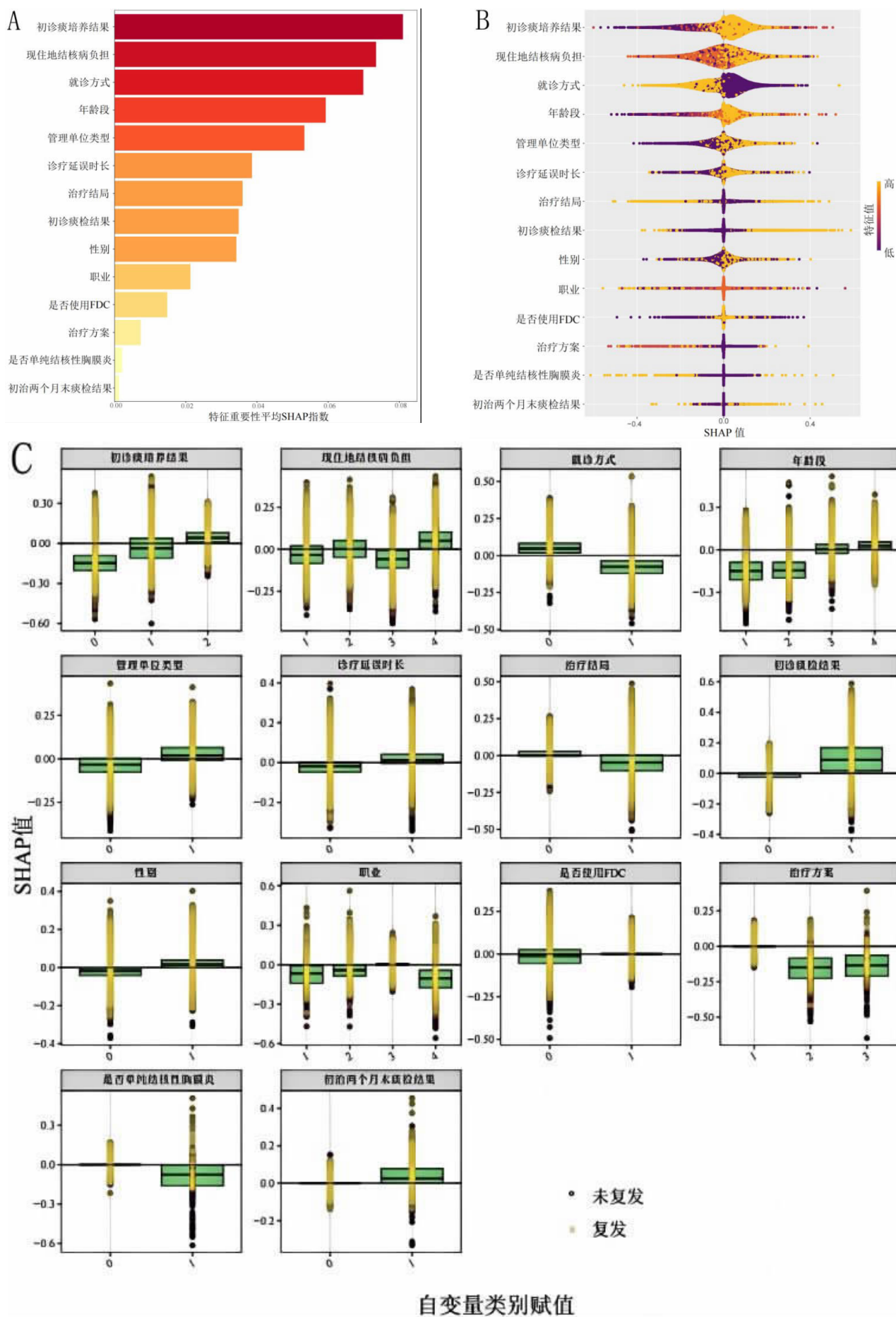
注:A 为训练集;B 为验证集。

图 2 复发预测模型训练集和验证集的校准曲线图

Figure 2 Calibration curves of recurrence prediction model for training and validation set

本研究结果显示,截至 2023 年末,2016—2022 年喀什地区成功治疗的初治结核病患者中复发者占 13.59%,高于我国其他地区^[10-11]。复发率男性高于女性,≥60 岁者高于其他年龄段者,农牧民高于其他职业者。该特征与多个地区类似,如韩国一项回顾性国家队列研究发现,糖尿病合并结核病患者中男性、中老年者是结核复发的独立预测因素^[12];中国衡阳市一项研究发现,男性肺结核患者复发风险是女性的 1.592 倍,≥60 岁者复发风险是其他年龄段的

8.175 倍,农民复发风险是其他职业者的 1.379 倍^[13]。男性复发率较高可能与职业暴露,吸烟及饮酒等行为因素相关,这些因素可削弱免疫功能并增加再感染风险;而≥60 岁人群的高复发率可能与年龄相关的免疫衰老及合并症(如糖尿病)有关,导致潜伏感染的再激活风险显著升高;农牧民的高复发率则可能与其职业特性相关,包括医疗资源可及性低、劳动强度大导致的营养状况较差,以及居住地结核病传播压力较高等综合因素^[14]。



注:A 这特征重要性排序;B 为 SHAP 蜂群图;C 为 SHAP 值的特征依赖图。

图 3 特征重要性排序、SHAP 蜂群图及 SHAP 值的特征依赖图

Figure 3 Feature importance ranking, SHAP swarm plot and SHAP value feature dependence plot

本研究构建的 7 种机器学习模型均具有较好的预测效能,其中决策树模型区分度最佳,校准度良好。决策树模型在处理复杂的监测数据时具有较高的适应性和解释性,对于临床医生制定个性化治疗方案具有重要意义。对决策树模型进行 SHAP 解释,SHAP 特征重要性排序结果显示,初诊痰培养结果、现驻地结核病负担和就诊方式是初治结核病患者复发的三大重要预测因子。痰培养是确诊肺结核的金标准之一,痰培养阳性患者复发风险升高,可能与初始菌量高、病情严重(如肺部空洞)及药物难以彻底清除病灶细菌相关;而痰培养结果未知的患者复发风险也显著升高,可能因缺乏病原学证据导致治疗方案不充分,尤其潜在耐药菌株未被识别和治疗,增加了复发风险。结核病负担严重地区通常伴随着广泛的结核分枝杆菌传播,提高治疗成功患者外源性再感染风险。同时,结核病高负担地区可能存在医疗资源不足、治疗不规范及耐药菌株流行等状况,进一步加剧复发风险。就诊方式主要分为健康体检与主动筛查等主动来源方式和转诊、追踪、直接就诊与推介等被动来源方式。主动方式通常由患者或医疗机构主动发起,患者在无明显症状时通过定期体检或筛查被发现。一项系统综述表明,主动来源是患者成功治疗后复发的保护因素^[1]。主动来源的患者通常在在疾病早期病情较轻时被发现,能够及时开始规范的抗结核治疗,减少了治疗延误和不规律服药的情况,这可能显著降低了复发风险。此外,管理单位为疾控中心、诊疗延误时长 ≥ 30 d、初诊痰检阳性、初治两个月末痰检阳性、使用 FDC 均在一定程度上增加初治结核病患者复发风险,而治愈、治疗方案为 2HRZE/7-10HRE 及单纯结核性胸膜炎降低复发风险,这在不同研究中均被报道^[15-16]。提示在结核病防控中,优化治疗管理、缩短诊疗延误、规范治疗方案以及加强高危人群的随访管理,是降低复发率的重要策略。

本研究有以下局限性。数据来源于《结核病管理信息系统》,可能因录入质量问题低估复发率;仅分析了病案中已登记且缺失较少的变量,缺乏对合并症(如 HIV 感染、糖尿病)及营养状况等潜在重要因素的探索;数据来自特定地区,可能存在地域偏差,模型虽通过内部验证,但未进行外部验证,泛化能力尚不明确;模型实际临床适用性仍缺乏充分验证。

综上,本研究利用喀什地区结核病常规监测数据,构建了基于 7 种机器学习算法的初治结核病患者复发预测模型,其中基于决策树的模型具有较好的预测效能,能够为结核病复发的个体化防控提供科学依据。此外,SHAP 模型解释分析表明,初诊痰培养结果、现驻地结核病负担和就诊方式是复发的重要预测

因子。

利益冲突声明 本研究不存在任何利益冲突

参考文献

- [1] 李涛, 赵雁林, 贾忠伟. 结核病复发及其影响因素研究进展[J]. 疾病监测, 2024, 39(4): 403-411.
Li T, Zhao YL, Jia ZW. Progress in research of tuberculosis recurrence and influencing factors[J]. Disease Surveillance, 2024, 39(4): 403-411.(In Chinese)
- [2] Teferi MY, El-Khatib Z, Boltana MT, et al. Tuberculosis treatment outcome and predictors in Africa: a systematic review and Meta-Analysis [J]. International Journal of Environmental Research and Public Health, 2021, 18(20): 10678.
- [3] 钟倩红, 马晓慧, 钟永辉, 等. 2014-2018 年佛山市肺结核患者复发情况及影响因素分析 [J]. 中国防痨杂志, 2022, 44(7): 690-697.
Zhong QH, Ma XH, Zhong YH, et al. Analysis of recurrence rate of pulmonary tuberculosis patients in Foshan and the influencing factors from 2014 to 2018 [J]. Chinese Journal of Antituberculosis, 2022, 44(7): 690-697.(In Chinese)
- [4] Hung CL, Chien JY, Ou CY. Associated factors for tuberculosis recurrence in Taiwan: a nationwide nested case-control study from 1998 to 2010[J]. PLOS One, 2015, 10(5): e0124822.
- [5] Luo Y, Xue Y, Song HJ, et al. Machine learning based on routine laboratory indicators promoting the discrimination between active tuberculosis and latent tuberculosis infection[J]. Journal of Infection, 2022, 84(5): 648-657.
- [6] Falcao I, Cardoso DL, Coutinho DSA, et al. Model for predicting drug resistance based on the clinical profile of tuberculosis patients using machine learning techniques [J]. PeerJ Computer Science, 2024, 10: e2246.
- [7] Kong HB, Li Y, Shen Y, et al. Predicting the risk of pulmonary embolism in patients with tuberculosis using machine learning algorithms [J]. European Journal of Medical Research, 2024, 29(1): 618.
- [8] Anh NK, Phat NK, Thu NQ, et al. Discovery of urinary biosignatures for tuberculosis and nontuberculous mycobacteria classification using metabolomics and machine learning [J]. Scientific Reports, 2024, 14(1): 15312.
- [9] Chen JF, Jiang YL, Li ZH, et al. Predictive machine learning models for anticipating loss to follow-up in tuberculosis patients throughout anti-TB treatment journey[J]. Scientific Reports, 2024, 14(1): 24685.
- [10] 彭红, 虞浩, 姜洁, 等. 江苏省初治结核病复发流行病学特征及影响因素[J]. 江苏预防医学, 2019, 30(4): 355-359.
Peng H, Yu H, Jiang J, et al. Epidemiological characteristics and associated factors with tuberculosis recurrence in Jiangsu province[J]. Jiangsu Journal of Preventive Medicine, 2019, 30(4): 355-359.(In Chinese)
- [11] Ruan QL, Yang QL, Sun F, et al. Recurrent pulmonary tuberculosis after treatment success: a population-based retrospective study in China[J]. Clinical Microbiology and Infection, 2022, 28(5): 684-689.
- [12] Eksombatchai D, Jeong D, Mok J, et al. Sex differences in the impact of diabetes mellitus on tuberculosis recurrence: a retrospective National cohort study[J]. International Journal of Infectious Diseases,

尽管既往的数据表明腰痛对女性造成的负担要大于男性,但根据本研究的模型预测,未来男性腰痛的发病率将上升,而女性的发病率将下降,这可能与我国社会生产模式转型及居民生活习惯转变有关。在 GBD 数据库中,腰痛有三大相关风险因素:职业工效学、吸烟和身体质量指数(body mass index,BMI)。近年来,职业工效学和吸烟对腰痛的影响有所下降^[12]。这是因为我国工业化进程导致过去许多重体力劳动被机器所取代,并且我国年轻群体的吸烟率在逐年下降^[13]。然而,高 BMI 对腰痛的影响却在逐年上升^[12]。据调查,在儿童和青少年阶段,中国男性的 BMI 及腰围普遍高于女性^[14]。而且,成年男性的 BMI 及腰围的逐年增长率也高于女性^[15]。因此,日益增长的 BMI 是男性腰痛的发病率上升的重要原因。

最后,本研究存在一定局限性。GBD 对疾病负担的估计主要依赖于预测模型,而非实际的观察数据。因此,在制定公共卫生政策时,应谨慎解读 GBD 的研究结果。

利益冲突声明 本研究不存在任何利益冲突

参考文献

- [1] GBD 2021 Low Back Pain Collaborators. Global, regional, and National burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021 [J]. LANCET RHEUMATOLOGY, 2023, 5(6): e316–e329.
- [2] Wang LF, Ye H, Li ZC, et al. Epidemiological trends of low back pain at the global, regional, and National levels [J]. European Spine Journal, 2022, 31(4): 953–962.
- [3] Gu Y, Wang ZM, Shi HW, et al. Global, regional, and National change patterns in the incidence of low back pain from 1990 to 2019 and its predicted level in the next decade [J]. International Journal of Public Health, 2024, 69: 1606299.
- [4] GBD 2021 Risk Factors Collaborators. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021[J]. Lancet, 2024, 403(10440): 2162–2203.
- [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] Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution [J]. Medicine Statistics, 1997, 16(7): 791–801.
- [7] Clegg LX, Hankey BF, Tiwari R, et al. Estimating average annual per cent change in trend analysis [J]. Medicine Statistics, 2009, 28(29): 3670–3682.
- [8] Yip Y, Ho SC, Chan S. Socio-psychological stressors as risk factors for low back pain in Chinese middle-aged women [J]. Journal of Advanced Nursing, 2001, 36(3): 409–416.
- [9] Carvalho MECC, Lima LC, de Lira Terceiro CA, et al. Lumbalgia na gestação low back pain during pregnancy [J]. Rev Bras Anestesiologia, 2017, 67(3): 266–270.
- [10] Amayri A, Khalayli N, Haj AD, et al. Low back pain in a sample of Syrian pregnant women: A cross-sectional study [J]. Health Sci Rep, 2023, 6(7): e1389.
- [11] Lu YM, Chen CH, Lue YJ. A cross-sectional study of disability and quality of Life in patients with low back pain: Focus on sex and gender[J]. Journal of Back and Musculoskeletal Rehabilitation, 2022, 35(1): 177–184.
- [12] 吴惠一,袁琴,张洋,等. 1990—2019 年我国肌肉骨骼疾病负担分析[J]. 中华疾病控制杂志,2023,27(6):655–661.
Wu HY, Yuan Q, Zhang Y, et al. Disease burden of musculoskeletal disorders in China from 1990 to 2019[J]. Chinese Journal of Disease Control & Prevention, 2023, 27(6): 655–661.(In Chinese)
- [13] Li SS, Meng LH, Chioloro A, et al. Trends in smoking prevalence and attributable mortality in China, 1991 – 2011[J]. Preventive Medicine, 2016, 93: 82–87.
- [14] 杜文雯,房红芸,周玮玮,等. 1982–2016 年中国 6–17 岁儿童青少年体质指数和腰围变化趋势[J]. 卫生研究,2025,54(2): 175–180.
Du WW, Fang HY, Zhou WW, et al. Trends of body mass index and waist circumference of children and adolescents aged 6–17 years in China from 1982 to 2016 [J]. Health Research, 2025, 54 (2): 175–180.(In Chinese)
- [15] Zhai Y, Fang HY, Yu WT, et al. Changes in waist circumference and abdominal obesity among Chinese adults over a ten-year period[J]. Biomedical and Environmental Sciences, 2017, 30(5): 315–322.

收稿日期:2025–03–03

(上接第 2316 页)

2023, 127: 1–10.

- [13] 周旭,王巧智,万燕萍,等. 衡阳市成功治疗的肺结核患者 5 年复发影响因素分析[J]. 预防医学,2022,34(3):268–271.
Zhou X, Wang QZ, Wan YP, et al. Factors affecting the five-year recurrence among successfully-treated pulmonary tuberculosis patients in Hengyang City [J]. Journal of Preventive Medicine, 2022, 34(3): 268–271.(In Chinese)
- [14] Mujtaba MA, Richardson M, Shahzad H, et al. Demographic and clinical determinants of tuberculosis and TB recurrence: a Double-Edged retrospective study from Pakistan [J]. Journal of

Tropical Medicine, 2022, 2022(1): 4408306.

- [15] Vega V, Cabrera-Sanchez J, Rodríguez S, et al. Risk factors for pulmonary tuberculosis recurrence, relapse and reinfection: a systematic review and meta-analysis [J]. BMJ Open Respir Res, 2024, 11(1): e002281.
- [16] Qiu BB, Wu ZC, Tao BL, et al. Risk factors for types of recurrent tuberculosis (reactivation versus reinfection): A global systematic review and meta-analysis [J]. International Journal of Infectious Diseases, 2022, 116: 14–20.

收稿日期:2025–03–12