

· 论著 ·

随机森林和决策树模型在轻型缺血性脑卒中患者复发预测中的应用分析

莫秋红 丁晓波 张岩波 李伟荣

030000 太原, 山西医科大学公共卫生学院(莫秋红、丁晓波、张岩波、李伟荣); 030000

太原, 山西省心血管病医院 山西医科大学附属心血管病医院神经内科(李伟荣)

通信作者: 李伟荣, Email: weironglee@163.com

DOI: 10.3969/j.issn.1009-6574.2024.02.001

【摘要】目的 构建轻型缺血性脑卒中(MIS)患者2年内复发的随机森林和决策树预测模型,并分析模型的临床应用价值。**方法** 回顾性收集2020年7月1日至12月31日于山西省心血管病医院神经内科就诊的520例MIS患者的病历资料,根据2年内是否复发将患者分为复发组和未复发组。基于缺失森林对数据进行填补,根据文献检索与专家讨论结果筛选预测变量并进行单因素分析,合成少数过采样技术-标称连续(SMOTE-NC)技术处理数据不平衡,采用贝叶斯优化十折交叉验证构建随机森林、决策树模型并与Logistic回归模型进行比较。基于受试者工作特征曲线下面积(AUC)、布里尔分数(BS)与校准曲线分别评价模型的区分度与校准度。对预测性能最好的模型采用SHAP模型解释预测结果。**结果** 2年内复发患者共93例(17.9%)。两组患者的年龄、吸烟、糖尿病、循环梗死部位、多发性脑梗死比例,以及舒张压、红细胞压积、血小板计数、低密度脂蛋白水平比较,差异有统计学意义($P < 0.05$)。Logistic回归模型、决策树模型与随机森林模型在测试集中,预测MIS患者2年内复发情况的AUC(95%CI)分别为0.764(0.691~0.835)、0.743(0.668~0.818)、0.892(0.843~0.941),BS分别为0.200、0.211、0.142,随机森林预测效果最好,准确度为0.822,灵敏度为0.818,阳性预测值为0.808,阴性预测值为0.835。SHAP分析结果显示,随机森林模型中重要性排序前5名的变量分别是年龄、低密度脂蛋白、吸烟、糖尿病、舒张压。**结论** 与决策树和Logistic回归模型相比,随机森林模型预测MIS 2年内复发的性能较好。

【关键词】 卒中; 轻型缺血性卒中; 复发; 随机森林; Logistic回归; 决策树

基金项目: 山西省重点研发计划项目(2021XM14)

Recurrence prediction of patients with minor ischemic stroke based on random forest and decision tree models Mo QiuHong, Ding Xiaobo, Zhang Yanbo, Li Weirong

School of Public Health, Shanxi Medical University, Taiyuan 030000, China (Mo QH, Ding XB, Zhang YB, Li WR); Department of Neurology, Shanxi Cardiovascular Hospital & Affiliated Cardiovascular Hospital of Shanxi Medical University, Taiyuan 030000, China (Li WR)

Corresponding author: Li Weirong, Email: weironglee@163.com

【Abstract】Objective To construct the random forest and decision tree prediction model for recurrence of minor ischemic stroke (MIS) within two years, and analyze the predicted performance of the models. **Methods** The medical records of 520 MIS patients who visited Department of Neurology of Shanxi Cardiovascular Hospital from July 1 to December 31, 2020 were retrospectively collected. Patients were divided into a recurrent group and a non-recurrent group based on whether they relapsed within two years. This study filled in the data through the missing forest. Based on literature search and expert discussion, predictive variables were selected and univariate analysis was conducted, and addressed data imbalance through the synthetic minority over-sampling technique-nominal continuity (SMOTE-NC). Random forest and decision tree models were constructed using Bayesian optimization 10-fold cross validation and compared with the Logistic regression model. The discrimination and calibration of the models were evaluated based on the area under the receiver operating characteristic curve (AUC), Brier score (BS), and calibration curve. The prediction results of the model with the excellent predictive performance were explained using the SHapley Additive exPlanations (SHAP) model. **Results** A total of 93 patients (17.9%) experienced recurrence within two years. There were statistical differences between the two groups in age, smoking, diabetes, location of circulatory infarction,

multiple cerebral infarction, diastolic pressure, hematocrit, platelet count, and low-density lipoprotein ($P < 0.05$). The *AUC* (95%*CI*) of the testing set of Logistic regression model, decision tree model, and random forest model for predicting recurrence within two years in patients with MIS were 0.764 (0.691, 0.835), 0.743 (0.668, 0.818), 0.892 (0.843, 0.941), and BS were 0.200, 0.211, and 0.142, respectively. The random forest model had the excellent prediction performance, with an accuracy of 0.822, a sensitivity of 0.818, a positive prediction value of 0.808, and a negative prediction value of 0.835. SHAP analysis showed that the top five variables in the random forest model were age, low-density lipoprotein, smoking, diabetes, and diastolic pressure. **Conclusions** Compared with decision tree model and Logistic regression model, the random forest model performs better in predicting the recurrence of MIS within two years.

【Key words】 Stroke; Minor ischemic stroke; Recurrence; Random forest; Logistic regression; Decision trees

Fund program: Key Research and Development Plan Project of Shanxi Province (2021XM14)

《中国脑卒中防治报告 2020》显示,缺血性脑卒中已成为我国成人致死、致残的首要原因,占有脑卒中的 60%~80%^[1]。轻型缺血性脑卒中(minor ischemic stroke, MIS)无致残性^[2],约占我国脑卒中的 30%^[3]。在实施标准化治疗后,仍有部分 MIS 患者复发。早期识别和进行危险分层对于预防 MIS 复发具有重要意义。目前,国内外多数研究采用 Logistic 回归模型和 Cox 模型等传统机器学习方法构建 MIS 复发预测模型^[4-6],这些模型使用的变量范围和临床实用性往往受限。新型机器学习可以处理将大量变量与高维数据结合的复杂数据,并有较好的模型性能,可用于筛查、诊断或评估预后。采用随机森林与决策树构建 MIS 复发的模型较为少见。基于此,本研究构建 MIS 患者 2 年内复发的新型机器学习随机森林和决策树模型,并与传统 Logistic 回归模型进行比较,以期对 MIS 复发的预防和早期干预提供辅助决策。

对象与方法

一、研究对象

回顾性收集 2020 年 7 月 1 日至 12 月 31 日于山西省心血管病医院神经内科就诊的 520 例 MIS 患者的病历资料。纳入标准:(1)符合《中国急性缺血性脑卒中诊治指南 2018》^[7]中急性缺血性脑卒中的诊断标准;(2)发病 7 d 内就诊并住院治疗^[8];(3)入院 NIHSS 评分 ≤ 5 分^[9];(4)一般资料、影像学与实验室检查等资料完整;(5)患者接受 2 年的随访评估。排除标准:(1)合并脑肿瘤、创伤性脑损伤或其他脑损伤、动脉瘤或动静脉畸形;(2)合并痴呆症或精神疾病。本研究排除了 25 例资料不完整、辅助检查不充分的患者,33 例因各种原因随访不足 2 年的患者,最终纳入 520 例 MIS 患者,根据 2 年内是否复发脑卒中分为复发组和非复发组。本研究已获得山西省心血管病医院伦理委员会审批。

二、方法

1. 观察指标:通过文献检索与神经内科专家讨论,从患者一般资料、实验室检查结果(在患者入院 24 h 内采集)、影像学检查结果等资料中筛选出 24 个关键变量。一般资料包括性别、年龄、体重指数(body mass index, BMI)、饮酒情况、吸烟情况、糖尿病、冠心病、NIHSS 评分、舒张压、收缩压;实验室检查结果包括白细胞计数、红细胞压积、嗜中性粒细胞百分比、血小板计数、纤维蛋白原、低密度脂蛋白、高密度脂蛋白、同型半胱氨酸、血清肌酐、凝血酶原时间;影像学检查结果包括脑白质病变、多发性脑梗死、循环梗死部位、颅内动脉狭窄。经治疗出院后,通过住院记录及电话随访调查患者 2 年内脑卒中是否复发,复发性脑卒中定义为原有神经缺失症状和体征好转或消失后,再次出现新的神经功能缺失症状;经头颅 CT 和(或)MRI 检查证实有新发的缺血性病灶;排除进展性卒中或病情恶化所致^[7]。

2. 数据预处理:为减少缺失数据导致的偏差,排除缺失值超过 20% 的变量^[10],采用缺失森林进行数据填补,单因素分析进行变量筛选,将 $P < 0.05$ 的变量作为自变量纳入模型。采用合成少数过采样技术-标称连续(Synthetic minority oversampling technique-nominal continuous, SMOTE-NC)技术进行数据平衡,SMOTE-NC 技术是对 SMOTE 的高级改进,它对标称特征进行了中位数计算,对混合数据进行了最近邻计算,能同时处理连续变量与分类变量的混合数据集^[11]。采用分层抽样的方法,从复发患者和非复发患者中分别抽取 80% 样本组成训练集,在十折交叉验证下构建决策树、随机森林、Logistic 回归模型,将剩余的 20% 样本作为测试集用于评价模型性能。

3. 统计学方法:采用 SPSS 25.0 软件与 R 软件 4.3.0 版本对研究数据进行整理和统计分析,由 2 名研究者核对录入数据。采用 Shapiro-Wilk 检验方法进行正态分布检验,符合正态分布的计量资料采用均数 \pm 标准差($\bar{x} \pm s$)表示,组间比较采用独立样本 *t*

检验;不符合正态分布的计量资料采用中位数和四分位数 [$M(P_{25}, P_{75})$]表示,组间比较采用 Wilcoxon 秩和检验。计数资料用频数、百分数(%)表示,组间比较采用 χ^2 检验。双侧检验, $P < 0.05$ 表示差异有统计学意义。采用受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under curve, AUC)评价模型的区分度,使用校准曲线及布里尔分数(Brier score, BS)评价模型的校准程度,采用灵敏度、特异度、准确性、阳性预测值、阴性预测值对模型进行比较,对预测性能最好的模型采用 SHAP 进行模型变量重要性解释,计算最优模型中各变量的贡献,确定变量的重要程度和对模型的正或负贡献^[12]。

结 果

1. 两组 MIS 患者的一般资料比较: 93 例(17.9%)患者在 2 年随访内卒中复发。复发组与非复发组患者年龄、循环梗死部位,吸烟、合并糖尿病、多发性脑梗死比例,舒张压、红细胞压积、血小板计数、低密度脂蛋白水平比较,差异有统计学意义($P < 0.05$),见表 1。

2. 机器学习模型表现: 将单因素分析筛选出的 9 个变量纳入 3 种机器学习模型中进行测试, 3 种机器学习模型测试集中 AUC 值从高到低依次排列为随机森林模型、Logistic 回归模型和决策树模型,见表 2、图 1。测试集校准曲线中 BS 从低到高依次为随机森林模型、Logistic 回归模型和决策树模型,见图 2。十折交叉验证指标平均值图显示,随机森林模型的十折交叉验证指标平均值最高,见图 3。

3. 基于随机森林模型的 SHAP 变量重要性分析: 在随机森林模型中,各变量平均权重的绝对值从高到低依次为年龄(0.889)、低密度脂蛋白(0.072)、吸烟(0.068)、糖尿病(0.064)、舒张压(0.062)、循环梗死部位(0.060)、红细胞压积(0.050)、血小板计数(0.045)、多发性脑梗死(0.012),见图 4。分类变量中,前循环与后循环梗死两者都有、糖尿病、吸烟、多发性脑梗死患者的 shapley 值为正值,为增加 MIS 患者 2 年内的复发风险的变量,见图 5。连续变量中,年龄、低密度脂蛋白、舒张压越大,shapley 值越大,为增加 MIS 患者 2 年内的复发风险的变量。见图 6。

表 1 两组轻型缺血性卒中患者的一般资料、实验室及影像学检查结果比较

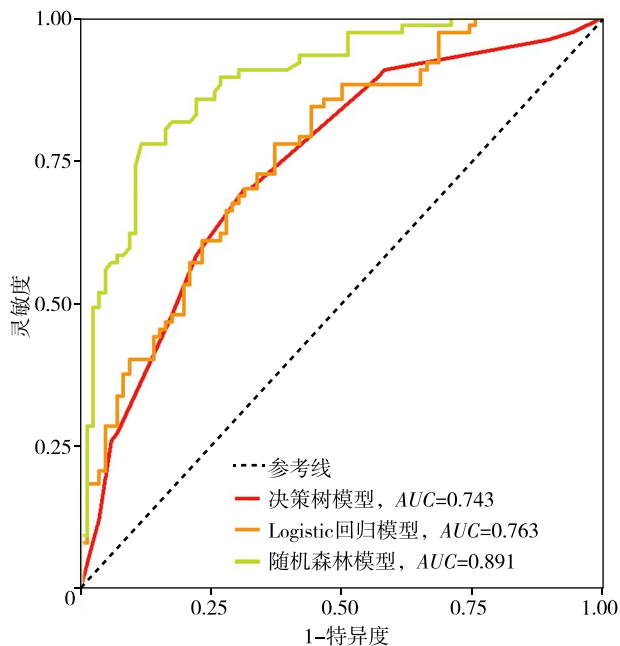
项目	非复发组(n=427)	复发组(n=93)	t/Z/ χ^2 值	P值
女性[例(%)]	121(28.3)	27(29.0)	0.018	0.893
年龄[岁, $M(P_{25}, P_{75})$]	63.0(55.0, 70.0)	66.0(58.0, 74.0)	-2.624	0.009
BMI[kg/m^2 , $M(P_{25}, P_{75})$]	24.7(22.9, 26.7)	24.6(22.6, 26.6)	-0.106	0.916
吸烟[例(%)]	219(51.3)	63(67.7)	8.330	0.004
饮酒[例(%)]	178(41.7)	40(43.0)	0.055	0.815
合并糖尿病[例(%)]	112(26.2)	44(47.3)	16.163	< 0.001
合并冠心病[例(%)]	122(28.6)	31(33.3)	0.834	0.361
脑白质病变[例(%)]	386(90.4)	85(91.4)	0.089	0.765
循环梗死部位[例(%)]				
前循环	249(58.4)	44(47.3)		
后循环	121(28.3)	21(22.6)	15.688	< 0.001
两者都有	57(13.3)	28(30.1)		
颅内动脉狭窄[例(%)]	263(61.6)	66(71.0)	2.880	0.890
多发性脑梗死[例(%)]	222(52.0)	61(65.5)	5.695	0.017
NIHSS 评分[分, $M(P_{25}, P_{75})$]	2.0(0, 3.0)	2.0(0.5, 3.0)	-0.149	0.882
收缩压[mmHg, $M(P_{25}, P_{75})$]	140.0(130.0, 152.0)	144.0(132.0, 164.0)	-1.628	0.103
舒张压[mmHg, $M(P_{25}, P_{75})$]	82.0(78.0, 90.0)	86.0(80.0, 94.0)	-2.199	0.028
白细胞计数[$\times 10^9/\text{L}$, $M(P_{25}, P_{75})$]	6.5(5.5, 7.7)	6.6(5.6, 7.7)	-0.526	0.599
红细胞压积(% , $\bar{x} \pm s$)	42.2 \pm 4.3	43.3 \pm 4.8	-2.172	0.030
血小板计数[$\times 10^9/\text{L}$, $M(P_{25}, P_{75})$]	195.0(162.0, 235.0)	210.0(176.0, 251.0)	-2.135	0.033
嗜中性粒细胞百分比(% , $\bar{x} \pm s$)	63.3 \pm 9.3	64.2 \pm 9.6	-0.881	0.379
纤维蛋白原[g/L, $M(P_{25}, P_{75})$]	3.0(2.6, 3.4)	3.0(2.7, 3.5)	-0.957	0.339
凝血酶原时间[s, $M(P_{25}, P_{75})$]	15.0(14.4, 15.7)	15.1(14.5, 15.8)	-0.433	0.665
血清肌酐[$\mu\text{mol}/\text{L}$, $M(P_{25}, P_{75})$]	65.2(57.2, 72.9)	67.2(58.8, 75.5)	-1.161	0.246
同型半胱氨酸[$\mu\text{mol}/\text{L}$, $M(P_{25}, P_{75})$]	22.0(17.0, 30.0)	21.0(16.0, 28.0)	-1.024	0.306
高密度脂蛋白[mmol/L, $M(P_{25}, P_{75})$]	0.9(0.8, 1.1)	0.9(0.8, 1.0)	-0.601	0.548
低密度脂蛋白[mmol/L, $M(P_{25}, P_{75})$]	2.5(2.1, 3.0)	2.7(2.3, 3.3)	-3.135	0.002

注: BMI 体重指数; NIHSS 美国国立卫生研究院卒中评分; 1 mmHg=0.133 kPa

表2 3种模型训练集与测试集预测轻型缺血性脑卒中患者2年内复发情况的性能表现

模型	数据集	AUC(95%CI)	灵敏度	特异度	阳性预测值	阴性预测值	准确度
Logistic回归模型	训练集	0.791(0.757 ~ 0.824)	0.720	0.727	0.704	0.743	0.724
	测试集	0.764(0.691 ~ 0.835)	0.662	0.721	0.680	0.705	0.693
决策树模型	训练集	0.836(0.806 ~ 0.867)	0.840	0.727	0.735	0.835	0.781
	测试集	0.743(0.668 ~ 0.818)	0.701	0.686	0.667	0.720	0.693
随机森林模型	训练集	0.984(0.977 ~ 0.992)	0.925	0.950	0.944	0.934	0.938
	测试集	0.892(0.843 ~ 0.941)	0.818	0.826	0.808	0.835	0.822

注: AUC 曲线下面积; CI 置信区间



注: ROC 受试者工作特征; AUC 曲线下面积

图1 3种模型预测轻型缺血性脑卒中患者2年内复发情况的ROC曲线

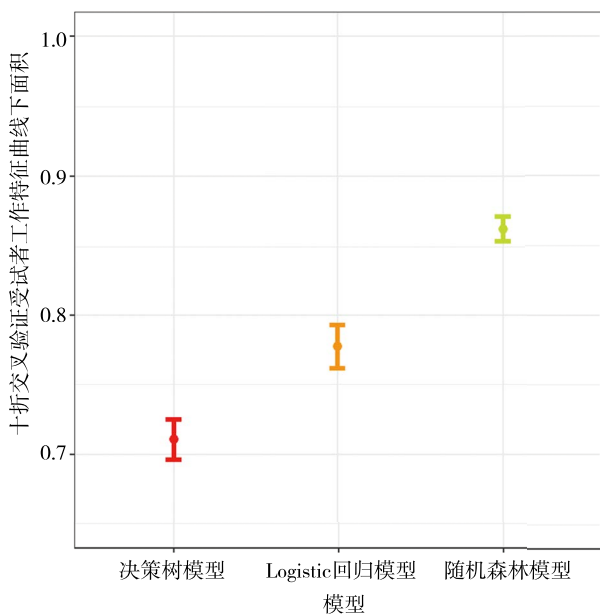
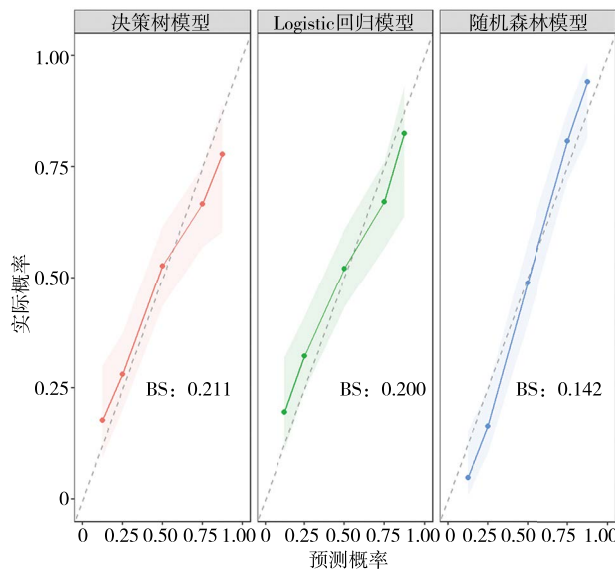


图3 3种模型预测轻型缺血性脑卒中患者2年内复发情况的十折交叉验证指标平均值图



注: BS 布里尔分数

图2 3种模型预测轻型缺血性脑卒中患者2年内复发情况的测试集校准曲线

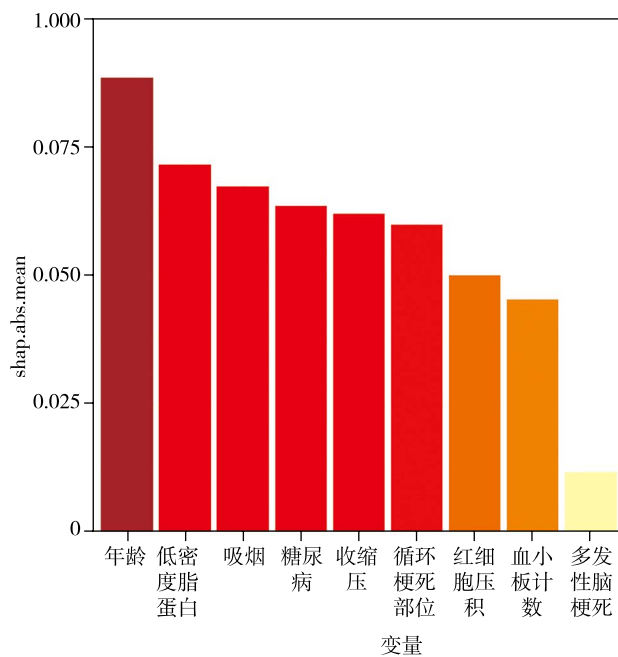
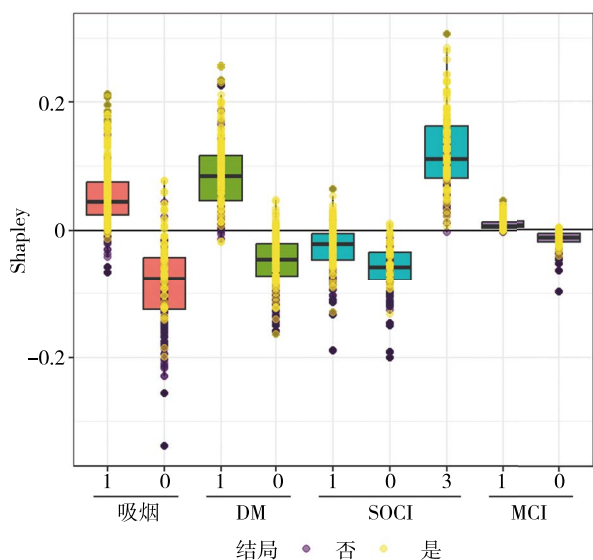
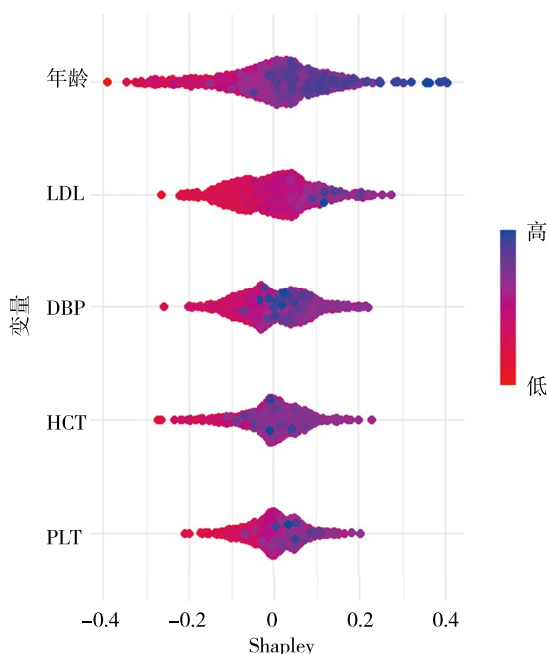


图4 9个预测变量在随机森林模型中的SHAP值



注: DM 糖尿病; SOCl 循环梗死部位; MCl 多发性脑梗死
图5 随机森林模型中4种分类型变量的SHAP图



注: LDL 低密度脂蛋白; DBP 舒张压; HCT 红细胞压积; PLT 血小板计数
图6 随机森林模型中5个连续型变量的SHAP图

讨论

本研究收集MIS患者的临床资料,以患者治疗出院后2年内脑卒中复发作为结局变量,分别采用Logistic回归、决策树与随机森林构建MIS患者的复发预测模型。本研究结果显示,3个模型的校准度均较好,说明3个模型预测患者的2年内复发风险与实际发生风险一致程度高,其中随机森林在测试集中的模型性能最好。

在随机森林模型中,各变量平均权重的绝对值越大,表明这些变量对模型的贡献越大,其在预测中的影响越显著。本研究结果显示,在随机森林模型的SHAP重要性分析中,排名前5名的变量分别为年龄、低密度脂蛋白、吸烟、糖尿病、舒张压,是MIS患者2年内脑卒中复发的影响因素。一项研究表明,脑年龄可能是一种有价值的非侵入性生物标志物,脑年龄较大与实际年龄较大和心室容积较大相关,预测的大脑年龄减去实际年龄越大,病变损伤越多,卒中后时间越长^[13]。既往研究表明低密度脂蛋白水平是与我国卒中患者复发的影响因素^[14]。低密度脂蛋白通过细胞外液将脂质转运到全身,使其可供身体细胞进行受体介导的内吞作用,如果低密度脂蛋白在动脉壁内被氧化,会导致颅内动脉粥样硬化,从而增加MIS复发风险。高血压易引起脑血管系统小血管的动脉粥样硬化和曲折,这些血管的病理变化导致颅内血管狭窄和灌注减少^[15],增加MIS复发风险。糖尿病属于血管危险因素,长期处于高血糖状态可导致血管内皮功能障碍、早期动脉硬化增加、全身炎症和毛细血管基底膜增厚,糖耐量受损会增加非糖尿病患者发生MIS的风险^[16]。这些不同的机制通过影响血管的结构、功能并调节免疫力,增加卒中患病风险。此外有研究发现,吸烟通过促进血栓形成和动脉血管收缩减少脑血流量来增加脑卒中的短期风险^[17-18]。

目前,国内外较少对MIS患者2年内卒中复发情况进行预测,MIS复发预测模型大多基于30d、半年以及1年的短期随访,并大多采用传统机器学习方法。本研究采用决策树和随机森林两种新型机器学习模型对MIS患者2年内卒中复发情况构建了预测模型,显示新型机器学习随机森林模型性的预测效能较优,进一步采用SHAP对模型进行解释,解释性能可提高对机器学习模型的理解和接受度,可协助临床医生更好地理解预测模型结果,从而为MIS患者做出更个性化与合理化的临床决策。

本研究也存在一定的局限性:(1)本研究为单中心研究,纳入的样本量较小,未进行外部验证,结果可能存在偏倚,未来需通过多中心、大规模研究进一步进行外部验证;(2)未纳入临床应用中新型的生物标志物,如已被证实与MIS患者临床预后有关的可溶性CD40配体、脂蛋白相关磷脂酶A2活性、外周血基质金属蛋白酶9等^[19-21]。

综上所述,随机森林模型在对MIS患者2年内卒中复发进行预测时,效果好于决策树与Logistic回归模型。随着图像信息和基因组学的加入,随机森林等机器学习算法可能会帮助医生对MIS患者复发预测情况进行较为准确的判断,并辅助医生进行相应的临床诊疗操作^[22]。

利益冲突 文章所有作者共同认可文章无相关利益冲突

作者贡献声明 研究设计与实施、数据分析为莫秋红、李伟荣, 论文撰写、图表绘制、模型构建为莫秋红, 论文修订为张岩波、李伟荣, 数据收集与整理为莫秋红、丁晓波

参 考 文 献

- [1] 王陇德, 彭斌, 张鸿祺, 等. 《中国脑卒中防治报告 2020》概要[J]. 中国脑血管病杂志, 2022, 19(2): 136-144. DOI: 10.3969/j.issn.1672-5921.2022.02.011.
- Wang LD, Peng B, Zhang HQ, et al. Summary of China Stroke Prevention Report 2020[J]. Chin J Cerebrovasc Dis, 2022, 19(2): 136-144.
- [2] Wang YJ, Li ZX, Gu HQ, et al. China Stroke Statistics: an update on the 2019 report from the National Center for Healthcare Quality Management in Neurological Diseases, China National Clinical Research Center for Neurological Diseases, the Chinese Stroke Association, National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention and Institute for Global Neuroscience and Stroke Collaborations[J]. Stroke Vasc Neurol, 2022, 7(5): 415-450. DOI: 10.1136/svn-2021-001374.
- [3] Guan T, Ma J, Li M, et al. Rapid transitions in the epidemiology of stroke and its risk factors in China from 2002 to 2013 [J]. Neurology, 2017, 89(1): 53-61. DOI: 10.1212/WNL.0000000000004056.
- [4] 李菁晶, 鞠奕, 王春雪, 等. 缺血性小卒中患者 1 年卒中复发的相关因素分析: 来自中国颅内动脉粥样硬化研究的数据[J]. 中国卒中杂志, 2015, 10(8): 643-649. DOI: 10.3969/j.issn.1673-5765.2015.08.003.
- Li JJ, Ju Y, Wang CX, et al. Analysis of factors associated with 1-year stroke recurrence in patients with ischemic stroke: data from the Chinese Intracranial Atherosclerosis Study[J]. Chin J Stroke, 2015, 10(8): 643-649.
- [5] 袁文霞, 韩刚, 雷涛. 缺血性小卒中患者再次卒中危险因素分析[J]. 贵州医药, 2020, 44(7): 1121-1122. DOI: 10.3969/j.issn.1000-744X.2020.07.046.
- Yuan WX, Han G, Lei T. Analysis of risk factors for recurrent stroke in patients with ischemic stroke[J]. Guizhou Medicine, 2020, 44(7): 1121-1122.
- [6] Tian Y, Pan Y, Yan H, et al. Coexistent cerebral small vessel disease and multiple infarctions predict recurrent stroke[J]. Neurol Sci, 2022, 43(8): 4863-4874. DOI: 10.1007/s10072-022-06027-6.
- [7] 中华医学会神经病学分会, 中华医学会神经病学分会脑血管病学组. 中国急性缺血性脑卒中诊治指南 2018 [J]. 中华神经科杂志, 2018, 51(9): 666-682. DOI: 10.3760/ema.j.issn.1006-7876.2018.09.004.
- Cerebrovascular Disease Group, Chinese Society of Neurology, Chinese Medical Association. Guidelines for the diagnosis and treatment of acute ischemic stroke of the Stroke Branch of the Chinese Society of Neurology, 2018 [J]. Chin J Neurol, 2018, 51(9): 666-682.
- [8] Amarenco P, Lavallée PC, Monteiro Tavares L, et al. Five-year risk of stroke after TIA or minor ischemic stroke[J]. N Engl J Med, 2018, 378(23): 2182-2190. DOI: 10.1056/NEJMoa1802712.
- [9] Lin X, Lin S, Cui X, et al. Prediction-driven decision support for patients with mild stroke: a model based on machine learning algorithms[J]. Front Neurol, 2021, 12: 761092. DOI: 10.3389/fneur.2021.761092.
- [10] Zhang K, Fang Y, Fan H, et al. A nomogram for predicting the in-hospital risk of recurrence among patients with minor non-cardiac stroke[J]. Curr Med Res Opin, 2022, 38(4): 487-499. DOI: 10.1080/03007995.2022.2038488.
- [11] 杨弘, 田晶, 王可, 等. 混合型缺失数据填补方法比较与应用[J]. 中国卫生统计, 2020, 37(3): 395-399.
- Yang H, Tian J, Wang K, et al. Comparison and application of hybrid missing data filling methods[J]. Chinese Journal of Health Statistics, 2020, 37(3): 395-399.
- [12] Wang K, Gu L, Liu W, et al. The predictors of death within 1 year in acute ischemic stroke patients based on machine learning[J]. Front Neurol, 2023, 14: 1092534. DOI: 10.3389/fneur.2023.1092534.
- [13] Lv J, Zhang M, Fu Y, et al. An interpretable machine learning approach for predicting 30-day readmission after stroke[J]. Int J Med Inform, 2023, 174: 105050. DOI: 10.1016/j.ijmedinf.2023.105050.
- [14] Liew SL, Schweighofer N, Cole JH, et al. Association of brain age, lesion volume, and functional outcome in patients with stroke[J]. Neurology, 2023, 100(20): e2103-e2113. DOI: 10.1212/WNL.000000000000207219.
- [15] Xu J, Zhang X, Jin A, et al. Trends and risk factors associated with stroke recurrence in China, 2007-2018 [J]. JAMA Netw Open, 2022; 5(6): e2216341. DOI: 10.1001/jamanetworkopen.2022.16341.
- [16] Turana Y, Tenglawan J, Chia YC, et al. Hypertension and stroke in Asia: a comprehensive review from HOPE Asia[J]. J Clin Hypertens (Greenwich), 2021, 23(3): 513-521. DOI: 10.1111/jch.14099.
- [17] Vermeer SE, Sandee W, Algra A, et al. Impaired glucose tolerance increases stroke risk in nondiabetic patients with transient ischemic attack or minor ischemic stroke[J]. Stroke, 2006, 37(6): 1413-1417. DOI: 10.1161/01.STR.0000221766.73692.0b.
- [18] Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: an update[J]. J Am Coll Cardiol, 2004, 43(10): 1731-1737. DOI: 10.1016/j.jacc.2003.12.047.
- [19] Rogers RL, Meyer JS, Judd BW, et al. Abstention from cigarette smoking improves cerebral perfusion among elderly chronic smokers[J]. JAMA, 1985, 253(20): 2970-2974.
- [20] Li J, Wang Y, Lin J, et al. Soluble CD40L is a useful marker to predict future strokes in patients with minor stroke and transient ischemic attack[J]. Stroke, 2015, 46(7): 1990-1992. DOI: 10.1161/STROKEAHA.115.008685.
- [21] Lin J, Zheng H, Cucchiara BL, et al. Association of Lp-PLA2-A and early recurrence of vascular events after TIA and minor stroke[J]. Neurology, 2015, 85(18): 1585-1591. DOI: 10.1212/WNL.0000000000001938.
- [22] Zhang M, Meng X, Pan Y, et al. Predictive values of baseline matrix metalloproteinase 9 levels in peripheral blood on 3-month outcomes of high-risk patients with minor stroke or transient ischemic attack[J]. Eur J Neurol, 2022, 29(10): 2976-2986. DOI: 10.1111/ene.15342.
- [23] 覃伟, 吕蕊芮, 王子尧, 等. 基于 logistic 回归和随机森林的急性缺血性卒中 3 个月预后预测模型的构建[J]. 现代预防医学, 2021, 48(2): 193-197.
- Qin W, Lyu XR, Wang ZY, et al. Construction of 3-month prognosis prediction model for acute ischemic stroke based on logistic regression and random forest[J]. Modern Preventive Medicine, 2021, 48(2): 193-197.

(收稿日期: 2023-10-11)

(本文编辑: 郑圣洁)