

• 论著 •

基于机器学习构建急性脑卒中后 1 年癫痫预测模型及其外部验证

周文考¹ 赵方俐¹ 邱星强¹ 杨玉娟² 王婷婷² 黄灵炎¹

¹ 厦门大学附属翔安医院急诊医学科, 福建厦门 361100; ² 厦门市第五医院重症医学科, 福建厦门 361100

周文考现在厦门大学附属翔安医院医务室, 福建厦门 361100

通信作者: 黄灵炎, Email: zwk15060726235@126.com

【摘要】 目的 通过机器学习寻找急性脑卒中后 1 年继发癫痫的最优预测算法, 使用该算法筛选指标建立列线图模型并进行外部验证, 以实现继发性癫痫的精准预测。**方法** 以 2019 年 6 月至 2023 年 6 月在厦门大学附属翔安医院急诊医学科就诊的 870 例急性脑卒中患者的数据进行模型建立 (模型组), 以同期在厦门市第五医院就诊的 435 例急性脑卒中患者数据进行机器学习算法和列线图的外部验证 (外部验证组)。根据脑卒中后 1 年是否发生继发性癫痫将患者分为对照组和癫痫组, 比较两组患者基线资料、脑卒中位置、脑血管情况、并发症、血液学指标和美国国立卫生研究院卒中量表 (NIHSS) 评分等。用逻辑回归、CN2 规则归纳、K-近邻、自适应提升算法、随机森林、梯度提升、支持向量机、朴素贝叶斯和神经网络等 9 种机器学习算法对数据进行深度学习, 评估各算法的性能, 并通过受试者工作特征曲线 (ROC 曲线) 的曲线下面积 (AUC) 评估最优模型算法。用逻辑回归分析影响急性脑卒中后继发癫痫的危险因素, 选取前 10 名因素建立列线图模型, 用 ROC 曲线评估该模型的预测效果并进行外部验证。**结果** 模型组 870 例急性脑卒中患者 1 年后继发癫痫 29 例。用 9 种机器学习算法对模型组数据进行分析并进行外部数据验证, 其中逻辑回归模型具有最佳效果和外延性效能, 其 AUC 为 0.923。对模型组数据进行的单因素逻辑回归分析显示, 血小板计数、白细胞计数、红细胞计数、糖化血红蛋白 (HbA1c)、C-反应蛋白 (CRP)、三酰甘油、高密度脂蛋白 (HDL)、天冬氨酸转氨酶 (AST)、丙氨酸转氨酶 (ALT)、活化部分凝血活酶时间 (APTT)、凝血酶时间、D-二聚体、纤维蛋白原、肌酸激酶 (CK)、肌酸激酶同工酶 (CK-MB)、乳酸脱氢酶 (LDH)、血清钠、乳酸、阴离子间隙、NIHSS 评分、脑疝、脑室旁卒中和颈内动脉斑块是影响急性脑卒中后 1 年继发性癫痫的危险因素; 进一步多因素逻辑回归分析显示, 白细胞计数、HDL、纤维蛋白原、乳酸和脑疝是其独立危险因素 (优势比 (OR) 分别为 1.837、198.039、47.025、11.559、70.722, 均 $P < 0.05$)。对外部验证组数据进行的单因素逻辑回归分析显示, 血小板计数、白细胞计数、CRP、三酰甘油、APTT、D-二聚体、纤维蛋白原、CK、CK-MB、LDH、NIHSS 评分和脑疝是影响急性脑卒中后 1 年继发性癫痫的危险因素; 进一步多因素逻辑回归分析显示, APTT 和脑疝是其独立危险因素 (OR 值分别为 0.587、116.193, 均 $P < 0.05$)。以脑疝、脑室旁卒中、颈内动脉斑块、白细胞计数、三酰甘油、凝血酶时间、D-二聚体、血清钠、乳酸和 NIHSS 评分等 10 项因素构建列线图模型, 其中模型组的 AUC 为 0.908, 外部验证组的 AUC 为 0.864。**结论** 通过机器学习算法构建的急性脑卒中后 1 年癫痫预测模型中, 逻辑回归算法具有最佳的预测性能, 以逻辑回归模型筛选的危险因素构建的列线图模型预测价值较高, 并得到外部验证, 具有较强的临床应用价值和外延性。

【关键词】 急性脑卒中; 癫痫; 机器学习; 预测模型; 外部验证

DOI: 10.3760/cma.j.cn121430-20241225-01069

Construction and external validation of a machine learning-based prediction model for epilepsy one year after acute stroke

Zhou Wenkao¹, Zhao Fangli¹, Qiu Xingqiang¹, Yang Yujuan², Wang Tingting², Huang Lingyan¹

¹Department of Emergency, Xiang'an Hospital of Xiamen University, Xiamen 361100, Fujian, China; ²Department of Intensive Care Unit, the Fifth Hospital of Xiamen, Xiamen 361100, Fujian, China

Zhou Wenkao is working on the Department of Medical Room, Xiang'an Hospital of Xiamen University, Xiamen 361100, Fujian, China

Corresponding author: Huang Lingyan, Email: zwk15060726235@126.com

【Abstract】 Objective To identify the optimal machine learning algorithm for predicting post-stroke epilepsy (PSE) within one year following acute stroke, establish a nomogram model based on this algorithm, and perform external validation to achieve accurate prediction of secondary epilepsy. **Methods** A total of 870 acute stroke patients admitted to the emergency department of Xiang'an Hospital of Xiamen University from June 2019 to June 2023 were enrolled for model development (model group). An external validation cohort of 435 acute stroke patients admitted to the Fifth Hospital of Xiamen during the same period was used to validate the machine learning algorithms and nomogram model. Patients were classified into control and epilepsy groups based on the development of PSE within one year. Clinical and laboratory data, including baseline characteristics, stroke location, vascular status, complications,

hematologic parameters, and National Institutes of Health Stroke Scale (NIHSS) score, were collected for analysis. Nine machine learning algorithms such as logistic regression, CN2 rule induction, K-nearest neighbors, adaptive boosting, random forest, gradient boosting, support vector machine, naive Bayes, and neural network were applied to evaluate predictive performance. The area under the curve (AUC) of receiver operator characteristic curve (ROC curve) was used to identify the optimal algorithm. Logistic regression was used to screen risk factors for PSE, and the top 10 predictors were selected to construct the nomogram model. The predictive performance of the model was evaluated using the ROC curve in both the model and validation groups. **Results** Among the 870 patients in the model group, 29 developed PSE within one year. Among the nine algorithms tested, logistic regression demonstrated the best performance and generalizability, with an AUC of 0.923. Univariate logistic regression identified several risk factors for PSE, including platelet count, white blood cell count, red blood cell count, glycated hemoglobin (HbA1c), C-reactive protein (CRP), triglycerides, high-density lipoprotein (HDL), aspartate aminotransferase (AST), alanine aminotransferase (ALT), activated partial thromboplastin time (APTT), thrombin time, D-dimer, fibrinogen, creatine kinase (CK), creatine kinase-MB (CK-MB), lactate dehydrogenase (LDH), serum sodium, lactic acid, anion gap, NIHSS score, brain herniation, periventricular stroke, and carotid artery plaque. Further multivariate logistic regression analysis showed that white blood cell count, HDL, fibrinogen, lactic acid and brain herniation were independent risk factors [odds ratio (OR) were 1.837, 198.039, 47.025, 11.559, 70.722, respectively, all $P < 0.05$]. In the external validation group, univariate logistic regression analysis showed that platelet count, white blood cell count, CRP, triacylglycerol, APTT, D-dimer, fibrinogen, CK, CK-MB, LDH, NIHSS score, and cerebral herniation were risk factors for PSE one year after acute stroke. Further multiple logistic regression analysis showed that APTT and cerebral herniation were independent predictors (OR were 0.587 and 116.193, respectively, both $P < 0.05$). The nomogram model, constructed using 10 key variables—brain herniation, periventricular stroke, carotid artery plaque, white blood cell count, triglycerides, thrombin time, D-dimer, serum sodium, lactic acid, and NIHSS score—achieved an AUC of 0.908 in the model group and 0.864 in the external validation group. **Conclusions** The logistic regression-based prediction model for epilepsy one year after acute stroke, developed using machine learning algorithms, showed optimal predictive performance. The nomogram model based on the logistic regression-derived predictors showed strong discriminative power and was successfully validated externally, suggesting favorable clinical applicability and generalizability.

【Key words】 Acute stroke; Epilepsy; Machine learning; Predictive model; External validation

DOI: 10.3760/cma.j.cn121430-20241225-01069

急性脑卒中是导致成人残疾和死亡的原因之一。全球每年约有 1 000 万人因脑卒中导致死亡,而幸存者则面临长期的残疾风险^[1-2]。继发性癫痫是脑卒中的常见并发症之一,约 10%~20% 的急性脑卒中在 1 年内发生癫痫^[3]。癫痫的发生不仅增加了患者的病死率,还显著降低了患者的生活质量,因此,早期预测和干预癫痫的发生对改善脑卒中患者的预后至关重要。随着机器学习技术的进步,基于机器学习的预测模型逐渐成为提高预测准确性的重要工具^[4]。机器学习能够从复杂的临床数据中提取隐藏的规律,实现对疾病风险的精准预测^[5]。然而,目前针对急性脑卒中后继发性癫痫的预测研究相对较少。本研究通过机器学习方法构建急性脑卒中后 1 年癫痫预测模型,利用临床和实验室数据,探索最优的预测模型。

1 资料与方法

1.1 研究对象: 回顾性纳入 2019 年 6 月至 2023 年 6 月在厦门大学附属翔安医院急诊医学科诊治的 870 例急性脑卒中患者的数据进行模型建立(模型组),根据脑卒中后 1 年是否发生继发性癫痫将患者分为对照组和癫痫组。此外,纳入同期在厦门市第五医院就诊的 435 例急性脑卒中患者的数据作为外

部验证(外部验证组)。

1.1.1 入选标准: ① 符合《中国急性脑卒中临床研究规范共识 2018》脑卒中诊断标准^[6]; ② 符合《2017 年国际抗癫痫联盟癫痫发作和癫痫新分类的简要解读》癫痫诊断标准^[7]; ③ 脑卒中为首次发作。

1.1.2 排除标准: ① 临床病例资料不完整; ② 既往存在癫痫或正使用抗癫痫药物; ③ 存在严重基础性疾病。

1.2 伦理学: 本研究经过厦门大学附属翔安医院(审批号: XAHLL2024045)以及厦门市第五医院(审批号: XMWY-K-101)伦理委员会审批同意。

1.3 样本量计算与提取: 根据样本量计算公式,对本中心数据计算,急性脑卒中后 1 年发生癫痫的概率 P 值为 4.30%, α 水准为 0.05, 此时 $z=1.96$, 允许误差为 2%, 计算样本量为 395。在样本量公式计算基础之上再增加 10%~20%, 弥补实际调查过程中失访带来的样本量损失, 按照 10% 的失访率计算, 最终样本量为 435。故模型建立提取样本量为 870, 外部验证提取样本量为 435。

1.4 观察指标: ① 基本资料: 患者年龄、性别等; ② 卒中位置: 根据颅脑磁共振等检查评估卒中所在位置, 如额叶、顶叶、颞叶、枕叶、岛内、基底神经

节、内囊、脑干、小脑、脑室旁、半卵圆区、皮质下叶、丘脑等；③ 血管状态：评估相关血管情况，包括是否发生颈总动脉斑块、颈内动脉斑块；④ 并发症：卒中后是否发生尿毒症、深静脉血栓、脑疝和脑积水，卒中前是否存在脂肪肝、糖尿病、高血压、冠心病、心房颤动、高尿酸血症、高脂血症和高蛋白血症等；⑤ 血液学指标：包括血常规和血生化等指标；⑥ 国家卫生研究院卒中量表(National Institutes of Health Stroke Scale, NIHSS)：NIHSS 评分是一种评估卒中患者神经功能缺损程度的标准化量表，包括 11 个项目，总分在 0~42 分，分数越高，代表卒中患者神经功能障碍越严重。

1.5 统计学方法：使用 SPSS 23.0 软件对数据进行统计学分析。R 4.3.1 语言进行机器学习及预测模型构建与验证。计数资料以例(率)表示，两组间比较采用 χ^2 检验， $n<5$ 用 Fisher 确切概率法。正态分布的计量数据以均数 $\pm$ 标准差($\bar{x} \pm s$)表示，采用独立样本 t 检验。使用 R 4.3.1 语言对模型数据进行逻辑回归、CN2 规则归纳、K-近邻、自适应提升算法、随机森林、梯度提升、支持向量机、朴素贝叶斯和神经网络等 9 种机器学习算法深度分析，并使用外部验证数据进行外延性分析，评估不同算法在独立数据集上的泛化能力；用受试者工作特征曲线(receiver operator characteristic curve, ROC 曲线)的曲线下面积(area under the curve, AUC)评估最优

机器学习模型。用单因素及多因素逻辑回归分别分析影响模型数据和外部验证数据中继发性癫痫的危险因素。为去除拟合不佳风险，从模型数据回归分析中以优势比(odds ratio, OR)为依据筛选前 10 名有效变量进行模型构建，以 $OR \geq 1.25$ 为截断值。使用与其对应的外部验证变量进行验证。用 "rms" 程编辑包对相关危险因素进行模型构建，"pROC" 程编辑包对模型进行外部验证。用 ROC 曲线验证列线图模型的可信度， $0.5 < AUC \leq 0.7$ 表示模型低度可信， $0.7 < AUC \leq 0.9$ 表示模型中度可信， $AUC > 0.9$ 表示模型高度可信。双侧检验，检验水准 $\alpha = 0.05$ ， $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 模型数据差异性分析(表 1~2)：870 例急性脑卒中患者中 1 年后继发癫痫 29 例。与对照组比较，癫痫组脑干、脑室旁受累的比例明显升高，颈内动脉斑块检出率明显升高(均 $P < 0.05$)，其余脑卒中位置和血管状态在两组之间差异无统计学意义(均 $P > 0.05$)。两组间并发症发生情况差异均无统计学意义(均 $P > 0.05$)。两组间血小板计数、白细胞计数、红细胞计数、糖化血红蛋白(glycosylated hemoglobin, HbA1c)、C-反应蛋白(C-reactive protein, CRP)、三酰甘油、高密度脂蛋白(high density lipoprotein, HDL)、天冬氨酸转氨酶(aspartate aminotransferase, AST)、丙氨酸转氨酶(alanine aminotransferase,

表 1 模型数据中急性脑卒中后 1 年是否发生癫痫两组患者卒中位置、血管状态和并发症情况比较								
指标	对照组 (n=841)	癫痫组 (n=29)	t/ χ^2 值	P 值	指标	对照组 (n=841)	癫痫组 (n=29)	χ^2 值 P 值
年龄(岁, $\bar{x} \pm s$)	87.78 $\pm$ 1.29	88.03 $\pm$ 1.35	1.037	0.300	前循环缺血	47(5.6)	2(6.9)	0.090 0.764
男性[例(%)]	377(44.8)	11(37.9)	0.540	0.463	后循环缺血	127(15.1)	4(13.8)	0.037 0.846
NIHSS 评分(分, $\bar{x} \pm s$)	10.57 $\pm$ 2.63	12.41 $\pm$ 2.44	3.714	<0.001	颈总动脉斑块	227(27.0)	9(31.0)	0.232 0.630
卒中位置[例(%)]					颈内动脉斑块	56(6.7)	5(17.2)	4.815 0.028
额叶	49(5.8)	4(13.8)	3.110	0.078	颈外动脉斑块	16(1.9)	2(6.9)	3.451 0.063
顶叶	37(4.4)	2(6.9)	0.408	0.523	大脑中动脉斑块	32(3.8)	2(6.9)	0.713 0.398
颞叶	33(3.9)	3(10.3)	2.914	0.088	大动脉粥样硬化	243(28.9)	10(34.5)	0.425 0.515
枕叶	20(2.4)	1(3.4)	0.136	0.712	并发症[例(%)]			
岛叶	13(1.5)	1(3.4)	0.641	0.423	卒中后尿毒症	9(1.1)	0(0)	0.314 0.575
基底神经节	52(6.2)	1(3.4)	0.367	0.545	卒中后深静脉血栓	91(10.8)	2(6.9)	0.452 0.501
内囊	1(0.1)	0(0)	0.035	0.853	卒中后脑疝	9(1.1)	0(0)	0.575 0.575
脑干	7(0.8)	2(6.9)	10.070	0.002	卒中后脑积水	7(0.8)	1(3.4)	2.106 0.147
小脑	18(2.1)	0(0)	0.634	0.426	卒中前脂肪肝	112(13.3)	5(17.2)	0.371 0.543
脑室旁	42(5.0)	4(13.8)	4.334	0.037	卒中前糖尿病	300(35.7)	10(34.5)	0.017 0.895
半卵圆区	24(2.9)	2(6.9)	1.580	0.209	卒中前高血压	643(76.5)	19(65.5)	1.844 0.174
皮质下叶	103(12.2)	7(24.1)	3.589	0.058	卒中前冠心病	573(68.1)	21(72.4)	0.237 0.626
丘脑	8(1.0)	1(3.4)	1.707	0.191	卒中前房颤	175(20.8)	10(34.5)	3.131 0.077
血管状态[例(%)]					卒中前高尿酸血症	123(14.6)	6(20.7)	0.816 0.366
椎动脉缺血	20(2.4)	0(0)	0.706	0.401	卒中前高脂血症	117(13.9)	3(10.3)	0.300 0.584
基底动脉缺血	8(1.0)	0(0)	0.278	0.598	卒中前高蛋白血症	210(25.0)	11(37.9)	2.485 0.115

注：NIHSS 为美国国立卫生研究院卒中量表

表 2 模型数据中急性脑卒中后 1 年是否发生癫痫两组患者实验室数据比较($\bar{x} \pm s$)

指标	对照组 (n=841)	癫痫组 (n=29)	t 值	P 值	指标	对照组 (n=841)	癫痫组 (n=29)	t 值	P 值
血小板计数($\times 10^9/L$)	179.11 $\pm$ 25.04	160.72 $\pm$ 31.01	3.855	<0.001	血尿酸($\mu\text{mol/L}$)	347.35 $\pm$ 81.87	325.49 $\pm$ 88.24	1.411	0.159
白细胞计数($\times 10^9/L$)	8.50 $\pm$ 1.21	10.38 $\pm$ 2.80	7.747	<0.001	凝血酶原时间(s)	13.99 $\pm$ 0.78	14.02 $\pm$ 0.51	0.207	0.836
红细胞计数($\times 10^9/L$)	3.96 $\pm$ 0.24	4.08 $\pm$ 0.19	2.693	0.007	APTT(s)	37.70 $\pm$ 3.09	36.09 $\pm$ 1.95	2.790	0.005
HbA1c(%)	6.76 $\pm$ 0.94	6.29 $\pm$ 0.71	2.650	0.008	凝血酶时间(s)	16.16 $\pm$ 0.56	16.71 $\pm$ 0.42	5.264	<0.001
CRP(mg/L)	23.53 $\pm$ 20.66	59.72 $\pm$ 53.28	5.827	<0.001	INR(%)	1.08 $\pm$ 0.07	1.09 $\pm$ 0.05	0.787	0.432
三酰甘油(mmol/L)	1.28 $\pm$ 0.40	1.44 $\pm$ 0.31	2.129	0.034	D-二聚体(mg/L)	1.76 $\pm$ 1.11	4.93 $\pm$ 4.76	12.074	<0.001
LDL(mmol/L)	2.42 $\pm$ 0.37	2.43 $\pm$ 0.24	0.160	0.873	纤维蛋白原(g/L)	3.86 $\pm$ 0.40	3.60 $\pm$ 0.67	3.535	<0.001
HDL(mmol/L)	1.25 $\pm$ 0.15	1.37 $\pm$ 0.24	3.900	<0.001	CK(U/L)	183.89 $\pm$ 196.21	990.32 $\pm$ 803.17	17.720	<0.001
AST(mmol/L)	25.75 $\pm$ 9.11	38.33 $\pm$ 29.17	6.418	<0.001	CK-MB(U/L)	14.17 $\pm$ 7.02	30.58 $\pm$ 21.92	10.930	<0.001
ALT(mmol/L)	20.36 $\pm$ 9.07	24.64 $\pm$ 7.96	2.511	0.012	LDH(U/L)	220.81 $\pm$ 48.29	319.84 $\pm$ 123.79	9.998	<0.001
TBIL($\mu\text{mol/L}$)	15.63 $\pm$ 4.25	16.72 $\pm$ 2.63	1.366	0.172	血清钠(mmol/L)	137.45 $\pm$ 1.65	138.22 $\pm$ 1.46	2.481	0.013
白蛋白(g/L)	38.28 $\pm$ 2.28	38.92 $\pm$ 2.26	1.490	0.137	血清钾(mmol/L)	3.89 $\pm$ 0.15	3.84 $\pm$ 0.16	1.609	0.108
血尿素氮(mmol/L)	7.21 $\pm$ 1.30	6.82 $\pm$ 1.43	1.593	0.111	乳酸(mmol/L)	2.33 $\pm$ 0.36	2.74 $\pm$ 0.46	5.975	<0.001
肌酐($\mu\text{mol/L}$)	114.70 $\pm$ 93.41	89.72 $\pm$ 26.49	1.437	0.151	阴离子间隙(mmol/L)	11.87 $\pm$ 1.42	12.53 $\pm$ 1.48	2.468	0.014

注: HbA1c 为糖化血红蛋白, CRP 为 C-反应蛋白, LDL 为低密度脂蛋白, HDL 为高密度脂蛋白, AST 为天冬氨酸转氨酶, ALT 为丙氨酸转氨酶, TBIL 为总胆红素, APTT 为活化部分凝血酶原时间, INR 为国际标准化比值, CK 为肌酸激酶, CK-MB 为肌酸激酶同工酶, LDH 为乳酸脱氢酶

表 3 外部验证数据中急性脑卒中后 1 年是否发生癫痫两组患者卒中位置、血管状态和并发症情况比较

指标	对照组 (n=423)	癫痫组 (n=12)	t/ χ^2 值	P 值	指标	对照组 (n=423)	癫痫组 (n=12)	χ^2 值	P 值
年龄(岁, $\bar{x} \pm s$)	85.23 $\pm$ 0.47	85.00 $\pm$ 0.43	1.705	0.089	前循环缺血	24(5.7)	0(0)	0.721	0.396
男性[例(%)]	198(46.8)	4(33.3)	0.852	0.356	后循环缺血	75(17.7)	1(8.3)	0.715	0.398
NIHSS 评分(分, $\bar{x} \pm s$)	10.45 $\pm$ 3.28	12.75 $\pm$ 2.14	2.417	0.016	颈总动脉斑块	117(27.7)	4(33.3)	0.187	0.665
卒中位置[例(%)]					颈内动脉斑块	41(9.7)	1(8.3)	0.025	0.875
额叶	26(6.1)	1(8.3)	0.096	0.757	颈外动脉斑块	8(1.9)	0(0)	0.231	0.631
顶叶	22(5.2)	1(8.3)	0.229	0.633	大脑中动脉斑块	12(2.8)	0(0)	0.350	0.554
颞叶	25(5.9)	0(0)	0.752	0.386	大动脉粥样硬化	129(30.5)	5(41.7)	0.683	0.409
枕叶	20(2.4)	1(3.4)	0.320	0.572	并发症[例(%)]				
岛叶	12(2.8)	0(0)	0.350	0.554	卒中后尿毒症	4(0.9)	0(0)	0.115	0.735
基底神经节	24(5.7)	0(0)	0.721	0.396	卒中后深静脉血栓	60(14.2)	3(25.0)	1.102	0.294
内囊	0(0)	0(0)	Fisher	1.000	卒中后脑疝	2(0.5)	1(8.3)	10.527	0.001
脑干	5(1.2)	0(0)	0.143	0.705	卒中后脑积水	6(1.4)	0(0)	0.173	0.678
小脑	9(2.1)	0(0)	0.261	0.610	卒中前脂肪肝	49(11.6)	3(25.0)	1.996	0.158
脑室旁	25(5.9)	0(0)	0.752	0.386	卒中前糖尿病	162(38.3)	3(25.0)	0.876	0.349
半卵圆区	14(3.3)	0(0)	0.410	0.522	卒中前高血压	338(79.9)	7(58.3)	3.309	0.069
皮质下叶	50(11.8)	0(0)	1.603	0.206	卒中前冠心病	261(61.7)	7(58.3)	0.056	0.813
丘脑	4(0.9)	0(0)	0.115	0.735	卒中前房颤	83(19.6)	1(8.3)	0.954	0.329
血管状态[例(%)]					卒中前高尿酸血症	61(14.4)	2(16.7)	0.048	0.827
椎动脉缺血	13(3.1)	1(8.3)	1.037	0.309	卒中前高脂血症	58(13.7)	2(16.7)	0.086	0.770
基底动脉缺血	5(1.2)	0(0)	0.143	0.705	卒中前高蛋白血症	108(25.5)	3(25.0)	0.002	0.967

注: NIHSS 为美国国立卫生研究院卒中量表

ALT)、活化部分凝血活酶时间(activated partial thromboplastin time, APTT)、凝血酶时间、D-二聚体、纤维蛋白原、肌酸激酶(creatine kinase, CK)、肌酸激酶同工酶(MB isoenzyme of creatine kinase, CK-MB)、乳酸脱氢酶(lactate dehydrogenase, LDH)、血清钠、乳酸、阴离子间隙等血液学指标及 NIHSS 评分差异有统计学意义(均 $P < 0.05$)。

2.2 外部验证数据差异性分析(表 3~4): 435 例急性脑卒中患者中 1 年后继发癫痫 12 例。两组卒中

位置和血管状态差异均无统计学意义(均 $P > 0.05$)。两组间卒中后脑疝发生率、血小板计数、白细胞计数、CRP、三酰甘油、AST、APTT、凝血酶时间、D-二聚体、纤维蛋白原、CK、CK-MB、LDH 及 NIHSS 评分差异均有统计学意义(均 $P < 0.05$)。

2.3 最优算法选择(表 5): 对模型数据进行 9 种机器学习算法分析,并进行外部数据验证,其中逻辑回归算法具有最佳效果和外延性效能,其 AUC 为 0.923,其次分别为神经网络、梯度提升、随机森林、

表 4 外部验证数据中急性脑卒中后 1 年是否发生癫痫两组患者实验室数据比较 ($\bar{x} \pm s$)								
指标	对照组 (n=423)	癫痫组 (n=12)	t 值	P 值	指标	对照组 (n=423)	癫痫组 (n=12)	t 值 P 值
血小板计数($\times 10^9/L$)	182.78 $\pm$ 22.34	163.53 $\pm$ 23.29	2.939	0.003	尿酸($\mu\text{mol/L}$)	339.56 $\pm$ 70.43	320.04 $\pm$ 65.43	0.948 0.344
白细胞计数($\times 10^9/L$)	8.65 $\pm$ 1.13	10.82 $\pm$ 2.11	6.354	<0.001	凝血酶原时间(s)	14.06 $\pm$ 1.23	14.08 $\pm$ 0.47	0.035 0.972
红细胞计数($\times 10^9/L$)	4.01 $\pm$ 0.25	4.09 $\pm$ 0.13	1.151	0.250	APTT(s)	38.13 $\pm$ 2.65	36.08 $\pm$ 1.58	2.675 0.008
HbA1c(%)	6.76 $\pm$ 1.00	6.26 $\pm$ 0.67	1.732	0.084	凝血酶时间(s)	16.11 $\pm$ 0.57	16.70 $\pm$ 0.44	3.541 <0.001
CRP(mg/L)	28.65 $\pm$ 23.49	69.83 $\pm$ 54.17	5.685	<0.001	INR(%)	1.08 $\pm$ 0.08	1.08 $\pm$ 0.05	0.379 0.705
三酰甘油(mmol/L)	1.26 $\pm$ 0.30	1.50 $\pm$ 0.26	2.675	0.008	D-二聚体(mg/L)	1.66 $\pm$ 0.98	5.41 $\pm$ 4.92	10.297 <0.001
LDL(mmol/L)	2.48 $\pm$ 0.41	2.58 $\pm$ 0.22	0.832	0.406	纤维蛋白原(g/L)	3.95 $\pm$ 0.40	3.58 $\pm$ 0.62	3.189 0.002
HDL(mmol/L)	1.28 $\pm$ 0.17	1.31 $\pm$ 0.12	0.510	0.610	CK(U/L)	204.61 $\pm$ 205.58	914.38 $\pm$ 851.67	9.930 <0.001
AST(mmol/L)	25.90 $\pm$ 14.43	34.40 $\pm$ 8.66	2.027	0.043	CK-MB(U/L)	14.83 $\pm$ 10.33	27.21 $\pm$ 17.28	4.003 <0.001
ALT(mmol/L)	21.08 $\pm$ 10.76	25.29 $\pm$ 8.49	1.345	0.179	LDH(U/L)	223.06 $\pm$ 52.12	311.32 $\pm$ 120.24	5.491 <0.001
TBIL($\mu\text{mol/L}$)	15.38 $\pm$ 3.53	17.10 $\pm$ 2.72	1.675	0.095	血清钠(mmol/L)	137.45 $\pm$ 2.01	138.34 $\pm$ 0.82	1.532 0.126
白蛋白(g/L)	38.08 $\pm$ 2.10	39.24 $\pm$ 2.21	1.887	0.060	血清钾(mmol/L)	3.83 $\pm$ 0.16	3.75 $\pm$ 0.08	1.863 0.063
血尿素氮(mmol/L)	7.03 $\pm$ 1.42	6.60 $\pm$ 1.17	1.037	0.301	乳酸(mmol/L)	2.50 $\pm$ 0.42	2.70 $\pm$ 0.19	1.626 0.105
血肌酐($\mu\text{mol/L}$)	143.96 $\pm$ 164.35	101.44 $\pm$ 55.00	0.894	0.372	阴离子间隙(mmol/L)	11.97 $\pm$ 1.44	12.47 $\pm$ 0.89	1.194 0.233

注:HbA1c 为糖化血红蛋白,CRP 为 C-反应蛋白,LDL 为低密度脂蛋白,HDL 为高密度脂蛋白,AST 为天冬氨酸转氨酶,ALT 为丙氨酸转氨酶,TBIL 为总胆红素,APTT 为活化部分凝血酶原时间,INR 为国际标准化比值,CK 为肌酸激酶,CK-MB 为肌酸激酶同工酶,LDH 为乳酸脱氢酶

朴素贝叶斯、支持向量机、K-近邻、CN2 规则归纳、自适应提升算法。

表 5 9 种深度机器学习模型对急性脑卒中患者 1 年后继发性癫痫的预测价值						
模型	AUC	准确率	F1 分数	精确率	召回率	Matthews 相关系数
逻辑回归	0.923	0.977	0.977	0.977	0.977	0.572
CN2 规则归纳	0.743	0.956	0.957	0.958	0.956	0.218
K-近邻	0.797	0.972	0.971	0.970	0.972	0.442
自适应提升算法	0.694	0.956	0.960	0.964	0.956	0.328
随机森林	0.846	0.982	0.980	0.979	0.982	0.604
梯度提升	0.888	0.972	0.973	0.975	0.972	0.526
支持向量机	0.869	0.984	0.981	0.984	0.984	0.640
朴素贝叶斯	0.872	0.841	0.892	0.965	0.841	0.225
神经网络	0.903	0.986	0.986	0.985	0.986	0.723

注:AUC 为曲线下面积

2.4 模型数据逻辑回归分析(表 6):单因素逻辑回归分析显示,血小板计数、白细胞计数、红细胞计数、HbA1c、CRP、三酰甘油、HDL、AST、ALT、APTT、凝血酶时间、D-二聚体、纤维蛋白原、CK、CK-MB、LDH、血清钠、乳酸、阴离子间隙、NIHSS 评分、脑疝、脑室旁卒中和颈内动脉斑块是影响急性脑卒中后 1 年继发性癫痫的危险因素(均 $P<0.05$);进一步多因素逻辑回归分析显示,白细胞计数、HDL、纤维蛋白原、乳酸和脑疝是影响急性脑卒中后 1 年继发性癫痫的独立危险因素(均 $P<0.05$)。

2.5 外部验证数据逻辑回归分析(表 7):单因素逻辑回归分析显示,血小板计数、白细胞计数、CRP、三酰甘油、APTT、D-二聚体、纤维蛋白原、CK、CK-MB、LDH、NIHSS 评分和脑疝是影响急性脑卒中后

表 6 模型数据急性脑卒中患者 1 年后继发性癫痫危险因素的逻辑回归分析				
变量	单因素分析		多因素分析	
	OR(95%CI)	P 值	OR(95%CI)	P 值
血小板计数	0.967(0.952~0.983)	<0.001	0.964(0.928~1.001)	0.058
白细胞计数	2.246(1.760~2.866)	<0.001	1.837(1.163~2.904)	0.009
红细胞计数	10.566(1.901~58.730)	0.007	0.276(0.008~10.038)	0.482
HbA1c	0.420(0.221~0.796)	0.008	0.348(0.076~1.592)	0.174
CRP	1.031(1.021~1.040)	<0.001	1.010(0.947~1.076)	0.767
三酰甘油	2.293(1.059~4.963)	0.035	2.501(0.313~19.956)	0.387
HDL	29.335(4.835~177.968)	<0.001	198.039(1.732~22 638.827)	0.029
AST	1.035(1.017~1.053)	<0.001	1.144(0.964~1.357)	0.125
ALT	1.034(1.006~1.063)	0.016	1.024(0.885~1.184)	0.750
APTT	0.806(0.698~0.932)	0.004	0.855(0.639~1.144)	0.291
凝血酶时间	3.388(1.942~5.911)	<0.001	4.536(0.899~22.882)	0.067
D-二聚体	1.593(1.392~1.822)	<0.001	1.651(0.629~4.332)	0.309
纤维蛋白原	0.215(0.090~0.519)	<0.001	47.025(3.398~650.696)	0.004
CK	1.003(1.002~1.004)	<0.001	1.000(0.997~1.004)	0.847
CK-MB	1.072(1.046~1.100)	<0.001	0.946(0.751~1.191)	0.635
LDH	1.013(1.009~1.017)	<0.001	1.015(0.978~1.053)	0.439
血清钠	1.317(1.061~1.634)	0.012	1.419(0.848~2.375)	0.183
乳酸	9.007(3.860~21.018)	<0.001	11.559(2.653~50.362)	0.001
阴离子间隙	1.316(1.060~1.633)	0.013	0.761(0.385~1.504)	0.432
NIHSS 评分	1.250(1.109~1.409)	<0.001	1.162(0.886~1.525)	0.277
脑疝	8.825(1.751~44.487)	0.008	70.722(5.099~980.985)	0.002
脑室旁卒中	3.044(1.013~9.146)	0.047	7.562(0.629~90.849)	0.111
颈内动脉斑块	2.920(1.073~7.946)	0.036	3.146(0.480~20.623)	0.232

注:HbA1c 为糖化血红蛋白,CRP 为 C-反应蛋白,HDL 为高密度脂蛋白,AST 为天冬氨酸转氨酶,ALT 为丙氨酸转氨酶,APTT 为活化部分凝血酶原时间,CK 为肌酸激酶,CK-MB 为肌酸激酶同工酶,LDH 为乳酸脱氢酶,NIHSS 为美国国立卫生研究院卒中量表,OR 为优势比,95%CI 为 95% 可信区间

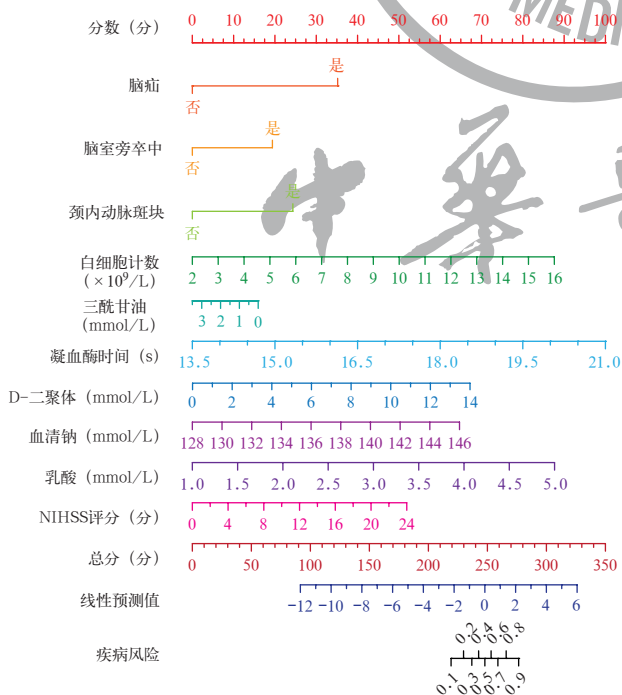
1 年继发性癫痫的危险因素(均 $P<0.05$);进一步多因素逻辑回归分析显示,APTT 和脑疝是其独立危险因素(均 $P<0.05$)。

表 7 外部验证数据急性脑卒中患者 1 年后继发性癫痫危险因素的逻辑回归分析

变量	单因素分析		多因素分析	
	OR (95%CI)	P 值	OR (95%CI)	P 值
血小板计数	0.962 (0.938 ~ 0.986)	0.002	0.979 (0.929 ~ 1.032)	0.438
白细胞计数	2.548 (1.742 ~ 3.728)	<0.001	1.572 (0.709 ~ 3.488)	0.266
CRP	1.028 (1.014 ~ 1.042)	<0.001	1.052 (0.961 ~ 1.152)	0.269
三酰甘油	5.959 (1.527 ~ 23.264)	0.010	2.229 (0.167 ~ 29.756)	0.544
AST	1.012 (0.996 ~ 1.027)	0.134		
APTT	0.686 (0.527 ~ 0.892)	0.005	0.587 (0.361 ~ 0.954)	0.032
D-二聚体	1.783 (1.394 ~ 2.282)	<0.001	1.079 (0.334 ~ 3.487)	0.900
纤维蛋白原	0.086 (0.021 ~ 0.360)	<0.001	0.226 (0.003 ~ 18.749)	0.509
CK	1.003 (1.002 ~ 1.004)	<0.001	1.004 (0.999 ~ 1.009)	0.127
CK-MB	1.030 (1.007 ~ 1.052)	0.009	0.759 (0.478 ~ 1.206)	0.243
LDH	1.010 (1.003 ~ 1.016)	0.004	1.015 (0.980 ~ 1.053)	0.404
NIHSS 评分	1.130 (1.013 ~ 1.262)	0.029	0.942 (0.735 ~ 1.208)	0.640
脑疝	19.136 (1.612 ~ 227.149)	0.019	116.193 (5.025 ~ 2 686.676)	0.003

注: CRP 为 C-反应蛋白, AST 为天冬氨酸转氨酶, APTT 为活化部分凝血酶原时间, CK 为肌酸激酶, CK-MB 为肌酸激酶同工酶, LDH 为乳酸脱氢酶, NIHSS 为美国国立卫生研究院卒中量表, OR 为优势比, 95%CI 为 95% 可信区间; 空白代表无此项

2.6 列线图模型建立与外部验证: 在模型数据中, 有 23 种影响卒中继发性癫痫的危险因素, 纳入 $OR \geq 1.250$ 的危险因素, 排除红细胞计数 (OR 值过大)、HDL (OR 值过大) 和阴离子间隙 (敏感度不高且应用不频繁) 后, 纳入脑疝、脑室旁卒中、颈内动脉斑块、白细胞计数、三酰甘油、凝血酶时间、D-二聚体、血清钠、乳酸和 NIHSS 评分等 10 项因素构建列线图模型, 见图 1。

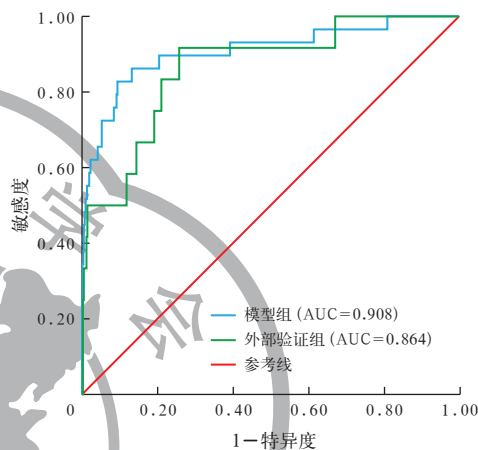


注: NIHSS 为美国国立卫生研究院卒中量表

图 1 预测急性脑卒中患者 1 年后继发性癫痫的列线图模型

ROC 曲线分析显示 (图 2), 在模型数据中, 列线图模型预测急性脑卒中后 1 年继发性癫痫的敏感度为 82.8%, 特异度为 90.6%, 准确度为 0.903; AUC 为 0.908, 95% 可信区间为 0.838 ~ 0.979。

在外部验证数据中, 列线图模型预测的敏感度为 58.3%, 特异度为 88.2%, 准确度为 0.874; AUC 为 0.864, 95% 可信区间为 0.753 ~ 0.975, 提示该预测模型具有较高可信价值。



注: ROC 曲线为受试者工作特征曲线, AUC 为曲线下面积

图 2 在模型组和外部验证组数据中列线图模型预测急性脑卒中患者 1 年后继发性癫痫的 ROC 曲线

3 讨论

本研究中通过机器学习算法构建了急性脑卒中后 1 年继发性癫痫预测模型, 并进行了外部验证。结果显示, 在 9 种机器学习模型中, 逻辑回归模型的 AUC 为 0.923, 表现出最佳的诊断性能, 优于其他机器学习算法。ROC 曲线显示, 以逻辑回归分析筛选出的危险因素建立的列线图模型的 AUC 为 0.908 (0.838 ~ 0.979), 经外部验证的 AUC 为 0.864 (0.753 ~ 0.975), 表明该模型在不同数据集上均能提供较高的预测准确性。逻辑回归模型凭借其较高的诊断性能, 进一步验证了其作为急性脑卒中后癫痫预测工具的可靠性与临床应用价值。

急性脑卒中后癫痫的发生机制复杂。脑卒中的局部损伤, 特别是脑皮质或深部脑结构的损伤, 可能导致神经元的兴奋性改变, 引起癫痫发作^[8-9]。脑卒中后出现的脑水肿、血脑屏障破坏和神经炎症反应也被认为是癫痫发生的潜在机制。脑疝、脑室旁卒中以及颈内动脉斑块形成等因素, 可能加剧脑部损伤, 进一步增加癫痫的发生风险^[10-11]。

与传统的统计方法相比, 机器学习能够处理大规模、复杂的临床数据, 并从中挖掘潜在的规律。

在本研究中,逻辑回归模型由于其良好的拟合度和高效的计算性能,在预测急性脑卒中后癫痫的发生中展示出最优的表现能力。通过深度学习,逻辑回归模型表现优异,而且在外部验证组中也能保持较高的 AUC 值,进一步证明其在癫痫预测中的可靠性和有效性。因此,逻辑回归作为最优模型的优势,提供了一个精确且具临床应用前景的预测工具。

通过逻辑回归模型筛选出 10 个影响急性脑卒中后癫痫发生的重要因素,包括脑疝、脑室旁卒中、颈内动脉斑块、白细胞计数、三酰甘油、凝血酶时间、D-二聚体、血清钠、乳酸和 NIHSS 评分。这些因素被认为是与卒中后癫痫发生密切相关的临床指标。脑疝^[12]和脑室旁卒中^[13]代表更严重的脑损伤,可增加癫痫的发生风险。颈内动脉斑块可导致脑血流不畅,进一步增加癫痫的发生概率^[14]。白细胞计数、三酰甘油、D-二聚体和乳酸等炎症及代谢指标的升高提示系统性炎症反应或代谢异常,这些均可促使神经元兴奋性增加,进而引起癫痫发作^[15]。血清钠^[16]和 NIHSS 评分^[17]则是反映脑卒中中严重程度的重要生理指标,也对癫痫的预测具有重要意义。凝血酶时间延长提示可能存在的凝血功能异常,影响脑卒中后神经系统的稳定性,增加出血风险及炎症反应,从而诱发或加重脑卒中后癫痫的发生与发展^[18]。

为了进一步提高预测的便捷性和临床适用性,本研究使用列线图模型对急性脑卒中后 1 年继发性癫痫发生风险进行可视化表达,医生可以依据患者的具体指标,如脑疝、血清钠等,快速判断其癫痫风险等级,并采取针对性的干预措施。此外,列线图模型在临床操作中具有较高的可操作性和实用性,有助于提高预测的准确性和对患者管理的效率。

尽管本研究取得了较好的预测效果,但仍存在一些不足之处。首先,数据来源仅限于两家医院,样本量有限,可能影响模型的泛化能力。其次,本研究主要依赖于静态临床数据,未考虑指标动态变化对癫痫预测的影响。未来的研究可以扩大样本量,收集更多来自不同地区、不同医院的患者数据,以增强模型的普适性。

综上,本研究表明脑疝、脑室旁卒中、颈内动脉斑块、白细胞计数、三酰甘油、凝血酶时间、D-二聚体、血清钠、乳酸和 NIHSS 评分是急性脑卒中后 1 年继发性癫痫的主要危险因素。基于机器学习及上述危险因素构建的急性脑卒中后 1 年继发性癫痫

可视化预测模型具有较高的预测性能,且得到了外部验证,便于临床快速评估癫痫的发生风险。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Ebbesen MQB, Dreier JW, Andersen G, et al. Stroke and risk of epilepsy: a danish nationwide register-based study [J]. *Stroke*, 2024, 55 (4): 972–982. DOI: 10.1161/STROKEAHA.123.044203.
- [2] 吴怀标, 张昊, 郭成军. 急性脑出血患者并发脑源性多器官功能障碍综合征风险预测模型的建立 [J]. *中华危重病急救医学*, 2024, 36 (9): 966–971. DOI: 10.3760/ema.j.cn121430-20240202-00109.
- [3] Nakken KO, Sætre E, Hlauschek G, et al. Post-stroke epilepsy [J]. *Tidsskr Nor Laegeforen*, 2024, 144 (9). DOI: 10.4045/tidsskr.24.0103.
- [4] Freiman S, Hauser WA, Rider F, et al. Post-stroke epilepsy: from clinical predictors to possible mechanisms [J]. *Epilepsy Res*, 2024, 199: 107282. DOI: 10.1016/j.epilepsyres.2023.107282.
- [5] Liu JX, He HY, Wang YLX, et al. Predictive models for secondary epilepsy in patients with acute ischemic stroke within one year [J]. *Elife*, 2024, 13: RP98759. DOI: 10.7554/eLife.98759.
- [6] 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. 中国急性脑卒中临床研究规范共识 2018 [J]. *中华神经科杂志*, 2018, 51 (4): 247–255. DOI: 10.3760/ema.j.issn.1006-7876.2018.04.003.
- [7] 丛璐璐, 赵宗茂. 2017 年国际抗癫痫联盟癫痫发作和癫痫新分类的简要解读 [J]. *河北医科大学学报*, 2018, 39 (9): 993–995, 1001. DOI: 10.3969/j.issn.1007-3205.2018.09.001.
- [8] Jiang ZZ, Sun YN, Wang ZY, et al. Causal relations between ischemic stroke and epilepsy: a bidirectional Mendelian randomization study [J]. *Heliyon*, 2024, 10 (11): e32532. DOI: 10.1016/j.heliyon.2024.e32532.
- [9] Shah SP, Heiss JD. Artificial intelligence as a complementary tool for clinical decision-making in stroke and epilepsy [J]. *Brain Sci*, 2024, 14 (3): 228. DOI: 10.3390/brainsci14030228.
- [10] Koeppe M, Trinka E, Löscher W, et al. Prevention of epileptogenesis: are we there yet? [J/OL]. *Curr Opin Neurol*, 2024 [2024-12-20]. <https://pubmed.ncbi.nlm.nih.gov/38345421/>. [published online ahead of print February 13, 2024]. DOI: 10.1097/WCO.0000000000001256.
- [11] Wang YZ, Zuo HZ, Li W, et al. Cerebral small vessel disease increases risk for epilepsy: a Mendelian randomization study [J]. *Neurol Sci*, 2024, 45 (5): 2171–2180. DOI: 10.1007/s10072-023-07221-w.
- [12] Lin R, Yu YY, Wang Y, et al. Risk of post-stroke epilepsy following stroke-associated acute symptomatic seizures [J]. *Front Aging Neurosci*, 2021, 13: 707732. DOI: 10.3389/fnagi.2021.707732.
- [13] Abraira L, Santamarina E, Cazorla S, et al. Blood biomarkers predictive of epilepsy after an acute stroke event [J]. *Epilepsia*, 2020, 61 (10): 2244–2253. DOI: 10.1111/epi.16648.
- [14] Phan J, Ramos M, Soares T, et al. Poststroke seizure and epilepsy: a review of incidence, risk factors, diagnosis, pathophysiology, and pharmacological therapies [J]. *Oxid Med Cell Longev*, 2022, 2022: 7692215. DOI: 10.1155/2022/7692215.
- [15] Abraira L, López-Maza S, Quintana M, et al. Exploratory study of blood biomarkers in patients with post-stroke epilepsy [J]. *Eur Stroke J*, 2024, 9 (3): 763–771. DOI: 10.1177/23969873241244584.
- [16] Freiman S, Hauser WA, Rider F, et al. Post-stroke seizures, epilepsy, and mortality in a prospective hospital-based study [J]. *Front Neurol*, 2023, 14: 1273270. DOI: 10.3389/fneur.2023.1273270.
- [17] Alet M, Márquez F, Marone A, et al. Predictive factors for the development of epilepsy after ischemic stroke [J]. *J Stroke Cerebrovasc Dis*, 2022, 31 (12): 106858. DOI: 10.1016/j.jstrokecerebrovasdis.2022.106858.
- [18] Pacione D, Blei F, Devinsky O, et al. Coagulation abnormalities in children undergoing epilepsy surgery [J]. *J Neurosurg Pediatr*, 2011, 7 (6): 654–659. DOI: 10.3171/2011.3.

(收稿日期: 2024-12-25)

(责任编辑: 保健媛 李银平)