

基于影像组学构建 2 型糖尿病脑梗死后 再发脑梗死的风险预测模型

孙欣^{1,2} 魏东升¹ 许允发² 刘婧¹ 安杨¹ 张兰^{3△}

【摘要】目的 使用机器学习及影像组学构建 2 型糖尿病脑梗死后再次发脑梗死的风险预测模型。**方法** 采用分层随机抽样方法回顾性分析 2020 年 1 月至 2023 年 5 月沈阳市第二中医医院中风病区收治的 2 型糖尿病合并急性脑梗死患者 90 例。按 7:3 的比例划分 90 例患者为训练组(63 例,3 年内未再发脑梗死 26 例,3 年内再发脑梗死 37 例)与测试组(27 例,3 年内未再发脑梗死 17 例,3 年内再发脑梗死 10 例),使用 3D-slicer 软件提取脑梗死区域图像影像组学特征。使用主成分分析可视化初次梗死及再梗死影像组学特征分布情况。采用最小绝对收缩和选择算子回归分析及支持向量机递减特征消除进行特征筛选,使用 logistic 多因素回归构建预测模型,在训练组与测试组中应用受试者操作特征(ROC)曲线对模型进行验证,评价影像组学特征鉴别糖尿病脑梗死与再发脑梗死的效能。最后使用 Spearman 相关系数进行鉴别特征与临床理化数据的相关分析。**结果** 2 型糖尿病患者初发与再发两组间舒张压、空腹血糖、甘油三酯、胆固醇及糖化血红蛋白具有统计学差异($P<0.05$)。3D-slicer 软件共提取出 837 个影像组学特征。主成分分析中 PC1 和 PC2 共解释了 87.1% 的总体变异,两组间存在较为明显的分离趋势。LASSO 模型筛选出 10 个特征,SVM-RFE 筛选出 5 个特征,两种机器学习共有 5 个交集影像组学特征,使用 5 个影像组学特征用于构建 2 型糖尿病脑梗死后再次发脑梗死的风险预测模型。训练组中预测模型的 ROC 曲线下面积(AUC)为 0.642,测试组中 AUC 为 0.639。Spearman 相关性分析发现,风险特征与糖化血红蛋白、舒张压、甘油三酯、胆固醇、血糖具有显著相关性($r_s>0.5,P<0.05$)。**结论** 基于机器学习及影像组学构建 2 型糖尿病脑梗死后再次发脑梗死的风险预测模型具有良好的鉴别效能。

【关键词】 影像组学 机器学习 糖尿病脑梗死 再发脑梗死

【中图分类号】 R195.1

【文献标识码】 A

DOI 10.11783/j.issn.1002-3674.2024.02.023

糖尿病属于慢性疾病,患病率逐年升高,从 1980 年流行病学调查的 0.67% 到 2017 年已增加到 11.2%^[1]。糖尿病患者可合并严重的心脑血管疾病,其中合并急性脑梗死是致残致死率较高的并发症之一,因此关于糖尿病合并脑梗死的预防、治疗及预后备受关注,也是科学研究难点与热点之一。糖尿病高血糖可促进过度氧化、凋亡等发生,促发炎症级联反应,血管内皮损伤严重,血小板黏附、聚集,形成血栓,闭塞血管。既往研究表明,糖尿病是急性脑梗死以及再发脑梗死的危险因素^[2-4],而且患脑梗死次数与预后相关,多次发病患者大多遗留严重的后遗症。近年来机器学习是一类经典的预测模型,已在慢性病发病风险识别以及危险因素研究中得到了广泛应用^[5]。本文旨在通过影像组学及机器学习方法总结出糖尿病患者脑梗死再发特征以指导临床更好地诊治预防糖尿病患者复发脑梗死。

头部 CT 检查作为脑梗死诊断必不可少的检查方法之一,不仅可以直观地显示脑梗死病变部位,同时也是神经系统鉴别诊断的重要检查,因此在临床广泛应用。影像组学是利用 CT、CTA、MRI、PET 等检查图像,利用高通量的方法提取并分析影像学的特征,之后再基于图像的分割、提取筛选的特征及构建预测模型

等机器学习算法来实现疾病的诊断、评估、预测等临床应用^[6]。目前基于头部 CT 影像组学预测糖尿病脑梗死再发梗死的研究较少。本研究通过使用机器学习及影像组学分析,提取出糖尿病脑梗死患者初次头部 CT 图像高通量信息、提取筛选特征并构建模型,探讨影像组学模型在预测糖尿病脑梗死再发脑梗死风险的价值。

随着人工智能(artificial intelligence, AI)的不断发展,在多个领域等的应用,机器学习(machine learning, ML)是人工智能的一种新形式,结合影像组学应用在临床的多个方面,两者结合构建的预测模型具有较高的特异性,对临床工作有一定的参考价值。本研究将利用机器学习基于影像组学特征构建糖尿病脑梗死后再次发脑梗死的风险预测模型,为临床提供理论支撑。

资料与方法

1. 一般资料

采用分层随机抽样方法回顾性收集 2020 年 1 月至 2023 年 5 月沈阳市第二中医医院中风病区收治的 2 型糖尿病合并急性脑梗死患者病例,将患者按年龄分为 3 层,50~60 岁,60~70 岁,>70 岁,在每层中随机选取 30 例患者。按 7:3 的比例划分 90 例患者为训练组(63 例,3 年内未再发脑梗死 26 例,3 年内再发脑梗死 37 例)与测试组(27 例,3 年内未再发脑梗死 17 例,3 年内再发脑梗死 10 例)。

2 型糖尿病诊断、急性脑梗死诊断符合《中国 2 型糖尿病防治指南(2020 年版)》^[1]《中国急性缺血性脑

1. 辽宁中医药大学研究生院(110847)

2. 沈阳市第二中医医院内分泌科、影像科

3. 辽宁中医药大学附属医院专家门诊

△通信作者:张兰, E-mail: zhanglanliujing@163.com

卒中诊治指南 2018》及《中国脑血管病影像应用指南 2019》^[7];并排除有脑外伤及手术史、急性脑出血、1 型糖尿病、严重心肝肾功能异常、肿瘤和自身免疫性疾病、脑血管病支架植入治疗的患者及正在妊娠者。

2. 脑梗死区域影像学数据提取及预处理

通过 3D-slicer 勾画出脑梗死区域,CT 值在 20 ~ 30 HU 之间的体素被定义为梗死区域^[2-3]。使用 3D-slicer 软件分析提取出糖尿病脑梗死患者影像组学特征,特征类型包括:一阶特征、灰度共现矩阵(gray level co-occurrence matrix, GLCM)、灰度共生矩阵(gray level dependence matrix, GLDM)、灰度大小区域矩阵(gray level size zone matrix, GLSZM)和灰度游程矩阵(gray level run length matrix, GLRLM),以及邻域灰度差矩阵(neighbouring gray tone difference matrix, NGTDM)。

3. 两组影像特征组成成分分析

使用“factoextra”包,进行主成分分析(principal component analysis, PCA),观察 837 个影像特征在糖尿病脑梗死与再发脑梗死组间分布是否具有差异性。

4. 机器学习筛选候选指标

使用最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)及支持向量机递归特征消除(SVM-RFE)十折交叉验证法进行筛选。取 LASSO 及 SVM-RFE 结果中交集特征用于训练集中糖尿病脑梗死后再梗风险预测模型列线图,并使用 ROC 曲线下面积(area under curve, AUC),评估鉴别模型可信度,最后在测试集中评估训练集中 AUC 值可信度。

5. 预测指标相关性分析

为了明确鉴别模型特征之间是否存在相关性,使用“GGally”包及“Pearson”相关系数进行分析鉴别模型特征之间存在的相关性。

6. 预测指标与临床理化指标的相关性

首先,分离出风险模型特征在 90 位患者中的数值,并分别匹配收缩压、舒张压、空腹血糖、甘油三酯、

胆固醇、高密度脂蛋白、低密度脂蛋白及糖化血红蛋白对应的中位数。最后使用 ggplot2、ggpubr 及 ggExtra 包构建线性相关图。

7. 统计学方法

采用 SPSS 22.0、R 4.2.1 及 3Dslicer 软件数据进行分析。计量资料用 $\bar{x} \pm s$ 表示,观察组与对照组进行独立样本 *t* 检验,计数资料用频数表示,进行 χ^2 检验。对机器学习显示有统计学意义的影像数据的变量纳入多因素 logistic 回归分析。以 $P < 0.05$ 为差异有统计学意义。

结 果

1. 基本资料

两组之间比较,年龄、性别、收缩压、高密度脂蛋白及低密度脂蛋白无统计学差异($P > 0.05$)。舒张压、空腹血糖、甘油三酯、胆固醇及糖化血红蛋白具有统计学差异($P < 0.05$),详见表 1。

表 1 初发脑梗死组和再发脑梗死组临床特征比较

	初发脑梗死组 (<i>n</i> = 26)	再发脑梗死组 (<i>n</i> = 37)	<i>P</i>
年龄(岁)	61.58 ± 10.04	60.57 ± 10.85	0.618
性别(M/F)	66.13%	68.19%	0.831
收缩压(mmHg)	143.00 ± 17.22	137.98 ± 19.35	0.156
舒张压(mmHg)	83.11 ± 11.75	88.74 ± 13.22	0.021
空腹血糖(mmol/L)	9.15 ± 2.57	11.31 ± 4.39	0.037
甘油三酯(mmol/L)	2.37 ± 1.03	2.88 ± 1.03	0.013
胆固醇(mmol/L)	4.67 ± 1.27	5.11 ± 0.76	0.037
高密度脂蛋白(mmol/L)	1.16 ± 0.26	1.11 ± 0.20	0.219
低密度脂蛋白(mmol/L)	2.93 ± 0.70	2.89 ± 0.63	0.758
糖化血红蛋白(%)	8.67 ± 1.62	9.53 ± 1.94	0.013

2. 脑梗死图像特征提取

通过 3D-slicer 分析后共获得 837 个梗死区域影像组学特征,其中一阶特征 162 个、GLCM 特征 213 个、GLDM 特征 126 个、GLSZM 特征 144 个、GLRLM 特征 144 个、NGTDM 特征 45 个。梗死区域影像特征数值通过 R 软件进行 Z-score 归一化处理(表 2)。

表 2 影像组学各类特征数量

	Original	Wavelet							
		HHH	HHL	HLH	HLL	LHH	LHL	LLH	LLL
一阶特征	18	18	18	18	18	18	18	18	18
GLCM	24	24	24	24	23	24	23	24	23
GLDM	14	14	14	14	14	14	14	14	14
GLSZM	16	16	16	16	16	16	16	16	16
GLRLM	16	16	16	16	16	16	16	16	16
NGTDM	5	5	5	5	5	5	5	5	5

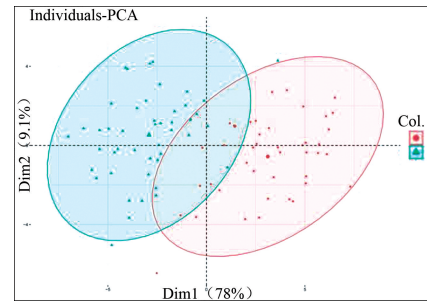
3. 两组间主成分分析

初次梗死与再发脑梗死两组间 PCA 共解释了两组间 87.1% 的差异性,两组影像学特征具有一定程度的聚类,然而初梗组与再梗组特征存在明显的分离趋势,这表明两组的脑梗死区域影像特征存在一定程度的差异性(图 1)。

4. 特征提取

提取特征后,采用 LASSO 和 SVM-RFE 两种机器学习方法进行特征筛选,在 LASSO 得到 10 个特征(wavelet.LHH_glszm_ZonePercentage, wavelet.LHH_glcm_MCC, wavelet.LHH_glcm_Imc2, wavelet.HHL_fir-storder_RobustMeanAbsoluteDeviation, wavelet.LLL_gl-

cm_JointEntropy, wavelet.LHL_glcm_Autocorrelation, wavelet.HHH_firstorder_Uniformity, wavelet.LHL_glrlm_ShortRunHighGrayLevelEmphasis, wavelet.LHL_gldm_LargeDependenceHighGrayLevelEmphasis, wavelet.HHH_gldm_GrayLevelVariance)、在 SVM-RFE 中得到 6 个特征(wavelet.LHH_glcm_MCC, wavelet.LLL_glcm_JointEntropy, wavelet.LHH_glcm_Imc2, wavelet.HHH_firstorder_Uniformity, wavelet.HHH_gldm_GrayLevelVariance, wavelet.HHH_firstorder_Entropy)。通过交集的方式得到两种机器学习共有 5 个交集特征(见图 2、3、4)。



红色表示再发脑梗死;绿色表示初发脑梗

图 1 初发脑梗死组和再发脑梗死组影像组学特征主成分分析

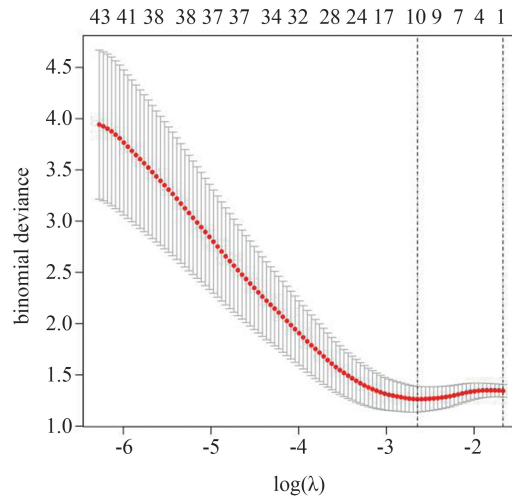
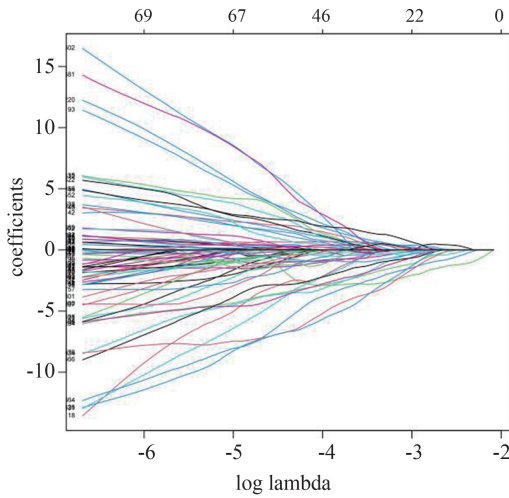


图 2 LASSO 分析筛选特征

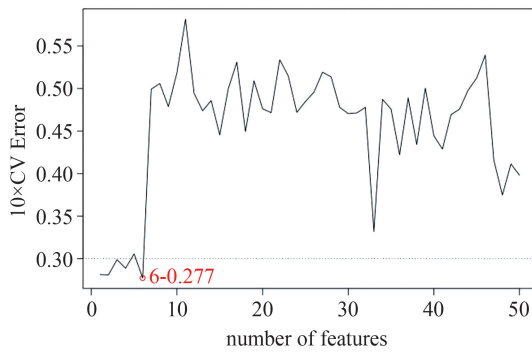


图 3 SVM-RFE 算法筛选特征

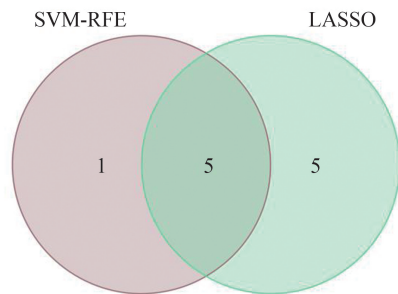


图 4 两种机器学习结果交集

5. 模型构建及验证

利用 5 个影像组学特征通过 logistic 多因素回归方法构建影像组学鉴别模型。

在列线图中 wavelet.HHH_firstorder_Entropy、wavelet.HHH_firstorder_Uniformity、wavelet.LHH_gl-

cm_MCC 及 wavelet.LLL_glcm_JointEntropy 与糖尿病患者发生再次梗死风险呈正相关, wavelet.LHH_glcm_Imc2 与与糖尿病患者发生再次梗死风险呈负相关。在训练集及测试集中灵敏度、特异度、精度、召回率、准确性、F 值、kappa 系数及 AUC 值的差距在±0.15 间,表明存在数据过拟合可能性较低,同时风险预测模型具有一定的可信度。风险预测相关分析结果显示, wavelet.LHH_glcm_MCC 与 wavelet.LHH_glcm_Imc2 存在明显的正向相关关系, wavelet.HHH_firstorder_Uniformity 与 wavelet.HHH_firstorder_Entropy 存在明显的负向相关关系。(图 5、6、7,表 3)

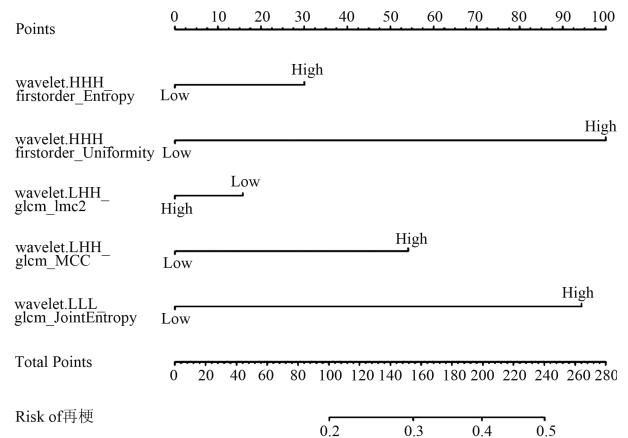


图 5 风险预测模型列线图

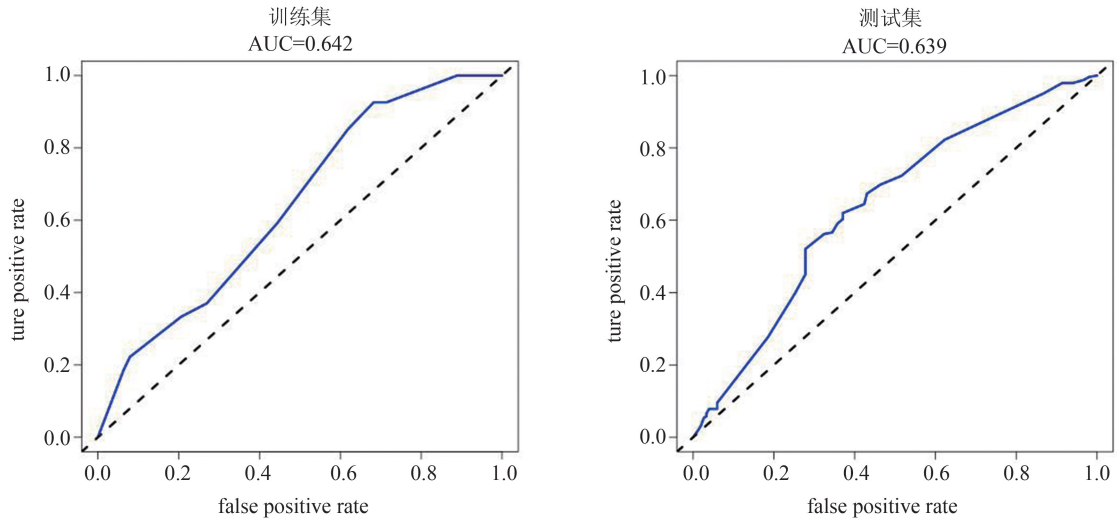


图6 训练集及测试集 AUC 图

表3 测试集及训练集中灵敏度、特异度、精度、召回率、F 值、准确性、kappa 系数及 AUC 值

分组	灵敏度	特异度	精度	召回率	F 值	准确性	kappa 系数	AUC 值
训练集	0.75	0.617	0.643	0.741	0.692	0.680	0.532	0.642
测试集	0.74	0.617	0.656	0.750	0.676	0.659	0.430	0.634

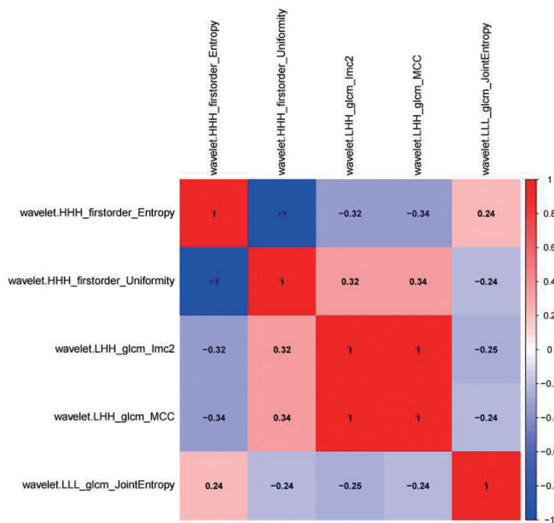


图7 预测模型特征相关性热图

6. 临床指标与鉴别模型特征线性相关分析

通过预测模型特征与收缩压、舒张压、空腹血糖、糖化血红蛋白、血脂(甘油三酯、胆固醇、高密度脂蛋白、低密度脂蛋白)相关性分析,结果显示共存在 4 种相关性,其中包括 3 个正相关,1 个负相关(表 4)。

讨论

大量研究证明,糖尿病患者极易发生脑梗死,并且在初发梗死后,仍然有相当一部分患者在几年内再次发生脑梗死^[8]。相较于初发的患者而言,再次梗死的患者不仅在死亡几率上大幅度增加,并且再梗死患者的预后及生存质量也在急速下降^[9]。近年来许多研究者致力于发现在临床上是否具有特异性指标可以标记出初发梗死患者再梗死的标志物,然而到目前为止仍未有可靠的指标^[10]。

表4 鉴别模型特征与理化指标相关性

	收缩压		舒张压		血糖		糖化血红蛋白		甘油三酯		胆固醇		高密度脂蛋白		低密度脂蛋白	
	Y	P	Y	P	Y	P	Y	P	Y	P	Y	P	Y	P	Y	P
wavelet.LHH_glcm_MCC	0.06	0.78	0.25	0.07	-0.014	0.1	-0.18	0.071	0.02	0.60	-0.01	0.79	-0.23	0.35	0.09	0.78
wavelet.LHH_glcm_Imc2	-0.12	0.20	-0.18	0.10	-0.47	0.19	-0.20	0.06	-0.05	0.08	-0.09	0.20	-0.01	0.71	-0.25	0.07
wavelet.LLL_glcm_JointEntropy	-0.013	0.13	0.50	0.001	0.17	0.06	0.33	0.02	0.3	0.03	0.22	0.12	0.04	0.29	0.05	0.51
wavelet.HHH_firorder_Entropy	0.10	0.34	0.31	0.04	0.25	0.47	0.27	0.70	0.40	0.10	0.22	0.19	0.27	0.54	0.10	0.29
wavelet.HHH_firorder_Uniformity	0.21	0.56	0.28	0.09	0.40	0.98	0.42	0.04	0.40	0.03	0.18	0.70	0.18	0.37	0.04	0.57

头部 CT 检查可以快速获得高质量的影像图像,而且具有方便快捷、无创伤、可重复等优点,在临床被广泛应用,是脑梗死诊断、鉴别诊断、指导治疗及评估预后重要检查手段。然而,从图像上并不能得到有效的指标用于再次梗死风险的预测。近年来随着影像组学的发展,许多领域的研究已经证实了其在临床预测、

诊断及预后中显示出良好的效能^[11-13]。本研究提取 837 个影像组学特征经过降维处理及两种机器学习后共得到 5 个最优特征,并使用最优特征构建列线图,列线图 AUC 值提示,在梗死预测模型具有较好的可行性。

预测模型中的 5 个特征分属两个图像特征范畴,

分别是灰度共现矩阵以及一阶特征 (first order)。GL-CM 和 First order 能在图像中反映出图像纹理以及灰度的均匀性及复杂性。最近一项实验证明,在炎症浸润下,细胞及组织图像灰度值和纹理与炎症程度密切相关,且随着炎症浸润程度的加深,灰度分布范围及纹理越不均匀^[14]。通过 Spearman 系数分析,结果显示 Joint Entropy、Entropy 以及 Uniformity 与血压、糖化血红蛋白以及甘油三酯存在显著相关性。大量研究证实高血压、高血脂及高血糖的状态下,血管壁损伤并呈现出炎症状态,然而在初期血管壁及周围组织并未出现显著损伤时,在常规检查中难以发现炎症浸润^[15-16]。然而在图像中,能通过灰度值及纹理的算法推测出组织是否具有炎症样改变。炎症浸润下,水样物质析出,组织灰度值下降,并且炎症浸润同时也会造成组织结构变化而出现纹理不均匀性及复杂性。因此在本研究中发现,高血压、高血脂及高血糖状态下患者梗死灶中灰度值及纹理呈现出复杂性及不均匀性,并且在后期更倾向于再次梗死的发生,应当给予这种患者更多关注。

本研究存在以下不足:①本研究中测试集及验证集 AUC 值较低,在未来研究中仍需扩大样本量进行进一步分析;②虽然数据集分为训练集及测试集,测试结果表明模型不存在数据过拟合,但数据来源为单中心,尚未进行外部验证;③再梗死样本量相对不足,可能会产生一定误差;④在相关性分析中临床指标与影像组学特征存在相关性,但是并未有基础研究做支撑,在未来的研究中需要进一步分析。综上所述,利用机器学习结合影像组学建立的诊断模型能够有效地预测糖尿病患者初发脑梗死后再次梗死的风险,且模型泛化能力好,对临床工作有一定的指导意义,建议临床推广使用。

参 考 文 献

[1] 中国 2 型糖尿病防治指南(2020 年版)(上).中国实用内科杂志, 2021,41(8):668-695.
[2] 郭淑红,黄静.再发脑梗塞相关及独立危险因素分析.西部中医药,

2015,28(1):45-47.

- [3] 梁佐堂,涂毅,李继锋,等. 128 排螺旋 CT 在脑梗死患者脑血管狭窄临床诊断中的价值.中国 CT 和 MRI 杂志,2021,19(11):11-13.
[4] 郭慧,陈永华. 再发脑梗塞的相关危险因素分析.中国急救医学, 2018,38(z1):117.
[5] 吴亚飞,方亚.机器学习方法在慢性病研究中的应用进展.中国卫生统计,2020,37(4):624-628.
[6] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology, 2016,278(2):563-577.
[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.中国脑血管病影像应用指南 2019.中华神经科杂志, 2020,53(4):250-268.
[8] Chen M, Luo W, Li J, et al. Clinical Characteristics and Outcomes of Acute Ischemic Stroke in Patients with Type 2 Diabetes: A Single-Center, Retrospective Study in Southern China. Int J Endocrinol, 2021;5517228.
[9] Shang L, Wang X, Sun H, et al. Effects of the triple therapy of carnosine glycoside, edaravone, and Xueshuantong in hemorrhagic cerebral infarction. Am J Transl Res, 2022,14(2):1024-1033.
[10] Chen LL, Wang WT, Zhang S, et al. Cohort study on the prognosis of acute cerebral infarction in different circulatory systems at 1-year follow-up. BMC Cardiovasc Disord, 2021,21(1):521.
[11] Conti A, Duggento A, Indovina I, et al. Radiomics in breast cancer classification and prediction. Semin Cancer Biol, 2021,72:238-250.
[12] Choi YS, Bae S, Chang JH, et al. Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics. Neuro Oncol, 2021,23(2):304-313.
[13] von Schacky CE, Wilhelm NJ, Schafer VS, et al. Development and evaluation of machine learning models based on X-ray radiomics for the classification and differentiation of malignant and benign bone tumors[J]. Eur Radiol, 2022,32(9):6247-6257.
[14] Wang Y, Huang N, Yang Z. Revealing the Role of Zinc Ions in Atherosclerosis Therapy via an Engineered Three-Dimensional Pathological Model. Adv Sci (Weinh), 2023,10(18):e2300475.
[15] Zhan Y, Brown C, Maynard E, et al. Ets-1 is a critical regulator of Ang II-mediated vascular inflammation and remodeling. J Clin Invest, 2005,115(9):2508-2516.
[16] Lachowicz S, Swieca M, Pejcz E. Improvement of Health-Promoting Functionality of Rye Bread by Fortification with Free and Microencapsulated Powders from *Amelanchier alnifolia* Nutt. Antioxidants (Basel), 2020,9(7):614.

(责任编辑:邓妍)