

· 论 著 ·

基于影像组学预测颅内原发性中枢神经系统淋巴瘤患者Bcl-2、c-Myc表达状态的应用价值

李 颢¹, 黄海涛¹, 李辉虎¹, 李 繁¹, 万 运²

1. 茂名市人民医院核医学科, 广东 茂名 525400;

2. 信宜市人民医院放射科, 广东 信宜 525300

[摘要] 目的: 基于多参数磁共振成像 (magnetic resonance imaging, MRI) 及多算法机器学习模型识别原发性中枢神经系统淋巴瘤 (primary central nervous system lymphoma, PCNSL) 中的Bcl-2和c-Myc双表达淋巴瘤 (double-expression lymphoma, DEL)。方法: 回顾并分析2021年1月—2024年1月茂名市人民医院及信宜市人民医院颅内PCNSL患者的临床资料, 根据免疫组织化学染色法检测Bcl-2和c-Myc蛋白的表达分为DEL组和非共表达组 (nDEL组)。对T1加权对比增强成像、弥散加权成像及水抑制成像进行手工分割肿瘤并提取影像组学特征。应用重复最小绝对值收缩和选择算子筛选特征, 然后使用15种机器学习算法及参数调整、自定义参数组合、10折交叉验证构建鉴别模型。结果: 共纳入200例颅内PCNSL患者 (81例DEL)。训练集、内部验证集和外部验证集之间年龄、性别、是否出血坏死、肿瘤位置、瘤周水肿、肿瘤最大径、肿瘤数量、是否侵犯脑膜或室管膜差异无统计学意义。根据组内相关系数 (>0.75) 初步获取2 895个稳定性影像学特征, 筛选得到16个特征, 极端梯度提升模型及梯度提升机模型鉴别效能最优, 验证集最高曲线下面积 (area under curve, AUC) 为0.91。结论: 基于多参数MRI及多算法机器学习模型在DEL检测中有良好的应用前景。

[关键词] 原发性中枢神经系统淋巴瘤; 双表达淋巴瘤; 机器学习; 影像组学

中图分类号: R739.41; R445.2 文献标志码: A

DOI: 10.19732/j.cnki.2096-6210.2025.03.011

基金项目: 茂名市科技计划项目 (220402114550456)。

利益冲突: 作者声明无利益冲突。

伦理批件: PJ2022MI-K034-01。

知情同意: 有。

引用本文: 李 颢, 黄海涛, 李辉虎, 等. 基于影像组学预测颅内原发性中枢神经系统淋巴瘤患者Bcl-2、c-Myc表达状态的应用价值 [J]. 肿瘤影像学, 2025, 34(3): 273-281.

Funding: Maoming Science and Technology Plan Project (220402114550456).

Conflicts of interest: authors declare no conflicts of interest.

Ethical approval: PJ2022MI-K034-01.

Informed consent: available.

Cite this article: LI X, HUANG H T, LI H H, et al. The application value of radiomics-based prediction of Bcl-2 and c-Myc expression status in patients with intracranial primary central nervous system lymphoma [J]. Oncoradiology, 34(3): 273-281.

The application value of radiomics-based prediction of Bcl-2 and c-Myc expression status in patients with intracranial primary central nervous system lymphoma LI Xie¹, HUANG Haitao¹, LI Huihu¹, LI Fan¹, WAN Yun²

(1. Department of ECT, Maoming People's Hospital, Maoming 525400, Guangdong Province, China; 2. Department of Radiology, Xinyi People's Hospital, Xinyi 525300, Guangdong Province, China)

Correspondence to: HUANG Haitao E-mail: 18506682343@163.com

[Abstract] **Objective:** To investigate the value of radiomics based on multiparametric magnetic resonance imaging (MRI) and multiple machine learning algorithms in predicting the expression of Bcl-2 and c-Myc in patients with encephalic primary central nervous system lymphoma (PCNSL). **Methods:** The clinical data of patients with intracranial PCNSL in Maoming People's Hospital and Xinyi People's Hospital from January 2021 to January 2024 were reviewed and analyzed. Based on the expression of Bcl-2 and c-Myc proteins detected by immunohistochemical staining, patients were divided into the double-expression lymphoma (DEL) group and the non-double-expression group (nDEL group). Tumors were manually segmented on MRI images to extract radiomic

features. Repeated least absolute shrinkage and selection operator (repeated-LASSO) was applied to select features, followed by the construction of classification models using 15 machine learning algorithms with parameter tuning, custom parameter combinations, LASSO, and 10-fold cross-validation. **Results:** There were no statistically significant differences in age, gender, presence of hemorrhage or necrosis, tumor location, peritumoral edema, maximum diameter, number of tumors, and presence of meningeal or ependymal invasion between training set, internal validation set, and external validation set. A preliminary set of 2 895 stable radiomic features was obtained based on an intraclass correlation coefficient ($ICC > 0.75$). Repeated-LASSO selected 16 features. The eXtreme Gradient Boosting (XGboost) model and gradient boosting machine (GBM) models, showed the best performance, with the highest area under curve (AUC) of 0.91 in the validation set. **Conclusion:** Multiparametric MRI combined with multiple machine learning algorithms shows great potential for detecting DEL in PCNSL.

[**Key words**] Primary central nervous system lymphoma; Double-expression lymphoma; Machine learning; Radiomics

原发性中枢神经系统淋巴瘤 (primary central nervous system lymphoma, PCNSL) 是罕见的非霍奇金淋巴瘤, 可累及脑、脊髓、脑膜等, 占有原发性脑肿瘤的3%。PCNSL的主要亚型为弥漫性大B细胞淋巴瘤 (diffuse large B-cell lymphoma, DLBCL) (>90%)。30%左右的DLBCL患者存在Bcl-2和c-Myc蛋白高表达, 根据世界卫生组织对造血和淋巴组织肿瘤的分类, 称之为双表达淋巴瘤 (double-expression lymphoma, DEL) [1]。在PCNSL患者中, 双表达状态影响其对化疗药物的敏感性及预后。此前研究 [2-4] 发现, 经高剂量甲氨蝶呤、长春新碱、丙卡嗪和放疗联合治疗的PCNSL患者中, 相比非双表达淋巴瘤 (nondouble-expression lymphoma, nDEL) 患者, DEL患者预后不良比例更大, 双表达状态被认为是不良预后的可靠预测因子。有1种或2种蛋白高表达的患者在5年内的进展和/或死亡风险比没有高表达的患者增加5~13倍 [4]。在PCNSL中, 部分或全部肿瘤切除是不必要的, 因为基于高剂量甲氨蝶呤的化疗是主要的治疗方案。穿刺活检是判断目标基因是否表达的“金标准”, 是有创性操作并有取样不足的风险。因此, 探寻无创分辨分子亚型的方法对于明确PCNSL的DEL状态具有重要意义。然而, 不同分子亚型在常规影像学表现以及临床特征方面, 几乎不存在显著差异。

本研究旨在基于多参数磁共振成像 (magnetic resonance imaging, MRI) 影像组学特征建立影像组学模型来识别PCNSL中的DEL和nDEL状态。

1 资料和方法

1.1 研究对象

回顾并纳入2021年1月—2024年1月茂名市人民医院及信宜市人民医院2个独立医疗中心共200例患者。纳入标准: ① 经病理学检查证实为PCNSL; ② Bcl-2和c-Myc免疫组织化学染色结果完整; ③ 患者免疫功能正常 (无先天性或获得性免疫缺陷病史); ④ 18岁以上; ⑤ 术前4周内进行MRI检查。排除标准: ① MRI检查中T2液体衰减反转恢复 (T2 fluid-attenuated inversion-recovery, T2 FLAIR)、弥散加权成像、T1加权对比增强成像 (T1-weighted contrast-enhanced, T1CE) 不完整或MRI图像存在明显伪影; ② MRI检查前进行淋巴瘤相应治疗。本研究经茂名市人民医院伦理委员会批准 (PJ2022MI-K034-01)。

1.2 仪器与方法

采用美国GE公司的Signa Creator 1.5 T、Signa Architect 3.0 T, 德国Siemens公司的Verio 3.0 T, 上海联影医疗科技股份有限公司的uMR 780 3.0 T、uMR 770 3.0 T MRI扫描仪。T2 FLAIR, 轴位, 重复时间 (repetition time, TR) /回波时间 (echo time, TE) /反转时间 = 8 527 s/162 s/2 100 s, 视野 (field of view, FOV) 24 cm × 24 cm, 矩阵 = 288 × 224, 激励次数 (number of excitation, NEX) 2。弥散加权成像 (diffusion-weighted imaging, DWI), 轴位, TR/TE = 3 000 s/65.5 s, FOV = 24 cm × 24 cm, b = 0

及1 000 s/mm²，矩阵=160×160，NEX 2；并计算表观弥散系数（apparent diffusion coefficient, ADC）。颅脑T1CE，以钆双胺注射液为对比剂，采用高压注射器以0.1 mmol/kg剂量经肘静脉以2 mL/s速率注射。扫描参数：T1加权成像（T1-weighted imaging, T1WI）轴位，TR 300~1 800 ms，TE 6~15 ms，NEX 2~4，FOV=25 cm×25 cm，层厚1~7 mm。

1.3 肿瘤分割和特征提取及筛选

MRI图像需进行预处理降低MRI扫描仪和序列参数差异的影响，包括N4偏置场校正、重采样体素大小为（0.7 mm×0.7 mm×4 mm）、体素强度值调整为100信号强度、Z-score标准化。图像导入3D-Slicer（v5.2.2）中，将FLAIR及ADC序列配准到T1CE序列上，先在T1CE上逐层勾画感兴趣区（region of interest, ROI），然后结合T1CE分别对FLAIR及ADC调整ROI。ROI包括肿瘤实体成分和坏死或囊变部分，但出血除外，尽量避免瘤周水肿或其他组织，多发病灶（病灶数量≥2个）则勾画最大的2~3个病灶。使用开源工具包Pyradiomics（<https://pyradiomics.readthedocs.io/en>）提取特征，5种图像类型（Original, Wavelet, LoG, Square, SquareRoot, Logarithm），特征类型包括一阶特征（共19个）、形状特征（共27个）、灰度共生矩阵特征（共24个）、灰度级大小区域矩阵特征（共16个）、灰度游程矩阵（gray-level run-length matrix, GLRLM）特征（共16个）、相邻灰度差矩阵（共5个）、灰度依赖矩阵（gray-level dependence matrix, GLDM）特征（共14个）。由1名MRI影像诊断医师（5年工作经验）双盲法独立手动勾画ROI，1名MRI影像诊断医师（7年工作经验）随机挑选20例独立手动勾画ROI，将这两组ROI进行特征提取，得到两组影像学特征然后进行组内相关系数（intraclass correlation coefficient, ICC）一致性检验。

T2 FLAIR、ADC、T1CE 3个序列的特征通过以下步骤进行筛选：① 选取ICC>0.75的初步特征；② 经t检验P<0.05的特征；③ 合并3个序列特征，应用重复最小绝对值收缩和选择算子

（least absolute shrinkage and selection operator, LASSO）（重复次数1 000，特征迭代次数>0.1）筛选。

1.4 模型构建

使用15种机器学习算法建模，包括神经网络、逻辑回归、线性判别分析、二次判别分析、k近邻、决策树、随机森林、极端梯度提升（extreme gradient boosting, XGboost）、岭回归、LASSO回归、弹性网络回归、支持向量机、梯度提升机（gradient boosting machine, GBM）、逐步逻辑回归、朴素贝叶斯。并通过R语言的caret包进行参数调整、自定义参数组合、LASSO特征筛选、10折交叉验证进行组合，筛选最优的3个模型，并进行性能评估。

1.5 肿瘤形态学评估

评估所有患者肿瘤形态学：① 根据病变中心确定肿瘤位置并分为两组，幕上或幕下，外周或深部（基底节区，丘脑，脑干和小脑）；② 在T1CE图像中测量肿瘤最大层面的最大径（单位mm）；③ 坏死或囊变部分为T2 FLAIR高信号为主T1CE低信号；④ 瘤周水肿分度（3度，轻度水肿宽度2 cm内，中度水肿宽度大于2 cm未超出同侧大脑半球1/2，重度水肿超出同侧大脑半球1/2）；⑤ 有无侵犯邻近脑膜或室管膜；⑥ 单发或多发病灶。

1.6 统计学处理

采用R语言4.4.0、R Studio软件进行分析。符合正态分布的计量资料以 $\bar{x}\pm s$ 表示，计数资料以n（%）表示。应用方差分析ANOVA、Kruskal-Wallis检验及 χ^2 检验来评估训练集和验证集之间的临床及影像学数据的统计学差异，P<0.05为差异有统计学意义。计算每个机器学习模型的灵敏度、特异度、准确度、F1评分和曲线下面积（area under curve, AUC）。当AUC值相同时，F1得分较高的模型为最优模型。

2 结 果

2.1 患者临床及影像学资料

200例患者共81例DEL（40.5%），其中茂名

市人民医院160例,按照8:2随机分为训练集与内部验证集,信宜市人民医院40例患者作为外部验证集,在训练集、内部验证集和外部验证集之

间,年龄、性别、是否出血坏死、肿瘤位置、瘤周水肿、肿瘤最大径、肿瘤数量、是否侵犯脑膜或室管膜差异无统计学意义(表1,图1)。

表1 训练集、内部验证集和外部验证集临床及影像学资料

Tab. 1 Clinical and imaging data of the training set, internal validation set, and external validation set

项目	训练集 (n=128)	内部验证集 (n=32)	外部验证集 (n=40)	统计量	P值
DEL	49 (38.2)	14 (43.7)	18 (45.0)	$\chi^2=0.73$	0.69
年龄/岁	58.88 ± 12.50	57.77 ± 12.70	60.37 ± 11.28	$F=0.49$	0.61
性别					
男性	58 (45.3)	15 (46.8)	19 (47.5)	$\chi^2=0.14$	0.93
女性	70 (54.7)	17 (53.2)	21 (42.5)		
出血坏死	78 (60.9)	20 (62.5)	21 (52.5)	$\chi^2=1.12$	0.57
病灶位置					
幕上	112 (87.5)	29 (90.6)	35 (87.5)	$\chi^2=0.93$	0.63
幕下	16 (12.5)	3 (9.4)	5 (12.5)		
外周	58 (45.3)	17 (53.1)	23 (57.5)	$\chi^2=1.42$	0.49
深部	70 (54.7)	15 (46.9)	17 (42.5)		
瘤周水肿				$H=1.38$	0.50
轻度	21 (16.4)	5 (15.6)	5 (12.5)		
中度	79 (61.7)	23 (71.9)	31 (77.5)		
重度	28 (21.9)	4 (12.5)	4 (10.0)		
肿瘤最大径/mm	52.92 ± 13.10	53.64 ± 12.58	51.75 ± 13.82	$F=0.21$	0.81
单发	67 (52.3)	17 (53.1)	29 (72.5)	$\chi^2=5.31$	0.07
多发	61 (47.7)	15 (46.9)	11 (27.5)		
侵犯脑膜或室管膜	60 (46.8)	11 (34.3)	11 (27.5)	$\chi^2=5.11$	0.08

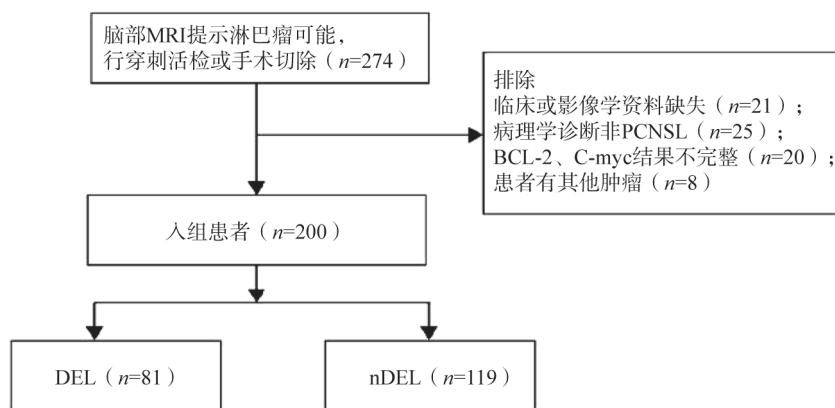


图1 研究对象纳入流程图

Fig.1 Flowchart for inclusion of participants

2.2 特征分析

经 $ICC > 0.75$ 得到2 895个影像学特征(3个序列各965个特征),应用LASSO筛选后得到16个特征(T1CE 8个,ADC 6个,T2 FLAIR 2个)(表2)。

2 895个影像学特征的Pearson相关系数的分层聚类热图见图2A,从整个热图中可以看出特征相关性较低,从树形图中可以看出不同类别的

特征之间存在关联。图2B为所有相关系数绝对值的直方图,其中91.1%的相关系数值小于或等于0.5。16个特征权重以ADC、T1CE较为显著(表2,图2E)。箱型图(图2D)展示了16个特征的分布情况,DEL组的特征值更倾向灰度不均匀和纹理更复杂;特征6、7的DEL组与nDEL组间差异无统计学意义,但与其他特征具有较好的相关性,特征之间具有相互依赖性(图2C)。

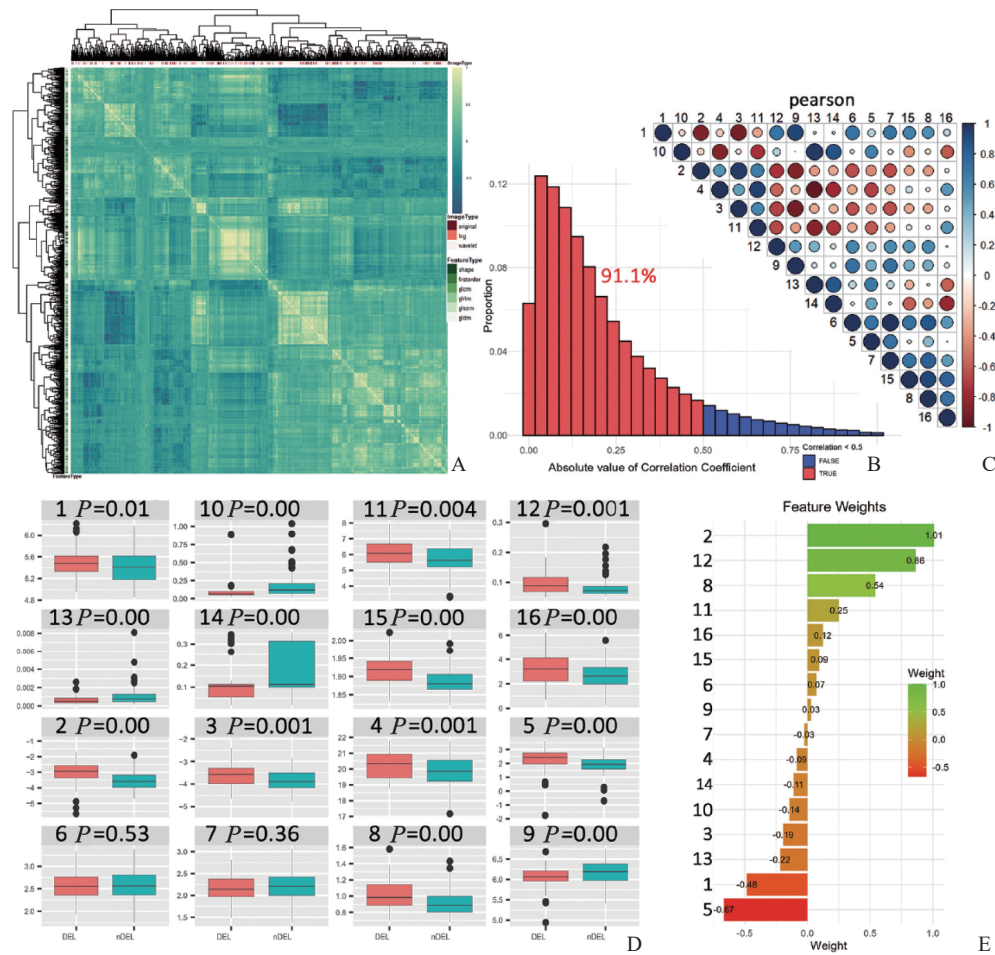


图2 影像组学特征相关性与重要性分析

Fig.2 Analysis of correlation and importance of radiomic features

A: 特征分层聚类热图; B: 特征相关系数绝对值直方图; C: 特征相关性热图; D: 特征箱型图; E: 特征权重图。

表2 筛选得到的影像组学特征

Tab. 2 Selected radiomic features

ID	特征	特征表现	权重
1	T1CE_original_firstorder_Median (中位数)	高值通常表示组织密度更高, 可能有更多的血流或活跃的细胞	-0.48
2	T1CE_log.sigma.1.0.mm.3D_firstorder_Median (中位数)	高值通常表示组织密度更高, 可能有更多的血流或活跃的细胞	1.01
3	T1CE_log.sigma.3.0.mm.3D_firstorder_Median (中位数)	高值通常表示组织密度更高, 可能有更多的血流或活跃的细胞	-0.19
4	T1CE_log.sigma.1.0.mm.3D_firstorder_TotalEnergy (总能量)	高总能量可能表示肿瘤中有很多活跃的细胞, 或者血流丰富	-0.09
5	ADC_log.sigma.1.0.mm.3D_firstorder_Mean (均值)	高值表示肿瘤的组织密度更高, 异质性增高	-0.67
6	ADC_log.sigma.1.0.mm.3D_firstorder_MeanAbsoluteDeviation (平均绝对误差)	反映了灰度值的变异性, 数值越大变异性大, 内部结构不均匀	0.07
7	ADC_log.sigma.1.0.mm.3D_firstorder_RobustMeanAbsoluteDeviation (鲁棒平均绝对偏差)	反映了灰度值的变异性, 数值越大变异性大, 内部结构不均匀	-0.03
8	ADC_wavelet.HLL_firstorder_Entropy (熵)	内部结构复杂的组织熵值往往较高, 反映高异质性	0.54

续表

ID	特征	特征表现	权重
9	T1CE_wavelet.LLL_firstorder_10Percentile (特征值的百分之10的值)	高值可能意味着肿瘤的异质性较低	0.03
10	T1CE_original_gldm_SmallDependenceLowGrayLevelEmphasis (小依赖低灰度强调)	高值表示图像中有很多小而暗的区域, 纹理较细腻	-0.14
11	T1CE_log.sigma.3.0.mm.3D_gldm_GrayLevelNonUniformity (灰度不均匀性)	高值意味纹理不均匀	0.25
12	T1CE_wavelet.LHL_gldm_DependenceNonUniformityNormalized (依赖不均匀性标准化)	高值可能提示肿瘤内部结构的异质性较高	0.86
13	FLAIR_original_gldm_SmallDependenceLowGrayLevelEmphasis	数值越大, 表示纹理越细腻	-0.22
14	FLAIR_wavelet.HHL_glrlm_ShortRunLowGrayLevelEmphasis (短运行低灰度强调)	高值可能反映出组织中有许多小而密集的细胞群, 纹理较细腻	-0.11
15	ADC_wavelet.LHL_gldm_DependenceEntropy (依赖熵)	高依赖熵表示图像的不确定性较高, 内部组织排列复杂和多样	0.09
16	ADC_wavelet.HLL_glszm_SmallAreaHighGrayLevelEmphasis (小区域高灰度强调)	高值可能意味着肿瘤中有许多活跃的细胞群或血流丰富的区域	0.12

注: 特征1-9为强度特征: 与图像灰度值有关, 描述图像的亮度和对比度; 特征10-16为纹理特征: 用于描述图像中像素的空间分布和模式; ID为特征新编号, 方便引用。

2.3 分类模型的构建及性能

使用神经网络、逻辑回归、线性判别分析、二次判别分析、k近邻、决策树、随机森林、XGboost、岭回归、LASSO回归、弹性网络回归、支持向量机、GBM、逐步逻辑回归、朴素贝叶斯15种机器学习算法建模, 并通过参数调整、自定义参数组合、LASSO特征筛选、10折

交叉验证得到多种组合, 根据内部、外部验证集AUC平均值排序(图3), 以XGBoost、GBM表现最佳, XGBoost结合10折交叉验证及LASSO构建的模型平均AUC最高, GBM模型在外部验证集中有最高AUC, 为0.91。两者模型性能接近(表3)。

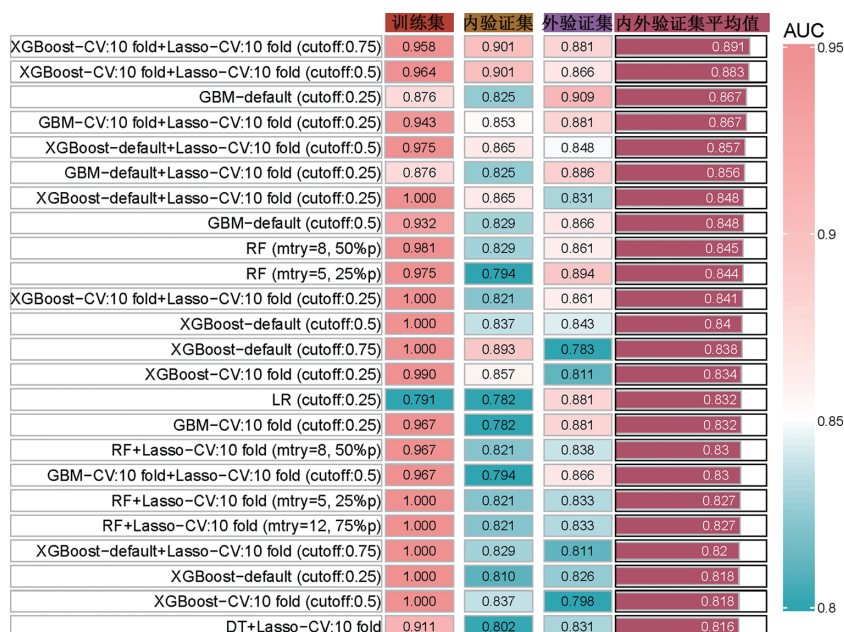


图3 根据内外验证集AUC平均值多模型排序

Fig.3 Ranking of multiple models by average AUC of internal and external validation sets

表3 模型的预测性能

Tab. 3 Predictive Performance of the Model

项目	AUC	准确度	精确度	召回度	F1分数	灵敏度	特异度	DeLong <i>P</i> 值
XGBoost								
训练集	0.96	0.93	0.92	0.94	1.00	0.90	1.00	0.00
内部验证集	0.90	0.91	0.92	0.86	0.89	0.86	0.94	0.02
外部验证集	0.88	0.88	0.81	0.94	0.87	0.94	0.82	0.17
GBM								
训练集	0.88	0.85	0.73	0.96	0.84	0.98	0.77	
内部验证集	0.83	0.81	0.72	0.92	0.81	0.93	0.72	
外部验证集	0.91	0.90	0.82	0.94	0.90	1.00	0.82	

DeLong检验: XGBoost与GBM对比。

3 讨论

本研究基于多参数MRI多机器学习算法识别颅内PCNSL中Bcl-2和c-Myc双表达状态, 筛选得到16个相关影像学特征, 3个最优模型对验证集DEL状态预测的AUC最高为0.91, 有望作为提高鉴别诊断DEL与nDEL准确性的无创诊断方法。

Bcl-2和c-Myc都是致癌基因。Bcl-2是明显抑制细胞凋亡的癌基因, c-Myc过表达可在许多肿瘤中出现, 其不是恶性肿瘤的充分条件, 常起到协同或放大作用。DEL是指Bcl-2和c-Myc蛋白均过表达, 可促进细胞增殖, 抵抗细胞凋亡, 帮助肿瘤细胞逃避免疫监视。类似分子重排(也称为双重打击淋巴瘤, DHL), DEL也预示着PCNSL的预后不良^[4], 一线治疗方案对DEL及Bcl-2或c-Myc单个蛋白高表达的nDEL有效率明显低于均未高表达的, DEL 5年内的进展和/或死亡风险最高是后者的13倍。基于这个结论, 本研究认为识别PCNSL中的DEL在诊疗及预后中十分迫切, 但颅内DEL在影像学特征和临床症状上没有明显差异, 需要病理组织切片染色进行区分。有研究尝试将多参数MRI应用于PCNSL的分子表达检测, Ozturk等^[5]的研究表明, $rADC_{mean}$ 可用于分辨PCNSL中的Bcl-2和c-Myc重排, 认为可能是重排导致病变细胞密度增加和细胞增殖。影像组学可以从传统影像中提取大量定量成像特征, 从而能够评估因视觉无法识别的内部肿瘤纹理。此前影像组学研究较多关注PCNSL与胶质瘤的

鉴别诊断及对PCNSL的治疗效果预测^[6-7], 近年有研究^[8]探讨纹理分析参数预测Bcl-2和c-Myc共表达的效能, 发现熵和能量参数有较好表现。Liu等^[9]基于多参数MRI及逻辑回归、支持向量机、k近邻、决策树和朴素贝叶斯算法构建机器学习模型预测DEL, 发现线性核支持向量机模型具有最优性能。

本研究从T2 FLAIR、ADC、T1CE图像中提取了2 895个原始特征, 整体特征间相关性较低, 不同类别的特征间存在关联; 其中91.1%的相关系数值 ≤ 0.5 , 即原始特征包含了描述肿瘤属性的全面信息, 冗余可以忽略不计。其后采用LASSO(重复次数1 000, 特征迭代次数 > 0.1)筛选后得到16个特征; LASSO具有较好的稳定性、重复性, 特征迭代次数与筛选特征个数相关, 参数值越小选择的特征越多对病灶的分析更全面。16个特征中有多个提示DEL组的肿瘤组织密度更高, 血供更丰富, 异质性更显著, 表现出更高的恶性程度, 如能量(特征4)、熵值(特征8、15)、均值及中位数(特征1~3、5)等, 这些参数高值通常与组织结构复杂、高异质性相关, 与此前研究^[8-10]相似。熵值是用于描述图像灰度分布随机性的一个特征参数。在内部结构复杂或质地不均匀的肿瘤组织中, 熵值通常较高, 这往往反映出肿瘤的高度异质性。较高的总能量通常意味着图像中存在更多高强度像素, 这可能表明肿瘤中有大量活跃的细胞或丰富的血流。均值和中位数的高值则表示肿瘤组织密度更大, 血供更丰富, 异质性增加, 意味着肿瘤的恶性程度更高^[10]。DEL中Bcl-2长期高表达抑制肿瘤细胞凋亡, 导致

细胞过度积聚,同时血管内皮细胞生长因子明显增高,肿瘤微血管过度密集,肿瘤细胞容易恶性转化,增殖速度加快,侵袭性增强,异质性更显著^[11]。我们发现特征6和7 DEL组与nDEL组间差异无统计学意义,但2个特征间及与其他特征具有较好的相关性,特征之间存在相互依赖性、协同作用。

为了构建更好的分类模型,我们应用神经网络、逻辑回归、线性判别分析、二次判别分析、k近邻、决策树、随机森林、XGBoost、岭回归、LASSO回归、弹性网络回归、支持向量机、GBM、逐步逻辑回归、朴素贝叶斯15种机器学习算法建模,并通过R语言的caret包调整参数、自定义参数组合、LASSO特征筛选、10折交叉验证得到超过200个模型,基本涵盖目前常用分类模型,根据内外部验证集AUC平均值排序,表现较好的其中包括XGBoost、GBM、随机森林、逻辑回归等及其多种组合。以XGBoost、GBM模型最优,其准确度、灵敏度、特异度均较好。该结果与Liu等^[9]发现相似,相对而言,我们增加了多种机器学习算法(共15种)以及样本量(200例),旨在获得更合适、泛化能力更强的预测模型;最后认为XGBoost、GBM算法更适合,两者均是集成学习方法,在处理复杂数据集时更加出色、具有较好的泛化能力。我们对比了XGBoost和GBM两种模型在不同数据集上的表现,XGBoost在多数数据集上的AUC高于GBM,这表明XGBoost在处理这些数据时具有更好的区分能力。尽管GBM在外部验证集上的AUC略高于XGBoost,但XGBoost在其他指标上表现更为均衡,特别是在精确度和召回率方面。DeLong检验显示XGBoost的AUC显著高于GBM,提示XGBoost在处理复杂数据集时具有更强泛化能力。

本研究局限性:在本研究中未纳入PCNSL中其他重要的分子生物学标志物,如Bcl-6、双打击DHL;DHL的诊断依赖于荧光原位杂交(fluorescence *in situ* hybridization, FISH),但此项检查耗时长、成本高,在数据收集中因缺失值较多,未能进行基因异常与蛋白表达及预后关

系分析。此外,虽然进行了外部验证,但数据较少,需要进行更多中心验证队列研究。

综上所述,在本研究中,基于ADC、T2 FLAIR、T1CE的多参数MRI及多算法机器学习模型在检测DEL状态方面表现良好,有望从分子水平反映颅内PCNSL的增殖活性,为PCNSL管理、治疗、预后评估和临床决策提供增量信息。

第一作者:

李 颢 (ORCID: 0009-0000-7167-3964), 硕士研究生, 主治医师。

通信作者:

黄海涛 (ORCID: 0000-0001-7107-2069), 硕士研究生, 主治医师, E-mail: 18506682343@163.com。

作者贡献声明:

李颢: 提出研究方向、设计研究方案、撰写论文; 黄海涛: 承担模型构建、技术开发、绘制图谱; 李辉虎: 确定研究对象范围、收集数据、采集、清洗与分析数据; 李繁: 研究方案可行性调查分析、参与论文修订、论文最终版本修订。万运: 患者数据收集及整理。

[参 考 文 献]

- [1] BRUNEAU J, MOLINA T J. Hematopathology encyclopedia of pathology [M]. Berlin: Springer, 2020.
- [2] KLANOVA M, KLENER P. BCL-2 proteins in pathogenesis and therapy of B-cell non-Hodgkin lymphomas [J]. Cancers (Basel), 2020, 12(4): 938.
- [3] KIM S, NAM S J, KWON D, et al. MYC and BCL2 overexpression is associated with a higher class of Memorial Sloan-Kettering Cancer Center prognostic model and poor clinical outcome in primary diffuse large B-cell lymphoma of the central nervous system [J]. BMC Cancer, 2016, 16: 363.
- [4] HATZL S, POSCH F, DEUTSCH A, et al. Immunohistochemistry for c-myc and Bcl-2 overexpression improves risk stratification in primary central nervous system lymphoma [J]. Hematol Oncol, 2020, 38(3): 277-283.
- [5] OZTURK K, SOYLU E, CAYCI Z. Differentiation between primary CNS lymphoma and atypical glioblastoma according to major genomic alterations using diffusion and susceptibility-weighted MR imaging [J]. Eur J Radiol, 2021, 141: 109784.
- [6] 罗 与, 白 岩, 王梅云. 影像组学在非霍奇金淋巴瘤诊断、疗效及预后评估中的研究进展 [J]. 临床放射学杂志, 2024, 43(4): 658-661.
LUO Y, BAI Y, WANG M Y. Research progress of imaging in diagnosis, curative effect and prognosis evaluation of non-Hodgkin's lymphoma [J]. J Clin Radiol, 2024, 43(4): 658-661.
- [7] GARABA A, ASLAM N, PONZIO F, et al. Radiomics for differentiation of gliomas from primary central nervous system lymphomas: a systematic review and meta-analysis [J]. Front

- Oncol, 2024, 14: 1291861.
- [8] 刘悦, 汪秀玲. 基于MRI的纹理分析预测颅内原发性中枢神经系统淋巴瘤患者Bcl-2, C-myc表达状态的应用价值 [J] . 临床放射学杂志, 2022, 41(4): 607-611.
LIU Y, WANG X L. The value of texture analysis based on MRI in predicting the expression of Bcl-2 and C-myc in patients with encephalic PCNSL [J] . J Clin Radiol, 2022, 41(4): 607-611.
- [9] LIU G L, ZHANG X Y, ZHANG N, et al. Detecting double expression status in primary central nervous system lymphoma using multiparametric MRI based machine learning [J] . J Magn Reson Imaging, 2024, 59(1): 231-239.
- [10] 董丽娜, 李梦双, 许倩, 等. 基于MRI的纹理分析预测胶质瘤患者Ki-67表达状态的应用价值 [J] . 临床放射学杂志, 2020, 39(8): 1478-1481.
DONG L N, LI M S, XU Q, et al. The value of texture analysis based on MRI in predicting the expression of Ki-67 in patients with glioma [J] . J Clin Radiol, 2020, 39(8): 1478-1481.
- [11] 秦卫. 骨髓细胞BCL-2、CD44、VEGF表达与急性白血病患者治疗效果的关系分析 [J] . 医学理论与实践, 2020, 33(22): 3823-3824.
QIN W. Analysis of the relationship between the expression of BCL-2, CD44 and VEGF in bone marrow cells and the therapeutic effect of patients with acute leukemia [J] . J Med Theory Pract, 2020, 33(22): 3823-3824.
- (收稿日期: 2025-01-16 修回日期: 2025-02-10)

《肿瘤影像学》2025年征订启事

《肿瘤影像学》自1992年创刊以来深受医学界欢迎, 1998年经中华人民共和国科学技术部、国家新闻出版署批准为国内外公开正式发行的期刊。杂志刊号: ISSN 2096-6210, CN 31-2087/R。采用优质铜版纸印制, A4开本, 64页/期, 双月刊。被中国学术期刊综合评价数据库、中国核心期刊(遴选)数据库、中国期刊全文数据库、美国化学文摘社(CAS)数据库、EBSCO学术论文数据库、美国乌利希国际期刊指南等收录, 是中国科技核心期刊。主要报道医学影像领域中科研成果、临床应用、综述、病例报告及与理工结合的有关论文等。

《肿瘤影像学》坚持学术性与科学性, 信息量大, 具有临床实用价值。是医院图书馆、影像科室及高等医药院校收存和使用的学术刊物, 是临床医学影像专业医务人员晋升中、高级职称的重要论文发表园地。欢迎各医学院校、医学图书馆、影像科室及个人向当地邮政局订阅。

本刊双月月末出版, 邮发代号4-653, 定价每期30元, 每年共180元整。

单位全称: 《肿瘤影像学》编辑部

通信地址: 上海市东安路270号复旦大学附属肿瘤医院10号楼415室

邮 编: 200032

电 话: (021)64188274

E - m a i l : zlyxx@zhongliuyingxiangxue.com

网 址 : www.zhongliuyingxiangxue.com

《肿瘤影像学》编辑部