

不同机器学习模型鉴别结肠型克罗恩病与溃疡性结肠炎的价值

杜晨 李翠平 王侠 樊梦思 孟帅 吴兴旺

[摘要] **目的** 比较不同的机器学习模型在鉴别结肠型克罗恩病(CCD)与溃疡性结肠炎(UC)中的价值。**方法** 收集2019年6月至2021年12月在安徽医科大学第一附属医院消化科就诊的有完整CT小肠成像(CTE)且经病理证实炎症性肠病(IBD)患者44例(CCD 25例,UC 19例)。利用ITK-Snap软件在静脉期病灶最明显肠段进行勾画,共计勾画106个病变肠段(CCD 58个、UC 48个)。利用AK软件提取勾画区影像组学特征,以7:3比例随机分为训练集和测试集;对训练集用Correlation_{xx}和MultiVariate_Logistic算法进行数据降维,筛选组间差异明显的影像组学特征构建6种机器学习模型,用测试集的特征对其进行验证。**结果** 175种组学特征中有4种组间差异有统计学意义($P < 0.05$)。6种模型中有4种模型曲线下面积均 > 0.90 。训练集中邻近算法(KNN)模型鉴别CCD与UC的受试者工作特征曲线下面积(AUC)为0.958(95% CI:0.917~0.992),准确率、特异度、灵敏度分别为87.7%、100%和72.7%;在测试集的AUC为0.904(95% CI:0.792~0.996),准确率、特异度和灵敏度分别为87.9%、88.9%和86.7%。**结论** 4种常用的机器学习模型在鉴别CCD与UC中均有良好的表现;其中KNN模型稳定性好,准确性更高。

[关键词] 结肠型克罗恩病;溃疡性结肠炎;影像组学;机器学习模型

doi:10.3969/j.issn.1000-0399.2023.01.004

The value of different machine learning models in discriminating colonic Crohn's disease from ulcerative colitis

DU Chen, LICuiping, WANG Xia, FAN Mengsi, MENG shuai, WU Xingwang

Department of Radiology, the First Affiliated Hospital of Anhui Medical University, Hefei 230022, China

Fund project: Scientific research project of academic and technical leaders in Anhui Province(No. 2021D299)

Corresponding author: Wu Xingwang, duobi2004@126.com

[Abstract] **Objective** To compare the value of different machine learning models in distinguishing colonic Crohn's disease (CCD) from ulcerative colitis (UC). **Methods** A total of 44 patients (25 CCD and 19 UC) with complete CT enterography (CTE) imaging and pathologically confirmed inflammation bowel diseases (IBD) were collected. Itk-snap software was used to outline the intestinal segments with the most obvious lesions at the venous phase imaging, and a total of 106 intestinal segments of lesions were delineated (58 CCD and 48 UC). The radiomics features of the delineation area were extracted by A. K. software and randomly divided into training set and test set in a ratio of 7:3. The Correlation_{xx} and MultiVariate_Logistic algorithms were used to reduce the dimensionality of the training set, and six machine learning models were constructed by screening the radiomics features with obvious differences between the groups, and then verified by the features of the test set. **Results** Four of the 175 radiomics characteristics differed significantly between groups. Four of the six models have areas under the curve > 0.90 . The Area Under Curve (AUC) (95% CI), accuracy, specificity, and sensitivity of KNN model in distinguishing CCD from UC in training set were 0.958 (0.917~0.992), 87.7%, 100% and 72.7%, respectively. In the test set, AUC (95% CI) was 0.904(0.792~0.996), accuracy, specificity and sensitivity were 87.9%, 88.9% and 86.7%, respectively. **Conclusion** The machine learning models have good performance in discriminating CCD and UC. KNN model has better stability and accuracy.

[Key words] Colonic Crohn's disease; Ulcerative colitis; Radiomics; Machine learning models

克罗恩病(Crohn's disease, CD)具有终身复发的倾向,表现为慢性反复交替感染和缓解的过程,没有诊断金标准^[1-2]。CD可以累及从口腔到直肠的任何部位,而溃疡性结肠炎(ulcerative colitis, UC)只累及结肠,当CD只累及结肠时,即结肠型CD(Colonic Crohn's disease, CCD),就难以和UC鉴别。二者的病

变分布、影像和内镜特征缺乏特异性,而CD和UC的治疗和预后有很大的区别。内镜结合病理检查是目前公认的鉴别诊断UC与CD的标准,病理发现特征性的非干酪样肉芽肿即可诊断CD。然而,该方法复杂有创,并且常需多次、多点、多部位活检,导致患者耐受性差。

基金项目:安徽省学术技术带头人科研项目(编号:2021D299)

作者单位:230022 安徽合肥 安徽医科大学第一附属医院放射科

通信作者:吴兴旺,duobi2004@126.com

常规的影像学鉴别 CCD 和 UC 有很大的局限性^[3]。2012 年荷兰学者 Lambin 等^[4]提出影像组学的概念,即通过高通量挖掘图像内部的信息,利用机器学习构建各种模型,在疾病治疗、预后判断中发挥越来越重要的价值^[5]。放射组学可以通过不同的机器学习算法输出客观分类和诊断模型,被称为医学成像与个性化医学之间的桥梁^[6]。有单纯运用组学鉴别 CD 与 UC,但是组学建模有很多学习方法,每一种机器学习方法的优缺点不同,在鉴别诊断中的效能各不相同^[7]。本研究希望通过比较常用的机器学习方法构建的预测模型在 CCD 和 UC 鉴别诊断中的优势,为影像组学选择一种更优的建模方法和为临床选择一种准确性高的基于 CT 影像组学的预测模型提供参考,旨在治疗前准确的判断 CCD 和 UC,为临床合理选择药物和治疗策略提供帮助。

1 资料与方法

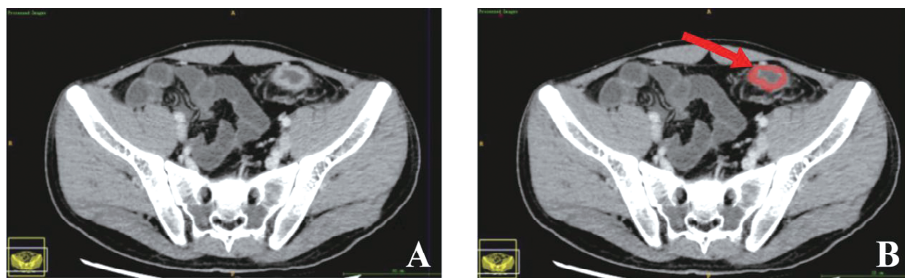
1.1 一般资料 收集 2019 年 6 月至 2021 年 12 月在安徽医科大学第一附属医院消化科就诊的经内镜或手术病理证实的 IBD 患者 44 例,其中男性 31 例、女性 13 例,平均年龄(34.55 ± 13.24)岁。

1.2 扫描方法 机型:GE Revolution CT。扫描参数:层厚 5 mm、层间距 5 mm、管电压 120 kV、管电流 215 mA、转

速 27.5 mm/rot、螺距 1.375:1,重建矩阵 512×512。患者检查前 12 h 禁食,清洁肠道,检查前 1 h 开始口服 2.5% 的等渗甘露醇溶液 1 800 mL,分 5 次服完(前 4 次每间隔 15 min 口服 400 mL,患者上扫描床前口服剩余 200 mL 以充盈胃腔)。对于无山莨菪碱使用禁忌证的患者于扫描前 10 min 肌注 20 mg 山莨菪碱。

1.3 纳入排除标准 纳入标准:①能正常配合 CT 检查的患者;②经活检病理检查或术后病理诊断为 CCD 或 UC;③图像质量良好,符合要求。排除标准:①碘过敏者,不能配合检查的患者;②图像不合格,如图像缺失、模糊、不完整等。

1.4 影像组学特征的提取与机器学习 由 1 名具有 5 年以上腹部 CT 诊断经验的医生使用 ITK-Snap 软件(version 3.6, www.itksnap.org/) 在静脉期图像病灶最明显层面沿增厚肠壁边缘手动勾画感兴趣区域(region of interest, ROI),见图 1。由 1 名从业 10 年以上的影像科医生对其进行验证,勾画时选取病变最明显且增厚肠壁边缘较清晰区域逐层勾画,避开肠壁外组织及肠腔,共勾画 106 个肠段(CCD 58 个,UC 48 个)。利用 AK 软件对提取的影像组学特征进行 Correlation 相关性研究,再以 Correlation_xx 和 MultiVariate_Logistic 算法进一步处理,筛选出最佳影像组学特征建立预测模型。



注:A,静脉期病灶 CTE 图像;B,静脉期病灶 CTE 图像标示 ROI 后;箭头示勾画后 ROI。

图1 ROI 勾画前后 CTE 图像示例

随机选择 70% 病例为训练集,30% 病例为测试集。在训练组分别选择 logistic 回归、贝叶斯算法(native Bayes, NB)、随机森林(random forest, RF)、决策树(decision tree, DT)、邻近算法(knearestneighbor, KNN)和支持向量机(support vector machine, SVM)模型进行数据处理,将测试集结果作为评估分类器特性的重要指标,并按照准确率选取最优化分类器模块。

1.5 统计学方法 采用 SPSS 25.0 统计软件进行统计分析。计量资料符合正态分布以 $\bar{x} \pm s$ 表示,两组间比较采用独立样本 t 检验;偏态分布计量资料用 $M(P_{25}, P_{75})$ 表示,比较采用 Mann-Whitney U 检验;计

数资料以百分比表示,组间比较采用 χ^2 检验。以 $P < 0.05$ 为差异具有统计学意义。对于 $P < 0.05$ 的参数绘制受试者工作特征(receiver operating characteristic, ROC)曲线,并计算曲线下面积(area under the curve, AUC)。

2 结果

2.1 一般资料 CCD 组患者男性 18 例、女性 7 例,年龄(27.9 ± 8.3)岁;UC 组患者男性 13 例、女性 6 例,年龄(43.3 ± 13.9)岁。性别差异无统计学意义($\chi^2 = 0.001, P = 0.976$),年龄差异有统计学意义($t = 0.970$,

$P < 0.001$)。

2.2 影像组学特征 在175种组学特征中筛选了4种组间差别显著的影像组学特性构建预测模型(见表1)。基于CT图像所构建影像组学模型的特点在训练集和测试集中的影像组学评分(Rad-score =

$$\text{wavelet-HHL_glcm_Contrast} \times -3.4901 + \text{lbp-3D-m1_firstorder_Kurtosis} \times 1.5833 + \text{wavelet-HLL_glcm_Imc1} \times -1.2739 + \text{lbp-3D-m2_firstorder_Range} \times 1.2317 + 0.3454$$

表1 经多变量逻辑回归筛选后用于构建影像组学模型的特征

影像组学	回归系数	标准误	Z 值	P 值
intercept	0.345	0.461	0.750	0.453
wavelet-HHL_glcm_Contrast	-3.490	1.097	-3.181	0.002
lbp-3D-m1_firstorder_Kurtosis	1.583	0.549	2.885	0.004
wavelet-HLL_glcm_Imc1	-1.274	0.539	-2.365	0.018
lbp-3D-m2_firstorder_Range	1.232	0.570	2.163	0.031

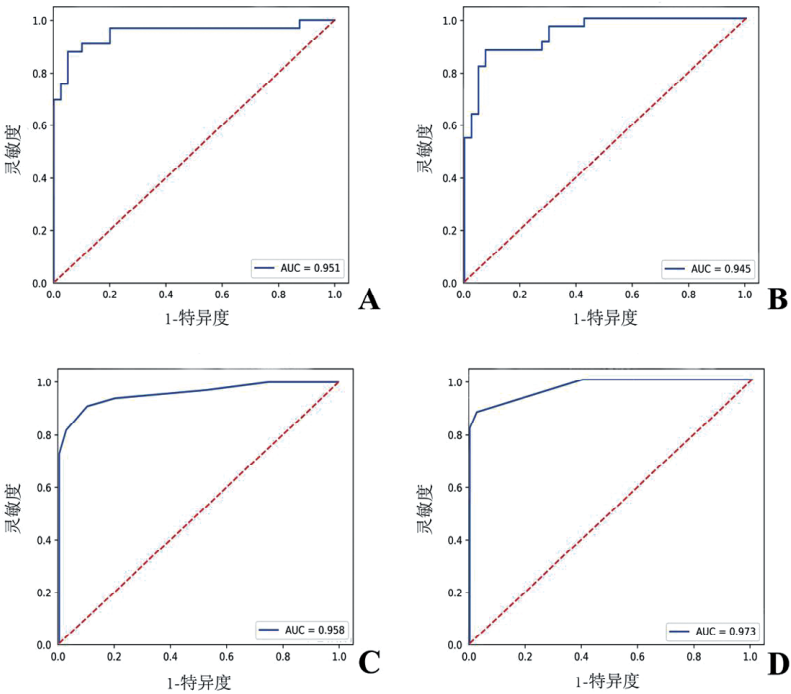
2.3 影像组学模型的效能 在4种影像组学预测模型的训练集中,曲线下面积(95% CI)、准确率、灵敏度及阴性预测值方面,DT模型都超过其他3类模型;特

异度和阳性预测值方面,KNN模型都超过其他3种模型(见表2)。训练集中建立的4种影像组学预测模型的ROC曲线见图2。

表2 4种影像组学预测模型在训练集中的诊断效能

模型	AUC(95% CI)	准确率	灵敏度	特异度	阳性预测值	阴性预测值
Logistic	0.951 (0.896 ~ 0.990)	0.904	0.848	0.950	0.933	0.884
NB	0.945 (0.901 ~ 0.980)	0.859	0.727	0.950	0.923	0.809
KNN	0.958 (0.917 ~ 0.992)	0.877	0.727	1.000	1.000	0.816
DT	0.973 (0.949 ~ 0.993)	0.932	0.879	0.975	0.967	0.907

注:Logistic 为 logistic 回归,NB 为贝叶斯算法,KNN 为邻近算法,DT 为决策树,AUC 为曲线下面积。



注:A,logistic 模型 ROC 曲线;B,NB 模型 ROC 曲线;C,KNN 模型 ROC 曲线;D,DT 模型 ROC 曲线

图2 训练集中4种影像组学预测模型的ROC曲线

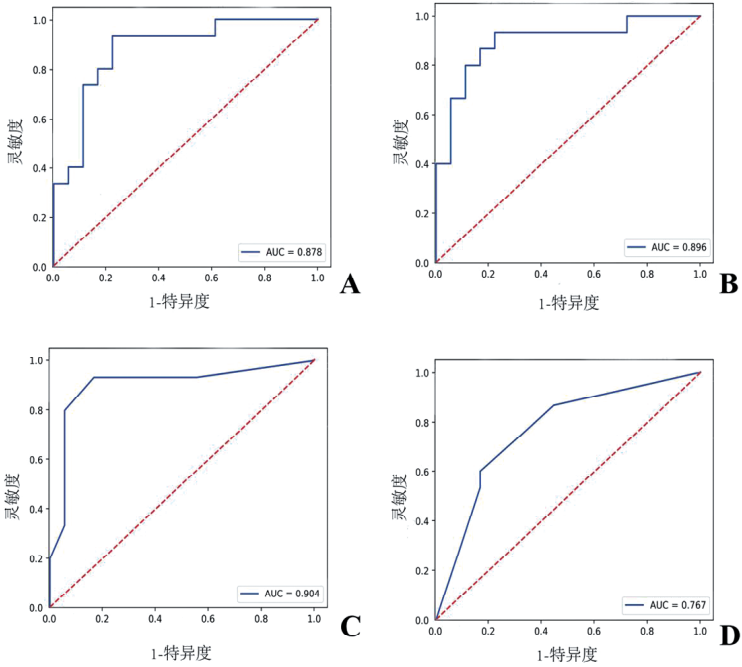
在4种影像组学预测模型的测试集中,AUC(95% CI)、准确率、灵敏度、特异度、阳性预测值、阴性预测值等方面,KNN模型都超过了其他3类模型,见

表3。测试集中建立的4种影像组学预测模型的ROC曲线见图3。

表 3 4 种影像组学预测模型在测试集中的诊断效能

模型	AUC(95% CI)	准确率	灵敏度	特异度	阳性预测值	阴性预测值
Logistic	0.878 (0.765 ~ 0.967)	0.788	0.800	0.778	0.730	0.824
NB	0.896 (0.788 ~ 0.981)	0.818	0.800	0.833	0.800	0.833
KNN	0.904 (0.792 ~ 0.996)	0.879	0.867	0.889	0.867	0.889
DT	0.767 (0.621 ~ 0.891)	0.727	0.600	0.833	0.750	0.714

注:Logistic 为 logistic 回归;NB 为贝叶斯算法;KNN 为邻近算法;DT 为决策树;AUC 为曲线下面积。



注:A,logistic 模型 ROC 曲线;B,NB 模型 ROC 曲线;C,KNN 模型 ROC 曲线;D,DT 模型 ROC 曲线。

图 3 测试集中 4 种影像组学预测模型的 ROC 曲线

3 讨论

CD 是一种胃肠道的慢性炎症性肉芽肿性病变,呈节段性分布,按照病变发生部位可分为多种亚型,其中 CCD 占有所有患者的 15% ~ 25%^[8-9];UC 是一种慢性非特异性炎性病变,呈连续性弥漫性分布^[10],二者有许多相似之处,因此早期的诊断对临床治疗及预后具有重要意义。

CCD 及 UC 各年龄阶段患者均可发病,但 CCD 常见于青少年患者^[11]。CCD 发病的高峰年龄稍年轻,男性略多于女性;UC 发病的高峰年龄较大,男女性别差异不大^[12]。本研究中 UC 组患者的年龄大于 CCD 组,两组之间差异有统计学意义,这可能是因为 UC 好发于青壮年,与免疫异常、感染、遗传等因素有关,两组之间性别差异并无统计学意义。

目前,临床上多通过肠镜后病理活检对 CCD 及 UC 做出诊断,但该检查有创,且操作难度大、费时、重复性较差^[13],易引起患者生理、心理上的不适及存在一些并发症发生的风险。UC 与 CD 的病变累及部位

不同,在鉴别诊断中具有重要意义,然 CCD 只累及结肠,与 UC 鉴别较为困难。因此,探寻一种无创、准确的诊断方法对鉴别 CCD 和 UC 有重要临床意义。随着计算机科学的进展,利用影像组学结合统计学和机器学习等大数据分析建模方法,可以很好地辅助临床诊断 CCD 与 UC^[14]。

本研究通过对 44 例(106 个肠段)IBD 患者的 CTE 图像甄选出 CCD 与 UC 组间差异明显的 4 个影像组学特点,进行了机器学习模型训练。在模型训练阶段选择了 6 种机器学习模型,其中 Logistic、NB、KNN、DT4 种模型在训练集中的分类效果较好,曲线下面积均 > 0.90。

logistic 回归训练速度较快,方法简单容易掌握,但其不易解决数据不均匀的问题,易产生欠拟合^[15]。对于数量较大的训练和查询,NB 具有较高的计算速度,而 NB 因为采用了关于样本属性独立性的假设,使得当与样品属性有关联时其有效性较差^[16]。本研究训练集及测试集中 logistic 回归及 NB 不如其他 2 类模型。

虽然 KNN 算法能够直接加入数据集中而不用经过重复训练,但当样本数量不平衡时,估计的误差也比较大^[17]。本研究训练集中,特异性和阳性预测值方面,KNN 模型都超过其他 3 种模型,测试集中, AUC (95% CI)、准确率、灵敏度、特异度、阳性预测值、阴性预测值等方面,KNN 模型均超过其他 3 类模型。

DT 能够可视化的数据分析,便于提炼出规律,在测试大数据集时,运算速度也相当快,但 DT 很容易忽略属性之间的相关产生过拟合^[18]。本研究训练集中, AUC(95% CI)、准确率、灵敏度及阴性预测值方面,DT 模型都超过其他 3 类模型,但在测试集中 AUC (95% CI)、准确率、灵敏度、阳性预测值及阴性预测值方面均不如其他 3 类,阳性预测值方面不如 NB 及 KNN。

结果表明,KNN 模型的稳定性较高,在测试集中准确率最高,AUC(95% CI)为 0.904(0.792~0.996),准确率为 87.9%。尽管基于 KNN 分类器所建立的组学预测模型,在训练集中的 AUC(95% CI)不如其他 3 类,但差异较小。同时该模型在测试集中鉴别诊断 CCD 与 UC 的准确率最高,整体表现较另外 3 种预测模型更优。KNN 无数据输入假定,对于处理样本集重叠更有优势^[19]。本研究中 CCD 和 UC 数据集之间出现了交叉重叠,故可能导致 KNN 算法检测效果明显高于其他诊断模式的因素。

研究的局限:①本研究样本量有限,需进行多中心的研究扩大数据集;②本研究未纳入除年龄和性别之外的其他临床研究信息,亟待进一步充实;③手动勾画 ROI,由于具有一定的主观偏倚,今后需开展自动分割技术;④本研究结果主要是根据病灶明显层面完成了 ROI 的勾画,有待开展多层面的 3D 勾画,以便进一步完善分析病灶的研究工作,进而更准确鉴别 CCD 与 UC。

综上所述,不同的机器学习模型在鉴别 CCD 与 UC 中,KNN 模型较为稳定,准确性最高,提示具有较好的临床应用前景。

参考文献

- [1] 柳慧,张玉洁,吴开春. 炎症性肠病临床研究现状及进展[J]. 临床内科杂志,2021,38(2):90-93.
- [2] 吴开春,梁洁,冉志华,等. 炎症性肠病诊断与治疗的共识意见(2018年·北京)[J]. 中国实用内科杂志,2018,38(9):796-813.
- [3] 涂慧娟,边联龙,黄永志,等. 克罗恩病 CT 小肠造影表现与活动指数的相关性[J]. 实用放射学杂志,2016,32

- (3):373-376,422.
- [4] LAMBIN P, RIOS-VELAZQUEZ E, LEI JENAAR R, et al. Radiomics: extracting more information from medical images using advanced feature analysis[J]. Eur J Cancer, 2012, 48(4): 441-446.
- [5] 蒋启红,任勇军. 影像组学在肝细胞癌诊疗中的研究进展[J]. 磁共振成像,2021,12(12):105-107,111.
- [6] 李文华,李天然,刘海丽. 肝癌影像组学临床应用研究进展[J]. 影像技术,2021,33(6):46-52.
- [7] LI H, MO Y, HUANG C, et al. An MSCT-based radiomics nomogram combined with clinical factors can identify Crohn's disease and ulcerative colitis. [J]. Ann Transl Med,2021,9(7):572.
- [8] 陈菁华. 炎症性肠病的 MRI 进展[J]. 实用放射学杂志,2017,33(9):1465-1467.
- [9] 周杰,李彪,孔德灿,等. 结肠型克罗恩病与溃疡性结肠炎的临床影像特征对比分析[J]. 中华炎性肠病杂志,2021,5(4):308-313.
- [10] 彭明洋,张丹凤,张卫东,等. CT 小肠造影对克罗恩病和溃疡性结肠炎的鉴别诊断价值[J]. 东南大学学报(医学版),2019,38(3):511-514.
- [11] 曾小明. 炎性肠病的诊断及难治性炎性肠病的药物治疗效果分析[J]. 临床合理用药杂志,2018,11(34):77-78.
- [12] 中华医学会消化病学分会炎症性肠病学组. 炎症性肠病诊断与治疗的共识意见(2018年,北京)[J]. 中华消化杂志,2018,38(5):292-311.
- [13] 郭素娟,肖威,邵成功,等. 多层螺旋 CT 小肠增强造影在诊断炎性肠病中的应用价值探讨[J]. 临床医学工程,2019,26(10):1315-1316.
- [14] 李双双,侯震,刘娟,等. 影像组学分析与建模工具综述[J]. 中国医学物理学杂志,2018,35(9):1043-1049.
- [15] 邓楠,罗幼喜. 函数型 Logistic 回归模型研究与应用[J]. 湖北工业大学学报,2022,37(1):115-120.
- [16] 程一芳. 数据挖掘中的数据分类算法综述[J]. 数字通信世界,2021(2):136-137,140.
- [17] 黄贤英,熊李媛,刘英涛,等. 基于类别特征改进的 KNN 短文本分类算法[J]. 计算机工程与科学,2018,40(1):148-154.
- [18] 李宏彬,贺太平. 三种决策树同源算法在肝部 B 超计算机辅助诊断中的应用比较[J]. 医学信息,2021,34(19):13-18.
- [19] 黄志成,叶钉利,胡乔治,等. 基于 CT 影像组学模型鉴别诊断小细胞肺癌与非小细胞肺癌[J]. 中国介入影像与治疗学,2021,18(8):474-478.

(2022-05-12 收稿)

(本文编校:张迪,崔月婷)