

论 著

基于CT平扫影像组学特征在预测胸腺上皮性肿瘤WHO简化病理分型中的价值

陈 杰 洪 悦 王 艳*

新疆维吾尔自治区人民医院放射影像中心
(新疆 乌鲁木齐 830001)

【摘要】目的 探讨基于CT平扫的影像组学特征在预测胸腺上皮性肿瘤WHO简化病理分型中的应用价值。**方法** 回顾性收集2010年1月-2022年3月由病理结果证实的胸腺上皮肿瘤(TETS)患者共57例,通过TETS的WHO简化病理分型,分为低危组(A、AB、B1型)23例、高危组(B2、B3、C型)34例,并按照8:2的比例,随机分成训练集和测试集。每个病灶均由两名放射科医生经过协商后,利用ITK-SNAP软件对兴趣区(ROI)进行勾画。使用Python v3.67提取放射组学特征,并使用Spearman相关系数和LASSO特征选择方法进行降维和筛选。在训练集中,应用支持向量机(SVM)、多层感知机(MLP)和逻辑回归(LR)构建术前诊断预测模型。采用受试者工作特征曲线(ROC)评估预测效果,并通过内部测试集验证预测模型。**结果** 共提取了1649个影像组学特征参数,经过Spearman相关系数筛选得到221个差异特征,并通过LASSO方法将其降维至12个组学特征。在测试集中,基于SVM、MLP和LR构建的术前预测模型分别表现出AUC值为0.800、0.868和0.971,其中LR模型具有更好的预测效果。**结论** 基于CT平扫影像组学特征构建的SVM、MLP和LR模型在预测胸腺上皮性肿瘤WHO简化病理分型方面展现出较好的预测潜力,其中LR模型具有更好的预测效果。

【关键词】 胸腺上皮肿瘤; WHO病理分型; 影像组学; 支持向量机; 多层感知机; 逻辑回归

【中图分类号】R736.3

【文献标识码】A

DOI:10.3969/j.issn.1672-5131.2024.01.023

Value of Simplified Pathological Classification of Thymus Epithelial Tumors Based on Plain CT Imaging Features

CHEN Jie, HONG Yue, WANG Yan*

Department of Radiology and Medical Imaging, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi 830001, Xinjiang Uygur Autonomous Region, China

ABSTRACT

Objective To investigate the utility of CT plain scan based radiomics features in predicting who simplified pathologic classification of thymic epithelial neoplasms. **Methods** a total of 57 patients with thymic epithelial tumors (TETS) confirmed by pathological findings from January 2010 to March 2022 were retrospectively enrolled and divided into low-risk group (types A, AB, and B1) 23 patients, high-risk group (types B2, B3, and C) 34 patients by who simplified pathological classification of TETS, and randomly divided into training set and test set according to the ratio of 8:2. Each lesion was delineated by a region of interest (ROI) using itk-snap software after consultation by two radiologists. Radiomics features were extracted using Python v3.67 and further feature dimensionality reduction, filtering was performed by using Spearman's correlation coefficient, least absolute shrinkage and selection operator method (lasso). In the training set, three machine learning methods including support vector machine (SVM), multi-layer perceptual machine (MLP), and logistic regression (LR) were selected to construct the diagnostic prediction model. The receiver operating characteristic curve (ROC) was employed to evaluate its predictive efficacy, and an internal test set was applied to validate the above prediction model. **Results** a total of 1649 radiomics feature parameters were obtained, 221 differential features were obtained by using Spearman correlation coefficients, and lasso was reduced to 12 omics features. The AUC values of the preoperative prediction model built by SVM, MLP and LR in the test set were 0.800, 0.868, 0.971, respectively, among which LR had better prediction. **Conclusion** in predicting the WHO simplified pathological classification of thymic epithelial tumors, the SVM, MLP, and LR models constructed based on CT plain imaging omics features have predictive potential, among which the LR model has better predictive effect.

Keywords: Thymic Epithelial Tumor; WHO Pathological Classification; Radiomics; Support Vector Machine; Multilayer Perceptron; Logistic Regression

胸腺上皮性肿瘤(thymic epithelial tumors, TETs)是发生在前上纵隔的最常见原发肿瘤。TETs具有生物学特性多样的特点,因此目前多认为TETs的病理学分型,可以作为其预后的独立危险因素^[1-5]。2015年,世界卫生组织(world health organization, WHO)对TETs的组织学分型进行了修订。为了便于实际临床应用及研究,可将TETs分型简化为低危组(A、AB、B1型)和高危组(B2、B3、C型)^[6-7]。CT是TETs术前的检查手段,而目前术前评估多基于增强CT的特征表现^[8-11],单纯平扫CT的应用价值不大。近年来,影像组学在各种疾病中的应用越来越多,是一种新的工具,已在多种肿瘤鉴别诊断中发挥作用^[12-14],可以通过对肿瘤分型、分级、预后评估等方面的有用信息进行提取、分析,目前应用越来越广泛。本研究旨在探讨基于CT平扫的影像组学特征,对于预测TETs的WHO简化病理分型中的价值,为临床术前诊断提供新的思路。

1 资料与方法

1.1 一般资料 本研究为一项回顾性、单中心研究。纳入了2010年1月至2022年3月我院经病理证实的TETs患者。

本研究的纳入标准包括:病理数据和临床数据完整;术前2周内行胸部CT平扫检查。排除标准包括:影像质量差,导致无法准确提取影像组学特征的情况;合并其他类型肿瘤者或者混合型胸腺瘤患者。最终,研究共纳入了57名患者,低危组23例,高危组34例;其中男性30例,女性27例,年龄范围为33~79岁,平均年龄(52.9±11.4)岁;以咳嗽(19例)、呼吸困难(6例)、胸痛(14例)、胸闷(12例)、重症肌无力(7例)、无症状者15例为主要临床症状。按照8:2比例随机划分为训练集和测试集,训练集45例,其中低危组16例、高危组29例;测试集12例,其中低危组5例、高危组7例。

1.2 方法 本研究采用GE Light Speed 64层螺旋VCT或Siemens Somatom FLASH双源CT机进行CT检查。扫描参数:管电压120kVp,管电流200-300mAs,球管旋转速度为0.5s/r,螺距1.375:1,准直64×0.6mm,探测器宽度40mm,显示野(FOV)36cm,CT图像采用512×512的矩阵,并且使用5mm的层厚和层距进行采集。患者取平卧位,单次深屏气时行胸部CT扫描。纵隔窗的窗位、窗宽设定为:50、350HU。

1.3 图像分割和CT图像组学特征提取

1.3.1 病灶ROI勾画及特征提取 所有的CT图像转换为DICOM格式并导出为NII格式。为了减少不同设备对图像组学特征稳定性的影响,对图像进行标准化和重采样后再提取特征,Resampling and size参数为:3、3、3。采用ITK-SNAP软件(<http://www.itksnap.org/>)对全肿瘤进行3D手工勾画,为确保病灶ROI的准确性,本研究将由两名放射科医

【第一作者】陈 杰,男,副主任医师,主要研究方向:CT、MRI影像诊断。E-mail: lamal2000@163.com

【通讯作者】王 艳,女,主任医师,主要研究方向:CT、MRI影像诊断。E-mail: drwangxj@163.com

师进行病灶ROI的勾画，其中一名医师拥有5年以上的工作经验，另一名医师拥有10年以上的工作经验，在不知病理结果的情况下协商后勾画ROI(图1)。通过使用Python v3.67中的PyRadiomics v3.0.1包(<https://pypi.org/project/pyradiomics/>)从勾画的ROI中提取放射组学特征。手工勾画的特征可以分为三类：一阶特征、形态学特征和纹理特征(包括灰度共生矩阵(gray-level co-occurrence matrix, GLCM)、灰度级带矩阵(gray level size zone matrix, GLSZM)等)。

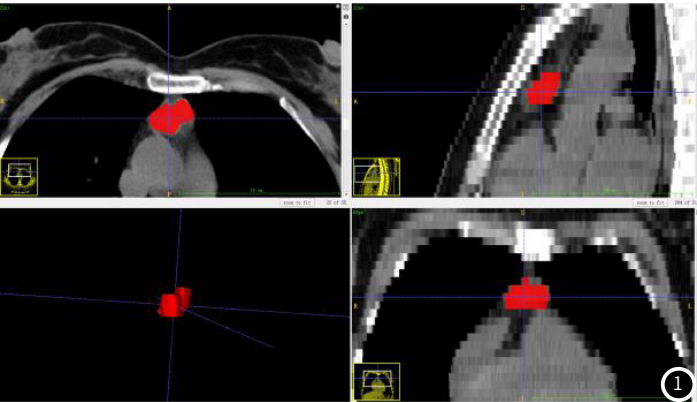


图1 ITK-SNAP软件勾画病灶，红色区域为病灶ROI及VOI。

1.3.2 影像组学特征的筛选 在进一步特征降维的过程中，本研究使用了Spearman相关系数和最小绝对收缩和选择算子方法(LASSO)。首先，通过计算Spearman相关系数，排除高度相关的特征，以减少冗余信息。接下来，采用LASSO方法将某些特征的权重系数缩小为0，通过使用交叉验证的方式找到能使LASSO回归模型误差最小的lambda(λ)值。保留权重系数不为0的影像组学特征。这些特征与它们的加权系数相乘，构成了影像组学标签。这些标签将被纳入影像组学模型的构建中，以进一步分析病灶的特征和预测结果。

1.3.3 预测模型的开发和验证 本研究选取了三种机器学习算法来构建CT影像组学特征模型。选取的模型包括支持向量机(support vector machines, SVM)、多层感知机(multilayer perceptron, MLP)、逻辑回归(logistic regression, LR)。通过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)量化评估模型的预测性能。并通过独立的内部测试集验证预测模型的效果。

1.4 统计学方法 采用SPSS 22.0软件，计量资料均以均数±标准差表示，进行正态性检验，数据符合正态分布，使用两独立样本t检验。单变量计量资料采用t检验，计数资料采用 χ^2 检验(或连续校正Yates卡方检验)。采用python(版本3.67)语言软件对影像组学数据进行分析。在相关性分析中，采用Spearman相关系数来评估特征之间的相关性，相关系数绝对值(|r|)>0.9表示高度相关。采用LASSO回归模型对具有差异的特征进行降维筛选。采用ROC曲线评价影像组学标签的预测性能，并计算敏感度、特异度。

2 结果

2.1 临床结果 本研究纳入了57例TETs患者，其中包括低危组23例(其中A型5例，AB型11例，B1型7例)和高危组34例(其中B2型9例，B3型7例，C型18例)。本研究的两组在年龄(55.00 ± 11.81 vs. 51.56 ± 11.04 , $t=1.112$)、性别(9/14 vs. 21/13, $\chi^2=2.819$)方面差异无统计学意义($P>0.05$)。57例患者中，通过健康体检发现肿瘤的共15例，其余42例主要以胸痛、胸部不适等症状为首发症状，7例患者伴有重症肌无力症状。低危组与高危组有临床症状(18/23 vs. 24/34, $\chi^2=0.417$)差异无统计学意义($P>0.05$)。

2.2 一般影像特征 CT扫描示病灶边缘多呈分叶状，平均直径为(5.8 ± 3.2)cm，低危组及高危组肿瘤在直径(6.1 ± 3.8 cm vs. 5.5 ± 2.8 cm, $t=0.644$)差异无统计学意义($P>0.05$)。肿瘤密度多表现为不均匀，其中13例可见囊变或坏死。平扫CT值范围31~56HU，平均(43.5 ± 6.1)HU。9例心包或血管侵犯，20例伴钙

化，15例伴有胸腔积液。

2.3 CT影像组学特征筛选 本研究共提取了1649个组学特征。首先，采用斯皮尔曼相关分析对数据进行降维，最终筛选出221个高度相关的特征。在去除低性能特征后，进一步使用LASSO和五折交叉验证进行特征筛选。通过调整lambda(λ)值，确定最优lambda值为0.06866(见图2A、2B)，并从中获得了13个主要特征及其相关权重系数(见图2C)。这些特征包括1个一阶特征和12个高阶纹理特征，用于构建模型。

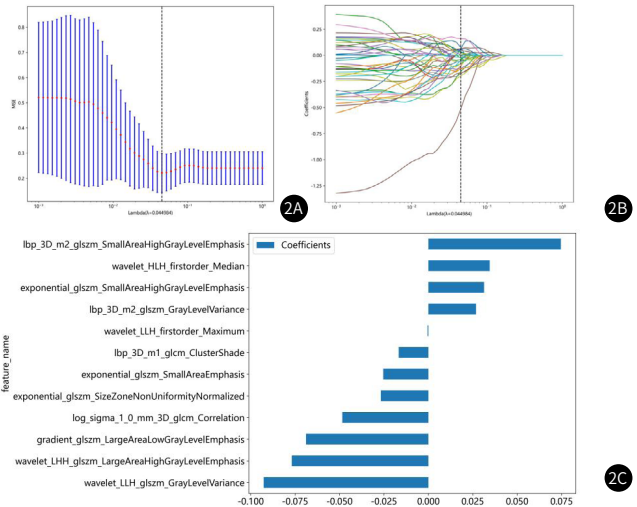


图2A-图2C 影像组学特征的筛选过程。图2A 使用LASSO回归和五折交叉验证对组学特征进行筛选，选择出了具有预测能力的特征。二项式偏差最小值对应的lambda(λ)值；图2B 每条曲线代表各自变量系数的变化，其中垂线对应a图中相应的最优 λ 值；图2C 特征权重条形图。

2.4 预测模型构建与比较 在本研究中，采用了三种常见的机器学习模型，包括支持向量机(SVM)、多层感知机(MLP)和逻辑回归(LR)，来构建影像组学的预测模型。并使用ROC曲线来评估这些模型的预测性能。(表1、图3A)。三种模型在训练集AUC值均较高，而在测试集LR模型具有较高的预测效能，AUC值为0.971，优于SVM及MLP模型(表1、图3B、图4A-C)。

表1 三种机器学习模型预测效能比较

模型	训练集				测试集			
	AUC	95%CI	敏感度	特异度	AUC	95%CI	敏感度	特异度
SVM	0.986	0.961~1.000	1.000	0.889	0.800	0.500~1.000	1.000	0.600
MLP	0.969	0.927~1.000	0.963	0.889	0.686	0.345~1.000	0.857	0.600
LR	0.947	0.867~1.000	1.000	0.889	0.971	0.892~1.000	0.857	1.000

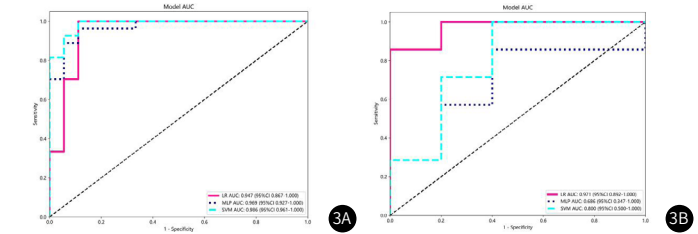


图3A-图3B 训练组及测试组中三种模型的对于的ROC曲线。图3A 训练组，三组模型均有较高的AUC值；图3B 测试组，LR模型具有较高的AUC值。

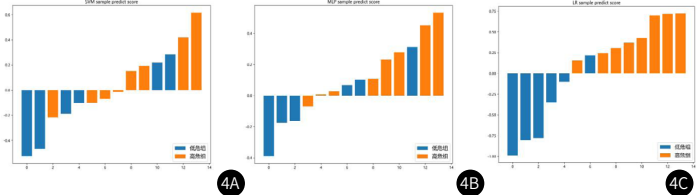


图4A-图4C 三种模型的样本预测直方图。图4A SVM模型；图4B MLP模型；图4C LP模型；基线以下的蓝色柱形表示预测正确的低危组样本，基线以上的红色柱形表示预测正确的高危组样本，而交叉部分表示预测错误的样本，LP模型整体预测效果佳。

3 讨论

本研究回顾性分析了TETs分型简化为低危组和高危组患者的CT平扫影像组学特征,并基于特征的筛选、降维后,筛选出与病理类型预测相关的13个最重要的影像组学特征,构建了由SVM、MLP、LR三种机器学习预测模型,用于预测TETs中低、高危组。三种模型在训练集中都有较好的诊断效能,AUC值均高于0.950以上;在测试集中,本研究观察到LR模型显示出良好的诊断性能,其AUC值高达0.971。这意味着该模型在预测TETs中低危组和高危组方面具有较高的准确性和区分能力。LR模型明显优于其他两种模型,对于临床TETs术前病理分型的预测有重要的辅助价值。作为TETs患者的首选检查,MSCT能直观地反映病灶的形态学及解剖学信息,但在实际工作中,诊断医师主观判断和经验可能对影像学特征的分析产生一定的影响。这是因为影像学是一门主观性较强的学科,医生在观察、分析和解释影像时会受到自身认知、经验和个人偏好的影响。因此,仅凭影像学的常规表现是无法很好地预测肿瘤恶性程度^[15-16],可重复性不稳定。有研究表明基于胸腺肿瘤体积、边界、强化程度等影像表现构建的CT模型预测恶性程度的准确率仅为70.8%^[6],CT增强影像组学具有区分低风险和高风险TETs的潜力^[17]。然而,在部分病人中不能完善增强CT的情况下,仅凭CT平扫,能提供给临床的有用信息就更少。本研究57例患者中,仅41例患者(71.9%)完成了增强CT检查。目前,影像组学能够提供肿瘤内基于灰度分布的微观异质性信息^[18-19],其优势在于不受主观影响及较好的可重复性,因此本研究选择CT平扫图像结合影像组学手段,构建预测模型。Yasaka K等^[20]人的研究表明,纹理分析模型在诊断性能方面具有优势,并比较了其视觉异质性分析的诊断性能,而纹理分析模型获得了较高的诊断性能。这可能是因为纹理分析可以量化地提取图像中的纹理特征,避免了主观因素的干扰,从而提高了诊断的客观性和准确性。任采月等^[6]基于纹理分析,建立了以MSCT征象、纹理参数联合预测模型,准确率为80.0%。但是上述研究中构建的预测模型并未使用常见算法。在本研究中,选择了SVM、MLP和LR三种常见的机器学习算法来构建诊断模型,并对它们的预测效能进行比较。结果显示,LR模型展现出较高的预测性能,其AUC值达到了0.971,优于SVM和MLP模型。LR是常用的有监督的二元分类模型,具有形式简单,模型可解释性好等优点^[21]。SVM在医学分类上应用广泛,作为有监督的二进制分类器,一般认为在二分类变量问题识别中具有优越性,能提供更好的分类性能^[22]。MLP是一种人工神经网络,其包含多个感知器^[23],它通常用于有监督学习算法,通过学习输入和输出变量之间的相关性来构建模型。一般认为SVM在预测二分类变量问题上都具有较好的效能。冯秀龙等^[24]研究中基于TETs的CT平扫图像,提取了1218个影像组学特征构建模型,结果显示14个模型中,LR、SVM模型在诊断方面均有较好的效能。而本研究提取出1649个影像组学特征构建模型,SVM、MLP模型均稍逊于LR模型,考虑可能与数据量、特征提取数量、降维方式等因素有关,尚有待探索。但LR模型在2组研究中均有较好的稳定表现。

根据病理分类的最优特征子集,所筛选出的特征多为GLSZM,考虑是由于肿瘤不同分型,肿瘤内部异质性差异不同,所以其代表的灰度值变异性差异不同。本研究,GLSZM特征对于预测胸腺瘤类型具有重要的价值,考虑可能和肿瘤细胞恶性程度增加,细胞体积及细胞致密有关。沈洁等^[25]研究发现,GLCM特征在预测B3型胸腺瘤和胸腺癌类型方面具有巨大的预测价值。He等^[26]也提出基于CT的影像组学可能成为治疗前准确区分TETs和淋巴瘤的有效工具。因此,今后研究中,仍需要进一步深入的研究来探索胸腺肿瘤病理分型与影像组学特征之间的关系。

本研究存在以下局限性:(1)本研究采用的是单中心回顾性分析,样本量较小。可能导致样本的代表性不够强,并限制了研究结果的泛化能力。(2)勾画ROI时,由于某些包块的边缘不清晰或部分容积效应的影响,造成精确度降低,特别是较小的肿瘤;(3)本研究构建模型构建,仅在TETs低、高危两组之间,缺乏胸腺各亚型间的对比;(4)由于增强病例数限制,本研究仅纳入了CT平扫图像,后期在增加病例数的基础上可考虑纳入CT增强图像。

综上所述,本研究结果初步表明三种机器学习模型具有术前预测TETs简化病理分型的应用前景,其中LR模型预测效能最高,可能有助于临床术前优化诊治方案。但模型的临床应用值及其稳定性需要进一步探索改进,以期获得更好的结果。

参考文献

- [1] Marx A, Chan JK, Coindre JM, et al. The 2015 World Health Organization classification of tumors of the thymus: continuity and changes [J]. J Thorac Oncol, 2015, 10(10): 1383-1395.
- [2] Veraa C, Janik S, Thanner J, et al. Clinical prognostic scores for patients with thymic epithelial tumors [J]. Sci Rep, 2019, 9(1): 18581.
- [3] 徐佳佳, 阴朋, 陈晓荣. CT在胸腺上皮肿瘤WHO新分类的诊断及鉴别诊断中的应用 [J]. 实用放射学杂志, 2021, 37(6): 918-922.
- [4] Tian W, Sun Y, Wu Q, et al. Surgical outcomes of 215 patients with thymic epithelial tumors: a single-center experience [J]. Thorac Cancer, 2020, 11(7): 1840-1847.
- [5] 许博, 王叶萍, 陈献国, 等. 胸腺瘤切除术后的重症肌无力发生情况及影响因素 [J]. 中华全科医学, 2020, 18(10): 1678-1681.
- [6] 任采月, 王升平, 张盛箭, 等. MSCT征象联合纹理分析在预测胸腺上皮性肿瘤WHO简化病理分型中的价值 [J]. 中国癌症杂志, 2020, 30(6): 468-474.
- [7] Jeong YJ, Lee KS, Kim J, et al. Does CT of thymic epithelial tumors enable us to differentiate histologic subtypes and predict prognosis? [J]. AJR Am J Roentgenol, 2004, 183(2): 283-289.
- [8] Yu C, Li T, Yang X, et al. Contrast-enhanced CT-based radiomics model for differentiating risk subgroups of thymic epithelial tumors [J]. BMC Med Imaging, 2022, 22(1): 37.
- [9] 李国梁, 吕明闯, 张印, 等. 多层螺旋CT对纵膈肿瘤病变患者诊断价值 [J]. 中国CT和MRI杂志, 2021, 19(07): 77-79.
- [10] 何强, 陆纪元, 谈炎欢. 多层螺旋CT扫描对B3型纵膈胸腺瘤诊断及术后复查的应用价值分析 [J]. 中国CT和MRI杂志, 2020, 18(01): 69-71.
- [11] 连海英, 朱祝. 良恶性胸腺瘤的CT诊断及鉴别诊断 [J]. 中国CT和MRI杂志, 2017, 15(11): 19-21.
- [12] Lambin P, Leijenaar R, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine [J]. Nat Rev Clin Oncol, 2017, 14(12): 749-762.
- [13] 王海波, 崔薇, 杨玮丽. 基于多序列MRI影像组学预测早期宫颈癌淋巴管侵犯的研究 [J]. 中华全科医学, 2021, 19(12): 2088-2092.
- [14] 余业洲, 赵红, 王龙胜, 等. 基于治疗前CT影像组学评估肺腺癌化疗疗效的研究 [J]. 中华全科医学, 2020, 18(4): 623-626, 701.
- [15] 杨志惠. 胸腺瘤CT影像学特点与病理的关系分析 [J]. 中国CT和MRI杂志, 2016, 14(9): 59-61.
- [16] 廖振涛, 曹显波, 于永军. 胸腺瘤合并重症肌无力的治疗 [J]. 罕少疾病杂志, 2010, 17(6): 28-29.
- [17] Chen X, Feng B, Li C, et al. A radiomics model to predict the invasiveness of thymic epithelial tumors based on contrast-enhanced computed tomography [J]. Oncol Rep, 2020, 43(4): 1256-1266.
- [18] 叶勇军, 胡玉敏, 孔春丽, 等. 基于CT平扫的影像组学列线图模型对高密度胸腺囊肿和低危胸腺瘤的鉴别诊断价值 [J]. 中国中西医结合影像学杂志, 2022, 20(4): 319-322.
- [19] 徐翠莲, 印宏坤, 伋立荣, 等. 深度学习结合影像组学的胸部CT新冠肺炎智能诊断模型研究 [J]. 罕少疾病杂志, 2021, 28(5): 16-20.
- [20] Yasaka K, Akai H, Nojima M, et al. Quantitative computed tomography texture analysis for estimating histological subtypes of thymic epithelial tumors [J]. Eur J Radiol, 2017, 92: 84-92.
- [21] 张艺梅, 丁香乾, 贺英, 等. 逻辑模型树算法性能分析与改进研究 [J]. 微型机与应用, 2014, 33(23): 25-28.
- [22] 戴宏, 符冉迪, 金炜. 基于影像组学与集成学习的脑胶质瘤分级预测 [J]. 宁波大学学报(理工版), 2021, 34(4): 28-34.
- [23] Pal SK, Mitra S. Multilayer perceptron, fuzzy sets, and classification [J]. IEEE Trans Neural Netw, 1992, 3(5): 683-697.
- [24] 冯秀龙. CT影像组学预测胸腺上皮性肿瘤简化病理分型与临床分期的研究 [D]. 空军军医大学, 2019.
- [25] 沈洁, 张怡帆, 何健. 基于增强CT影像组学特征鉴别诊断B3型胸腺瘤与胸腺癌 [J]. 临床肺科杂志, 2022, 27(10): 1539-1543.
- [26] He W, Xia C, Chen X, et al. Computed tomography-based radiomics for differentiation of thymic epithelial tumors and lymphomas in anterior mediastinum [J]. Front Oncol, 2022, 12: 869982.

(收稿日期: 2023-03-21)

(校对编辑: 姚丽娜)