

Differentiation of nodule or mass type pulmonary cryptococcosis and lung adenocarcinoma, lung tuberculosis based on plain CT scanning radiomics models

FAN Mengsi¹, ZHAO Hong^{1*}, CAO Hanbo², YU Yezhou¹, ZOU Liwei¹, DUAN Shaofeng³

(1. Department of Radiology, the Second Affiliated Hospital of Anhui Medical University, Hefei 230601, China; 2. Diagnostic Radiology Center, Zhoushan Hospital, Zhoushan 316000, China; 3. GE Healthcare China, Shanghai 210000, China)

[Abstract] **Objective** To explore the feasibility of differential diagnosis of nodule or mass pulmonary cryptococcosis (PC), lung adenocarcinoma and lung tuberculosis (TB) based on plain CT scanning radiomics prediction models. **Methods**

Plain CT data of 28 patients with nodule or mass type PC, 30 with pulmonary adenocarcinoma and 26 with lung TB were retrospectively analyzed. The texture features of lesions on CT images were extracted and selected to establish the optimized texture parameters between PC and lung adenocarcinoma, also between PC and lung TB. Then all samples were divided into training set and testing set according to ratio of 7:3. The random forest method was used to establish prediction model with the optimized texture parameters, and the model was used to evaluate training set data and verified with testing set data. The corresponding ROC curve was drawn, so as to evaluate the model's differential diagnosis efficiency. **Results** After screening, 7 optimized feature parameters were obtained between PC and lung adenocarcinoma, while 4 were obtained between PC and lung TB. The AUC, sensitivity, specificity, accuracy of the model for differentiating PC from lung adenocarcinoma was 0.96, 1.00, 0.78 and 0.89, respectively, while for differentiating PC from lung TB was 0.99, 0.88, 0.89 and 0.88, respectively. **Conclusion** Radiomics models based plain CT scanning can be used for differentiating and diagnosing nodule or mass PC from lung adenocarcinoma and lung TB.

[Keywords] lung; cryptococcosis; lung neoplasms; tuberculosis; artificial intelligence; radiomics; tomography, X-ray computed

DOI:10.13929/j.issn.1003-3289.2020.06.000

基于 CT 平扫影像组学模型鉴别结节/肿块型肺隐球菌病及肺腺癌与肺结核

樊梦思¹, 赵红^{1*}, 曹捍波², 余业洲¹, 邹立巍¹, 段绍峰³

(1. 安徽医科大学第二附属医院放射科, 安徽 合肥 230601; 2. 浙江大学舟山医院放射诊断中心, 浙江 舟山 316000; 3. GE Healthcare China, 上海 210000)

[摘要] **目的** 探讨基于 CT 平扫影像组学预测模型鉴别诊断结节/肿块型肺隐球菌(PC)与肺腺癌、肺结核(TB)的可行性。**方法** 回顾性分析 28 例结节/肿块型 PC 和 30 例肺腺癌、26 例 TB 的平扫 CT 资料,提取病灶纹理特征,对其进行特征选择,获得 PC 组与肺腺癌组、PC 组与肺 TB 组之间存在显著差异的特征参数。按 7:3 比例将所有样本分为训练集合测试集,采用随机森林法以较优特征参数建立预测模型,对训练集数据进行评估,之后于测试集数据进行验证;绘制相

[第一作者] 樊梦思(1993—),女,安徽蚌埠人,在读硕士,医师。研究方向:胸部影像诊断学。E-mail: 1048273239@qq.com

[通信作者] 赵红,安徽医科大学第二附属医院放射科,230601。E-mail: 178331090@qq.com

[收稿日期] 2019-06-18 **[修回日期]** 2020-01-09

应 ROC 曲线,评估模型的鉴别诊断效能。**结果** 针对 PC 和肺腺癌、PC 和肺 TB 分别获得 7 个和 4 个较优纹理特征参数。测试集验证结果显示模型鉴别 PC 与肺腺癌以及 PC 与肺 TB 的 AUC、敏感度、特异度、准确率分别为 0.96、1.00、0.78、0.89 及 0.99、0.88、0.89、0.88。**结论** 基于 CT 平扫图像影像组学可用于鉴别诊断结节/肿块型 PC 及肺腺癌与肺 TB。

[关键词] 肺; 隐球菌病; 肺肿瘤; 肺结核; 人工智能; 影像组学; 体层摄影术, X 线计算机

[中图分类号] R563; R814.42 **[文献标识码]** A **[文章编号]** 1003-3289(2020)06-0000-00

肺隐球菌病(pulmonary cryptococcosis, PC)是由新型隐球菌和格特(Gattii)隐球菌等引起的亚急性/慢性肺部真菌感染性疾病,主要发生于免疫功能受损者,也可见于免疫正常人群^[1-2],后者常表现为单发或多发肿块或结节^[3]。PC 患者临床症状缺乏特异性,血清隐球菌荚膜多糖抗原(cryptococcal capsular polysaccharide antigen, CrAg)检测对血清 HIV 检测阴性的 PC 患者敏感性不高^[4]。影像学上结节/肿块型 PC 与周围型肺癌、肺结核(tuberculosis, TB)存在部分重叠,临床易发生误诊。隐球菌对脑膜及神经组织有明显亲和性,可透过血脑屏障造成严重中枢神经系统感染^[5],因此及早确诊非常重要。影像组学采用人工智能提取数据特征化算法,将包含反映潜在病理生理特征的图像转变为高维度数据,之后对数据进行分析,以寻找能综合评价肿瘤表型的相关特征信息^[6-7],已广泛用于肿瘤诊断、分级、预后评估及预测复发等^[8-11]。本研究观察以影像组学预测模型鉴别诊断结节/肿块型 PC 与肺腺癌、肺 TB 的可行性。

1 资料与方法

1.1 一般资料 回顾性分析 2016 年 10 月—2019 年 5 月 28 例经病理证实的结节/肿块型 PC(PC 组)、30 例病理确诊肺腺癌(肺腺癌组)及 26 例肺 TB(TB 组)患者的胸部 CT 平扫资料。PC 组男 18 例,女 10 例,年龄 15~81 岁,中位年龄 53;血清 HIV 均为阴性。肺腺癌组男 14 例,女 16 例,年龄 31~82 岁,平均(61.9±12.7)岁。肺 TB 组男 20 例,女 6 例,年龄 20~71 岁,中位年龄 52.5 岁;经穿刺活检病理证实为肉

芽肿性病变,结合抗酸染色、实验室检查及临床资料确诊为肺 TB。所有患者接受胸部 CT 检查前均未经任何治疗。

1.2 仪器与方法 采用 GE LightSpeed 64 排 VCT 扫描仪,管电压 120 kV,管电流 164~320 mA,采用螺旋扫描模式,螺距 1.2,准直 128×0.625 mm;重建层厚 5 mm,层间隔 5 mm,行胸部 CT 平扫。

1.3 特征提取 将 CT 平扫软组织窗(窗宽:350 HU,窗位:50 HU)图像导入 ITK-SNAP(www.itk-snap.org)软件,于病变较大层面手动勾画 ROI,面积 1.0~2.5 cm²,尽量避免钙化、空洞、坏死区域(图 1~3)。将原始图像及 ROI 导入 A. K (Artificial Intelligence Kit) Version V3.2.0. R 后处理软件,调整体素为 1 mm×1 mm×1 mm,之后采用线性插值法对其重采样,以高斯滤波函数去噪、灰度离散化,对原始图像进行标准化重建及融合。计算病灶 ROI 的纹理特征,提取包括灰度直方图(Histogram)、灰度共生矩阵(gray level co-occurrence matrix, GLCM)、游程矩阵(run length matrix, RLM)、灰度区域大小矩阵(gray level size zone matrix, GLSZM)及 Haralick 特征在内的共 386 个特征。

1.4 图像和统计学分析 采用 R 软件 3.5.1 版(www.r-project.org)归纳上述数据,删除异常值所在列。按 7:3 比例将所有样本分为训练集测试集,训练集中含 19 例 PC、21 例肺腺癌、18 例肺 TB,测试集分别为 9、9、8 例。以单因素方差分析和秩和检验(ANOVA+MW)、Spearman 相关系数、最小绝对收缩

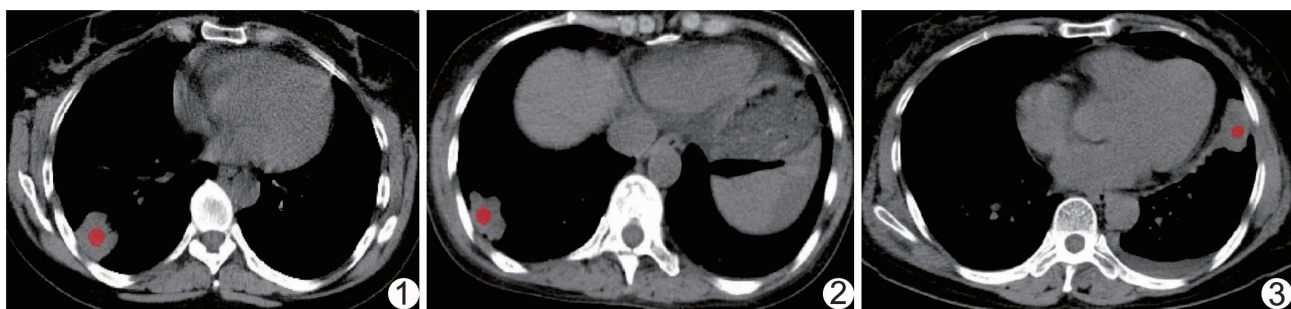


图 1 患者女,53 岁,PC 轴位平扫 CT 图(红色为 ROI)

图 2 患者女,40 岁,肺腺癌 轴位平扫 CT 图(红色为 ROI)

图 3 患者

女,54 岁,肺 TB 轴位平扫 CT 图(红色为 ROI)

和选择算子 (least absolute shrinkage and selection operator, Lasso) 回归 3 种方法联合进行特征选择和降维, 2 特征的 Spearman 相关系数设为 0.9, 获得 PC 组与肺腺癌组、PC 组与肺 TB 组之间存在显著差异的较优特征参数。采用随机森林法分别建立基于影像组学特征的预测模型, 并以训练集进行评估, 之后于测试集进行验证; 绘制相应的 ROC 曲线, 以 AUC、敏感度、特异度及准确率评价其鉴别效能。

2 结果

本组实性病变直径约 1~5 cm, 图像均无明显呼吸及光束硬化伪影干扰。

经特征选择获得针对 PC 与肺腺癌的 7 个较优特征, 分别为 Histogram 的第 10 百分位数 (Percentile10)、GLCM 的逆差距 (inverse difference moment)、逆差距_所有方向_步长 1_标准差 (inverse difference moment all direction_offset1_SD)、熵_所有方向_步长 7 (entropy_all direction_offset7)、熵_所有方向_步长 7_标准差 (entropy_all direction_offset7_SD)、集群阴影_角度 135_步长 1 (cluster shade_angle135_offset1) 及 RLM 的长行程优势_角度 135_步长 7 (long run emphasis_angle135_offset7)。采用随机森林法建立预测模型, 并进行评估和验证, ROC 曲线 (图 4A) 示 AUC 为 0.96 [95%CI (0.87, 1.00)], 截断值 0.97, 鉴别 PC 与肺腺癌的敏感度 1.00、特异度 0.78、准确率 0.89 (表 1)。随机森林模型自变量重要性排序见图 5A。

经特征选择, 针对 PC 与肺 TB 获得 4 个较优特征参数, 即 Histogram 的第 10 百分位数, GLCM 的相关性_角度 90_步长 4 (correlation_angle90_offset4), RLM 的长行程优势_所有方向_步长 1 (long run emphasis_all direction_offset1), 长行程优势_角度 0_步长 1 (long run emphasis_angle0_offset1)。随机森林建模后于训练集进行评估, 于测试集进行验证, ROC 曲线 (图 4B) 示 AUC 为 0.99 [95%CI (0.93, 1.00)], 截断值 0.57, 鉴别 PC 与肺 TB 的敏感度

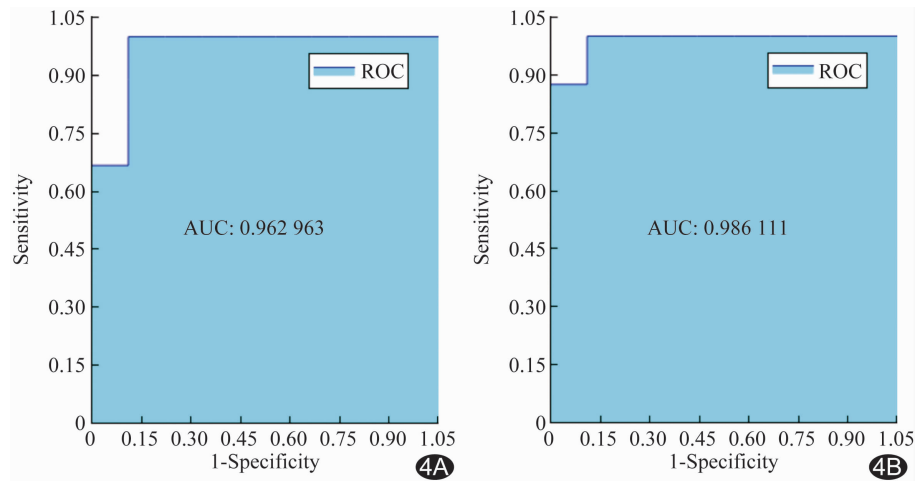


图 2 预测模型鉴别 PC 与肺腺癌(A)及 PC 与肺 TB(B)测试集的 ROC 曲线

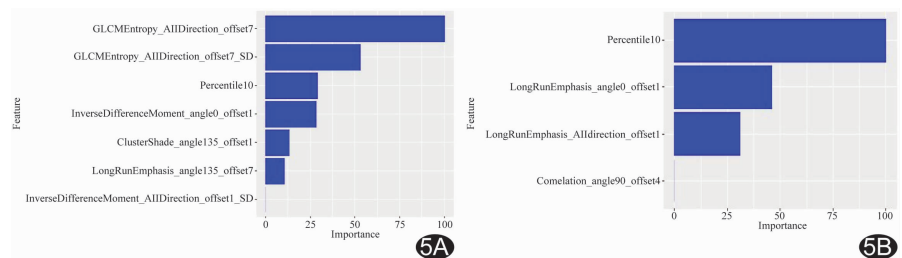


图 3 随机森林模型自变量重要性示意图 A. PC 组与肺腺癌组; B. PC 组与肺 TB 组

0.88、特异度 0.89、准确率 0.88 (表 1)。随机森林模型自变量重要性排序见图 5B。

表 1 测试集中随机森林模型鉴别诊断 PC 与肺腺癌及 PC 与肺 TB 的效能

鉴别项	AUC	敏感度	特异度	准确率	F1 评分
PC 与肺腺癌	0.96	1.00	0.78	0.89	0.90
PC 与肺 TB	0.99	0.88	0.89	0.88	0.88

3 讨论

PC 主要为通过呼吸道吸入隐球菌孢子至肺泡后引起, 易感人群多为获得性免疫缺陷综合征 (acquired immune deficiency syndrome, AIDS)、器官移植术、长期服用广谱抗生素及免疫力低下者等, 也可发生于免疫功能正常人群, 近年来发病率呈上升趋势, 尤其在无免疫功能缺陷或基础疾病人群中^[1-2]。血清 CrAg 检测对于 HIV 阴性 PC 患者的敏感度仅为 25%~56%^[4], 导致诊断困难。PC 临床症状及影像学表现均缺乏特异性, 免疫功能正常者多以结节或肿块为主, 多发于两肺下叶外周带, 易误诊为肺腺癌。PC 早期病理学表现为胶样病变, 免疫功能正常者后期进一步形成肉芽肿性病变, 易误诊为 TB 性肉芽肿^[12]。对于肺

腺癌,治疗方式主要包括手术、放射及化学治疗,而肺 TB 一般需 6~8 个月抗结核治疗,对免疫功能正常 PC 患者则需要应用氟康唑或伊曲康唑进行 3~12 个月的长效治疗,且隐球菌易侵入中枢神经系统,导致隐球菌性脑膜炎,病程长,预后差。

目前用于建立基于影像组学特征的预测模型的常用机器学习方法有随机森林、logistic 回归模型、支持向量机、人工神经网络、聚类分析及“leave-one out”交叉验证等。随机森林是机器学习中的一个决策树概念,表示假设为连续的“if-then”,训练时生成大量决策树,结合每棵树的个体决策获得最优分类,具有相对较低的过度拟合倾向,在影像组学机器模型中的稳定性和预测性较高^[13-14]。

本研究采用随机森林法筛选较优特征参数,分别建立预测模型并于训练集数据进行训练,再于测试集数据进行验证,结果示模型鉴别 PC 与肺腺癌的 AUC、敏感度、特异度、准确率分别为 0.96、1.00、0.78 及 0.89,鉴别 PC 与肺 TB 时分别为 0.99、0.88、0.89 及 0.88,表明基于随机森林的 CT 平扫图像影像组学可用于鉴别结节/肿块型 PC 与肺腺癌、肺 TB 性肉芽肿,并具有良好的诊断性能。经筛选可用于鉴别 PC 与肺 TB 的 7 个较优特征包括 Histogram 的第 10 百分位数,GLCM 的熵、逆差距、集群阴影、相关性以及 RLM 的长行程优势。第 10 百分位数指样本观测灰度列中有 10% 的灰度值小于或等于该值。熵描述共生矩阵的复杂性,其值越大表示共生矩阵越复杂,提示病灶越复杂、异质性越大。逆差距指病灶的同质性,其值越大则变化越小、局部越均匀。集群阴影与 GLCM 的一致性和偏度有关,其值越小,偏度越小,即差异性较小。相关性度量 GLCM 元素的线性依赖关系,若矩阵元素值相差很大,则相关性值小。长行程优势描述图像粗糙度或平滑度,在光滑图像上其值较大。本研究 PC 组第 10 百分位数高于肺腺癌组及肺 TB 组,意味着 CT 图像中表现为低灰度值的病灶炎性坏死区域相对较少,即低值区灰度值对 PC 有一定提示作用。PC 组逆差距、长行程优势高于肺腺癌组,而熵、集群阴影低于肺腺癌组,提示肺腺癌图像灰度不均匀。PC 组相关性、长行程优势高于 TB 组,说明 PC 密度较 TB 病变更均匀,其局部像素灰度级相差小。本研究结果示 PC 病灶实性部分的图像灰度较肺腺癌和肺 TB 更均匀、集中,可能与 PC 肉芽肿病灶内细胞呈弥漫性分布、很少形成结节,且坏死不彻底、存在网状纤维支架有关^[15]。本研究基于 CT 平扫图像,未采集 CT 增强

图像纹理特征,结果显示基于随机森林的影像组学方法具有较好的鉴别诊断效能,有利于患者免于接受增强 CT,从而减低辐射剂量。

综上所述,基于随机森林的影像组学方法可用于鉴别诊断结节/肿块型 PC 与肺腺癌、肺 TB,为临床准确诊断和个体化治疗提供依据。本研究尚存不足:①样本量少;②仅采集平扫软组织窗单层局局部病变图像,难以提取病变全部信息;③去除特征冗余时仅选择组间差异最大的特征,可能遗漏有潜在价值的特征参数。

[参考文献]

- [1] ZHANG Y, LI N, ZHANG Y, et al. Clinical analysis of 76 patients pathologically diagnosed with pulmonary cryptococcosis [J]. Eur Respir J, 2012, 40(5): 1191-1200.
- [2] JARVIS J N, HARRISON T S. Pulmonary cryptococcosis [J]. Semin Respir Crit Care Med, 2008, 29(2): 141-150.
- [3] LINDELL R M, HARTMAN T E, NADROUS H F, et al. Pulmonary cryptococcosis: CT findings in immunocompetent patients [J]. Radiology, 2005, 236(1): 326-331.
- [4] ZHOU Y, LIN PC, YE JR, et al. The performance of serum cryptococcal capsular polysaccharide antigen test, histopathology and culture of the lung tissue for diagnosis of pulmonary cryptococcosis in patients without HIV infection [J]. Infect Drug Resist, 2018, 11: 2483-2490.
- [5] 温海. 隐球菌感染诊治专家共识 [J]. 中国真菌学杂志, 2010, 5(2): 65-68, 86.
- [6] GILLIES R J, KINAHAN P E, HRICAK H. Radiomics: Images are more than pictures, they are data [J]. Radiology, 2016, 278(2): 563-577.
- [7] van GRIETHUYSEN J J M, FEDOROV A, PARMAR C, et al. Computational radiomics system to decode the radiographic phenotype [J]. Cancer Res, 2017, 77(21): e104-e107.
- [8] CONSTANZO J, WEI L, TSENG H H, et al. Radiomics in precision medicine for lung cancer [J]. Transl Lung Cancer Res, 2017, 6(6): 635-647.
- [9] KIM S, SHIN J, KIM D Y, et al. Radiomics on gadoteric acid-enhanced magnetic resonance imaging for prediction of postoperative early and late recurrence of single hepatocellular carcinoma [J]. Clin Cancer Res, 2019, 25(13): 3847-3855.
- [10] ARTZI M, BRESSLER I, BEN BASHAT D. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis [J]. J Magn Reson Imaging, 2019, 50(2): 519-528.
- [11] 王平, 裴旭, 殷小平, 等. 基于增强 CT 影像组学模型鉴别肾透明细胞癌与非透明细胞癌 [J]. 中国医学影像技术, 2019, 35(11): 1689-1692.
- [12] 张丽妍, 张冬蕾, 谭晓明, 等. 以坏死性肉芽肿为表现的肺隐球菌病 28 例临床影像分析 [J]. 中国真菌学杂志, 2017, 12(4):

- 212-215.
- [13] ZHANG B, HE X, OUYANG F, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma[J]. Cancer Lett, 2017, 403:21-27.
- [14] PARMAR C, GROSSMANN P, BUSSINK J, et al. Machine learning methods for quantitative radiomic biomarkers[J]. Sci Rep, 2015, 5:13087.
- [15] 易祥华,孔洁,朱美芳,等.原发性肺隐球菌病的病理诊断和超微结构观察[J].中华病理学杂志,2004,33(5):424-428.