

· 综述 ·

影像组学在预测非小细胞肺癌表皮生长因子受体基因突变中的应用

杨露, 欧阳治强, 单海燕, 陈龙, 朱钰, 褚吉祥, 廖承德

【摘要】 非小细胞肺癌表皮生长因子受体(EGFR)基因突变与否对患者治疗方案的选择以及预后评估具有重要的作用。目前,临床上主要依靠有创手段获得肿瘤的病理组织并进行分子检测来确定非小细胞肺癌患者的 EGFR 基因突变状况,但由于采样和分析的局限性,检测到的结果可能无法完全反应病灶部位肿瘤内和肿瘤间的异质性。影像组学可从影像图像中高通量地提取并分析大量影像组学特征数据,并将其与肿瘤基因表达相关联,可为临床诊断和治疗提供更多的参考依据。本文将对影像组学辅助诊断非小细胞肺癌 EGFR 基因突变的研究进行综述。

【关键词】 影像组学; 影像基因组学; 癌,非小细胞肺; 表皮生长因子受体基因; 突变

【中图分类号】 R730.26; R394.3; R394 **【文献标识码】** A

【文章编号】 1000-0313(2023)06-0783-05

DOI:10.13609/j.cnki.1000-0313.2023.06.020

开放科学(资源服务)标识码(OSID):



肺癌是世界上死亡率最高的恶性肿瘤^[1],非小细胞肺癌(non-small cell lung cancer, NSCLC)占全部肺癌病理类型的 85% 以上^[2]。近年来,第三代酪氨酸激酶抑制剂(tyrosine kinase inhibitors, TKIs)奥希替尼作为新型靶向药物,显著延长了完全切除的 I B~III A 期表皮生长因子受体(epidermal growth factor receptor, EGFR)基因突变阳性的 NSCLC 患者的无病生存期(disease-free survival, DFS)^[3],此外,奥希替尼作为治疗 EGFR 基因突变的晚期 NSCLC 患者的一线治疗药物,能使患者的神经系统进展或死亡风险降低 52%^[4]。然而,在治疗前确定患者是否存在 EGFR 基因突变及其突变亚型是临床开展靶向干预的重要前提。目前,对 NSCLC 患者 EGFR 基因型的检测主要是利用肿瘤病理组织或细胞学标本进行基因测序,基因检测技术在临床实践中虽然已经逐步开展,但由于费用高昂、检测周期长、准确性易受肿瘤异质性干扰等缺点^[5],临床亟需一种检测手段来辅助对 NSCLC 患者 EGFR 基因表型做出客观、准确的评价^[6,7]。影像组学可通过提取并分析海量的影像组学特征数据,选择那些与临床、病理、分子和遗传特征最显著相关的参数,通过统计分析建立相应的诊断或预测模型,达到辅

助临床诊断和预后评估的目的(图 1)^[8]。近年来,利用影像组学预测 NSCLC 患者基因突变的研究甚多,在影像组学实施的各个环节,包括图像获取、图像分割、特征提取、模型构建及分析都有可能对最终的结果产生影响。因此,本文将从基于 CT、PET/CT、MRI 影像组学实施的各个环节着手,综述其在预测 NSCLC 患者 EGFR 基因突变方面的研究进展。

基于 CT 影像组学

由于 CT 图像中隐藏着许多人眼无法识别的肿瘤生物学信息,利用一般的语义特征来预测基因型已不再满足临床诊断的需求,基于 CT 影像组学的研究方法在预测 NSCLC 患者 EGFR 表型中已取得一定成果。然而,在影像组学实施的各个环节都有可能对预测的客观性和准确性造成影响。

1. 扫描参数及重建方法对基于 CT 影像组学预测模型的影响

在保持生物信息的同时实现高度的可重复性是当前影像组学研究所面临的主要挑战之一。非常规成像变量如重建核的不一致会显著影响影像组学特征的泛化性。针对这一问题,Yoon 等^[9]利用卷积神经网络(convolutional neural networks, CNN)解决了 CT 重建核对影像组学研究的混杂影响,证明基于 CNN 的方法能够协调异质重建核对医学图像同质性的影响,在提高特征可重复性和泛化能力方面显示出了巨大的潜力。此外,有学者^[10]用贝蒂数(Betti numbers, BNs)鉴定了来自 3 个不同国家 NSCLC 患者的 EGFR 基因突变状况,结果表明基于 BNs 的方法对图像扫描

作者单位: 650118 昆明,昆明医科大学第三附属医院/云南省肿瘤医院放射科(杨露、欧阳治强、单海燕、朱钰、褚吉祥); PET/CT 中心(陈龙); 650051 昆明,昆明医科大学附属延安医院/昆明市延安医院(廖承德)

作者简介: 杨露(1997-),女,云南玉溪人,硕士研究生,主要从事肿瘤影像诊断工作。

通讯作者: 廖承德, E-mail: 846681160@qq.com

基金项目: 云南省卫生健康委员会-医学学科带头人培养计划(D-2018009); 云南省应用基础研究面上项目(202101AT070050)

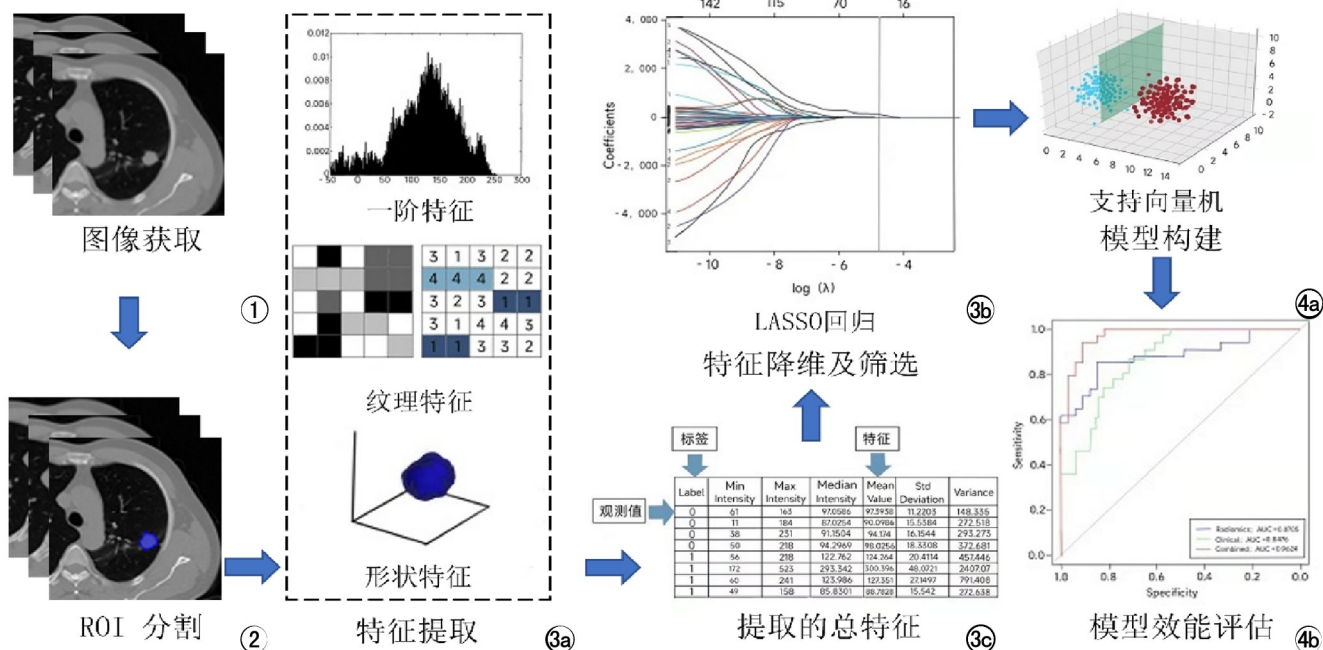


图 1~4 影像组学研究基本流程。图 1 影像图像的获取与预处理:利用体素水平的重采样、高斯滤波、空间配准等手段重建符合研究标准的医学影像。图 2 图像分割:手动或自动分割 ROI 以便对医学影像中的目标区域进行有针对性的处理。图 3 特征提取及降维:提取的影像组学特征主要有 4 大类,即形状特征、一阶特征、纹理特征、小波变换特征,因组学特征数量庞大、信息冗余,需使用最小收缩和选择算子 (least absolute shrinkage and selection operator, LASSO)、最大相关-最小冗余 (max-relevance and min-redundancy, mRMR) 以及 Pearson 相关性分析等方法对其进行降维处理以获取少数真正关键的组学特征。图 4 模型构建及分析:利用筛选出来的影像组学特征训练和验证普通机器学习或深度机器学习算法,从而建立分类或预测模型。

仪/扫描参数的变化具有鲁棒性,由此提出了一种新型稳健的影像基因组学研究方法。目前,对于 CT 图像获取的标准化仍然面临诸多的挑战,如何处理扫描参数不一致所带来的影响仍然是未来研究关注的重点。

2. 获取不同类型 CT 图像对预测模型的影响

不同类型的 CT 图像(平扫或增强)所反映的肿瘤生物学信息各有不同,利用影像组学从不同类型的 CT 图像中提取影像组学特征进行分析可能对后续模型的建立产生不同的影响。

Tu 等^[11]基于 404 例 NSCLC 患者术前 CT 平扫 (non-contrast-enhanced CT, NE-CT) 图像构建了用于预测 EGFR 基因突变的影像组学模型,与临床模型相比,进一步整合了肿瘤最大直径、位置、患者性别和影像组学特征的联合模型,其预测效能有了进一步的提升[训练队列和验证队列曲线下面积 (area under curve, AUC) 分别为 0.798 和 0.818],但该研究所建立的 EGFR 基因预测模型是基于患者术前 CT 平扫图像,对肿瘤异质性及血管生成等生物学特性的分辨力有限,预测效能有待进一步的提高。随后,HUO 等^[12]基于 608 例 NSCLC 患者术前胸部 CT 增强 (contrast-

enhanced CT, CE-CT) 图像构建了用于预测 EGFR 基因表型的梯度树增强 (extreme gradient tree boosting, XGB) 模型,与先前基于平扫图像的预测模型相比,该模型的预测准确性明显更高 (AUC 在训练集和验证集分别为 0.969, 0.886),其原因可能为 CT 增强图像能够提供更多与肿瘤血液供应、内部异质性等相关的影像信息,并且更有利于影像科医师对肿瘤边界进行准确界定。

之前的研究都仅限于对基于 CT 平扫或增强图像构建的影像组学模型进行孤立的比较,而未真正考虑将两者整合。近来,有研究将 NSCLC 患者术前 NE-CT 图像和 CE-CT 图像间无显著差异的影像组学特征定义为一般影像组学特征,并通过与吸烟史、肺气肿和间质性肺疾病等临床特征结合后构建了用于预测 EGFR 基因表型的模型,该模型无论是在 NE-CT 图像测试中还是在 CE-CT 图像测试中都取得了较高的预测效能^[13]。综上,基于不同类型 CT 图像的影像组学模型在预测 NSCLC 患者 EGFR 基因表型方面已取得了一定的进展,但是图像的获取离不开扫描的技术,各个研究之间扫描技术参差不齐,未来有望探索更好

的扫描方案来获得质量较佳的图像,以提高预测模型的实用价值。

3. 基于 CT 影像组学构建不同的人工智能模型

近年来,人工智能(artificial intelligence, AI)特别是深度学习技术已被越来越多地应用于医学图像的解读。在肺癌研究领域,由于患者术前胸部 CT 图像易获取以及 AI 的快速发展^[14],影像组学也得到了飞速发展。

最近,有研究者开发并优化了一个多任务人工智能模型包括深度学习(deep learning, DL)模型、影像组学(radiomics, RA)模型和结合 DL、RA 和临床特征的联合(joint, JO)模型。结果显示 JO 模型(AUC 为 0.928)在预测 EGFR 基因突变时的表现明显优于 DL 模型(AUC 为 0.842)^[15]。此外,有研究构建了一个用来预测 EGFR 基因表型和程序性死亡配体 1(programmed death ligand 1, PD-L1)表达状态的 3D-CNN 模型^[16],进一步加入梯度加权类激活映射后实现了预测模型的可视化,使得输入的 3D 图像可充分显示病变的特征,为 CNN 提供更加丰富的图像信息。随后有学者发现在 CNN 模型的基础上添加一个挤压与激发层则能进一步提升模型的效能^[17]。

总之,通过分层网络学习图像中的抽象特征,建立输入数据和输出标签之间映射的 AI 模型在预测 EGFR 基因表型时具备独特的优势,但值得注意的是尽管研究者们正努力开发一些基于 CNN 模型的可视化技术,但对深度学习模型本身所属的“黑箱”理论进行合理解释仍然面临诸多挑战^[18]。

基于 PET/CT 影像组学

随着分子影像技术的不断发展,正电子发射成像计算机断层成像(positron emission tomography/computed tomography, PET/CT)已被广泛应用于肺癌的临床诊断与预后评价。PET/CT 可同时反映病灶的形态学信息以及代谢情况,相较于常规 CT 检查能为诊断者提供更多有关肿瘤生物学行为的影像信息。

1. 基于不同分子探针获得的图像提取影像组学特征

目前临床应用最为广泛的代谢显像剂为¹⁸F-FDG,基于不同的分子探针获得的图像是否对预测模型有影响已有部分研究开展。Shiri 等^[19]从 150 例 NSCLC 患者术前的¹⁸F-FDG PET/CT、低剂量 CT 和 CE-CT 图像中提取的影像组学特征预测 EGFR 基因突变的 AUC 效能相较于峰值标准摄取值(peak standard uptake value, SUV peak)(AUC=0.69)提高到 0.75,此外,基于多变量机器学习模型在预测 EGFR 基因突变

中,AUC 进一步提高至 0.82,证明 PET/CT 的加入能使 EGFR 基因突变预测模型获得更多有关 NSCLC 肿瘤生物学行为的影像信息,进而提高预测准确性。当然,经过纳米技术加工的分子影像学探针所提供的影像信息则更为丰富且关键,最新的一项研究发现基于一种新的分子探针¹⁸F-MPG(¹⁸F-monopropylene glycol, ¹⁸F-MPG)PET 图像提取的影像组学特征在预测原发性肺癌和转移性肺癌 EGFR 基因突变时的准确率明显优于基于¹⁸F-FDG PET 图像所提取的特征(90%和 89.66% vs 76%和 82.75%)^[20],说明两种 PET 图像与 EGFR 基因表达间的相关性存在差异,但目前基于¹⁸F-MPG PET 的组学研究相对较少,相关预测模型的深度释义还有待进一步的探索。

2. 特征协调提高基于 PET/CT 影像组学预测模型的准确性

由于受到来自图像采集协议、处理/重建设置和成像扫描仪变化的影响,往往导致提取的影像组学特征具有可变性^[21,22],这种可变性将导致影像组学模型不可再现和不可靠。为了克服这些挑战,有研究者提出协调方法^[23],将 136 例经病理证实的 NSCLC 患者协调前后的 PET/CT 影像组学特征用于构建单变量和多变量预测模型,结果表明协调后的多变量预测模型预测 EGFR 基因突变的 AUC 平均范围从 0.87~0.90 增加到 0.92~0.94。鉴于此,后续研究可采用上述协调方法,整合来自不同中心的数据,基于更大且更协调的数据集开发泛化能力更强、精确度更高的预测模型。

3. 基于 PET/CT 影像组学构建的人工智能模型

随着大数据时代的到来,基于 PET/CT 影像组学已涌现出许多人工智能模型。Mu 等^[24]开发了一个基于¹⁸F-FDG PET/CT 的深度学习模型,使用该模型生成的深度学习评分(deep learning score, DLS)在训练集、内部验证集和外部测试集中预测 NSCLC 患者 EGFR 基因突变的准确率分别为 81.1%、82.8%和 78.5%。无独有偶,Koyasu 等^[25]使用随机森林(random forest, RF)和 XGB 两种机器学习算法构建了 NSCLC 患者组织学亚型和 EGFR 基因突变的预测模型,与 RF 模型相比 XGB 能够完美拟合多种成像特征信息,在一定程度上发挥了更好的预测能力。当前,基于 PET/CT 影像组学的 AI 模型在预测 NSCLC 相关分子表型方面已初露头角,有望成为付诸临床实践的先头兵。

基于 MRI 影像组学构建预测 NSCLC 患者 EGFR 突变的模型

MRI 在肺癌患者中主要用于显示颅脑、脊柱等部位的转移瘤,而用于肺部原发灶的成像较少。肿瘤在

转移过程中原发肿瘤和转移瘤也可能存在时间和空间的异质性,获取肿瘤不同部位的图像来进行预测分析对后续指导治疗同等重要。

一项回顾性研究^[26]中研究人员基于肺癌脑转移瘤 T₂ 液体衰减反转恢复序列(T₂ fluid attenuated inversion recovery, T₂ FLAIR)图像最终选取了 9 个影像组学特征并用于预测模型的构建,该模型在预测 EGFR 基因突变中显示出良好的辨别力(验证队列 AUC 为 0.987,独立测试队列 AUC 为 0.871)。另一项从基于 110 名确诊为原发性肺腺癌患者胸椎转移瘤的 T₁WI(T₁-weighted, T₁WI)、T₂WI(T₂-weighted, T₂WI)和 T₂ 加权脂肪抑制(T₂-weighted fat-suppressed, T₂FS)3 种磁共振序列中提取影像组学特征并结合患者吸烟史数据构建了预测 EGFR 基因突变的诺谟图^[27],该诺谟图在训练队列和验证队列中同样展现出了较好的预测能力,AUC 分别为 0.816 和 0.788。唐等^[28]从 74 例肺腺癌患者肺部原发灶多序列 MRI 图像[T₂WI,扩散加权成像(diffusion-weighted imaging, DWI)序列,表观扩散系数(apparent diffusion coefficient, ADC)序列]中筛选了 16 个影像组学特征用于构建 EGFR 基因表型预测模型,其预测 EGFR 基因突变的敏感度 53.1%,特异度 92.9%,符合率 75.7%,AUC 为 0.826,结合性别因素构建的联合模型预测符合率提高到 78.9%。先前的多项研究皆证明基于多部位多参数 MRI 的影像组学在预测 NSCLC 患者 EGFR 基因突变方面存在巨大的潜能,如何将原发肿瘤及转移瘤基因表型相关联形成精确度更高、实用性更强的双模态/多模态诊断模型是今后研究的重点与难点。

不足与展望

目前,利用影像组学预测 NSCLC 患者 EGFR 基因突变的研究尚处于初步阶段,无创性预测肿瘤患者的基因表型仍然是未来的希望与挑战。在影像组学应用于临床实践之前,还有很多的工作要做。首先,所有的生物标记物,无论是从 CT、PET/CT、MRI 的数据中发展出来的,还是结合在一起都面临着分析方法和分析的潜在偏倚来源。影像组学的独特之处是在图像获取、处理和成像特征量化过程中可能存在的误差和偏差。标准化也不总是可行的,未来的努力还需要广泛访问来自真实世界数据的大量图像以转化为临床实践。其次,大部分研究的焦点都集中于病灶本身,未来研究的任务是将更多瘤周图像信息纳入诊断模型的构建;第三,研究样本量有限,且在研究样本量相对于所评估的影像学特征数量较少的时候,有时常常导致预测模型的过拟合,这也是各个研究之间预测模型在比

较时存在的问题之一。后续研究的重点在于扩大样本量以及进行多中心的研究来提高影像组学模型的泛化性与实用性。

影像组学是医学、计算机科学和统计学的交叉领域。影像组学分析需要一个多学科的研究团队,包括多个专业领域的个人来解释对临床有意义的结果。随着对精准医疗的需求越来越高,通过不断优化影像组学实施的各个环节,其在预测基因突变方面仍然充满着无限希望。

参考文献:

- [1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.
- [2] Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: Epidemiology, screening, diagnosis, and treatment [J]. Mayo Clin Proc, 2019, 94(8): 1623-1640.
- [3] Wu YL, Tsuboi M, He J, et al. Osimertinib in resected EGFR-mutated non-small-cell lung cancer[J]. N Engl J Med, 2020, 383(18): 1711-1723.
- [4] Reungwetwattana T, Nakagawa K, Cho BC, et al. CNS response to osimertinib versus standard epidermal growth factor receptor tyrosine kinase inhibitors in patients with untreated EGFR-mutated advanced non-small-cell lung cancer[J]. J Clin Oncol, 2018, 36: 3290-3297. DOI: 10.1200/JCO.2018.78.3118
- [5] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data[J]. Radiology, 2016, 278(2): 563-577.
- [6] Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review[J]. JAMA Oncol, 2016, 2(12): 1636-1642.
- [7] Napel S, Mu W, Jardim-Perassi BV, et al. Quantitative imaging of cancer in the postgenomic era: radio (geno) mics, deep learning, and habitats[J]. Cancer, 2018, 124(24): 4633-4649.
- [8] Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis[J]. Eur J Cancer, 2012, 48(4): 441-446.
- [9] Yoon JH, Sun SH, Xiao M, et al. Convolutional neural Network addresses the confounding impact of CT reconstruction kernels on radiomics studies[J]. Tomography, 2021, 7(4): 877-892.
- [10] Ninomya K, Arimura H, Chan WY, et al. Robust radiogenomics approach to the identification of EGFR mutations among patients with NSCLC from three different countries using topologically invariant Betti numbers[J]. PLoS One, 2021, 16(1): e0244354. DOI: 10.1371/journal.pone.0244354.
- [11] Tu W, Sun G, Fan L, et al. Radiomics signature: A potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology[J]. Lung Cancer, 2019, 132: 28-35. DOI: 10.1016/j.lungcan.2019.03.025
- [12] Huo JW, Luo TY, Diao L, et al. Using combined CT-clinical radiomics models to identify epidermal growth factor receptor mutation subtypes in lung adenocarcinoma[J]. Front Oncol, 2022, 12: 846589. DOI: 10.3389/fonc.2022.846589
- [13] Yang X, Liu M, Ren Y, et al. Using contrast-enhanced CT and

- non-contrast-enhanced CT to predict EGFR mutation status in NSCLC patients-a radiomics nomogram analysis[J]. *Eur Radiol*, 2022, 32(4):2693-2703.
- [14] Huynh E, Hosny A, Guthier C, et al. Artificial intelligence in radiation oncology[J]. *Nat Rev Clin Oncol*, 2020, 17(12):771-781.
- [15] Wang C, Ma J, Shao J, et al. Predicting EGFR and PD-L1 status in NSCLC patients using multitask AI system based on CT images[J]. *Front Immunol*, 2022, 13:813072. DOI: 10.1155/2021/5499385
- [16] Wang C, Xu X, Shao J, et al. Deep learning to predict EGFR mutation and PD-L1 expression status in non-small-cell lung cancer on computed tomography images [J]. *J Oncol*, 2021, 2021:5499385. DOI:10.3389/fonc.2020.598721
- [17] Zhang B, Qi S, Pan X, et al. Deep CNN model using CT radiomics feature mapping recognizes EGFR gene mutation status of lung adenocarcinoma[J]. *Front Oncol*, 2020, 10:598721. DOI: 10.3389/fonc.2020.598721
- [18] Liu Z, Wang S, Dong D, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges[J]. *Theranostics*, 2019, 9(5):1303-1322.
- [19] Shiri I, Maleki H, Hajianfar G, et al. Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms[J]. *Mol Imaging Biol*, 2020, 22(4):1132-1148.
- [20] Li H, Gao C, Sun Y, et al. Radiomics analysis to enhance precise identification of epidermal growth factor receptor mutation based on positron emission tomography images of lung cancer patients [J]. *J Biomed Nanotechnol*, 2021, 17(4):691-702.
- [21] Edalat-Javid M, Shiri I, Hajianfar G, et al. Cardiac SPECT radiomic features repeatability and reproducibility: A multi-scanner phantom study[J]. *J Nucl Cardiol*, 2021, 28(6):2730-2744.
- [22] Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges[J]. *Magn Reson Imaging*, 2012, 30(9):1234-1248.
- [23] Shiri I, Amini M, Nazari M, et al. Impact of feature harmonization on radiogenomics analysis: prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images[J]. *Comput Biol Med*, 2022, 142:105230. DOI: 10.1016/j.compbiomed.2022.105230
- [24] Mu W, Jiang L, Zhang J, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics[J]. *Nat Commun*, 2020, 11(1):5228.
- [25] Koyasu S, Nishio M, Isoda H, et al. Usefulness of gradient tree boosting for predicting histological subtype and EGFR mutation status of non-small cell lung cancer on ¹⁸F FDG-PET/CT[J]. *Ann Nucl Med*, 2020, 34(1):49-57.
- [26] Wang G, Wang B, Wang Z, et al. Radiomics signature of brain metastasis: prediction of EGFR mutation status[J]. *Eur Radiol*, 2021, 31(7):4538-4547.
- [27] Ren M, Yang H, Lai Q, et al. MRI-based radiomics analysis for predicting the EGFR mutation based on thoracic spinal metastases in lung adenocarcinoma patients[J]. *Med Phys*, 2021, 48(9):5142-5151.
- [28] 唐兴, 白国艳, 王虹, 等. 基于多序列 MRI 影像组学预测肺腺癌 EGFR 基因表型[J]. *放射学实践*, 2021, 36(8):1010-1015.

(收稿日期:2022-03-31 修回日期:2022-09-12)