

头颈部疾病影像组学和人工智能的艰难探索与展望

曲晓霞, 鲜军舫

【中图分类号】R445.2;R814.42 【文献标志码】A 【文章编号】1000-0313(2021)08-0949-03

DOI:10.13609/j.cnki.1000-0313.2021.08.001

影像组学(radiomics)和医学影像人工智能(artificial intelligence, AI)在全身各部位或器官疾病的显示、诊断和鉴别诊断、分期和分级、术前评估、疗效监测和预后等方面的研究广泛开展^[1-2],影像 AI 产品也几乎涉及所有器官或疾病,部分产品已取得国家药监局三类医疗器械注册证^[3],显著提高了诊疗水平和效率,促进了个性化诊疗的实施。头颈部器官和结构多,病变发病率较低,虽然全国各地临床和影像科都希望借助影像组学和 AI 来规范和提高头颈部影像学水平,对影像组学和 AI 在头颈部的研究和推广应用寄予了很高的期望,但由于头颈部情况特殊,与其它部位相比,影像组学和 AI 的研究相对困难、步履维艰,AI 研究需要的样本量大,故头颈部疾病的 AI 相关研究开展起来更难。此外,绝大部分影像组学和 AI 的研究是基于技术导向和论文导向,而不是基于目标和问题导向;同时,大部分影像医师对影像组学参数及含义缺乏了解,也极大地阻碍了其发展。因此,我们要积极转变意识,认真总结和分析,找到问题和症结,推进头颈部影像组学和医学影像 AI 的发展。

影像组学及其在头颈部的研究进展

1. 影像组学的参数及内涵

影像组学采用高通量特征提取算法对影像学数据进行定量分析,提取全部特征,然后对这些影像组学特征进行定量分析,挑选出能预测诊断、疗效和预后等的特征集合,即“影像组学标签(radiomic signature)”。影像组学是将影像的定量分析与机器学习等方法有机结合起来,充分应用影像组学标签建立预测诊断、疗效和预后等的模型,突破了以往仅仅只是形态学和半定量分析的局限性和瓶颈,为临床决策和精准诊疗提供个性化的量化信息^[2]。影像组学包括影像数据采集、标注与分割、特征提取和定量分析、特征选择、模型建

立及临床验证等步骤,尤其是临床验证和推广应用是影像组学的灵魂和目的,没有临床验证和推广应用,即使再好的影像组学标签和模型,也都只是空中楼阁和海市蜃楼。

影像组学特征可以分为形状特征、一阶直方图特征以及二阶直方图或纹理特征^[2]。形状特征指选取感兴趣区大小的特征及与球体相似程度的特征,包括体积、表面积、最大径、有效直径、表面体积比、致密度、偏心度和球形度等。一阶直方图特征指与体素强度分布有关的特征,可反映体素的对称性、均匀性以及局部强度分布的变化,包括最小值、最大值、平均数或中位数、标准差、偏度和峰度等,常由直方图分析计算获得。二阶直方图或纹理特征是指体素空间分布强度等级的特征,主要包括灰度共生矩阵(gray level co-occurrence matrix, GLCM)和灰度级长矩阵(gray level run-length matrix, GLRLM)等。灰度共生矩阵是表示灰度值的矩阵,包括熵(与异质性有关)、能量(与均匀性有关)、对比度、同质性(与局部灰度均衡性有关)、不相似性和相关性。灰度级长矩阵是指在图像的二维矩阵中预设方向上强度相同的连续体素的长度^[2]。

2. 影像组学在诊断和鉴别诊断方面的研究进展

基于 MRI 或 CT 的影像组学特征显著提升了腮腺^[4-5]、泪腺^[6]、眼球^[7]、眼眶^[8]、鼻窦^[9]和甲状腺^[10]等部位良恶性病变的鉴别能力,敏感度为 75%~100%,特异度为 88%~93%。基于 CT 影像组学特征建立的逻辑回归模型鉴别头颈部淋巴结转移和淋巴瘤的 AUC 分别达到 0.987(训练组)和 0.938(训练组),具有很高的诊断效能^[11]。本期发表的《基于 CT 影像组学模型鉴别头颈部良、恶性淋巴结》^[12]一文中,CT 影像组学模型显著提高了对头颈部淋巴结良恶性的鉴别诊断效能,ROC 曲线下面积(AUC)达到了 0.958。

颈部淋巴结转移是甲状腺乳头状癌复发及远处转移的高危因素,术前准确检出淋巴结转移对选择治疗方案和评估预后具有重要价值。本期发表的《基于 CT 增强图像影像组学预测甲状腺乳头状癌颈部淋巴结转移》^[13]的研究中,结果表明基于 CT 增强图像使用非线性判别分析的诊断模型的错判率仅 3.08%,敏感度和特异度分别达 98.46%和 95.38%。本期发表

作者单位:100730 北京,首都医科大学附属北京同仁医院放射科

作者简介:曲晓霞(1986—),女,山西忻州人,博士研究生,工程师,主要从事医学影像处理分析及人工智能。

通讯作者:鲜军舫, E-mail: cjr_xianjunfang@vip.163.com
基金项目:北京市医院管理中心“登峰”计划专项经费资助(DFL20190203);北京市医管局临床医学发展专项“扬帆计划”(ZYLX201704)

的《4D-CT 影像组学模型鉴别甲状旁腺良性病变与淋巴结的应用初探》^[14]一文的结果表明,基于 CT 平扫和增强扫描及联合诊断模型在鉴别甲状旁腺腺瘤或增生与颈部淋巴结中的诊断效能较高,AUC 为 0.869~0.962、敏感度为 0.813~0.934、特异度为 0.693~0.933。

基于 MRI 的灰度共生矩阵纹理特征预测舌癌患者颈部淋巴结转移的研究结果显示,熵用于预测是否转移的 AUC 为 0.90,阈值为 7.19,敏感度为 80.0%,特异度为 86.7%;厚度的 AUC 为 0.81,阈值为 11.19mm,敏感度为 78.0%,特异度为 81.7%。而影像医师基于 MRI 表现诊断颈部淋巴结转移的敏感度仅 65.0%,特异度为 80.0%^[15]。

3. 影像组学在肿瘤分级及分期方面的研究进展

一项基于磁共振表观扩散系数(ADC)图的纹理分析预测舌和口底鳞状细胞癌组织学分级(I、II、III级)的结果表明:ADC 值第 10 百分位数(ADC₁₀)、熵及联合模型的 ROC 曲线下面积(AUC)分别为 0.72、0.75 和 0.81^[16],预测能力优于单一的 ADC 值。

影像组学在喉及下咽癌、舌和口底鳞状细胞癌的分期方面也取得了较好的效果。喉及下咽鳞状细胞癌侵犯甲状软骨的临床分期为 T₄ 期,常需进行全喉切除术。基于增强 CT 的影像组学特征预测喉及下咽鳞状细胞癌侵犯甲状软骨的 AUC 为 0.876~0.905,显著高于放射科医师的诊断效能(AUC 为 0.721)^[17]。

4. 影像组学在预后预测方面的研究进展

基于增强 CT 影像组学特征的列线图预测喉鳞状细胞癌术后总生存率的 C 指数在训练队列和验证队列分别为 0.817 和 0.913,预测效能明显提高^[18]。基于 T₂WI 和增强后 T₁WI 的影像组学特征显著提高了晚期鼻咽癌预后的预测能力,C 指数为 0.761,优于 T₂WI 或增强后 T₁WI 的影像组学特征,也优于 TNM 分期^[19]。

5. 头颈部影像组学研究的局限性及改进措施

颈部淋巴结转移对于预后和是否实施颈淋巴结清扫非常关键,因此,影像组学在颈部淋巴结转移诊断方面的研究比较多,诊断效能较高,但实际情况可能并非如此,由于对颈部淋巴结转移的判断标准是以区域淋巴结全部转移或无转移来定的,结果的可靠性需要采用淋巴结影像与病理点对点对照的前瞻性研究方案来验证和纠正。此外,相关文献显示影像组学分析在两种或几种具体疾病方面的诊断效能也比较高,但与临床实际场景不一致,应该是将该部位的所有病变纳入研究对象,在今后的研究设计中要重点关注上述结果的可重复性和可靠性的问题。

在组学分析过程中,需在病变或结构内勾画感兴趣区(ROI),勾画方法有手动法、半自动法和自动法。

手动勾画法的准确性较高,对不规则病变或结构的边界勾画相对精准,但受人为主观因素的影响较大,可重复性相对较低,耗时较长,效率较低,而且对于大样本量研究,手动勾画的工作量更大,投入的人力、物力和财力更多。与手动法相比,半自动或自动勾画标注法的可重复性较高,效率也比较高。其中,半自动法为目前影像组学研究中勾画 ROI 的主要方法。但是头颈部结构精细,病变小,部分病变常累及邻近的不规则间隙、神经和血管,采用半自动勾画法的效果不佳^[20],通常需采用手动勾画。因此,头颈部病变的影像组学研究开展起来比较困难,这可能是头颈部病变影像组学发表的文章较少、临床工作中应用更难的主要原因之一。怎样提高头颈部病变或结构的勾画效率和可重复性是头颈部影像组学研究和应用的重大挑战,是头颈部影像组学未来能否顺利发展和提升的关键点之一。

与其它领域的组学研究一样,人们寄希望于影像组学能够获得疾病的生物标志物,尽管研究报道影像组学特征对诊断、疗效评估和预后预测有较高的敏感度、特异度和符合率,但在临床实际工作中需要验证和检测。验证分为内部验证和外部验证,内部验证可消除过拟合形成的结果虚高的情况。

AI 在头颈部的研究进展

鼻骨区是颅面部外伤中最常受累部位,但由于鼻骨区体积较小且解剖结构细微,该部位骨折容易被漏诊及误诊。本期发表的《鼻骨区骨折深度学习模型的建立和临床效能评估》^[21]一文中使用 CT 图像深度学习法建立 AI 模型,对鼻骨区骨折进行探测,结果显示:AI 模型检出骨折的敏感度为 86.64%,特异度为 41.99%,符合率为 68.71%,能够提高低年资医师对鼻骨区骨折的诊断效能,缩小与高年资医师之间的差距。

甲状腺癌术前对甲状腺结节良、恶性进行判断至关重要。本期发表的《基于 CT 双期增强图像的深度学习迁移学习模型对甲状腺良恶性结节的分类研究》^[22]一文中,对比分析了 3 种深度迁移学习(deep transfer learning, DTL)模型(VGG19、ResNet50 和 DenseNet201)对甲状腺良恶性结节的鉴别诊断效能,结果显示 DenseNet201 模型获得了最好的训练和测试结果,在训练集和测试集中的最高预测符合率分别为 1.00 和 0.98。

由于 AI 技术要求样本量较大,所以其在头颈部的研究和试验产品相对较少。尽管如此,相信通过努力,AI 可通过对头颈部疾病的检测和测量以及通过结构化标准报告来规范影像报告,将会为临床提供诊断、治疗前评估、疾病监测和预后的定量信息,助力个性化临床决策和精准诊疗。

头颈部疾病影像组学和 AI 的展望

头颈部病变的发病率相对较低,样本量一般较小,与其它部位常见病变相比,其影像组学和 AI 算法结果的可靠性和可重复性更要值得重视,希望通过严格规范的多中心研究获得较大样本量病例,来提高结果的可靠性和可重复性。第二,与其它部位一样,目前对这些通过影像组学或 AI 算法获得的影像学参数或标志物的生物学意义尚知之甚少,有的可能完全不知道。尽管文献报道过一些影像学参数或标志物与肿瘤分级、炎性细胞浸润、基因表达、基因突变和信号传导通路等具有相关性,但仍要进一步研究影像组学或 AI 算法,解开“黑箱”之谜,弄清楚这些参数代表的临床和/或分子含义。第三,从知识维度来讲,知识可分为“明知识”和“潜知识”。“明知识”(或者称为“显知识”)是可以表达的知识,而“潜知识”(或者称为“默知识”)是可以感受但没法清楚表达的知识。而“暗知识”的概念是随着机器学习深入发展应运而生的^[23],是指机器能够明白、但人既不能清楚表达又不能够直观感受的知识,其扩大了以前的知识维度。在临床实际工作中,医师如何充分利用只有机器才能够明白的“暗知识”为头颈部疾病的精准诊疗、随访监测和预后预测提供更多信息,是头颈部 AI 和影像组学未来发展的方向和目标,也是其能够真正走进临床并充分发挥作用的必由之路和希望之路。第四,从目前研究现状和发表的论文来看,与其它部位一样,大多数头颈部疾病的影像组学和 AI 研究是基于技术驱动和论文导向,而真正基于目标导向和问题导向的研究相对较少。因此,我们既要重视影像医师基本功和临床思维的培养,又要进一步明确临床问题并开展“以患者为中心、以临床问题为导向”的 AI 和影像组学研究,充分挖掘和提取头颈部影像信息,为患者提供个性化临床决策和精准诊疗。

参考文献:

- [1] 许强,张其锐,卢光明. 新一代医学影像人工智能临床转化现状与挑战[J]. 中华放射学杂志, 2019, 53(11): 913-915.
- [2] 史张,刘崎. 影像组学技术方法的研究及挑战[J]. 放射学实践, 2018, 33(6): 91-94.
- [3] 刘士远. 医学影像人工智能发展趋势与挑战[J]. 中华放射学杂志, 2021, 55(7): 700-702.
- [4] 雷晓雯,程敬亮,冉云彩. ADC 全域直方图分析对腮腺多形性腺瘤和腺淋巴瘤的鉴别诊断价值[J]. 放射学实践, 2020, 35(8): 1005-1008.
- [5] 宋承汝,程敬亮,张勇,等. 高分辨率 DWI 表现扩散系数直方图鉴别腮腺常见肿瘤[J]. 放射学实践, 2019, 34(7): 749-754.
- [6] Guo J, Li Z, Qu X, et al. Value of MRI-based radiomics analysis for differentiation of benign and malignant epithelial neoplasms in the lacrimal gland: a retrospective study[J]. Acta Radiol, 2021, 62

- (6): 743-751. DOI: 10. 1177/0284185120940258.
- [7] Su Y, Xu X, Zuo P, et al. Value of MR-based radiomics in differentiating uveal melanoma from other intraocular masses in adults[J/OL]. Eur J Radiol, 2020, 131(2020): e109268. DOI: <https://doi.org/10.1016/j.ejrad.2020.109268>.
- [8] Guo J, Liu Z, Shen C, et al. MR-based radiomics signature in differentiating ocular adnexal lymphoma from idiopathic orbital inflammation[J]. Eur Radiol, 2018, 28(9): 3872-3881.
- [9] Wang X, Dai S, Wang Q, et al. Investigation of MRI-based radiomics model in differentiation between sinonasal primary lymphomas and squamous cell carcinomas[J]. Jpn J Radiol, 2021, 39(8): 755-762. DOI: 10. 1007/s11604-021-01116-6.
- [10] 孔丹,张建东,单文莉,等. CT 影像组学模型对甲状腺良恶性结节的鉴别价值[J]. 中华放射学杂志, 2020, 54(3): 187-191.
- [11] 张力,于淑靖,张迎,等. CT 影像组学在头颈部木村淋巴瘤病变与淋巴瘤鉴别中的应用[J]. 放射学实践, 2020, 35(2): 159-164.
- [12] 胡大涛,夏春华,李羚,等. 基于 CT 影像组学模型鉴别头颈部良、恶性淋巴结[J]. 放射学实践, 2021, 36(8): 965-970.
- [13] 刘妮,谢元亮,黄增发,等. 基于 CT 增强图像影像组学预测甲状腺乳头状癌颈部淋巴结转移[J]. 放射学实践, 2021, 36(8): 971-975.
- [14] 苏童,陈钰,赵伦,等. 4D-CT 影像组学模型鉴别甲状旁腺良性病变与淋巴结的应用初探[J]. 放射学实践, 2021, 36(8): 952-958.
- [15] 钟熹,江魁明,麦慧,等. 基于灰度共生矩阵的 MRI 纹理分析预测舌癌患者颈部淋巴结转移的价值初探[J]. 中华放射学杂志, 2018, 52(9): 649-654.
- [16] 任继亮,袁瑛,董迪,等. 术前表观扩散系数图纹理分析预测舌和口底鳞状细胞癌组织学分级的价值[J]. 中华放射学杂志, 2019, 53(4): 281-285.
- [17] Guo R, Guo J, Zhang L, et al. CT-based radiomics features in the prediction of thyroid cartilage invasion from laryngeal and hypopharyngeal squamous cell carcinoma[J/OL]. Cancer Imaging, 2020, 20(1): e81. [2020 Nov 11]. DOI: 10. 1186/s40644-020-00359-2.
- [18] Chen L, Wang H, Zeng H, et al. Evaluation of CT-based radiomics signature and nomogram as prognostic markers in patients with laryngeal squamous cell carcinoma[J/OL]. Cancer Imaging, 2020, 20(1): e28. [2020 Apr 22]. DOI: 10. 1186/s40644-020-00310-5.
- [19] Zhang B, Tian J, Dong D, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma[J]. Clin Cancer Res, 2017, 23(15): 4259-4269. DOI: 10. 1158/1078-0432.CCR-16-2910.
- [20] 周燕,许晓泉,苏国义,等. 双能 CT 动静脉期加权融合图像影像组学对甲状腺乳头状癌颈部淋巴结转移的诊断价值[J]. 中华放射学杂志, 2021, 55(7): 703-709.
- [21] 陶建华,陈聪,张怀宇,等. 鼻骨区骨折深度学习模型的建立和临床效能评估[J]. 放射学实践, 2021, 36(8): 959-964.
- [22] 孟名柱,潘昌杰,张铭,等. 基于不同的深度迁移学习模型对甲状腺 CT 良恶性结节的分类研究[J]. 放射学实践, 2021, 36(8): 976-980.
- [23] 王维嘉. 暗知识: 机器认知如何颠覆商业和社会[M]. 北京, 中信出版社, 2019. ISBN: 9787508692982

• 头颈部影像学专题 •

4D-CT 影像组学模型鉴别甲状旁腺良性病变与淋巴结的应用初探

苏童, 陈钰, 赵伦, 李秀丽, 王曼, 刘春晓, 朱振宸, 刘昊喆, 张竹花, 胡亚, 廖泉, 邢小平, 金征宇

【摘要】 目的:探讨基于 4D-CT 扫描的建立影像组学模型对甲状旁腺腺瘤或增生与颈部淋巴结的鉴别诊断价值。**方法:**回顾性将 80 例行术前 4D-CT 平扫和多期增强扫描且经手术病理证实的甲状旁腺腺瘤或增生患者纳入本研究。在甲状旁腺结节及同侧颈部 II 区淋巴结内逐层手动勾画感兴趣区, 获得其容积感兴趣区 (VOI), 基于平扫期、动脉期、静脉期和延迟期图像, 各提取 1262 个影像组学特征。在每个单独期相或其组合的基础上, 特征选择基于 L1-based 方法, 采用 L1 正则化的 Logistic 回归分析分别建立相应的影像组学模型, 并进行 10 折交叉验证。采用 ROC 曲线下面积 (AUC)、敏感性和特异性评价各模型的鉴别诊断效能。**结果:**分别基于平扫期、动脉期、静脉期、延迟期、平扫+动脉期、平扫+动脉+静脉期及平扫+动脉+静脉+延迟期图像建立 7 个影像组学模型, 各模型鉴别甲状旁腺腺瘤或增生与颈部淋巴结的诊断效能均较高 (训练组 AUC 为 0.949~0.962, 验证组为 0.869~0.904)。在验证集中, 基于静脉期的影像组学模型的 AUC 最大 (0.904), 相应的鉴别诊断敏感度为 0.886、特异度为 0.785、符合率为 0.835。**结论:**基于 4D-CT 平扫和多期增强扫描建立的影像组学模型对鉴别甲状旁腺腺瘤或增生与颈部淋巴结的准确性较高, 其中以基于静脉期图像的影像组学模型的诊断效能最佳。

【关键词】 影像组学; 体层摄影术, X 线计算机; 甲状旁腺病变; 淋巴结

【中图分类号】 R814.42; R736.2; R733.4 **【文献标识码】** A

【文章编号】 1000-0313(2021)08-0952-07

DOI:10.13609/j.cnki.1000-0313.2021.08.002

开放科学 (资源服务) 标识码 (OSID):



Application of radiomics model based on 4D-CT for preoperative discrimination benign lesions of parathyroid from lymph nodes SU Tong, CHEN Yu, ZHAO Lun, et al. Department of Radiology, Peking Union Medical College Hospital, Beijing 100730, China

【Abstract】 Objective: The purpose of this study was to discriminate parathyroid adenomas/hyperplasia from lymph nodes using radiomics models based on four-dimensional computed tomography (4D-CT). **Methods:** Preoperative 4D-CT images of eighty patients with parathyroid adenomas/hyperplasia were retrospectively included in this study. 1262 radiomic features from each single phase, as well as their combinations, were extracted. Based on each single phase and their combinations, the L1-based method was performed for feature selection, radiomics models were developed by using Logistic Regression with L1 regularization and evaluated by performing the repeated 10-fold cross validation. The performance of the radiomics models was assessed using the area under the receiver-operating characteristic curve (AUC). **Results:** Seven radiomics models were built based on plain, arterial, venous and delayed phases individually, and on combinations of plain+arterial phases, plain+arterial+venous phases, and plain+arterial+venous+delayed phases, respectively. All of the models showed a relatively high AUC for discriminating parathyroid adenomas/hyperplasia from lymph nodes, ranging from 0.949 to 0.962 in the training cohort and from 0.869 to 0.904 in the validation cohort. In the validation cohort, the model built on a single venous phase showed the highest AUC of 0.904, with sensitivity of 0.886, specificity of 0.785 and accuracy of 0.835. **Conclusion:** Radiomics models based on 4D CT

作者单位: 100730 北京, 中国医学科学院北京协和医院放射科 (苏童, 陈钰, 王曼, 刘春晓, 朱振宸, 刘昊喆, 张竹花, 金征宇), 基本外科 (胡亚, 廖泉), 内分泌科 (邢小平); 100000 北京, 北京深睿医疗 (赵伦, 李秀丽)

作者简介: 苏童 (1992-), 女, 山东聊城人, 博士研究生, 住院医师, 主要从事头颈部影像诊断工作。

通讯作者: 金征宇, E-mail: jinzy@pumch.cn; jinzy_pumch@126.com; 张竹花, E-mail: 15010147685@163.com

基金项目: 中国医学科学院中央级公益性科研院所基本科研业务费专项资金 (2019PT320008、2018PT32003、2017PT32004); 北京市科技计划课题 (Z201100005620008)

images may be a helpful approach for discriminating parathyroid adenomas/hyperplasia from lymph nodes. The radiomics model based on a single venous phase showed the best diagnostic efficacy.

【Key words】 Radiomics; Tomography, X-ray computed; Parathyroid lesions; Lymph node

原发性甲状旁腺功能亢进症是一种常见的内分泌疾病,多因甲状旁腺腺瘤或良性增生所致,以甲状旁腺激素分泌过多为特征,主要治疗方式为手术切除异常甲状旁腺组织。甲状旁腺腺瘤与增生虽然在病理学上表现不同,但均为良性结节,且对其临床治疗方法相似。随着影像诊断技术的发展,对病灶的术前定位更精确,微创甲状旁腺病变切除术成为大多数病例的首选治疗方法,术后并发症更少。因此,术前影像学检查对甲状旁腺病变的准确定位和诊断具有重要作用。

对责任病灶的准确定位以及对甲状旁腺区结节的准确性(与淋巴结或异位甲状腺组织的鉴别等)是甲状旁腺疾病影像诊断的主要挑战。^{99m}Tc-甲氧基异丁基异腈(^{99m}Tc-sestamibi, ^{99m}Tc-MIBI)显像和甲状旁腺超声检查是诊断甲状旁腺疾病的常规影像学方法,但存在相应的局限性:如果病灶没有典型的高摄取, MIBI 显像很难精确定位病灶;超声诊断高度依赖于操作者的经验和技能,且仅限于通过形态学特征区分甲状旁腺结节与颈部淋巴结。在 MIBI 显像和超声诊断困难或无法获得一致性结论时,CT 通常作为替代影像学检查方法^[1]。2006 年, Rodgers 等^[2]引入了 4D-CT 扫描方案并对其进行了广泛的研究。在甲状旁腺病变术前诊断方面,4D-CT 的敏感度为 70%~90%,与 MIBI 显像和超声相当;而在 MIBI 显像和超声诊断困难的复杂病例或复发性病变中,4D-CT 具有明显的诊断优势^[3]。甲状旁腺病变是富血供的组织,其 4D-CT 表现为快速增强和快速廓清的强化模式,不同于颈部淋巴结。鉴于甲状旁腺解剖位置的特点及异位甲状旁腺的好发部位^[3],主要应与位于甲状腺周围及上纵隔区域的颈部 VI 和 VII 区小淋巴结进行鉴别。但是,并非所有的甲状旁腺腺瘤均表现为典型的、不同于颈部淋巴结的快速增强和快速廓清的强化模式,若是不能意识到此强化特点的异质性,不可避免会漏诊误诊的结果^[4]。因此,除了增强模式外,若能综合整个病变的影像学数据建立预测模型,或可为甲状旁腺疾病的影像学鉴别诊断提供客观的辅助诊断信息。

影像组学是一种新兴的图像处理方法,通过提取高通量的定量特征来建立合适的模型,以辅助疾病诊断、治疗效果评估和预后预测。已有研究表明,CT 影像组学对于颈部淋巴结病变具有一定的诊断优势和预测价值^[5-6]。基于 4D-CT 鉴别甲状旁腺腺瘤或增生与颈部淋巴结的定量方法的相关研究较少,CT 影像组学模型具有应用前景。因此,本研究旨在探讨基于

4D-CT 的影像组学模型在鉴别甲状旁腺病变与淋巴结方面的诊断效能,同时对基于单期 CT 与多期 CT 图像的组学模型的诊断效能进行比较。

材料与方法

1. 病例资料

本研究获得本院伦理委员会批准。回顾性搜集 2014 年 9 月—2019 年 1 月在本院诊断为原发性甲状旁腺功能亢进的 105 例患者的病例资料。所有患者术前行 MIBI 显像和甲状旁腺超声检查,但在病变诊断上未获得阳性或一致性结果,随后行 4D-CT 检查。纳入标准:①术前 4 周内行标准的甲状旁腺 4D-CT 平扫和多期增强扫描;②在我院行手术切除并具备完整的病理学报告,病理诊断结果为甲状旁腺腺瘤或增生;③ 4D-CT 图像资料完整且图像质量良好、无明显伪影。最终共 80 例符合条件的患者被纳入后续分析,具体步骤和流程详见图 1。

2. CT 图像采集

使用 Siemens Somatom Definition Flash 双源 CT 机,采用标准化扫描方案,扫描期相包括平扫、动脉期、静脉期和延迟期。患者取仰卧位,扫描范围自颅底至气管分叉水平。扫描参数:实时动态曝光剂量调节(CARE Dose 4D)技术(参考管电流为 165 mA), 120 kV,层厚 4.0 mm,层间隔 4.0 mm,螺距 0.9,视野 259 mm×259 mm。扫描参数值均符合 2020 年版《头颈部 CT 检查和辐射剂量管理专家共识》推荐^[7]。经肘前静脉以 3.0 mL/s 的流率团注 90 mL 对比剂优维显(370 mg I/mL),动脉期、静脉期和延迟期增强扫描

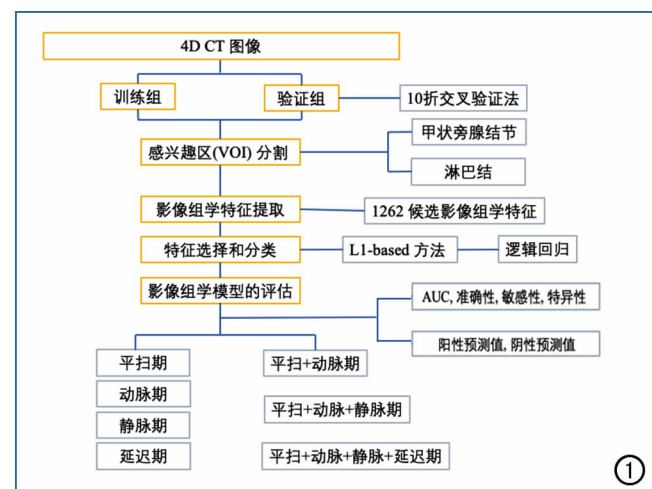


图 1 本研究的流程图。

的延迟时间分别 30、65 和 115 s^[8]。CT 检查的有效剂量为 7.44~12.07 mSv,平均(9.10±1.25) mSv,与文献报道的数值范围基本一致^[8-9]。

3. 影像组学特征提取

首先,由两位分别有 5 年和 10 年头颈部影像诊断经验的放射科医师在不知晓临床信息的前提下,在每例患者的各期 CT 图像上在甲状旁腺结节(包括腺瘤和增生腺体)内手动勾画感兴趣区,逐层勾画结节的实性成分,避开囊变或坏死区域,获得结节的容积感兴趣区(volume region of interest,VOI)。同时,选择同侧颈部 II 区的一个淋巴结作为自身对照,淋巴结的选择由两位放射科医师共同协商决定。淋巴结的选择标准是短径在 4~7 mm,卵圆形或椭圆形,边界清晰,密度及强化均匀。淋巴结 VOI 的获取方法与甲状旁腺结节一致。

将勾画了 VOI 的原始图像上传至在线分析平台(www.research.deepwise.com)。首先对数据进行重采样以抵消空间分辨率不一致带来的干扰,然后采用小波变换算法进行滤波以提高信噪比。然后,使用 PyRadiomics 工具,自每个 VOI 中提取 1262 个影像组学特征。这些特征分为 3 类:(1)一阶特征(描述图像强度);(2)形状特征(描述肿瘤的二维和三维形态);(3)纹理特征(描述肿瘤的内部和表面纹理),本研究中主要包括灰度共生矩阵、灰度大小带矩阵、灰度游程矩阵和灰度相关矩阵四大类相关参数。

此外,自所有病例中随机选取 30 例患者,由两位放射科医师按上述感兴趣区勾画方法,获得每例患者每一期相的 VOI 并提取其纹理参数,对两位医师提取的组学特征值的一致性进行评估(观察者间);每位放射科医师在 2~3 周后重复第一次的步骤,对 2 个时间点提取的组学特征值进行一致性评价(观察者内)。一致性评价采用组内相关系数(intraclass correlation coefficient,ICC),ICC>0.75 为一致性良好。

4. 特征选择和影像组学模型的建立

在每个单独期相或其组合的基础上,使用 L1-based 的方法对提取的组学特征进行筛选,选择系数非零的特征,并计算 logistic 回归积分得到影像组学特征系数。基于筛选的组学特征,采用 L1 正则化的 Logistic 回归分析分别建立相应的影像组学模型。正则化参数包括 penalty l1、C=0.15、拟合截距=true。

5. 影像组学模型的评价

在 10 倍交叉验证中,基于每个独立交叉的测试结果,联合计算得到各个影像组学模型诊断效能指标,评估和比较每个期相及其组合之间的效能差异。采用受试者工作特征(ROC)曲线分析每个组学模型的诊断效能,计算曲线下面积(AUC)及模型预测的符合率、

敏感度、特异度、阳性预测值和阴性预测值。

采用 Hosmer-Lemeshow 检验,在训练组和验证组中创建校准曲线,来反映影像组学模型预测概率与病理诊断结果的一致性^[10]。为了评估影像组学特征的明显增量效用,通过决策曲线分析来评估影像组学模型的临床实用性,决策曲线分析是通过估算该组队列中阈值概率范围内的净收益来进行的^[11]。

6. 统计学分析

使用 IBM SPSS Statistics 22.0 统计学分析软件。各组之间组学特征值的差异采用 Mann-Whitney U 非参数检验,以 $P<0.05$ 为差异有统计学意义。对符合正态分布的计量资料以平均值±标准差的形式表示,不符合正态分布的计量资料以中位数(最小值~最大值)表示。

结 果

1. 临床资料

本研究共纳入 80 例甲状旁腺单发结节患者,男 24 例,女 56 例,年龄 14~78 岁,平均(54.0±13.3)岁。其中,10 例(12.5%)为甲状旁腺结节样增生,70 例为甲状旁腺腺瘤;12 个(15%)结节内有囊变区。本组中甲状旁腺结节(n=80)和同侧 II 区淋巴结(n=80)的平均最大直径分别为 11.40(5.83~32.80)和 8.94(4.47~21.63) mm,差异有统计学意义($Z=-4.945, P<0.001$),平均体积分别为 599.75(116.71~7831.21)和 267.58(69.42~1458.63) mm³ ($Z=-5.368, P<0.001$)。

根据 Perrier 等^[12]提出的甲状旁腺结节位置分区方法,本组中 80 个甲状旁腺结节的位置分布见表 1。左侧结节多于右侧结节。有 52.5% 甲状旁腺病变位于 D、E 和 F 区,即甲状腺后下方和前上纵隔(图 2),容易与小淋巴结相混淆,因为颈部淋巴结也常见于这些部位。

表 1 甲状旁腺病变的 Perrier 定位 /个

分区	右侧	左侧	合计
A	12	15	27
B	2	0	2
C	9	1	10
D	2	3	5
E	7	17	24
F	8	0	8
G	1	3	4
合计	33	39	80

2. 影像组学特征

从 4D-CT 每期图像上分别提取了 1262 个影像组学特征。表 2 列举了每个期相所筛选影像组学特征的回归系数。基于每个单独期相的四个模型中回归系数绝对值最大的特征分别是 original_firstorder_Mean、

表 2 基于各期图像筛选的有鉴别价值的影像组学特征及其回归系数

影像组学特征	系数	P 值
平扫期		
original_firstorder_Mean	0.7433	<0.001
wavelet-LLL_firstorder_RootMeanSquared	0.3834	<0.001
wavelet-LHL_gldm_DependenceVariance	0.2894	<0.001
wavelet-HLH_glszm_ZoneEntropy	0.2892	<0.001
original_shape_Maximum2DDiameterSlice	0.2194	<0.001
wavelet-HLL_gldm_HighGrayLevelEmphasis	0.2124	<0.001
log-sigma-5-0-mm-3D_glcml_DifferenceEntropy	0.2061	<0.001
wavelet-LHH_glcml_JointEntropy	0.1191	<0.001
wavelet-LLL_firstorder_10Percentile	0.1117	<0.001
log-sigma-5-0-mm-3D_glszm_GrayLevelNonUniformity	0.1098	<0.001
log-sigma-1-0-mm-3D_glcml_Idn	0.1049	<0.001
wavelet-HHH_firstorder_Kurtosis	0.0901	<0.001
wavelet-LHH_glszm_ZoneEntropy	0.0589	<0.001
log-sigma-3-0-mm-3D_glszm_GrayLevelNonUniformityNormalized	-0.0475	0.006
log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis	0.0461	<0.001
wavelet-HLL_firstorder_Kurtosis	0.0025	<0.001
动脉期		
wavelet-LLL_firstorder_RootMeanSquared	0.6392	<0.001
original_firstorder_90Percentile	0.4193	<0.001
wavelet-LHH_glrml_RunVariance	0.3276	<0.001
log-sigma-1-0-mm-3D_glcml_ClusterShade	0.2992	<0.001
original_gldm_DependenceVariance	0.2387	<0.001
wavelet-HHH_firstorder_Kurtosis	0.2315	<0.001
wavelet-LHH_glszm_GrayLevelNonUniformity	0.2236	<0.001
wavelet-HHH_glszm_ZoneEntropy	0.1804	0.002
wavelet-LHL_glrml_LongRunEmphasis	0.1619	<0.001
log-sigma-5-0-mm-3D_glszm_SizeZoneNonUniformityNormalized	-0.1561	<0.001
log-sigma-5-0-mm-3D_firstorder_Skewness	0.1191	0.001
wavelet-LHH_glszm_ZoneEntropy	0.0956	<0.001
wavelet-LHH_glrml_RunEntropy	0.0486	<0.001
log-sigma-5-0-mm-3D_glszm_ZoneEntropy	0.0351	<0.001
wavelet-LHL_glrml_RunPercentage	-0.0102	<0.001
静脉期		
wavelet-LLL_firstorder_Median	0.6074	<0.001
original_shape_Maximum2DDiameterSlice	0.5099	<0.001
original_firstorder_Median	0.5050	<0.001
log-sigma-5-0-mm-3D_glcml_DifferenceEntropy	0.3537	<0.001
wavelet-HHH_firstorder_Kurtosis	0.2436	<0.001
wavelet-LHH_glrml_RunEntropy	0.1567	<0.001
wavelet-LHL_gldm_DependenceVariance	0.1464	<0.001
wavelet-LHH_glszm_ZoneEntropy	0.1300	<0.001
wavelet-LHH_glszm_GrayLevelNonUniformity	0.0613	<0.001
wavelet-LLL_firstorder_10Percentile	0.0432	<0.001
log-sigma-5-0-mm-3D_firstorder_Skewness	0.0396	<0.001
log-sigma-5-0-mm-3D_glcml_DifferenceVariance	0.0283	<0.001
wavelet-LLL_firstorder_Mean	0.0263	<0.001
original_firstorder_Mean	0.0191	<0.001
wavelet-LHH_glrml_ShortRunLowGrayLevelEmphasis	-0.0169	<0.001
log-sigma-4-0-mm-3D_glrml_GrayLevelNonUniformityNormalized	-0.0054	0.004
延迟期		
wavelet-LLL_firstorder_RootMeanSquared	1.1210	<0.001
wavelet-LHL_gldm_DependenceVariance	0.3378	<0.001
wavelet-HHH_firstorder_Minimum	-0.2975	0.022
log-sigma-5-0-mm-3D_glcml_DifferenceEntropy	0.2451	0.002
original_firstorder_Maximum	0.2034	<0.001
original_shape_Maximum2DDiameterSlice	0.1988	<0.001
log-sigma-5-0-mm-3D_firstorder_Kurtosis	0.1754	0.001
wavelet-HLH_glcml_MaximumProbability	-0.1654	<0.001
log-sigma-5-0-mm-3D_glszm_SmallAreaHighGrayLevelEmphasis	0.0193	<0.001
wavelet-LHL_firstorder_Kurtosis	0.0142	<0.001
wavelet-LLL_firstorder_Median	0.0716	<0.001
log-sigma-5-0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis	0.0331	0.015

wavelet-LLL_firstorder_RootMeanSquared、wavelet-LLL_firstorder_Median 和 wavelet-LLL_firstorder_RootMeanSquared。

一致性分析结果显示,基于各期图像筛选的影像组学特征的观察者间和观察者内一致性均较好,ICC 的取值范围分别为0.788~0.908 和 0.871~0.986。

3. 影像组学模型的诊断效能

在训练组和验证组中,基于每个期相及其组合建立的影像组学模型的诊断效能指标值见表 3,AUC 散点图见图 3。所有影像组学模型鉴别甲状旁腺腺瘤/增生与淋巴结的 AUC 和符合率均较高。训练组中,四期联合的组学模型的 AUC 最大,达 0.962,相应的鉴别诊断符合率为 0.873、敏感度为 0.813、特异度为 0.933;验证组中,基于静脉期的影像组学模型的各项诊断效能指标值最大,AUC 为 0.904,诊断符合率为

0.835。

基于每个单独期相及其组合的影像组学模型的校准曲线见图 4。在训练组和验证组中各影像组学模型预测甲状旁腺病变的概率与病理结果之间均具有良好的一致性(P 值范围分别为 0.115~0.424 和 0.133~0.778)。在验证组中,基于静脉期和基于平扫+动脉+静脉期的影像组学模型显示出相对较好的一致性。

基于每个单独期相及其组合的影像组学模型的决策树曲线(反映鉴别诊断能力)见图 5。决策曲线分析结果显示,在训练组的阈值概率的大部分范围内,基于多期相的模型比基于单独期相的模型在鉴别甲状旁腺腺瘤/增生与淋巴结方面获得了更大的净效益;在验证组中,基于平扫期和静脉期的模型具有相对最佳的净效益。此外,基于延迟期的模型在大多数阈值概率范围内的净效益最低。

表 3 基于各期相及其组合的影像组学模型的诊断效能指标值

期相	AUC	符合率	敏感度	特异度	阳性预测值	阴性预测值
训练组						
平扫	0.950	0.877	0.926	0.827	0.843	0.918
动脉	0.949	0.846	0.914	0.778	0.804	0.900
静脉	0.952	0.861	0.899	0.823	0.835	0.890
延迟	0.952	0.868	0.934	0.803	0.826	0.924
平扫+动脉	0.957	0.867	0.920	0.813	0.831	0.910
平扫+动脉+静脉	0.957	0.867	0.920	0.813	0.831	0.910
平扫+动脉+静脉+延迟	0.962	0.873	0.813	0.933	0.833	0.924
验证组						
平扫	0.892	0.815	0.864	0.765	0.787	0.849
动脉	0.874	0.796	0.864	0.728	0.761	0.843
静脉	0.904	0.835	0.886	0.785	0.805	0.873
延迟	0.886	0.796	0.868	0.724	0.759	0.846
平扫+动脉	0.874	0.787	0.880	0.742	0.742	0.853
平扫+动脉+静脉	0.874	0.780	0.867	0.693	0.739	0.839
平扫+动脉+静脉+延迟	0.869	0.787	0.867	0.707	0.747	0.841

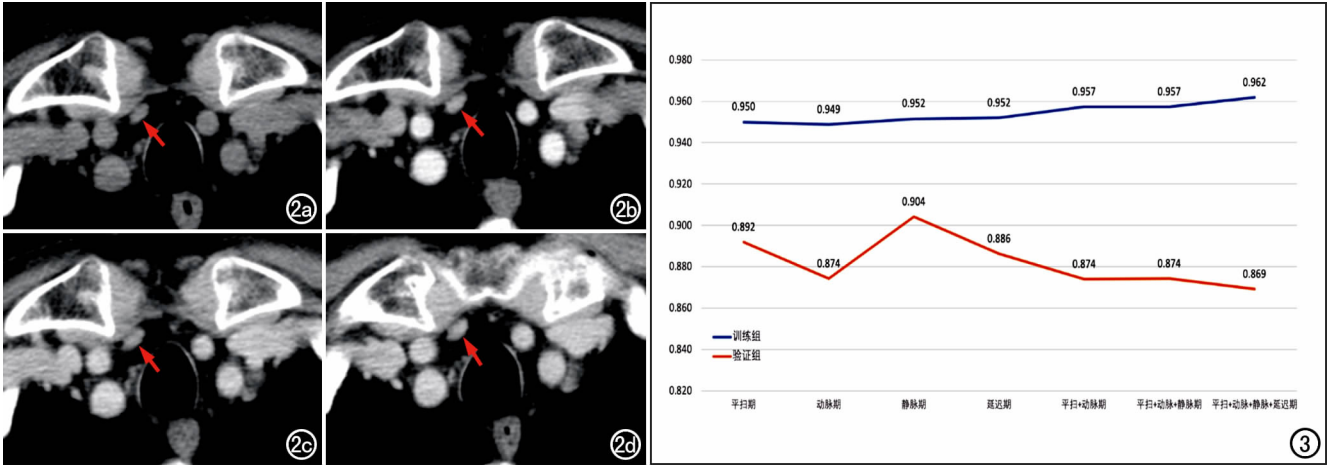


图 2 1 例呈不典型强化的右侧甲状旁腺腺瘤患者的横轴面 4D-CT 图像(窗位 250HU,窗宽 50HU)。a)CT 平扫显示胸骨后小结节(箭),直径 8mm,CT 值 40.0HU,难以与上纵隔小淋巴结区别;b)动脉期,病灶呈中度强化(箭),CT 值约 80.0HU;c)静脉期,病灶呈轻度强化(箭),CT 值为 61.9HU;d)延迟期,病灶持续呈轻度强化(箭),CT 值为 62.8HU。图 3 描述训练组和验证组的影像组学模型 AUC 的散点图。

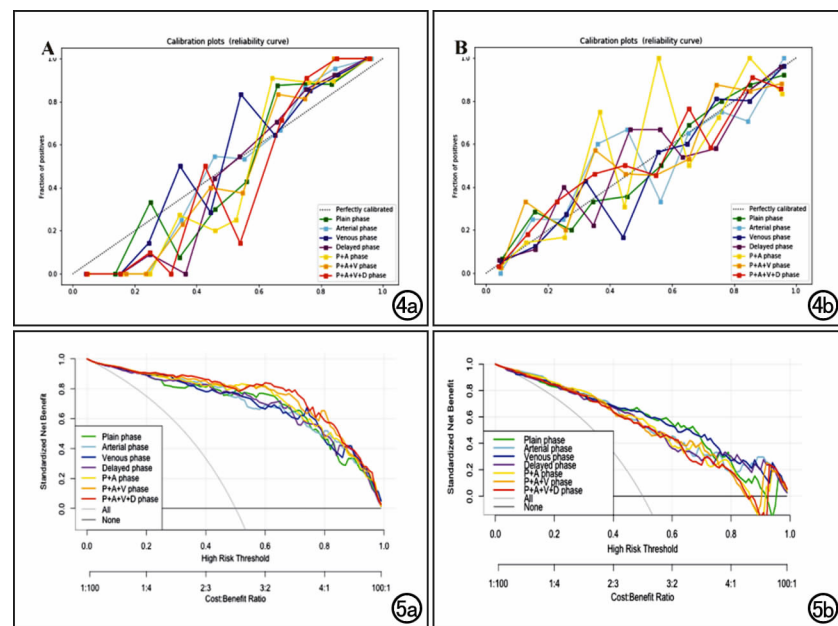


图 4 训练组和验证组影像组学模型的校准曲线, y 轴表示甲状腺腺瘤/增生的实际发生率; x 轴表示甲状腺腺瘤/增生的模型预测概率, 黑色虚线为性能完美, 而各条彩色实线代表的是基于每个期相及其组合的组学模型的实际预测性能, Hosmer-Lemeshow 检验显示所有模型的 P 值均大于 0.05。a) 训练组; b) 验证组。图 5 训练组和验证组中各影像组学模型的决策树曲线, y 轴表示净收益, x 轴表示阈值概率, 黑色水平线表示所有病变均诊断为甲状腺腺瘤/增生, 灰色曲线表示所有病变均为淋巴结的假设, 彩色实线表示如果从影像组学模型获得的阳性概率高于阈值概率, 则所有靶结节都被诊断为甲状腺腺瘤/增生。a) 训练组; b) 验证组。

讨论

该回顾性研究结果表明, 基于 4D-CT 的影像组学模型可作为鉴别甲状腺腺瘤/增生与颈部淋巴结的一种新型定量分析方法, 且具有较高的诊断准确性, 有助于甲状腺疾病的术前诊断; 基于增强扫描静脉期的影像组学模型的诊断效能最高。在临床实践中, 平扫期和动脉期在识别病变与周围组织结构之间的关系方面也提供了大量参考信息。因此, 考虑到减少辐射剂量, 而鉴于延迟期扫描未能显示出更高的诊断效能, 临床实践中可考虑不进行延迟期扫描。

与 MIBI 显像和超声相比, 4D-CT 在甲状腺腺瘤定位诊断方面的敏感度可达 55%~87%^[8,13], 尤其是在复杂病例中更具优势。尽管 4D-CT 在甲状腺腺瘤术前诊断中的应用越来越普遍, 但多期 CT 扫描的辐射剂量问题值得思考。有学者认为, 单期 CT 扫描在定位诊断方面的准确性并不亚于多期 CT 扫描, 尤其是对于在动脉期表现出独特的快速早期强化特点的病灶。有学者对传统 4D-CT 扫描的优化进行了一

系列探索, 并在甲状腺腺瘤病变的鉴别诊断方面取得了良好的效果^[8,14-17]。但是, 目前尚未见对于甲状腺腺瘤和淋巴结鉴别方面最佳扫描方案进行探讨的文献报道。

在多期扫描中, 动脉期对富血供腺瘤与淋巴结的鉴别诊断价值最大、诊断准确性最高。本研究结果也显示出相似的结果, 基于单一动脉期的影像组学模型获得了良好的鉴别诊断能力。但值得注意的是, 与单纯动脉期相比, 联合多期扫描建立的诊断模型, 大多数诊断效能指标值都有所提高, 提示经影像组学方法处理提取的 CT 图像定量数据的联合诊断具有附加价值。

本研究结果显示, 基于常规影像的增强模式可能不再是单一的参考指标, 在影像组学分析中, 其它的病变特征, 如本研究结果显示在影像组学特征中小波特征则更具权重。此外, 以往的研究大多采用定性分析的方法, 依赖于对多期扫描中病变强化模式的主观分析, 主观偏倚较强。本研究结果显示, 基于单一静脉期的影像组学模型的 AUC 值 (0.904) 最大, 相应的诊断符合率为 0.835、敏感度为 0.886、特异度为 0.785。

我们认为上述结果是合理的, 因为无论是甲状腺腺瘤还是淋巴结的强化, 在静脉期比动脉期更加均匀, 基于静脉期建立的诊断模型可以充分反映病变整体的异质性。此外, 动脉期存在的邻近头颈部静脉的放射状伪影以及病变内囊性成分等都会增加病灶分割的难度。影像组学特征 (如强度、形状、纹理或小波特征参数) 可以提供关于病变异质性的定量数据, 与病灶的其它特征 (如大小、强化程度等) 是不同的且具有互补性, 并提供了一种客观的定量描述方法。

此外, 基于平扫期的影像组学模型亦具有相对较高的诊断效能 (训练组: AUC = 0.950, 符合率 = 0.877; 验证组: AUC = 0.892, 符合率 = 0.815)。因此, 对于不能进行增强 CT 检查的患者, 如既往有对比剂过敏史、严重肾功能衰竭或对辐射剂量存在顾虑的患者, CT 平扫也能成为一种有价值的选择。

此外, 本研究结果显示基于延迟期或延迟期与其它期联合的影像组学模型并没有获得更好的诊断效能, 出于减少辐射暴露剂量的考虑, 笔者认为在临床实践中可不进行延迟期的扫描。一些机构将扫描方案调整为 3 期, 不再进行延迟期扫描, 结果显示对甲状腺

腺瘤定位诊断的敏感度并无明显降低^[16,18]。

本研究存在一定的局限性。首先,本研究为单中心回顾性研究,样本量相对较小,因此对研究结果需结合临床实际进行谨慎解读。但是,本研究中采用的是标准 4D-CT 扫描方案,在大型医疗机构中普遍应用^[19],这使得未来进行多中心大样本量的前瞻性研究成为可能,从而可进一步验证本研究获得的初步结果。第二,本研究的入组患者均为首诊且术前 MIBI 显像与超声检查结果无一致性结论的患者,而对于 4D-CT 扫描更具诊断优势的,尤其是顽固性/复发性疾病或多发病灶的复杂病例,在此尚未涉及,值得进一步研究。第三,本研究中选取的是甲状旁腺病变同侧的 II 区良性淋巴结作为自身对照,而未选取与甲状旁腺位置更接近的 IV、VI 或 VII 区淋巴结,是因为这些部位的淋巴结通常很小,勾画 VOI 存在困难,且容易受到颈部静脉对比剂伪影的干扰。影像组学方法是通过分析 VOI 的内部特征信息,提取淋巴结的一致性数据,不因位置改变而受到影响,因此笔者认为选取 II 区淋巴结也是合理的。第三,所有病灶分割都是由相同的放射科医师来完成的,可能会存在主观偏倚。我们试图通过随机分配所有图像、至少间隔 2~3 周进行重新勾画来减少此偏倚。

综上所述,基于 4D-CT 图像的影像组学模型可用于鉴别甲状旁腺腺瘤/增生与颈部淋巴结,是一种有效的定量分析方法。基于静脉期的影像组学模型显示出最好的诊断效能,而基于平扫期的影像组学模型或可为复杂情况的患者提供替代选择;考虑到减少辐射剂量,延迟期扫描可被省略。

参考文献:

- [1] 宋坤,陈杭美,王华. CT 对甲状旁腺腺瘤的诊断及其应用价值[J]. 放射学实践,2012,27(1):105-106.
- [2] Rodgers SE, Hunter GJ, Hamberg LM, et al. Improved preoperative planning for directed parathyroidectomy with 4-dimensional computed tomography[J]. Surgery, 2006, 140(6):932-940.
- [3] Kukar M, Platz TJ, Schaffner TA, et al. The use of modified four-dimensional computed tomography in patients with primary hyperparathyroidism: an argument for the abandonment of routine sestamibi single-positron emission computed tomography (SPECT)[J]. Ann Surg Oncol, 2015, 22(1):139-145.
- [4] Phitayakorn R, McHenry CR. Incidence and location of ectopic abnormal parathyroid glands[J]. Am J Surg, 2006, 191(3):418-423.
- [5] Bahl M, Sepahdari AR, Sosa JA, et al. Parathyroid adenomas and hyperplasia on four-dimensional CT scans: three patterns of enhancement relative to the thyroid gland Justify a three-phase protocol[J]. Radiology, 2015, 277(2):454-462.
- [6] 赵泓博,尹昶丽,刘畅,等. 基于 CT 双期增强影像组学预测甲状腺乳头状癌淋巴结转移[J]. 放射学实践,2021,36(4):458-463.
- [7] 张力,于淑靖,张迎,等. CT 影像组学在头颈部木村病淋巴结病变与淋巴瘤鉴别中的应用[J]. 放射学实践,2020,35(2):159-164.
- [8] 中华医学会放射学分会头颈学组,中华医学会影像技术分会辐射防护学组,头颈部 CT 检查和辐射剂量管理专家共识[J]. 中华放射学杂志,2020,54(9):827-838.
- [9] Hunter GJ, Schellingerhout D, Vu TH, et al. Accuracy of four-dimensional CT for the localization of abnormal parathyroid glands in patients with primary hyperparathyroidism [J]. Radiology, 2012, 264(3):789-795.
- [10] Starker LF, Mahajan A, Bjorklund P, et al. 4D parathyroid CT as the initial localization study for patients with de novo primary hyperparathyroidism[J]. Ann Surg Oncol, 2011, 18(6):1723-1728.
- [11] Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited[J]. Crit Care Med, 2007, 35(9):2052-2056.
- [12] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models[G]. Med Decis Making, 2006, 26(6):565-574.
- [13] Perrier ND, Edeiken B, Nunez R, et al. A novel nomenclature to classify parathyroid adenomas[J]. World J Surg, 2009, 33(3):412-416.
- [14] Silver E, Sadeghi N, Knoll S, et al. Utility of single-phase computed tomography in identifying parathyroid adenomas: a feasibility study[J]. Curr Probl Diagn Radiol, 2018, 47(2):90-93.
- [15] Noureldine SI, Aygun N, Walden MJ, et al. Multiphase computed tomography for localization of parathyroid disease in patients with primary hyperparathyroidism: how many phases do we really need[J]. Surgery, 2014, 156(6):1300-1306.
- [16] Beland MD, Mayo-Smith WW, Grand DJ, et al. Dynamic MDCT for localization of occult parathyroid adenomas in 26 patients with primary hyperparathyroidism[J]. AJR, 2011, 196(1):61-65.
- [17] Hunter GJ, Ginat Dt Fau-Kelly HR, Kelly Hr Fau-Halpern EF, et al. Discriminating parathyroid adenoma from local mimics by using inherent tissue attenuation and vascular information obtained with four-dimensional CT: formulation of a multinomial logistic regression model[J]. Radiology, 2014, 270(1):168-175.
- [18] Kelly HR, Hamberg LM, Hunter GJ. 4D-CT for preoperative localization of abnormal parathyroid glands in patients with hyperparathyroidism: accuracy and ability to stratify patients by unilateral versus bilateral disease in surgery-naïve and re-exploration patients[J]. AJNR, 2014, 35(1):176-181.
- [19] Hoang JK, Sung WK, Bahl M, et al. How to perform parathyroid 4D CT: tips and traps for technique and interpretation [J]. Radiology, 2014, 270(1):15-24.

(收稿日期:2021-03-27 修回日期:2021-06-25)

· 头颈部影像学专题 ·

鼻骨区骨折深度学习模型的建立和临床效能评估

陶建华, 陈聪, 张怀宇, 曲晓霞, 郭健, 鲜军舫

【摘要】 目的:探讨基于深度学习算法建立的 AI 模型对鼻骨区骨折的检出效能。**方法:**回顾性搜集 1000 例无骨折患者和 1080 例有骨折患者的鼻骨 CT 图像。以 2 位具有 15 年以上鼻骨区骨折诊断经验的高年资医师一致认定的有或无骨折的诊断结果作为金标准。使用三维特征 2.5D 金字塔网络 (FPN) 深度学习算法建立 AI 模型并完成鼻骨区骨折定位的训练和测试。采用敏感度、特异度和符合率评估 AI 模型、人工独立检出 (2 位低年资 A、B 和 1 位高年资医师 C) 以及 AI 辅助人工检出三种方式对鼻骨区骨折的检出效能。**结果:**AI 模型检出骨折的敏感度为 86.64%, 特异度为 41.99%, 符合率为 68.71%。AI 辅助低年资医师 A 和 B 检出骨折的敏感度和符合率较人工独立检出分别提高了 24.35%、23.54% 和 17.81%、16.97%, 差异均有统计学意义 ($P < 0.05$); AI 辅助高年资医师对骨折的检出效能与人工独立检出相比, 差异无统计学意义 ($P > 0.05$)。**结论:**人工智能模型能够提高低年资医师对鼻骨区骨折的诊断效能, 缩小与高年资医师之间的差距。

【关键词】 人工智能; 深度学习; 鼻骨; 骨折; 诊断效能

【中图分类号】 R814.42; R683.5 **【文献标识码】** A **【文章编号】** 1000-0313(2021)08-0959-06

DOI: 10.13609/j.cnki.1000-0313.2021.08.003

开放科学 (资源服务) 标识码 (OSID):



Establishment and evaluation of the clinical efficacy of deep learning model for detection of nasal bone fracture on CT images TAO Jian-hua, CHEN Cong, ZHANG Huai-yu, et al. Department of Radiology, Beijing Tongren Hospital, Capital Medical University, Beijing 100730, China

【Abstract】 Objective: To establish a detection model of nasal bone fracture based on deep learning of artificial intelligence (AI) and to evaluate its detection efficiency. **Methods:** The original axial CT images of nasal bone in 1000 patients without fracture and 1080 patients with fracture were collected retrospectively. The diagnostic results of two senior radiologists with more than 15 years of work experience specialized in nasal bone fracture were taken as gold standard. AI model established with 3D Feature Pyramid Network (FPN) deep learning method was used to train and test the localization of nasal bone area, while 2.5D FPN method was applied to train the location of nasal bone fracture area. The sensitivity, specificity and accuracy were employed to evaluate the performance on detecting nasal bone fracture by AI model, radiologists (2 junior and 1 senior) and the radiologists with aid of AI model. **Results:** The sensitivity, specificity and accuracy of AI model in detecting nasal bone fracture were 86.64%, 41.99% and 68.71%, respectively. The sensitivity and accuracy of junior doctor (A and B) assisted by AI model in fracture detection showed 24.35%, 23.54%, and 17.81%, 16.97% increase when compared with junior doctor without AI assistant with statistically significant difference ($P < 0.05$). The sensitivity, specificity and accuracy of AI model combined with senior doctor in fracture detection were not significantly different ($P > 0.05$) compared with those of independent detection. **Conclusion:** The AI detection model of nasal bone fracture can improve the sensitivity and accuracy of fracture detection for junior radiologist, thus can effectively reduce the difference in detection performance between junior and senior radiologists.

【Key words】 Artificial intelligence; Deep learning; Nasal bone; Fracture; Diagnostic efficacy

作者单位: 100730 北京, 首都医科大学附属北京同仁医院放射科 (陶建华, 张怀宇, 曲晓霞, 郭健, 鲜军舫); 100730 北京, 汇医慧影科技有限公司 (陈聪)

作者简介: 陶建华 (1973—), 女, 北京人, 硕士, 副主任医师, 主要从事为头颈部影像诊断。

通讯作者: 鲜军舫, E-mail: ejr_xianjunfang@vip.163.com

基金项目: 证据科学教育部重点实验室 (中国政法大学) 开放基金资助课题 (2018KFKT05); 北京市医院管理局临床医学发展专项 (ZYLX201704); 北京市医院管理中心“登峰”计划专项 (DFL20190203)

鼻骨区骨质菲薄,位置浅在,是颅面部外伤中最常受累部位。由于鼻骨区体积较小,解剖结构细微,该部位骨折容易被漏诊及误诊^[1-3]。在临床急诊工作中,颅面部外伤患者常涉及到法医鉴定等刑事问题,在工作量大、时间紧和视觉疲劳的情况下,能否正确判断鼻骨区骨折显得尤为重要^[4-5]。深度学习是一种人工智能(artificial intelligence, AI)方法,通过学习样本数据的内在规律和表示层次,通过组合低层特征形成更加抽象的表示高层属性的特征,使机器模仿人类的视听和思考活动。近年来人工智能技术在疾病诊断及预后评估方面逐渐发挥出较大作用,基于深度学习的 AI 模型构建的肺结节、颈部血管和冠状动脉疾病检测系统已逐步应用于临床实践^[6-9]。此外,虽然 AI 方法已应用于四肢骨、骨盆、椎体和肋骨骨折的检测且已取得较好的进展^[10-12],但是在颅面部骨折方面的临床应用尚未见到相关报道。本课题拟基于大样本的鼻骨区骨折和非骨折影像数据,建立鼻骨区骨折的人工智能辅助诊断模型,探讨此模型在辅助影像医师初步筛查骨折部位、减少医师因过度疲劳导致可能的漏诊和误诊,提高诊断准确性方面的临床应用价值。

材料与方法

1. 一般资料

回顾性分析 2018 年 1 月—2019 年 8 月因鼻骨区外伤来我院急诊行鼻骨 CT 检查的 2080 例患者的 CT 资料。经 2 位高年资医师(工作年限>15 年)共同认定(金标准),其中 1000 例无明确鼻骨区骨折,1080 例有明确鼻骨区骨折,所有患者均无鼻骨区和副鼻窦区肿瘤。无骨折组中,男 667 例,女 333 例,年龄 18~75 岁,平均(38.2±8.2)岁;骨折组中,男 625 例,女 455

例,年龄 18~80 岁,平均(36.5±7.1)岁。

本研究经我院伦理委员会批准(TRECKY2019-014)。

2. CT 扫描方法

使用 Philips brilliance 16 排螺旋 CT 机,扫描范围自鼻根部到鼻尖下方层面,扫描参数:120 kV, 200 mAs,螺距 0.438,矩阵 512×512,层厚 0.8 mm,层间距 0.4 mm,骨算法重建。每例患者获得原始骨算法图像 200~300 帧,设置图像的窗宽为 4000 HU、窗位为 700 HU。屏蔽患者个人信息后将所有图像上传到 AI 工作站进行分析,研究步骤和流程见图 1。

3. 图像标注

无骨折组中,在每例患者的原始 CT 图像中选取鼻根部层面、鼻骨区中间层面和上颌骨额突最上端这 3 个层面,使用白色矩形框(可隐藏)对鼻骨进行标注,矩形框内应包括双侧鼻骨、上颌骨额突和鼻中隔前部。1000 例中 800 例用于 AI 模型的鼻骨区定位训练,100 例用于 AI 模型的测试,100 例用于 AI 辅助下不同医师对骨折检出效能的评估。

骨折组中,在每例患者鼻骨区原始 CT 图像上采用矩形框对所有层面的骨折部位进行标注。先由低年资医师进行标注,然后由具有 15 年以上鼻骨区骨折诊断经验的高年资医师进行审核和确认。矩形框内应包含骨折线,面积尽量小。1080 例中 680 例用于训练,240 例用于 AI 测试,160 例用于 AI 辅助下医师对骨折检出效能的评估。

4. 基于深度学习的 AI 模型的训练和测试

鼻骨区的定位训练和测试:本研究中使用 3D 特征金字塔网络(Feature Pyramid Network, FPN)深度学习算法^[13],基于无骨折组的鼻骨区标注图像进行定

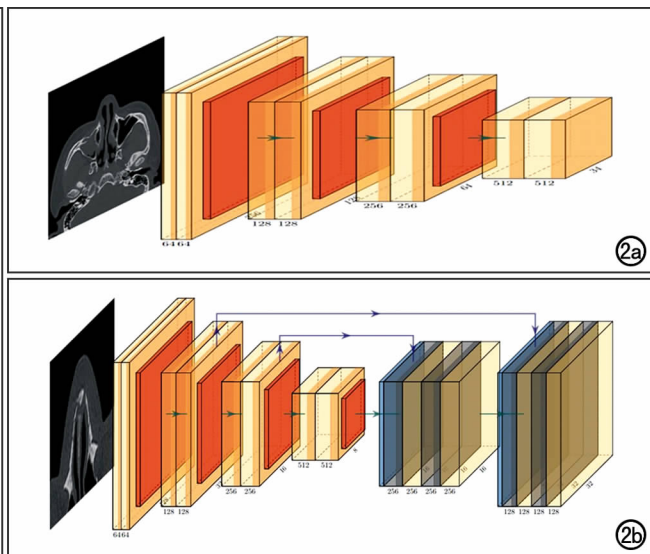
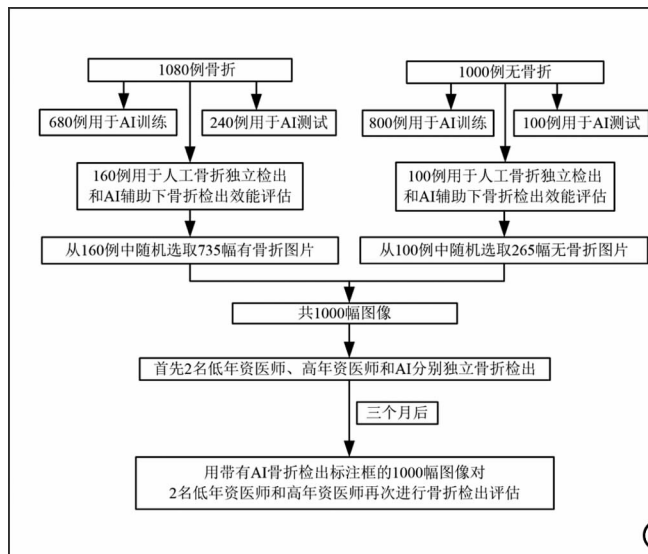


图 1 本研究步骤和方法的流程图。图 2 基于 FPN 算法的深度学习对鼻骨图像学习过程的示意图。

a)鼻骨区定位训练;b)鼻骨区骨折的定位训练。

位训练来建立 AI 模型,具体过程见图 2a。鼻骨的标注从鼻骨最上端起始位置到上颌骨额突结束的位置,统一将尺寸归化为矩阵 $128 \times 256 \times 256$,使用 3 个卷积特征的残差网络跳转链接方式和降采样,将 CT 图像经过统一化处理,最终的特征是 512 通道,矩阵 $16 \times 32 \times 32$ 。经 100 例测试,鼻骨区定位符合率达 100%。

鼻骨区骨折的定位训练和测试:对鼻骨区骨折的定位训练过程如图 2b 所示。训练样本包括 680 例有骨折和 200 例无骨折(从用于 AI 模型的鼻骨区定位训练的 800 例无骨折中随机选取)患者的原始 CT 图像。本文设计了 2.5D 的 FPN 方法实现网络鼻骨区骨折定位,将相邻的 3 层切片堆叠起来作为输入图像,

扩大空间信息,使骨折集中在 $5 \times 5 \sim 12 \times 12$ 像素中显示。网络在特征提取时将最后一层进行反卷积堆叠采样,提供更多的空间信息。选取 3560 帧图像(从 240 例骨折病例中随机选取 2560 帧有骨折图像,从 100 例无骨折病例中随机选取 1000 帧无骨折图像,对 AI 模型的骨折定位能力进行测试。

5. 人工和 AI 辅助对骨折检出效能的比较

选取 1000 帧原始图像(从 160 例骨折病例中随机选取 735 帧有骨折图像,从 100 例无骨折病例中随机选取 265 帧无骨折图像),首先由 2 位低年资医师(工作年限 < 3 年)、1 位高年资医师(工作 > 10 年)和 AI 模型分别独立对图像进行分析和标注。3 个月后,由

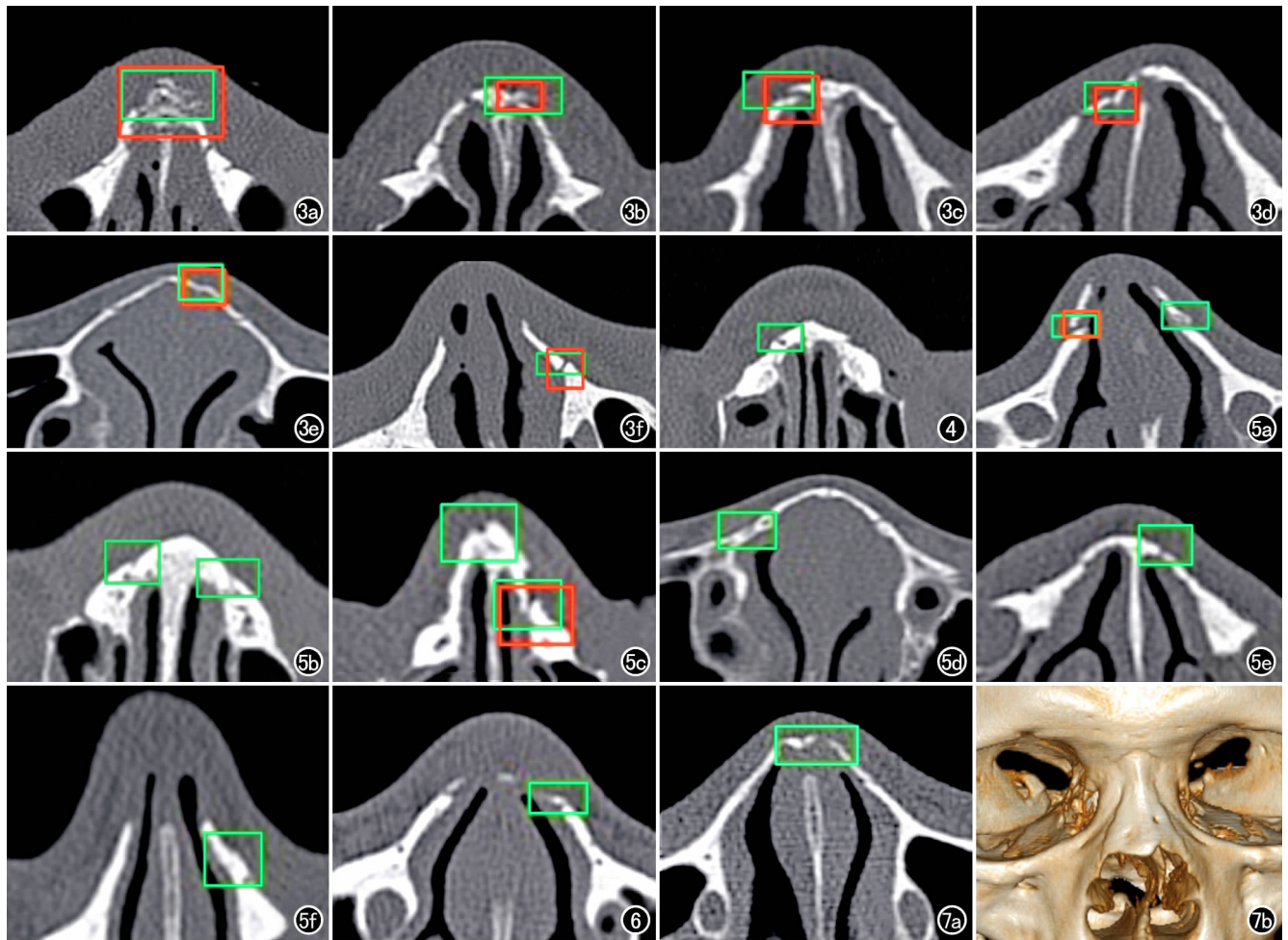


图 3 人工(红色框)与 AI 模型(绿色框)标注骨折的一致性较高。a)双侧鼻骨粉碎骨折,骨性鼻中隔前部骨折;b)左侧鼻骨骨折,断端成角;c)右侧鼻骨骨折,断端分离;d)右侧鼻骨骨折,断端分离;e)左侧鼻骨轻微骨折、略塌陷,骨折线不清晰;f)左侧上颌骨额突线性骨折,断端无移位。

图 4 AI 模型把右侧血管沟(鼻骨孔)误诊为骨折。图 5 AI 模型把透亮点状或线状血管沟、缝间骨、鼻颌缝误诊为骨折(绿框)。a) AI 模型(左侧绿框)将左侧鼻颌缝内缝间骨误诊为骨折,人工(红框)和 AI 模型(右侧绿框)均检出右侧鼻颌缝分离;b) AI 模型将左侧鼻颌缝和右侧血管沟误诊为骨折;c) AI 模型(左上绿框)将右侧血管沟误诊为骨折,同时人工(右下绿框)和 AI 模型(红框)均检出左侧上颌骨额突骨折;d) AI 模型将右侧缝间骨和血管沟均误诊为骨折;e) AI 模型将左侧血管沟误诊为骨折;f) AI 模型将左侧鼻颌缝误诊为骨折。图 6 AI 模型把鼻骨末端的鼻颌缝误诊为骨折(绿框)。图 7 无鼻骨骨折患者。a)横轴面 CT 图像, AI 模型将左侧缺损型鼻骨末端误诊为骨折(绿框);b) VR 图像示左侧缺损型鼻骨末端。

上述 3 位医师分别对 AI 模型标注过的 CT 图像再次进行分析和观察,对骨折部位进行标注。

在评估过程中,AI 标注框和人工标注框在 x、y 轴方向的重叠率均大于 50%且重叠面积大于 25%视为正确识别(图 3)。标注框正确识别骨折定义为真阳性;标注框未正确识别骨折定义为假阴性;标注框识别错误定义为假阳性(包括骨折组和无骨折组);在无骨折图像上无标注框定义为真阴性。

6. 统计学方法

使用 SPSS 24.0 统计学软件,采用四格表或 R×C 表 Pearson 卡方检验对人工标注与 AI 辅助下人工标注骨折部位的敏感度、特异度和符合率进行比较。采用 ROC 曲线来分析各种标注方式对骨折的检出效能,通过 Delong 检验比较各种标注方式的 ROC 曲线下面积(AUC)。以 $P<0.05$ 为差异有统计学意义。

结 果

在鼻骨骨折的测试集中,AI 模型检出骨折的敏感度为 86.64%(2218/2560),特异度为 41.99%(721/1717),符合率为 68.71%(2939/4277);AI 模型检出的假阳性骨折数为 996 个,常见位置依次为血管沟、鼻颌缝和鼻骨末端(图 4~7),分别占 48.99%、29.91%和 21.10%。

三位医师单独及在 AI 模型辅助下对鼻骨骨折的诊断效能指标及统计分析结果见表 1~3。两位低年资医师在 AI 辅助下检出鼻骨骨折的敏感度和符合率均较独立检出有显著提高,分别提高了 24.35%、

23.54%和 17.81%、16.97%,差异均有统计学意义($P<0.05$),AUC 亦有明显提高($P<0.05$),但特异度的变化不明显($P>0.05$)。高年资医师在 AI 模型辅助下对鼻骨骨折的检出敏感度、特异度和符合率与独立检出比较,差异均无统计学意义($P>0.05$)。在 AI 模型辅助下 2 位低年资医师检出骨折的敏感度和符合率与高年资医师间的差距明显缩小,敏感度的差距从 29.39%和 25.45%分别缩小到 6.13%和 3.00%,符合率的差距从 28.07%和 23.46%缩小到 11.21%和 7.44%。在 3 位医师的各 2 种诊断方式中,以 AI 模型辅助下高年资医师的诊断效能最高,AUC 最大(图 8)。

讨 论

传统深度学习模型需要百万级的数据集,在训练数据相对较少的条件下,本研究采用迁移学习的预训练模型,使用左右翻转和直方图随机扰动策略进行训练,每次训练时对数据进行随机变换以实现数据的扩充,从而解决数据集较小的问题。对鼻骨区骨折的检出属于对微小目标的识别,骨折区域在整个标准框立方体中占比太小,导致标注区域的噪声过多,训练效果较差。针对这种特殊情况,本研究采用 2.5D 网络,将相邻的前后两层堆叠起来作为输入图像,从而可降低噪声并弥补空间信息的不足。对于无明显移位的轻微鼻骨区骨折,除了局部信息,还需要更大空间视野来辨识骨折、骨缝和血管沟,本研究中使用特征金字塔网络,在特征提取时将最后一层进行反卷积堆叠来进行采样,这种方法可以提供更多的空间信息,增加对骨

表 1 低年资医师 A 及在 AI 模型辅助下对鼻骨骨折诊断效能的比较

效能指标	医师 A	医师 A+AI 模型	χ^2/Z 值	P 值
敏感度	68.16%(501/735)	92.51%(680/735)	19.91	<0.001
特异度	57.72%(198/343)	59.55%(187/314)	0.08	0.774
符合率	64.84%(699/1078)	82.65%(867/1049)	8.42	0.004
AUC [#]	0.673(0.644~0.702)	0.777(0.751~0.802)	11.06	<0.001

注: # 括号内数据为 95%CI。

表 2 低年资医师 B 及在 AI 模型辅助下对鼻骨骨折诊断效能的比较

效能指标	医师 B	医师 B+AI 模型	χ^2/Z 值	P 值
敏感度	72.10%(530/735)	95.64%(703/735)	21.43	<0.001
特异度	63.52%(209/329)	64.14%(195/304)	0.00	1.000
符合率	69.45%(739/1064)	86.42%(898/1039)	8.29	0.004
AUC [#]	0.714(0.685~0.741)	0.808(0.782~0.831)	9.55	<0.001

注: # 括号内数据为 95%CI。

表 3 高年资医师 C 及在 AI 模型辅助下对鼻骨骨折诊断效能的比较

效能指标	医师 C	医师 C+AI 模型	χ^2/Z 值	P 值
敏感度	97.55%(717/735)	98.64%(725/735)	0.34	1.000
特异度	81.35%(240/295)	81.78%(238/291)	0.33	1.000
符合率	92.91%(957/1030)	93.86%(963/1026)	0.08	1.000
AUC [#]	0.895(0.875~0.913)	0.902(0.883~0.920)	0.09	0.931

注: # 括号内数据为 95%CI。

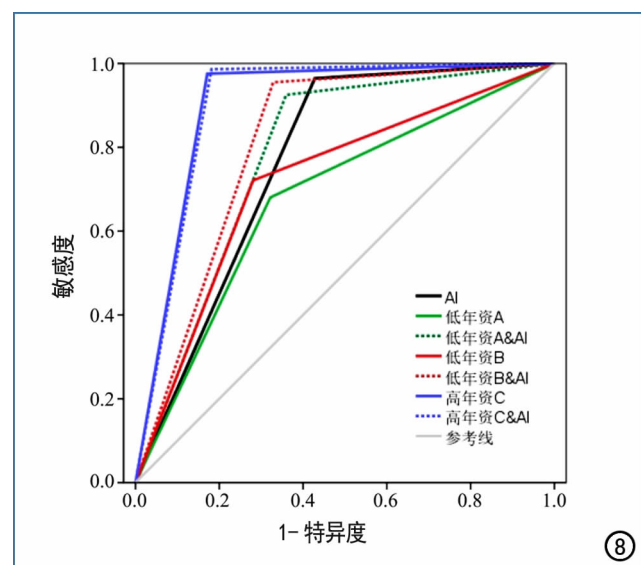


图 8 低年资和高年资医师独立及在 AI 模型辅助下检出骨折的 ROC 曲线,以 AI 模型辅助下高年资医师的诊断效能最高,AUC 最大。

折、骨缝和血管沟的识别。

低年资医师在 AI 模型的辅助下检出骨折的敏感度和符合率均有显著提高,AUC 亦有明显增加,说明 AI 模型可以帮助低年资医师对骨折进行初步筛查。急诊值班工作中,工作量大,外伤严重且复杂,在这种高强度脑力劳动和容易视觉疲劳的环境下,AI 模型辅助筛查鼻骨区骨折,可以在一定程度上减少漏诊,提高诊断效能。此外 AI 模型缩小了低年资医师与高年资医师检出骨折的敏感度和符合率之间的差距,间接减少了高年资医师审核报告的时间。

值得注意的是 AI 模型检出骨折的假阳性率较高(图 4~7),假阳性标注出现的部位主要为鼻颌缝(占 44.19%)、血管沟(占 41.31%)和鼻骨末端(占 14.50%),这些区域也是临床实际工作中最容易被影像医师与骨折相混淆的部位。在横轴面图像上,鼻颌缝中下部分常常出现缝间骨征象,单独观察某一层面容易误认为骨折碎片。鼻骨区血管沟孔较多,除了固定出现的鼻骨孔以外,鼻骨和上颌骨额突还存在其它的一些血管孔沟,且常不对称出现,在单独某个层面上也极易与骨折相混淆。鼻骨末端的形态多变,有 M 型、锯齿型、单侧缺损型和拱型等多种形态,在横轴面图像上,锯齿型和单侧缺损型的鼻骨末端常表现为双侧不对称的局部缺损,容易被误认为碎片和单侧骨折。因此,在临床实践中,应连续上下多层面、多角度并结合 VR 图像来观察鼻颌缝、鼻骨孔、血管沟位置和鼻骨末端形态等,有助于将上述解剖变异与鼻骨骨折进行鉴别。AI 模型评估鼻骨骨折时出现假阳性的主要原因是 AI 模型更多关注单层的局部信息,对某一

结构上下层面的连续变化的空间信息观察不足;其次的原因是用于训练基于深度学习算法的 AI 模型的样本数量不够大。低年资医师检出假阳性骨折的主要原因也是受到鼻颌缝、血管沟和鼻骨末端不规则形态的干扰。

本研究中 AI 模型检出骨折的敏感度为 86.64%,低年资医师平均约为 70.00%,低于 AI 模型,差距主要存在于鼻骨区的轻微线性骨折,分析原因主要是低年资医师的工作时间短、对线性骨折经验不足所致。鼻骨的骨质较菲薄,轻微的线性骨折往往仅能见到局部骨质有线状透亮影,断端通常无明显移位。在工作量大、时间紧的急诊工作中,医师在视觉疲劳的情况下,对于轻微的线性骨折更容易漏诊,可以利用容积再现(volume rendering, VR)技术进行图像重组,获得鼻骨及邻近区域的三维立体图像,从而能对鼻骨区进行全面观察,提高对线性骨折的检出率,减少漏诊。

鼻骨区包括鼻骨、上颌骨额突和骨性鼻中隔,均属于不规则骨,和规则的四肢长骨和扁骨相比,解剖结构复杂,其形态和邻近血管的变异较多,这给应用 AI 模型来检测鼻骨区骨折带来困难和挑战。本研究基于深度学习算法初步建立了检测鼻骨区骨折的 AI 模型,这是将 AI 技术应用于颅面部骨折的新尝试。在 AI 模型辅助下,低年资医师对鼻骨骨折的检出效能有一定程度的提高,不足之处在于 AI 标注框的面积普遍较大、精准度不高,检出骨折的假阳性率仍较高。针对上述问题,笔者拟在后续研究中通过增大训练 AI 模型的样本量和多样性、进行多中心研究和优化算法,来提高 AI 模型对解剖结构和骨折的空间识别能力、减少假阳性率,并通过缩小标注框来提高骨折定位的精准度,增加诊断医师对 AI 骨折检出模型的信任度,使其能更好地应用于颅面部骨折的筛查和法医鉴定中。

参考文献:

- [1] 邵迎吉,张炜,李峻,等. CT 影像上易误诊为鼻骨骨折的正常结构[J]. 临床放射学杂志, 2009, 28(8): 1073-1075. DOI: CNKI: SUN: LCFS. 0. 2009-08-011.
- [2] 敖振杰,张得旺,黄炎标,等. 鼻区骨细微结构及其骨折的 MSCT 表现[J]. 实用放射学杂志, 2012, 28(3): 358-362, 383. DOI: 10. 3969/j. issn. 1002-1671. 2012. 03. 010.
- [3] 王珮华. 我国鼻骨骨折诊治的现状与建议[J]. 中华耳鼻咽喉头颈外科杂志, 2020, 55(3): 191-194. DOI: 10. 3760/cma. j. issn. 1673-0860. 2020. 03. 001.
- [4] 梁建军,吕金星,殷占英. 鼻部骨折的法医学鉴定[J]. 北京警察学院学报, 2017, 5(3): 97-103. DOI: 10. 16478/j. cnki. jbjpc. 20170524. 002.
- [5] 沈靓,沈宇. 法医影像学深入应用探讨[J]. 广东公安科技, 2019, 137(3): 69-71. DOI: CNKI: SUN: GDGA. 0. 2019-03-019.
- [6] Gong L, Jiang S, Yang Z, et al. Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation net-

- works[J]. Int J Comput Assist Radiol Surg, 2019, 14(11):1969-1979. DOI:10.1007/s11548-019-01979-1.
- [7] Li L, Liu Z, Huang H, et al. Evaluating the performance of a deep learning-based computer-aided diagnosis (DL-CAD) system for detecting and characterizing lung nodules: comparison with the performance of double reading by radiologists[J]. Thorac Cancer, 2019, 10(2):183-192. DOI:10.1111/1759-7714.12931.
- [8] Saxena A, Ng E, Lim ST. Imaging modalities to diagnose carotid artery stenosis: progress and prospect[J]. Biomed Eng Online, 2019, 18(1):66-71. DOI:10.1186/s12938-019-0685-7.
- [9] Moon JH, Lee DY, Cha WC, et al. Automatic stenosis recognition from coronary angiography using convolutional neural networks [J/OL]. Comput Methods Programs Biomed, 2021, 198:e105819. DOI:10.1016/j.cmpb.2020.105819. Epub 2020 Nov 2.
- [10] Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians[J]. Proc Natl Acad Sci USA, 2018, 115(45):11591-11596. DOI:10.1073/pnas.1806905115.
- [11] Urakawa T, Tanaka Y, Goto S, et al. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network[J]. Skeletal Radiol, 2019, 48(2):239-244. DOI:10.1007/s00256-018-3016-3.
- [12] Wani IM, Arora S. Computer-aided diagnosis systems for osteoporosis detection: a comprehensive survey[J]. Med Biol Eng Comput, 2020, 58(9):1873-1917. DOI:10.1007/s11517-020-02171-3.
- [13] Jin Z, Liu B, Chu Q, et al. SAFNet: a Semi-anchor-free network with enhanced feature pyramid for object detection[J/OL]. IEEE Trans Image Process, 2020. DOI:10.1109/TIP.2020.3028196. PMID:33026987.

(修改日期:2021-04-17 修回日期:2021-06-30)

《放射学实践》杂志微信公众平台开通啦!

遵照同行评议、价值导向、等效应用原则,国内各大学会、协会、组织机构通过科技工作者推荐、专家评议、结果公示等规定程序,《放射学实践》杂志入选中国科协发布 10 大领域《我国高质量科技期刊分级目录》业内认可的较高水平期刊。《放射学实践》杂志入选 2020 年版北京大学和北京高校图书馆期刊工作研究会共同主持的国家社会科学基金项目“学术期刊评价及文献计量学研究”研究成果——《中国核心期刊要目总览》。

《放射学实践》杂志微信公众平台立足于准确地传递医学影像领域的最新信息,致力于为关注医学影像领域的广大人士服务。欢迎大家通过微信平台,以文字、图片、音频和视频等形式与我们互动,分享交流最新的医学影像资讯。您还可以通过微信平台免费阅读及搜索本刊所有发表过的论文,投稿作者可以查询稿件状态等。

您可以通过以下方式关注《放射学实践》杂志微信公众平台:

1. 打开微信,通过“添加朋友”,在搜索栏里直接输入“放射学实践”进行搜索。
2. 在“查找微信公众号”栏里输入“放射学实践”即可找到微信公众号,点击“关注”,添加到通讯录。
3. 打开微信,点击“扫一扫”,手机镜头对准下面的二维码,扫出后点击关注即可。



关注有惊喜!

• 头颈部影像学专题 •

基于 CT 影像组学模型鉴别头颈部良、恶性淋巴结

胡大涛, 夏春华, 李羚, 高斌

【摘要】 目的:探讨 CT 影像组学模型在头颈部淋巴结良恶性鉴别诊断中的应用价值。**方法:**回顾性分析 2010—2019 年本院经病理或穿刺活检方法证实为良性或恶性淋巴结的 200 例患者的临床和 CT 资料,其中良性组 105 例,恶性组 95 例。将全部患者随机分为训练集(133 例)和测试集(67 例),运用 Mazda 软件提取淋巴结的 CT 影像组学特征,使用 LASSO 方法降维后,建立影像组学标签(score)。基于有鉴别意义的变量,包括临床指标及淋巴结短径(Size)、增强动脉期 CT 值(Z)和影像组学标签等,采用多因素 Logistic 回归分析,分别建立基于影像组学和非影像组学(Size+Z)的预测模型,使用校准曲线观察两种模型的拟合情况。**结果:**经过特征降维后,将获得的 $S(2, -2)$ Correlat、 $S(0, 3)$ InvDfMom 和 $S(4, 0)$ Contrast 这 3 个影像组学特征用于建立影像组学标签,其在训练集和测试集中鉴别良恶性淋巴结的 AUC 分别为 0.884 和 0.749。经 Logistic 回归法建立的影像组学预测模型在训练集和测试集中的 AUC(分别为 0.958 和 0.908)均大于非影像组学预测模型(分别为 0.847 和 0.806),差异均有统计学意义($P < 0.05$),且两种模型的校准曲线与理想曲线显示出良好的拟合效果。**结论:**CT 影像组学模型在头颈部淋巴结良恶性的鉴别诊断中具有一定的应用价值。

【关键词】 头颈部; 淋巴结转移; 体层摄影术, X 线计算机; 影像组学

【中图分类号】 R814.42; R733.4 **【文献标志码】** A **【文章编号】** 1000-0313(2021)08-0965-06

DOI:10.13609/j.cnki.1000-0313.2021.08.004

开放科学(资源服务)标识码(OSID):

**Differentiation of benign and malignant lymph nodes in the head and neck based on CT radiomics model**

HU Da-tao, XIA Chun-hua, LI Ling, et al. Department of Radiology, the Third Affiliated Hospital of Anhui University, Anhui 230009, China

【Abstract】 Objective: The purpose of this study was to investigate the application value of CT-based radiomics model in the differential diagnosis of benign and malignant lymph nodes in the head and neck regions. **Methods:** The clinical data and CT images of 200 patients with benign or malignant lymph nodes confirmed by pathology or puncture biopsy in our hospital from 2010 to 2019 were retrospectively analyzed. There were 105 patients in the benign group and 95 patients in the malignant group. All patients were randomly divided into training set ($n=133$) or validation set ($n=67$). Mazda software was used to extract the radiomics features of lymph nodes based on CT images, and LASSO method was used to reduce the dimensionality and establish the radiomics score. After multivariate logistic regression analysis, radiomics-based and non radiomics-based (Size+Z) diagnostic models were established from relevant variables shown statistical difference between the benign and malignant lymph nodes, including the clinical index of patients and short diameter of (Size), CT values in enhanced arterial phase (Z) and radiomics score of lymph nodes. Then calibration curve was used to observe the fitting of the two models. **Results:** After feature dimensionality reduction, three radiomics features including $S(2, -2)$ Correlat, $S(0, 3)$ InvDfMom and $S(4, 0)$ Contrast were selected out for establishing the radiomics score. ROC curve showed that the AUC of the score in the training cohort and the test cohort was 0.884 and 0.749, respectively. The radiomics diagnostic model showed better discrimination efficacy than that of non-radiomics model in training cohort (AUC:0.958 vs. 0.908, $P < 0.05$) and validation cohort (AUC:0.847 vs. 0.806, $P < 0.05$), the calibration curve and ideal curve of the two models show good fitting effect. **Conclusion:** CT-based radiomics model is valuable in the dif-

作者单位:230009 合肥,安徽医科大学第三附属医院影像科(胡大涛,夏春华,高斌),五官科(李羚)

作者简介:胡大涛(1990—),男,安徽滁州人,硕士研究生,主要从事头颈部肿瘤的影像诊断与鉴别诊断。

通讯作者:夏春华, E-mail: xiachunhua3775@sina.com

ferential diagnosis of benign and malignant lymph nodes in the head and neck regions.

【Key words】 Head and neck; Lymph nodes; Tomography, X-ray computed; Radiomics

头颈部恶性肿瘤包括鼻咽癌、口咽癌、下咽癌、淋巴瘤及甲状腺癌等一系列肿瘤,全球范围内其发病率位居肿瘤发病率的第 7 位^[1]。头颈部淋巴结转移是影响头颈部恶性肿瘤预后的首要因素,因此准确评估淋巴结的良、恶性,对制订治疗计划和判断预后有重要作用。淋巴结的形状、大小、边界及有无坏死等目前依旧是临床判断淋巴结良、恶性的主要依据^[2],其中淋巴结短径 ≥ 1.0 cm 具有重要鉴别诊断价值^[3]。但是临床上发现肿大淋巴结也可能因炎性反应性增生所致,而正常大小的淋巴结亦有可能已发生肿瘤浸润。淋巴结的大小和发生坏死是最重要的影像学评估标准,但是单纯依靠淋巴结来评估淋巴结良、恶性的准确性并不高^[4]。常规 CT 及 MRI 平扫对判断淋巴结性质的准确性偏低,尤其对较小淋巴结,当其短径 < 1.0 cm 时这 2 种方法的错判率和漏诊率较高。尽管增强 CT 或 MRI 提高了对淋巴结中心坏死的检出率,对鉴别淋巴结的良、恶性有一定价值,但仍旧缺乏特异性。近年来,随着影像检查技术的快速发展,逐步由常规形态学诊断拓展到更先进的功能成像、分子影像等,功能成像在鉴别淋巴结良、恶性中的作用已逐渐成为研究热点。研究较多的功能成像技术主要有 DWI 和表观扩散系数(ADC)、超顺磁性氧化铁增强 MRI、CT 灌注成像及氢质子波谱成像等,这些技术在一定程度上提高了诊断的准确性。如周镇源等^[5]的研究表明颈部淋巴结的坏死区与实性区 ADC 值的比值有助于鉴别颈部淋巴结病变的良恶性。但上述功能成像技术也存在一定的缺点:如 DWI 图像的清晰度及分辨率较差,在头颈部扫描时可产生伪影;超顺磁性氧化铁的主要缺点是可用性有限,并且缺少标准的人体射频线圈和成像方法^[6]。目前在临床诊疗中,对疾病的诊断主要依赖于影像医师根据病灶的部位、形态、大小、密度、信号和强化方式等形态学征象进行主观判读。随着医学的发展,医疗模式逐渐转向精准医学模式,传统的影像信息利用方法不能满足精准医学的发展需要,亟需发展新的图像分析方法,充分挖掘影像图像中蕴含的深层信息,以实现肿瘤患者的精准诊疗。

影像组学(radiomics)是通过高通量地从医学影像中提取大量影像信息,采用自动化算法自 ROI 内提取出大量的特征信息,深度挖掘影像中蕴含的信息,筛选出具有特征性的数字化信息,并建立相关分类及预测模型,进而定量描述肿瘤表型,从而在肿瘤的诊断、治疗和预后评估等方面发挥巨大作用^[7]。近年来,影像组学已在肿瘤分类及表型分析、肿瘤治疗疗效评估

预后预测等方面取得了相当丰富的研究成果^[8-10]。影像组学分析不仅具有无创性,而且能够重复使用数据信息,能对病灶整体进行分析,从而有效避免了穿刺活检及病理检查存在的取样误差等问题。我们基于 CT 影像资料,将影像组学技术应用于头颈部淋巴结良、恶性的评估,旨在解决临床医师难以成功定性咽后等深部间隙内淋巴结这一难题,提高影像诊断水平。

材料与方法

1. 病例资料

回顾性分析本院 2010—2019 年经手术病理或穿刺活检证实为良性或恶性淋巴结的 200 例患者的临床和术前 CT 图像(平扫及增强扫描动脉期)。患者年龄 19~80 岁,平均 (51.86 ± 14.31) 岁。恶性淋巴结组 95 例,其中 6 例为淋巴瘤,51 例为鼻咽癌伴淋巴结转移,18 例为甲状腺癌伴淋巴结转移,20 例为口咽和喉咽部、舌部、腭部及颊面部恶性肿瘤伴淋巴结转移;良性淋巴结组 105 例,其中 4 例经临床确诊为淋巴结结核,其余 101 例均为淋巴结清扫术后病理切片或穿刺活检证实为淋巴结转移阴性。

纳入标准:①经病理或穿刺活检证实为良性或恶性淋巴结,于术前 1~2 周在本院行 CT 检查;②淋巴结短径 ≥ 5 mm;③CT 检查前未进行过放化疗等相关治疗;④图像清晰,图像质量满足诊断和分析要求。

2. 检查方法及图像选择

使用 GE Discovery CT750 HD 宝石能谱 CT 机,所有患者行头颈部 CT 平扫及双期增强扫描,扫描参数:300 mA,120 kV,层厚 3.0~5.0 mm,矩阵 512 $\times$ 512,扫描范围自眉弓上缘至颈根部水平。增强扫描使用高压注射器注射对比剂碘海醇(350 mg I/mL),注射流率 2.5~3.0 mL/s,总量 80~100 mL,注射对比剂后 25~30 s 行动脉期扫描、60~70 s 行静脉期扫描。平扫及增强扫描的层面需保持一致。

3. 临床资料的搜集和图像分析

由 2 位临床诊断经验丰富的影像科医师阅片分析所有患者的 CT 图像并提取主要临床资料,包括患者的年龄和性别、横轴面图像中靶淋巴结最大层面上淋巴结短径、平扫及动脉期 CT 值。将所有受试者随机分为 2 组,训练集 133 例(70 例良性+63 例恶性),测试集 67 例(35 例良性+32 例恶性)。

4. 影像组学特征的提取和筛选

由一位具有多年工作经验的影像科医师进行影像组学分析和测量。首先,选取每例患者靶淋巴结最大

层面的横轴面平扫及增强图像(平扫跟增强图像务必是同一层面),将其自 PACS 系统中以 DICOM 格式导入开源 MaZda 4.6 软件(www.eletel.p.lodz.pl/programy/mazda/);图像导入后先行 $\mu \pm 3\sigma$ (μ 为灰度平均值, σ 为灰度值标准差)进行灰度标准化处理,减少图像不同明暗度及对比度对灰度值的影响;然后,采用手动分割方法于靶淋巴结内勾画 ROI,ROI 边缘距病灶边缘 1~2 mm,避开钙化、周边血管等区域。在病灶勾画完成后,利用 Mazda 软件提取包括游程矩阵、灰度共生矩阵、灰度直方图、绝对梯度、小波变换和自回归模型在内的共 275 种纹理特征。

在训练集中使用选择算子和最小绝对收缩(selection operator and least absolute shrinkage, LASSO)算法进行降维,并通过 10 折交叉验证法筛选出 3~5 个与头颈部淋巴结良恶性预测最具相关性的影像组学特征。

5. 预测模型的建立、验证和评估

在训练集中应用 LASSO 方法并结合 10 倍交叉验证法去除冗余特征,获得最优组学特征,将其 LASSO 回归系数带入线性方程计算组学评分并建立影像组学标签;运用多因素 Logistic 回归分析各诊断因子(性别、年龄、病灶短径、平扫 CT 值、动脉期 CT 值、影像组学标签),分别建立非影像组学和影像组学诊断模型,采用受试者工作特征曲线(ROC)及曲线下面积(AUC)和校准曲线(calibration curves)评价各诊断模型在训练集和测试集中的诊断效能和拟合度。

最后将影像组学标签(score)联合独立诊断因子(Size、Z)制作成可量化赋分的诺模图预测工具:根据相应变量值在 Points 线段上赋分,计算出总得分,再垂直对应到 Low weight rate 线段上,以概率形式预测目标淋巴结的良、恶性。

6. 统计学方法

使用 SPSS24.0 和 R3.4.4 软件(<https://www.r-project.org/>)对所有数据进行统计学分析。本研究中计量资料均不符合正态分布,故采用 U 检验;分类变量采用 χ^2 检验。以 $P<0.05$ 为差异具有统计学意义。

结 果

1. 临床和常规 CT 资料

训练集和测试集中良、恶性组患者的临床和常规 CT 特征的比较见表 1。训练集和测试集中良性组与恶性组之间患者年龄、性别及病灶平扫 CT 值的差异均无统计学意义($P>0.05$)。动脉期强化 CT 值、淋巴结短径及影像组学标签在良性组与恶性组之间的差异具有统计学意义($P<0.05$)。

2. 特征选择、影像组学标签建立及诊断效能评价

在训练集中通过应用 LASSO 回归模型并结合 10 倍交叉验证法去除冗余特征,选择 AUC 最大处的 λ 值来筛选特征,最终从 275 个组学特征中筛选出 3 个最优组学特征(图 1),分别为 S(2,-2)Correlat、S(0,3)InvDfMom 和 S(4,0)Contrast。将这 3 个特征用于建立影像组学标签,表达式为 $y = 12.846714 \times S(2,-2)Correlat + (-22.142526) \times S(0,3)InvDfMom + 3.500306 \times S(4,0)Contrast + (-41.876957)$ 。在训练集和测试集中影像组学标签判别良、恶性淋巴结的 AUC 值分别为 0.884(95%CI: 0.830~0.937)和 0.749(95%CI: 0.634~0.864),诊断效能良好(图 2)。

3. 预测模型建立和诊断效能评价

将所有临床诊断因子和影像组学标签进行多因素 Logistic 回归分析,结果显示淋巴结短径(size)、动脉期增强 CT 值(Z)和影像组学标签(score)可作为判别淋巴结性质的独立诊断因子(表 2),并由此分别被纳入非影像组学模型(联合 size 和 Z)及影像组学联合诊断模型(score 联合 size 和 Z)。

采用 ROC 曲线对两种模型在训练集和测试集中的诊断效能进行评估,结果见表 3。综上影像组学诊断模型比非影像组学表现出更好的诊断效能($P=0.025, P=0.033$)(图 3)。影像组学预测模型在训练集和测试集中的校准曲线对比理想曲线显示出了良好的拟合效果(训练集 R^2 为 0.958,测试集 R^2 为 0.975),优于非影像组学诊断预测模型(训练集 R^2 为 0.930,测试集 R^2 为 0.918),详见图 4。

将影像组学标签(score)联合独立诊断因子依据 Logistic 回归系数权重制作可量化赋分的 nomogram 预测工具,计算出总得分后对应到 Low weight rate 线段上得出预测概率(图 5)。本研究结果以各项得分总和 68 分为临界点,超过此临界点预测为恶性,且总分

表 1 训练集和测试集中良、恶性组临床资料及常规 CT 征象的比较

变量	训练组				测试组			
	良性组	恶性组	Z/ χ^2 值	P 值	良性组	恶性组	Z/ χ^2 值	P 值
年龄/岁	51.51±14.03	55.10±16.10	-2.694	0.619	50.09±13.78	48.13±10.55	-1.244	0.711
性别比(男/女)	12/58	29/34	12.977	0.436	7/28	4/28	0.685	0.782
淋巴结短径/mm	7.27±1.86	13.28±6.01	-9.468	0.048	7.61±1.99	12.69±5.03	-6.561	0.034
平扫 CT 值/HU	27.60±9.02	43.38±11.76	-8.521	0.387	26.67±8.30	40.69±6.68	-5.916	0.098
动脉期 CT 值/HU	47.00±14.18	69.75±20.91	-8.958	0.036	43.36±11.75	65.44±15.23	-5.938	0.012
影像组学标签	-0.6677±0.7383	0.4531±0.5860	-7.625	<0.001	-0.4006±0.6338	0.3331±0.5097	-6.504	0.002

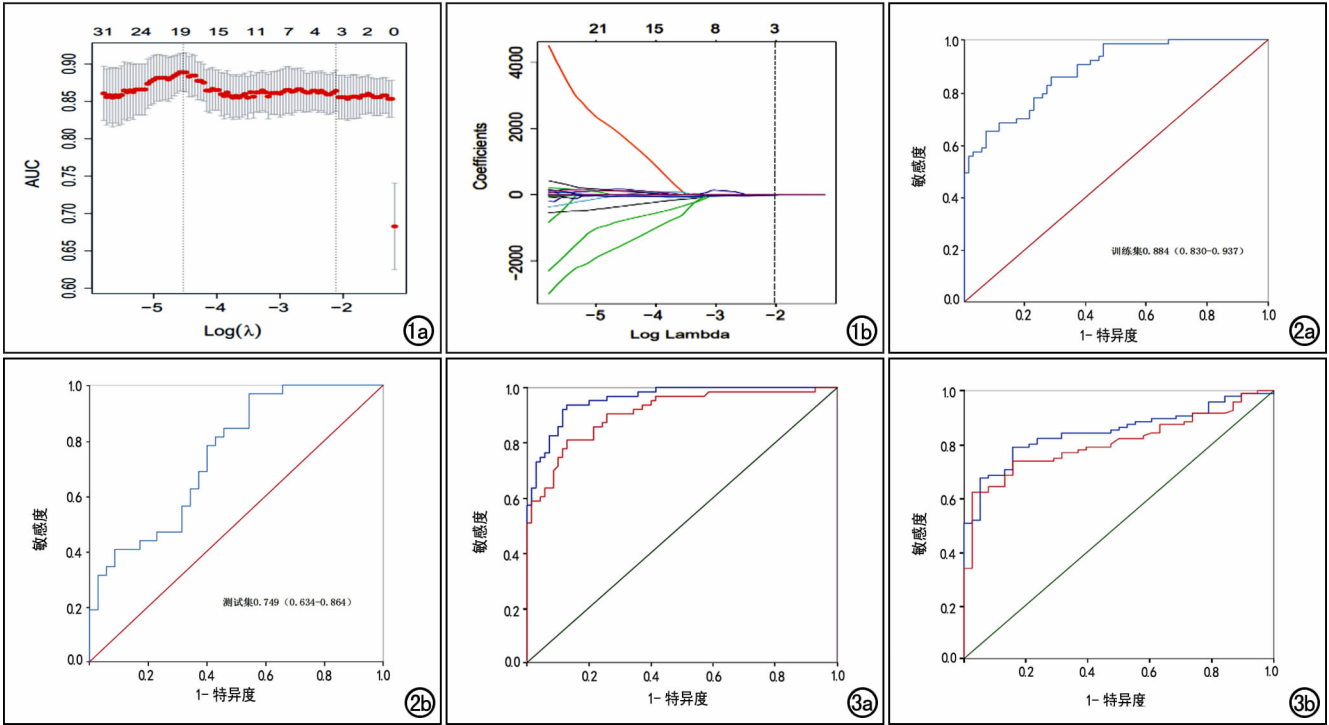


图 1 影像组学特征筛选和影像组学标签的建立。a) LASSO 模型结合 10 折交叉验证选择参数 λ, 虚线对应最优参数 λ 以及 AUC 和特征个数; b) 275 个影像组学特征在 LASSO 模型中的系数图, 虚线处对应最优参数 λ, 共筛选得到 3 个最优非零系数特征。图 2 影像组学标签的 ROC 曲线。a) 训练集, AUC=0.884; b) 测试集, AUC=0.749。图 3 预测模型的 ROC 曲线。a) 影像组学模型在训练集和测试集中的 AUC 分别为 0.985 和 0.908; b) 非影像组学模型在训练集合测试集中的 AUC 分别为 0.847 和 0.806。

值越大, 恶性概率越高。

讨论

CT 扫描在头颈部良、恶性淋巴结的术前和术后诊断中具有良好的诊断效能^[11]。目前淋巴结大小和坏死的证据仍然是最重要的影像学诊断标准, 但是淋巴结大小这一指标显示出很大的可变性, 而且容易受到观察者对于影像特征解读的主观因素的影响, 并不

能有效鉴别出较的淋巴结是否存在恶变。这就迫切需要一种新的技术手段, 近年来影像组学技术的应用不断深入, 它不仅可以规避观察者本身对于影像特征解读的主观偏向, 还能深度挖掘并整合图像中大量的人眼无法识别和区分的数字化信息, 从而能明显提高对头颈部转移性淋巴结的诊断水平, 减少漏判或误判率。目前而言, 影像组学应用最多的领域是肿瘤的辅助诊断, 尤其是对肿瘤良、恶性的辅助判别; 其次是对肿瘤

表 2 多因素 Logistic 回归分析结果

变量	OR(95%CI)	偏回归系数	标准误	P 值	Wald Z 值
年龄	1.0359(0.8283~6.3638)	0.0378	0.0236	0.1101	1.60
性别	3.3576(0.6952~16.2150)	1.2112	0.8034	0.1317	1.51
淋巴结短径	5.5647(1.2886~24.0300)	0.3433	0.1493	0.0215	2.30
平扫 CT 值	2.8726(0.6039~13.6640)	0.0555	0.0419	0.1848	1.33
动脉期 CT 值	6.8263(1.1644~40.0200)	0.0600	0.0282	0.0333	2.13
影像组学标签	31.6900(6.0882~104.9500)	2.3805	0.5797	0.0000	4.11

表 3 两种诊断模型的诊断效能

指标	影像组学联合诊断模型		非影像组学诊断模型	
	训练集	测试集	训练集	测试集
AUC	0.958(0.930~0.987)	0.908(0.859~0.958)	0.847(0.782~0.912)	0.806(0.734~0.879)
敏感度	0.918	0.818	0.794	0.728
特异度	0.876	0.857	0.832	0.832
符合率	0.924	0.883	0.824	0.788

注: 括号内数据为 95%CI。

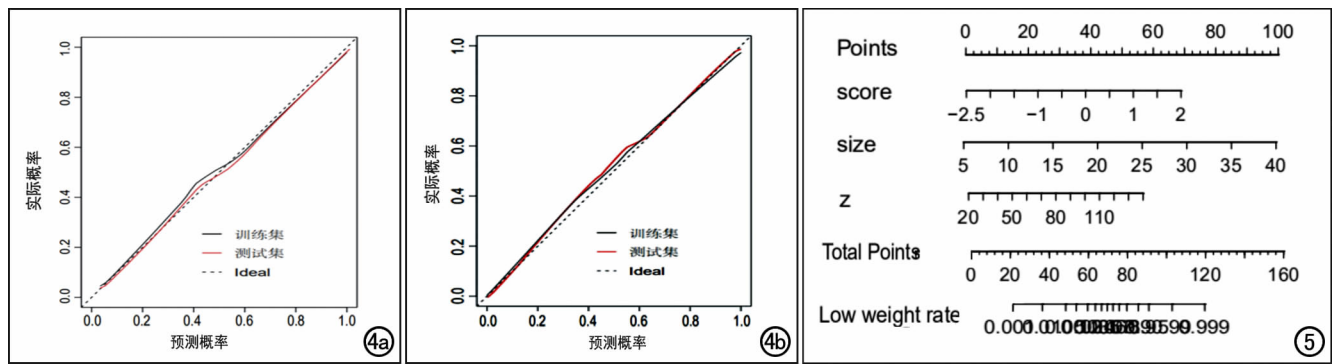


图 4 诊断模型在训练集和测试集的校准曲线(校准度曲线中 X 轴代表预测模型值, Y 轴表示真实值, 曲线越接近对角线, 说明吻合度越好)。a) 影像组学诊断模型; b) 非影像组学诊断模型。图 5 CT 影像组学预测模型的诺模图。将所有变量得分相加计算总得分, 再垂直对应到 Low weight rate 线段, 得出预测概率。

病灶的发展趋势进行预测。

目前将影像组学应用于淋巴结的相关研究大多是围绕纵隔、腋窝及腹盆腔淋巴结, 利用多参数 MRI、PET-CT 及超声成像, 对原发肿瘤的侵袭性、疗效和预后等方面进行探讨, 少有基于 CT 影像组学对头颈部淋巴结良恶性进行鉴别诊断的相关研究。如施奕倩等^[12]的研究中发现基于 CT 的影像组学对头颈部鳞状细胞癌患者总生存期的预测具有一定价值, 有潜力成为预测头颈部鳞状细胞癌患者预后的潜在生物学标志物。Chen 等^[13]的研究中纳入了 115 例直肠癌患者, 利用直肠腔内超声、剪切波弹性成像和 CT 图像, 提取直肠肿瘤、淋巴结和周围组织的组学特征, 并对 3 种方法分别建立影像组学标签, 使用多元 Logistic 回归分析与临床数据进行进一步验证, 其结果显示在训练组和验证组中, 淋巴结转移患者的所有 3 个影像组学标签 AUC 值均显著高于淋巴结转移阴性患者 ($P < 0.05$)。单嫣娜等^[14]利用动态增强 MRI 影像组学技术研究乳腺癌腋窝淋巴结转移, 经过降维成功筛选出 5 个最佳影像组学特征, 它们独立预测腋窝淋巴结转移的 AUC 为 0.747~0.931。王伟镇等^[15]运用影像组学方法预测甲状腺乳头状癌颈部中央区淋巴结转移, 结果超声横、纵双切面纹理特征预测的准确性较单切面高, 且远高于常规超声检查的准确率。Wu 等^[16]使用磁共振 T₂WI 图像筛选出 9 种影像特征用于预测淋巴结转移, 其预测效能 AUC 值达到 0.887, 并与临床诊断因子结合组成混合模型的诊断效能 AUC 值为 0.890。以上各研究成果均表明了影像组学方法在淋巴结转移的诊断和预测中取得了令人满意的效果。

然而, 大多数研究所采用的影像工具: MRI 为多序列多参数成像, 耗时相对较长, 在不同参数和序列图像中筛选影像组学特征及最优模型的过程较为复杂, 不利于常规应用和推广; PET-CT 费用高昂, 对于小病灶显示不佳, 而且注射用药物对身体有一定影响。而

基于 CT 的影像组学方法, 仅基于 CT 值数据提取纹理特征, 相对而言快速、简单和实用。如孔丹等^[17]利用影像组学软件于患者术前平扫及增强 CT 图像提取出 378 个特征, 经过筛选得到 16 个最优影像组学特征用于建立甲状腺良恶性结节的鉴别诊断模型, 其在训练组和测试组中的 ROC 曲线下面积分别为 0.92 和 0.90, 诊断符合率为 88.2%。沙雪等^[18]基于平扫、动脉期和静脉期 CT 图像, 建立鉴别非小细胞肺癌纵隔淋巴结转移的影像组学模型, 并探讨不同模型的诊断效能, 他们共提取了 841 个影像组学特征, 使用 LASSO 算法筛选降维后, 基于各时相 CT 影像组学特征和两个时相 CT 影像组学特征的差值来建立诊断模型, 结果显示基于 CT 各时相图像建立的影像组学模型均可辅助临床来判断淋巴结的性质, 其中以基于平扫 CT 图像的影像组学模型的 AUC 最高, 结合 CT 平扫及动脉期增强图像建立的影像组学模型具有更高的鉴别敏感度和阴性预测值。张力等^[19]将影像组学工具应用于头颈部木村病淋巴结病变和淋巴瘤的鉴别研究中, 建立的 CT 影像组学模型具有良好的诊断效能。以上这些研究大多只给出影像组学特征单独的诊断效能, 未利用影像组学特征建立可供量化分析的综合预测模型, 也未在不同数据集中对其进行验证。

本研究团队使用影像组学方法建立并验证了一个结合 CT 强化值、淋巴结短径和影像组学标签的诺模图诊断预测模型, 并量化输出预测值。通过回归分析, 显示淋巴结短径(size)和增强动脉期 CT 值(Z)对判别淋巴结的良恶性具有统计学意义 ($P < 0.05$)。另外, 采用影像组学方法最终筛选出 3 个最具鉴别意义的组学特征, 建立影像组学标签, 根据组学特征的 LASSO 系数, 以逆差矩(InvDfMom)的贡献最大, 该特征参数反映局部灰度均匀性, 其值越高表明该区域在灰度值之间缺乏变化, 这一结果可能提示肿瘤的异质性跟图像灰度的均匀性有关联。而且 Yip 等^[20]的研究结果

显示组学特征中以 InvDfMom(逆差矩)与非小细胞肺癌体细胞突变最具相关性,同样验证了上述结果。在此基础上,我们进一步采用 ROC 曲线对影像组学标签(score)的诊断效能进行了分析,在训练集和测试集中其 AUC 均大于 0.80,随后将 score 与动脉期强化 CT 值(Z)、淋巴结短径(size)联合建立影像组学综合诊断模型,该模型在两个数据集中的诊断效能同样表现优秀。

本研究存在以下不足之处:(1)部分患者注射造影剂时间快慢不等,可能对病灶动脉期强化 CT 值有些许影像。(2)所有 ROI 都是在横轴面 CT 图像上手动勾画,没有获得病灶容积 ROI 来提取病灶特征,因此所获的影像信息可能不全面。已有相关研究证实对肿瘤病灶采用三维纹理特征提取所构建的诊断模型更加准确^[21]。(3)没有将多期或动态增强扫描图像纳入本研究,它们或许较单纯动脉期扫描能提供更多影像学信息,这也是本研究的一大缺憾。但本次研究已积累部分经验,有望在以后的研究中进一步完善。

参考文献:

- [1] Siegel RL, Miller KD, Jemal A, et al. Cancer statistics, 2020[J]. CA Cancer J Clin, 2020, 70(1): 7-30. DOI: 10.3322/caac.21590.
- [2] 满育平, 马隆佰, 吴春梅, 等. 3.0T 磁共振 ADC 值及 DCE-MRI 定量分析对颈部良恶性淋巴结鉴别诊断[J]. 放射学实践, 2019, 34(6): 619-623. DOI: 10.13609/j.cnki.1000-0313.2019.06.005.
- [3] Peltenburg B, de Keizer B, Dankbaar JW, et al. Prediction of ultrasound guided fine needle aspiration cytology results by FDG PET-CT for lymph node metastases in head and neck squamous cell carcinoma patients[J]. Acta Oncol, 2018, 57(12): 1687-1692. DOI: 10.1080/0284186X.2018.1529426.
- [4] Osarogiagbon RU, Ramirez RA, Wang CG, et al. Size and histologic characteristics of lymph node material retrieved from tissue discarded after routine pathologic examination of lung cancer resection specimens[J]. Ann Diagnost Pathol, 2014, 18(3): 136-139. DOI: 10.1016/j.anndiagpath.2014.02.004.
- [5] 周镇源, 刘志锋, 蔡金辉, 等. 表观扩散系数鉴别颈部淋巴结病变诊断价值[J]. 实用放射学杂志, 2019, 35(8): 1220-1224. DOI: 10.3969/j.issn.1002-1671.2019.07.005.
- [6] Philips BWJ, Stijns RCH, Rietsch SHG, et al. USPIO-enhanced MRI of pelvic lymph nodes at 7T: preliminary experience[J]. Eur Radiol, 2019, 29(12): 6529-6538. DOI: 10.1007/s00330-019-06277-7.
- [7] 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. DOI: 10.1016/j.ejca.2011.11.036.
- [8] Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer[J]. J Clin Oncol, 2016, 34(18): 2157-2164. DOI: 10.1200/JCO.2015.65.9128.
- [9] Hu J, Zhao YJ, Li MC, et al. Machine-learning-based computed tomography radiomic analysis for histologic subtype classification of thymic epithelial tumours[J/OL]. Eur J Radiol, 2020, 126(5): e108929. DOI: 10.1016/j.ejrad.2020.108929.
- [10] Bulens P, Couwenberg A, Intven M, et al. Predicting the tumor response to chemoradiotherapy for rectal cancer: model development and external validation using MRI radiomics[J]. Radiother Oncol, 2020, 142(1): 246-252. DOI: 10.1016/j.radonc.2019.07.033.
- [11] Cho SJ, Suh CH, Baek JH, et al. Diagnostic performance of CT in detection of metastatic cervical lymph nodes in patients with thyroid cancer: a systematic review and meta-analysis[J]. Eur Radiol, 2019, 29(9): 4635-4647. DOI: 10.1007/s00330-019-06036-8.
- [12] 施奕倩, 袁瑛, 任继亮, 等. 基于 CT 的影像组学在头颈部鳞状细胞癌预后中的应用[J]. 实用放射学杂志, 2019, 35(12): 1910-1913. DOI: 10.3969/j.issn.1002-1671.2019.12.006.
- [13] Chen LD, Liang JY, Wu H, et al. Multi-parametric radiomics improve prediction of lymph node metastasis of rectal cancer compared with conventional radiomics[J]. Life Sci, 2018, 208(9): 55-63. DOI: 10.1016/j.lfs.2018.07.007.
- [14] 单嫣娜, 龚向阳, 丁忠祥, 等. 动态增强 MRI 影像组学特征预测乳腺癌腋窝淋巴结转移的价值[J]. 中华放射学杂志, 2019, 53(9): 742-747. DOI: 10.3760/cma.j.issn.1005-1201.2019.09.006.
- [15] 王伟镇, 李颖嘉. 影像组学预测甲状腺乳头状癌中央区淋巴结转移[J]. 南方医科大学学报, 2019, 39(9): 1094-1098. DOI: 10.12122/j.issn.1673-4254.2019.09.15.
- [16] Wu S, Zheng J, Li Y, et al. Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer[J]. E Bio Med, 2018, 34(8): 76-84. DOI: 10.1016/j.ebiom.2018.07.029.
- [17] 孔丹, 张建东, 单文莉, 等. CT 影像组学模型对甲状腺良恶性结节的鉴别价值[J]. 中华放射学杂志, 2020, 54(3): 187-191. DOI: 10.3760/cma.j.issn.1005-1201.2020.03.003.
- [18] 沙雪, 巩贯忠, 仇清涛, 等. 基于 CT 影像组学鉴别非小细胞肺癌纵隔转移性淋巴结的模型研究[J]. 中华放射医学与防护杂志, 2020, 40(2): 150-155. DOI: 10.3760/cma.j.issn.0254-5098.2020.02.014.
- [19] 张力, 于淑靖, 张迎, 等. CT 影像组学在头颈部木村病淋巴结病变与淋巴瘤鉴别中的应用[J]. 放射学实践, 2020, 35(2): 159-164. DOI: 10.13609/j.cnki.1000-0313.2020.02.007.
- [20] Yip SS, Kim J, Coroller TP, et al. Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer[J]. J Nucl Med, 2017, 58(4): 569-576. DOI: 10.2967/jnumed.116.181826.
- [21] 赵海波, 覃丽兰, 铁超恩, 等. 二维和三维 MR 纹理特征提取在脑肿瘤影像组学模型构建中的对比研究[J]. 实用放射学杂志, 2019, 35(8): 1328-1331. DOI: 10.3969/j.issn.1002-1671.2019.08.031.

(收稿日期: 2020-11-27 修回日期: 2021-02-27)

· 头颈部影像学专题 ·

基于 CT 增强影像组学模型预测甲状腺乳头状癌颈部淋巴结转移

刘妮, 谢元亮, 黄增发, 王翔

【摘要】 目的:探讨基于 CT 增强图像的组学特征预测甲状腺乳头状癌(PTC)颈部淋巴结转移的可行性。**方法:**回顾性分析 130 个经病理证实的 PTC 结节(伴颈部淋巴结转移 65 个,无淋巴结转移 65 个)的 CT 增强扫描图像,应用 Mazda 纹理分析软件对 CT 图像进行预处理,并提取 PTC 结节的一阶和高阶纹理特征。应用 Fisher 相关系数、分类误差概率与平均相关系数(POE+ACC)和交互信息(MI)三种降维方法分别对纹理特征进行筛选,每种方法筛选出 10 个最佳纹理特征。应用 Mazda 软件的 B11 模块中的原始数据分析(RDA)、主成分分析(PCA)、线性判别分析(LDA)和非线性判别分析(NDA)四种统计学方法分别进行数据的进一步降维,计算不同统计学方法、不同降维方法组合下建立的影像组学模型的错判率、敏感度和特异度。**结果:**最佳纹理参数主要来源于二阶及高阶纹理特征(灰度游程矩阵、灰度共生矩阵和小波转换)。不同降维方法组合同一统计学方法时所获模型的错判率较为接近,范围为 3.08%~16.92%。三种降维方法分别与非线性判别分析(NDA)组合时,所获模型的诊断结果一致,与另外 3 种统计学方法相比敏感度和特异度最高,分别为 98.46%、95.38%,错判率最低(3.08%)且差异均有统计学意义($P<0.05$)。**结论:**基于 CT 增强图像的 PTC 结节的二阶及高阶纹理特征对预测颈部淋巴结转移是可行的,不同的降维方法和统计学方法组合获得的组学模型诊断效能不同,推荐使用非线性判别分析与任一降维方法组合的组学模型。

【关键词】 纹理分析; 甲状腺肿瘤; 体层摄影术, X 线计算机; 对比增强扫描

【中图分类号】 R814.42; R736.1 **【文献标志码】** A **【文章编号】** 1000-0313(2021)08-0971-05

DOI:10.13609/j.cnki.1000-0313.2021.08.005

开放科学(资源服务)标识码(OSID):



Preliminary study on radiomics based on CT contrast-enhanced images in predicting cervical lymph node metastasis in patients with papillary thyroid carcinoma LIU Ni, XIE Yuan-liang, HUANG Zeng-fa, et al. Department of Radiology, the Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430014, China

【Abstract】 Objective: To explore the feasibility of radiomics features based on contrast-enhanced CT images in predicting the cervical lymph node metastasis in patients with papillary thyroid carcinoma (PTC). **Methods:** The contrast-enhanced CT images of 130 patients with PTC (65 with CLNM and 65 without CLNM) confirmed by pathology were analyzed retrospectively. Texture analysis software (Mazda) was used to extract the texture features of the thyroid nodules in the two groups. Then the extracted texture parameters were dimensionally reduced, and the 10 optimal texture features were selected by Fisher coefficient, classification error probability with average correlation coefficients (POE+ACC) and mutual information coefficient (MI) in Mazda software. Then four statistic methods were used to further reduce the dimensionality respectively by using B11 software in Mazda software. The misdiagnosis rates of radiomics models from different combinations of statistic methods with dimensionality reduction methods were calculated. **Results:** The ten optimal texture features were all second-order or high-order texture features (from gray-level run matrix, gray-level co-occurrence matrix or wavelet transform). The misdiagnosis rates selected out by three dimensionality reduction methods with the same statistic method were closed to each other, ranged from 3.08% to 16.92%; NDA showed the lowest misdiagnosis rates (3.08%) and the highest sensitivity and specificity (98.46% and 95.38%, respectively) with statistically significant difference ($P<0.05$) when com-

作者单位: 430014 武汉, 华中科技大学同济医学院附属武汉市中心医院

作者简介: 刘妮(1988—), 女, 湖北天门人, 硕士, 主治医师, 主要从事头颈部疾病的影像学诊断和研究工作。

通讯作者: 王翔, E-mail: wangxiang385@aliyun.com

pared with the other three statistic methods. **Conclusion:** It is feasible for texture analysis in predicting the CLNM in PTC patients with second-order or high-order texture features. The different combinations of statistic methods and dimensionality reduction methods have different efficacy. NAD combined with either dimensionality reduction method is recommended, which has the lowest misdiagnosis rate and the high-highest sensitivity and specificity.

【Key words】 Texture analysis; Thyroid tumor; Tomography, X-ray computed; Contrast enhanced scan

甲状腺恶性结节的发病率呈逐年上升趋势,手术切除为主要治疗方式^[1],多数研究者认为颈部淋巴结转移(cervical lymph node metastasis, CLNM)是甲状腺乳头状癌(papillary thyroid carcinoma, PTC)复发及远处转移的高危因素^[2]。因此,术前准确检出 CLNM 对选择治疗方案、评估预后具有重要价值。超声是甲状腺术前检查的常规方法,但检出 CLNM 的敏感性较低^[3]。影像学组学作为一项新兴技术,可以高通量地从影像数据中提取大量肉眼无法识别的特征信息,对影像数据进行更深层次的挖掘、预测和分析^[4]。目前应用影像组学预测甲状腺癌伴有 CLNM 的相关研究较少,因此本研究旨在探讨 CT 组学特征对 PTC 伴有颈部淋巴结转移的预测价值。

材料与方法

1. 研究对象及分组

搜集 2019 年 6 月—2019 年 12 月在本院经手术病理证实(无论超声或 CT 提示有无淋巴结转移,均行可疑区域颈部淋巴结清扫术,术后获取病理检查结果)且行 CT 增强扫描的甲状腺结节患者的病例资料。纳入标准:(1)为避免因设备不同造成图像参数和质量等的差异,仅纳入使用 Siemens 双源 CT 机进行检查的患者;(2)为避免甲状腺多个或弥漫性结节影响对责任病灶的确定,仅纳入甲状腺内仅单个结节灶的患者;(3)为避免不同病理学类型的生物学及影像学差异,仅纳入病理证实为 PTC 的患者。排除标准:(1)结节长径小于 1 cm;(2)CT 图像伪影较大,影响对甲状腺结节的形态观察和范围测量;(3)病灶内有大面积钙化或囊变(钙化及囊变面积占病灶面积的 1/2 或以上)。最终纳入 130 例 PTC 患者,女 92 例,男 38 例;年龄 27~72 岁,平均(60±13)岁;结节直径 1.00~4.23 cm,平均(2.12±0.61)cm;CLNM 组 65 例,无淋巴结转移(nCLNM)组 65 例。

2. CT 检查方法

使用 Siemens Somatom Definition Flash CT 机,扫描范围自颅底至胸廓入口水平,所有患者行静脉期增强扫描,经肘静脉注射非离子型对比剂碘海醇(0.3 mg I/mL),剂量 1 mL/kg,注射流率 3.0 mL/s,

随后以相同流率注射 30 mL 生理盐水,延迟时间 50 s,其它扫描参数:100 keV,自动 mAs,二代迭代重建技术,螺距 0.984,视野 30 cm × 30 cm,矩阵 512 × 512,层厚和层间距均为 5.00 mm,重建层厚 1.25 mm。

3. 纹理分析方法

将 CT 增强图像(DICOM 格式)导入 Mazda 4.6 软件(波兰 Institute of Electronics 公司)进行影像组学分析。首先,由两位高年资主治医师在肿瘤最大层面手动勾画 ROI,意见不一致时请一位副主任医师阅片,最终协商确定。自每个 ROI 提取 6 大类近 300 个纹理参数,详见表 1。

表 1 Mazda 软件显示的纹理特征参数来源及分类

纹理特征类别	数目	纹理特征
直方图	9	均值、峰度、百分位数(1%、10%、50%、90%、99%)、方差、偏度
灰度共生矩阵	220	角二阶矩、对比度、相关系数、平方和、均值和、熵和、熵、均差、逆差距、熵差、变异系数(4 个分布方向共 20 对像素点,每对像素点均包含以上 11 个纹理特征)
灰度游程矩阵	20	短游程增强、长游程增强、分数游程、灰度不均匀性度量、游程长度不均匀性度量(这 5 类特征均包含 4 个分布方向:垂直、水平、45°和 135°)
梯度模型	5	梯度均值、方差、峰度、偏度、非零梯度百分比
自回归模型	5	01~04、σ
小波转换	16	小波转换能量系数包括低频部分分解系数、垂直方向分解系数、水平方向分解系数和对角线方向分解系数共 4 类,每类又包含 4 个子集

其中,直方图纹理特征为一阶纹理参数,灰度共生矩阵(grey-level co-occurrence matrix, GLCOM)、灰度游程矩阵(grey-level run length matrix, GLRLM)、梯度模型、自回归模型和小波转换(wavelet transform)类的纹理特征属于二阶及高阶纹理参数。然后,使用 Mazda 软件提供的 3 种降维方法,即 Fisher 相关系数、分类误差概率和平均相关系数(classification error probability and average correlation coefficients, POE+ACC)及交互信息(mutual information, MI),分别筛选出能鉴别两组的 10 个最佳纹理特征。随后,

分别使用 Mazda 软件中的 B11 模块中提供的 4 种统计方法,包括原始数据分析(raw data analysis,RDA)、主成分分析(principal component analysis,PCA)、线性判别分析(linear discriminant analysis,LDA)和非线性判别分析(nonlinear discriminant analysis,NDA)进一步对纹理特征进行分类和筛选。将不同统计学方法与不同降维方法组合,分别建立相应的影像组学模型,采用错判率这一指标来描述各个模型的诊断效能,错判率越小则说明该方法提取的纹理特征在鉴别 PTC 有淋巴结转移中的准确性越高、价值越大。

4. 统计学分析

根据 Mazda 软件的输出结果,使用 SPSS 25.0 统计软件进行统计学分析,采用 Fisher 确切概率法(四格表中至少有一个格子的理论频数<1)对同一种降维方法与不同统计学方法(NDA、RDA、PCA 和 LDA)组合建立的组学模型、同一种统计学方法与不同降维方法组合建立的模型对 CLNM 评估所得的错判率均进行统计学检验。以 $P<0.05$ 为差异有统计学意义。

结 果

1. 最佳纹理特征的筛选

每种降维方法筛选出 10 个最佳纹理特征,共获得 30 个最佳纹理特征,其中 21 个特征属于 GLRLM 类,5 种属于小波转换类,3 种属于 GLCOM 类,1 种属于直方图参数。

在 3 种降维方法中均被筛选出的纹理特征有游程长度不均匀性 45°方向(45 degree run length non uniformity,45dgr_RLN)、游程长度不均匀性垂直方向(vertical RLN,Vertl_RLN)、垂直方向(Vertl_)灰度不均匀性(grey-level non uniformity,GN);在 3 种降维方法中被筛选出 2 次的纹理参数有水平方向(Horizonl,Horzl)_RLN、135dgr_RLN、Horzl_GN、135dgr_GN、45dgr_GN、小波转换系数 s-2 (Wavelet energy

with low pass and high pass frequency bands with scale factors 2,WavEnLH_s-2),前 8 种属于 GLRLM 类参数,最后一项属于小波转换类参数。

2. 组学模型的判别结果

不同统计学方法与不同降维方法组合下的组学模型(如图 1,以 MI 降维算法分别与 4 种统计学方法组合)对 PTC 合并 GLNM 的判别结果见表 2。分别与 4 种统计学方法组合,Fisher 系数法组学模型的错判率为 3.08%~16.92%,POE+ACC 法为 3.08%~15.38%,MI 法为 3.08%~14.68%。四种统计学方法(RDA、PCA、LDA、NDA 组合不同降维方法建立的组学模型中,以 NDA 法的判别结果最稳定,组合 3 种降维方法的 3 个组学模型之间的判别结果基本一致,且错判率最低(3.08%),敏感度和特异度最高(分别为 98.46%和 95.38%)。

表 2 不同特征分类法的预测结果

特征分类法	错判率	敏感度	特异度
Fisher/RDA	12.30%	87.69%	87.69%
Fisher/PCA	11.54%	89.23%	87.69%
Fisher/LDA	16.92%	81.54%	84.62%
Fisher/NDA	3.08%	98.46%	95.38%
POE+ACC/RDA	10.00%	89.23%	90.77%
POE+ACC/PCA	9.23%	90.77%	90.77%
POE+ACC/LDA	15.30%	83.77%	86.15%
POE+ACC/NDA	3.08%	98.46%	95.38%
MI/RDA	13.08%	86.15%	87.69%
MI/PCA	12.30%	89.23%	86.15%
MI/LDA	14.62%	84.62%	86.15%
MI/NDA	3.08%	98.46%	95.38%

3. 错判率的比较

同一种降维方法分别与 4 种统计学方法组合,所建立的 4 个组学模型之间错判率的差异有统计学意义($P<0.05$)。而每种统计学方法分别与 3 种降维方法进行组合,所建立的 3 个组学模型之间错判率的差异无统计学意义,RDA、PCA 和 LDA 组对应的 P 值分别为 0.73、0.71 和 0.87,而 NDA 组合的 3 种降维方法建立的组学模型之间错判率完全相同,均为 3.08%。

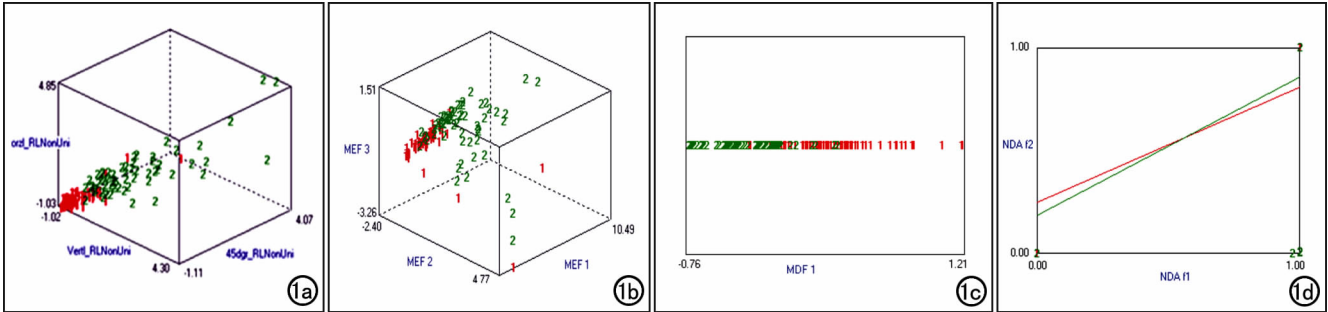


图 1 Mazda 软件中 B11 模块提供的 4 种统计学方法与 MI 降维算法组合建立的组学模型的分析结果。数字 1(红色)代表甲状腺乳头状癌伴颈部淋巴结转移,数字 2(绿色)代表甲状腺乳头状癌不伴淋巴结转移。两种颜色的数码字分的越开、重叠越少,提示该方法的鉴别能力越高、误判率越低。a)RDA;b)PCA;c)LDA;d)NDA。

讨 论

颈部淋巴结是 PTC 发生淋巴结转移的第一站,又称为前哨淋巴结。超声检查对颈部淋巴结的显示不佳,而仅通过颈部淋巴结的 CT 形态学特征来诊断是否发生转移有较大的局限性。在术前准确诊断 PTC 有无颈部淋巴结转移是临床上一直以来的热点和难点。有研究结果显示,PTC 有无淋巴结转移与原发灶的 CT 形态学特征及病理亚型有关^[5-6]。恶性肿瘤的生物学异质性与影像纹理特征的异质性有一定相关性,基于 CT 或 MRI 提取的纹理特征可用于评估肿瘤的性质^[7]。基于上述研究背景,本研究中应用纹理分析软件对 PTC 原发灶的 CT 增强图像进行纹理分析,提取 CT 图像上大量肉眼无法识别的病灶的特征信息,旨在为临床术前诊断 PTC 有无颈部淋巴结转移提供参考依据。据文献报道,当 ROI 的直径小于 1.0 cm 时,影像组学分析软件的准确性会下降^[8]。因此本研究中排除了结节长径小于 1.0 cm 的病例。

本研究结果显示,3 种纹理特征降维方法的错判率均较低,检出敏感度最高可达 98%,有文献报道,超声新技术(如超声造影、弹性成像、超声引导下细针穿刺抽吸细胞学检查等)在鉴别甲状腺癌颈部淋巴结转移的符合率可达 92%~97%^[9]。黄益龙等^[10]报道双能 CT 碘图评估甲状腺微小乳头状癌颈部淋巴结的符合率约 75.8%。王娴等^[11]报道多 b 值 DWI 鉴别甲状腺癌甲状腺外侵犯的阳性预测值约 77.6%。与以上几种方法的结果相比,本研究中采用的影像组学方法的诊断价值并不逊色,提示影像组学在鉴别 PTC 颈部淋巴结转移中有一定临床应用价值。

本研究中利用 3 种降维方法分别筛选出 10 个最佳纹理参数,45dgr_RLN、Vertl_RLN、Horzl_RLN、135dgr_RLN、Vertl_GN、Horzl_GN、135dgr_GN、45dgr_GN 和 WavEnLH_s-2 这几个纹理特征分别出现在 2 种或 3 种方法的筛选结果中,除最后一个参数属于小波转换类,前 8 个参数均属于灰度游程矩阵类。小波转换和灰度游程矩阵类的纹理参数均属于二阶及高阶纹理特征。在 30 个纹理参数中,仅有 1 个参数方差(variance)来源于直方图(一阶纹理参数),表明基于 PTC 结节 CT 图像的二阶及高阶纹理特征在预测颈部淋巴结转移中有较大价值,而一阶纹理参数的作用有限。Kim 等^[12]的研究结果显示,基于 PTC 原发灶提取的直方图参数(平均值、标准差、偏度、峰度和熵)对预测淋巴结转移无效。本研究结果进一步验证了上述研究结果。谢文君等^[13]认为 PTC 病理组织学亚型与颈淋巴结转移具有紧密的关系,可能的原因是不同病理组织学亚型,肿瘤细胞的致密性不一,或者伴有

CLNM 的 PTC 病灶的代谢较活跃,原发瘤灶易发生坏死致肿块密度不均,而灰度游程矩阵可以反映图像纹理的粗细程度和不均匀程度^[14],所以在本研究中灰度游程矩阵类别的纹理参数能在一定程度上预测颈部淋巴结转移,而一阶纹理特征不能很好地反映 PTC 原发灶的空间异质性改变。

目前大多数研究中关于纹理特征的提取是依赖于纹理分析软件,而对最佳纹理特征进行筛选分析常用的是传统的、单一的统计学分析方法^[15],这会导致模型的单一性,在一定程度上影响纹理分析的准确性。Mazda 软件自带 4 种模型的统计分析方法:线性分类(RDA、PCA、LDA)和非线性分类(NDA),可以对上述三种降维方法得到的纹理参数进行分析和评判。本研究结果显示,NDA 与 3 种降维方法组合的组学模型的分析结果稳定,且在各组中错判率最低,为 3.08%。表明非线性分类方法较线性分类方法更有效,错判率最低,这可能与 NDA 算法更适用于 CT 图像有关。因为 CT 图像数据通常是高维度、非线性数据,而 NDA 的特点是通过选择使用合适的支持向量机,把非线性、不可分离数据转化成三维空间分布的、可分离的线性数据^[16]。而 RDA、PCA、LDA 特征分类法是线性变换,在将非线性数据转换为线性数据方面具有局限性。

本研究中的主要局限性在于仅阐述了 CT 组学特征在预测甲状腺癌颈部淋巴结转移中有一定的预测价值,主要体现在二阶及高阶纹理参数中,但是没有进一步探讨具体参数的数值、阈值及曲线下面积,离广泛应用于临床还有一定的距离。影像组学在甲状腺病变的研究中,大多数处于对结节良、恶性的判断,对影像组学在预测甲状腺癌颈部淋巴结转移中作用的文献报道较少,且大多为超声资料,目前笔者尚未查阅到探讨 Mazda 软件在相关领域的研究报道,所以笔者先探讨 Mazda 软件在预测甲状腺癌淋巴结转移中是否有可行性,在可行的基础上,下一步将分析相关参数的数值、计算阈值及曲线下面积,并传统影像学方法进行比较,这些工作将在今后研究中逐步开展。

参考文献:

- [1] 林少帆,林黛英,吴先衡,等.逐步判别分析甲状腺良、恶性结节的 CT 鉴别诊断因素[J].放射学实践,2020,35(4):473-477.
- [2] 朱乔丹,欧笛,杨菁,等.甲状腺乳头状癌颈部淋巴结转移风险评估的研究现状[J/OL].中华医学超声杂志(电子版),2019,16(11):e880-e884. DOI:10.3877/cma.j.issn.1672-6448.2019.11.018.
- [3] 中华医学会放射学分会头颈学组.甲状腺结节影像检查流程专家共识[J].中华放射学杂志,2016,50(12):911-915.
- [4] 胡云婷,黄增发,谢元亮,等.一阶 CT 纹理分析在甲状腺良恶性结节鉴别中的价值[J].临床放射学杂志,2019,38(3):422-425.
- [5] Wei S, Lan X, Hao Z, et al. Risk factors for central lymph node

- metastasis in CN0 papillary thyroid carcinoma: a systematic review and meta-analysis[J/OL]. Plos One, 2015, 10(10): e139021. DOI: 10.1371/journal.pone.0139021
- [6] 葛明华, 徐栋, 杨安奎, 等. 甲状腺良性结节、微小癌及颈部转移性淋巴结热消融治疗专家共识(2018 版)[J]. 中国肿瘤, 2018, 27(10): 768-773.
- [7] 谢元亮, 杜丹, 谢伟, 等. DCE-MRI 纹理分析鉴别宫颈鳞癌与腺癌及预测分级的价值[J]. 放射学实践, 2019, 34(8): 835-840.
- [8] 郭炜, 罗德红, 赵燕凤, 等. CT 纹理分技术鉴别甲状腺良恶性结节可行性研究[J]. 国际医学放射学杂志, 2017, 40(1): 3-5, 36.
- [9] 周倩, 许萍. 超声评估甲状腺癌颈部转移淋巴结的研究进展[J]. 中国医学影像技术, 2019, 35(11): 1752-1756.
- [10] 黄益龙, 赵雯, 李浚利, 等. 双能 CT 碘图评估甲状腺微小乳头状癌颈淋巴结转移潜能的价值[J]. 中华放射学杂志, 2019, 53(8): 655-660.
- [11] 王炯, 湛业荣, 蒋鹏程, 等. 多 b 值 MR 扩散加权成像对甲状腺乳头状癌甲状腺外侵犯的诊断价值[J]. 中华放射学杂志, 2018, 52(12): 897-902.
- [12] Kim SY, Lee E, Nam SJ, et al. Ultrasound texture analysis: association with lymph node metastasis of papillary thyroid microcarcinoma[J/OL]. Plos One, 2017, 12(4): e176103. DOI: 10.1371/journal.pone.0176103
- [13] 谢文君, 付天红, 朱莉艳. PTC 组织学亚型与颈淋巴结转移及近期预后的关系[J]. 中华全科医学, 2014, 12(5): 737-738.
- [14] 陈立鹏, 刘伟锋, 赵钊. 基于 CT 平扫图像的纹理分析鉴别肉芽肿性肺曲霉菌病与肺隐球菌病[J]. 中国医学计算机成像杂志, 2018, 24(5): 441-445.
- [15] 宋文龙, 郭大静, 陈倩羽, 等. 基于 MR 增强图像肝细胞癌直方图纹理分析[J]. 放射学实践, 2019, 34(12): 1348-1353.
- [16] 王笑, 所世腾, 朱炯, 等. 肠道 CT 图像纹理分析和非线性判别分析在结直肠癌及溃疡性结肠炎鉴别诊断中的应用[J]. 上海交通大学学报(医学版), 2018, 38(6): 624-631.

(收稿日期: 2020-12-06 修回日期: 2021-02-18)

欢迎订阅 2021 年《放射学实践》

《放射学实践》是由国家教育部主管, 华中科技大学同济医学院主办, 与德国合办的全国性影像学学术期刊, 创刊至今已 36 周年。遵照同行评议、价值导向、等效应用原则, 国内各大学、协会、组织机构通过科技工作者推荐、专家评议、结果公示等规定程序, 《放射学实践》杂志入选中国科协发布 10 大领域《我国高质量科技期刊分级目录》业内认可的较高水平期刊。《放射学实践》杂志再次入选 2020 年版北京大学和北京高校图书馆期刊工作研究会共同主持的国家社会科学基金项目“学术期刊评价及文献计量学研究”研究成果——《中国核心期刊要目总览》。

本刊坚持服务广大医学影像医务人员的办刊方向, 关注国内外影像医学的新进展、新动态, 全面介绍 X 线、CT、磁共振、介入放射及放射治疗、超声诊断、核医学、影像技术学等医学影像方面的新知识、新成果, 受到广大影像医师的普遍喜爱。

本刊为中国科学引文数据库(CSCD)核心库来源期刊、《中文核心期刊要目总览》收录期刊、中国科技核心期刊、中国高质量科技期刊分级目录临床医学领域医学影像学期刊, 并在中国学术期刊分区中位列 Q1 区。

主要栏目: 论著、继续教育园地、专家荐稿、研究生展版、图文讲座、本刊特稿、实验研究、传染病影像学、影像技术学、外刊摘要、学术动态、请您诊断、病例报道、知名产品介绍、信息窗等。

本刊为月刊, 每册 25 元, 全年定价 300 元。

国内统一刊号: ISSN 1000-0313/CN 42-1208/R 邮政代号: 38-122

电话: (027) 69378385

E-mail: fsxsjzz@163.com 网址: <http://www.fsxsj.net>

编辑部地址: 430199 武汉市蔡甸区中法新城同济医院专家社区别墅 C 栋

• 头颈部影像学专题 •

基于 CT 双期增强图像的深度学习迁移学习模型对甲状腺良恶性结节的分类研究

孟名柱, 潘昌杰, 张铭, 何光远, 沈栋, 陈罕奇

【摘要】 目的:评价基于 CT 双期增强图像的不同深度迁移学习(DTL)模型对甲状腺良恶性结节的分类效能。**方法:**采用相同程序架构和相同数据集对 3 种 DTL 模型(VGG19、ResNet50 和 DenseNet201)的分类诊断效能进行测试和评估。以不同模型在训练集和测试集中的最高预测符合率和在验证集中的符合率、召回率、F1 评分和受试者工作特性曲线(ROC)下面积作为评估模型效能的指标。**结果:**DenseNet201 模型获得了最好的训练和测试结果,在训练集和测试集中的最高预测符合率分别为 1.00 和 0.98;VGG19 模型用时最长,其在训练集和测试集中的预测符合率分别为 0.99 和 0.98,较 DenseNet201 略差;ResNet50 模型用时最短,但测试结果最差,在训练集和测试集中的最高符合率分别为 0.93 和 0.92。VGG19、ResNet50 和 DenseNet201 模型在验证集中的平均符合率为 0.96、0.92 和 0.98,召回率分别为 0.96、0.91 和 0.98,F1 评分分别为 0.96、0.91 和 0.98。DenseNet201 模型的 ROC 曲线下面积为 0.98,高于 VGG19 模型(0.95)和 ResNet50 模型(0.91)。**结论:**基于 DenseNet201 的 DTL 模型对甲状腺 CT 良恶性结节具有较高的分类效能,有助于提高影像诊断准确性。

【关键词】 迁移学习; 卷积神经网络; 甲状腺结节; 体层摄影术, X 线计算机

【中图分类号】 R **【文献标志码】** A **【文章编号】** 1000-0313(2021)08-0976-05

DOI:10.13609/j.cnki.1000-0313.2021.08.006

开放科学(资源服务)标识码(OSID):



Discrimination of benign and malignant thyroid nodules on CT images using different deep transfer learning models MENG Ming-zhu, PAN Chang-jie, ZHANG Ming, et al. Department of Radiology, the Second Hospital of Changzhou, Changzhou 213164, China

【Abstract】 Objective: To evaluate the use of different deep transfer learning (DTL) models to differentiate malignant from benign thyroid nodules on CT images. **Methods:** The diagnostic efficiencies of three DTL models (VGG19, ResNet50, DenseNet201) were tested and evaluated by using the same program architecture and data set. The highest accuracy of different models in training set and testing set, as well as the accuracy, recall rates, F1 scores and area under the subjects operating characteristic curves in validation set were taken as the performance indicators. **Results:** DenseNet201 model obtained the best training and testing results, the highest prediction accuracy of training set and testing set were 1.00 and 0.98. VGG19 model took the longest time, and were slightly worse in terms of training and testing prediction accuracy (0.99, 0.98) than DenseNet201 model. ResNet50 model took the shortest time for data analysis, but got the worst training and testing accuracy: 0.93 and 0.92, respectively. In validation set, the efficacy index values of the three models (VGG19, ResNet50, DenseNet201) were as follows: average accuracy of 0.96, 0.92 and 0.98, recall rate of 0.96, 0.91 and 0.98, F1 score of 0.96, 0.91 and 0.98, and area under the ROC curve of 0.95, 0.91 and 0.98). **Conclusion:** DTL model based on DenseNet201 has excellent efficacy for differentiating malignant from benign thyroid nodules on CT images, and can provide valuable information to improve the diagnostic accuracy.

【Key words】 Deep transfer learning; Convolutional neural networks; Thyroid nodules; Tomography, X-ray computed

作者单位: 213164 江苏, 常州市第二人民医院医学影像科

作者简介: 孟名柱(1984—), 男, 安徽凤阳人, 硕士研究生, 主治医师, 主要从事人工智能辅助影像诊断。

通讯作者: 陈罕奇, E-mail: chenhanqiab@sina.com

正确的术前诊断和临床分期对指导临床制订合理的甲状腺癌手术治疗方案具有重要意义,也直接影响甲状腺癌患者的预后,对于良性甲状腺结节的准确诊断也可避免不必要的手术。在临床实践中,由于甲状腺良性结节的发病率高于恶性结节,如对所有甲状腺结节进行活检将导致不必要的高活检率,而不会改善整体人群的健康状况^[1]。由于甲状腺结节患者的临床表现无特异性,影像学表现也存在“同病异影,异病同影”的特点,因此,及时、准确地对甲状腺结节进行诊断非常重要^[2]。随着 AI 技术的发展,影像学已经从主观感知技能转向客观科学^[3]。在传统的机器学习中,研究人员计算手工制作的数据特征,这些特征是使用预先定义各领域专业知识。这意味着我们需要为不同的任务设计不同的数据特性。例如,为甲状腺 CT 图像优化的特征可能对其它器官(如肺等)不起作用。在深度学习过程中,深度学习网络可以从数据学习到特定的网络特征,因此不需要指定预先设计的特征。这意味着人们可以应用相同的深度学习方法来解决许多不同的任务。然而,只有提供充足的数据进行训练,深度学习网络才能从数据中正确地学习特征,这很难在实践中进行应用^[4]。深度迁移学习(deep transfer learning, DTL)是采取预先训练好的深度学习网络并对其进行微调以学习新任务的过程,使得卷积神经网络可以在较小的数据集中加以应用,近年来已成为研究的热点^[5]。许多 DTL 模型已经应用于医学影像数据的分析,如图像分割、辅助对疾病的分类诊断及新辅助化疗评价等^[6-11]。

材料与方法

1. 一般资料

搜集 2016—2020 年在本院经手术病理证实的 610 例甲状腺结节患者的 CT 图像资料。良性组 308 例共 323 个结节(15 例患者为甲状腺左、右叶各有 1 个结节);男 96 例,女 212 例;年龄 16~80 岁,平均(49.5±13.7)岁;结节直径 5~46mm,平均(26.5±11.5)mm;病理类型包括甲状腺腺瘤 169 个(6 个为非典型性腺瘤),结节性甲状腺肿 141 个(5 个位为峡部),甲状腺囊肿 12 个,甲状腺脓肿 1 个。恶性组 302 例;男 89 例,女 213 例;年龄 21~79 岁,平均(48.4±13.6)岁;均为单发结节,直径(34.8±18.2)mm;病理类型包括乳头状癌 279 个(4 个位于峡部),滤泡性腺癌 16 个,髓样癌 6 个,淋巴瘤 1 个。

2. 检查方法

检查前向患者交代扫描中的注意事项,并签署碘对比剂注射知情同意书,除去头颈部的金属物品。使用 Siemens Somatom Definition Flash CT 扫描机,扫

描参数:层厚 2.5~3.0 mm,螺距 0.7~1.0。使用高压注射器经肘静脉注射非离子型碘对比剂碘克沙醇(320 mg I/mL),注射流率 2.0~2.5 mL/s,剂量 1.0 mL/kg。分别在注射对比剂后 20 和 40 s 启动动脉期和静脉期扫描。

分别在平扫、动脉期和静脉期图像中选择能够清晰显示病灶的层面,获取每个甲状腺结节的图像数量为 6~45 帧,最终获得良性组 CT 图像 5406 帧,恶性组 CT 图像 5340 帧。将良性组、恶性组图像分别按照 8:1:1 的比例随机分为训练集(良性组 4326 帧,恶性组 4290 帧)、测试集(良性组 54 帧,恶性组 525 帧)和验证集(良性组 540 帧,恶性组 525 帧)。

3. 建立 DTL 模型

计算机环境配置:采用 Windows 10 64 位操作系统,处理器为酷睿 i7 10700F,8 核心 16 线程,GPU 为 RTX2060,显存 6GB,程序运行时关闭其它所有应用程序。基本的 DTL 算法设计是在 Anaconda 开源软件中创建,基于 Python 编程语言的 TensorFlow2.0 和 keras2.2.4 开源编程包。选择二分类交叉熵损失函数(binary crossentropy),优化器选择 Adam,学习率设置为 0.0001,AlphaDropout 参数设置为 0.5,训练 Epoch 设置为 100,激活函数 relu,分类函数 sigmoid。在完成代码的编写之后,每个经过预先培训的深度学习网络体系结构(VGG19、ResNet50、DenseNet201)都被插入到代码相应位置,每种网络结构的训练和测试都使用相同的模型微调方法,以避免超参数不同对模型结果的影响,DTL 模型的中间层的可视化及模型简图见图 1~2。模型的每次训练和测试过程均被计算机记录下来,记录模型在训练和测试过程中获得的最高预测符合率和最小损失值。

4. 评估方法和分析指标

患者年龄和病灶直径以均数±标准差来表示,使用 SPSS 23.0 软件进行统计分析。最高符合率、最小损失值、平均分类符合率、召回率和 F1 评分及 ROC 曲线下面积为模型自动运算得出。平均分类符合率(Ac)、召回率(Rc)和 F1 评分为模型对验证集中良性组和恶性组总体数据验证结果的平均值,而不是模型在验证集中对良性组和恶性组的符合率、召回率和 F1 评分结果的平均数。计算公式分别如下:

$$Ac = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$PPV = \frac{TP}{TP + FP} \quad (2)$$

$$Rc = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times PPV \times Rc}{PPV + Rc} \quad (4)$$

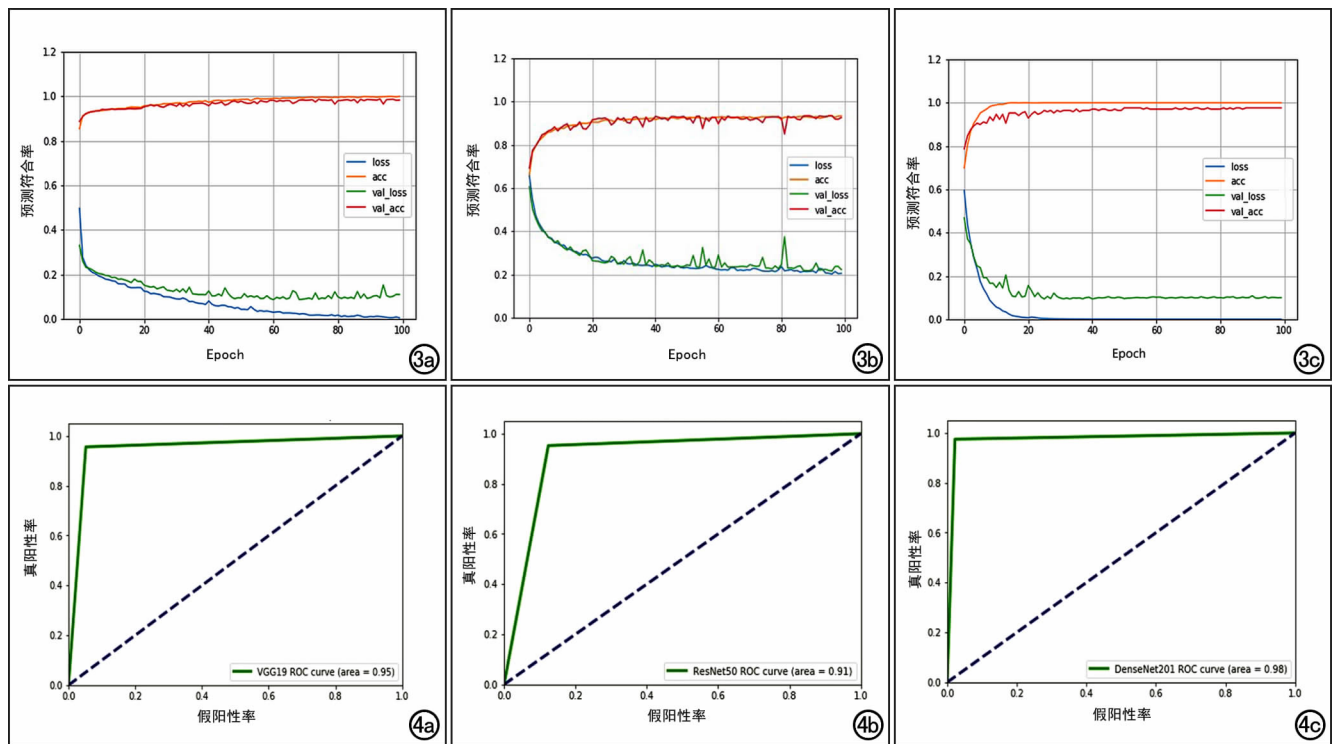


图3 三种 DTL 模型的学习曲线图(loss:训练集损失值,acc:训练集符合率;val_loss:测试集损失值;val_acc:测试集符合率)。DenseNet201 模型的损失值稳步下降和符合率稳步提高,提示模型分类效果比较理想。a) VGG19 模型; b) ResNet50 模型; c) DenseNet201 模型。图4 三种 DTL 模型在验证集中分类结果的 ROC 曲线示意图。a) VGG19 模型, AUC=0.95; b) ResNet50 模型, AUC=0.91; c) DenseNet201 模型, AUC=0.98。

2. DTL 模型的验证结果

在验证集中,3 个模型对良性和恶性病变的分类结果见表 2。以 DenseNet201 模型分类效能最佳,各项指标值均高于 VGG19 和 ResNet50 模型,而 ResNet50 模型分类效能最低(图 4)。

讨论

传统的图像特征提取包括图像的纹理、直方图特征等,这些特征的提取具有一定的复杂性和局限性。影像组学的发展为人工智能在影像领域应用奠定了基础^[10]。深度学习是机器学习的分支范畴,以卷积神经网络为常见网络模型,在多个工业领域中得到了广泛的应用,但最主要的应用领域还是关于图像识别,比如人脸识别、无人驾驶技术等。卷积神经网络与普通的神经网络的区别在于,前者包含了一个自动提取图像特征的卷积层和池化层,卷积层由一些矩形排列神经元组成,每个神经元可以与邻层的神经元建立连接,且同一特征平面的神经元共享同一卷积核^[11]。卷积核的大小一般以随机的小数矩阵形式初始化。卷积核所带来的直接好处是减少网络各层之间的连接,同时又降低了过拟合的风险。池化可以看作一种特殊的卷积过程。卷积和池化大大简化了模型复杂度,减少了模

型的参数,这使得深度学习不需要使用价格昂贵的图形处理器,而在普通的中央处理器上就可以实现。需要指出的是如果实验数据量较大,还是要在专业的计算机设备上运行。

深度迁移学习是利用原先基于大数据训练好的卷积神经网络模型为基础,并对其进行微调以学习新的任务的过程。本研究显示,基于 3 种(VGG19, ResNet50, DenseNet201)常用的卷积神经网络搭建的 DTL 模型对相同甲状腺 CT 图像数据进行训练时所用的时间差异较大,每个模型在训练集和测试集中的预测符合率也各不相同。ResNet50 模型所用时间最短(32 分钟),但预测符合率最差(训练集 0.93、测试集 0.92)。VGG19 所用时间最长(65 分钟),但是其预测符合率(训练集 0.99、测试集 0.98)低于 DenseNet201(训练集 1.00、测试集 0.98)。并且在验证集中 DenseNet201 模型较 VGG19 和 ResNet50 模型也显示出了其优越性。DenseNet201 模型在验证集的 ROC 曲线下面积达 0.98。这个研究结果与国外学者 Jaiswal 等^[5]的研究结果一致,但与 Blaivas 等^[12]的研究结果有相似、也有相反的结果。相似的结果是 ResNet50 模型用时均比 VGG19 和 DenseNet201 用时短;不一致的地方是本研究结果中 DenseNet201 模型

的预测符合率高于 VGG19 和 ResNet50 模型,这与 Blaivas 等^[12]的研究结果恰恰相反。分析原因可能有以下几点:①所用的数据集不同,不同 DTL 模型适合于研究不同的分类任务。Blaivas 等^[12]使用的是即时超声图像数据,而本研究使用的是甲状腺 CT 图像数据;②DTL 网络模型搭建时的微调方式可能不一样。本研究在原先网络模型基础之上构建了 4 个全连接层,神经元数量前 3 层分别为 1000、500 和 100,最后一层是分类输出层,神经元数量为 1;③超参数的设置可能不一样。本研究微调的前 3 层激活函数是 relu,分类输出层激活函数为 sigmoid,并且使用了 Alpha-Dropout 函数防止过拟合。对于②、③两点,Blaivas 等^[12]在文献中并没有给出具体的说明。说明同一个网络模型对不同的数据集的预测符合率和训练时间会表现出相当大的差异,不同的网络模型对同一组数据分析时,其预测符合率和训练时间也会各不相同。需要指出的是本研究 Epoch 设置为 100,如果经过更多的 Epoch,VGG19 和 ResNet50 模型的预测符合率可能也会有所提高,这有待于进一步的研究。

从历史上看,随着网络层数的加深,训练的误差率也会增加^[13]。本研究发现,对甲状腺 CT 图像数据,并不是越新和越深的卷积神经网络效果就越好。按出现时间先后和网络深度分别为 VGG19(2014 年,19 层)、ResNet50(2015 年,50 层)和 DenseNet201(2017 年,201 层)。虽然 VGG19 出现的时间较 ResNet50 早,而且网络深度不及 ResNet50,但是基于 VGG19 的 DTL 模型在甲状腺 CT 图像数据训练和测试中的结果都优于 ResNet50 模型。这个结果与我们的常规思维恰恰相反,分析原因可能是因为随着网络层数的增加,会面临梯度消失的问题,也就是经过的层数越多,则前面的信息就会渐渐减弱和消散,造成模型的预测符合率降低。DenseNet201 之所以有这么好的分类效果,也是由其自身的密集网络结构特征所决定的。DenseNet201 网络的组织形式和 ResNet50 类似,但是将原先 ResNet 的单元模块进行了替换,下采样过程也不相同。DenseNet201 进一步扩展了网络连接,对于网络的任意一层,该层前面所有层的特征图都是这层的输入,而该层的特征图又是后面所有层的输入。这就赋予了 DenseNet201 如下的优点:①减轻了梯度消失问题;②增强了 feature map 的传播,利用更充分了;③大大减少了参数量,在达到相同精度时,DenseNet201 的参数量和计算量均为 ResNet 的一半左右^[11,13]。

本研究的不足之处是缺少足够多的数据用于训练。DTL 为甲状腺良恶性结节的分类诊断提供更为丰富的信息,为临床制定治疗方案提供新的影像学依

据。相信随着研究的深入,在不久的将来会有符合率更高、鲁棒性更强的网络模型展现给医学界。

参考文献:

- [1] Foley WD. Thyroid nodules by US;more imaging and/or more intervention[J]. Radiology, 2019, 291(1): 100-101. DOI: 10. 1148/ radiol. 2019190189.
- [2] Farahati J, Mader U, Gilman E, et al. Changing trends of incidence and prognosis of thyroid carcinoma[J]. Nuklearmedizin, 2019, 58(2): 86-92. DOI: 10. 1055/a-0859-7454.
- [3] 韩英, 何生, 姜增誉, 等. 人工智能在乳腺影像领域的应用现状[J]. 放射学实践, 2019, 34(7): 813-816. DOI: 10. 13609/j. cnki. 1000-0313. 2019. 07. 020.
- [4] Lee SH, Park H, Ko ES. Radiomics in breast imaging from techniques to clinical applications: a review[J]. Korean J Radiol, 2020, 21(7): 779-792. DOI: 10. 3348/kjr. 2019. 0855.
- [5] 邹颖, 夏爽. 深度学习和影像组学在头颈部恶性肿瘤影像学中的研究进展[J]. 中华放射学杂志, 2020, 54(10): 1021-1024. DOI: 10. 3760/cma. j. cn112149-20191022-00858.
- [6] Jaiswal A, Gianchandani N, Singh D, et al. Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning[J/OL]. J Biomol Struct Dynam, 2020; e1-e8. DOI: 10. 1080/07391102. 2020. 1788642. [https://doi.org/10. 1080/07391102. 2020. 1788642](https://doi.org/10.1080/07391102.2020.1788642).
- [7] Cui R, Liu M. Hippocampus analysis by combination of 3D DenseNet and shapes for Alzheimer's disease diagnosis[J]. IEEE J Biomed Health Inform, 2019, 23(5): 2099-2107. DOI: 10. 1109/jbhi. 2018. 2882392.
- [8] Zhang X, Liang M, Yang Z, et al. Deep learning-based radiomics of B-mode ultrasonography and shear-wave elastography: improved performance in breast mass classification[J]. Front Oncol, 2020, 10: 1621. DOI: 10. 3389/fonc. 2020. 01621. [https://sci-hub. se/ 10. 3389/fonc. 2020. 01621](https://sci-hub.se/10.3389/fonc.2020.01621).
- [9] Sutton EJ, Onishi N, Fehr DA, et al. A machine learning model that classifies breast cancer pathologic complete response on MRI post-neoadjuvant chemotherapy[J]. Breast Cancer Res, 2020, 22(1): 57-68. DOI: 10. 1186/s13058-020-01291-w.
- [10] Yang S, Jiang L, Cao Z, et al. Deep learning for detecting corona virus disease 2019 (COVID-19) on high-resolution computed tomography: a pilot study[J]. Ann Translat Med, 2020, 8(7): 450-450. DOI: 10. 21037/atm. 2020. 03. 132.
- [11] 戴正行, 胡春洪, 王希明, 等. 基于 DenseNet 网络深度学习法 CT 图像人工智能分析技术判断肺结节良恶性[J]. 放射学实践, 2020, 35(4): 484-488. DOI: 10. 13609/j. cnki. 1000-0313. 2020. 04. 017.
- [12] 袁紫旭, 徐挺洋, 姚建华, 等. 人工智能在恶性肿瘤诊治中的应用[J]. 中华实验外科杂志, 2019, 36(2): 203-207. DOI: 10. 3760/cma. j. issn. 1001-9030. 2019. 02. 002.
- [13] Huang G, Liu Z, Maaten Lvd, et al. Densely connected convolutional networks[C/OL]. 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018. <https://arxiv.org/pdf/1608.06993.pdf>.

(收稿日期: 2021-06-02 修回日期: 2021-07-12)