

# 人工智能与医学影像融合发展:机遇与挑战

朱文珍, 胡琼洁

【关键词】 人工智能; 深度学习; 影像组学; 医学图像

【中图分类号】 R814.42; R445.2 【文献标识码】 A 【文章编号】 1000-0313(2019)09-0938-04

DOI:10.13609/j.cnki.1000-0313.2019.09.001

人工智能(Artificial intelligence, AI)是基于计算机来模拟人类的思维过程和智能行为的一门学科,随着 AI 技术的发展,目前已成为涉及计算机科学、心理学、哲学和语言学等多学科交叉的一门新兴前沿学科<sup>[1,2]</sup>。近几年来随着深度学习算法的出现、计算能力的指数级增长、丰富的大数据资源和基于训练的自主学习方法,以及计算机具有条件反射等类脑能力而发展出复杂人工智能,使得新一代 AI 技术迎来了爆发式的发展和应用。人工智能赋能医疗行业,在虚拟医师助理、病历与文献分析、药物研发、基因测序和影像辅助诊断、精准医学等方面都取得了令人惊喜的成果。其中,医学影像与人工智能的结合是最具发展前景的领域。2012年以后,随着深度卷积神经网络技术的兴起和应用,AI 在计算机视觉领域的发展取得了突破。计算机视觉基于图像识别,可以对医学影像数据进行深入分析,获取更多有价值的信息。通过大量数据的训练和学习,使其分析能力不断提升,从而在精准诊断方面显示出广阔的应用前景。目前在肿瘤检出、自动结构式报告、定性和定量诊断、肿瘤提取和放疗靶器官勾画等方面已有较多的临床应用和研究<sup>[1,2]</sup>。当人工智能方法作为辅助医师的工具无缝集成到临床工作流程中时,通过提供预先筛选的图像和确定的特征可更准确地进行影像学评估,且可重复性高,能显著提高工作效率,减少误诊和漏诊,并可对疗效进行监测<sup>[1-3]</sup>。

## 人工智能在肿瘤影像学领域的应用及研究

人工智能尤其是深度学习算法在医学影像领域应用最为成熟的为肿瘤影像,在肿瘤的预防、早期发现、诊断和干预、预后评估等方面提供了可能<sup>[1,2,4]</sup>。在肿瘤影像学中,人工智能在三个主要的临床任务中发挥了巨大的作用,即肿瘤的检测、表征和监测<sup>[2]</sup>。

### 1. 肿瘤的检测

作者单位:430030 武汉,华中科技大学同济医学院附属同济医院放射科

作者简介:朱文珍,二级教授,主任医师,博士生导师,主要从事神经影像学临床和研究工作。

在传统的临床工作流程中,放射科医师依靠视觉感知来识别图像上可能的异常,依靠知识、经验和对正常影像表现的理解,在一定程度上属于主观决策。近年来,多项研究表明使用人工智能软件后影像科医师对病灶的检出能力明显提高<sup>[1,5]</sup>,其中在肺结节检出的临床应用已经较为成熟。Jeon 等<sup>[6]</sup>的研究中 7 位影像医师先独自对 134 份胸部 CT 影像进行阅片,后参考 CAD 自动检测结果进行修改,结果表明在 CAD 辅助下结节的平均发现率从最初的 77% 提升到了 84%。Lee 等<sup>[5]</sup>发现 CAD 对孤立性结节的诊断敏感度高于医生肉眼观察。2018 年, Silva 等<sup>[7]</sup>对基于意大利多中心肺癌筛查数据库中的 2303 例使用人工智能软件,其 GGN 检出率达 90.2% (175/194)。Kooi 等<sup>[8]</sup>利用卷积神经网络(CNNs)深度学习检测乳腺钼靶 X 线摄影中的病变,其具有较高的敏感性,与人类阅读水平相近。此外, AI 对 MRI 脑转移瘤和前列腺病变的检出同样具有较高的灵敏度<sup>[9,10]</sup>。在模式识别的背景下,具有可疑图像特征的区域 AI 可突出显示并呈现给阅片者,从而有效避免对病灶的遗漏,因此 AI 可应用于人群的初始肿瘤筛查。但需要指出是,目前的 AI 软件都是针对特定的应用迭代定制的,因此需要仔细权衡其特异性,放射科医师必须对其输出结果进行仔细评估。如何在保证较高敏感性的同时有效降低检测的假阳性率,是 AI 未来研究的重点之一。

### 2. 表征

表征,又称为特征描述,是泛指病灶分割、肿瘤诊断和分期及基因组学的特定术语<sup>[2]</sup>。这些任务是通过量化病灶的影像学特征(如大小、范围和内部结构等)来完成的。对人类无法解释的肿瘤的大量特征,通过人工智能的方法,可进行复杂的非线性关系分析,来发现各变量间的相关性<sup>[1,2]</sup>。人工智能技术可以自动识别病灶的各种特征,同时将它们作为图像的生物标志。这些生物标志可应用于临床决策和对肿瘤的干预,包括风险评估、鉴别诊断、预后评估和治疗反应的监测等<sup>[2,4]</sup>。对器官和病变组织进行特异和准确的分割至关重要。在肿瘤放疗过程中,必须精确地分割肿瘤和非肿瘤组织的范围,才能制订精确的放疗计划。有研

究表明,在基于超声和MRI的病灶分割任务中,深度学习技术与放射科医师的表现不相上下<sup>[11-15]</sup>。

对于肿瘤良恶性的鉴别,人工智能同样表现出其优势。以目前研究较为成熟的肺结节为例,由于良性和恶性结节的CT表现具有相似性,使得人类很难准确预测肺部病灶的良恶性。Awai等<sup>[16]</sup>研究提取了结节的大小、密度、形状、钙化体积和空腔体积,开发了新的AI结节分析系统,对结节的恶性概率打分,对31个直径小于3cm的肺结节进行验证,结果显示它可以显著提高低年资影像医生的诊断准确度。Wu等<sup>[17]</sup>利用CT图像定量纹理特征能够很好地区分良恶性肺结节。此外,近年来,不少学者利用人工智能方法探讨肿瘤的病理亚型与影像学的相关性。Coroller等<sup>[18]</sup>通过建立深度学习模型进行1级脑膜瘤与2、3级脑膜瘤的鉴别。我们的研究团队Zhang等<sup>[19]</sup>利用基于机器学习多参数MRI直方图和纹理特征优化了胶质瘤病理分级以及判定肿瘤增殖程度。人工智能在肿瘤分期,如肿瘤淋巴结转移(TNM)的研究也取得了一些突破。对于PET-CT中淋巴结转移的分类任务,深度学习比放射科医师具有更高的敏感性,但特异性较低<sup>[20]</sup>。Kirienko等<sup>[21]</sup>利用卷积神经网络实现了肺癌基线PET/CT TNM中T参数的自动评估,可作为辅助肺癌患者分期的工具。

人工智能在肿瘤基因组学方面的研究近年来也取得了突飞猛进的发展。我们的研究团队Su等<sup>[22]</sup>研究表明多模态MRI影像组学能够有效预测胶质瘤亚型和KI-67水平,预测胶质瘤的生物学行为。此外基于MRI纹理分析可区分星形细胞瘤与1p/19q共缺失少突胶质细胞瘤两种类型低级别胶质瘤<sup>[23]</sup>。Korfiatis等<sup>[24]</sup>发现MRI纹理特征可作为预测胶质母细胞瘤MGMT甲基化状态的生物标志物。Chang等<sup>[25]</sup>利用卷积神经网络实现了在MRI图像判断低级别与高度胶质瘤IDH状态的可能。Yang等<sup>[26]</sup>和Wang等<sup>[27]</sup>均证实了深度学习模型可无创预测肺腺癌EGFR突变状态,为临床决策提供帮助。

### 3. 肿瘤治疗监测

疾病监测对肿瘤的诊断和治疗反应的评估至关重要。人工智能在监测肿瘤随时间的变化方面也发挥着作用,无论是在自然史上还是在治疗反应中。基于人工智能的监测方法能够捕捉到随着时间的推移病灶的大量不同图像的鉴别特征,包括纹理的细微变化和对象内部的异质性,这些特征超出了人类读者所能测量到的<sup>[2]</sup>。我们的研究团队Zhang等<sup>[28]</sup>发现常规MRI图像纹理分析能准确预测低度胶质瘤的早期恶性转化。CT影像纹理特征异质性参数能够预测非小细胞肺癌的第一轮化疗效果<sup>[29]</sup>。但该领域仍处于起步阶

段,将成为未来研究的方向。除了成像,微创生物标记物,如循环肿瘤DNA也正在开发,可用于肿瘤的诊断和疾病的纵向跟踪<sup>[30]</sup>。微创生物标记物结合医学影像图像,再基于人工智能的综合诊断,为更准确地评估预后和实时监测疾病提供精准的医学依据。

### 人工智能在医学影像学其它领域的应用

人工智能在非肿瘤性疾病的筛查、诊断与鉴别诊断及预后评估等方面也显示出一定的临床价值。Lee等<sup>[31]</sup>设计了具有区域异常表征的可解释阿尔茨海默病诊断模型。此外,研究发现人工智能技术可对更薄的皮层厚度及更小的海马亚区进行鉴别诊断<sup>[32]</sup>。Chilamkurthy等<sup>[33]</sup>研究发现深度学习算法可以准确地识别需要紧急干预治疗的头部CT异常表现,包括各种类型的颅内出血、颅骨骨折和中线移位等,这为自动分诊提供了可能。相关学者应用深度学习网络实现脑外伤急诊患者全自动检测系统,有助于放射科和急诊科医师减少诊断时间和人为错误<sup>[32]</sup>。Cheng等<sup>[34]</sup>应用深度学习算法可实现骨盆X线髋部骨折的检测和可视化,可能有助于紧急筛查和评估。此外,Tao等<sup>[35]</sup>基于深度学习方法实现了准确和全自动量化多中心电影MR图像左心室的功能。

人工智能有望影响临床放射工作流程中的其它基于图像的任务,包括图像采集之后的预处理步骤,以及随后的报告和综合诊断。深度学习模式可以显著改善图像质量。Xu等开发了一个完全卷积编码器-解码器网络,采用跳过连接和残差学习策略来恢复PET-MRI高分辨率<sup>[32]</sup>。Ardila等<sup>[36]</sup>通过随访验证训练和病理佐证,建立一个端到端方法,回顾性使用输入的CT数据来执行定位和肺癌风险分类任务,该模型具有重要的潜在临床意义,可提高工作流程的效率和诊断一致性。人工智能自动化报告可以形成更标准化的术语,提高工作效率。Wang等<sup>[37]</sup>研究提示基于深度学习的智能排版和结构化报告系统可优化临床影像工作全流程,显著提升影像科的工作效率和诊断准确性。然而,目前自动化报告缺乏严格和统一的标准,尚需在今后进一步完善。

### 人工智能在医学影像领域面临的挑战

#### 1. 标准化数据和数据库的建立

数据仍然是人工智能系统最核心和最关键的组成部分,包括图像存档和通信系统(picture archiving and communication systems, PACS)和医学数字成像和通信(digital imaging and communications in medicine, DICOM)在内的图像数据标准化存储便于访问和检索。然而,这些数据很少在标签、注释、分隔和质量保

证等方面进行整理,医疗数据的管理需要训练有素的专业人员,时间和成本上都很昂贵,成为了开发自动化临床解决方案人工智能模型的一个主要瓶颈<sup>[2]</sup>。标准化数据,特别是多种成像模式和解剖位置的数据,在医学领域尤其重要<sup>[38]</sup>。公开的医疗数据量越来越多,这是一个令人鼓舞的进步,但是我们不仅需要增加公开访问的数据库,而且需在性能、普遍性和可重复性方面进行优化。同时鼓励大学或者大学机构、专业团体和政府机构克服某些基本的技术、法律和伦理问题,共享经过验证的数据,促进智能合作<sup>[1,39]</sup>。

## 2. 人工智能深度学习模式——“黑匣子”

另一个限制是人工智能深度学习模式对特定结果的可解释性及对失败的预知能力<sup>[1,2]</sup>。深度学习直接从数据中学习识别图像中有助于预测结果的特定特征,其是高度假设的,导致缺乏对深度学习如何得出某些结论的理解。由于缺乏透明性,因此很难预测故障或将特定结论广泛地推广到不同的成像硬件、扫描协议和患者群体。毫不奇怪,许多应用于影像学的无法解释的人工智能深度学习系统被称为“黑匣子”。

## 3. 数据安全和伦理问题

使用患者数据来训练这些人工智能系统可能会引发数据安全和隐私方面的伦理问题。符合《健康保险可携性与责任法》的存储系统促进了更严格的隐私保护。相关研究探索了一种系统,它可以让多个实体不共享输入数据集,只共享训练过的模型的情况下联合训练 AI 模型<sup>[40]</sup>。

此外,当前人工智能工具的一个共同特点是能较好地执行单一任务,无法处理多个任务。放射科医生是人工智能培训过程中的关键要素,贡献知识并监督效率,然而真正懂人工智能的医生相对较少。我们影像科医师需要充分利用这些资源,促使 AI 成为更好的工具,发挥他们最大的作用。只有这样,我们才能做到不仅能监督结果,还能采用验证手段来解释其背后的原因,以及发现潜在的、可能被忽视的隐藏信息,更高效地服务于临床。

### 参考文献:

- [1] Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology[J]. *Nat Rev Cancer*, 2018, 18(8): 500-510.
- [2] Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications[J]. *CA Cancer J Clin*, 2019, 69(2): 127-157.
- [3] McDonald RJ, Schwartz KM, Eckel LJ, et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload[J]. *Acad Radiol*, 2015, 22(9): 1191-1198.
- [4] Wang X, Guo J, Gu D, et al. Tracking knowledge evolution, hotspots and future directions of emerging technologies in cancers research: a bibliometrics review[J]. *J Cancer*, 2019, 10(12): 2643-2653.
- [5] Lee JW, Goo JM, Lee HJ, et al. The potential contribution of a computer-aided detection system for lung nodule detection in multidetector row computed tomography[J]. *Invest Radiol*, 2004, 39(11): 649-655.
- [6] Jeon KN, Goo JM, Lee CH, et al. Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening CT[J]. *Invest Radiol*, 2012, 47(8): 457-461.
- [7] Silva M, Schaefer-Prokop CM, Jacobs C, et al. Detection of subsolid nodules in lung cancer screening: complementary sensitivity of visual reading and computer-aided diagnosis[J]. *Invest Radiol*, 2018, 53(8): 441-449.
- [8] Kooi T, Litjens G, Van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions[J]. *Med Image Anal*, 2017, 35(1): 303-312.
- [9] Grøvik E, Yi D, Iv M, et al. Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI[J]. *J Magn Reson Imaging*, 2019 May 2. DOI: 10.1002/jmri.26766.
- [10] Alkadi R, Taher F, El-Baz A, et al. A deep learning-based approach for the detection and localization of prostate cancer in T<sub>2</sub> magnetic resonance images[J]. *J Dig it Imaging*, 2019, 32(5): 793-807.
- [11] Raudaschl PF, Zaffino P, Sharp GC, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015[J]. *Med Phys*, 2017, 44(5): 2020-2036.
- [12] Wang Y, Zhong Z, Hua J. Deep organ net: on-the-fly reconstruction and visualization of 3D/4D lung models from single-view projections by deep deformation Network[J]. *IEEE Trans Vis Comput Graph*, 2019 Aug 22. DOI: 10.1109/TVCG.2019.2934369.
- [13] Zhang R, Huang L, Xia W, et al. Multiple supervised residual network for osteosarcoma segmentation in CT images[J]. *Comput Med Imaging Graph*, 2018, 63(1): 1-8.
- [14] Jia H, Xia Y, Song Y, et al. 3D APA-net: 3D adversarial pyramid anisotropic convolutional Network for prostate segmentation in MR images[J]. *IEEE Trans Med Imaging*, 2019 Jul 11. DOI: 10.1109/TMI.2019.2928056.
- [15] Karimi D, Zeng Q, Mathur P, et al. Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images[J]. *Med Image Anal*, 2019, 57(10): 186-196.
- [16] Awai K, Murao K, Ozawa A, et al. Pulmonary nodules: estimation of malignancy at thin-section helical CT-effect of computer-aided diagnosis on performance of radiologists[J]. *Radiology*, 2006, 239(1): 276-284.
- [17] Wu W, Pierce LA, Zhang Y, et al. Comparison of prediction models with radiological semantic features and radiomics in lung cancer diagnosis of the pulmonary nodules: a case-control study[J]. *Eur Radiol*, 2019 May 21. DOI: 10.1007/s00330-019-06213-9.
- [18] Coroller TP, Bi WL, Huynh E, et al. Radiographic prediction of meningioma grade by semantic and radiomic features[J]. *PLoS*

- ONE, 2017, 12(11):e0187908.
- [19] Zhang X, Yan LF, Hu YC, et al. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features[J]. *Oncotarget*, 2017, 8(29): 47816-47830.
- [20] Wang H, Zhou Z, Li Y, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from <sup>18</sup>F-FDG PET/CT images[J]. *EJNMMI Res*, 2017, 7(1):11.
- [21] Kirienko M, Sollini M, Silvestri G, et al. Convolutional neural Networks promising in lung cancer T-parameter assessment on baseline FDG-PET/CT[J]. *Contrast Media Mol Imaging*, 2018, 30(10):1382309.
- [22] Su C, Jiang J, Zhang S, et al. Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behavior[J]. *Eur Radiol*, 2019, 29(4): 1986-1996.
- [23] Zhang S, Chiang GC, Magge RS, et al. MRI based texture analysis to classify low grade gliomas into astrocytoma and 1p/19q codeleted oligodendroglioma[J]. *Magn Reson Imaging*, 2019, 57(4):254-258.
- [24] Korfiatis P, Kline TL, Coufalova L, et al. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas[J]. *Med Phys*, 2016, 43(6):2835-2844.
- [25] Chang K, Bai HX, Zhou H, et al. Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging[J]. *Clin Cancer Res*, 2018, 24(5):1073-1081.
- [26] Yang X, Dong X, Wang J, et al. Computed tomography-based radiomics signature: A potential indicator of epidermal growth factor receptor mutation in pulmonary adenocarcinoma appearing as a subsolid nodule[J]. *Oncologist*, 2019 Apr 1. DOI: 10.1634/theoncologist.2018-0706.
- [27] Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning[J]. *Eur Respir J*, 2019, 53(3). DOI: 10.1183/13993003.00986-2018.
- [28] Zhang S, Chiang GC, Magge RS, et al. Texture analysis on conventional MRI images accurately predicts early malignant transformation of low-grade gliomas[J]. *Eur Radiol*, 2019, 29(6): 2751-2759.
- [29] Ravanelli M, Farina D, Morassi M, et al. Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: prediction of the response to the first-line chemotherapy[J]. *Eur Radiol*, 2013, 23(12): 3450-3455.
- [30] Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA[J]. *Nat Rev Cancer*, 2017, 17(4):223-238.
- [31] Lee E, Choi JS, Kim M, et al. Alzheimer's disease neuroimaging initiative. Toward an interpretable Alzheimer's disease diagnostic model with regional abnormality representation via deep learning[J]. *Neuroimage*, 2019, 22(8):116113.
- [32] 李诗卉, 吴迪, 周依然, 等. RSNA2018人工智能及分子影像在中枢神经系统的研究进展[J]. *放射学实践*, 2019, 34(3):248-251.
- [33] Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study[J]. *Lancet*, 2018, 392(10162):2388-2396.
- [34] Cheng CT, Ho TY, Lee TY, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs[J]. *Eur Radiol*, 2019, 29(10):5469-5477.
- [35] Tao Q, Yan W, Wang Y, et al. Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study[J]. *Radiology*, 2019, 290(1):81-88.
- [36] Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography[J]. *Nat Med*, 2019, 25(8):1319.
- [37] Wang Y, Yan F, Lu X, et al. IILS: Intelligent imaging layout system for automatic Imaging report standardization and intra-interdisciplinary clinical workflow optimization[J]. *EBioMedicine*, 2019, 44(6):162-181.
- [38] Purushotham S, Meng C, Che Z, et al. Benchmarking deep learning models on large healthcare datasets[J]. *J Biomed Inform*, 2018, 83(7):112-134.
- [39] Chang K, Balachandar N, Lam C, et al. Distributed deep learning networks among institutions for medical imaging[J]. *J Am Med Inform Assoc*, 2018, 25(8):945-954.
- [40] Shokri R, Shmatikov V. Proceedings of the 22nd ACM SIGSAC conference on computer and communications security[C]. Denver: 22nd ACM SIGSAC Conference, 2015:1310-1321.

(收稿日期:2019-09-02)

## · 人工智能影像学专题 ·

## 基于三维卷积神经网络肺结节深度学习算法模型临床效能初步评估

王祥, 李清楚, 邵影, 邹勤, 孙安, 陈彦博, 陈如谭, 高耀宗, 刘士远, 萧毅

**【摘要】 目的:**为评价人工智能模型的应用价值,本研究在专家共识的基础上建立了肺结节标准测试集,对前期建立的一种基于三维卷积神经网络肺结节深度学习算法模型进行验证,评价该模型的临床效能和限度。**方法:**基于胸部 CT 肺结节数据标注与质量控制专家共识建立标准测试数据集,对前期建立的基于三维卷积神经网络的肺结节深度学习算法模型及传统 CAD 系统(Siemens syngo. via VB 3.0 和 Philips ISP V8)进行检验,在肺结节检出灵敏度、精准度以及平均每例假阳性个数等多个指标方面进行优效验证。**结果:**针对测试数据集中的肺结节,Syngo. via 工作站检出灵敏度为 36%,精准度为 69%,平均每例假阳性 1.2 个;Philips ISP 工作站肺结节检出灵敏度为 34%,精准度为 73%,平均每例假阳性 0.9 个;三维卷积神经网络的肺结节深度学习算法模型检出灵敏度为 90%,精准度为 71%,平均每例假阳性 2.8 个。**结论:**该三维卷积神经网络算法模型相较于传统 CAD 系统,肺结节检出灵敏度显著提升。由于训练数据集的偏倚等问题,灵敏度仍有进一步提升的空间。通过针对性地补充训练数据集,如增加磨玻璃结节的比重,可进一步提升肺结节检出灵敏度。改进之后的模型有望成为影像医生肺癌筛查工作的得力助手。

**【关键词】** 肺结节;人工智能;卷积神经网络;计算机辅助筛查;体层摄影术,X 线计算机;磨玻璃结节

**【中图分类号】** TP18;R734.2;R814.42 **【文献标识码】** A

**【文章编号】** 1000-0313(2019)09-0942-05

DOI:10.13609/j.cnki.1000-0313.2019.09.002

开放科学(资源服务)标识码(OSID):



### A preliminary clinical evaluation of a 3D convolutional neural network based deep learning system

WANG Xiang, LI Qing-chu, SHAO Ying, et al. Department of Radiology, Changzheng Hospital of Naval Military Medical University, Shanghai 200003, China

**【Abstract】 Objective:** To assess the application value of artificial intelligence model, a standardized testing set constructed according to the expert consensus were used to validate a prior developed 3D convolutional neural network (3D CNN) algorithm and evaluate its clinical efficacy and limitation. **Methods:** A standardized testing dataset was constructed based on the expert consensus on the rule and quality control of pulmonary nodule annotation based on thoracic CT and used to test a prior developed 3D CNN algorithm and two traditional CAD systems (Siemens syngo. via VB 3.0 and Philips ISP V8. Several metrics). Several evaluation indicators, such as detection sensitivity, precision and average number of false positives per subject (average FP), were calculated. **Results:** Testing with the standardized chest CT dataset, the sensitivity, precision and average FP of Syngo. via VB 3.0 were 36%, 69%, and 1.2, respectively. The three indicators of Philips ISP V8 were 34%, 73%, and 0.9, respectively. The studied 3D CNN algorithm obtained a sensitivity of 90%, a precision of 71%, and an average FP of 2.8. **Conclusion:** Compared to the traditional CAD systems, the 3D CNN algorithm shows a significant improvement on detection sensitivity of lung nodules. However, the detection sensitivity still needs further improvement due to the unbalanced training dataset. Studying with a more diversified training dataset, e. g., more included ground-glass nodules, may further improve the detection sensitivity. The

作者单位:200003 上海,海军军医大学(原第二军医大学)附属长征医院影像科(王祥,李清楚,邹勤,孙安,陈如谭,刘士远,萧毅);201210 上海,上海联影智能医疗科技有限公司(邵影,陈彦博,高耀宗)

作者简介:王祥(1992-),男,山东临沂人,硕士,住院医师,主要从事胸部影像诊断工作。

通讯作者:萧毅, E-mail: cz-xiaoyi@smmu.edu.cn

基金项目:上海市科学技术委员会基金项目(NO:17411952400);国家重点研发计划政府间合作项目(NO:2016YFE0103000);上海市卫计委智慧医疗项目(NO:2018ZHYL0101);科技部国家重点研发计划(NO:2018YFC0116404);

AI system can help the radiologists in lung cancer screening in a more efficient way after improvement.

**【Key words】** Lung nodule; Artificial intelligence; Convolutional neural network; Computer-aided detection; Tomography, X-ray computed; Ground-glass nodule

肺癌是全球范围内发病率及死亡率最高的恶性肿瘤<sup>[1,2]</sup>,患者早期往往没有任何临床症状而被忽视,研究表明肺癌患者五年生存率约为10%~16%<sup>[3]</sup>,大多数患者检出时已经失去了早期根治的机会。早期发现、早期诊断及早期治疗可以显著提高肺癌患者五年生存率<sup>[4]</sup>。低剂量胸部CT具有较高的空间分辨率、扫描时间短及可对病变进行三维重建等优点,广泛应用于肺结节的检出及鉴别诊断。然而,CT筛查数量急剧增加及扫描层厚越来越薄,影像医生工作强度显著提升,临床工作中漏诊及误诊现象时有发生<sup>[5,6]</sup>。因此一个准确的肺结节自动检测系统将给影像科医生提供巨大帮助,有望真正成为医生的好帮手。

传统肺结节计算机辅助检查技术虽然可以提高肺结节检测的效率和准确性,但是由于该方法对磨玻璃结节检出率低、假阳性率高等问题,因此在临床应用并未常规使用<sup>[7,8]</sup>。随着医疗大数据库的建立及计算机硬件水平的发展,深度学习算法在医疗诊断领域优势明显,为实现医学影像的自动智能诊断提供了新的契机<sup>[9]</sup>。目前,虽然部分研究者采用深度学习算法对肺结节进行自动检测<sup>[10,11]</sup>,但大部分算法还是基于2D或2.5D影像<sup>[12]</sup>,没有有效利用CT影像中的三维信息。因此,笔者建立了基于三维卷积神经网络的肺结节自动检出模型,并对其在独立测试集上进行验证。

## 材料与方法

### 1. 病例资料

本研究经长征医院伦理委员会批准并免除了患者知情同意书。本研究属于回顾性研究,所涉及到的患者信息均进行了脱敏处理。所有数据分为训练集及测试集两部分,训练集用于训练三维卷积神经网络,测试集用于验证三维卷积神经网络的效能。训练数据集来自2018~2019年多家医院以及临床机构收集到的回顾性数据。训练数据集分别来自体检、门诊及住院患者中采集的胸部CT影像。

病例纳入标准:①行薄层CT(层厚 $\leq 1$  mm)检查;②术前未有治疗史;③术后病理报告完善;④所有结节大小均 $\leq 3$  cm;⑤薄层CT上无明显空洞。

病例排除标准:①肺叶扫描不全;②图像存在严重伪影;③图像存在缺层或断层;④图像不符合DICOM 3.0协议。

通过纳入及排除标准的筛选,最终7144例患者共45662个结节纳入训练集。7144例患者中女4000例(56%),年龄20~96岁,平均(51.0 $\pm$ 14.7)岁;男3144例(44%),年龄19~94岁,平均(52.2 $\pm$ 14.3)岁。

测试数据集来自长征医院2012年7月至2017年3月回顾性纳入的400例患者(男173例,女227例,年龄27~80岁)。采用与训练集相同的纳入标准与排除标准,最终经过筛选96例患者共733个肺结节纳入测试集。所有人组病例均为多发结节,每个结节视为独立结节,本研究没有考虑不同结节之间的相关性影响。训练集与测试数据集中肺结节大小和类型的分布见表1。

表1 数据集中不同大小、不同类型结节的分布

类型	训练数据集		测试集	
	结节个数(个)	患者例数(例)	结节个数(个)	患者例数(例)
实性	32223	6510	277	80
<5mm	26925	6132	240	74
5~10mm	4390	2590	35	25
>10mm	908	830	2	1
磨玻璃	3589	2291	390	95
<5mm	1677	1264	153	71
5~10mm	1394	1098	178	82
>10mm	518	496	59	42
胸膜	4681	2824	39	24
钙化	5169	2320	27	19
汇总	45662	7144	733	96

### 2. 图像采集

测试集数据扫描范围自肺尖至后肋膈角下缘,包括两侧胸壁、腋窝。患者吸气末屏气扫描,避免呼吸运

表2 测试集5台CT机器详细扫描参数

相关参数	CT16(1)	CT128(2)	GE64(3)	CT128(4)	CT256(5)
厂家	Toshiba	Philips	General Electric	Philips	Philips
管电压(kV)	120	120	120	120	120
管电流(mA)	50~150	50~150	Auto	50~150	50~150
探测器准直(mm)	16 $\times$ 0.5	64 $\times$ 0.625	64 $\times$ 0.625	64 $\times$ 0.625	128 $\times$ 0.625
矩阵	512 $\times$ 512	512 $\times$ 512	512 $\times$ 512	512 $\times$ 512	512 $\times$ 512
重建层厚(mm)	1.000	1.000	1.000	1.000	0.625

动伪影。所有测试病例均行胸部薄层 MSCT 扫描, CT 图像采集来自 5 台 CT 机器, 详细扫描参数见表 2。

### 3. 标准测试数据集

参照胸部 CT 肺结节数据标注与质量控制专家共识<sup>[13]</sup>, 建立本次肺结节测试数据集。对入组的 733 个结节的 CT 影像学特征在标准肺窗(窗宽 1500 HU, 窗位 -450 HU)进行标注, 每个结节均由 3 位胸部影像诊断医师(5 年以上胸部亚专业工作经验)同步盲法诊断并标注, 判断结节类型, 当诊断结果不一致时, 由第 4 位高年资亚专业国内权威影像医师作为仲裁专家进行判定, 最后汇总意见作为结节诊断及标注金标准。本测试数据库根据结节大小分为 3 组: <5 mm、5~10 mm 和 >10 mm。结节密度分为钙化结节与非钙化结节(实性结节、亚实性结节)。其中, 亚实性结节进一步分为纯磨玻璃密度结节与混杂磨玻璃结节, 纯磨玻璃密度结节指肺实质内圆型或类圆形边界清楚的密度增高影, 其内血管及支气管显示清晰, 通过纵隔窗判断有无实性肿瘤成分, 无实性成分者为纯磨玻璃密度结节, 反之判断为混杂磨玻璃密度结节<sup>[14]</sup>。特殊部位的结节主要纳入胸膜结节, 定义为与胸膜广基底相连的圆形、类圆形或不规则形的局灶性密度增高影。结节尺寸测量以肺部 CT 肺结节数据标注与质量控制专家共识为标准, 在标准肺窗测量, 选取结节最大横截面长径及短径计算其平均直径, 平均直径为(长径+短径)/2<sup>[13]</sup>, 并记录检出结果。本次测试集总纳入 733 个肺结节, 其中实性结节 277 个, 磨玻璃结节 390 个, 胸膜结节 39 个, 钙化结节 27 个。肺结节大小范围为 3~30 mm, 平均肺结节直径为 6.5 mm, 其中恶性结节 145 个, 均经病理证实。

### 4. CAD 性能统计学指标

本研究从肺结节检出的灵敏度(Sensitivity), 精确度(Precision)以及平均每例假阳性个数来衡量肺结节筛查算法的综合性能。真阳性检出(True Positive)指当 CAD 系统检出的肺结节定位框与金标准中任意一个肺结节定位框有重合时, 则这个肺结节检出被认为是一个真阳性检出; 反之, 则认为这个肺结节检出是假阳性检出。CAD 系统的肺结节灵敏度定义为 CAD 系统真阳性个数/金标准肺结节总个数; 精确度定义为 CAD 系统真阳性个数/CAD 系统检出肺结节的总个数; 平均每例假阳性个数定义为总共假阳性检出数量/总共测试集的病例个数。肺结节检出的灵敏度反映了 CAD 系统对肺结节的检出率, 精确度和平均每例假阳性个数体现了 CAD 系统肺结节检出的假阳性。

分别使用本研究建立的三维卷积神经网络模型, Siemens syngo.via(VB 3.0)及 Philips ISP(V8)CAD

软件进行标准测试数据集中肺结节的检出。通过与医生确立的肺结节“金标准”对比, 计算三类方法在肺结节检出的灵敏度、精确度以及平均每例患者假阳性检出的个数, 同时比较对于不同大小、不同类型肺结节检出的灵敏度。

## 结果

在测试集 733 个肺结节中, 三维卷积神经网络算法模型共检出 658 个结节, 另外有 273 个为假阳性检出。整体肺结节检出灵敏度为 90%, 精确度为 71%, 平均每例患者假阳性检出 2.8 个。验证的三维卷积神经网络的自由响应受试者工作特征(free-response receiver operating characteristics, fROC)曲线<sup>[15]</sup>见图 1, 绿色框在 fROC 曲线所对应的点是该三维卷积神经网络的临床应用的操作点。

Siemens syngo.via 工作站 CAD 软件共检出 263 个结节, 另外有 117 个为假阳性检出, 整体肺结节检出灵敏度为 36%, 精确度为 69%, 平均每例患者假阳性检出 1.2 个。在同样的假阳性检出下, fROC 曲线显示三维卷积神经网络的灵敏度为 67%, z 检验表明该灵敏度显著优于 syngo.via 工作站( $P < 0.05$ )。

Philips ISP 工作站 CAD 软件共检出 249 个结节, 另外有 90 个为假阳性检出, 整体肺结节检出灵敏度为 34%, 精确度为 73%, 平均每例患者假阳性检出 0.9 个(表 3)。在同样的假阳性检出下, 三维卷积神经网络的灵敏度为 59%, z 检验表明该灵敏度显著优于 Philips ISP( $P < 0.05$ )。

表 3 基于标准测试集的肺结节检测各项指标

检出方法	灵敏度 (%)	精确度 (%)	平均每例假阳性检出
三维卷积神经网络	90	71	2.8
Siemens syngo.via	36	69	1.2
Philips ISP	34	73	0.9

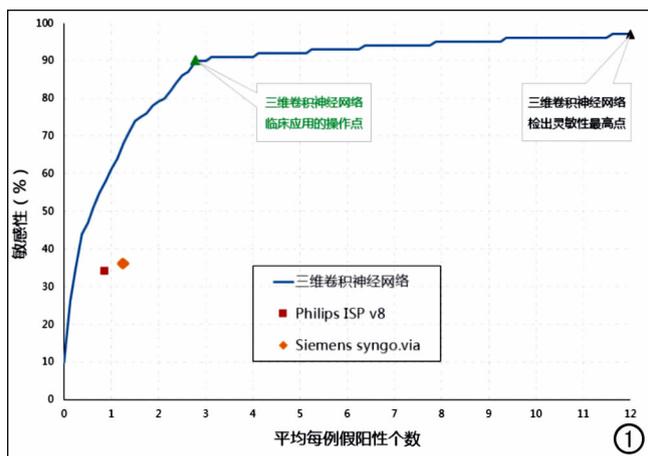


图 1 三维卷积神经网络的 fROC 曲线。

表4 测试集中不同大小、不同类型的肺结节检出个数及灵敏度

检出方法	实性			磨玻璃			胸膜	钙化	汇总
	<5mm	5~10mm	>10mm	<5mm	5~10mm	>10mm			
三维卷积神经网络	237(99)	35(100)	2(100)	109(71)	157(88)	52(88)	39(100)	27(100)	658(90)
Siemens Syngo. via	38(16)*	35(100)	1(50)	21(14)*	99(56)*	40(68)*	9(23)*	20(74)*	263(36)*
Philips ISP	35(15)*	32(91)*	2(100)	16(11)*	84(47)*	49(83.1)	13(33)*	18(67)*	249(34)*

注:表中数据分别表示肺结节检出个数及灵敏度(括号);\*表示在该类型的肺结节检出上,三维卷积神经网络显著优于所比较方法,差异有统计学意义(Z检验, $P<0.05$ )。

针对不同大小、不同类型的肺结节进行分层统计,分别计算三维卷积神经网络、Siemens syngo. via 以及 Philips ISP 对于各类结节检出的灵敏度。对于三维卷积神经网络,<5 mm 的实性结节检出 237 个(99%),5~10 mm 实性结节检出 35 个(100%),>10 mm 实性结节检出 2 个(100%);<5 mm 的磨玻璃结节检出 109 个(71%),5~10 mm 的磨玻璃结节检出 157 个(88%),>10 mm 的磨玻璃结节检出 52 个(88%);胸膜结节检出 39 个(100%);钙化结节检出 27 个(100%)(表 4)。

通过与 syngo. via 以及 Philips ISP 进行对比,可以看出三维卷积神经网络几乎在所有类别的肺结节检出上都有更高的灵敏度,尤其在<5 mm 的实性结节、磨玻璃结节、胸膜和钙化结节上,其灵敏度的优势更为明显。

## 讨论

肺结节早期检出并明确诊断对肺结节的临床管理至关重要。我们前期基于回顾性数据建立了基于三维卷积神经网络算法的肺结节深度学习算法模型,所建立的肺结节检出算法是基于特征金字塔网络的检测框架<sup>[16]</sup>,采用三维卷积操作来替代传统的二维卷积操

作。整个算法的流程图如图 2 所示,通过多个检测网络对胸部 CT 肺结节进行检出。基于深度学习计算的肺结节影像特征,每个检测网络会对图像中的任意一个位置是否存在肺结节进行判断。只有所有检测网络一致认为存在肺结节的位置,才会被算法输出给医生进行审阅。通过结合多个深度学习检测网络,该算法模型可以有效并准确地检出肺结节的位置。通过对模型的准确性进行独立验证,并与传统 CAD 系统(Siemens syngo. via VB 3.0 和 Philips ISP V8)在肺结节检出灵敏度、精准度以及平均每例假阳性个数等多个指标方面进行优效验证,结果显示该模型基于三维卷积神经网络,可以充分反映每枚结节的三维空间信息,较传统 CAD 模型具有较明显的优势。基于三维卷积神经网络算法的肺结节自动检测模型较传统 CAD 肺结节检出系统具有更高的灵敏度,整体肺结节检出灵敏度为 90%,精准度为 71%,平均每例患者假阳性检出 2.8 个,尤其对于<5mm 的肺结节的检出提升效果更为显著。在不考虑结节大小对模型假阳性影响的情况下,对结节假阳性率的控制稍优于 Li 等<sup>[17]</sup>的研究结果,该算法模型在实性结节,胸膜结节及钙化结节检出上有着较高的灵敏度,但在磨玻璃结节检出上灵敏度仍有不足。主要原因是由于磨玻璃结节在训练集中

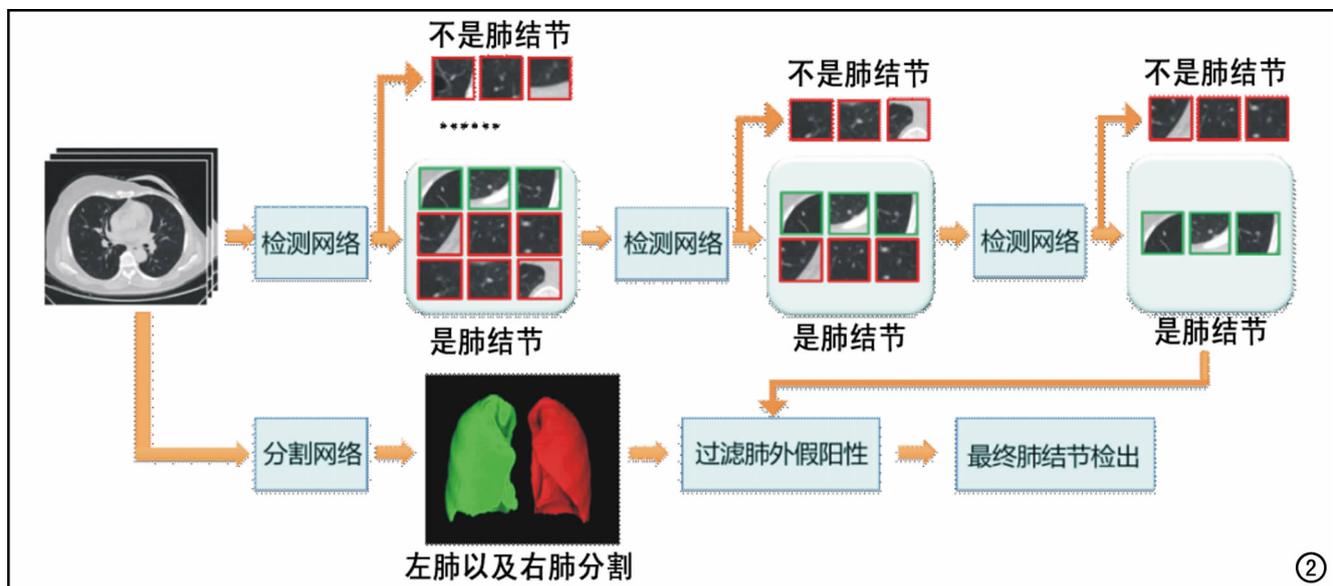


图2 基于级联学习的肺结节辅助筛查流程图,绿色方块表示肺结节,红色方块表示非肺结节。

仅占 7.8%，大量标注结节为实性结节，因此训练得到的算法模型偏向于学习实性结节的影像特征，对实性结节表现灵敏，而对样本量较少的磨玻璃结节则相对不灵敏。

通过对测试集假阳性结节进行回顾性研究，笔者发现导致误诊的主要原因包括扩张支气管内分泌物、迂曲增粗的血管、正常小叶核心结构、马赛克样局限气体滞留及一些不同病因导致的炎性感染病变等。同样对于密度较淡磨玻璃结节、隐藏在血管旁结节、肺门旁结节、气管内结节及胸膜旁结节是主要的漏检原因。对于一款肺癌筛查软件来说，较高的假阳性率在肺结节临床筛查中尚可接受，但假阴性率需尽力避免，因此，本研究模型虽然灵敏度较高，但对于临床应用来说仍有较大的提升空间。

通过对本研究的初级模型的检测结果可以看出，由于深度学习算法需要大量带有肺结节标注的胸部 CT 影像数据来进行训练，数据的多样性以及标注质量直接决定了算法的效能；从而反映出一个深度学习的算法模型，其训练集数据的量和数据的标注质量至关重要，同时训练集的数据分布也直接影响着模型的鲁棒性<sup>[18]</sup>。

本研究有以下局限性：①本研究属于回顾性研究且测试集样本量相对较少，存在选择偏倚；②没有对结节特殊位置进行详细分类；③不同类别的样本量不均衡，会导致算法训练性能下降，这是机器学习中常见的问题<sup>[19]</sup>。下一步我们将建立经过质控达标的标准测试集来验证深度学习模型的效能，并对漏诊结节分析其漏诊原因，有针对性地增加模型训练集分类权重进一步迭代优化。

综上所述，三维卷积神经网络算法模型相较于传统 CAD 系统，在肺结节检出灵敏度上有显著提升，通过对漏诊结节的补充训练迭代，有望在临床上提高医生的阅片效率并降低漏诊率，成为影像医生的助手。当然，这仅仅是用于结节筛查，肺结节鉴别诊断模型有待进一步研发验证。

#### 参考文献：

[1] Jacobs CD, Jafari ME. Early results of lung cancer screening and radiation dose assessment by low-dose CT at a community hospital[J]. Clin Lung Cancer, 2017, 18(5): 327-331.

[2] Torre LA, Siegel RL, Jemal A. Lung cancer statistics[J]. Adv Exp Med Biol, 2016, 893(1): 1-19.

[3] Henschke CI. Early lung cancer action project: overall design and

findings from baseline screening[J]. Cancer, 2000, 89(S11): 2474-2482.

[4] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018[J]. CA Cancer J Clin, 2018, 60(5): 277-300.

[5] 胡琼洁, 陈冲, 王玉锦, 等. 实习医师以共同阅片和第二阅片者模式使用计算机辅助检测系统在低剂量 CT 中的应用研究[J]. 放射学实践, 2018, 33(10): 1022-1028.

[6] Sui Y, Wei Y, Zhao D. Computer-aided lung nodule recognition by SVM classifier based on combination of random undersampling and SMOTE[J]. Comput Math Methods Med, 2015, 12(1): 1-13.

[7] Murphy K, Ginneken BV, Schilham AM, et al. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification[J]. Med Image Anal, 2009, 13(5): 757-770.

[8] Messay T, Hardie RC, Rogers SK. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery[J]. Med Image Anal, 2010, 14(3): 390-406.

[9] 金征宇. 人工智能医学影像应用: 现实与挑战[J]. 放射学实践, 2018, 33(10): 989-991.

[10] Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications[J]. Clin Radiol, 2019 Jun 12. pii: S0009-9260(19)30210-7. DOI: 10.1016/j.crad.2019.04.017.

[11] Setio AAA, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge[J]. Med Image Anal, 2017, 42(12): 1-13.

[12] Gong L, Jiang S, Yang Z, et al. Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks[J]. Int J Comput Assist Radiol Surg, 2019, 26(4): 1-11.

[13] 中国食品药品检定研究院, 中华医学会放射学分会心胸学组. 胸部 CT 肺结节数据标注与质量控制专家共识(2018)[J]. 中华放射学杂志, 2019, 53(1): 9-15.

[14] 邹勤, 刘士远, 管宇, 等. 超高分辨率 CT 对 3cm 以下肺磨玻璃密度结节的诊断价值[J]. 临床放射学杂志, 2017, 36(4): 484-488.

[15] Bandos AI, Rockette HE, Gur SD. Area under the free-response ROC curve (FROC) and a related summary index[J]. Biometrics, 2009, 65(1): 247-256.

[16] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. Las Vegas: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 936-944.

[17] Li X, Guo F, Zhou Z, et al. Performance of deep-learning-based artificial intelligence on detection of pulmonary nodules in chest CT[J]. Chin J Lung Cancer, 2019, 22(6): 336-340.

[18] 萧毅, 刘士远. 肺结节影像人工智能技术现状与思考[J]. 肿瘤影像学, 2018, 27(4): 8-11.

[19] Japkowicz N, Stephen S. The class imbalance problem: A systematic study[J]. Intell Data Anal, 2002, 6(5): 429-449

(收稿日期: 2019-07-29 修回日期: 2019-08-11)

## 基于术前分期 CT 的影像组学标签预测三阴性乳腺癌

张文, 何兰, 范志豪, 黄晓媚, 杨晓君, 梁长虹, 刘再毅

**【摘要】** 目的:探讨基于术前分期 CT 的影像组学标签在预测三阴性乳腺癌分子分型中的附加价值。方法:回顾性收集 2016 年 1 月至 2018 年 5 月经手术病理证实且均为临床术前评估分期需行常规胸部 CT 增强扫描的 481 例肿块型乳腺浸润性癌患者,按照样本量 1:2 随机抽样选取三阴性乳腺癌与非三阴性乳腺癌共计 150 例患者(90 例作为训练组,60 例作为验证组)。所有患者均经免疫组织化学检测,获得乳腺癌分子分型。对所有患者基于病灶三维图像提取影像组学特征,并采用 Lasso logistic 回归模型进行特征降维及筛选,以建立影像组学标签。采用 ROC 曲线评价影像组学标签对三阴性乳腺癌的鉴别诊断效能。结果:由 5 个关键影像组学特征构成的影像组学标签与乳腺癌三阴性分子分型相关( $P < 0.0001$ )。建立的影像组学标签对于鉴别三阴性乳腺癌具有较好的预测效能,其在训练组和验证组的 ROC 曲线下面积(AUC)分别为 0.766(95% CI:0.743~0.789)和 0.758(95% CI:0.718~0.798)。结论:基于术前分期 CT 建立的影像组学标签有助于三阴性与非三阴性乳腺癌的鉴别,这是术前常规胸部增强 CT 扫描在辅助临床分期之外的附加价值,可为临床治疗决策提供参考。

**【关键词】** 乳腺肿瘤; 三阴性乳腺癌; 影像组学; 体层摄影术, X 线计算机; 分子分型; 预测

**【中图分类号】** R737.9; R814.42 **【文献标识码】** A **【文章编号】** 1000-0313(2019)09-0947-05

DOI:10.13609/j.cnki.1000-0313.2019.09.003

开放科学(资源服务)标识码(OSID):



**Preoperative staging CT-based radiomics signature in predicting triple-negative breast cancer** ZHANG Wen, HE Lan, FAN Zhi-hao, et al. The Second School of Clinical Medicine, Southern Medical University, Guangzhou 510515, China

**【Abstract】 Objective:** To explore the additional value of preoperative staging CT-based radiomics signature in predicting molecular subtype of triple-negative breast cancer. **Methods:** From January 2016 to May 2018, 481 patients with pathological confirmed invasive breast cancer (mass type) were enrolled in the study. All patients were performed routine staging enhanced chest CT before surgery. A total of 150 cases were randomly selected according to the ratio of 1:2 (triple-negative vs non-triple-negative breast cancer). Among the 150 cases, 90 cases were divided into the training dataset and 60 cases were divided into the validation dataset. All patients obtained detailed information about molecular subtypes of breast cancer by immunohistochemistry. The radiomic features were extracted based on volume of the interest in all patients and Lasso logistic regression model was used for dimensionality reduction, feature selection and construction of the radiomics signature. The receiver operation curve (ROC) was used to evaluate the performance of the radiomics signature in predicting triple-negative breast cancer. **Results:** The radiomics signature composed of 5 radiomic features was associated with the triple-negative molecular subtype of breast cancer ( $P < 0.0001$ ). The constructed radiomics signature obtained good predictive performance in identifying triple-negative breast cancer, with the area under the ROC curve (AUC) of 0.766 (95% CI:0.743~0.789) and 0.758 (95% CI:0.718~0.798) in the training and validation dataset, respectively. **Conclusion:** The constructed radiomics signature based on preoperative staging CT could be helpful in differentiating between triple-negative and non-triple-negative breast cancer, which can provide additional value of routine preoperative staging CT for clinical

作者单位:510515 广州,南方医科大学第二临床医学院(张文、刘再毅);510080 广州,广东省人民医院放射科 广东省医学科学院(张文、何兰、范志豪、黄晓媚、杨晓君、梁长虹、刘再毅)

作者简介:张文(1988-),男,广东珠海人,硕士研究生,主治医师,主要从事胸腹部影像诊断工作。

通讯作者:刘再毅, E-mail: zyliu@163.com

基金项目:国家重点研发计划(2017YFC1309100);广东省省级科技计划项目(2017B020227012)

cal treatment decisions.

**【Key words】** Breast neoplasms; triple negative breast cancer; Radiomics; Tomography, X-ray computed; Molecular subtypes; Prediction

乳腺癌是女性最常见的恶性肿瘤,也是女性恶性肿瘤中病死率最高的疾病<sup>[1]</sup>。研究表明,决定早期乳腺癌治疗方案、提示疗效与预后的并非组织病理学类型,而是其分子分型<sup>[2,3]</sup>。三阴性乳腺癌(triple negative breast cancer, TNBC)占乳腺癌的 10%~20%<sup>[4]</sup>,是一种异质性疾病,其生物学特征不同于其他类型的乳腺癌,缺乏分子靶向治疗的有效靶点且预后不良,目前化疗仍是临床上 TNBC 唯一有效的治疗方案<sup>[5]</sup>。因此,早期辨别 TNBC 与非三阴性乳腺癌(non triple negative breast cancer, NTNBC)对取得最佳诊疗方案具有重要临床意义。目前临床主要依靠免疫组织化学明确乳腺癌的分子分型,但由于对肿瘤组织进行活检取样是一种具有侵入性的方法,同时取样和分析时存在局限性<sup>[6]</sup>。相比之下,影像检查能够无创且从整体上反映肿瘤特征,一方面在基因水平上对每个亚型内的差异进一步分析,另一方面能够多次动态评价治疗效果<sup>[7]</sup>。

根据美国国立综合癌症网络(National Comprehensive Cancer Network, NCCN)指南,胸部 CT 增强扫描已被列入乳腺癌患者术前常规检查之一。不同于乳腺 X 线摄影、超声及磁共振检查对乳腺病灶的直观评估,胸部 CT 增强扫描的主要目的是用于辅助临床分期。通常认为 CT 检查对微小钙化的显示不及乳腺 X 线摄影,对乳腺囊实性病变的诊断准确性不及超声,对乳腺良、恶性病变的鉴别也无显著特异性。但随着影像组学的发展<sup>[8,9]</sup>,我们假设这种新的手段可以在乳腺癌患者术前分期 CT 中提取出以往肉眼无法直接观测到的信息,而这些信息有可能与肿瘤的分子分型、生物学特性等存在一定的关系,从而使术前常规 CT 扫描在辅助临床分期之外,提供对病灶特征评估有益的附加信息,目前国内外尚未见相关报道。

本研究拟基于影像组学的方法,在不额外增加患者辐射剂量和经济负担的前提下,进一步挖掘乳腺癌术前常规 CT 图像数据中蕴含的信息,旨在探讨术前分期 CT 在预测三阴性乳腺癌分子分型中的附加价值。

## 材料与方法

### 1. 研究对象

本研究经医院伦理委员会批准,并免除知情同意要求。回顾性收集 2016 年 6 月至 2018 年 1 月经我院病理组织活检或手术确诊的乳腺癌患者。病例纳入标准:①经手术和/或活检组织病理学证实的肿块型乳腺

浸润性癌;②具备用于术前评估乳腺癌临床分期且未经治疗的胸部增强 CT 扫描图像;③免疫组织化学资料完整,可获得分子分型。病例排除标准:①图像质量不佳,影响病灶观察及勾画;②相关临床资料或病理信息不完整;③特殊类型的乳腺癌。

### 2. 免疫组化获得分子分型

所有患者均进行了乳腺肿瘤活检或手术,获得的病理标本均进行免疫组织化学检测,包括雌激素受体(estrogen receptor, ER)、孕激素受体(progesterone receptor, PR)、人表皮生长因子受体 2(human epidermal growth factor receptor2, HER2)和 Ki-67。将乳腺癌分为以下 4 型:Luminal A 型、Luminal B 型、HER2 过表达型和三阴性(triple negative, TN)<sup>[10]</sup>。

### 3. CT 扫描方案

采用 Brilliance iCT 256 (Philips Healthcare, Cleveland OH, USA)、Somatom Definition Flash (Siemens Healthcare, Forchheim, Germany)、Philips Ingenuity CT (Philips Healthcare, Cleveland, Ohio, USA)及 LightSpeed VCT (GE Medical systems, Milwaukee, Wis, USA)扫描仪进行术前胸部 CT 增强扫描。CT 扫描参数:管电压 120 kVp,管电流 150 mAs; Brilliance iCT 256、Somatom Definition Flash 和 Philips Ingenuity CT 旋转时间为 0.5 s, LightSpeed VCT 旋转时间为 0.4 s;探测器准直分别为 128×0.625 mm、2×64×0.6 mm、64×0.625 mm 及 64×0.625 mm;视野为 350 mm×350 mm,矩阵为 512×512,重建层厚为 1.0 mm 或 1.25 mm。CT 扫描方案:首先行常规 CT 平扫,采集平扫 CT 图像;其次行增强 CT 扫描,采用高压注射器(Ulrich CT Plus 150, Ulrich Medical, Ulm, Germany)经静脉注射非离子型对比剂优维显 80~100 mL(370 mg I/mL),流率 4 mL/s,分别于注射对比剂后 18 s、35 s 采集肺动脉期、主动脉期 CT 图像。

### 4. 影像组学特征提取

采用基于 MATLAB 2014a (Mathworks, Natick, MA, USA)软件进行影像组学分析。兴趣区(region of interest, ROI)的放置由两位放射科医生(分别具有 3 年和 7 年的乳腺影像诊断经验)使用 itk-SNAP 软件(<http://www.itksnap.org/pmwiki/pmwiki.php>)完成。选取主动脉期图像,剔除病灶显示不清或非肿块型病灶,沿肿块病灶边缘逐层手动勾画 ROI,ROI 应尽量包括整个病灶,注意避开病灶内坏死、钙化及气体

影等,最终得到整个瘤体感兴趣体积(volume of interest, VOI)。第一位医生(医生1)放置 ROI 并进行影像组学特征提取一次,1周后再进行第二次 ROI 放置及特征提取,用于测量者自身一致性的评估;另一位医生(医生2)仅进行 ROI 放置及特征提取一次,与医生1的第一次测量进行测量者间的一致性评估。

所提取的影像组学特征主要分为三类:①一阶统计特征;②形态学特征;③纹理特征。每例患者的术前 CT 图像共提取 182 个影像组学特征(表 1)。

表 1 提取的影像组学特征

特征类型	特征名称
一阶灰度直方图特征	能量、熵、最小值、最大值、平均值、中位数、平均绝对偏差、均方差、标准偏差、偏度、峰度、方差、均匀度
形态学特征	表面积、周长、凹度、体素量、最大 3D 直径、球度
灰度纹理特征	灰度共生矩阵 GLCM、灰度游程矩阵 GLRLM、灰度区域大小矩阵 GLSZM、灰度差异矩阵 GLDM、邻域灰度差异矩阵(NGTDM)

## 5. 统计学分析

患者临床病理基线资料的比较:对患者临床病理基线资料训练组与验证组间的差异性比较依据数据特征采用卡方检验(年龄)、独立样本  $t$  检验或 Mann-Whitney U 检验进行统计学分析。

观察者自身及之间的一致性评估:采用组内和组间相关系数(intra and interclass correlation coefficients, ICCs)对医生 1 自身和两位测量者间的一致性进行评价。采用医生 1 的两次测量数据计算测量者自身的 ICC;采用医生 1 的第一次测量结果与医生 2 的测量结果计算测量者间的 ICC。ICC>0.75 认为其一致性好。

特征筛选和影像组学标签的建立:采用 Lasso-logistic 回归模型筛选基于动脉期 VOI 提取的对乳腺癌三阴性状态具有预测意义的影像组学特征。通过选取的特征与对应加权系数乘积的线性组合依次形成每例患者的影像组学标签,并计算相应值。

影像组学标签预测效能的评价:应用受试者工作特征(receiver operator characteristic, ROC)曲线的曲线下面积(area under curve, AUC)评估建立的影像组学标签对乳腺癌三阴性状态的鉴别预测效能,AUC 越大证明其诊断效能越高。计算相应的 ROC 曲线下面积(AUC)、敏感度、特异度,均包括 95% 置信区间的计算。以上分析的界值(cut-off points)以最大约登指数(youden index)(敏感度+特异度-1)确定。

采用 R 软件(版本:3.0.1; <http://www.Rproject.org>)进行统计学分析。加载相应的程序包或函数完成相应的统计检验。以  $P<0.05$  为差异有统计学意义。

## 结果

### 1. 人口学及临床特征

共纳入 481 例肿块型乳腺浸润性癌病例形成研究队列,其中 TNBC 50 例,根据病例号,采用随机数字法,按照样本量 1:2 随机抽样选取三阴性乳腺癌与非三阴性乳腺癌共计 150 例患者(90 例作为训练组,60 例作为验证组)进行分析,患者均为女性。TNBC 患者的年龄为 23~81 岁,中位年龄 49 岁;NTNBC 患者的年龄为 26~73 岁,中位年龄 49 岁,两组患者的年龄差异无统计学意义( $P>0.05$ ,表 2)。

表 2 乳腺癌患者的临床资料及病理结果比较

临床特征	训练组	验证组	$t/\chi^2$ 值	$P$ 值
年龄(岁)	49.93±11.90	49.95±11.05	-0.009	0.993
月经状态(n,%)			1.394	0.238
绝经前	61(67.8)	35(58.3)		
绝经后	29(32.2)	25(41.7)		
病理分级(n,%)			4.197	0.123
I	8(8.9)	1(1.7)		
II	44(48.9)	27(45)		
III	38(42.2)	32(53.3)		
分子分型(n,%)			1.227	0.754
Luminal A 型	11(12.2)	6(10)		
Luminal B 型	30(33.3)	17(28.3)		
HER2 过表达型	19(21.2)	17(28.3)		
三阴性(TNBC)	30(33.3)	20(33.4)		

### 2. 观察者自身及之间的一致性评估

基于医生 1 与医生 2 提取的特征计算得出测量者间的一致性较好(ICC 为 0.785~0.998),基于医生 1 两次提取的特征计算得出测量者内的一致性较好(ICC 为 0.812~0.998)。因此,下述分析均基于医生 1 第一次提取的特征。

### 3. 影像组学标签的建立

通过 Lasso-Logistic 回归模型共筛选出具有非零系数的 5 个影像组学特征(图 1)。基于这些特征与之相应的系数乘积的线性组合构建影像组学标签(公式 1)。训练组中, TNBC 组患者的影像组学标签值为 -0.496(四分位数间距 -0.707, -0.222),高于 NTNBC 组患者的 -0.864(四分位数间距 -1.031, -0.634),差异有统计学意义( $P<0.001$ );验证组中, TNBC 组和 NTNBC 组患者的影像组学标签值分别为 -0.603(四分位数间距 -0.902, -0.346)和 -0.899(四分位数间距 -1.010, -0.768),差异也具有统计学意义( $P<0.001$ )。

$$\text{影像组学标签值} = -2.61548 + 0.010178 \times \text{compactness1} - 50.3070 \times \text{GLZM\_ZP} + 0.11981 \times \text{GLSZM\_SAHGLE} + 0.26238 \times \text{band\_max} + 1.43099 \times \text{band\_mean} \quad (1)$$

### 4. 预测效能评估

本研究建立的影像组学标签对乳腺癌分子分型三阴性具有较好的预测效能。训练组中的 AUC 为 0.766(95% CI:0.743~0.789),敏感度为 0.667,特异

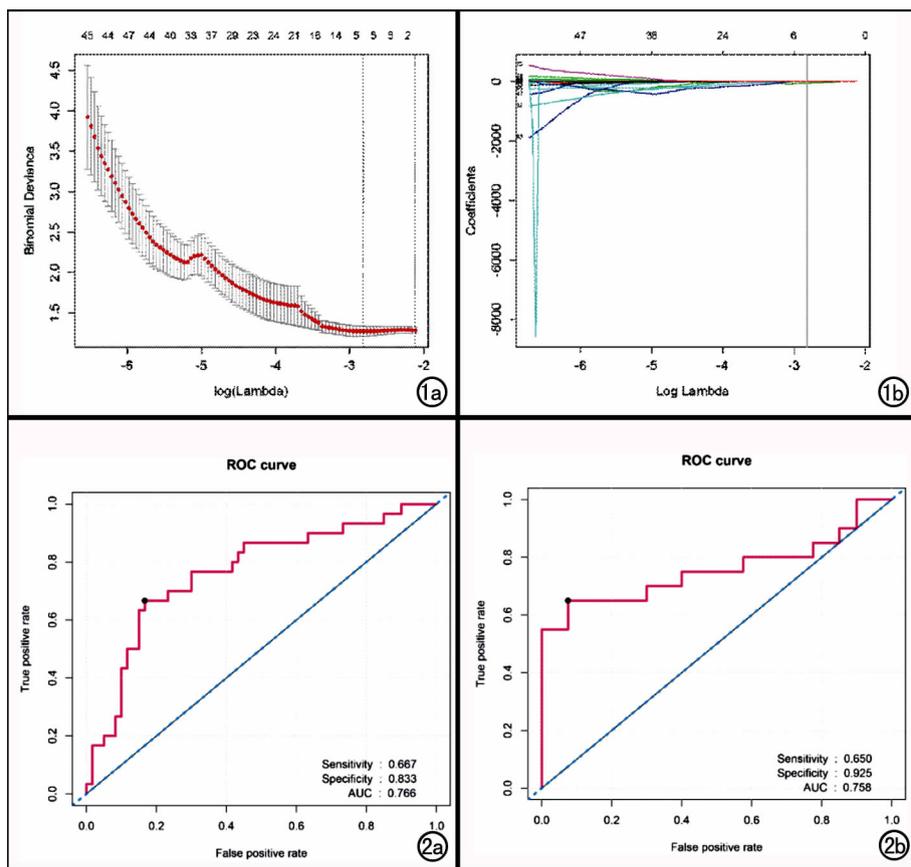


图1 采用 Lasso-Logistic 回归模型对影像组学特征进行筛选。a) 纵坐标为二项式偏差(binomial deviance), 横坐标为  $\log(\lambda)$ 。模型中使用十倍交叉验证方法, 通过调节不同参数  $\lambda$  实现模型的二项式偏差最小, 从而筛选出与三阴性乳腺癌相关联的特征集合。本研究共筛选出 5 个特征; b) 为特征筛选过程的系数收敛图。纵坐标表示特征在模型中的各自系数, 横坐标为  $\log(\lambda)$ 。垂直线对应采用十折交叉验证后拟合筛选出的非零特征, 共筛选出 5 个特征。图2 影像组学标签对乳腺癌三阴性状态预测的 ROC 曲线。a) 训练组的 ROC 曲线; b) 验证组的 ROC 曲线。

度为 0.833, 阳性预测值为 0.655, 阴性预测值为 0.820; 在验证组中的 AUC 为 0.758 (95% CI: 0.718~0.798), 敏感度为 0.65, 特异度为 0.925, 阳性预测值为 0.800, 阴性预测值为 0.822(图 2)。

## 讨论

通过本次研究, 我们首次探讨了基于术前分期 CT 的影像组学标签在术前鉴别 TNBC 与 NTNBC 的价值, 可以为临床治疗决策提供参考信息。本研究回顾性分析的图像, 均为我院已有的用于乳腺癌术前分期的胸部增强 CT 图像, 并没有额外增加患者的辐射剂量与经济负担, 这对于术前分期 CT 获得的主观特征, 影像组学方法对图像信息的挖掘, 具有附加的临床价值。

目前, 乳腺癌临床治疗方案的制定主要依据分子

分型, TNBC 的生物学特征不同于其他类型的乳腺癌, 相同治疗方案的 TNBC 患者其反应及预后也不尽相同, 其原因是由于肿瘤的异质性<sup>[11]</sup>。由于肿瘤生长过程中时间和空间存在差异性, 病理活检有时并不能代表完全的肿瘤组织特点, 而影像组学标签由图像中提取得到的量化特征构成, 与肿瘤病灶的主观定性特征相比, 可以更全面、无创地定量描述肿瘤异质性<sup>[12-14]</sup>。本研究联合了 5 个 CT 图像特征构建影像组学标签, 作为预测乳腺癌分子分型的影像生物标志, 成功将 TNBC 与 NTNBC 患者进行分层 ( $P < 0.0001$ )。此外, 本研究提取得到大量影像组学特征, 在解决高通量计算中普遍面临的预测因子筛选问题中, 采用了一种惩罚估计技术——Lasso logistic 回归模型, 以实现变量筛选, 最终得到了由 5 个影像组学特征组成的影像组学标签, 显示了较好的预测效能。在此之前, 相关学者基于乳腺 X 线摄影、磁共振的影像组学鉴别 TNBC 与 NTNBC 做了较多研究<sup>[15-17]</sup>, 但都没有验证。本研究通过将训练组得到的预判能力阈值用于验证组进行验证, 使实验结果更可靠。同时, 我们发现用于术前评估患者临床分期的 CT 图像, 可以通过影像组学的方法进一步挖掘并获得图像特征, 这些特征构建的标签在预测三阴性乳腺癌中有较高的特异度(训练组为 0.833, 验证组为 0.925)。对比乳腺癌其他分子分型, TNBC 对靶向治疗及内分泌治疗效果差, 对化疗敏感性较高, 而化疗具有潜在的毒性, 为了使化疗得到合理的运用, 降低潜在风险, 影像组学标签的高特异度可能会为乳腺癌个体化治疗带来额外的价值。分期 CT 属于临床常规检查, 不会增加患者的辐射剂量和经济负担, 可为临床诊疗提供额外的信息, 因此具有一定的临床应用价值。

本研究有一定的局限性: 首先, 这是对单个机构获得的图像进行回顾性分析。在后续研究中, 需增大样本量, 利用多中心数据对本研究结果进行验证; 其次, 本研究基于临床分期 CT 进行数据挖掘, 未针对乳腺

病灶进行数据采集优化。

综上所述,基于术前分期 CT 建立的影像组学标签有助于乳腺癌分子分型及 TNBC 与 NTNBC 的鉴别,这是术前常规胸部增强 CT 扫描在辅助临床分期之外的附加临床价值,可为临床治疗决策提供参考。

#### 参考文献:

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018; GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA Cancer J Clin*, 2018, 68(6): 394-424.
- [2] Harbeck N, Gnant M. Breast cancer [J]. *Lancet*, 2017, 389(10074): 1134-1150.
- [3] Arvold ND, Taghian AG, Niemierko A, et al. Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy[J]. *J Clin Oncol*, 2011, 29(29): 3885-3891.
- [4] Reis-Filho JS, Tutt AN. Triple negative tumours: a critical review [J]. *Histopathology*, 2008, 52(1): 108-118.
- [5] Denkert C, Liedtke C, Tutt A, et al. Molecular alterations in triple-negative breast cancer—the road to new treatment strategies [J]. *Lancet*, 2017, 389(10087): 2430-2442.
- [6] Guo Y, Hu Y, Qiao M, et al. Radiomics analysis on ultrasound for prediction of biologic behavior in breast invasive ductal carcinoma [J]. *Clin Breast Cancer*, 2018, 18(3): 335-344.
- [7] 高微波, 朱海涛, 孙应实. 乳腺癌影像基因组学研究现状与进展 [J]. *中华放射学杂志*, 2017, 51(12): 990-992.
- [8] Valdora F, Houssami N, Rossi F, et al. Rapid review: radiomics and breast cancer [J]. *Breast Cancer Res Treat*, 2018, 169(2): 217-229.
- [9] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data [J]. *Radiology*, 2016, 278(2): 563-577.
- [10] Park S, Koo JS, Kim MS, et al. Characteristics and outcomes according to molecular subtypes of breast cancer as classified by a panel of four biomarkers using immunohistochemistry [J]. *Breast*, 2012, 21(1): 50-57.
- [11] 马文娟, 季宇, 郝玉娟, 等. 乳腺癌 X 线摄影的影像组学特征 [J]. *中国医学影像学杂志*, 2018, 26(7): 515-518.
- [12] 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.
- [13] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach [J]. *Nat Commun*, 2014, 6(5): 4006.
- [14] Liang C, Huang Y, He L, et al. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I ~ II and stage III ~ IV colorectal cancer [J]. *Oncotarget*, 2016, 7(21): 31401-31412.
- [15] Agner SC, Rosen MA, Englander S, et al. Computerized image analysis for identifying triple-negative breast cancers and differentiating them from other molecular subtypes of breast cancer on dynamic contrast-enhanced MR images: a feasibility study [J]. *Radiology*, 2014, 272(1): 91-99.
- [16] Ma W, Zhao Y, Ji Y, et al. Breast cancer molecular subtype prediction by mammographic radiomic features [J]. *Acad Radiol*, 2019, 26(2): 196-201.
- [17] Wang J, Kato F, Oyama-Manabe N, et al. Identifying triple-negative breast cancer using background parenchymal enhancement heterogeneity on dynamic contrast-enhanced MRI: A pilot radiomics study [J]. *PLoS One*, 2015, 10(11): e0143308.

(收稿日期: 2019-07-24 修回日期: 2019-08-16)

## · 人工智能影像学专题 ·

## 基于深度残差网络研发辅助诊断软件用于 X 线胸片分类诊断

张晓东, 孙兆男, 任昕, 周宇, 周雯, 李建辉, 谢辉辉, 刘婧, 张虽虽, 李津书, 王霄英

**【摘要】 目的:**研究以深度残差网络(ResNet)为基础架构建立深度学习模型,对 X 线胸片(CXR)做出“有发现”与“无发现”鉴别诊断的可行性。**方法:**回顾性收集 2017 年 1 月 1 日至 2018 年 7 月 1 日的连续 CXR 图像及诊断报告,经过数据清洗后分为“无发现”组(无任何异常发现,诊断印象为“两肺心膈未见异常”,共 9765 例)与“有发现”组(诊断印象中提及了一种以上影像所见,共 9956 例)。使用 ResNet152(152 layers)作为二分类模型的基础架构,结合 Grad-CAM 技术生成模型激活热图,训练二分类模型。数据随机分为训练集(70%)、调优集(20%)和测试集(10%)。以测试集的预测结果检测 CXR 二分类模型的效能。**结果:**在测试集中(“有发现”者 1018 例,“无发现”者 995 例),CXR 二分类模型鉴别“有发现”与“无发现”的精确度分别为 0.885 和 0.894,召回率分别为 0.898 和 0.880,F1-分数分别为 0.891 和 0.887,ROC 曲线下面积均为 0.96。**结论:**使用 CXR 二分类模型可对 X 线胸片做出“无发现”与“有发现”的预测。

**【关键词】** 深度学习;人工智能;深度残差网络;胸部 X 线片;用例;结构化报告

**【中图分类号】** TP18;R814.41 **【文献标识码】** A **【文章编号】** 1000-0313(2019)09-0952-06

DOI:10.13609/j.cnki.1000-0313.2019.09.004

开放科学(资源服务)标识码(OSID):



### Development of a ResNet-based CADx software for classification diagnosis in chest X-ray images

ZHANG Xiao-dong, SUN Zhao-nan, REN Xin, et al. Department of Radiology, Peking University First Hospital, Beijing 100034, China

**【Abstract】 Objective:** To evaluate the feasibility of residual deep neural network (ResNet) based deep learning model in the differentiation between "no significant finding" and "any significant finding" cases in chest X-ray (CXR) images. **Methods:** From January 1 2017 to July 1 2018, a consecutive cohort were retrospectively collected in this study. All the images were divided into the "no significant finding" group (defined as no significant finding was detected and reported) and the "any significant finding" group (defined as any significant findings that were detected or reported). Finally, 9765 cases were collected in "no significant finding" group and 9956 cases were collected in "any significant finding" group. The ResNet 152 (152 layers) was used as the framework of the binary classification model and the Grad-CAM (gradient class activation maps) was used to generate model activation heating maps. The cohort was randomly divided into training (70%), validation (20%), and testing dataset (10%). The results obtained in testing dataset was considered as the performance of the model. **Results:** In the testing dataset, 1018 cases were classified as "any significant finding", while 995 cases were classified as "no significant finding". The precision, recall and F1-scores were 0.885 vs 0.894, 0.898 vs 0.880 and 0.891 vs 0.887, respectively in terms of two group (any significant finding vs no significant finding). The areas under ROC (AUC) were both 0.96. **Conclusion:** The CXR binary classification model could be useful in the prediction of "no significant finding" and "any significant finding" for CXR images.

**【Key words】** Deep learning; Artificial intelligence; residual deep neural network; Chest X-ray; Use case; Structured report

作者单位:100034 北京,北京大学第一医院医学影像科(张晓东,孙兆男,任昕,周宇,周雯,李建辉,谢辉辉,刘婧,李津书,王霄英);100011 北京,北京赛迈特锐医学科技有限公司(张虽虽)

作者简介:张晓东(1979-),男,山西人,博士,副主任技师,主要从事医学图像处理工作。

通讯作者:王霄英, E-mail: cjr.wangxiaoying@vip.163.com

胸部 X 线片 (chest X-ray, CXR) 是诊断呼吸系统疾病的重要影像检查方法, 在医学影像科日常工作中占很大比例。使用 CXR 进行检查的临床场景包括门诊检查、术前常规检查、健康体检、急诊检查等<sup>[1]</sup>; 在前三种情况下, CXR 结果多数为正常, 影像检查的任务是检出少数的异常者, 再做进一步处理 (如行胸部 CT 检查等)。近年来人工智能 (artificial intelligence, AI) 领域的技术开始应用于 CXR 的质控、图像分割和诊断等工作<sup>[2-6]</sup>, 初步结果认为在适应证明确的情况下, AI 对完成影像检查任务是有帮助的。本研究拟使用深度学习方法研发影像辅助诊断工具, 对门诊、术前常规和体检人群的 CXR 进行“无发现”、“有发现”的分类 (以下简称二分类), 以提高医学影像科医师的工作效率。

## 材料与方 法

本研究获得了本院伦理审查委员会的批准, 按照本单位 AI 项目研发规范开展研究工作。

### 1. 用例定义

根据本单位 AI 项目管理方法, 首先定义研发 CXR 二分类 AI 模型的用例 (Use Case)<sup>[7]</sup>, 包括以下内容: CXR 二分类 AI 模型的 ID、临床问题、场景描述、模型在实际工作中的调用流程、模型输入输出数据结构等。AI 模型返回结果定义为“无发现”和“有发现”两类。“无发现”的定义: CXR 上未见应写入报告的影像发现, 诊断印象为两肺心膈未见异常。“有发现”的定义: CXR 上有需要在报告中提及的任何影像

所见, 包括 (但不限于) 胸壁、肺、纵隔、横膈、上腹部的影像所见, 诊断印象中提及了一种以上的影像所见。

### 2. 研究队列建立

回顾性收集 CXR 数据建立二分类模型研发队列。从本院 PACS 系统中检索 2017 年 7 月 1 日至 2018 年 7 月 1 日的连续资料, 由 2 位住院医师根据登记信息、设备信息、影像报告和 CXR 图像进行筛选, 再由 1 位高年资医师再次读片, 对图像进行最终分类 (表 1)。

经过数据筛选建立研究队列, 获得“无发现”患者的 CXR 图像共 9765 幅, “有发现”患者的 CXR 图像共 9956 幅。以患者为单位随机分为训练集 (training dataset, 70%)、调优集 (validation dataset, 20%) 和测试集 (testing dataset, 10%), 各个数据集中的患者信息是不同的, 具体见表 2。

表 2 CXR 二分类模型训练过程中的数据分配 (例)

分类	训练集	调优集	测试集	合计
有发现	6903	2035	1018	9956
无发现	6795	1975	995	9765

### 3. 数据处理

CXR 二分类模型输入图像大小为  $512 \times 512$ , 模型输出为“无发现”、“有发现”二分类概率, 并可通过 Grad-CAM 方法得到该模型的激活热图。

模型训练的初始权重来源于 ImageNet 预训练 ResNet152 模型<sup>[8]</sup>。图像预处理部分包括以下步骤: 原始图像大小缩放至  $512 \times 512$ , 并对图像像素值按照 ImageNet 数据集的均值和方差进行归一化处理。扩

表 1 CXR 二分类模型研发数据筛选方法

筛选依据	条件	分类*	合格值	不合格值
登记信息	检查类型	0;1	X 线	其他
	检查项目		胸部正位	胸片侧位、胸片正侧位、床旁胸片等
	年龄		>18 岁	≤17 岁
设备信息	图像数量	0;1	1 张	0 张, ≥2 张
	诊断报告		具备	不具备
	检查设备		CR; DR	其他
报告信息	影像表现	0	“肺野”、“纵隔”、“横膈”、“胸壁”、“下颈部”、“上腹部”正常	术后改变、乳腺术后、支架术后、起搏器术后; PICC、IVAP 等置入物; 膈抬高; 叉状肋、脊柱侧弯、颈 7 横突过长、胸廓萎陷; 陈旧肋骨骨折; 甲状腺增大; 间位结肠、胆囊结石等
			1	胸部疾病 (结核、结节、占位等); 胸部病变 (渗出、实变、纤维索条影、肺气肿、肺不张、钙化、陈旧病变等); 纵隔形态异常; 胸膜病变, 胸壁病变等
	诊断印象 (报告)	0	双肺心膈未见异常、未见明显异常、未见明确异常、无异常; 两肺/心肺未见异常、未见明显异常、未见明确异常、无异常; 肺心膈未见异常; 未见## 疾病, 且影像表现中“双肺野清晰”等	/
		1	不符合 0 者均入选	/
CXR 图像	投照体位	0;1	正确	不正确
	曝光条件		适当	不适当
	体表金属		无或少量	明显
	诊断正确		是	否

注: \* 0=无发现; 1=有发现

增方法:随机旋转、缩放、平移等。

#### 4. 模型训练

在 NVIDIA GPU (P100, Nvidia Corporation, Santa Clara, CA) 上利用 Pytorch (<https://pytorch.org/>) 深度学习 Python 库, 在本院收集整理临床患者 CXR 图像数据集上进行开发和训练。

二分类模型基于 152 层的卷积神经网络模型, 以 Resnet 为卷积神经网络 (convolutional neural network, CNN) 深度学习模型的基础架构<sup>[9]</sup>, 结合梯度类激活图 (gradient class activation maps, Grad-CAM) 方法<sup>[10]</sup>, 建立可以同时显示分类结果及激活热图的 CXR 二分类模型 (图 1)。

训练和调优分别在训练集和调优集数据上进行, 训练好的模型在测试集上验证。用测试集的混淆矩阵 (confusion matrix) 评价二分类模型的效能, 得到将 CXR 分为“无发现”与“有发现”的精确度 (precision)、召回率 (recall)、F1-分数 (F1-score) 及受试者工作特征 (receiver operating characteristic, ROC) 曲线的曲线下面积。将 CXR 二分类模型预测结果返回到结构化报告中, 以保证 AI 模型在临床工作流程中的可用性。

## 结果

CXR 二分类模型对测试集数据预测的混淆矩阵 (图 2) 显示, 分类的精确度、召回率和 F1-分数均较高

(表 3), ROC 曲线下面积均为 0.96 (图 3)。在测试集 1018 个“有发现”的病例中, 模型正确识别了其中 914 个, 有 104 个未被识别出来; 995 个“无发现”的病例中, 模型正确识别了其中 876 个, 其中 119 个被误认为“有发现”。以“有发现”为阳性, 模型的假阴性率为 10.6%, 假阳性率为 11.5%。经与影像专家的读片结果进行对照, 发现假阴性者的主要“发现”种类为: 少量纤维索条影、小范围钙化、心脏轮廓稍饱满、主动脉结轻度突出、肺门影稍重、肺纹理轻度增多、肋骨变异等。

表 3 CXR 二分类模型在测试集中的分类效能

分类	例数	精确度	召回率	F1-分数
有发现	1018	0.885	0.898	0.891
无发现	995	0.894	0.880	0.887

CXR 二分类模型返回了激活图 (图 4~6), 可以看到图像中的病灶区域与预测为“有发现”的激活区域基本吻合, 而如果将该图像预测为正常的话, 其激活高亮区域大部分散在分布, 与解剖学上的特征区域没有相关性。

根据用例设计要求, 将 CXR 二分类模型结果直接返回到结构化报告中, 生成相应的结果。对分类是“无发现”的, 直接生成“正常”报告, 包括常规影像表现和诊断印象, 不返回激活图。对分类是“有发现”的, 激活“异常”控件, 并返回激活图 (图 7)。

## 讨论

胸部 X 线检查是呼吸系统疾病重要的影像检查方法之一, 是医学影像科工作中最常见的临床任务。在综合性医院, 大多数常规体检和呼吸系统疾病筛查均首选 CXR<sup>[11]</sup>, 且在许多诊断任务中相当多的影像诊断结果是“无发现”。使用辅助诊断软件将“有发现”与“无发现”分开, 可提高医生的工作效率, 有利于工作流程的优化。值得注意的是, 本研究中此模型召回率和 F1 值不足 0.9, 将面临着 10% 左右的漏诊率和误诊率, 尤其是漏诊将给临床应用价值带来较大影响。经与影像专家的读片结果进行对比, 发现漏诊的情况主要集中于轻微影像改变, 对临床结局影响不大, 所以仍需在工作中持续改进。当前的改进方法是在实际工作中发现的预测错误的数据持续性收集、整理、标注, 用于模型

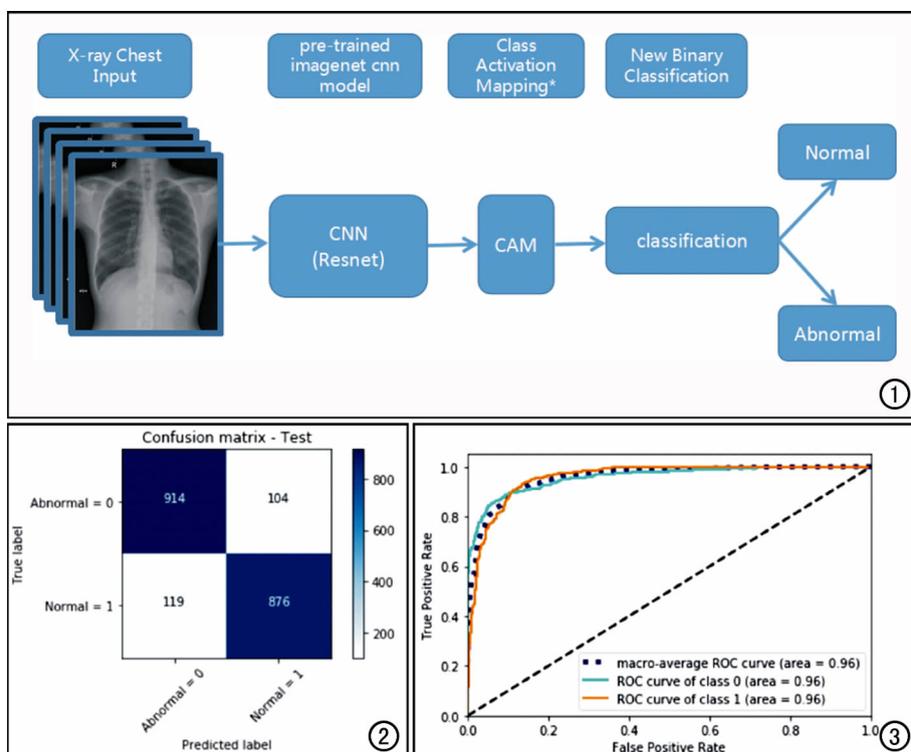


图 1 基于 ResNet152 基础架构并结合 Grad-CAM 的 CXR 二分类模型训练过程。图 2 CXR 二分类模型在测试集中的混淆矩阵。图 3 CXR 二分类模型在测试集中的 ROC 曲线。

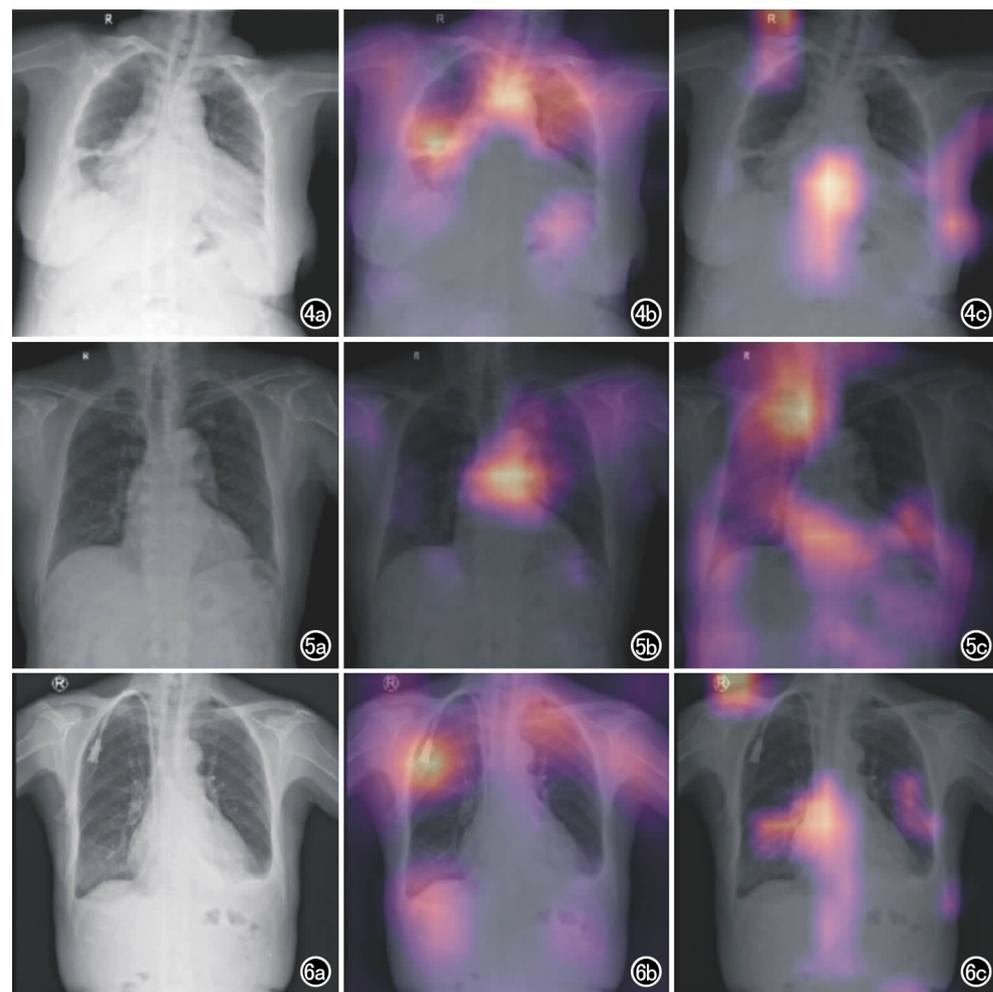


图4 心衰合并右肺病变及胸腔积液,CXR二分类预测结果为“有发现”。a)CXR原图;b)“有发现”的激活图,高亮区域为纵隔、心尖、右下肺,与影像征象异常的区域基本吻合;c)预测结果如果是“无发现”,则激活区域在肺外以及脊柱区域,与图像的关键解剖位置不相关。图5 主动脉增宽,CXR二分类预测结果为“有发现”。a)CXR原图;b)“有发现”的激活图,高亮区域为中纵隔主动脉区域,与影像征象异常的区域基本吻合;c)预测结果如果是“无发现”,则激活区域散在分布于整个图像中,与图像的关键解剖位置不相关。图6 深静脉置管,双侧胸腔积液,肺内散在小片渗出,CXR二分类预测结果为“有发现”。a)CXR原图;b)“有发现”的激活图,高亮区域为深静脉置管、双下胸部和左上肺,与影像征象异常的区域基本吻合;c)预测结果如果是“无发现”,则激活区域散在分布于整个图像中,无特殊规律,与图像的关键解剖位置不相关。

的进一步训练,以实现模型的持续迭代。另一方面,将患者入组人群进一步细分,统一专家诊断标准,都有可能提高模型的准确率。

本研究结果显示,深度学习模型可以用于CXR的辅助诊断,这与既往研究结果相似<sup>[12,13]</sup>。对于医学图像的分类问题,模型的可解释性非常重要,有利于对模型分类结果的准确性评估。近年来,相关学者对于深度卷积神经网络模型的可解释性做了很多探索,其中CAM方法是对不同类别的卷积层生成的特征图加权求和而得到激活热图,通过该激活热图可以对模型分

类的结果进行解释<sup>[14]</sup>。Grad-CAM是CAM技术的扩展,可以适用于任何CNN架构。本研究中CXR二分类模型利用Grad-CAM来生成分类激活热图,在该热图上可以展示输入图像中的哪些区域对于得到该分类结果是重要的激活区域。

本研究的数据筛选过程是由医生回顾性读片完成的。对于大样本模型训练,多是从报告中用自然语言处理(natural language processing, NLP)方法提取信息<sup>[15,16]</sup>,优点是处理大量影像数据时可节省医生时间,但报告内容常欠规范,不完整,因此需要大量的数据,即便是大量数据其准确性仍欠佳<sup>[17]</sup>。本研究由影像医生进行数据筛选,且由高年资医生再次回顾性阅片确定图像分类,虽然在数据清洗过程中人力成本较高,但结果较好,用不太多的数据训练得到的模型,准确性仍很高<sup>[18]</sup>。

AI模型研发和泛化过程中应考虑技术问题<sup>[19]</sup>。本研究CXR二分类模型训练过程中未对设备进行筛选,使用了连续数据。CXR图像来自本单位在实际临床工作中使用的多种DR设备,由不同技师完成摄片工

作,未根据设备、人员进行分组。本研究结果证明了不同DR设备、不同技师采集的图像均可用于模型的训练,且无论来源于何种DR设备的图像,其验证集数据的分类预测均达到较好的结果。胸部X线检查有明确的技术规范,经过常规培训的技师都可按照规范完成日常工作,操作一致性强;且现代DR设备有自动曝光功能,可自动设置最佳投照条件并调整图像的窗水平,因此图像预处理难度不大,可适用于多种AI模型的训练<sup>[19]</sup>。由于常规CXR基本可以保证图像质量,

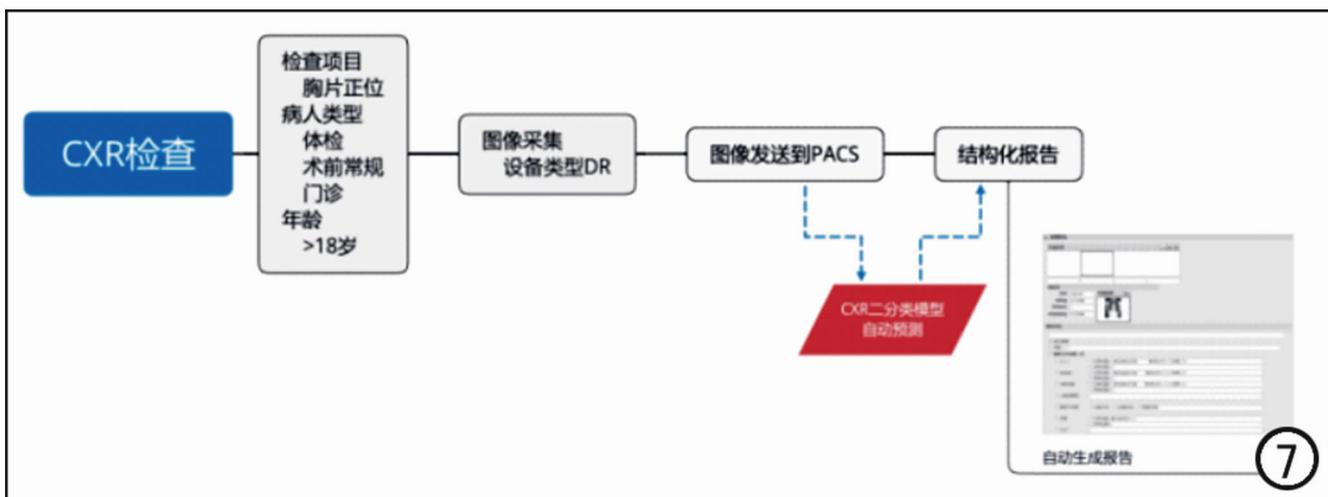


图7 CXR二分类模型与结构化报告整合的流程。

来源于不同设备的图像性质差别不大,从这个角度考虑,在CXr二分类模型泛化过程中,没有图像采集技术上的风险。

从实际场景出发设定AI研发目标,是项目设计的关键问题,决定了模型临床应用的适用性<sup>[20]</sup>。本研究CXr二分类模型训练过程中定义的分类输出结果为“无发现”和“有发现”。在最初的用例设计中,拟将返回的二分类值定义为“正常”和“异常”,但在实际建立队列过程中,发现“肺心膈正常”并不是“正常”。在“肺心膈正常”时,图像中如存在有临床意义的其他征象,也需要报告出来;如“单侧乳腺切除术后,胸壁软组织缺失,两肺心膈未见异常”,再如“PICC置入术后,两肺心膈未见异常”等,这些术后改变、置入物等信息,都是CXr检查应发现并报告给临床医生的。因此,为了提高模型训练效率,本研究中将“肺心膈正常”且合并其他任何需报告的情况归为“无发现”;将有疾病影像表现,或“肺心膈正常但合并其它有临床意义的所见”情况归为“有发现”。

AI临床应用的一个基本要求是与临床流程整合。在符合法规、伦理的框架下,国内外技术人员做过很多探索<sup>[21]</sup>,均认为把AI模型作为独立第三方软件使用的形式不是最优解。笔者认为将AI结果直接返回到临床实际使用的结构化报告,是较好的解决方案<sup>[22]</sup>。为了提高临床工作效率,在将CXr二分类结果返回到报告时,如果模型预测结果为“无发现”,则激活“两肺心膈未见异常”的报告内容,此时报告是完全自动生成的。如果模型预测结果为“有发现”,则在报告中激活的是“异常”控件,需要医生进一步对异常进行描述,这时报告是半自动生成的。本单位临床实际工作中CXr“无发现”的情况较常见,占到工作量的一半以上,因此如将CXr二分类模型接入报告,在分类准确

率足够高时,能明显提高临床工作效率。而且由模型返回结果自动生成的报告,语言规范、结构清晰,对未来提取报告信息做其他研究分析也有利<sup>[23]</sup>。

本研究的局限性:①CXr二分类模型在“有发现”情况下提供的信息不是具体疾病,临床价值受限,仍需进一步优化,训练用于多种疾病诊断的CXr多分类AI模型<sup>[3]</sup>,针对不同体位<sup>[2]</sup>、不同年龄、不同申请目的时对多种常见疾病做出诊断<sup>[24,25]</sup>;②模型研发后应在实际临床工作中进行验证,开展前瞻性、随机对照临床试验,证明使用AI模型的临床获益;③CXr二分类模型在本单位经过验证后,应尝试将其泛化<sup>[26]</sup>,应用于不同医疗机构、不同临床场景下,进一步提高其适用性。

综上所述,使用深度学习模型可对X线胸片做出“无发现”与“有发现”的分类诊断,在常规临床工作中使用模型分检病例,有利于提高医学影像科的工作效率。

致谢:感谢北京赛迈特锐医学科技有限公司张欣静、郭嘉欣在数据整理工作中的贡献。

#### 参考文献:

- [1] American College of Radiology. ACR Appropriateness Criteria [EB/OL]. <https://www.acr.org/Clinical-Resources/ACR-Appropriateness-Criteria>, 2019.
- [2] Kim TK, Yi PH, Wei J, et al. Deep learning method for automated classification of anteroposterior and posteroanterior chest radiographs[J]. J Digit Imaging, 2019 Apr 10. DOI: 10.1007/s10278-019-00208-0.
- [3] Novikov AA, Lenis D, Major D, et al. Fully convolutional architectures for multiclass segmentation in chest radiographs[J]. IEEE Trans Med Imaging, 2018, 37(8): 1865-1876.
- [4] Wei Y, Liu YB, Lin LY, et al. Lung field segmentation in chest radiographs from boundary maps by a structured edge detector[J]. IEEE J Biomed Health Inform, 2018, 22(3): 842-851.
- [5] Candemir S, Antani S. A review on lung boundary detection in

- chest X-rays[J]. *Int J Comput Assist Radiol Surg*, 2019, 14(4): 563-576.
- [6] Mao C, Yao L, Pan Y, et al. Deep generative classifiers for thoracic disease diagnosis with chest X-ray images[C]. Madrid: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018:1209-1214.
- [7] American College of Radiology Data Science Institute. AI use cases 101: what radiologists need to know [EB/OL]. <https://www.acrdsi.org/Blog/AI-Use-Cases-101>, 2019.
- [8] Jia D, Wei D, Richard S, et al. Imagenet: a large-scale hier-archical image database in computer vision and pattern recognition[C]. Florida: Fontainebleau Resort, Miami Beach, 2009: 248-255.
- [9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Las Vegas: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [10] Rahimi RS, Rockey DC. Overuse of head computed tomography in cirrhosis with altered mental status[J]. *Am J Med Sci*, 2016, 351(5): 459-466.
- [11] Heo SJ, Kim Y, Yun S, et al. Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual Workers' health examination data[J]. *Int J Environ Res Public Health*, 2019, 16(2): 250.
- [12] Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the cheXNeXt Algorithm to practicing radiologists[J]. *PLoS Med*, 2018, 15(11): e1002686.
- [13] Singh R, Kalra MK, Nitiwarangkul C, et al. Deep learning in chest radiography: detection of findings and presence of change [J]. *PLoS One*, 2018, 13(10): e0204155.
- [14] Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]. Las Vegas: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2921-2929.
- [15] Chen MC, Ball RL, Yang L, et al. Deep learning to classify radiology free-text reports[J]. *Radiology*, 2018, 286(3): 845-852.
- [16] 马帅, 吴静云, 刘义, 等. 自然语言处理对泌尿系结石 CT 特征的提取: 与人工诊断结果的对照研究[J]. *放射学实践*, 2016, 31(12): 1151-1154.
- [17] Wang C, Elazab A, Jia F, et al. Automated chest screening based on a hybrid model of transfer learning and convolutional sparse denoising autoencoder[J]. *Biomed Eng Online*, 2018, 17(1): 63.
- [18] Qin C, Yao D, Shi Y, et al. Computer-aided detection in chest radiography based on artificial intelligence: a survey[J]. *Biomed Eng Online*, 2018, 17(1): 113.
- [19] Lakhani P. Deep convolutional neural networks for endotracheal tube position and X-ray image classification: challenges and opportunities[J]. *J Digit Imaging*, 2017, 30(4): 460-468.
- [20] Labaki WW, Han MK. Artificial intelligence and chest imaging. Will deep learning make us smarter? [J]. *Am J Respir Crit Care Med*, 2018, 197(2): 148-150.
- [21] Parker W, Lee B, Nicolaou S. AI is bringing USB back: implementing a beta chest X-ray neural network[J]. *J Digit Imaging*, 2019 Jun 25. DOI: 10.1007/s10278-019-00247-7.
- [22] 曹敏, 郭小超, 张来运, 等. 医学影像结构式报告的现状和展望 [J]. *放射学实践*, 2016, 31(12): 1130-1132.
- [23] Huesch MD, Cherian R, Labib S, et al. Evaluating report text variation and informativeness: natural language processing of CT chest imaging for pulmonary embolism[J]. *J Am Coll Radiol*, 2018, 15(3): 554-562.
- [24] Pasa F, Golkov V, Pfeiffer F, et al. Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization[J]. *Sci Rep*, 2019, 9(1): 6268.
- [25] Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study[J]. *PLoS Med*, 2018, 15(11): e1002697.
- [26] Zech JR, Badgeley MA, Liu M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study [J]. *PLoS Med*, 2018, 15(11): e1002683.

(收稿日期: 2019-07-31 修回日期: 2019-08-12)

## AI 对肺磨玻璃结节筛查及定性的临床应用研究

蔡雅倩, 张正华, 韩丹, 黄建强, 李浚利, 全文凤

**【摘要】 目的:**探讨人工智能(AI)对肺磨玻璃结节(GGN)筛查及定性的临床应用价值。**方法:**搜集行胸部 CT 平扫的 200 例患者(共 1230 个 GGN),比较 A 组(住院医师)、B 组(AI)、C 组(住院医师结合 AI)对诊断 GGN 的敏感度、误判率、漏诊率、阳性预测值和平均诊断时间。将其中经手术病理证实的 137 例 GGN 患者按其病理结果分为良性组(54 例)、恶性组(83 例),比较两组间 AI 量化参数的差异,对有统计学差异的参数行 ROC 曲线分析,再以病理结果为因变量,各指标为自变量行 Logistic 回归分析。**结果:**B 组误诊率高于 A、C 两组,阳性预测值小于 A、C 两组;A 组漏诊率高于 B、C 两组,敏感度小于 B、C 两组,差异均有统计学意义( $P$  值均  $<0.05$ )。GGN 良、恶性结节的长径、最大面积、体积、平均 CT 值、最大 CT 值和恶性概率差异均有统计学意义( $P$  值均  $<0.05$ ),对各参数行 ROC 曲线分析,曲线下面积(AUC)均大于 0.7,Logistic 回归分析显示长径和最大面积是 GGN 恶变的独立危险因素。**结论:**AI 协助医生阅片可明显提高肺 GGN 检出敏感度,并可降低误诊率和漏诊率,同时对 GGN 的良恶性预判具有一定参考价值。

**【关键词】** 人工智能;肺磨玻璃结节;体层摄影术,X 线计算机;筛查;诊断,计算机辅助

**【中图分类号】** R814.42;R563.9;TP18 **【文献标识码】** A

**【文章编号】** 1000-0313(2019)09-0958-05

DOI:10.13609/j.cnki.1000-0313.2019.09.005

开放科学(资源服务)标识码(OSID):



**A clinical study of AI on the detection and classification of pulmonary ground glass nodules** CAI Ya-qian, ZHANG Zheng-hua, HAN Dan, et al. Department of Medical Imaging, First Affiliated Hospital of KunMing Medical University, KunMing 650032, China

**【Abstract】 Objective:** To explore the clinical application value of artificial intelligence (AI) in the detection and classification of pulmonary ground glass nodules (GGNs). **Methods:** A total of 1230 GGNs in 200 patients with chest CT plain scan were collected. The detection sensitivity, false positive, range of missed diagnosis, positive predictive value and average diagnosis time were compared in the following groups: group A (resident), group B (AI) and group C (resident combined with AI). 137 GGNs confirmed by surgery and pathology were divided into benign group (54 cases) and malignant group (83 cases) according to their pathological reports. The differences of AI quantification parameters between the two groups were compared, ROC curve analysis was performed for the parameters with statistical differences. Then, logistic regression analysis was performed considering the pathological results as the dependent variables and the parameters as independent variables. **Results:** The false positive of group B was higher than that of group A and C, while the positive predictive value was lower than that of group A and C. The rate of missed diagnosis of group A was higher than that of group B and C, while the detection sensitivity was lower than that of group B and C ( $P < 0.05$ ). There were statistically significant differences in longest diameter, maximum area, volume, average CT value, maximum CT value and malignant rate between benign and malignant groups ( $P < 0.05$ ). All parameters obtained AUCs greater than 0.7 after ROC curve analysis and the longest diameter and maximum area were identified as independent risk factors for malignant transformation of GGN after logistic regression analysis. **Conclusion:** AI-aided CT reviewing can significantly improve the detection sensitivity and reduce the false positive, range of missed diagnosis in terms of pulmonary GGNs. Meanwhile, it has po-

作者单位:650032 昆明,昆明医科大学第一附属医院医学影像科

作者简介:蔡雅倩(1991-),女,湖北荆州人,硕士研究生,主要从事胸部影像诊断工作。

通讯作者:韩丹, E-mail: kmhandan@sina.com

基金项目:云南省教育厅科学研究基金资助项目(2019J1229)

tential reference value for the prediction of benign and malignant GGN.

**【Key words】** Artificial intelligence; Pulmonary ground glass nodules; Tomography, X-ray computed; Screening; Diagnosis, computer-assisted

肺癌的发病率、病死率已位居我国恶性肿瘤第一位,早期诊断和治疗可明显提高患者生存质量,延长生存期。随着高分辨率 CT (high resolution CT, HRCT)在肺癌早期筛查中的普及,图像数量倍增、小结节显示率提高及定量测量使阅片工作量显著增加,高强度工作致影像科医生易产生视觉疲劳,加之经验不足等,不可避免导致小结节的漏诊率增高,特别是磨玻璃结节(ground glass nodule, GGN),密度浅淡,漏诊率更高。以往研究表明持续存在的 GGN 可能提示癌前病变、微浸润性或浸润性肿瘤的可能<sup>[1]</sup>,因此在大量图像资料中准确筛查出 GGN 并早期定性,已成为临床迫切需求。基于深度学习的人工智能(artificial intelligence, AI)凭借其准确的算法模型,不仅能在短时间内检出 GGN,而且能对其进行定性分析,预判其良恶性,已在临床广泛试用,但对其准确性目前尚无明确定论。本研究旨在探讨 AI 对 GGN 筛检和定性诊断的临床应用价值。

## 材料与方法

### 1. 研究对象

搜集 2018 年 12 月—2019 年 6 月间行胸部 CT 平扫的 200 例患者。病例纳入标准:①结节直径 $\leq 3$  cm;②CT 图像层厚为 1 mm。病例排除标准:①肺部弥漫性病变;②图像有严重伪影。137 例 GGNs 患者经手术病理证实,其中良性 54 例,恶性 83 例。54 例良性病变患者中,男 25 例,女 29 例,平均年龄(52.24 $\pm$ 8.47)岁,包括不典型腺瘤样增生 6 例,肺泡间隔纤维化 12 例,结核 11 例,炭末沉积 6 例,炎症 19 例;83 例恶性病变患者中,男 30 例,女 53 例,平均年龄(54.64 $\pm$ 8.45)岁,包括浸润性腺癌 44 例,微浸润性腺癌 29 例,原位癌 10 例。

### 2. CT 扫描方法

采用 Siemens Somatom Definition AS 128 层螺旋 CT 机、Siemens Somatom Definition Flash 及联影 UCT 760 行胸部 HRCT 平扫。Siemens Somatom Definition AS 128 层螺旋 CT 机及 Siemens Somatom Definition Flash 扫描参数:管电压 100 kV,管电流 80 mAs, CarekV semi,自动管电流调制,螺距 0.938,重建卷积函数采用 B70f。联影 UCT 760 扫描参数:管电压 100 kV,管电流 80 mAs,自动管电流调制,螺距 0.938,肺窗滤波函数 B-SHARP-C。重建层厚 1 mm,层间距 1 mm。扫描范围从肺尖至肺底全部区

域,两侧包括胸壁、腋窝。采用肺窗进行图像分析(窗宽 1500 HU,窗位-400 HU)。

### 3. 图像分析处理

由两位从事胸部影像诊断的高年资医师对 200 例患者的胸部 CT 图像同时进行阅片分析,定量结节个数,两者意见不一致时,讨论后达成统一意见。根据阅片方式不同将 200 例患者分成 3 组:A 组由住院医师单独阅片,B 组由 AI(医学影像辅助诊断软件 Dr. Wise Lung Analyzer,版本:V1.1.0.1,型号:MIDS-PNA)单独阅片,C 组由住院医师结合 AI 综合阅片。记录每组的阅片时间及结节检出数,并以两位高年资医师的诊断结果为标准,分别计算 A、B、C 三组的误诊率、漏诊率、敏感度、阳性预测值,误诊率=误诊数/检出数 $\times 100\%$ ,漏诊率=漏诊数/结节总数 $\times 100\%$ ,敏感度=诊断正确数/结节总数 $\times 100\%$ ,阳性预测值=诊断正确数/检出数 $\times 100\%$ 。

将经手术病理证实的 137 例患者的肺窗薄层图像调入进行 AI 分析(图 1),得到相关量化参数:结节长径、短径、最大面积、体积、平均 CT 值、最大 CT 值、最小值 CT 值及恶性概率。长径和短径分别为结节在三维空间内的最长径和最短径。恶性概率为 AI 基于计算机深度学习分析肺结节特征后得出结节为恶性的概率。

### 4. 统计学分析

采用 SPSS21.0 软件包进行统计学分析。采用  $\chi^2$  检验和  $t$  检验比较 A、B、C 三组间肺 GGN 检出效能的敏感度、误诊率、阳性预测值和平均诊断时间;比较 GGN 良、恶性组间一般临床资料和 AI 量化参数的差异,采用受试者工作特征(receiver operator characteristic, ROC)曲线分析其诊断效能,并采用 Logistic 回归分析其恶变的独立危险因素。以  $P < 0.05$  为差异有统计学意义。

## 结果

### 1. 肺 GGN 检出情况比较

由两位从事胸部影像诊断的高年资医师对 200 例患者的胸部 CT 图像进行分析,确认有 1230 个结节。A 组检出 854 个结节,误诊 52 个,漏诊 428 个;B 组检出 1476 个结节,误诊 320 个,漏诊 74 个;C 组检出 1476 个结节,误诊 63 个,漏诊 49 个。三组的平均诊断时间见表 1,两两比较差异均有统计学意义( $P$  值均  $< 0.05$ )。统计学分析结果显示 B 组误诊率高于 A、C

组,阳性预测值小于 A、C 组;A 组漏诊率高于 B、C 组,敏感度低于 B、C 组,差异均有统计学意义( $P$  值均  $<0.05$ ,表 1)。

表 1 三种阅片方式肺结节检出情况比较

指标	A 组	B 组	C 组
误诊率(%)	6.08	21.68*	5.06
漏诊率(%)	34.80*	6.02	3.98
敏感度(%)	65.20*	93.98	96.02
阳性预测值(%)	93.91	78.32*	94.94
诊断时间(min)	7.10±1.68*	0.10±0.03*	3.66±1.22*

注:\*表示与其它组比较,差异具有统计学意义( $P<0.05$ )。

## 2. 良恶性肺 GGN 患者的一般临床特征及 AI 定量参数比较

GGN 良、恶性患者组间年龄、性别、短径、最小 CT 值差异均无统计学意义( $P$  值均  $>0.05$ ),长径、最大面积、体积、平均 CT 值、最大 CT 值及恶性概率差异均有统计学意义( $P$  值均  $<0.05$ ,表 2)。

## 3. 诊断效能分析

表 2 GGN 良恶性组间临床特征及 AI 量化参数比较

参数	良性组	恶性组	$t$ 值/ $\chi^2$ 值	$P$ 值
年龄(岁)	54.33±8.53	55.84±6.86	-1.143	0.255
性别(男/女)	25/29	30/53	1.403	0.236
长径(mm)	8.43±3.64	14.88±5.23	-8.505	$<0.001$
短径(mm)	2.19±0.67	2.08±0.81	0.839	0.403
最大面积( $\text{mm}^2$ )	39.13±23.97	121.23±71.76	-9.629	$<0.001$
体积( $\text{mm}^3$ )	166.99±122.98	888.32±615.25	-10.368	$<0.001$
平均 CT 值(HU)	-634.74±100.89	-524.81±151.77	-4.688	$<0.001$
最大 CT 值(HU)	-142.81±191.93	58.36±201.73	-5.813	$<0.001$
最小 CT 值(HU)	-990.43±44.81	-1002.54±34.24	1.691	0.094
恶性概率(%)	53.98±35.68	80.06±20.13	-4.889	$<0.001$

对良、恶性组间有统计学差异的各个参数进行 ROC 曲线分析(图 2),以曲线下面积(area under curve, AUC)大于 0.7 的参数为自变量,以病理结果良恶性为因变量,进行二元 Logistic 回归分析,结果显示长径[优势比(odds ratio, OR)=4.232,  $P<0.001$ ]和最大面积(OR=0.896,  $P=0.008$ )是 GGN 恶变的独立危险因素。

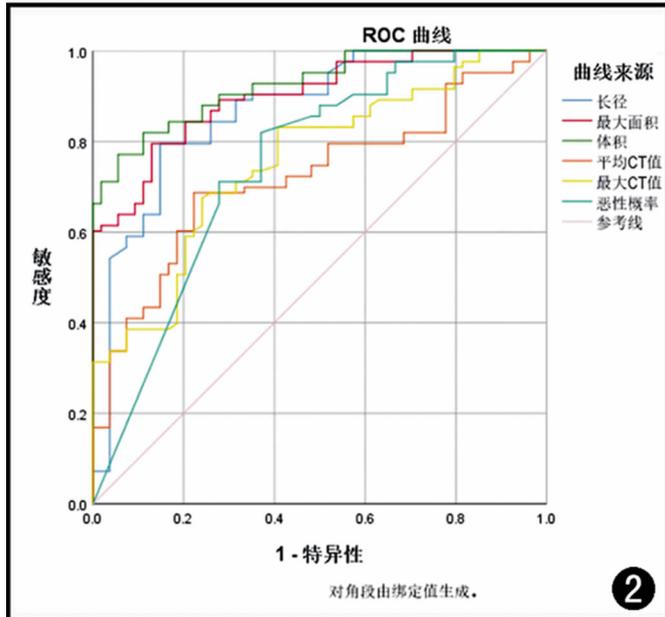
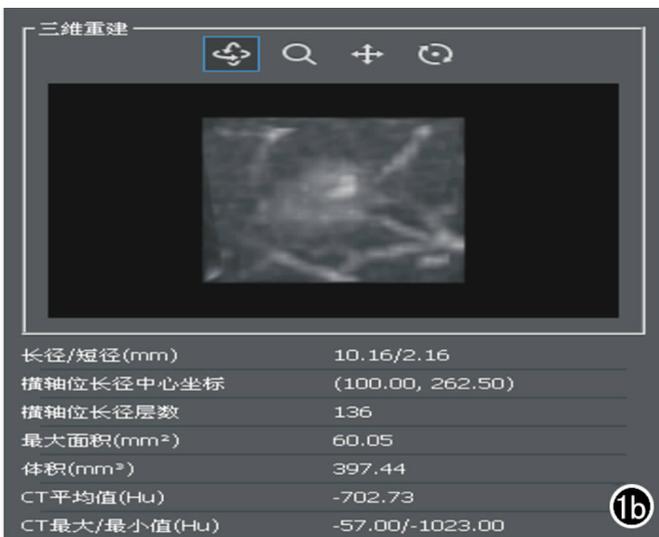
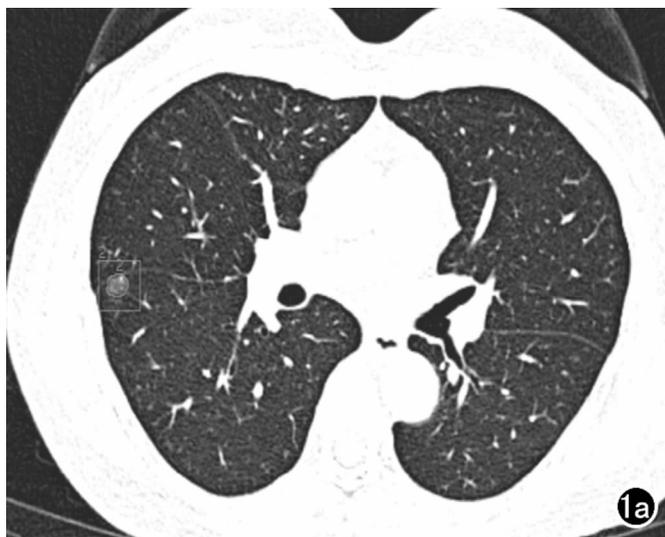


图 1 AI 自动筛选肺结节并对其进行定量分析,计算机自动测量其长径、短径、最大面积、平均 CT 值、最大 CT 值、最小 CT 值,并对其进行准确定位。a) AI 对肺内结节进行筛检、勾画结节轮廓并标注序号; b) AI 直接显示肺结节的各参数。  
图 2 GGN 各量化参数的 ROC 曲线。

表3 GGN良、恶性组的ROC曲线分析

指标	阈值	敏感度	特异度	AUC	P值
长径(mm)	10.34	0.795	0.852	0.867	<0.001
最大面积(mm <sup>2</sup> )	61.12	0.795	0.870	0.899	<0.001
体积(mm <sup>3</sup> )	358.62	0.771	0.944	0.925	<0.001
平均CT值(HU)	-573.44	0.687	0.778	0.722	<0.001
最大CT值(HU)	19.50	0.675	0.759	0.756	<0.001
恶性概率(%)	80.50	0.819	0.630	0.747	<0.001

## 讨论

肺癌是我国最常见的恶性肿瘤,病死率高,但研究表明 IA 期肺癌患者中,原位癌和微浸润性癌术后五年生存率可接近 100%<sup>[2,3]</sup>,因此肺癌的早期发现、早期诊断及治疗显得尤为重要。早期筛查已成为肺癌防治的重要手段,其中 HRCT 是国际公认的有效手段。肺结节筛查图像诊断较为简单,但随着人们健康体检意识不断加强,胸部 HRCT 筛查人群日益增多,图像数量倍增,影像医师的工作压力也日趋增大,漏诊、误诊的风险不断增加。AI 的出现不仅降低了早期肺癌的漏诊率,也在一定程度上解放了影像科医生的双眼,使其可将有限的精力用于更高难度的工作中。AI 是现阶段各行各业研究的热点,目前已广泛运用于影像、临床及病理等多个医学领域<sup>[4,5]</sup>,主要集中在发现异常、量化测量及鉴别诊断等方面<sup>[6,7]</sup>。而用于肺结节筛查和定性的辅助诊断软件国内外各有不同,主要是由于模型算法不同使其结果的灵敏度、特异度出现一定偏差<sup>[8]</sup>。本研究旨在探讨我院引入的深睿公司医学影像辅助诊断软件对肺 GGN 检出和定性诊断的价值。

本研究结果显示 A 组的漏诊率明显高于 B、C 两组,而敏感度明显低于 B、C 两组,正是由于医生在进行大量重复阅片过程中不可避免地产生视觉疲劳,大大影响结节的检出情况。B 组的误诊率明显高于 A、C 两组,而阳性预测值明显低于 A、C 两组,原因为 AI 主要容易将局部稍增粗的肺纹理误判为 GGN,表明 AI 的算法有待进一步完善和提高。C 组(住院医师结合 AI)的阅片诊断水平接近于高年资医师,明显提高了 GGN 的检出率,减少了漏诊风险,又弥补了 AI 假阳性率高的缺点。另外,AI 平均 6s 能完成 1 例阅片,而一般人工阅片需要 7 min,人工结合 AI 大大提高了影像科医生的工作效率和诊断准确性,值得广泛推广,与胡琼洁等<sup>[9]</sup>和邵亚军等<sup>[10]</sup>的研究结果一致。

以往较多研究探讨肺 GGN 良、恶性间三维特征的差异,对其良恶性预判具有一定参考价值,但测量繁琐、费时,AI 弥补了这些缺点,可直接提取结节并对其各指标进行准确量化,方便、快捷,减少了人工测量的误差,可重复性强。另外 AI 可根据其模型算法计算出结节的恶性概率,影像医生可参考 AI 提供的三维

特征参数及恶性概率对其进行定性分析,提高诊断效率。本研究中,GGN 良、恶性组间结节短径和最小 CT 值差异无统计学意义,长径、最大面积、体积、平均 CT 值、最大 CT 值及恶性概率差异均有统计学意义,与 Shi 等<sup>[11]</sup>的研究结果一致。回归分析结果显示长径和最大面积是 GGN 恶变的危险因素,AUC 分别为 0.867、0.899,当临界值分别为 10.34 mm、61.12 mm<sup>2</sup>时,敏感度均为 0.795,特异度分别为 0.852、0.870。Kitami 等<sup>[12]</sup>和熊廷伟等<sup>[13]</sup>的研究结果表明直径是 GGN 恶变的危险因素,Yang 等<sup>[14]</sup>的研究结果也表明基于直径的 logistic 回归模型用于区分良、恶性 GGN 的准确率达 78.7%,AUC 为 0.861,敏感度和特异度分别为 78.0%、80.0%。最大面积是 GGN 恶变的独立危险因素以往少有文献报道,可能与本研究中部分 GGN 整体体积缩小、局部恶变面积增大有一定关系,需扩大样本量进一步证实。关于 GGN 的 CT 值以往研究结果各有不同,矫娜等<sup>[15]</sup>研究表明恶性 GGN 的平均 CT 值及最大 CT 值均明显高于良性 GGN,而 Xiang 等<sup>[16]</sup>认为两者差异并无统计学意义,本研究结果表明两者间差异有统计学意义,病理机制可能是由于恶性 GGN 内肿瘤细胞沿着肺泡间隔贴壁生长,导致肺泡腔内气体减少,或由于肺泡塌陷、纤维化等导致病灶内局部密度增高<sup>[17]</sup>。Logistic 回归分析结果显示恶性概率不是 GGN 恶变的危险因素,表明 AI 对 GGN 的定性预判能力有待提高。但是,本研究发现良、恶性两组 GGN 的恶性概率差异有统计学意义,当恶性概率大于 80%时,提示 GGN 恶性的可能性较大,具有一定参考价值。因此在诊断工作中重点观察 AI 提供的 GGN 长径、最大面积、体积、CT 值、恶性概率等参数,并结合常规形态学进行综合分析,可明显提高 GGN 定性诊断的准确率。

本研究存在以下局限性:①需扩大样本量,进一步对结果进行验证;②未引入随访数据等重要指标说明 AI 在肺结节随访中的价值,有待进一步研究。

综上所述,在肺 GGN 的检出及良、恶性鉴别诊断方面,AI 可在短时间内有效提高 GGN 检出的敏感度,并对其三维特征参数进行准确量化,计算其恶性概率,且结果可信度较高,是影像科医生可靠而实用的有力助手,可在临床中广泛应用。但目前 AI 的假阳性率仍较高,需进一步改善其算法以提高特异度。

## 参考文献:

- [1] 秦福兵,陆友金.纯磨玻璃结节定量 CT 鉴别肺浸润性腺癌与浸润前病变和微浸润腺癌[J].中国医学计算机成像杂志,2016,22(1):22-26.
- [2] Wilshire CL, Louie BE, Manning KA, et al. Radiologic evaluation of small lepidic adenocarcinomas to guide decision making in surgi-

- cal resection[J]. *Ann Thorac Surg*, 2015, 100(3):979-988.
- [3] Zhou J, Li Y, Zhang Y, et al. Solitary ground-glass opacity nodules of stage IA pulmonary adenocarcinoma: combination of <sup>18</sup>F-FDG PET/CT and high-resolution computed tomography features to predict invasive adenocarcinoma [J]. *Oncotarget*, 2017, 8(14):23312-23321.
- [4] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis[J]. *Med Image Anal*, 2017, 42(9):60-88.
- [5] Qin C, Yao D, Shi Y, et al. Computer-aided detection in chest radiography based on artificial intelligence: a survey[J]. *Biomed Eng Online*, 2018, 17(1):113.
- [6] Gong L, Jiang S, Yang Z, et al. Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks[J]. *Int J Comput Assist Radiol Surg*, 2019 Apr 26. DOI: 10.1007/s11548-019-01979-1.
- [7] Valente IR, Cortez PC, Neto EC, et al. Automatic 3D pulmonary nodule detection in CT images: A survey[J]. *Comput Methods Programs Biomed*, 2016, 124(2):91-107.
- [8] Shaffie A, Soliman A, Fraiwan L, et al. A generalized deep learning-based diagnostic system for early diagnosis of various types of pulmonary nodules[J]. *Technol Cancer Res Treat*, 2018, 17(1):1533033818798800.
- [9] 胡琼洁, 陈冲, 王玉锦, 等. 实习医师以共同阅片和第二阅片者模式使用计算机辅助检测系统在低剂量 CT 中的应用研究[J]. *放射学实践*, 2018, 33(10):1022-1028.
- [10] 邵亚军, 张宝荣, 郭佑民, 等. 计算机辅助工具对肺结节检测效能的研究[J]. *实用放射学杂志*, 2018, 34(9):1430-1433.
- [11] Shi Z, Deng J, She Y, et al. Quantitative features can predict further growth of persistent pure ground-glass nodule[J]. *Quant Imaging Med Surg*, 2019, 9(2):283-291.
- [12] Kitami A, Sano F, Hayashi S, et al. Correlation between histological invasiveness and the computed tomography value in pure ground-glass nodules[J]. *Surg Today*, 2016, 46(5):593-598.
- [13] 熊廷伟, 李川, 龚明福, 等. MSCT 在肺孤立性磨玻璃结节鉴别诊断中的价值[J]. *中华肺部疾病杂志(电子版)*, 2018, 11(4):401-404.
- [14] Yang Y, Wang WW, Ren Y, et al. Computerized texture analysis predicts histological invasiveness within lung adenocarcinoma manifesting as pure ground-glass nodules[J]. *Acta Radiol*, 2019, 28(2):284185119826536.
- [15] 矫娜, 吴明祥, 龚静山, 等. 计算机辅助诊断定量分析表现为磨玻璃样结节的肺原位腺癌与非典型腺瘤样增生[J]. *中国 CT 和 MRI 杂志*, 2015, 13(6):29-31.
- [16] Xiang W, Xing Y, Jiang S, et al. Morphological factors differentiating between early lung adenocarcinomas appearing as pure ground-glass nodules measuring  $\leq 10$ mm on thin-section computed tomography[J]. *Cancer Imaging*, 2014, 14(11):33.
- [17] 曹恩涛, 于红, 范丽, 等. 纯磨玻璃密度结节肺腺癌的 CT 三维定量分析[J]. *中华放射学杂志*, 2016, 50(12):940-945.

(收稿日期:2019-07-29 修回日期:2019-08-14)

## 影像组学在胰腺肿瘤病变影像学评估中的研究进展

杨采薇, 蒋涵羽, 刘曦娇, 宋彬

**【摘要】** 随着影像学诊断手段的不断丰富,胰腺肿瘤的检出率也随之提高。人工智能在近年来不断应用于医学影像中,为肿瘤影像的发展开启了新时代。腹部肿瘤影像在不断地发掘影像中隐藏的信息,影像组学在胰腺肿瘤病变中的研究聚焦于胰腺肿瘤的诊断与鉴别诊断、生物学行为的评价、预后评估以及放化疗的疗效评估等方面,以期提高对肿瘤病变的诊断效能,更精准地监测肿瘤的进展、预测肿瘤的预后,为个体化及精准化治疗做出贡献。本文旨在综述影像组学在胰腺肿瘤病变影像学评估中的研究现状、存在的问题以及对未来的初步展望。

**【关键词】** 影像组学; 医学影像学; 胰腺肿瘤; 预后; 疗效评估; 诊断, 鉴别

**【中图分类号】** R735.9; TP18 **【文献标识码】** A **【文章编号】** 1000-0313(2019)09-0963-06

DOI:10.13609/j.cnki.1000-0313.2019.09.006

开放科学(资源服务)标识码(OSID):



人工智能(Artificial Intelligence, AI)是指研究开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门信息科学<sup>[1-2]</sup>。人工智能在医学影像中的运用主要是基于影像大数据的挖掘、提取、搜索和运用的图像识别及深度学习,提供对人类日常工作不可视的模式挖掘,影像组学(radiomics)是此类工作模式的代表。

### 影像组学简介及其优势

影像组学一词最早由美国学者 Gillies 于 2010 年首次提出<sup>[3]</sup>,用于描述从影像图像中提取定量特征,并于 2012 年由荷兰学者 Lambin 进一步完善影像组学的概念<sup>[4]</sup>,即“从放射图像中高通量提取影像特征并创建高维数据集”。影像组学通过结合影像图像识别、提取、量化及处理高通量提取出的影像特征及表型<sup>[5]</sup>。

影像组学在医学影像中运用广泛,其在中枢神经系统、乳腺、肺、肝、肾、胰腺、胃部、结肠、血液系统及前列腺等部位的肿瘤均有一定的研究,在肝纤维化、肺纤维化、心血管疾病以及中枢神经系统痴呆等病变的领域也有所涉猎。影像组学在医学影像应用最成熟的领域为肿瘤影像,其在肿瘤检出、定性诊断、自动结构化报告、肿瘤提取、肿瘤放疗靶器官勾画等方面已经有较多的科研试验和临床研究<sup>[6]</sup>,已广泛应用于肺结节和肺癌筛查、乳腺癌筛查以及前列腺癌影像诊断等临床诊疗工作中<sup>[7]</sup>。

然而,目前影像组学的产生基本机制和生物基础

还有待验证。肿瘤微环境,包括瘤周的纤维化状态,增殖、代谢、缺氧状态以及组织或肿瘤的血管生成等与影像组学的关系仍处于亟待研究的状态。尽管肿瘤的成分和生物学行为复杂,肿瘤在细胞和分子水平上可见的改变与影像学信号难以直接关联,但这些因素可能导致病变呈现出肉眼难以直接观察的图像变化,从而促成影像组学的存在。无论肿瘤的物理、形态学和生物学性质如何,影像组学在评估病变和组织方面与现有的其它技术相比具有根本优势,这些优点包括:①影像组学可以提供有关肿瘤行为和反应在时间演变和空间演变上的具体信息;②与活检或实验室检查所需的样本消耗相比,影像组学的样本具有永久保存性;③与活检组织的病变异质性抽样误差相比,影像组学能够直观地观察病变的总体特征和潜在地观察病变器官及毗邻器官的背景特征;④无创性成像而非组织活检的相对安全性。因此,影像组学技术有可能应用于早期检测、治疗监测、预后预测和生物标记物的发现<sup>[8]</sup>。

对于胰腺肿瘤来说,目前的整体治疗效果欠佳,并且胰腺肿瘤复杂多样,发生发展的分子机制也尚不明确,缺乏早期诊断及鉴别诊断方法。胰腺癌起病隐匿,恶性程度高,早期发现困难,导致多数患者在初诊时已无手术机会<sup>[9]</sup>;胰腺囊性肿瘤种类繁多,性质不一,术前难以准确区分良恶性,导致部分患者接受影响预后的不恰当的治疗方案;胰腺神经内分泌肿瘤异质性强,临床表现多样,容易误诊误治。胰腺肿瘤诊治水平的提高仍有赖于对其发生发展机制的深入了解以及诊断与治疗方法的进步<sup>[10]</sup>。

目前影像组学在胰腺肿瘤方面的研究主要关注于胰腺常见肿瘤的诊断与鉴别诊断、生物学行为评价及分层分级、恶性度评价及预后预测、治疗反应与疗效评

作者单位:646000 成都,四川大学华西医院放射科  
 作者简介:杨采薇(1996-),女,四川成都人,硕士研究生,主要从事腹部影像诊断工作。  
 通讯作者:宋彬, E-mail: cjr\_songbin@vip.163.com  
 基金项目:国家自然科学基金面上项目(81771797)

估等方面。

## 影像组学在胰腺肿瘤影像学中的研究现状

### 1. 胰腺肿瘤的诊断及生物学行为评价

现有影像组学在胰腺肿瘤研究的诊断方面侧重于胰腺神经内分泌瘤(pancreatic neuroendocrine neoplasms, PNENs)的分级诊断及鉴别诊断,以及胰腺浆液性囊腺瘤的鉴别诊断及治疗方案选择,为传统影像学在诊断胰腺肿瘤时遇到的难题提供新的解决思路和方案;对于交界性肿瘤生物学评价的研究为其个体化、精准化的临床诊疗方案提供了诊断效能更高的术前预测手段,能够更好地管控肿瘤的生物学进展。

胰腺神经内分泌瘤的诊断、鉴别诊断及分级:PNENs 占所有胰腺肿瘤的 2%~10%,PNENs 起源于多能干细胞并表现出内分泌症状<sup>[11]</sup>。近十年来 PNENs 的发病率有升高趋势<sup>[12]</sup>。2010 年 WHO 指南依据 Ki-67 指数及核分裂数将 PNENs 分为 G1、G2、G3 这 3 个等级<sup>[13]</sup>。欧洲神经内分泌瘤协会指出,PNENs 的分级评估与其生物学行为、预后预测及治疗方式的选择密切相关<sup>[14]</sup>。

PNENs 典型影像表现为边缘清楚的富血供实性肿瘤,在增强扫描动脉期明显强化,CT 值高于周围正常胰腺组织,并且不伴导管扩张及血管套征象<sup>[15,16]</sup>。然而近期研究报道,高达 41.5%的 PNENs 为乏血供肿瘤,特别是 G2 及 G3 级别的肿瘤<sup>[17,18]</sup>,乏血供肿瘤影像学表现不典型,与胰腺癌鉴别诊断困难。Li 等<sup>[19]</sup>利用 CT 纹理分析特别是参数直方分布图鉴别不典型的乏血供 PNENs 与胰腺导管腺癌,发现了偏度(skewness)、均数(mean)、百分位数(median)等有价值的纹理参数。此外,胰腺神经内分泌癌(pancreatic neuroendocrine carcinoma, PNEC)与胰腺导管腺癌在 CT 上常常难以鉴别。近期有研究证实动脉期 PNEC 相对于胰腺导管腺癌通常表现为较低熵(entropy)和较高均匀度(uniformity)( $P < 0.05$ )。此研究证实传统征象中动脉期及门脉期的增强程度鉴别诊断两者的效能最高,其受试者工作特征(receiver operating characteristic, ROC)曲线的曲线下面积(area under curve, AUC)分别为 0.98 和 0.99,但熵和均匀度的鉴别诊断效能(AUC:0.71~0.72)也达到了可接受程度<sup>[20]</sup>。相关研究则证实了 CT 传统征象及定量组学分析可鉴别 PNENs 与胰腺转移性肾细胞癌<sup>[21]</sup>,传统征象如肿瘤钙化、孤立肿块及主胰管扩张等是 PNENs 的特异性征象,但其敏感度欠缺,而纹理参数中的熵在保证一定总体准确度的基础上提高了敏感度(sensitivity, SEN),其 AUC 为  $0.77 \pm 0.06$ ,SEN 为 71.4%,特异度(specificity, SPE)为 79.1%。而对于富血供

PNENs 与胰腺内副脾的鉴别, Lin 等<sup>[22]</sup>认为胰腺内副脾常表现为动脉期不均质强化及门脉期与脾脏一致的强化(强化程度均强于富血供 PNENs),而纹理特征中的熵及均匀度在中标准差值(1.5)及高标准差值(2.5)中对于鉴别富血供 PNENs 与胰腺内副脾均有统计学意义( $P < 0.01$ ),且其诊断效能高于传统征象,均匀度在中标准差值(1.5)及高标准差值(2.5)中的鉴别诊断效能:AUC 为 0.82~0.89,SEN 为 85.0%~95.0%,SPE 为 75.0%~83.3%。PNENs 的鉴别诊断虽然得到了一些有意义的结果,如熵及均匀度等代表肿瘤异质性的纹理参数,但还需要进一步的研究结果验证其结果的可靠性及真实性。

虽然已有许多研究证实 CT、MRI 及 PET 对于术前诊断 PNENs 及其分级有一定价值,但上述无创性影像检查方法对其早期诊断、鉴别及分级的准确度仍有待提高<sup>[23]</sup>。D'Onofrio 等<sup>[24]</sup>对 100 例 PNENs 患者的术前增强 CT 图像进行回顾性分析,发现 CT 纹理分析中的峰度(Kurtosis)在 G1、G2、G3 级肿瘤中差异有统计学意义(G1 vs G2,  $P = 0.0063$ ; G2 vs G3,  $P = 0.0004$ ; G1 vs G3,  $P < 0.0001$ ),而熵在 G1 与 G3 ( $P = 0.0013$ )、G2 与 G3 级别 PNENs ( $P = 0.0084$ )中差异也有统计学意义。峰度和熵对于各级别 PNENs 的诊断效能也有价值,峰度诊断 G3 级别 PNENs 的 AUC 达 0.924,SEN、SPE 分别为 82%、85%;熵诊断 G3 级别 PNENs 的 AUC 达 0.732,SEN、SPE 分别为 82%、64%。Choi 等<sup>[25]</sup>进行了类似研究,结果显示对于 G2 及 G3 级别 PNENs 的预测,CT 纹理分析的诊断效能高于传统 CT 征象(AUC 分别为 0.744 和 0.738),但两者差异无统计学意义;并且对这两个级别 PNENs 的预测,CT 纹理分析得到了有意义的参数——较低的球状性(sphericity)、较高的偏度及较低的峰度,级别越高的 PNENs,其囊变、钙化、坏死的程度及范围增加,其异质性也相对增高。Canellas 等<sup>[26]</sup>研究认为 CT 纹理参数中熵与肿瘤分级相关,较高熵在中-高级别 PNENs 中的比值比(odd ratio, OR)是低级别 PNENs 的 3.7 倍( $P = 0.008$ ),且熵高于 4.65 的 PNENs 患者的无病生存期与熵低于 4.65 的 PNENs 患者相比差异有统计学意义。基于 MRI 的纹理分析对于 PNENs 的分级有一定意义<sup>[23]</sup>,基于 T<sub>2</sub>WI 序列的 MRI 纹理参数有逆差矩(inverse difference moment)、能量(energy)、相关(correlation)和熵差(difference entropy),基于 DWI 的 MRI 纹理参数有相关、对比度(contrast)、逆差矩、最大灰度(maxintensity)和熵差,该 6 种纹理参数的 AUC 在 0.703~0.989 之间。但以上研究均为回顾性研究,样本量较小,且有图像扫描技术参数不统一、手动勾画兴趣区(region of

interest, ROI)造成的观察者间差异等问题存在,导致其研究的准确度及模型泛化能力受影响。

胰腺浆液性囊腺瘤的鉴别诊断及治疗方案选择:随着多种诊断手段的不断发展,近二十年来胰腺囊性瘤的检出率不断提高<sup>[27]</sup>,但是由于传统影像学检查在术前对胰腺囊性瘤的生物学行为、侵袭性以及演进方面仅能提供有限的信息,导致对胰腺囊性瘤的手术管理方式往往过于激进<sup>[28]</sup>。目前研究发现,胰腺囊性瘤主要分为四种亚型:浆液性囊性瘤(serous cystic neoplasms, SCNs)、导管内乳头状粘液瘤(intraductal papillary mucinous neoplasms, IPMNs)、粘液性囊性瘤(mucinous cystic neoplasms, MCNs)以及实性假乳头状瘤(solid pseudopapillary neoplasms, SPNs)。其中 SCNs 是一类较为良性和惰性的分型,生长缓慢并且恶性进展的风险较低<sup>[29,30]</sup>。Wei 等<sup>[31]</sup>利用基于计算机辅助支持系统(computer-aided diagnosis, CAD)的组学分析 CT 图像,用支持向量机(support vector machine, SVM)筛选出有意义的纹理指标,纹理参数在训练集中得出的诊断效能:AUC 为 0.767, SEN 为 0.686, SPE 为 0.709;在验证集中得出的诊断效能:AUC 为 0.837, SEN 为 0.667, SPE 为 0.818。由此可见,CT 组学能提高 SCNs 与其它种类胰腺囊性瘤的鉴别诊断效能。

## 2. 胰腺肿瘤的预后评估、恶性程度评价及疗效预测

影像组学在胰腺肿瘤的疗效预测、评价及预后评估中的研究主要集中于胰腺导管内乳头状粘液瘤(IPMNs)的风险度分层、恶性程度预测及胰腺癌的预后、放化疗的疗效评估等方面。

IPMNs 的风险度分层及恶性程度预测:1982 年 Ohashi 第一次提出了过去未被识别的胰腺囊性瘤的新病理描述——IPMNs<sup>[32]</sup>。相较于生物学行为较好,临床大多数建议保守治疗随访的 SCNs, IPMNs 被认为是侵袭性较强且生物学行为较差的一类胰腺肿瘤,其为恶性或可能为恶性的程度高<sup>[33-35]</sup>。IPMNs 的病理进展为低级别肿瘤(LG-grade)-高级别肿瘤(HG-grade)-侵袭性肿瘤,而 IPMN 逐步进展演变为胰腺癌的过程占有胰腺癌构成的 20%~30%<sup>[36-38]</sup>。非侵袭性 IPMNs 外科切除后的 5 年生存率高达 90%~100%<sup>[39]</sup>,而 IPMNs 进展为侵袭性 IPMNs 后的 5 年生存率仅 30%<sup>[40]</sup>,可见其肿瘤的恶性程度很高<sup>[41]</sup>,故在早期诊断高级别 IPMNs 以及在 IPMNs 进展为侵袭性 IPMNs 前行根治性切除非常重要。仙台共识指南以及福冈指南不断更新,根据病变大小及影像学形态对其进行预测,但其数据研究来源具有选择偏倚——恶性程度高而进行手术切除的病变被纳入较

多,并且未结合肿瘤标志物、肿瘤代谢等临床指标进行预测。故现有指南的预测效能不足,对一些良性病变的误切导致其阳性预测值较低<sup>[42]</sup>,而影像组学的出现为 IPMNs 危险度分层及恶性程度的预测提供了新的方向。

Hanania 等<sup>[43]</sup>利用 CT 纹理分析灰度共生矩阵对低级别 IPMNs 和高级别 IPMNs 进行预测,通过计算机提取出 360 个纹理参数进行降维、筛选、建模,最后筛选出 14 个参数,其预测 IPMN 恶性程度分级最有效参数的 AUC 达 0.82, SEN、SPE 分别为 85%、68%,在交叉验证设计研究中其最优的 logistic 回归得出的 AUC 为 0.96, SEN、SPE 高达 97%、88%。同样,Chakraborty 等<sup>[44]</sup>采用较大样本(n=103)对 IPMNs 的高风险组及低风险组进行了纹理分析预测,同样得出对分组有意义的纹理参数,其最佳 AUC 为 0.77,而纹理参数结合临床指标的预测效能 AUC 达 0.81。Permeth 等<sup>[45]</sup>的研究结果显示,影像组学与 miRNA 基因组学分类器(miRNA genomic classifier, MGC)结合的模型较传统病理学更能预测 IPMNs 的生物学行为及潜在恶性程度,该结合模型预测效能高于各自单独的预测效能, AUC 达 0.92,并且 SEN、SPE、阳性预测值(positive predictive value, PPV)、阴性预测值(negative predictive value, NPV)分别为 83%、89%、88%、85%。影像组学技术可以结合或补充其它已有手段来预测 IPMNs 的恶性程度,并获得更准确的结果,为 IPMNs 的治疗、管理、随访、预后等提供更为可靠的方式。

胰腺癌的预后及放化疗后的疗效评估:据《2013 年中国肿瘤登记年报》统计,胰腺癌位列我国男性恶性肿瘤发病率第 8 位,人群恶性肿瘤死亡率第 7 位,在全球范围内均呈快速上升趋势<sup>[46]</sup>。胰腺癌 5 年生存率约为 6%<sup>[47]</sup>,相较于其他上皮性癌而言,其 5 年生存率如此低是因为大部分胰腺癌患者在诊断时已处于晚期,大多数患者早期无特异性症状直到疾病进展至中晚期<sup>[10]</sup>。外科切除曾被认为是提高手术患者 15%~20% 总生存率(overall survival, OS)的唯一潜在根治性治疗手段<sup>[48]</sup>,约有 20% 的胰腺癌患者确诊后可进行外科切除<sup>[49]</sup>,但行切除术的胰腺癌患者复发率很高,导致行切除术后的胰腺癌患者的 5 年生存率也仅有 25%<sup>[47]</sup>。

如今,可切除肿瘤的术后辅助治疗有望延长患者的总体生存时间,并且对于原本可能切除的病变术前新辅助治疗使得其获得根治性切除手术的可能性<sup>[49]</sup>。早期外科切除术前增强 CT 被认为是评估胰腺癌的可切除性、分级分期及肿瘤进展的常规影像学手段<sup>[50,51]</sup>,源于 CT 的组学研究也因此有望成为评估

肿瘤预后的可行性手段。对于可切除的胰腺癌患者, Eilaghi 等<sup>[52]</sup>研究发现, 胰腺癌病变组织纹理特征中的差异性 (tumordissimilarity) 及归一化逆差 (inverse differencenormalized) 的值显著高于相应胰腺实质背景的上述纹理特征的值, 相较于肿瘤信号及肿瘤大小等传统 CT 征象而言, 纹理特征对于总生存率的预测效能更高。Yun 等<sup>[53]</sup>运用纹理分析的直方图和灰度共生矩阵回顾性研究了 88 例胰头癌的术前 CT, 发现纹理特征在复发组与非复发组中差异有统计学意义, 并且在 COX 风险模型多元回归分析中, 较低标准差 (standard deviation)、较低对比度、较高相关与较差的无病生存率 (progression free survival, PFS)、淋巴结转移显著相关。

对于不可切除或可能切除的胰腺癌患者, Sandra-segaran 等<sup>[54]</sup>运用影像组学纹理分析的方法回顾性分析 60 例化疗患者, 通过 Kaplan-Meier 生存分析, 发现高峰期 ( $>0.565$ ) 及高正像素平均值 (mean values of positive pixels, MPP) ( $>31.625$ ) 与预后更差的总生存率有关, 但影响不可切除胰腺癌患者总生存率最为重要的因素还是肿瘤转移。在 COX 风险模型多元回归分析中, 肿瘤转移与总生存率、无进展生存率显著相关。Cheng 等<sup>[55]</sup>也对不可切除的胰腺癌化疗前 CT 进行组学研究, 发现了可预测无病生存率及总生存率的相关纹理参数, 如特定过滤值下的偏度、肿瘤标准差。Campbell 等<sup>[56]</sup>研究证实, 在转移性胰腺癌原发病灶的组学研究中, 熵值及峰度与染色体间异质性相关, 而胰腺背景实质的熵值及峰度也能提供转移性胰腺癌总生存率的预后信息。

除化疗之外, 立体定向全身放疗 (stereotactic body radiation therapy, SBRT) 是近年来针对不可切除胰腺癌的新兴治疗技术, 可利用 1~5 个分数的有限靶区实现高剂量放射精准定位, 因此在胰腺癌局部治疗中有较大优势<sup>[57]</sup>。Cozzi 等<sup>[58]</sup>评估了 CT 纹理分析在预测胰腺癌 SBRT 放疗治疗的总生存率及局部控制方面的能力, 并发现了在高风险组与低风险组中可观察到有统计学差异的纹理参数。与总生存率有关的纹理参数在训练集和验证集中均有统计学意义 ( $P=0.01$  和  $0.05$ , 一致性指数分别为  $0.73$  和  $0.75$ ); 与局部控制有关的纹理参数在训练集和验证集中均有统计学意义 ( $P=0.007$  和  $0.004$ , 一致性指数分别为  $0.69$  和  $0.75$ )。

综上所述, 影像组学对于预测胰腺癌的预后与评估放化疗的疗效具有巨大的潜在价值, 但目前的研究尚不完善, 需要进一步研究与验证。

## 影像组学研究及其在胰腺肿瘤影像中存在的问题

虽然影像组学的研究如火如荼, 但其在运用于临

床实践前需要解决的问题还很多。

病变自身方面。由于肿瘤的类型、大小、形态、解剖位置、涉及器官等方面都存在差异, 所以在不同肿瘤中的研究成果进度不一。对于胰腺肿瘤, 目前研究主要集中于胰腺恶性肿瘤中最常见的胰腺癌与其它肿瘤的鉴别诊断及其预后评估、治疗评价, 导管内乳头状粘液瘤的预后、生物学行为评价, 胰腺神经内分泌瘤的分级预测、鉴别诊断等, 而胰腺肿瘤分类繁多, 特别是胰腺囊性肿瘤分类繁杂, 对于胰腺少见及罕见肿瘤以及肿瘤合并炎症或其它疾病的影像组学研究还有所欠缺。

图像方面。由于扫描参数、扫描仪器的不统一, 图像质量不能保证; 图像采集、图像分割、图像感兴趣区勾画分割及纹理参数的提取都采用不同的方法标准, 以及图像后处理及统计学建模等采用的标准也无规范。胰腺作为腹膜后器官, 体积较小, 形态柔软, 在不同个体及年龄间表现出非常高的患者间解剖变异性, 导致其在定位及自动分割等方面的研究进展缓慢, 现多采用手动分割的方法。而且在组学纹理特征的统计中也存在选择偏倚, 不同的统计手段得出的纹理特征经不同的筛选方法得出的结果不一, 导致研究结果的可重复性有待商榷, 需要更加细致、更规范化、更标准的研究工作。

数据及方案设计方面。大数据、云数据无法获得, 信息壁垒导致深度学习的大规模研究无法开展实施, 这是大多数影像组学研究存在的问题, 目前大多数研究为单中心回顾性研究, 样本数量有限, 不仅限制了影像组学阈值标准的设定, 而且数据受限于样本量不足而使模型无法得到充分训练, 难以拟合疾病特征的真实状态, 过于复杂的算法也会增加模型的过拟合风险, 限制了其在前瞻性人群中的应用<sup>[59]</sup>, 并且还存在着缺乏外部及内部验证和交叉测试的问题。目前需要多中心大样本前瞻性研究来证实组学特征模型的可靠性和鲁棒性, 以待加入临床应用<sup>[60]</sup>。

模型的可解释性及语义方面。影像组学中的纹理特征为描述病变异质性的不可视特征, 属于一阶、二阶或高阶的统计学术语, 缺乏具体临床意义。并且, 目前大多数肿瘤及非肿瘤性疾病研究得出的有意义的纹理参数大多只对应具体的预测模型和诊断效能, 多数纹理特征尚缺乏与之对应的生物学特性之间的关联, 以及其合理的病理学解释, 导致其在重复研究以及临床运用中的可解释性受限<sup>[59]</sup>。如今尚缺乏对影像组学纹理的确切术语描述和标准化语义规范, 其在运用于临床报告中可能将遭遇困难。

## 影像组学在胰腺肿瘤中的未来展望

人工智能在多种疾病的定量成像、数据提取和模

型构建等方面具有极大的应用前景,并且具有预测性、诊断性、生物性和即时疗效评估性。目前影像组学已逐渐运用于胰腺常见肿瘤的诊断、生物学行为评价、预后预测、治疗反应与疗效评估中。未来影像组学在胰腺肿瘤方面潜在的应用领域包括但不限于:胰腺横截面图像的三维虚拟重建,以提供胰腺肿瘤病灶及瘤周相关的详细、全面信息,用于精确的手术计划和放化疗评估;放射性药物的靶向分子成像,专门针对胰腺肿瘤的宿主依赖细胞——这些放射性标记剂直接针对CA19-9、PAM-4、EGFR2等分子靶向标志物<sup>[8]</sup>。我们期待影像组学与基因组学、蛋白质组学等其它各种组学结合起来,为精准医疗、个体化医疗的实现提供平台。

#### 参考文献:

- [1] Peek N, Combi C, Marin R, et al. Thirty years of artificial intelligence in medicine (AIME) conferences: A review of research themes[J]. *Artif Intell Med*, 2015, 65(1): 61-73.
- [2] 李健文, 卢光明. 人工智能在肾肿瘤影像中的应用[J]. *肿瘤影像学*, 2018, 27(4): 256-260.
- [3] Gillies RJ, Anderson AR, Gatenby RA, et al. The biology underlying molecular imaging in oncology: from genome to anatomy and back again[J]. *Clin Radiol*, 2010, 65(7): 517-521.
- [4] 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.
- [5] 马晓雯, 罗娅红. 影像组学在乳腺癌应用中的研究进展[J]. *磁共振成像*, 2018, 9(8): 637-640.
- [6] Gyawali B. Does global oncology need artificial intelligence? [J]. *Lancet Oncol*, 2018, 19(5): 599-600.
- [7] 金征宇. 人工智能医学影像应用: 现实与挑战[J]. *放射学实践*, 2018, 33(10): 989-991.
- [8] Korn RL, Rahmanuddin S, Borazanci E. Use of Precision Imaging in the Evaluation of Pancreas Cancer[M]. USA: Springer, 2019: 209-236.
- [9] Winter JM, Cameron JL, Campbell KA, et al. 1423 pancreaticoduodenectomies for pancreatic cancer: A single-institution experience [J]. *J Gastrointest Surg*, 2006, 10(9): 1199-1210.
- [10] 赵玉沛, 张太平, 曹喆. 胰腺肿瘤诊治的瓶颈与挑战[J]. *临床实用外科杂志*, 2018, 38(7): 719-721.
- [11] Zerbi A, Falconi M, Rindi G, et al. Clinicopathological features of pancreatic endocrine tumors: a prospective multicenter study in Italy of 297 sporadic cases[J]. *Am J Gastroenterol*, 2010, 105(6): 1421-1429.
- [12] Klöppel G. Classification and pathology of gastroenteropancreatic neuroendocrine neoplasms [J]. *Endocr Relat Cancer*, 2011, 18(Suppl 1): S1-16.
- [13] Bosman FT, Carneiro F, Hruban RH, et al. World health classification of tumours of the digestive system [M]. Lyon: IARC Press, 2010: 292-295.
- [14] Falconi M, Eriksson B, Kaltsas G, et al. Consensus guidelines update for the management of functional p-NETs (F-p-NETs) and non-functional p-NETs[J]. *Neuroendocrinol*, 2016, 103(2): 153-171.
- [15] Poultsides GA, Huang LC, Chen Y, et al. Pancreatic neuroendocrine tumors: radiographic calcifications correlate with grade and metastasis[J]. *Ann Surg Oncol*, 2012, 19(7): 2295-2303.
- [16] Rha SE, Jung SE, Lee KH, et al. CT and MR imaging findings of endocrine tumor of the pancreas according to WHO classification [J]. *Eur J Radiol*, 2007, 62(3): 371-377.
- [17] Humphrey PE, Alessandrino F, Bellizzi AM, et al. Non-hyperfunctioning pancreatic endocrine tumors: multimodality imaging features with histopathological correlation[J]. *Abdom Imaging*, 2015, 40(7): 2398-2410.
- [18] Kim JH, Eun HW, Kim YJ, et al. Pancreatic neuroendocrine tumour (PNET): Staging accuracy of MDCT and its diagnostic performance for the differentiation of PNET with uncommon CT findings from pancreatic adenocarcinoma[J]. *Eur Radiol*, 2016, 26(5): 1338-1347.
- [19] Li J, Lu J, Liang P, et al. Differentiation of atypical pancreatic neuroendocrine tumors from pancreatic ductal adenocarcinomas: Using whole-tumor CT texture analysis as quantitative biomarkers[J]. *Cancer Med*, 2018, 7(10): 4924-4931.
- [20] Guo C, Zhuge X, Wang Q, et al. The differentiation of pancreatic neuroendocrine carcinoma from pancreatic ductal adenocarcinoma: the values of CT imaging features and texture analysis[J]. *Cancer Imaging*, 2018, 18(1): 37.
- [21] van der Pol CB, Lee S, Tsai S, et al. Differentiation of pancreatic neuroendocrine tumors from pancreas renal cell carcinoma metastases on CT using qualitative and quantitative features[J]. *Abdom Radiol (NY)*, 2019, 44(3): 992-999.
- [22] Lin X, Xu L, Wu A, et al. Differentiation of intrapancreatic accessory spleen from small hypervascular neuroendocrine tumor of the pancreas: textural analysis on contrast-enhanced computed tomography[J]. *Acta Radiol*, 2019, 60(5): 553-560.
- [23] Guo C, Ren S, Chen X, et al. Pancreatic neuroendocrine tumor: prediction of the tumor grade using magnetic resonance imaging findings and texture analysis with 3T magnetic resonance[J]. *Cancer Manag Res*, 2019, 11(4): 1933-1944.
- [24] DOnofrio M, Ciaravino V, Cardobi N, et al. CT enhancement and 3D texture analysis of pancreatic neuroendocrine neoplasms[J]. *Sci Rep*, 2019, 9(1): 2176.
- [25] Choi TW, Kim JH, Yu MH, et al. Pancreatic neuroendocrine tumor: prediction of the tumor grade using CT findings and computerized texture analysis[J]. *Acta Radiol*, 2018, 59(4): 383-392.
- [26] Canellas R, Burk KS, Parakh A, et al. Prediction of pancreatic neuroendocrine tumor grade based on CT features and texture analysis[J]. *AJR Am J Roentgenol*, 2018, 210(2): 341-346.
- [27] Ng DZ, Goh BK, Tham EH, et al. Cystic neoplasms of the pancreas: current diagnostic modalities and management [J]. *Ann Acad Med Singapore*, 2009, 38(3): 251-259.
- [28] Brugge WR, Lauwers GY, Sahani D, et al. Cystic neoplasms of the pancreas[J]. *N Engl J Med*, 2004, 351(12): 1218-1226.
- [29] Galanis C, Zamani A, Cameron JL, et al. Resected serous cystic neoplasms of the pancreas: a review of 158 patients with recommendations for treatment[J]. *J Gastrointest Surg*, 2007, 11(7): 820-826.
- [30] Jais B, Rebours V, Malleo G, et al. Serous cystic neoplasm of the pancreas: a multinational study of 2622 patients under the aus-

- pices of the International Association of Pancreatology and European Pancreatic Club (European Study Group on Cystic Tumors of the Pancreas)[J]. *Gut*, 2016, 65(2):305-312.
- [31] Wei R, Lin K, Yan W, et al. Computer-aided diagnosis of pancreas serous cystic neoplasms: a radiomics method on preoperative MDCT images[J]. *Technol Cancer Res Treat*, 2019, 18(1):1533033818824339.
- [32] Balzano G, Zerbi A, Di Carlo V. Intraductal papillary mucinous tumors of the pancreas; incidence, clinical findings and natural history[J]. *JOP*, 2005, 6(1 Suppl):108-111.
- [33] Goh BK, Thng CH, Tan DM, et al. Evaluation of the Sendai and 2012 International Consensus Guidelines based on cross-sectional imaging findings performed for the initial triage of mucinous cystic lesions of the pancreas; a single institution experience with 114 surgically treated patients[J]. *Am J Surg*, 2014, 208(2):202-209.
- [34] Goh BK, Tan YM, Cheow PC, et al. Cystic neoplasms of the pancreas with mucin-production[J]. *Eur J Surg Oncol*, 2005, 31(3):282-287.
- [35] Goh BK, Tan YM, Yap WM, et al. Pancreatic serous oligocystic adenomas; clinicopathologic features and a comparison with serous microcystic adenomas and mucinous cystic neoplasms[J]. *World J Surg*, 2006, 30(8):1553-1559.
- [36] Adsay NV, Merati K, Basturk O, et al. Pathologically and biologically distinct types of epithelium in intraductal papillary mucinous neoplasms; delineation of an "intestinal" pathway of carcinogenesis in the pancreas[J]. *Am J Surg Pathol*, 2014, 28(7):839-848.
- [37] Tanaka M, Kobayashi K, Mizumoto K, et al. Clinical aspects of intraductal papillary mucinous neoplasm of the pancreas[J]. *J Gastroenterol*, 2005, 40(7):669-675.
- [38] Wu J, Matthaei H, Maitra A, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development[J]. *Sci Transl Med*, 2011, 3(92):92ra66.
- [39] Sohn TA, Yeo CJ, Cameron JL, et al. Intraductal papillary mucinous neoplasms of the pancreas: an updated experience[J]. *Ann Surg*, 2004, 239(6):788-797.
- [40] Koh YX, Chok AY, Zheng HL, et al. Systematic review and meta-analysis comparing the surgical outcomes of invasive intraductal papillary mucinous neoplasms and conventional pancreatic ductal adenocarcinoma[J]. *Ann Surg Oncol*, 2014, 21(8):2782-2800.
- [41] Freeny PC, Saunders MD. Moving beyond morphology: new insights into the characterization and management of cystic pancreatic lesions[J]. *Radiology*, 2014, 272(2):345-363.
- [42] Tanaka M, Chari S, Adsay V, et al. International consensus guidelines for management of intraductal papillary mucinous neoplasms and mucinous cystic neoplasms of the pancreas[J]. *Pancreatology*, 2006, 6(1-2):17-32.
- [43] Hanania AN, Bantis LE, Feng Z, et al. Quantitative imaging to evaluate malignant potential of IPMNs[J]. *Oncotarget*, 2016, 7(52):85776-85784.
- [44] Chakraborty J, Midya A, Gazit L, et al. CT radiomics to predict high-risk intraductal papillary mucinous neoplasms of the pancreas[J]. *Med Phys*, 2018, 45(11):5019-5029.
- [45] Permut JB, Choi J, Balarunathan Y, et al. Combining radiomic features with a miRNA classifier may improve prediction of malignant pathology for pancreatic intraductal papillary mucinous neoplasms[J]. *Oncotarget*, 2016, 7(52):85785-85797.
- [46] 杨尹默, 刘子文, 廖泉, 等.《胰腺癌诊治指南(2014)》热点问题解读[J]. *中华外科杂志*, 2014, 52(12):891-893.
- [47] Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014[J]. *CA Cancer J Clin*, 2014, 64(1):9-29.
- [48] Ferrone CR, Pieretti-Vanmarcke R, Bloom JP, et al. Pancreatic ductal adenocarcinoma; long-term survival does not equal cure[J]. *Surgery*, 2012, 152(3 Suppl 1):S43-49.
- [49] Gillen S, Schuster T, Meyer Zum Büschenfelde C, et al. Preoperative/neoadjuvant therapy in pancreatic cancer: a systematic review and meta-analysis of response and resection percentages[J]. *PLoS Med*, 2010, 7(4):791-802.
- [50] Tamm EP, Bhosale PR, Lee JH. Pancreatic ductal adenocarcinoma; ultrasound, computed tomography, and magnetic resonance imaging features[J]. *Semin Ultrasound CT MR*, 2007, 28(5):330-338.
- [51] Diehl SJ, Lehmann KJ, Sadick MV. Pancreatic cancer; value of dual-phase helical CT in assessing resectability[J]. *Radiology*, 1998, 206(2):373-378.
- [52] Eilaghi A, Baig S, Zhang Y, et al. CT texture features are associated with overall survival in pancreatic ductal adenocarcinoma-a quantitative analysis[J]. *BMC Med Imaging*, 2017, 17(1):38.
- [53] Yun G, Kim YH, Lee Y, et al. Tumor heterogeneity of pancreas head cancer assessed by CT texture analysis; association with survival outcomes after curative resection[J]. *Sci Rep*, 2018, 8(1):7226.
- [54] Sandrasegaran K, Lin Y, Asare-Sawiri M, et al. CT texture analysis of pancreatic cancer[J]. *Eur Radiol*, 2019, 29(3):1067-1073.
- [55] Cheng SH, Cheng YJ, Jin ZY, et al. Unresectable pancreatic ductal adenocarcinoma; role of CT quantitative imaging biomarkers for predicting outcomes of patients treated with chemotherapy[J]. *Eur J Radiol*, 2019, 113(4):188-197.
- [56] Campbell DH, Barrett M, Ramanathan RK, et al. Quantitative textural analysis (QTA) in CT imaging; identifying markers for genetic instability and overall survival in cohort of previously treated metastatic pancreatic cancer (mPC) [J]. *Cancer Res*, 2014, 74(19 S):1885.
- [57] Swaminath A, Chu W. Stereotactic body radiotherapy for the treatment of medically inoperable primary renal cell carcinoma; current evidence and future directions[J]. *Can Urol Assoc J*, 2015, 9(3):275-280.
- [58] Cozzi L, Comito T, Fogliata A, et al. Computed tomography based radiomic signature as predictive of survival and local control after stereotactic body radiation therapy in pancreatic carcinoma[J]. *PLoS One*, 2019, 14(1):e0210758.
- [59] 唐磊. 进展期胃癌的影像学评估:从图像到大数据影像组学[J]. *中华胃肠外科杂志*, 2018, 21(10):1106-1112.
- [60] Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction[J]. *Radiology*, 2018, 286(3):800-809.