❖胸部影像学

Comparison on convolutional neural network classification model and radiologists in differentiating invasive lung adenocarcinoma

YIN Ke[1], ZHANG Jiuquan[1], WU Jianlin[2*], BA Wenjuan[3], LIN Lin[2], SHEN Jing[2], XIONG Jingtong[4], ZHANG Weijie[5]

(1. *Department of Radiology, Chongqing University Cancer Hospital, Chongqing* 400030, *China*; 2. *Department of Radiology, Affiliated Zhongshan Hospital of Dalian University, Dalian* 116001, *China*; 3. *Department of Radiology, Affiliated Hospital of Yangzhou University, Yangzhou* 225003, *China*; 4. *Department of Radiology, the Second Hospital of Dalian Medical University, Dalian* 116023, *China*; 5. *Shaanxi Weinan Shenzhou Dexin Medical Imaging Technology Co., Ltd., Weinan* 714099, *China*)

[**Abstract**] **Objective** To compare the diagnostic efficiency of convolutional neural network (CNN) classification model based on chest CT and attending physicians of radiology in distinguishing pre-invasive lesions (PIL) and invasive adenocarcinoma (IA) manifested as lung ground glass nodule (GGN). **Methods** Chest CT of 1 086 patients with 1 214 pathologically confirmed PIL or IA were retrospectively analyzed. GGN were divided into training group (n=1 092) and verification group (n=122) at the ratio of 9∶1. Isotropic preprocessing of the training data and random sampling and filling of 3D image blocks were performed, so as to establish CNN classification model to identify PIL and IA in GGN and verify using verification group. The verification group was classified by 2 radiologists (physician 1 and 2) to identify PIL or IA manifested as lung GGN. Receiver operating characteristic (ROC) curve was used to compare the efficacy of CNN classification model and 2 radiologists in classifying the verification group. **Results** ROC curve showed that the area under the curve (AUC) of CNN classification model and physician 1 and 2 was 0.866, 0.742 and 0.769, respectively. The diagnostic sensitivity of CNN classification model was 84.81%, significantly higher than that of physician 1 (67.09%, χ^2=11.352, P<0.001) and physician 2 (74.68%, χ^2=18.473, P<0.001), while the specificity was not statistically different between physician 1 and 2 (both P>0.05). **Conclusion** The efficacy of CNN classification model for classifying PIL and IA manifested as lung GGN was higher than that of attending physicians of radiology.

[**Keywords**] lung neoplasms; adenocarcinoma of lung; neoplasm invasiveness; artificial intelligence

DOI: 10.13929/j.issn.1003-3289.2021.09.015

对比卷积神经网络分类模型与放射科医师鉴别浸润性肺腺癌的效能

尹　柯[1], 张久权[1], 伍建林[2*], 巴文娟[3], 林　琳[2], 沈　晶[2], 熊婧彤[4], 张伟杰[5]

(1. 重庆大学附属肿瘤医院影像科, 重庆　400030; 2. 大连大学附属中山医院放射科, 辽宁 大连　116001; 3. 扬州大学附属医院放射科, 江苏 扬州　225003; 4. 大连医科大学附属第二医院放射科, 辽宁 大连　116023; 5. 陕西渭南神州德信医学成像技术有限公司, 陕西 渭南　714099)

[**摘　要**] **目的**　比较基于胸部 CT 建立的卷积神经网络(CNN)分类模型与放射科主治医师分类磨玻璃结节(GGN)中

[**基金项目**] 重庆市科卫联合医学科研项目(2019ZDXM007)。

[**第一作者**] 尹柯(1993—),男,四川绵阳人,硕士,医师。研究方向:胸部影像诊断。E-mail: yinke93@163.com

[**通信作者**] 伍建林,大连大学附属中山医院放射科,116001。E-mail: cjr.wujianlin@VIP.163.com

[**收稿日期**] 2020-12-21　[**修回日期**] 2021-06-20

的浸润前病变(PIL)与浸润性腺癌(IA)的效能。**方法** 回顾性分析 1 086 例经病理确诊 PIL 或 IA 患者的胸部 CT,共 1 214 枚 GGN,按 9∶1 比例将其分为训练组(n=1 092)及验证组(n=122)。对训练组数据进行各向同性预处理和 3D 图像块随机采样和填充,建立 CNN 分类模型,将训练组 GGN 分类为 PIL 或 IA,并于验证组进行验证。由 2 名放射科主治医师(医师 1 和 2)对验证组 GGN 进行分类,鉴别 PIL 与 IA。采用受试者工作特征(ROC)曲线比较 CNN 分类模型与 2 名放射科主治医师对验证组进行分类的效能。**结果** ROC 曲线显示,CNN 分类模型、医师 1 及 2 鉴别 PIL 与 IA 的曲线下面积(AUC)分别为 0.866、0.742 及 0.769;CNN 分类模型诊断敏感度(84.81%)显著高于医师 1(67.09%,χ^2=11.352,P<0.001)、2 (74.68%,χ^2=18.473,P<0.001),而特异度与医师 1、2 差异均无统计学意义(P 均>0.05)。**结论** CNN 分类模型鉴别 GGN 中的 PIL 与 IA 的效能优于放射科医师。

[**关键词**] 肺肿瘤;肺腺癌;肿瘤侵袭性;人工智能
[**中图分类号**] R734.2;R814.42 [**文献标识码**] A [**文章编号**] 1003-3289(2021)09-1338-05

2011 年,国际肺癌研究协会、美国胸科学会和欧洲呼吸学会[1]基于高分辨率 CT-病理学相关性研究提出新的肺腺癌国际多学科分类体系,以供临床医师预测肺腺癌组织病理学亚型、患者预后及规划临床干预措施[2]。近年来,胸部影像学深度学习(deep learning,DL)图像分析成为研究热点,主要用于评估 CT 所检出的良、恶性肺结节[3-6],并日渐成熟[8-9],现阶段面临的挑战是利用 DL 分类不同性质肺结节。微浸润腺癌(minimally invasive adenocarcinoma,MIA)与不典型腺瘤性增生(atypical adenomatous hyperplasia,AAH)、原位腺癌(adenocarcinoma in situ,AIS)同为浸润前病变(pre-invasive lesion,PIL)[10-12];对 PIL 均可行亚肺叶切除术,术后 5 年生存率近 100%,而对浸润性腺癌(invasive adenocarcinoma,IA)则推荐行肺叶切除加淋巴结清扫或取样术,术后 5 年生存率 60%～88%,故术前有效区分两者对选择手术方式及改善患者预后具有重要意义[9,13-14]。本研究比较基于胸部 CT 建立的卷积神经网络(convolutional neural network,CNN)分类模型与放射科主治医师鉴别磨玻璃结节(ground glass nodule,GGN)中的 PIL 与 IA 的效能。

1 资料与方法

1.1 一般资料 回顾性分析 2013 年 1 月—2018 年 12 月 1 086 例经手术病理确诊 PIL 或 IA 患者的胸部 CT;男 414 例,女 672 例,年龄 18～78 岁,平均(55.5±12.2)岁;共检出 1 214 枚 GGN,其中 974 例检出 1 枚、96 例检出 2 枚、16 例检出 3 枚。纳入标准:①术前 1 个月内接受胸部薄层 CT 检查;②组织病理学证实为 PIL 或 IA;③GGN 最长径≤30 mm。排除标准:①术前接受化学或放射治疗;②图像质量差及存在肺纤维化。按 9∶1 比例将所有 GGN 分为 CNN 分类模型训练组(1 092 枚)及验证组(122 枚)。训练组包括 490 枚 AAH、AIS 及 MIA 和 602 枚 IA;验证组包括 43 枚 AAH、AIS 及 MIA 和 79 枚 IA。

1.2 仪器与方法 采用 Siemens Somatom Definition 双源 CT、Siemens Somatom Definition AS 64 层 CT 或 Siemens Somatom Sensation 16 排 CT 仪,于患者吸气末行胸部扫描,范围自胸廓入口至肺底、包含肾上腺,管电压 100～140 kV,管电流 200～280 mAs,层厚及间隔均为 5 mm,FOV 350 mm×350 mm,矩阵 512×512;扫描结束后采用骨算法将原始图像重建为层厚 1 mm 图像。

1.3 图像处理 将原始数据转换为各向(即轴位、冠状位、矢状位)同性格式,即采用 1×1×1 mm^3 分辨率和 3 次样条插值生成各向同的 DICOM 图像,以降低 CT 扫描变化对算法学习的影响;并以多个分辨率和大小不同但中心相同的 3D 立体结节图像块分析各向同性 DICOM 图像。

1.4 训练 CNN 分类模型 采用开源软件 Keras 库构建 CNN 模型,其多尺度 CNN 架构由 3 层分支组成(图 1),各分支图像输入中心相同而大小不同,分别为 72 mm×72 mm×72 mm、36 mm×36 mm×36 mm 及 18 mm×18 mm×18 mm。通过池化和/或上采样操作将所有图像统一调整为 72×72×72 体素。将归一化后图像分别传递至骨干网络进行级联,再传递至全连接层和线性整流单元,通过归一化指数函数激活输出层,获得每个类别的概率。骨干网络为 CNN 架构的关键组成部分,常以一系列具有不同数目滤波器的卷积块组成基础骨干网络,建立多个密集块组成的密集型骨干网络(图 2A),每个骨干网络由 3 个密集块组成,每个密集块含 4f 个 1×1×1 卷积核和 f 个 3×3×3 卷积核,采用“瓶颈”技术串联卷积核的输入特征和输出结果(图 2B)。以 CNN 分类模型对验证组进行验证。

1.5 放射科医师评估效能 由 1 名具有 3 年影像学诊断经验的住院医师整理并提供验证组 GGN 在 CT 图像中的位置列表;再由分别具有 6 年及 8 年影像学诊

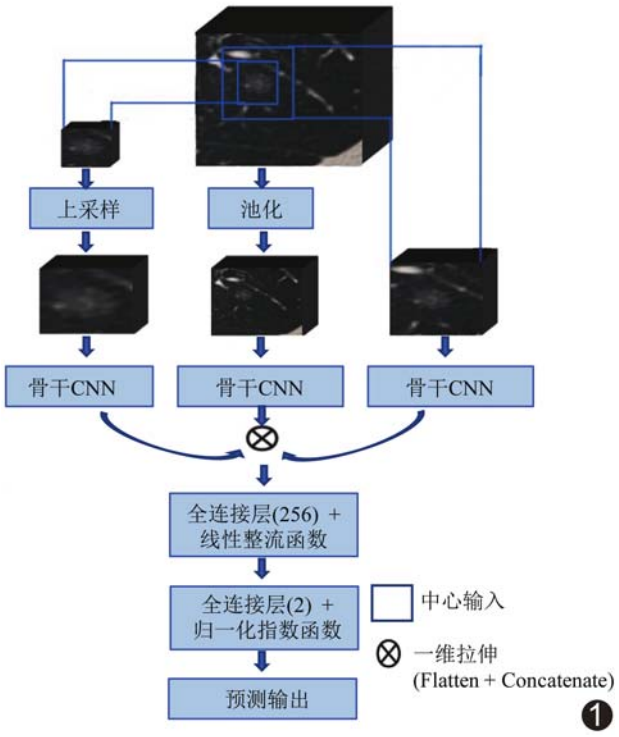

图 1　多尺度 CNN 架构由 3 层分支组成，分支以给定结节为中心输入图像，具有不同尺寸，以最大池化和上采样操作将其调整为统一的 72×72×72 体素

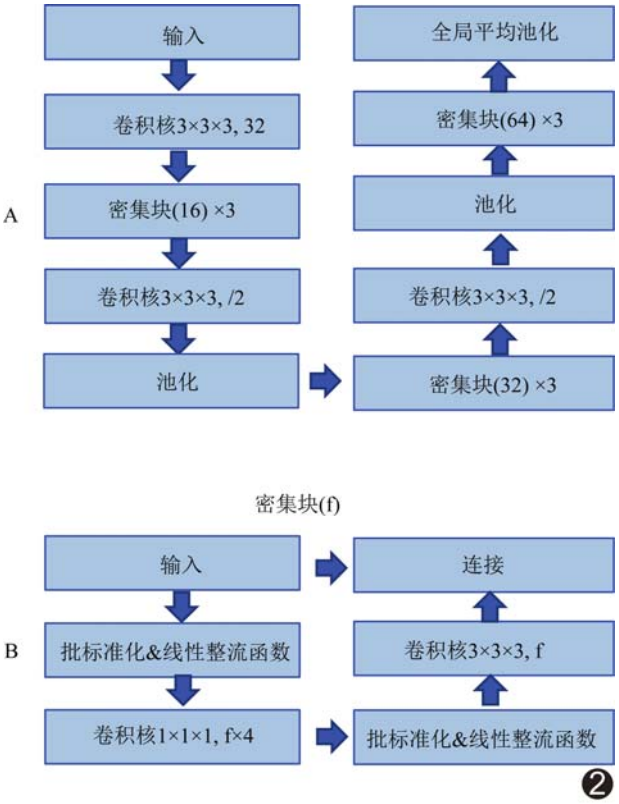

图 2　骨干网络示意图　A. 密集型骨干网络示意图；B. 密集块构成示意图

断经验的主治医师 1 及 2（非胸部影像学专长）根据结节位置独立评估其为 PIL 或 IA 的可能性，并设定五个信心级别，即：肯定是 PIL、可能是 PIL、不确定、可能是 IA 以及肯定是 IA。

1.6　统计学分析　采用 SPSS 20.0 统计分析软件。将“肯定是 IA”和“可能是 IA”置信水平视为阳性结果，将其他水平视为阴性结果。绘制受试者工作特征（receiver operating characteristic，ROC）曲线，比较 CNN 分类模型及 2 名放射科医师鉴别验证组 PIL 与 IA 的效能，获得其曲线下面积（area under the curve，AUC）、敏感度、特异度和总体分类准确率。以 χ^2 检验比较 CNN 分类模型与放射科医师诊断敏感度及特异度的差异。$P<0.05$ 为差异有统计学意义。

2　结果

ROC 曲线（图 3）结果显示，CNN 分类模型、医师 1 及 2 鉴别验证组 PIL 与 IA 的 AUC 分别为 0.866、0.742 及 0.769，敏感度分别为 84.81%、67.09% 及 74.68%，特异度分别为 88.37%、81.40% 及 79.07%。CNN 分类模型的诊断敏感度显著高于医师 1（$P<0.001$）、2（$P<0.001$）；CNN 分类模型诊断的特异度与医师 1、2 差异均无统计学意义（P 均 >0.05），见表 1。

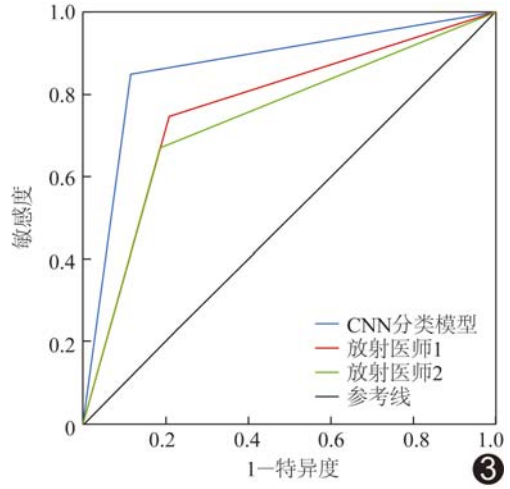

图 3　CNN 分类模型与医师 1、2 鉴别 PIL 与 IA 的 ROC 曲线

3　讨论

治疗 PIL 的方式与 IA 有所不同，故早期发现 GGN 并以无创方法有效鉴别其为 PIL 或 IA 对患者生存及预后具有重要价值[10-12]。本研究建立 CNN 分类模型，将 GGN 分类为 PIL 与 IA，其 AUC、总体分类准确率及敏感度均高于 2 名放射科医师，提示其诊断效能可能较初、中级放射科医师更高，具有良好临床应

表 1 CNN 分类模型与 2 名医师鉴别验证组 GGN 为 PIL 或 IA 的 ROC 曲线结果

分类方法	敏感度(%)	特异度(%)	AUC 值	准确率(%)
CNN 分类模型	84.81	88.37	0.866	86.06
医师 1	67.09	81.40	0.742	72.13
医师 2	74.68	79.07	0.769	76.23
χ^2 值	11.352*	1.710*	—	—
	18.473#	1.243#	—	—
P 值	<0.001*	0.191*	—	—
	<0.001#	0.265#	—	—

注：*：CNN 分类模型与医师 1 比较；#：CNN 分类模型与医师 2 比较

用前景。本研究构建 CNN 分类模型的诊断敏感度及总体分类准确率比传统机器学习算法[11]更佳，逻辑回归敏感度为 81.1%、总体分类准确率为 81.5%，随机森林的敏感度为 80.7%、总体分类准确率为 83.0%，决策树的敏感度为 72.8%、总体分类准确率为 76.7%，而 AdaBoost 算法的敏感度 74.4%、总体分类准确率 82.1%。

既往研究基于高分辨率 CT[10,12,15-16]或放射组学[17-19]预测早期肺腺癌浸润程度，其所提取的 CT 征象及放射组学信息均依赖人为预处理、CT 征象及放射组学信息分析，导致再现性降低。基于 CNN 的 DL 分类模型直接来自 CT 图像中学习预测特征的成像信息，简化多级分析流程，且可重复性更高。本研究中 CNN 分类模型的总体分类准确率与上述研究相似或更高，表明 CNN 分类模型可于未预先提取特定征象或信息的情况下实现较准确的 PIL 与 IA 分类，具有良好的临床实用性、重复性和普及性。此外，本研究结果显示，CNN 分类模型的敏感度显著高于 2 名放射科主治医师，提示 CNN 分类模型对于准确评估 GGN 中的 IA 表现更佳，更有利于临床及时干预和处理 IA 患者。本研究中的 2 名放射科医师均为中级职称，其专长并非胸部影像学诊断，亦未曾专注于研究肺磨玻璃样腺癌或接受相关专业培训，仅根据既往工作经验自行分析判断分类结果，导致诊断敏感度较低。采用 CNN 分类算法不仅可充分利用多维度 DICOM 数据信息，还能通过学习肺腺癌浸润程度分类中更多未知关键特征信息来充当记忆网络，从而对 GGN 分类更加稳定，并可及时调整学习策略而更加灵活智能，对分类较疑难 GGN 亦能保持较高准确性，弥补青年医师诊断经验的不足。

本研究 CNN 分类算法的主要缺陷是数据本身及数据量不足问题。理想情况下，应采用相同扫描参数采集所有 CT 数据，但在实践中难以实现。为此本研究在算法训练过程中采用多种数据处理方式，如翻转、缩放、平移和旋转等，以尽量降低数据异质性。此外，本研究数据仅来自两家医疗机构，具有一定局限性，有待多中心采集数据，进一步增强模型的泛化能力和鲁棒性，提高 CNN 分类模型的敏感度和总体准确率。

总之，CNN 分类模型对 GGN 分类为 PIL 与 IA 的效能优于放射科主治医师，有望为临床鉴别诊断 GGN 提供新的智能化辅助手段。

[参考文献]

[1] TRAVIS W D, BRAMBILLA E, NOGUCHI M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma[J]. J Thorac Oncol, 2011,6(2):244-285.

[2] MOON Y, LEE K Y, PARK J K. The prognosis of invasive adenocarcinoma presenting as ground-glass opacity on chest computed tomography after sublobar resection[J]. J Thorac Dis 2017,9(10):3782-3792.

[3] CIOMPI F, de HOOP B, van RIEL S J, et al. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box[J]. Med Image Anal, 2015, 26(1):195-202.

[4] NIBALI A, HE Z, WOLLERSHEIM D. Pulmonary nodule classification with deep residual networks[J]. Int J Comput Assist Radiol Surg, 2017,12(10):1799-1808.

[5] SONG Q, ZHAO L, LUO X, et al. Using deep learning for classification of lung nodules on computed tomography images[J]. J Healthc Eng, 2017,2017:8314740.

[6] CIOMPI F, CHUNG K, van RIEL S J, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning[J]. Sci Rep, 2017,7:46479.

[7] HUA K L, HSU C H, HIDAYATI S C, et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique[J]. Onco Targets Ther, 2015,8:2015-2022.

[8] SUN W, ZHENG B, QIAN W. Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis[J]. Comput Biol Med, 2017, 89:530-539.

[9] LIU S, WANG R, ZHANG Y, et al. Precise diagnosis of intraoperative frozen section is an effective method to guide resection strategy for peripheral small-sized lung adenocarcinoma[J]. J Clin Oncol, 2016,34(4):307-313.

[10] 尹柯，伍建林，邱太春. 高分辨率 CT 征象诊断浸润性肺腺癌的模型建立[J]. 中国医学影像学杂志，2019,27(11):824-828.

[11] MEI X, WANG R, YANG W, et al. Predicting malignancy of pulmonary ground-glass nodules and their invasiveness by

random forest[J]. J Thorac Dis, 2018,10(1):458-463.
[12] 赵娇,李建华,费佳,等.肺磨玻璃样结节:有助于预判肺腺癌浸润性的 CT 征象[J].放射学实践,2018,33(4):57-62.
[13] MOON Y, LEE K Y, PARK J K. The prognosis of invasive adenocarcinoma presenting as ground-glass opacity on chest computed tomography after sublobar resection[J]. J Thorac Dis 2017,9(10):3782-3792.
[14] KODAMA K, HIGASHIYAMA M, OKAMI J, et al. Oncologic outcomes of segmentectomy versus lobectomy for clinical T1a N0 M0 non-small cell lung cancer[J]. Ann Thorac Surg, 2016,101:504-511.
[15] 尹柯,巴文娟,汤敏,等.多层感知器模型预测纯磨玻璃结节肺腺癌浸润程度[J].中国医学影像技术,2020,36(11):1652-1656.
[16] 邱太春,尹柯,汝晓双,等.基于 CT 征象的 Logistic 回归模型预测纯磨玻璃样肺腺癌恶性浸润程度的临床研究[J].影像诊断与介入放射学,2019,28(1):54-59.
[17] ALPERT J B, RUSINEK H, KO J P, et al. Lepidic predominant pulmonary lesions (LPL): CT-based distinction from more invasive adenocarcinomas using 3D volumetric density and first-order CT texture analysis[J]. Acad Radiol, 2017, 24(12):1604-1611.
[18] HWANG I P, PARK C M, PARK S J, et al. Persistent pure ground-glass nodules larger than 5 mm: Differentiation of invasive pulmonary adenocarcinomas from preinvasive lesions or minimally invasive adenocarcinomas using texture analysis[J]. Invest Radiol, 2015,50(11):798-804.
[19] YUNLANG S, LEI Z, HUIYUAN Z, et al. The predictive value of CT-based radiomics in differentiating indolent from invasive lung adenocarcinoma in patients with pulmonary nodules[J]. Eur Radiol, 2018,28(12):5121-5128.

2020 版中国科技期刊引证报告相关数据
——《中国医学影像技术》

由中国科学技术信息研究所主持的“2019 中国科技论文统计结果发布会”于 2020 年 12 月在北京举行。《中国医学影像技术》杂志在《2020 版中国科技期刊引证报告(核心版)》的相关数据为:

1 文献来源量:399 篇;
2 基金论文比:0.49;
3 核心总被引频次:3031;
4 核心影响因子:1.189;
5 学科扩散指标:17.21;
6 学科影响指标:1.00;
7 综合评价总分:72.9。