

陈新花, 黄晓旗, 李建龙, 等. 基于两种 AI 软件对肺结节的诊断价值评价研究[J]. CT 理论与应用研究, 2023, 32(4): 493-499. DOI:10.15953/j.ctta.2022.087.
CHEN X H, HUANG X Q, LI J L, et al. Evaluation of diagnostic value of pulmonary nodules based on two AI softwares[J]. CT Theory and Applications, 2023, 32(4): 493-499. DOI:10.15953/j.ctta.2022.087. (in Chinese).

基于两种 AI 软件对肺结节的诊断价值评价研究

陈新花¹, 黄晓旗¹, 李建龙^{1✉}, 郭佑民²

1. 延安大学附属医院, 陕西 延安 716000
2. 西安交通大学第一附属医院, 西安 710000

摘要: 目的: 探讨两种 AI 软件在 $\geq 5\text{mm}$ 肺结节中的临床应用价值。方法: 选取 2021 年 6 月至 2021 年 10 月延安大学附属医院体检肺结节患者 92 例 (共 483 个结节)。AI 软件检测到的结节由影像学医师进行评估并记录其个数及结节类型; 人工阅片由两名高年资影像学医师进行视觉评估, 并以此作为识别结节的金标准。计算两个软件的检出率、假阳性率和假阴性率, 评价两种 AI 软件的结节诊断价值; 卡方检验和 Fisher 精确检验来比较不同软件与金标准之间的差异; 最后, 评价两种 AI 软件联合对肺结节的诊断价值。结果: 软件 A 和软件 B 结节检出率为 92.1% 和 87.0%; 软件 A 与人工阅片的吻合度一般 ($\text{Kappa} = 0.213$), 软件 B 与人工阅片的吻合度较弱 ($\text{Kappa} = 0.150$); 软件 A 相比人工阅片对实性结节和钙化结节的检出有统计学差异; 软件 B 相比人工阅片纯磨玻璃结节的检出有统计学差异; 联合两种 AI 软件结节的检出率为 97.1%, 两种软件的联合与人工阅片比, 结节类型的检出没有统计学差异。两种 AI 软件联合与人工阅片的吻合度较好 ($\text{Kappa} = 0.439$)。结论: 两种 AI 软件联合会提高结节诊断及分类分析的能力; 推荐联合诊断的方法用于临床, 也为进一步提升 AI 软件数据集的同质化管理提供佐证。

关键词: 人工智能; 计算机体层成像; 肺结节; 检测

DOI:10.15953/j.ctta.2022.087 中图分类号: R814; TP 317 文献标识码: A

随着低剂量 CT 的广泛普及, 肺结节的检出率逐步提升。部分肺结节可能会进一步进展为肺癌^[1], 肺癌已经变成危害人类健康最常见的恶性肿瘤之一^[2]。肺结节的早期筛查、定期随访有利于肺癌的早期干预, 可以明显延长患者的生存年限。但是大量结节增加了影像医师的工作量, 肺结节易漏诊及误诊^[3]。人工智能 (artificial intelligence, AI) 可以减轻医生工作量, 减少漏诊率^[4-5]。但由于 AI 软件众多, 其识别结节的准确性及同质性尚无明确定论^[6]。

本研究旨在探讨两种 AI 软件在 $\geq 5\text{mm}$ 肺结节检测中临床应用价值, 寻找不同人工智能软件分析结节类型的准确指标, 观察联合两种软件是否能提高结节检测及分类分析的能力。这也为软件制造商需要进一步加强软件数据集的同质化管理提供证据。

1 资料与方法

1.1 研究对象

回顾性收集 2021 年 6 月至 2021 年 10 月在延安大学附属医院检出肺结节的 92 例患者, 共筛查出肺结节 483 个。纳入标准: ① 符合《胸部 CT 肺结节数据集构建及质量控制专家共识 (2021 版)》^[7] 相关诊断标准; ② 至少包含 1 枚 $\geq 5\text{mm}$ 肺结节; ③ 排除有肺炎和运动伪影的图像。

1.2 CT 扫描技术

所有检查由联影 UCT-760 128 层进行扫描。扫描过程中, 患者取仰卧位, 从肺尖到肺底进行全肺扫描。扫描参数: 管电压 120 kV, 自动管电流 80 mAs, 准直 40 mm, 螺距 1.0875, 转速 0.5 s/r, 旋转时间 0.5 s, 最小层厚 0.625 mm, 矩阵 1024×1024 , 标准算法重建。

收稿日期: 2022-05-18。

1.3 肺结节检测

使用人工智能软件 A(图 1)、软件 B(图 2) 进行肺结节的自动检测。将 92 例胸部 CT 原始数据传输至工作站, 软件自动批量进行肺结节的识别和标记。软件 A(型号: MIDS-PNAB V1.2.0.1), 算法由 2D 与 3D 深度卷积网络^[8]组成; 软件 B(型号: SW001.001), 主要由计算机辅助诊断(computer aided diagnosis, CAD)算法组成, 包括 5 大主要部分: 肺分割、候选结节的产生、特征提取、分类、相关参数计算。



图 1 软件 A 肺结节识别界面

Fig.1 The pulmonary nodule recognition interface of software A



图 2 软件 B 肺结节识别界面

Fig.2 The pulmonary nodule recognition interface of software B

1.4 结节标准认定

由两名高年资影像学医师(从事影像诊断工作 10 年以上)对 92 例肺结节患者的胸部 CT 原始数据进行诊断, 参照《胸部 CT 肺结节数据集构建及质量控制专家共识(2021 版)》^[7]诊断标准进行识别, 以两人的一致性意见作为结节参考标准。意见不统一时, 由高级职称医师组织共同讨论, 以此作为诊断金标准^[9]。

在人工智能软件中, 结合 CT 平扫和三维重建来观察结节, 以此来确定 AI 标记结果是否与肺内结节的定义一致, 从而判断结节的数量和类型。结节根据密度的不同分为实性结节、磨玻璃结节(ground-glass nodule, GGN)、钙化结节(图 3), 其中磨玻璃结节包括纯磨玻璃结节(pure ground-glass nodule, pGGN)和部分实性结节(partial solid nodule, PSN)。纯磨玻璃结节指密度均匀一致的磨玻璃样改变, 在胸部 CT 上, 该结节中能够观察到支气管、血管的形态和走向; 部分实性结节也称为混合磨玻璃结节, 指病灶遮盖部分肺实质的结节, 其为恶性结节的可能性最大^[10-11]。

1.5 统计学方法

采用 SPSS 26.0 进行统计学分析。符合正态分布的计量资料以($\bar{x} \pm s$)表示, 不符合正态分布的计量资料用 $M(Q_1, Q_3)$ 表示。计数资料用 % 表示, 两种 AI 软件与人工阅片比较采用卡方检验或 Fisher 确切概率法; AI 结节检出率与人工阅片一致性评价采用 Kappa 检验; 分别计算两种软件及两种软件联合对于肺结节的检出率、假阴性率、假阳性率。 $P < 0.05$ 为差异具有统计学意义。

2 结果

2.1 不同类型结节分布

经两位高年资影像医师人工阅片, 92 例患者胸部 CT 所检测出的结节共 483 个, 其中实性结节 295 个(61.1%), PSN 6 个(1.24%), mGGN 68 个(14.1%), 钙化结节 114 个(23.6%, 表 1)。

2.2 两种 AI 软件对 483 个肺结节检出情况的比较

软件 A 结节检出率 92.1%, 软件 A 与人工阅片比较有统计学差异, 软件 A 与人工阅片的一致性一般(Kappa 值 = 0.213)。软件 B 的结节检出率 87.0%, 软件 B 与人工阅片比较有统计学差异(表 2), 软件 B 和人工阅片的一致性较弱(Kappa 值 = 0.150)。软件 A 假阳性率较高(图 4), 软件 B 假阴性率较高(图 5)。

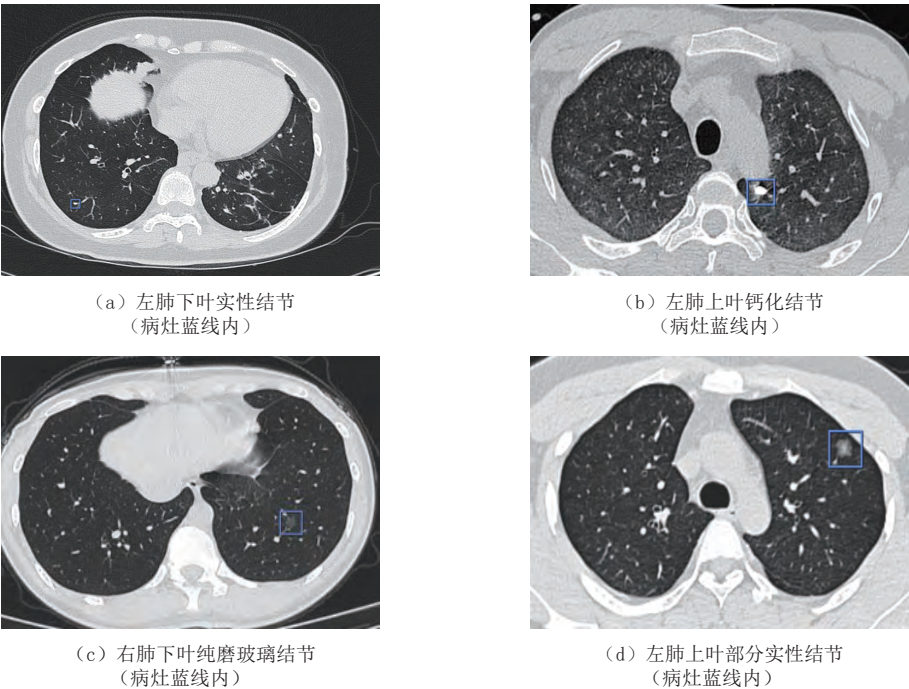


图 3 肺结节的不同类型

Fig.3 Different types of pulmonary nodules

表 1 两种软件及两种软件联合对肺结节的检出情况

Table 1 The detection of pulmonary nodules by two separate kinds of software and their combined effect

阅片方法	检出情况		
	检出率/%	假阳性率/%	假阴性率/%
软件 A	92.1	3.52	7.87
软件 B	87.0	2.69	13.0
软件 A 联合软件 B	97.1	1.04	2.91

表 2 两种 AI 软件及两种 AI 软件的联合对结节检出率的差异

Table 2 The differences in nodule detection rate between the two artificial intelligence software and their combined effect

软件	N (%)	统计检验		
		χ^2	P	Kappa
软件 A-人工阅片	445 (92.1)	25.519	0.000	0.213
软件 B-人工阅片	420 (87.0)	16.472	0.000	0.150
软件 A 联合软件 B-人工阅片	469 (97.1)	102.697	0.000	0.439

2.3 两种 AI 软件对不同类型结节检出情况比较

两种软件对于实性结节及磨玻璃结节的检出率有统计学差异（表 3）；软件 A 与人工阅片比实性结节和钙化结节的检出有统计学差异；软件 B 和人工阅片在磨玻璃结节的检测上有统计学差异（表 4）。

2.4 两种 AI 软件对于肺结节的联合检测

两种 AI 软件的联合增加了不同类型肺结节的检出率，尤其是实性结节和磨玻璃结节。两种软件的联合使不同类型结节的检出率与人工阅片相比无统计学差异（表 4），两种软件的联合与人工阅片

比结节检出数无统计学差异(表2),两种软件的联合与人工阅片之间的一致程度优于单独使用一种软件(Kappa = 0.439)。

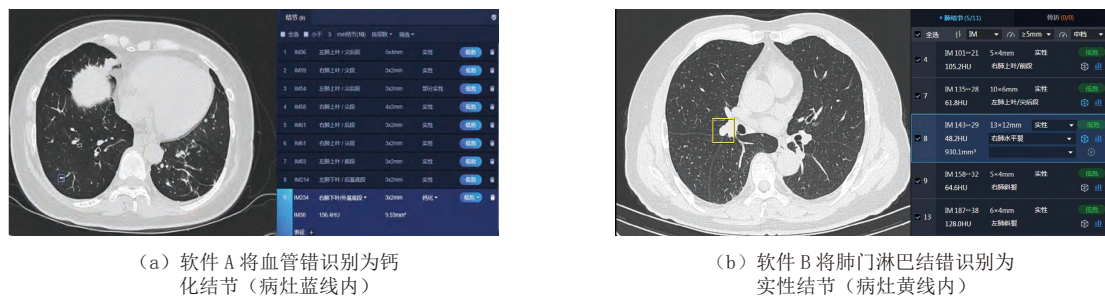


图 4 两种软件识别的假阳性结节
Fig.4 False positive nodules identified by the two software

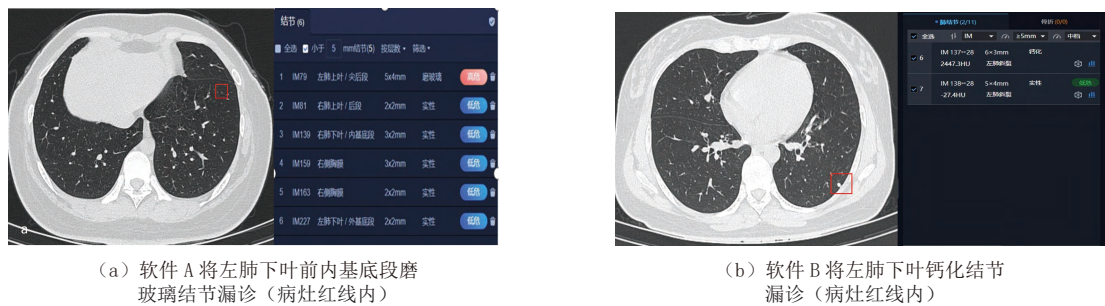


图 5 两种软件识别的假阴性结节
Fig.5 False negative nodules identified by the two software

表 3 两种 AI 软件对不同类型结节的检出情况
Table 3 The detection of different types of nodules by the two artificial intelligence software

变量	N	结节检测软件		P
		软件 A (%)	软件 B (%)	
实性结节	295	279 (94.6)	254 (86.1)	0.030
部分实性结节	6	6 (100.0)	3 (50.0)	0.000
纯磨玻璃结节	68	56 (82.4)	63 (92.6)	0.079
钙化结节	114	104 (91.2)	100 (87.7)	0.098

表 4 两种 AI 软件及两种 AI 软件联合识别不同类型结节的差异
Table 4 The differences between the two artificial intelligence software and their combined effect in identifying different types of nodules

变量	实性结节		部分实性结节		纯磨玻璃结节		钙化结节	
	χ^2	P	χ^2	P	χ^2	P	χ^2	P
软件 A-人工阅片	10.693	0.001	—	0.538	0.030	0.862	14.809	0.000
软件 B-人工阅片	0.800	0.371	—	0.182	6.741	0.009	0.556	0.456
软件 A 联合软件 B-人工阅片	0.381	0.537	—	0.464	0.271	0.603	0.000	1.000

注:—为 Fisher 确切概率法没有相应的统计量。

3 讨论

CT 的广泛使用使肺结节的检出率增加,但随着图像数量的增多,影像医师易受主观因素影响而

容易对某些结节漏诊及误诊^[12]。以 GGN 为主的肺癌具有“惰性”发展、远处转移少的特点,且预后良好,5 年生存率 100%^[13-14]。所以,GGN 的精确识别对于患者后续诊治非常重要。现有的 GGN 管理基于结节大小、特征和生长速度。AI 软件能够有效筛查肺结节,协助医师发现早期肺癌并根据结节的大小及结节类型制定合理的随访时间并采取不同的干预措施。但由于日常工作中用于辅助诊断的 AI 软件不同,肺结节检测的准确性存在一定差异^[6]。本研究发现两种 AI 软件对于不同类型的结节诊断价值不同,尤其是实性结节及磨玻璃结节。基于此,本研究进一步研究了两种 AI 软件的联合诊断价值。

本研究发现两种 AI 软件联合使用会筛选出更多的阳性结节,减少了结节的漏诊,虽然在一定程度上增加了时间成本,但以往研究表明 AI 较人工阅片用时明显缩短,每张胸片平均用时约为 0.28 s^[15];即使在两种 AI 软件同时阅片的情况下,工作效率也较人工阅片明显提高,本文主要探讨了 AI 软件在 ≥ 5 mm 肺结节的检测价值,两种软件的使用会增加结节的检出率,尤其是实性结节及磨玻璃结节,这更有助于早期肺癌的检出。因为只有在结节检出率提升的情况下,才能进一步对结节的良恶性性质进行判别。

李欣菱等^[16]及刘娜等^[17]的研究证明肺结节 AI 检测软件的假阳性率较高,这与我们的研究结果基本相符。本研究软件 A 将 17 个结节错识别为阳性结节,其中 6 个(35%)表现为增粗、聚集的血管^[16],7 个(41%)表现为肺内纤维索条影、瘢痕和实变影以及局限性肺不张,1 个(6%)表现为肺门旁淋巴结,3 个(18%)表现为突向肺野内的骨性结构、胸膜褶皱。两种 AI 软件对结节的检出具有差异性,可能由于样本量不足导致,也可能由于不同深度学习模型算法不同,软件 A 基于 2D 与 3D 卷曲神经网络算法,软件 B 基于 CAD 算法。王祥等^[18]研究表明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,但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。因而后期软件需要进一步优化算法,加强各个软件对于结节识别的训练,保证不同软件对同一结节识别的同质性,便于临床对结节的随访及决策,避免临床工作中患者的过度检查,减轻患者的经济及心理负担。

左玲子等^[19]研究表明引起纯磨玻璃密度假阳性肺结节的常见原因有胸膜结节、索条影、血管增粗、分叉和肺小叶结构。本研究中软件 B 将 3 例部分实性结节被错识别为纯磨玻璃结节,这可能是由于结节本身性质、成像参数、AI 软件诊断性能所导致的结果。

两种软件检出率较人工阅片比还有一定差距,因此,需要软件扩大样本量,减少漏诊及误诊,提高 AI 诊断的准确性^[19]。最后,在本研究中发现两种软件的联合使用可以明显提高不同类型肺结节的识别及检出率,特别是磨玻璃结节及实性结节,这有利于患者在低剂量 CT 基线扫描时早期肺癌的精准识别。

4 总结

本研究的局限性:① 本研究样本来源于健康体检人群,部分实性结节的例数较少,今后应纳入不同人群进行分析;② 金标准的确定存在一定偏倚性;③ 两种软件的算法不同可能会影响结果,同等算法下肺结节的检出情况有待进一步研究。

综上所述,两种 AI 软件对肺结节类型的检出有差异,因此需要对 AI 软件进一步优化算法,加强其对于结节识别的同质性,并提升软件数据集的管理,便于临床对结节的随访及决策。两种 AI 软件联合诊断的方法是早期肺癌准确识别的一种可行优化方案。

参考文献

- [1] 中华医学会肿瘤学分会,中华医学会杂志社. 中华医学会肿瘤学分会肺癌临床诊疗指南(2021 版)[J]. 中华医学杂志, 2021, 101(23): 1725-1757.
- [2] LEIGH J, COLLABORATION G B O D C. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life years for 29 cancer groups, 1990 to 2017[J]. JAMA Oncology, 2019, 5(12): 1749-1768. DOI:10.1001/jamaoncol.2019.2996.

- [3] YUAN S, WEI Y, ZHAO D. Computer-aided lung nodule recognition by SVM classifier based on combination of random undersampling and SMOTE[J]. *Computational and Mathematical Methods in Medicine*, 2015: 1-13. DOI:10.1155/2015/368674.
- [4] AZUMA M, NAKADA H, TAKEI M, et al. Detection of acute rib fractures on CT images with convolutional neural networks: Effect of location and type of fracture and reader's experience[J]. *Emergency Radiology*, 2022, 29(2): 317-328. DOI:10.1007/s10140-021-02000-6.
- [5] MASSALHA S, CLARKIN O, THORNHILL R, et al. Decision support tools, systems, and artificial intelligence in cardiac imaging[J]. *Canadian Journal of Cardiology*, 2018, 34(7): 827-838. DOI:10.1016/j.cjca.2018.04.032.
- [6] 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]. *Technology in Cancer Research & Treatment*, 2018, 17: 1533033818798800.
- [7] 中华医学会放射学分会, 中国食品药品检定研究院, 国家卫生健康委能力建设与继续教育中心, 等. 胸部 CT 肺结节数据集构建及质量控制专家共识[J]. *中华放射学杂志*, 2021, 55(2): 104-110. DOI:10.3760/cma.j.cn112149-20200713-00915.
Chinese Society of Radiology Chinese Medical Association, National Institutes for Food and Drug Control, National Health Commission Capacity Building and Continuing Education Center, et al. Expert consensus on the construction and quality control of thoracic CT datasets for pulmonary nodules[J]. *Chinese Journal of Radiology*, 2021, 55(2): 104-110. DOI:10.3760/cma.j.cn112149-20200713-00915. (in Chinese).
- [8] HASSANZADEH T, ESSAM D, SARKER R. Evolutionary deep attention convolutional neural networks for 2D and 3D medical image segmentation[J]. *Journal of Digital Imaging*, 2021, 34(6): 1387-1404. DOI:10.1007/s10278-021-00526-2.
- [9] 张正华, 蔡雅倩, 韩丹, 等. 人工智能与不同级别医师对肺结节检出效能的对照研究[J]. *中国医学影像学杂志*, 2020, 28(9): 662-665. DOI:10.3969/j.issn.1005-5185.2020.09.005.
ZHANG Z H, CAI Y Q, HAN D, et al. Artificial intelligence and various levels of physicians in detection of pulmonary nodules: A comparative study[J]. *Chinese Journal of Medical Imaging*, 2020, 28(9): 662-665. DOI:10.3969/j.issn.1005-5185.2020.09.005. (in Chinese).
- [10] TSUTANI Y, MIYATA Y, NAKAYAMA H, et al. Prognostic significance of using solid versus whole tumor size on high-resolution computed tomography for predicting pathologic malignant grade of tumors in clinical stage IA lung adenocarcinoma: A multicenter study[J]. *The Journal of Thoracic and Cardiovascular Surgery*, 2012, 143(3): 607-612. DOI:10.1016/j.jtcvs.2011.10.037.
- [11] LEE H J, GOO J M, LEE C H, et al. Nodular ground-glass opacities on thin-section CT: Size change during follow-up and pathological results[J]. *Korean Journal of Radiology*, 2007, 8(1): 22-31. DOI:10.3348/kjr.2007.8.1.22.
- [12] KANG G X, LIU K, HOU B B, et al. 3D multi-view convolutional neural networks for lung nodule classification[J]. *Plos One*, 2017, 12(11): e0188290. DOI:10.1371/journal.pone.0188290.
- [13] SHIGEFUKU S, SHIMADA Y, HAGIWARA M, et al. Prognostic significance of ground-glass opacity components in 5-year survivors with resected lung adenocarcinoma[J]. *Annals of Surgical Oncology*, 2021, 28(1): 148-156. DOI:10.1245/s10434-020-09125-x.
- [14] MENG Y, LIU C L, CAI Q, et al. Contrast analysis of the relationship between the HRCT sign and new pathologic classification in small ground glass nodule-like lung adenocarcinoma[J]. *Radiology Medical*, 2019, 124(1): 8-13. DOI:10.1007/s11547-018-0936-x.
- [15] 刘凯, 张荣国, 涂文婷, 等. 深度学习技术对胸部 X 线平片亚实性结节的检测效能初探[J]. *中华放射学杂志*, 2017, 51(12): 918-921. DOI:10.3760/cma.j.issn.1005-1201.2017.12.006.
LIU K, ZHANG R G, TU W T, et al. A preliminary investigation on pulmonary subsolid nodule detection using deep learning methods from chest X-rays[J]. *Chinese Journal of Radiology*, 2017, 51(12): 918-921. DOI:10.3760/cma.j.issn.1005-1201.2017.12.006. (in Chinese).
- [16] 李欣菱, 郭芳芳, 周振, 等. 基于深度学习的人工智能胸部 CT 肺结节检测效能评估[J]. *中国肺癌杂志*, 2019, 22(6): 336-340. DOI:10.3779/j.issn.1009-3419.2019.06.02.
LI X L, GUO F F, ZHOU Z, et al. Performance of deep-learning-based artificial intelligence on detection of pulmonary nodules in chest CT[J]. *Chinese Journal of Lung Cancer*, 2019, 22(6): 336-340. DOI:10.3779/j.issn.1009-3419.2019.06.02. (in Chinese).
- [17] 刘娜, 赵正凯, 邹佳瑜, 等. 基于人工智能的胸部 CT 肺结节检出及良恶性诊断效能评估[J]. *CT 理论与应用研究*, 2021, 30(6): 709-715. DOI:10.15953/j.1004-4140.2021.30.06.06.
LIU N, ZHAO Z K, ZOU J Y, et al. Evaluation of detection and diagnostic efficiency of pulmonary

- nodules by chest CT based on artificial intelligence[J]. CT Theory and Applications, 2021, 30(6): 709-715. DOI:10.15953/j.1004-4140.2021.30.06.06. (in Chinese).
- [18] 王祥, 李清楚, 邵影, 等. 基于三维卷积神经网络肺结节深度学习算法模临床效能初步评估[J]. 放射学实践, 2019, 34(9): 942-946. DOI:10.13609/j.cnki.1000-0313.2019.09.002.
- WANG X, LI Q C, SHAO Y, et al. A preliminary clinical evaluation of a 3D convolutional neural network based deep learning system[J]. Radiologic Practice, 2019, 34(9): 942-946. DOI:10.13609/j.cnki.1000-0313.2019.09.002. (in Chinese).
- [19] 左玲子, 黄艳. 人工智能在体检肺 CT 中检出的假阳性结节研究[J]. 中国医疗设备, 2021, 36(10): 177-180. DOI:10.3969/j.issn.1674-1633.2021.10.041.
- ZUO L Z, HUANG Y. Study of false positive nodules detected by artificial intelligence in lung CT examination[J]. China Medical Devices, 2021, 36(10): 177-180. DOI:10.3969/j.issn.1674-1633.2021.10.041. (in Chinese).

Evaluation of the Diagnostic Value of Pulmonary Nodules Based on Two AI Software

CHEN Xinhua¹, HUANG Xiaoqi¹, LI Jianlong^{1✉}, GUO Youmin²

1. Affiliated Hospital of Yan'an University, Yan'an 716000, China

2. The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710000, China

Abstract: Objective: To explore the clinical value of two kinds of AI detection software in ≥ 5 mm pulmonary nodules. Methods: A total of 92 patients with pulmonary nodules (483 nodules) were selected from the affiliated Hospital of Yan'an University between June and October 2021. The nodules detected by AI software were evaluated and the number and type of nodules were recorded by two senior radiologists. Two senior imaging doctors then evaluated the manual film reading, which was used as the gold standard for nodule recognition. Subsequently, the detection rate and false positive and negative rates of the two software were calculated, and the nodule detection value of the two AI software was evaluated. Additionally, the chi-square and Fisher precision tests were used to compare the differences between the different software and the gold standard. Finally, the diagnostic value of the combination of the two kinds of AI software for pulmonary nodules was calculated. Results: The detection rates of software A and software B nodules were 92.1% and 87.0%, respectively. Moreover, the coincidence degree between software A and manual reading was general ($Kappa = 0.213$), while that between software B and manual reading was weak ($Kappa = 0.150$). There was also a significant difference in the detection of solid nodules and calcified nodules between software A and manual reading, as well as between software B and pure ground glass nodules. The detection rate of nodules with the combined two kinds of AI software was 97.1%. However, compared with manual reading, there was no significant difference in the detection of nodule types. The combination of the two AI software had a good agreement with manual reading ($Kappa = 0.439$). Conclusion: The combination of two kinds of AI software improved the ability of nodule detection and classification analysis. Furthermore, the method of joint diagnosis is recommended for clinical use and it provides evidence for further improving the homogenization management of AI data sets.

Keywords: artificial intelligence; CT; pulmonary nodules; detection



作者简介: 陈新花, 女, 延安大学影像医学与核医学专业硕士研究生, 主要从事胸部与骨关节影像诊断, E-mail: 523423054@qq.com; 李建龙[✉], 男, 延安大学附属医院放射科主任医师、副教授、硕士生导师, 主要从事胸部与骨关节影像诊断, E-mail: yaljl1116@163.com。