

# 人工智能与不同级别医师对肺结节检出效能的对照研究

张正华, 蔡雅倩, 韩丹\*, 周小君, 黄益龙, 段慧

昆明医科大学第一附属医院医学影像科, 云南昆明 650000; \*通讯作者 韩丹 [kmhandan@sina.com](mailto:kmhandan@sina.com)

【基金项目】云南省教育厅科学研究基金项目 (2019J1229)

**【摘要】目的** 探讨基于深度学习的人工智能 (AI) 与不同级别医师 CT 肺结节检测能力的差异。**资料与方法** 收集胸部 CT 平扫 550 例, 以 2 名从事胸部影像诊断的主任医师共同阅片检出肺结节的数目、位置、大小为参照标准, 比较 A 组 (5 名实习医师)、B 组 (5 名住院医师)、C 组 (5 名主治医师)、D 组 (AI) 对肺结节的检出率、假阳性率和平均阅片时间。**结果** 2 名主任医师共确定 5520 枚结节。A、B、C、D 组检出率分别为 40.5%、65.0%、89.4% 和 94.3%, 差异有统计学意义 ( $\chi^2=5067.866$ ,  $P<0.001$ ); 假阳性率分别为 20.0%、13.0%、10.0% 和 14.0%。D 组与 C 组胸膜下结节检出率差异无统计学意义 ( $\chi^2=0.611$ ,  $P=0.434$ ); D 组外周性及中心性结节检出率明显高于其他组, 差异有统计学意义 ( $P<0.05$ )。D 组与 C 组大结节检出率差异无统计学意义 ( $\chi^2=2.980$ ,  $P=0.084$ ); D 组中等结节及小结节检出率明显高于其他 3 组, 差异有统计学意义 ( $P<0.05$ )。4 组平均阅片时间分别为 (14.9±2.5) min、(6.9±2.1) min、(5.0±1.5) min、(0.1±0.0) min, 差异有统计学意义 ( $P<0.001$ )。**结论** AI 能在短时间内有效检出肺结节, 其检出率略高于主治医师, 可作为肺结节筛查的有效辅助工具。

**【关键词】** 肺结节; 人工智能; 体层摄影术, X 线计算机; 病例对照研究

**【中图分类号】** R563.1; R445.3 **【DOI】** 10.3969/j.issn.1005-5185.2020.09.005

## Artificial Intelligence and Various Levels of Physicians in Detection of Pulmonary Nodules: A Comparative Study

ZHANG Zhenghua, CAI Yaqian, HAN Dan\*, ZHOU Xiaojun, HUANG Yilong, DUAN Hui

Department of Medical Imaging, First Affiliated Hospital of Kunming Medical University, Kunming 650000, China; \*Address Correspondence to: HAN Dan; E-mail: [kmhandan@sina.com](mailto:kmhandan@sina.com)

**【Abstract】Purpose** To explore the ability of deep learning-based artificial intelligence (AI) in the detection of pulmonary nodules compared with physicians at different levels. **Materials and Methods** 550 patients with chest CT scans were enrolled in this study. The consensus of pulmonary nodules detected by two positive physicians who were diagnosed with chest imaging was used as the reference standard. The location, size and density of lung nodules were recorded. The detection rate, false positive rate and average reading time of pulmonary nodules were compared among group A (5 interns), group B (5 resident physicians), group C (5 attending physicians) and group D (AI). **Results** The two chief physicians determined a total of 5520 nodules. The detection rates of groups A, B, C and D were 40.5%, 65.0%, 89.4% and 94.3%, respectively; and the difference was statistically significant ( $\chi^2=5067.866$ ,  $P<0.001$ ); the false positive rates were 20.0%, 13.0%, 10.0% and 14.0%, respectively. There was no significant difference in the detection rate of subpleural nodules between group D and C ( $\chi^2=0.611$ ,  $P=0.434$ ); the detection rate of peripheral and central nodules in group D was significantly higher than that of other groups, and the difference was statistically significant ( $P<0.05$ ). There was no significant difference in the detection rate of large nodules between group D and C ( $\chi^2=2.980$ ,  $P=0.084$ ); the detection rate of medium and small nodules in group D was significantly higher than that of the other three groups, and the difference was statistically significant ( $P<0.05$ ). The average reading time of the 4 groups was (14.9±2.5) min, (6.9±2.1) min, (5.0±1.5) min and (0.1±0.0) min, respectively, the difference was statistically significant ( $P<0.001$ ). **Conclusion** AI can effectively detect lung nodules in a short period of time, and its detection rate is slightly higher than that of the attending physician, and it can be used as an effective auxiliary tool for lung nodule screening.

**【Key words】** Pulmonary nodules; Artificial intelligence; Tomography, X-ray computed; Case-control studies

Chinese Journal of Medical Imaging, 2020, 28 (9): 662-665

随着高分辨率 CT (HRCT) 在肺癌早期筛查应用的普及, 图像数量倍增及对小结节显示率增高使阅片工作量剧增。高强度工作可导致影像科医师视觉疲劳,

加之经验不足等, 不可避免地导致小结节的漏诊和误诊。如何在人力资源不足的情况下, 在大量图像中准确检出肺结节并早期定性成为临床的迫切需求。基于

深度学习的人工智能 (artificial intelligence, AI) 凭借其较准确的算法模型可快捷地检出结节, 并已开始临床试用。然而目前有关 AI 对肺结节检测效能的相关研究报道较少, 且尚无明确定论<sup>[1-2]</sup>。本研究拟通过对比不同级别医师与 AI 对肺结节的检出能力, 评价 AI 对肺结节检出的敏感性。

## 1 资料与方法

**1.1 研究对象** 随机抽取 2018 年 10 月—2019 年 6 月昆明医科大学第一附属医院行胸部 CT 平扫患者 550 例, 其中男 382 例, 女 168 例; 年龄 18~78 岁, 平均 (58.6±11.7) 岁。纳入标准: ①肺结节 1~40 枚, 且直径≤3 cm; ②CT 图像层厚为 1 mm。排除标准: ①肺部弥漫性病变; ②按 1~5 分制评价图像质量, 排除评分≤3 分者。评分细则: 1 分, 无法显示正常结构, 完全不能满足诊断要求; 2 分, 肺纹理及分支模糊, 无法满足诊断要求; 3 分, 边缘模糊, 肺纹理及分支尚可, 基本可以满足诊断要求; 4 分, 边缘略模糊, 肺纹理及分支清晰, 满足诊断要求; 5 分, 边缘清晰, 肺纹理及分支清晰, 完全满足诊断要求<sup>[3]</sup>。

### 1.2 仪器与方法

**1.2.1 CT 检查** 采用 Siemens Somatom Definition AS 128、Siemens Somatom Definition Flash 及联影 UCT 760 CT 机行胸部 HRCT 扫描。扫描参数: ①Siemens Somatom Definition AS 128 层螺旋 CT 机及 Siemens Somatom Definition Flash: 管电压 100 kV, 管电流 80 mAs, 半自动管电压调制, 自动管电流调制, 螺距 0.938, 重建卷积函数采用 B70f。②联影 UCT 760: 平扫管电压 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。观察采用肺窗窗宽 1500 Hu、窗位 -400 Hu。

**1.2.2 结果判读** 由 2 名从事胸部影像诊断的主任医师采用盲法对 550 例患者的胸部 CT 图像共同阅片, 分析结节的数目、位置、大小。当 2 名主任医师判读结果不一致时, 请第 3 名主任医师共同协商确定一致性意见作为参照标准, 以此计算肺结节的检出率和假阳性率。

根据肺结节位置分为胸膜下结节 (与胸膜相连)、中心性结节 (距肺门 2.0 cm 以内)、外周性结节 (肺门区以外, 但不与胸膜相连); 根据结节最大径分为小结节 (<0.5 cm)、中等结节 (0.5~1.0 cm)、大结

节 (>1.0 cm)。

**1.2.3 分组及测定方法** 根据阅片医师不同分成 4 组: A 组 5 名 5 年制本科实习生 (第 5 年临床实习, 已从事肺部 CT 诊断工作 3 个月); B 组 5 名住院医师 (从事住院医师工作 3 年); C 组 5 名主治医师 (从事主治医师工作 3 年); D 组为 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) 软件。阅片内容: ①每组结节检出的数目、位置、大小, 并与参照标准对照, 根据公式 (1)、(2) 分别计算每组肺结节的检出率和假阳性率; ②医师阅片时间指医师在 3 兆巨鲨专业显示屏逐层观察全部肺窗图像, 得出每例患者结节数目、位置、大小的时间; AI 阅片时间是点击 AI 软件显示屏显示结节数目、位置、大小的时间。分别比较 4 组肺结节检出情况和阅片时间。

$$\text{检出率} = \frac{\text{真阳性结节}}{\text{结节总数}} \times 100\% \quad (1)$$

$$\text{假阳性率} = \frac{\text{假阳性结节}}{\text{结节总数}} \times 100\% \quad (2)$$

**1.3 统计学方法** 采用 SPSS 23.0 软件, 符合正态分布的计量资料以  $\bar{x} \pm s$  表示。肺结节检出率的组内和组间比较采用  $\chi^2$  检验; 各组阅片时间比较采用秩和检验或单因素方差分析, 两两比较采用 LSD 法。P<0.05 表示差异有统计学意义。

## 2 结果

**2.1 参照标准** 2 名主任医师共检出 5520 枚结节。每例患者肺结节数为 1~32 枚, 平均 (10.0±1.3) 枚, 结节直径 0.2~2.8 cm。

**2.2 结节检出情况比较** A、B、C 组内 5 名医师肺结节检出率差异无统计学意义 ( $\chi^2=9.266$ 、7.130、5.618,  $P>0.05$ )。4 组平均检出率依次为 A 组<B 组<C 组<D 组, 差异有统计学意义 ( $\chi^2=5067.866$ ,  $P<0.001$ ; 表 1), 其中 D 组的假阳性率高于 B、C 组, 典型病例见图 1。

A、B、C 组胸膜下、外周性及中心性结节检出率依次为 A 组<B 组<C 组, 差异有统计学意义 ( $P<0.05$ ); D 组与 C 组胸膜下结节检出率差异无统计学意义 ( $\chi^2=0.611$ ,  $P=0.434$ ), D 组外周性及中心性结节检出率明显高于其他 3 组, 差异有统计学意义 ( $P<0.05$ )。在不同大小结节检测方面: A、B、C 组大结节、中等结节及小结节检出率均依次为 A 组<B 组<C 组, 差异有统计学意义 ( $P<0.05$ ); D 组与 C 组大结节检出率差异无统计学意义 ( $\chi^2=2.980$ ,  $P=0.084$ ), D 组中等结节及小结节检出率明显高于其他 3 组, 差异有统计学意义 ( $P<0.05$ )。见表 1。

表 1 各组不同肺结节位置与大小检出率比较 (%)

分组	位置			大小			平均检出率	假阳性率
	胸膜下	外周性	中心性	小结节	中等结节	大结节		
A 组	40.0*	40.6*	40.3*	37.5*	41.0*	60.2*	40.5	20.0
B 组	61.8*	65.7*	59.7*	62.3*	66.2*	78.5*	65.0	13.0
C 组	85.0	90.7*	70.2*	87.0*	91.0*	98.5	89.4	10.0
D 组	86.4	96.0	82.3	91.2	98.7	96.8	94.3	14.0

注：与 D 组比较，\* $P<0.05$

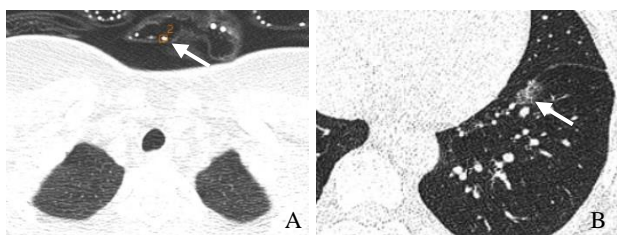


图 1 AI 将体外衣物误判为结节(箭),致假阳性率较高(A);左肺下叶前基底段胸膜下磨玻璃结节(箭)容易漏诊,术后病理证实为腺癌(B)

2.3 阅片时间 各组组内阅片时间差异均无统计学意义( $P>0.05$ );各组阅片时间分别为(14.9±2.5)min、(6.9±2.1)min、(5.0±1.5)min、(0.1±0.0)min,差异有统计学意义( $P<0.001$ )。D 组阅片时间明显短于 A、B、C 组,差异均有统计学意义( $P<0.001$ )。

### 3 讨论

本研究结果显示人工阅片肺结节的检出率随着医师级别增高依次递增,且假阳性率依次递减,表明肺结节的检出需要一定的临床经验水平。在实际工作中,由于人员配备不均衡,部分医院缺乏高年资医师。高效的肺结节检测对肺癌的风险评估具有重要意义,快速准确定位肺结节的确切位置具有重要意义和挑战性。近年研究涉及基于传统计算机视觉算法计算机辅助诊断系统(CAD)和 AI。CAD 方案大多比较复杂和耗时,需要更多的图像处理模块,且相关敏感度及特异度较低<sup>[4-5]</sup>;AI 是一种基于卷积神经网络的肺结节检测模型,其敏感度和特异度较传统的 CAD 高,目前在肺结节检出及定性方面应用较多,但以往研究对不同模型算法的 AI 探讨较多,大多基于其原理、算法步骤等展开讨论<sup>[6-10]</sup>,很少从临床及实用性评估其在临床工作中运用的准确性及弊端。2017 年 Fleischner 肺结节指南总原则指出,结节随访的重要指标包括数目及大小<sup>[11]</sup>。因此,本研究主要探讨 AI 辅助诊断软件对肺结节位置、数目及大小的检测效能。

本研究结果显示, AI 组肺结节检出率为 94.3%,高于以往研究报道的 CAD 对肺结节的检出率<sup>[12-13]</sup>。金文忠等<sup>[12]</sup>报道以 6 名资深影像专业医师作为参照标准, AI 肺结节检出率平均达 96.6%,与本研究结果相似。本研究通过与不同级别的医师检出情况对照显示, AI 的检测水平略高于本院主治医师水平,若影像医师在临床工作中参考 AI 检测结果,可大大减少肺结节的漏诊,特别是对于实习医师及住院医师而言。对于不同位置结节的检出情况,主治医师组胸膜下结节检出率与 AI 组差异无统计学意义。由于胸膜下结节表现为与胸膜相连的结节影,易于观察。AI 组外周及中心性结节检出率显著高于人工阅片,由于该区域肺纹理相对较多,结节与血管断面相似,加之长时间人工阅片导致注意力难以集中。对于不同大小结节的检出情况,主治医师组大结节检出率与 AI 组差异无统计学意义,两组大结节检出率为 98.5%、96.8%。刘晓鹏等<sup>[1]</sup>以病理结果为“金标准”, AI 的肺结节(1~3 cm)检出率为 96.4%,表明 AI 在大结节检出方面具有较完善的算法。AI 组中等结节及小结节检出率明显高于人工阅片,其可能原因为结节较小且人工阅片易疲劳,容易漏诊。AI 肺结节的高检出率使得临床工作中肺结节的检出率增加;而李欣菱等<sup>[14]</sup>研究认为其对较小(特别是<0.5 cm)结节的临床意义有待商榷,若完全按照 AI 的结果进行临床干预会增加医疗负担及过度治疗。因此,对肺结节的判断及干预需要医师参考 AI 结果并进一步分析。

本研究发现, AI 检出结节总数大于参照标准。由于 AI 的高敏感性而出现假阳性率较高(14.0%)的特点,表明目前 AI 的重点集中在阳性结节的学习上,而对假阳性结节的学习尚需完善。本研究中 AI 容易将血管截面及肺外部分结构标注成结节(如体外衣物、胸锁关节及胃肠道内容物等),易于辨认,在临床工作中需影像医师进行复诊。有研究采用训练大规模数据集的方式自动化学学习和提升,可明显降低假阳性率<sup>[15]</sup>。尽管 AI 组肺结节检出率高于人工阅片,但也会遗漏



一些结节, AI 未检出结节的可能原因包括: ①结节较小; ②结节位于胸膜下; ③结节位于血管旁; ④结节密度浅淡; ⑤不明原因。在实际工作中发现, 对于密度较淡且位于胸膜下的磨玻璃结节 AI 更易漏诊, 而这些磨玻璃结节往往是恶性结节。因此, 在临床工作中应注意观察胸膜下、肺门及血管旁区域, 防止漏诊。而对于那些不明原因的未检出结节, 可能与软件的分割算法有关<sup>[11]</sup>。因此, 基于深度学习的 AI 算法仍需有针对性地改善模型并进一步优化参数, 以提高肺结节的检出率并降低假阳性率。以往研究表明 AI 较人工阅片时间明显缩短, 每张胸片平均用时约为 0.28 s<sup>[16]</sup>; 而本研究显示 AI 阅 CT 片时间仅为 (0.1±0.0) min, 大大提高了工作效率。Li 等<sup>[17]</sup>报道对于多发肺结节患者, AI 能快速直接显示结节的位置、数目及大小, 为后续诊疗计划提供参考依据。因此, AI 协助影像科医师可提高阅片的速度和诊断准确率, 使更多的患者从中受益。

总之, AI 对于肺结节的检出率较高, 其水平略高于主治医师, 作为肺结节筛查的有效辅助工具, 在临床上可广泛应用。随着 AI 算法的不断完善及多中心大样本研究的开展, 有望进一步提高 AI 的肺结节检出效能。

### 参考文献

- [1] 刘晓鹏, 周海英, 胡志雄, 等. 人工智能识别技术在 T1 期肺癌诊断中的临床应用研究. 中国肺癌杂志, 2019, 22(5): 319-323.
- [2] 金文忠, 陆耀, 汪阳. 基于人工智能的胸部 CT 智能辅助诊断系统在 LDCT 数据集上的应用研究. 中国医学计算机成像杂志, 2018, 24(5): 373-377.
- [3] 徐秋贞, 邵海磊, 吕燕, 等. 胸部能谱成像模式与常规 CT 扫描的辐射剂量及图像质量的仿真体模研究. 中华放射医学与防护杂志, 2017, 37(12): 957-961.
- [4] Freedman MT. Comment on "maximum-intensity- projection and computer-aided-detection algorithms as stand-alone reader devices in lung cancer screening using different dose levels and reconstruction kernels". Am J Roentgenol, 2017, 208(3): W132.
- [5] Young S, Lo P, Kim G, et al. The effect of radiation dose reduction on computer-aided detection (CAD) performance in a low-dose lung cancer screening population. Med Phys, 2017, 44(4): 1337-1346.
- [6] Wang S, Zhou M, Liu Z, et al. An automatic detection system of lung nodule based on multi-group patch-based deep learning network. Med Image Anal, 2017, 8(40): 172-183.
- [7] Tu X, Xie M, Gao JJ, et al. Automatic categorization and scoring of solid, part-solid and non-solid pulmonary nodules in CT images with convolutional neural network. Sci Rep, 2017, 7(1): 8533.
- [8] Liu JK, Jiang HY, Gao MD, et al. An assisted diagnosis system for detection of early pulmonary nodule in computed tomography images. J Med Syst, 2017, 41(2): 30.
- [9] 王风, 王磊, 李因, 等. 基于三维卷积神经网络深度学习的肺结节良恶性的鉴别诊断. 中国医学影像学杂志, 2019, 27(10): 779-782, 787.
- [10] Fu B, Wang GS, Wu M, et al. Influence of CT effective dose and convolution kernel on the detection of pulmonary nodules in different artificial intelligence software systems: a phantom study. Eur J Radiol, 2020, 126: 108928.
- [11] Macmahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the fleischner society 2017. Radiology, 2017, 284(1): 228-243.
- [12] 罗红兵, 周鹏, 青浩渺, 等. 计算机辅助检测系统在低剂量 CT 肺癌筛查中非钙化肺结节检出方法的研究. 肿瘤预防与治疗, 2017, 30(1): 33-38.
- [13] 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. 实用放射学杂志, 2018, 34(9): 1430-1433.
- [14] 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. 中国肺癌杂志, 2019, 22(6): 336-340.
- [15] Huang P, Park S, Yan R, et al. Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. Radiology, 2018, 286(1): 286-295.
- [16] 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. 中华放射学杂志, 2017, 51(12): 918-921.
- [17] Li X, Hu B, Li H, et al. Application of artificial intelligence in the diagnosis of multiple primary lung cancer. Thorac Cancer, 2019, 10(11): 2168-2174.

【收稿日期】2020-03-26 【修回日期】2020-06-12

(本文编辑 闻浩)