

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^[1],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 ≤ 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 χ^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/ χ^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(mm ²)	39.13±23.97	121.23±71.76	-9.629	<0.001
体积(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。

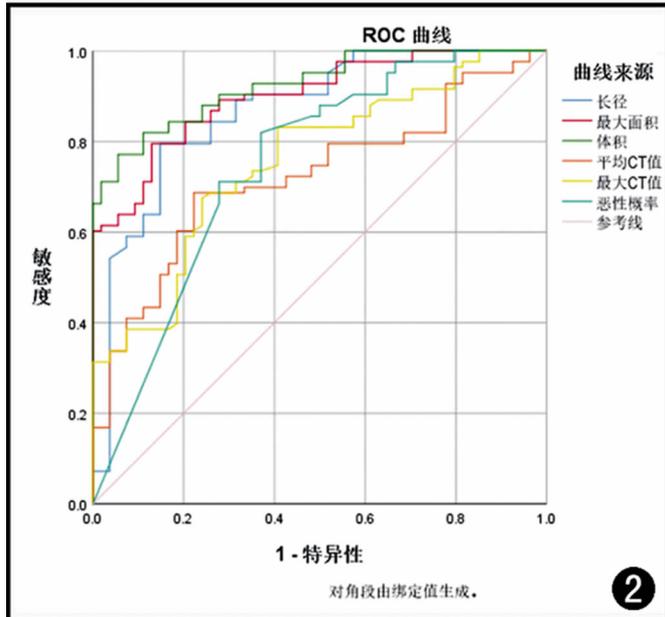


图 1 AI 自动筛选肺结节并对其进行定量分析,计算机自动测量其长径、短径、最大面积、平均 CT 值、最大 CT 值、最小 CT 值,并对其进行准确定位。a) AI 对肺内结节进行筛检、勾画结节轮廓并标注序号; b) AI 直接显示肺结节的各参数。图 2 GGN 各量化参数的 ROC 曲线。

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 ²)	61.12	0.795	0.870	0.899	<0.001
体积(mm ³)	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%^[2,3],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^[4,5],主要集中在发现异常、量化测量及鉴别诊断等方面^[6,7]。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^[8]。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^[9]和邵亚军等^[10]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^[11]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 mm²时,敏感度均为 0.795,特异度分别为 0.852、0.870。Kitami 等^[12]和熊廷伟等^[13]的研究结果表明直径是 GGN 恶变的危险因素,Yang 等^[14]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^[15]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^[16]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^[17]。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 ¹⁸F-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 ≤ 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)