

基于人工智能的 CT 影像预测肺腺癌浸润性: Meta 分析

黄文君, 周秀秀, 周陶胡, 葛艳明, 董鹏, 范丽

【摘要】 目的:评价基于人工智能(AI)的 CT 影像预测肺腺癌浸润性的诊断性能。**方法:**通过 PubMed、Embase、Cochrane 图书馆、Web of Science、中国知网、SinoMed、万方和维普等数据库,检索 2011 年 1 月 1 日—2021 年 6 月 30 日公开发表的基于 AI 的 CT 影像预测肺腺癌浸润性的所有文章。以诊断试验的纳入和排除标准筛选文章并提取关键特征信息。通过软件 MetaDiSc 1.4 和 Stata 16.0 进行 Meta 分析。以浸润性腺癌作为阳性结果、非浸润性腺癌作为阴性结果,计算合并后的敏感度、特异度、阳性似然比、阴性似然比和诊断比值比,绘制集成受试者操作特征(SROC)曲线并得出曲线下面积(AUC)。基于 Meta 回归探究异质性的来源。使用敏感性分析验证 Meta 分析结果的可靠性。**结果:**共 12 项研究入选,包括 4066 例患者的 4325 枚肺结节,磨玻璃结节占 97%。研究间存在异质性,故指标合并采取随机效应模型。合并后的敏感度、特异度、阳性似然比、阴性似然比和诊断比值比分别为 0.86、0.82、4.55、0.19 和 28.31,AUC 为 0.9110。Meta 回归结果表明 AI 算法的不同是异质性的来源,敏感性分析显示 Meta 分析结果的可靠度高。**结论:**基于 AI 的 CT 影像对肺腺癌浸润性有较高的预测能力,可向医生提供更确切的诊疗信息,进而优化患者的治疗方案。

【关键词】 人工智能; 体层摄影术,X 线计算机; 腺癌; 肺肿瘤; Meta 分析

【中图分类号】 R814.42; R734.2; R-05 **【文献标志码】** A **【文章编号】** 1000-0313(2022)11-1359-09

DOI:10.13609/j.cnki.1000-0313.2022.11.005

开放科学(资源服务)标识码(OSID):



Predicting the invasiveness of lung adenocarcinoma using CT images based on artificial intelligence: a meta-analysis HUANG Wen-jun, ZHOU Xiu-xiu, ZHOU Tao-hu, et al. School of Medical Imaging, Weifang Medical University, Shandong 261000, China

【Abstract】 Objective: To evaluate the diagnostic performance of artificial intelligence (AI) using CT images for predicting lung adenocarcinoma invasiveness. **Methods:** Systematically retrieved all published articles on AI-based CT imaging for predicting lung adenocarcinoma invasiveness from January 1, 2011, to June 30, 2021, through PubMed, Embase, Cochrane Library, Web of Science, CNKI, SinoMed, Wanfang Database, and VIP series database. Articles were screened with inclusion and exclusion criteria of diagnostic tests and characteristic information was extracted. Meta-analysis was performed by using Metadisc 1.4 software and Stata 16.0 software. Sensitivity, specificity, positive/negative likelihood ratio, and diagnostic ratio were measured for the combined results with invasive adenocarcinoma as the positive result and non-invasive adenocarcinoma as the negative result. And the summary receiver operating characteristic (SROC) curve was plotted and the area under the curve (AUC) was derived. The sources of heterogeneity was explored based on Meta-regression. Sensitivity analysis was used to validate the reliability of the Meta-analysis results. **Results:** Twelve studies were enrolled, including 4325 pulmonary nodules in 4066 patients, with ground glass nodules accounting for 97%. The indexes were combined in a random-effects model due to heterogeneity existed in those studies. The combined sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and diagnostic ratio were 0.86, 0.82, 4.55, 0.19, and 28.31, respectively, with an AUC of 0.9110. Meta-regression analysis indicated that different AI algorithms may be the source of heterogeneity, and sensitivity analysis shows

作者单位: 261000 山东, 潍坊医学院医学影像学院(黄文君、周陶胡、葛艳明、董鹏); 261000 山东, 潍坊医学院附属医院医学影像科(葛艳明、董鹏); 200003 上海, 海军军医大学第二附属医院放射诊断科(周秀秀、范丽)

作者简介: 黄文君(1992—), 男, 四川南充人, 硕士研究生, 主要从事胸部疾病的影像诊断及人工智能方向的研究。

通讯作者: 范丽, E-mail: fanli0930@163.com

基金项目: 国家自然科学基金(81871321, 81930049, 82171926); 上海市青年科技英才扬帆计划(20YF1449000); 上海长征医院 2020 年度院创新型临床研究项目(2020YLCYJ-Y24); 上海长征医院金字塔人才工程

high reliability of Meta-analysis results. **Conclusion:** AI-based CT imaging has a robust predictive power for lung adenocarcinoma invasiveness, which can provide doctors with more precise diagnosis and treatment information, and thus optimize the therapeutic schedule.

【Key words】 Artificial intelligence; Tomography, X-ray computed; Adenocarcinoma; Lung neoplasms; Meta-analysis

国际癌症研究机构更新的最新版 GLOBOCAN 2020 指出, 现今肺癌仍是癌症死亡的主要原因, 同时国内肺癌的整体发病率(56.3/10 万)和死亡率(49.4/10 万)也都居于首位^[1]。作为最常见的肺癌亚型, 肺腺癌占有肺癌病例的一半以上^[2], 可见防控肺腺癌是肺癌防控中的重要一环。随着常规 CT 的全面覆盖和低剂量 CT 筛查的普及, 许多肺腺癌能在早期即被检出^[3-4]。目前很多肺腺癌或其前体病变的 CT 影像表现是持续存在的磨玻璃结节(ground glass nodule, GGN)^[5]。据 2015 版世卫组织(WHO)分类, 肺腺癌包含浸润前病变——非典型腺瘤样增生(atypical adenomatous hyperplasia, AAH)和原位腺癌(adenocarcinoma in situ, AIS), 微浸润性腺癌(minimally invasive adenocarcinoma, MIA)及浸润性腺癌(invasive adenocarcinoma, IAC)^[6]。针对不同的病理亚型, 患者的手术方式和 5 年无病生存率(disease free survival, DFS)明显不同, IAC 患者较非 IAC(包括 AAH、AIS 和 MIA)患者需要更大范围的肺组织切除术且 5 年 DFS 显著降低^[7-9]。因此通过无创的 CT 影像预测肺腺癌浸润性能为临床医师制定最优治疗方案提供更全面的参考依据。近年来人工智能(artificial intelligence, AI)的快速发展能够利用 CT 影像鉴别 IAC 和非 IAC, 主要通过 CT 影像组学和深度学习(deep learning, DL)两种方法实现^[10]。目前基于此两种方法区分 IAC 和非 IAC 的研究已有不少, 但是不同研究间的样本量、研究质量及 AI 算法的差异却使得 AI 诊断的性能参差不齐。因此, 本文对基于 AI 的 CT 影像预测肺腺癌浸润性的能力进行了 Meta 分析, 旨在为其临床应用提供更有力的支撑。

材料与方法

1. 文献检索策略

检索数据库: PubMed、Embase、Cochrane 图书馆、Web of Science、中国知网(CNKI)、SinoMed(CBM)、万方和维普等在线数据库。文献发表时间: 2011 年 1 月 1 日—2021 年 6 月 30 日。检索内容: 国内外公开发表的基于 AI 的 CT 影像预测肺腺癌浸润性的所有文章。中文检索关键词包括: 肺结节、磨玻璃结节、纯磨玻璃结节、亚实性结节、混杂磨玻璃结节、腺癌、肺癌、CT、计算机断层扫描、人工智能、深度学习、

机器学习、神经网络、随机森林、支持向量机、电脑辅助诊断、病理、浸润、分类、识别、预测等; 英文检索关键词包括: pulmonary nodule, ground glass nodule, pure ground glass nodule, mixed ground glass nodule, sub-solid nodule, adenocarcinoma, lung cancer, CT, computed tomography, AI, deep learning, machine learning, computer-aided diagnosis, neural network, random forest, support vector machine, pathology, invasive, classify, predict 等。此外, 还包括这些词的同义词或近义词, 如部分实性结节, 非实性结节, 肺部肿瘤, part-solid nodule, non-solid nodule, lung nodule, lung neoplasm 等。

2. 纳入与排除标准

检索的文章使用 Endnote 9 软件管理。纳入标准: ①以汉语或英语公开发表的文献; ②以 CT 为影像检查方法, 通过 AI 进行诊断并取得诊断结果的患者; ③以病理学检查结果为肺腺癌(WHO 2015 版)为金标准, 纳入的病理亚型包括 IAC、MIA、AIS 和 AAH; ④病灶直径 ≤ 3 cm。排除标准包括: ①重复发表的文献; ②无法获取全文、数据不完整或研究内容不同; ③病变直径 > 3 cm 或对病变大小未作出说明; ④其他类型的肺癌或良性病变; ⑤文章类型为综述、Meta 分析、动物实验、会议、通信、述评等。

3. 纳入文献质量评价

QUADAS-2 诊断准确性试验质量评价工具用于文献质量评判^[11]。先由 3 位研究人员(分别有 1 年、3 年和 5 年影像诊断工作经验)独自进行预评估, 然后由专家组(由 3 位有 15 年以上影像诊断工作经验的医生组成)汇总评估结果作出最终决定。评估主要包含“偏倚风险判断”和“临床适用性评价”两部分。前者从病例选择、待评价试验、金标准、病例流程和进展状况四个方面进行评判, 分为“是”、“否”、“不清楚”三级; 后者主要判断病例选择、待评价试验、金标准与待评价项目的匹配程度, 分为“高”、“低”、“不清楚”三级。

4. 数据提取

先由 2 位研究人员提取数据, 再由另 1 位研究人员完成核对。当产生不同意见时, 交由专家组讨论并作出最终决定。提取的内容包括: ①一般特征: 第一作者、出版年份、资料来源、患者总数、性别、肺结节总数、结节类型和病理类型(IAC/非 IAC)。②诊断特征: AI

算法/模型、提取特征、样本分组、模型分组;以 IAC 为阳性结果,非 IAC 为阴性结果,直接提取或经计算得出真阳性(true positive, TP)、假阳性(false positive, FP)、假阴性(false negative, FN)和真阴性(true negative, TN)的样本量,以金标准为参照组成四格表数据。此外,纳入研究一般将数据分为训练组和测试组两部分,前者用于对 AI 诊断模型进行训练,后者用于最终模型的性能检验。因此,本 Meta 分析仅提取测试组的数据。

5. 统计分析

统计分析软件包括 RevMan 5.3、MetaDiSc 1.4 和 Stata16.0。RevMan 5.3 用于绘制风险偏倚评价的条形图。MetaDiSc 1.4 用于检测阈值效应、异质性分析、合并指标的计算以及亚组分析。阈值效应引起的异质性使用敏感度对数与(1-特异度)对数之间的 Spearman 相关系数进行检测;非阈值效应引起的异质性使用 Cochran-Q 及 I^2 检验进行检测。组间异质性较低时($I^2 < 50\%$),采用固定效应模型合并指标;组间异质性较高时($I^2 \geq 50\%$),采用随机效应模型合并指标。评价指标包含:合并敏感度(pooled sensitivity, Sen 合并)、合并特异度(pooled specificity, Spe 合并)、合并阳性似然比(pooled positive likelihood ratio, PLR 合并)、合并阴性似然比(pooled negative likelihood ratio, NLR 合并)、合并诊断比值比(pooled diagnosis odds ratio, DOR 合并)、集成受试者操作特征(summary receiver operating characteristic, SROC)曲线以及曲线下面积(area under curve, AUC),其中 DOR 合并是主要合并指标。亚组分析依照研究对象的特点分组,通过 Meta 回归判断异质性来源,绘制各组的 SROC 曲线并得到 AUC 值,计算各组的 Sen 合并、Spe 合并、DOR 合并。Stata16.0 用于敏感性分析和发表偏倚检验。分别剔除敏感性分析筛出的高敏数据,重新绘制 SROC 曲线并比较 AUC 值来验证 AI 诊断性能的稳定性。Deeks 检验用于评估发表偏倚。以 $P < 0.05$ 为差异有统计学意义。

结 果

1. 纳入文献概述

通过检索共获得文献 1207

篇,12 篇文献(中文 4 篇,英文 8 篇)进入最终的分析^[12-23]。文献筛选流程见图 1,文献质量评价结果见图 2。纳入文献的一般特征和诊断特征分别见表 1 和表 2,共纳入 4066 例患者的 4325 枚肺结节,训练组和测试组的结节数分别为 3144 枚和 1181 枚,共提取了 34 组四格表数据。在性别构成上,10 篇文献根据患者总数列出,其中男性 1079 人,女性 2207 人;另有 2 篇根据结节总数列出,其中男性肺结节 297 枚,女性肺结节 526 枚。有 10 篇文献详细说明了结节类型,包括 2979 枚纯磨玻璃结节(pure ground glass nodules, pGGN)、797 枚混杂磨玻璃结节(mixed ground glass nodules, mGGN)和 113 枚实性结节;其余 2 篇仅指出结节类型为 GGN,共 436 枚。

2. Meta 分析结果

阈值效应和异质性检验:Spearman 相关系数为 0.154($P = 0.386 > 0.05$),意味着本研究不存在阈值效应。DOR 合并的 Cochran-Q 检验得出 Cochran-Q = 79.09($P < 0.001$),说明有非阈值效应造成的异质性;且 Sen 合并、Spe 合并、PLR 合并、NLR 合并、DOR 合并的 I^2 值分别为 71.6%、72.5%、71.7%、66.4%、58.3%, I^2 均 $> 50\%$,进一步说明了较高异质性的存在,故各指标需采取随机效应模型合并。

各指标合并结果:AI 模型的 Sen 合并 = 0.86 [95% 置信区间(95% CI):0.85~0.88], Spe 合并 =

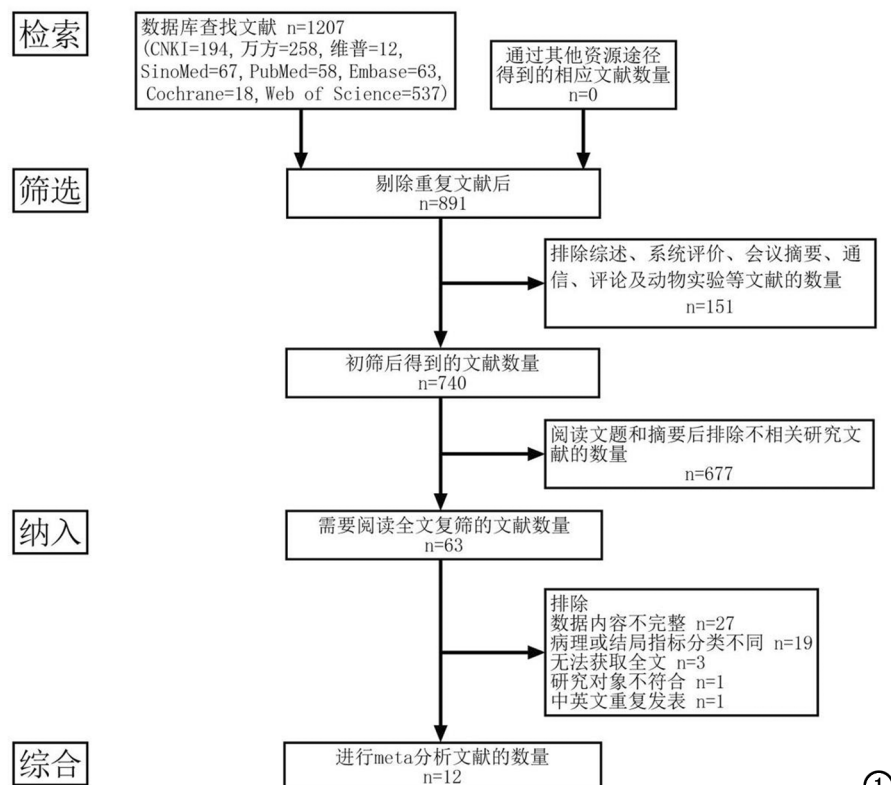


图 1 文献筛选流程图。

表 1 纳入研究的一般特征

纳入研究	资料来源	患者总数/ 肺结节总数	性别 (男/女)	结节类型	病理类型 IAC/ 非 IAC
Cui 2020 ^[12]	2011 年 11 月至 2017 年 10 月在天津医科大学肿瘤研究所和医院接受治疗的患者	437/437	109/328	pGGN;233 mGGN;204	116/321
Zhang 2021 ^[13]	2015 年 5 月至 2020 年 8 月在南京医科大学第一附属医院和复旦大学附属华东医院接受治疗的患者	709/709	201/508	pGGN	192/517
Cho 2020 ^[14]	三星医疗中心胸外科的肺癌外科登记数据库、釜山国立大学医院 2003 至 2013 年接受 GGN 完全切除的患者	233/233	105/128	GGN	126/107
Wu 2020 ^[15]	2013 年 1 月至 2017 年 10 月在三家不同医院治疗的患者	582/623	218/405 [*]	pGGN;210 mGGN;300 实性;113	391/232
Xu 2020 ^[16]	2015 年 1 月至 2017 年 10 月在邵逸夫爵士医院接受治疗的患者	275/322	72/203	pGGN	172/150
Cai 2021 ^[17]	2016 年 12 月至 2019 年 9 月在扬州苏北人民医院接受治疗的患者	136/136	37/99	pGGN	39/97
Park 2021 ^[18]	2010 年 8 月至 2016 年 6 月在蔚山大学医学院峨山医学中心接受治疗的患者	423/501	181/242	pGGN;208 mGGN;293	336/165
Xiong 2021 ^[19]	2012 年 11 月至 2018 年 6 月在大连医科大学第一附属医院接受治疗的患者	198/200	79/121 [*]	pGGN	120/80
Pan 2018 ^[20]	2013 年 1 月至 2017 年 10 月在广州医科大学附属第一医院接受治疗的患者	176/179	68/108	pGGN	76/103
Tao 2020 ^[21]	2013 年 2 月至 2018 年 5 月在中国人民解放军总医院第一医学中心接受治疗的患者	219/240	91/128	pGGN	96/144
Yin 2020 ^[22]	2015 年 1 月至 2018 年 10 月在大连大学附属中山医院、大连医科大学附属第二医院和大连市中心医院接受治疗的患者;2019 年 6 月至 8 月在大连大学附属中山医院、大连医科大学附属第二医院接受治疗的患者	482/542	148/334	pGGN	229/313
Zhang 2018 ^[23]	2015 年 1 月至 2017 年 12 月在河北医科大学第四医院进行过低剂量 CT 肺癌检查的人群	196/203	67/129	GGN	92/111

注: * 性别组成是根据结节数目统计的。

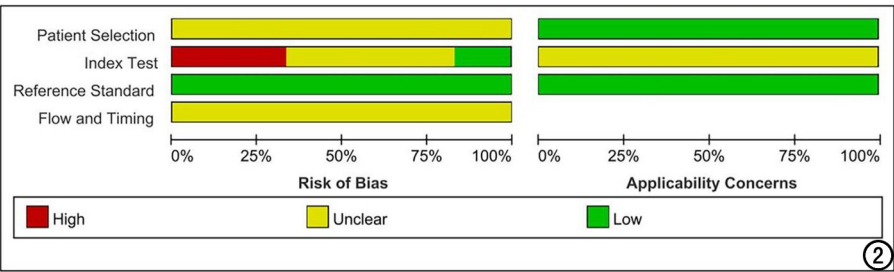


图 2 纳入文献质量评估结果(QUADAS-2)。

0.82(95%CI:0.80~0.84),PLR 合并=4.55(95%CI:3.61~5.75),NLR 合并=0.19(95%CI:0.15~0.23),DOR 合并=28.31(95%CI:19.76~40.54),总体 AUC 为 0.9110(图 3~6)。

亚组分析:依据不同的 AI 算法将 34 组数据分成 3 类,分别为随机森林(random forest,RF)14 组、支持向量机(support vector machine,SVM)7 组以及“其他”(Others)13 组;依据测试组样本数量是否≥50 例分成 2 类(是:26 组;否:8 组)。基于此两种分类进行 Meta 回归,分析显示 AI 算法的差异($P=0.0028$)或许是产生异质性的原因之一,样本量的差异($P=0.2761$)无法作为异质性来源的依据。重新合并按不同算法分组的各项指标(表 3),RF 组的 Sen 合并(0.91)、DOR 合并(52.91)和 AUC(0.9412)最高,而 Others 组的 Spe 合并(0.84)最高。

敏感性分析和发表偏倚:图 7 显示第 12、13 和 28 组数据存在较强敏感性,分别剔除这 3 组数据后的 AUC 为 0.9077、0.9074 和 0.9138,全部剔除这 3 组数据后的 AUC 为 0.9063,该结果与总体 AUC(0.9110)差别不大,说明此次 Meta 分析敏感性较低,结果稳定可靠。Deeks 偏倚检验结果显示

$P=0.06>0.05$,表明漏斗图基于回归线对称,无发表偏倚。

讨 论

随着 AI 在医学领域应用的不断深化和细分,在肺腺癌诊断方面,基于影像组学和 AI 的良恶性鉴别、恶性程度分级等研究已广泛开展^[24]。然而,由于组学模型泛化不易,纹理特征的鲁棒性不同,DL 又需要比组学模型更多的样本用于训练,导致不同研究间的 AI 诊断性能差异显著^[25-26]。GGN 作为初期肺腺癌的首要影像表现,AI 对其浸润性的预测同样受上述因素制约,所以本文对基于不同 AI 算法的 CT 影像预测肺腺癌浸润性的诊断能力进行 Meta 分析。最终纳入 12 项研究的 4325 枚肺结节,其中绝大多数(97%,4212/

表 2 纳入研究的诊断特征

纳入研究	AI 算法	提取特征	样本分组	模型(序号)	TP	FP	FN	TN
Cui 2020 ^[12]	CAD	CT 图像特征	437(291 训练组+146 测试组)	CAD(1)	31	32	5	78
Zhang 2021 ^[13]	3D-CNN	整个病灶的感兴趣容积区域	709(532 训练组+177 测试组)	3D-CNN(2)	34	13	14	116
Cho 2020 ^[14]	LR	影像组学特征	233(186 训练组+47 测试组)	LR1(3)	34	1	4	8
	LR	影像组学和病灶边缘特征	233(186 训练组+47 测试组)	LR2(4)	34	0	4	9
	SVM	影像组学特征	233(186 训练组+47 测试组)	SVM1(5)	30	1	8	8
	SVM	影像组学和病灶边缘特征	233(186 训练组+47 测试组)	SVM2(6)	32	1	6	8
	RF	影像组学特征	233(186 训练组+47 测试组)	RF1(7)	31	1	7	8
	RF	影像组学和病灶边缘特征	233(186 训练组+47 测试组)	RF2(8)	33	1	5	8
	RF	病灶体积和影像组学特征	474(331 训练组+143 测试组 1)	1RV(9)	73	15	13	42
	RF	临床和语义、体积及影像组学特征	474(331 训练组+143 测试组 1)	1CSRV(10)	74	13	12	44
	RF	术中冰冻、体积和影像组学特征	474(331 训练组+143 测试组 1)	1FSRV(11)	78	3	8	54
	RF	术中冰冻、体积特征	474(331 训练组+143 测试组 1)	1FSV(12)	85	8	1	49
	RF	临床和语义、体积、术中冰冻及影像组学特征	474(331 训练组+143 测试组 1)	1CSFSRV(13)	84	9	2	48
	RF	病灶体积和影像组学特征	480(331 训练组+149 测试组 2)	2RV(14)	100	14	5	30
	RF	临床和语义、体积及影像组学特征	480(331 训练组+149 测试组 2)	2CSRV(15)	98	12	7	32
	RF	术中冰冻、体积和影像组学特征	480(331 训练组+149 测试组 2)	2FSRV(16)	93	2	12	42
Wu 2020 ^[15]	RF	术中冰冻、体积特征	480(331 训练组+149 测试组 2)	2FSV(17)	97	3	8	41
	RF	临床和语义、体积、术中冰冻及影像组学特征	480(331 训练组+149 测试组 2)	2CSFSRV(18)	97	5	8	39
Xu 2020 ^[16]	SVM+RF	临床和放射学定量特征	322(258 训练组+64 测试组)	con(19)	28	12	6	18
	SVM+RF	影像组学特征	322(258 训练组+64 测试组)	rad(20)	28	9	6	21
	SVM+RF	临床、放射学定量和影像组学特征	322(258 训练组+64 测试组)	mix(21)	31	11	3	19
Cai 2021 ^[17]	SVM	影像组学特征	136(95 训练组+41 测试组)	SVM(22)	11	3	1	26
Park 2021 ^[18]	3D-CNN	病灶的三维边界框并生成掩膜	501(411 训练组+90 测试组)	3D-CNN(23)	32	6	13	39
Xiong 2021 ^[19]	LR	影像组学特征	200(140 训练组+60 测试组)	LR(24)	30	6	6	18
	SVM	影像组学特征	200(140 训练组+60 测试组)	SVM(25)	32	9	4	15
	NBC	影像组学特征	200(140 训练组+60 测试组)	NBC(26)	22	4	14	20
	KNN	影像组学特征	200(140 训练组+60 测试组)	KNN(27)	29	7	7	17
	DT	影像组学特征	200(140 训练组+60 测试组)	DT(28)	29	12	7	12
	RF	影像组学特征	200(140 训练组+60 测试组)	RF(29)	31	11	5	13
Pan 2018 ^[20]	LR	影像组学特征	179(125 训练组+54 测试组)	LR(30)	15	3	8	28
Tao 2020 ^[21]	3D-CNN	CT 图像的感兴趣容积区域	240(192 训练组+48 测试组)	3D-CNN(31)	14	2	6	26
Yin 2020 ^[22]	LR	CT 图像特征	542(442 训练组+100 测试组)	LR(32)	28	5	13	54
	ANN	CT 图像特征	542(442 训练组+100 测试组)	ANN(33)	33	4	8	55
Zhang 2018 ^[23]	KNN	影像组学特征	203(141 训练组+62 测试组)	KNN(34)	20	2	8	32

注:CAD:计算机辅助诊断;3D-CNN:三维卷积神经网络;LR:逻辑回归;SVM:支持向量机;RF:随机森林;NBC:朴素贝叶斯;KNN:K-近邻;DT:决策树;ANN:人工神经网络;TP:真阳性;FP:假阳性;FN:假阴性;TN:真阴性。

表 3 各亚组合并结果

组别	Sen 合并(95%CI)	Spe 合并(95%CI)	DOR 合并(95%CI)	AUC
随机森林	0.91(0.89~0.92)	0.83(0.79~0.86)	52.91(28.58~97.59)	0.9412
支持向量机	0.85(0.80~0.89)	0.71(0.64~0.78)	14.46(8.22~25.43)	0.8942
其他	0.77(0.73~0.81)	0.84(0.80~0.87)	17.85(11.67~27.30)	0.8464

4325)为 GGN。此外,无论是按患者数还是按结节数统计,女性均占大多数,分别为 67%(2207/3286)和 64%(526/823),与文献报道的近年显著升高的女性肺腺癌发病率相符^[27]。

本次分析无阈值效应,但研究间异质性较高,故对各指标使用随机效应模型合并。DOR 合并为主要合并指标,代表诊断试验结果与疾病的关系水平;当 DOR 合并>1 时,其值越大,该试验的诊断能力就越高^[28]。本文 DOR 合并=28.31,说明了 AI 对肺腺癌

浸润性较高的诊断性能。Sen 合并、Spe 合并分别为 86%、82%,说明 AI 确定 IAC 和识别非 IAC 的能力都相对较高。PLR 合并显示 AI 正确识别 IAC 的概率是错误识别概率的 4.55 倍,NLR 合并显示错误识别非 IAC 的概率是正确识别概率的 0.19 倍。总体 AUC 为 0.9110,也证明 AI 的诊断性能优异。

根据 Meta 回归结果,AI 算法的差异可推断为异质性的来源。从表 3 可知,RF 组的 Sen 合并、DOR 合并和 AUC 均高于另外两组,说明基于 CT 影像组学特

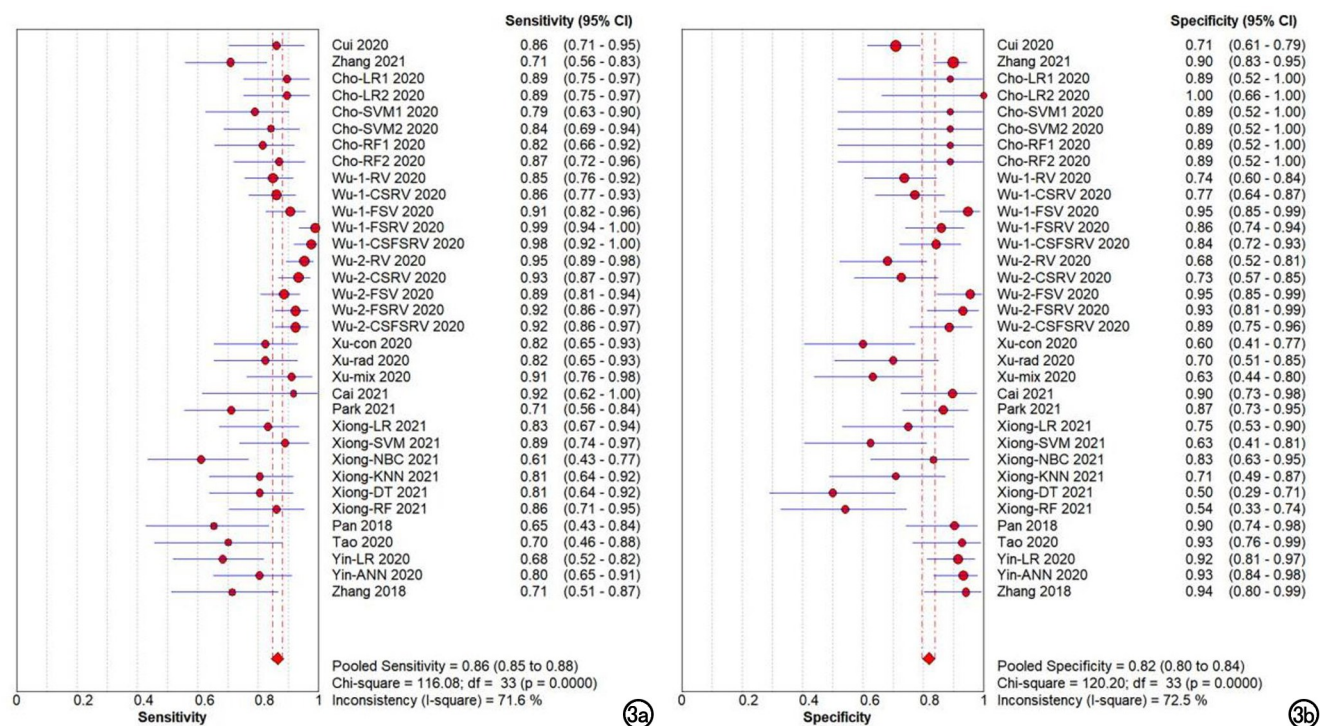


图 3 a) AI 模型的合并敏感度; b) 合并特异度。

征构建的 AI 模型中, RF 算法对 IAC 的诊断性能好于其他机器学习 (machine Learning, ML) 算法, 可能代表了其更好的泛化能力。当然, RF 的高诊断性能也不能排除是因为纳入分析组数较多、代表性较强导致的结果。需要指出的是, 这也不能说明 ML 要比 DL 的诊断性能更好; 相反, 使用基于 DL 的卷积神经网络 (convolutional neural networks, CNN) 进行肺结节分

类在多数场合下是优于现有的计算机辅助诊断技术的^[29-31], 采用 CNN 对肺腺癌分类的研究也逐年增多, 在诊断性能上已达到甚至略超过高年资放射科医生的水平^[32-34]。从本文看, 仅有三组四格表数据是基于 CNN 算法的结果, 因而在分组中被归入了“其他”一组, 且样本量较少, 均导致其诊断价值体现不足。此外, 在文献筛选过程中, 部分无法提取四格表信息的基

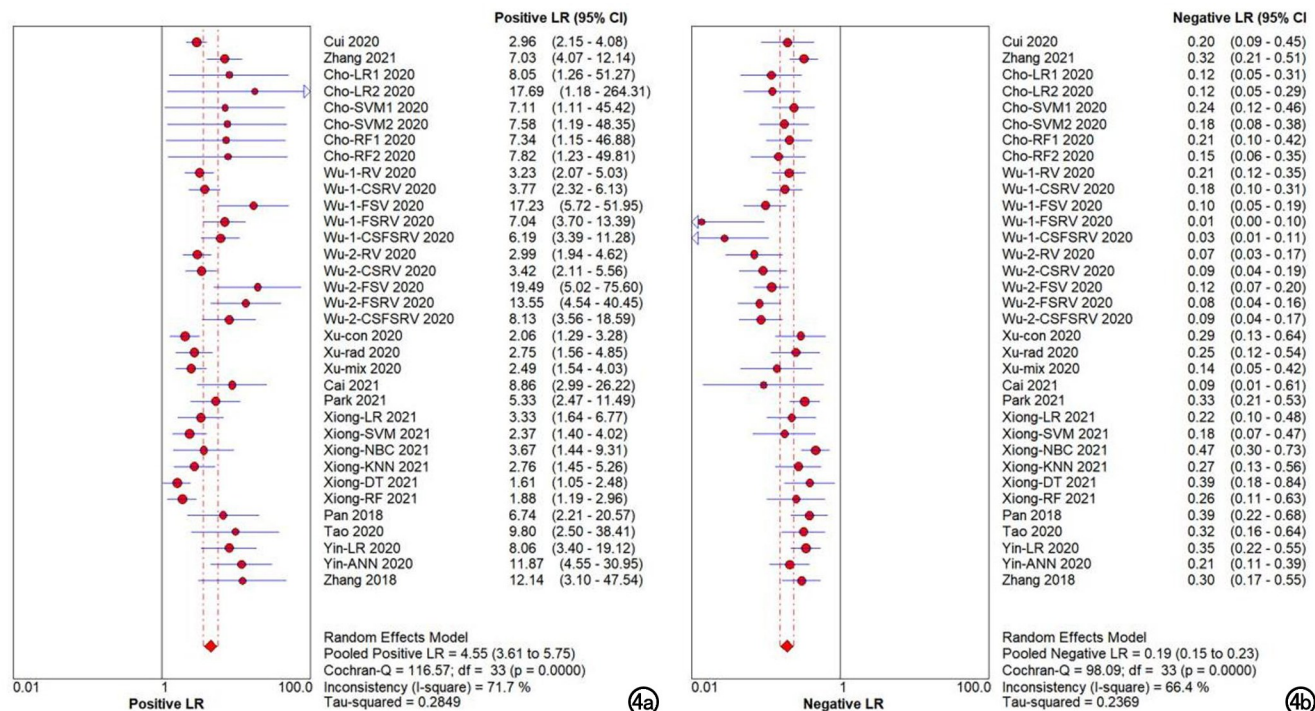


图 4 a) AI 模型的合并阳性似然比; b) 合并阴性似然比。

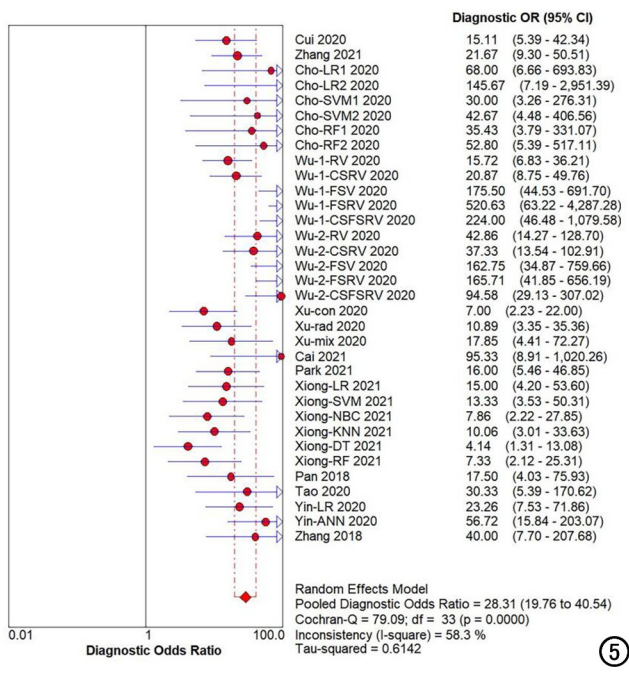


图 5 AI 模型的合并诊断比值比。

于 DL 的研究也被排除,同样对其诊断价值的判断存在影响。但是,本次亚组分析已证实了不同 AI 算法是引起异质性的原因,而且最终 SROC 曲线的 AUC 为 0.9110,排除较高敏感性数据后的 AUC 也都在此值附近轻微波动,显示了综合了 ML 和 DL 的 AI 在预测肺腺癌浸润性的能力上的高效性和稳定性。

当前 AI 取得的成就固然可喜,但依然处在初步验证和应用阶段,还有相当的改进空间,需要与临床医生密切配合才能发挥其价值。如本文中, AI 对肺腺癌浸润性的诊断仍存在一定的漏诊率和误诊率(分别为 14% 和 18%),而对组学特征的进一步挖掘、扩大训练样本量或进一步提升有限数据集上算法的性能都是可行的解决方法。这些工作都需要医生工作的支持,其中放射科医生在判断新兴 AI 技术的临床应用价值上有重要作用,并在整合互补成像技术信息与临床数据的工作中担任指导者角色,从而能提供更完善的诊断

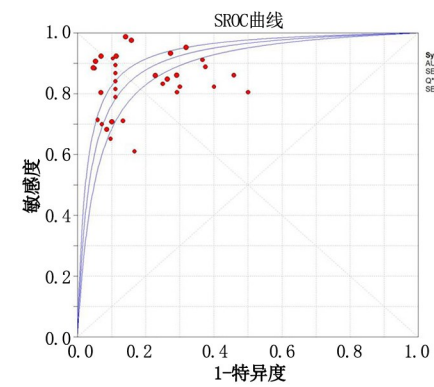


图 6 AI 模型的 SROC 曲线。

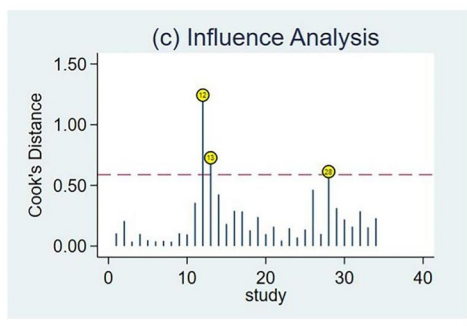


图 7 敏感性分析结果。

信息^[26]。此外,即便 AI 已能独立诊断,但最终仍需要放射科医生签发报告并承担法律风险^[35],因此,如果能通过立法明确医生与提供 AI 服务的公司间各自的责任范围,也能促进 AI 诊断服务的普及。最后,目前医生对 AI 的积极态度也为二者的深度协同提供了有利前提^[36]。

本文存在以下不足:①收集的是二分类数据(IAC 与非 IAC),其他二分类方法(如浸润前病变和浸润性病变)、多分类方法(如对 IAC 的病理亚型细分)以及包含了良性结节的分类方式均被排除,可能会对诊断价值的评判产生影响。但如前文所言,以 IAC 和非 IAC 进行区分更具有临床意义,因而基于本文分类所得的诊断价值可能更具实用性。②因部分研究无法提取四格表数据或所含样本量较少,其代表的算法的诊断价值可能无法被真实反映。③尽管 Meta 回归显示不同的 AI 算法是异质性产生的原因,但是其他方面,如患者间的差异、CT 设备或扫描协议的不一致、提取特征的数量差距或组学特征的不同等,都可能是出现异质性的因素,有待深入研究。④仅纳入了以中、英两种语言发表的文章,其他语种的文献或被漏掉而造成偏倚。

综上所述,本 Meta 分析显示基于 AI 的 CT 影像对肺腺癌浸润性有较高的预测能力,能为临床上制定随访策略或手术方案提供更详细的参考依据,具有实际应用的价值。但 AI 诊断目前仍有一定的漏诊率和误诊率,获取更详实的患者资料、采用结构化的影像报告、收集多中心的大量样本、深度挖掘纹理特征和优化神经网络架构等办法将有助于进一步提升 AI 诊断工具的性能。同时, AI 与医生的深度协同能促进工作效率的提高,增加医疗资源利用率,为实现肺癌的早筛早诊早治提供一条更科学高效的途径。

参考文献:

[1] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.

[2] Shi JF, Wang L, Wu N, et al. Clinical characteristics and medical service utilization of lung cancer in China, 2005 — 2014: Overall design and results from a multicenter retrospective epidemiologic survey [J]. Lung Cancer, 2019, 128:91-100.

[3] National Lung Screening Trial Research Team, Church TR, Black WC, et al. Results of initial low-dose computed tomographic screening for lung cancer[J]. N Engl J Med, 2013,

- 368(21);1980-1991.
- [4] Maldonado F, Boland JM, Raghunath S, et al. Noninvasive characterization of the histopathologic features of pulmonary nodules of the lung adenocarcinoma spectrum using computer-aided nodule assessment and risk yield (CANARY)——a pilot study[J]. J Thorac Oncol, 2013, 8(4): 452-460.
 - [5] Son JY, Lee HY, Kim JH, et al. Quantitative CT analysis of pulmonary ground-glass opacity nodules for distinguishing invasive adenocarcinoma from non-invasive or minimally invasive adenocarcinoma: the added value of using iodine mapping[J]. Eur Radiol, 2016, 26(1): 43-54.
 - [6] Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification[J]. J Thorac Oncol, 2015, 10(9): 1243-1260.
 - [7] Pedersen JH, Saghir Z, Wille MM, et al. Ground-glass opacity lung nodules in the Era of lung cancer CT screening: radiology, pathology, and clinical management[J]. Oncology, 2016, 30(3): 266-274.
 - [8] Hattori A, Hirayama S, Matsunaga T, et al. Distinct clinicopathologic characteristics and prognosis based on the presence of ground glass opacity component in clinical stage IA lung adenocarcinoma[J]. J Thorac Oncol, 2019, 14(2): 265-275.
 - [9] Ye T, Deng L, Xiang J, et al. Predictors of pathologic tumor invasion and prognosis for ground glass opacity featured lung adenocarcinoma[J]. Ann Thorac Surg, 2018, 106(6): 1682-1690.
 - [10] Avanzo M, Stancanello J, Pirrone G, et al. Radiomics and deep learning in lung cancer[J]. Strahlenther Onkol, 2020, 196(10): 879-887.
 - [11] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies[J]. Ann Intern Med, 2011, 155(8): 529-536.
 - [12] Cui X, Heuvelmans MA, Fan S, et al. A subsolid nodules imaging reporting system (SSN-IRS) for classifying 3 subtypes of pulmonary adenocarcinoma[J]. Clin Lung Cancer, 2020, 21(4): 314-325.
 - [13] Zhang T, Wang Y, Sun Y, et al. High-resolution CT image analysis based on 3D convolutional neural network can enhance the classification performance of radiologists in classifying pulmonary non-solid nodules[J]. Eur J Radiol, 2021, 141: 109810.
 - [14] Cho HH, Lee G, Lee HY, et al. Marginal radiomics features as imaging biomarkers for pathological invasion in lung adenocarcinoma[J]. Eur Radiol, 2020, 30(5): 2984-2994.
 - [15] Wu G, Woodruff HC, Sanduleanu S, et al. Preoperative CT-based radiomics combined with intraoperative frozen section is predictive of invasive adenocarcinoma in pulmonary nodules: a multi-center study[J]. Eur Radiol, 2020, 30(5): 2680-2691.
 - [16] Xu F, Zhu W, Shen Y, et al. Radiomic-based quantitative CT analysis of pure ground-glass nodules to predict the invasiveness of lung adenocarcinoma[J]. Front Oncol, 2020, 10: 872.
 - [17] Cai J, Liu H, Yuan H, et al. A radiomics study to predict invasive pulmonary adenocarcinoma appearing as pure ground-glass nodules[J]. Clin Radiol, 2021, 76(2): 143-151.
 - [18] Park S, Park G, Lee SM, et al. Deep learning-based differentiation of invasive adenocarcinomas from preinvasive or minimally invasive lesions among pulmonary subsolid nodules[J]. Eur Radiol, 2021, 31(8): 6239-6247.
 - [19] Xiong Z, Jiang Y, Che S, et al. Use of CT radiomics to differentiate minimally invasive adenocarcinomas and invasive adenocarcinomas presenting as pure ground-glass nodules larger than 10mm[J]. Eur J Radiol, 2021, 141: 109772.
 - [20] 潘小环. 薄层 CT 及影像学鉴别表现为纯磨玻璃结节的肺腺癌不同病理亚型的研究[D]. 广东: 广州医科大学, 2018.
 - [21] 陶雪敏, 方瑞, 吴重重, 等. 深度学习模型对纯磨玻璃结节肺腺癌病理亚型的预测分析[J]. 中国医学科学院学报, 2020, 42(4): 477-484.
 - [22] 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.
 - [23] 张晶. 早期肺腺癌低剂量 CT 定量测量指标及影像学特征与病理的相关研究[D]. 石家庄: 河北医科大学, 2018.
 - [24] 周小君, 韩丹, 张正华, 等. 基于人工智能和影像组学的胸部 CT 在早期肺腺癌中的研究进展[J]. 放射学实践, 2021, 36(7): 934-937.
 - [25] Chassagnon G, Vakalopoulou M, Paragios N, et al. Artificial intelligence applications for thoracic imaging[J]. Eur J Radiol, 2020, 123: 108774.
 - [26] Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications[J]. Clin Radiol, 2020, 75(1): 13-19.
 - [27] Byun J, Schwartz AG, Lusk C, et al. Genome-wide association study of familial lung cancer[J]. Carcinogenesis, 2018, 39(9): 1135-1140.
 - [28] 李幼平. 实用循证医学[M]. 北京: 人民卫生出版社, 2018: 539-549.
 - [29] Causey JL, Zhang J, Ma S, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans[J]. Sci Rep, 2018, 8(1): 9286.
 - [30] Lyu J, Ling SH. Using multi-level convolutional neural Network for classification of lung nodules on CT images[J]. Annu Int Conf IEEE Eng Med Biol Soc, 2018, 2018: 686-689.
 - [31] Xie Y, Xia Y, Zhang J, et al. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT[J]. IEEE Trans Med Imaging, 2019, 38(4): 991-1004.
 - [32] Gong J, Liu J, Hao W, et al. A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images[J]. Eur Radiol, 2020, 30(4): 1847-1855.
 - [33] Wang J, Chen X, Lu H, et al. Feature-shared adaptive-boost deep learning for invasiveness classification of pulmonary subsolid nodules in CT images[J]. Med Phys, 2020, 47(4): 1738-1749.
 - [34] Gong J, Liu J, Li H, et al. Deep learning-based stage-wise risk stratification for early lung adenocarcinoma in CT images: a multi-center study[J]. Cancers (Basel), 2021, 13(13): 3300.
 - [35] Vogel L. Rise of medical AI poses new legal risks for doctors[J]. CMAJ, 2019, 191(42): E1173-E1174.
 - [36] Oh S, Kim JH, Choi SW, et al. Physician confidence in artificial intelligence: an online mobile survey[J]. J Med Internet Res, 2019, 21(3): e12422.

(收稿日期: 2021-10-08 修回日期: 2022-02-28)