# AI辅助CT在肺结节良、恶性诊断中的系统评价

张 蕾[1a]，杨嘉琪[1a]，辛小霞[2]，符文杰[3]，张常青[2*]，樊景春[1b*]

1 甘肃中医药大学 a 第一临床医学院，b 循证医学研究中心，甘肃 兰州 730000；2 甘肃省人民医院新区分院 放射科，甘肃 兰州 730207；3 甘肃省人民医院 放射科，甘肃 兰州 730000

**摘要：目的** 评价人工智能（AI）辅助计算机断层扫描（CT）诊断肺结节良、恶性的价值。**方法** 通过检索PubMed、The Cochrane Library、Embase、Web of Science、中国知网、中国生物医学文献服务系统，搜集AI结合CT评估肺结节良、恶性的相关研究。对符合纳入、排除标准的文献使用QUADAS-2量表进行方法学质量评估。运用RevMan5.3和Stata18.0进行数据整合与分析。分组比较AI与临床医师在肺结节性质判别中的诊断效能，用$Z$检验比较2组的差异，并构建综合受试者操作特征曲线；用异质性检验分析研究间差异的影响因素，以AI系统类型分组进行亚组分析，比较不同系统间敏感度、特异度等指标的差异。**结果** AI辅助CT诊断与医生阅片的合并诊断效能指标敏感度、特异度、阳性似然比、阴性似然比、诊断比值比，及综合受试者操作特征曲线下面积，差异均无统计学意义（$P$>0.05）。**结论** AI辅助CT在肺结节性质判别中展现出与传统医生阅片相当的诊断效能，敏感度略高，提示AI辅助CT在降低漏诊率方面或有潜力，但需大规模临床验证。

**关键词：**计算机断层扫描；人工智能；肺结节；人工智能辅助诊断；深度学习；Meta分析

**中图分类号：**R816.4 **文献标识码：**A DOI：10.13885/j.issn.2097-681X.T20250032

# Comparative performance of AI and radiologists in diagnosing pulmonary nodules on CT

ZHANG Lei[1a], YANG Jiaqi[1a], XIN Xiaoxia[2], FU Wenjie[3],
ZHANG Changqing[2*], FAN Jingchun[1b*]

1 a The First Clinical Medical College, b Center for Evidence-based Medicine, Gansu University of Chinese Medicine, Lanzhou 730000, China; 2 Radiology Department, New District Branch of Gansu Provincial Hospital, Lanzhou 730207, China; 3 Department of Radiology, Gansu Provincial Hospital, Lanzhou 730000, China

**Abstract: Objective** To evaluate artificial intelligence (AI)-assisted computed tomography (CT) in diagnosing benign and malignant pulmonary nodules. **Methods** Systematic searches were conducted through PubMed, The Cochrane Library, Embase, Web of Science, the China National Knowledge Infrastructure and China Biology Medicine disc for studies on AI, integrated CT evaluation of pulmonary nodules. By searching PubMed, The Cochrane Library, Embase, Web of Science, the China National Knowledge Infrastructure and China Biology Medicine disc to collect relevant studies on AI combined with CT evaluation of benign and malignant pulmonary nodules. According to the established inclusion and exclusion criteria, the QUADAS-2 scale was used to evaluate the methodological quality of literature that met the requirements. RevMan5.3 and Stata18.0 were used for data integration and analysis, comparing pooled diagnostic metrics between AI sys-

**收稿日期：**2025-05-27 **修回日期：**2025-06-29 **接受日期：**2026-01-21

**基金项目：**甘肃省自然科学基金资助项目（23JRRA1292）

**作者简介：**张常青，男，主任医师，研究方向为胸部影像诊断，e-mail：13893212070@126.com，通信作者；
樊景春，女，副教授，研究方向为临床流行病学与循证医学，e-mail：fan_jc@126.com，通信作者
*共同通信作者

tems and radiologists, specificity, sensitivity, negative likelihood ratio, positive likelihood ratio, and diagnostic odds ratio. Inter-group differences were statistically tested using Z-tests. Summary receiver operating characteristic curves were constructed to visualize diagnostic accuracy, with heterogeneity explored through statistics. Subgroup analyses stratified by AI systems were conducted to compare the differences in sensitivity, specificity, and other indicators between different systems. **Results** Pooled metrics demonstrated the following comparisons (AI-assisted CT vs. radiologist interpretation): sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio, and summary receiver operating characteristiccurve area under the curve. Statistical comparisons revealed no significant inter-group differences (all $P>0.05$). **Conclusion** AI-assisted CT demonstrates diagnostic efficacy comparable to conventional radiologist interpretation in pulmonary nodule characterization. The marginally higher sensitivity suggests potential utility for reducing missed diagnoses, though large-scale clinical validation remains warranted.

**Keywords:** computed tomography; artificial intelligence; pulmonary nodule; artificial intelligence-assisted diagnosis; deep learning; Meta-analysis

肺癌是全球由癌症致死的主要原因，约占所有癌症新发病例的11.4%和癌症死亡的18.0%[1]。肺癌是男性癌症发病和死亡的第三大常见原因，是女性癌症死亡的第二大常见原因[2]。作为最大的发展中国家，中国肺癌的发病率和死亡率是发达国家的3~4倍，2022年肺癌新发病例高达106.06万例，死亡病例达257.42万例[3]。肺癌死亡率随着时间的推移而增加，凸显了其对公共卫生威胁的严重性。随着中国工业化、城市化和人口老龄化加剧，肺癌已严重威胁到中国居民的健康，并成为阻碍中国经济发展的主要公共卫生问题，由于肺癌早期通常无明显症状，待出现咳嗽、体重减轻、胸痛等明显症状时，癌细胞可能已扩散，导致诊断时多为晚期，治疗难度增大。目前仅有部分高危人群会进行低剂量螺旋计算机断层扫描（computed tomography, CT）进行筛查，但筛查的普及程度和可及性有限，在一些偏远、欠发达地区缺乏相应设备和技术人员，无法开展大规模筛查。而通过早期筛查发现结节并排除恶性是降低肺癌死亡率的关键。早期发现肺结节提示潜在的早期肺癌或其他肺部病变。通过及时检测和干预，可以显著提高治愈率，避免病情恶化。早期肺癌的5年生存率可达90%以上，而晚期的则为10%~15%，所以肺结节筛查、早期诊断对于提高肺癌治愈率、降低死亡率、减少误诊和提供治疗机会具有重要意义。研究[3]表明，亚临床期肺癌多表现为肺部微小局灶性病灶（即肺结节），评估影像学特征成为原发性肺癌筛查的核心环节。肺结节指肺内直径≤3 cm的类圆形病灶，影像表现为密度增高的实性或亚实性（含磨玻璃）阴影，边界清晰，不伴肺不张、肺门淋巴结肿大和胸腔积液[3]。研究[4]显示约70%的偶发肺结节属于良性病变，因此准确鉴别肺结节的性质对于肺癌的早期筛查具有重要意义。

目前，胸部X线片、CT、磁共振成像、正电子发射断层扫描/CT等多种检查方法可用于肺结节检测[3,5-6]。然而，肺结节的形态和大小各异，传统CT结果需靠医生专业知识和临床经验判读，大量CT图像加大了影像和临床医师的工作量进而影响阅片效果，存在一定的漏诊率和误诊率。人工智能（artificial intelligence, AI）辅助诊断技术问世后，极大地提升了肺结节的检出效率，AI在CT影像分析上的速度远超传统人工阅片[7]，其应用不仅大幅减轻了放射医师的工作强度，还提高了筛查的可靠性[8]，医疗行业正加速推进AI肺结节检测技术的落地。

## 1 资料与方法

### 1.1 纳入与排除标准

纳入标准：（1）研究类型为基于AI联合CT评估肺结节性质的非随机对照试验文献；（2）研究对象为应用AI辅助CT并取得确切诊断的患者；（3）诊断方法为各类AI算法分析影像资料与医生阅片的分组诊断，并以组织病理学诊断为金标准。（4）评价指标为特异度、敏感度、阴性似然比、阳性似然比、诊断比值比、受试者操作特征曲线下面积（area under the curve, AUC）。

排除标准：（1）无法获取原始数据或关键数

据缺失的研究文献。（2）重复发表文献；（3）非实验研究类文献（包括系统综述、个案报道、会议论文及技术说明类文章）。

### 1.2 文献检索策略

使用计算机检索文献数据库包括 PubMed、The Cochrane Library、Embase、Web of Science、中国知网和中国生物医学文献服务系统。检索年限为建库至为2024年7月。英文主题词检索："Solitary Pulmonary Nodule/diagnostic imaging" OR "Multiple Pulmonary Nodules/diagnostic imaging" OR "Artificial Intelligence"。主题词检索：肺结节、人工智能。

### 1.3 文献筛选、数据提取、纳入研究的质量评价

#### 1.3.1 文献筛选流程

（1）初步去重：将各数据库检索到的文献导入EndNote软件，利用软件的自动查重功能移除重复记录。对于潜在重复文献，保留唯一记录，移除重复项。（2）初筛（标题/摘要筛选）：由两名研究者独立审阅剩余文献的标题和摘要，根据预先制定的纳入排除标准初步筛选。排除明显不符合要求的研究（如非AI+CT诊断肺结节、非诊断性研究、非原始研究等）。筛选过程中如有分歧，通过讨论或咨询第三位研究者解决。（3）全文筛选：对通过初筛的文献，由上述两名研究者独立获取并精读全文，严格按照纳入、排除标准进行最终筛选。记录排除原因（如数据缺失、无金标准对照、非实验性研究等）。分歧通过讨论或第三位研究者裁决解决。（4）最终纳入：经过以上步骤，确定最终纳入系统评价的文献。

#### 1.3.2 数据提取

（1）制定提取表格：根据研究目的，预先设计标准化的数据提取表格。（2）独立双人提取：由两名研究者独立从最终纳入的每篇文献中提取数据，填入表格。（3）提取内容：提取的信息包括但不限于：第一作者、发表年份、研究设计、研究对象特征、AI系统名称及类型、医生阅片方式、金标准、诊断性能四格表数据（真阳性数、假阳性数、假阴性数、真阴性数）、其他相关指标（如AUC、敏感度、特异度等）。（4）交叉核对：两名研究者完成各自提取后，交叉核对。对于提取不一致的数据，通过复核原文和讨论达成共识。如仍存在争议，由第三位研究者裁定。

#### 1.3.3 纳入研究的质量评价

评价工具：采用QUADAS-2（Quality Assessment of Diagnostic Accuracy Studies-2）量表对所有纳入研究进行方法学质量评价。该量表从病例选择、待评价试验、参考标准、病例流程与进展情况四个领域评估偏倚风险。

独立双人评价：由两名研究者独立使用QUADAS-2量表对每篇纳入研究进行评价。

分歧解决：评价过程中出现的分歧，通过讨论解决。若讨论后仍无法达成一致，由第三位研究者做出最终判断。

#### 1.3.4 质量控制措施

预先培训：在筛选、提取和质量评价工作开始前，对所有参与的研究者进行统一培训，确保其对纳入/排除标准、数据提取项目定义、QUADAS-2量表各条目的理解完全一致。

双人独立操作：文献筛选、数据提取、质量评价三个核心环节均采用"双人独立"方式进行，有效降低主观误差。

交叉核对与讨论：在筛选和提取阶段，独立操作后进行交叉核对，对不一致处进行讨论并复核原文，确保数据准确。

分歧仲裁机制：明确规定当双人独立操作出现分歧且讨论无法解决时，由指定的资深研究者进行仲裁，保证决策的权威性。

过程记录：详细记录每一步筛选的排除原因、数据提取过程中的疑问及解决方式、质量评价中的分歧及裁决结果，确保过程可追溯。

预实验：在正式筛选和数据提取前，可选取少量代表性文献进行预筛选和预提取，校准研究者间的理解和操作一致性。

QUADAS-2应用：使用国际公认的QUADAS-2工具进行质量评价，提供了结构化和标准化的评价框架，提高了评价的客观性和可比性。

### 1.4 统计学分析

使用RevMan5.3和Stata18.0进行数据分析。使用$Q$检验和$I^2$判断异质性：当$I^2<50\%$时，判定为低度异质性，选用固定效应模型；否则采用随机效应模型；列出包含假阴性数、真阴性数、真阳性数、假阳性数等参数的四格表，分别计算AI组和医生组的阳性似然比、阴性似然比、综合特

异度、综合灵敏度，及综合诊断比值比等指标，并计算其95% *CI*。绘制综合受试者操作特征曲线及其曲线下的面积量化诊断效能；采用*Z*检验比较AI组与医生组间的合并效应量差异。AUC的诊断效能可分为4个等级：AUC<0.5提示无临床诊断意义；0.7>AUC≥0.5为有限诊断价值；0.9>AUC≥0.7为具有良好诊断价值；AUC≥0.9则表明诊断效能极佳[9]；采用亚组分析探寻引起异质性的因素。以纳入研究使用的AI系统作为分组依据进行亚组分析，采用随机效应模型，比较不同系统间敏感度、特异度等指标的差异。

## 2 结果

### 2.1 文献筛选流程及结果

执行检索策略后，初步获得2 697篇文献，在完成严格的筛选程序后，共纳入8篇文献（中文5篇，英文3篇），合计1 151例研究对象。详见图1所示的筛选流程及结果，研究基本特征数据详见表1。

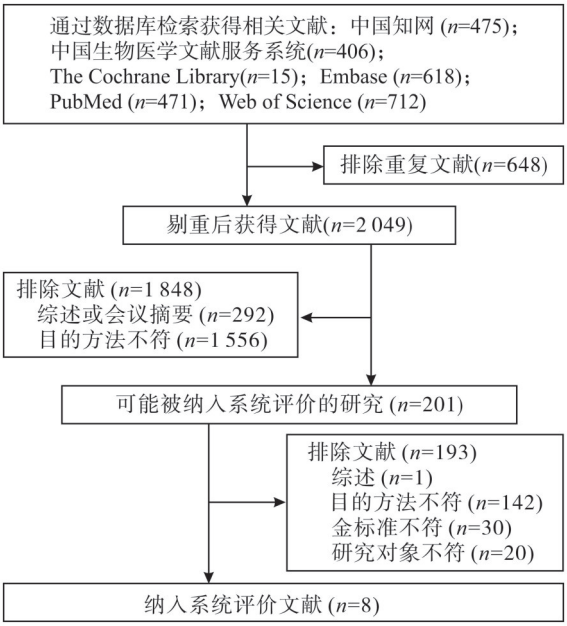


图1 文献筛选流程图

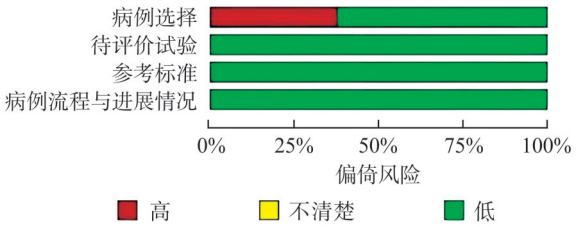


图2 质量评价结果

### 2.2 纳入研究的质量评价

纳入研究的质量评价结果见图2。纳入研究的偏倚风险主要集中于病例选择环节（32.5%高风险），可能影响内部真实性；适用性问题轻微（仅12.5%病例选择存在担忧），整体外部有效性可接受。

### 2.3 Meta分析结果

#### 2.3.1 合并效应量

使用Stata18.0分别绘制AI组与医生组的森林图

表1 纳入研究基本特征统计

| 文献 | 组别 | 真阳性数/例 | 假阳性数/例 | 假阴性数/例 | 真阴性数/例 | 人工智能系统 |
|---|---|---|---|---|---|---|
| [10] | 人工智能 | 159 | 8 | 12 | 15 | 不详 |
| | 医生 | 151 | 8 | 20 | 15 | |
| [11] | 人工智能 | 62 | 30 | 4 | 194 | 医准智能 |
| | 医生 | 52 | 3 | 14 | 221 | |
| [12] | 人工智能 | 57 | 10 | 10 | 111 | 数坤智能 |
| | 医生 | 59 | 8 | 9 | 112 | |
| [13] | 人工智能 | 92 | 3 | 6 | 68 | 锐影系统 |
| | 医生 | 91 | 4 | 4 | 70 | |
| [14] | 人工智能 | 36 | 4 | 2 | 38 | 不详 |
| | 医生 | 35 | 6 | 3 | 36 | |
| [15] | 人工智能 | 83 | 5 | 4 | 15 | 深睿医疗 |
| | 医生 | 80 | 4 | 7 | 16 | 健康服务 |
| [16] | 人工智能 | 213 | 27 | 38 | 82 | LC CT AI诊断系统 |
| | 医生 | 195 | 5 | 56 | 104 | |
| [17] | 人工智能 | 71 | 19 | 2 | 8 | uAI, ChestCare |
| | 医生 | 76 | 13 | 2 | 9 | |

（图3~5）和综合受试者操作特征曲线（图4~6）。基于CT影像的AI辅助诊断系统诊断肺结节良恶性的合并敏感度、合并特异度、合并阳性似然比、合并阴性似然比、合并诊断比值比和综合受试者操作特征曲线AUC分别为0.93（95% *CI*：[0.89，0.95]）、0.81（95% *CI*：[0.66，0.91]）、4.9（95% *CI*：[2.6，9.4]）、0.09（95% *CI*：[0.06，0.14]）、53（95% *CI*：[24，120]）和0.95（95% *CI*：[0.92，0.96]），见图3、4；基于CT影像的医生诊断肺结节良恶性的合并敏感度、合并特异度、合并阳性似然比、合并阴性似然比、合并诊断比值比和综合受试者操作特征曲线AUC分别为0.90（95% *CI*：[0.84，0.94]）、0.89（95% *CI*：[0.75，0.96]）、8.1（95% *CI*：[3.5，18.6]）、0.12（95% *CI*：[0.08，0.18]）、70（95% *CI*：[32，151]）和0.94（95% *CI*：[0.92，0.96]），见图5、6。

2.3.2　AI与医生诊断效能指标比较

AI辅助CT与医生诊断肺结节在敏感度、特异度、似然比、诊断比值比，及综合受试者操作特征曲线AUC等核心指标上差异均无统计学意义（$P>0.05$），诊断效能指标比较结果见表2。

2.3.3　异质性检验结果

$I^2>50\%$，存在显著的异质性，因每个研究使用的AI算法和CT不同，所以从发表时间、研究地区、CT扫描厚度、医生组阅片人数、CT管电压、

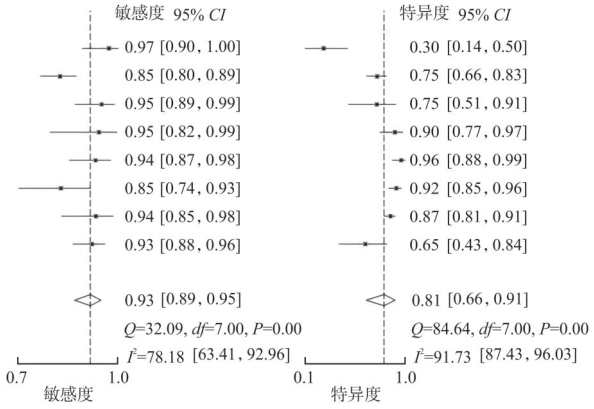


图3　人工智能辅助计算机断层扫描敏感度和特异度

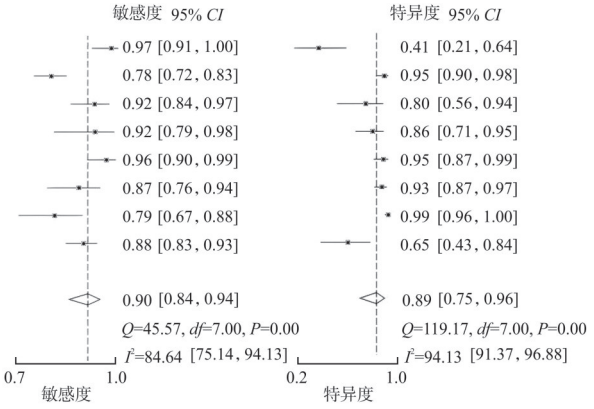


图5　医生辅助计算机断层扫描敏感度和特异度

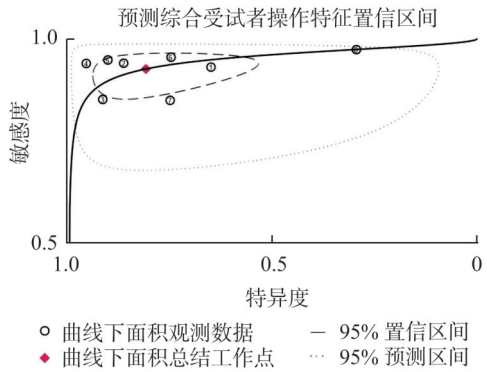


图4　人工智能辅助计算机断层扫描综合受试者操作特征曲线

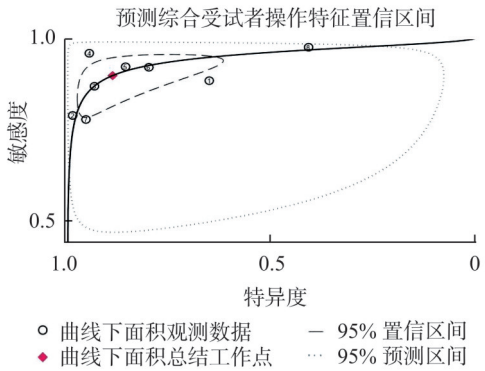


图6　医生辅助计算机断层扫描综合受试者操作特征曲线

**表2　AI与医生诊断效能指标的统计学比较结果**

| 组别 | 敏感度 95% *CI* | 特异度 95% *CI* | 阳性似然比 95% *CI* | 阴性似然比 95% *CI* | 诊断比值比 95% *CI* | 曲线下面积 95% *CI* |
|---|---|---|---|---|---|---|
| 人工智能组 | 0.93[0.89，0.95] | 0.81[0.66，0.91] | 4.9[2.9，9.4] | 0.09[0.06，0.14] | 53[24，120] | 0.95[0.92，0.96] |
| 医生组 | 0.90[0.84，0.94] | 0.89[0.75，0.96] | 8.1[3.5，18.6] | 0.12[0.08，0.18] | 70[32，151] | 0.94[0.92，0.96] |
| 差异 | 0.03[0.02，0.08] | 0.08[0.21，0.05] | 3.2[9.8，3.4] | 0.03[0.10，0.04] | 17[81，47] | 0.01[0.02，0.04] |
| *P*值 | 0.265 | 0.218 | 0.340 | 0.407 | 0.601 | 0.500 |

研究人数方面进行Meta回归分析，未分析出异质性的来源。综合受试者操作特征曲线呈现显著弓形分布，表明诊断阈值变异是异质性的主要来源。

2.3.4 AI系统亚组分析结果

亚组分析显示，AI系统的性能存在显著分化，表现为敏感度、特异度和AUC的差异。各组纳入研究数量、样本量及合并效应量结果见表3。

**表3 人工智能系统亚组分析表**

| 系统名称 | 研究数/篇 | 样本量/例 | 合并敏感度95%*CI* | 合并特异度95%*CI* | 曲线下面积95%*CI* |
|---|---|---|---|---|---|
| 医准智能 | 1 | 194 | 0.88[0.82,0.92] | 0.83[0.78,0.87] | 0.90[0.87,0.93] |
| 数坤人工智能 | 1 | 111 | 0.97[0.94,0.99] | 0.79[0.72,0.85] | 0.96[0.94,0.98] |
| 锐影系统 | 1 | 68 | 0.85[0.79,0.90] | 0.84[0.79,0.88] | 0.89[0.86,0.92] |
| 深睿医疗 | 1 | 38 | 0.93[0.89,0.96] | 0.80[0.74,0.85] | 0.92[0.89,0.95] |
| LC CT系统 | 1 | 15 | 0.85[0.79,0.90] | 0.75[0.68,0.81] | 0.88[0.85,0.91] |
| 联影智能 | 1 | 36 | 0.91[0.87,0.94] | 0.86[0.81,0.90] | 0.95[0.93,0.97] |
| 未明确系统 | 2 | 274 | 0.93[0.89,0.96] | 0.82[0.70,0.90] | 0.87[0.82,0.92] |

## 3 讨论

本研究结果表明，AI辅助CT在肺结节性质判别中的综合诊断效能与医生阅片相当。虽然AI组呈现出略高的敏感度，医生组显示出更佳的特异度，但这些差异均无统计学意义（$P$>0.05），提示当前AI仅可作为辅助诊断工具，尚不能完全替代临床医师的专业判读。

对纳入研究的异质性分析，综合受试者操作特征曲线散点呈现典型的弓形分布模式，提示存在由诊断阈值差异导致的显著异质性；基于$I^2$>50%提示的显著异质性，选择随机效应模型合并分析数据。由于各研究采用的AI算法和CT设备存在差异，本研究通过Meta回归分析潜在异质性来源。结果显示，发表时间、研究地区、医师组样本量和研究总人数差异均无统计学意义，表明异质性可能源于其他未检测因素。

亚组分析发现，AI系统间存在明显的性能分化现象：侧重敏感性的系统（如数坤AI、深睿医疗）敏感度达0.93~0.97，与钟丽茹等[18]报告的数坤AI最高敏感度（0.97）一致，但特异度（0.79~0.80）低于沈小钦等[19]研究中同类系统的均值（0.85），可能导致假阳性风险增加；平衡型系统（如联影智能）在敏感度（0.91）与特异度（0.86）间取得最优协同，其AUC与沈小钦等[19]报道的联影系统效能（0.94）相符，综合诊断价值突出；侧重特异性的系统（如医准、锐影）特异度（0.83~0.84）接近钟丽茹等[18]中联影AI的最高值（0.86），但敏感度（0.85~0.88）相对受限；综合效能较低的系统（如LC CT系统）各项指标均显著低于其他组（敏感度0.85，特异度0.75），临床应用需审慎评估。

造成该现象的技术根源可能包括深度学习网络架构的多样性、训练数据分布的差异性、特征选择方法的特异性，以及验证流程的不一致性等多方面因素[9, 20~21]。

本研究存在一定局限性：纳入的文献数量有限（8篇），总样本量相对较小（1 151）例，可能降低统计检验效能，影响其对广泛人群的代表性；纳入研究使用了不同的AI算法和CT设备，存在显著异质性，且Meta回归分析未能明确识别异质性的具体来源，一定程度上影响了结果的稳定性和可比性；仅纳入中、英文文献，可能遗漏其他语种的相关研究，对系统性有一定影响。

综上所述，Meta分析结果显示，AI辅助CT在肺结节良、恶性鉴别中展现出与放射科医师相当的总体诊断效能。亚组分析揭示不同AI系统存在特征性性能分化，其中平衡型系统展现出最优综合诊断价值。当前证据支持AI作为临床辅助工具的定位，其敏感度优势对降低肺结节漏诊率具有潜在意义。但该研究存在一定的局限性，需大规模规范化临床验证。

**利益冲突**　所有作者声明不存在利益冲突。

**作者贡献**　张　蕾：设计实验，实施研究，采集数据，分析、解释数据，起草文章，对文章的知识性内容作批评性审阅，统计分析；杨嘉琪：设计实验，采集数据；辛小霞：采集数据；符文杰：采集数据，对文章的知识性内容作批评性审阅；张常青：设计实验，实施研究，对文章的知识性内容作批评性审阅，提供研究经费；樊景春：对文章的知识性内容作批评性审阅，并给予技术支持。

## 参 考 文 献


[1] 王培宇，黄祺，王少东，等.《全球癌症统计数据2022》要点解读[J]. 中国胸心血管外科临床杂志，2024，31(7)：933-954.

[2] THAI A A，SOLOMON B J，SEQUIST L V，et al. Lung cancer[J]. The lancet，2021，398(10299)：535-554.

[3] MAZZONE P J，LAM L. Evaluating the patient with a pulmonary nodule：a review[J]. JAMA，2022，327(3)：264-273.

[4] LARICI A R，FARCHIONE A，FRANCHI P，et al. Lung nodules：size still matters[J]. European respiratory review，2017，26(146)：170025.

[5] 陆晓骞，宗凌燕，沈琪. 正电子发射断层扫描/计算机断层扫描(PET/CT)成像在肿瘤筛查中的应用进展[J]. 中国辐射卫生，2023，32(1)：66-69，74.

[6] LIU H，CHEN R H，TONG C，et al. MRI versus CT for the detection of pulmonary nodules：a meta-analysis[J]. Medicine，2021，100(42)：e27270.

[7] VENKADESH K V，SETIO A A A，SCHREUDER A，et al. Deep learning for malignancy risk estimation of pulmonary nodules detected at low-dose screening CT[J]. Radiology，2021，300(2)：438-447.

[8] 董来东，黄果. 基于CT影像的人工智能辅助诊断系统对4771例肺癌诊断价值的系统评价与Meta分析[J]. 中国胸心血管外科临床杂志，2021，28(10)：1183-1191.

[9] 程顺达，程颖，孙士江. 基于机器学习的肿瘤智能辅助诊断方法[J]. 电子科技，2022，35(5)：56-59.

[10] DU W，HE B，LUO X J，et al. Diagnostic value of artificial intelligence based on CT image in benign and malignant pulmonary nodules[J]. Journal of oncology，2022，2022：5818423.

[11] 耿然. 基于深度学习的人工智能影像辅助诊断系统对肺结节的诊断效能评价[J]. 新乡医学院学报，2022，39(11)：1031-1035.

[12] 贺文俊，张周兵，胡永松，等. 基于人工智能的CT定量分析在肺结节患者中的诊断价值及效能ROC曲线研究[J]. 现代仪器与医疗，2023，29(3)：89-92.

[13] 刘亚龙，安小康，郭明杰，等. 基于人工智能技术的肺部CT扫描在肺结节筛查中的应用分析[J]. 医学理论与实践，2023，36(11)：1918-1919.

[14] 王彬，顾文权，蔡孙美，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]. 影像研究与医学应用，2023，7(5)：95-97.

[15] FAN W，LIU H T，ZHANG Y，et al. Diagnostic value of artificial intelligence based on computed tomography (CT) density in benign and malignant pulmonary nodules：a retrospective investigation[J]. Peer J，2024，12：e16577.

[16] PAN W，FANG X T，ZANG Z Y，et al. Diagnostic efficiency of artificial intelligence for pulmonary nodules based on CT scans[J]. American journal of translational research，2023，15(5)：3318-3325.

[17] 袁飞，钟临锋，陈兴，等. 肺结节人工智能评估危险度辅助人工阅片的应用价值[J]. 分子影像学杂志，2024，47(2)：132-137.

[18] 钟丽茹，罗娜，贺露瑶. 多个人工智能辅助诊断系统对肺结节的诊断效能比较[J]. 影像研究与医学应用，2024，8(17)：26-29.

[19] 沈小钦，梁红，朱晓琼. CT图像人工智能辅助诊断系统鉴别肺结节良恶性的价值[J]. 中国辐射卫生，2024，33(5)：578-583.

[20] BERNHARDT M，JONES C，GLOCKER B. Investigating underdiagnosis of AI algorithms in the presence of multiple sources of dataset bias [EB/OL]. (2022-01-19) [2024-07-30]. arXiv：2201.07856. https://arxiv.org/abs/2201.07856.

[21] 关松林. 机器学习在疾病诊断中的应用[D]. 重庆：重庆大学，2019.


（责任编辑：王春燕）