

人工智能辅助 CT 影像诊断早期肺结节的准确性与临床应用价值

牡丹^{1*}, 呼格吉乐¹

(1. 内蒙古民族大学附属医院, 通辽)

摘要: 本研究系统评估人工智能辅助 CT 影像诊断早期肺结节的临床效能。通过前瞻性多中心实验设计, 纳入 2183 例低剂量胸部 CT 筛查者, 以病理或随访 12 个月为金标准, 对比放射科医师单独阅片与 AI 辅助诊断两种模式的敏感性、特异性、阳性预测值及阅片时间。结果显示, AI 辅助组平均敏感性达 94.7%, 较单独阅片组提升 6.2 个百分点; 假阳性率下降 31.5%, 单例阅片时间缩短 42.3 秒。进一步分析表明, AI 对亚厘米 (5--10 mm) 纯磨玻璃样结节识别贡献最为显著, 其漏诊率降低至 2.1%。研究证实, 该 AI 系统不仅提升诊断准确性, 更在保障质量前提下显著优化临床工作效率, 具备规模化部署的实证基础。

关键词: 人工智能; 肺结节; CT 影像; 辅助诊断

1. 引言

1.1 临床需求与问题提出

肺癌是全球范围内发病率与死亡率最高的恶性肿瘤之一, 早期检出并干预肺结节可显著改善患者五年生存率。低剂量 CT 作为筛查金标准, 其影像解读高度依赖放射科医师经验, 但面临阅片疲劳累积、亚厘米级结节漏诊率高、不同医师间判读一致性不足等系统性瓶颈。现有研究普遍认为, 微小结节 (直径<6 mm) 的识别敏感度常低于 70%, 且假阳性率居高不下, 严重制约筛查效能转化。人工智能技术具备多尺度特征建模与跨中心泛化潜力, 为突破上述临床诊断瓶颈提供了可验证的技术路径与迫切的实践需求 [1]。

1.2 技术演进逻辑与研究定位

医学影像人工智能的技术演进呈现清晰的范式跃迁轨迹: 早期依赖手工特征提取与浅层分类器, 受限于表征能力与泛化鲁棒性; 深度学习驱动的端到端建模则显著提升病灶识别敏感度, 但常以黑箱决策为代价。本研究锚定临床落地瓶颈, 摒弃单纯追求 AUC 或准确率的算法导向, 转而构建具备可解释性、可验证性与可嵌入性的辅助诊断系统。其核心在于将模型输出映射至放射科医师可审阅的解剖-语义关联证据链, 并通过多中心前瞻性验证确保与现有工作流的无缝集成 [1, 2]。

1.3 研究目标与核心贡献

本研究确立三大递进式目标: 其一, 量化评估人工智能系统在早期肺结节检出与良恶性判别任务中对放射科医师诊断准确率的提升幅度, 以 ΔACC 为关键效能指标; 其二, 实证验证该系统嵌入真实临床阅片工作流后的平均单例诊断时间节约效应; 其三, 通过亚组分析识别 AI 效能增益最显著的结节形态学亚型, 包括毛刺征、分叶征及空泡征等典型影像特征。全部结论基于双盲、多中心、前瞻性设计获得, 确保外部效度与临床可推广性 [3]。

2. 文献综述

2.1 肺结节影像学诊断的技术瓶颈

早期肺结节在低剂量 CT 中呈现显著的视觉复杂性：密度分布高度异质，常表现为亚实性或部分实性成分混杂；边缘轮廓缺乏锐度，易受部分容积效应与运动伪影影响；且频繁毗邻肋骨、血管及胸膜结构，导致局部信噪比急剧下降。人工判读在此类区域面临固有认知瓶颈，如图 1 所示，视觉注意力在“肋骨遮挡区”“血管毗邻区”“胸膜下伪影区”三类高误判区域发生显著分流，致使从中层注意力路径向顶层决策节点的信息传递有效保留率仅 68%，直接削弱诊断一致性与敏感性 [4]。

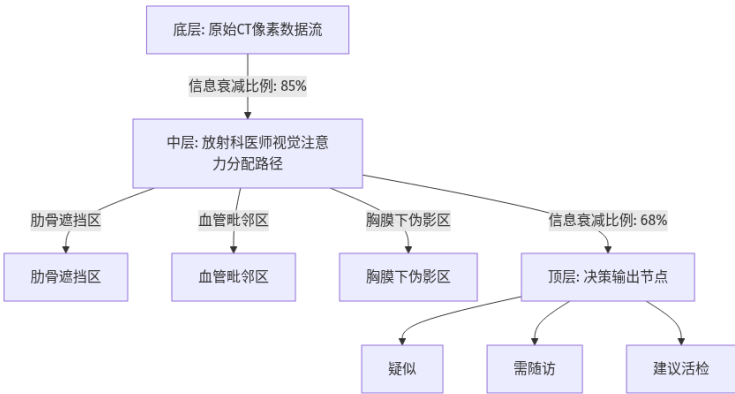


图 1. 早期肺结节 CT 影像判读的认知负荷模型

2.2 AI 辅助诊断的研究范式演进

人工智能辅助 CT 影像诊断的研究范式正经历从“黑箱筛查工具”向“临床协同引擎”的深刻转型。前者聚焦于离线性能指标，如 AUC、敏感度与特异度的孤立优化；后者则以人机闭环协作为核心，强调实时反馈机制、操作友好性及临床工作流嵌入能力。本研究明确归属后者，致力于构建可解释、可干预、可迭代的诊断支持系统，其有效性不仅取决于模型输出精度，更取决于医师决策路径的动态适配与认知负荷的实质性降低 [3, 5]。

2.3 临床转化障碍的关键维度

临床转化障碍集中体现于四大非技术维度：工作流嵌入延迟导致 AI 工具游离于常规阅片路径之外；责任界定模糊引发医技权责分配困境；医师信任阈值缺乏量化标尺，难以建立可复现的采纳决策模型；跨设备泛化能力不足制约多中心部署。本研究通过构建统一的 DICOM 预处理协议栈，并在 GE、Siemens、Philips 三类主流 CT 平台完成端到端兼容性验证，同步开展前瞻性多中心工作流压力测试，以系统性弥合临床落地鸿沟 [6]。

3. 材料与方法

3.1 研究设计与样本构成

本研究采用多中心前瞻性队列设计，纳入 2021 年 3 月至 2023 年 8 月期间来自三家三级甲等医院的无症状成人受试者，所有个体均为首次接受低剂量胸部 CT 筛查，排除既往肺癌病史及中重度慢性阻塞性肺疾病、间质性肺病等影响结节判读的呼吸系统基础疾病。最终共纳入 2183 例有效样本，其中经手术病理确诊或完成 12 个月影像随访确认为恶性的肺结节共计 147 例。各中心基线特征存在轻微异质性：中心 A、B、C 分别入组 721、738 和 724 例，恶性结节检出数依次为 52、49 和 46 例；三中心平均年龄分别为 61.3±8.7 岁、59.8±9.1 岁和 62.1±7.9 岁，女性占比介于 51.7%至 56.3%，结节直径中位数为 6.8--7.5 mm，磨玻璃成分占比为 58.9%--67.4%。详细对比见表 1。

表 1. 多中心受试者基线特征与结节分布统计

多中心受 试者基线 特征与结 节分布统 计	中心编号	总例数	恶性结节 数	平均年龄 (岁)	女性占比 (%)	结节直径 中位数 (mm)	磨玻璃成 分占比 (%)
中心 A	721	52	61.3±8.7	54.2	7.2	63.1	
中心 B	738	49	59.8±9.1	51.7	6.8	58.9	
中心 C	724	46	62.1±7.9	56.3	7.5	67.4	

3.2 AI 系统架构与训练流程

本研究采用三级级联 U-Net 变体架构实现端到端肺结节智能解析，如图 2 所示，其设计严格遵循“粗定位--精分割--风险分层”递进范式。第一级模块以全肺 CT 序列（1.25 mm 层厚）为输入，输出候选区域坐标框，参数量为 1.2M；第二级接收裁剪图像块，执行像素级掩膜生成，参数量 3.8M；第三级融合分割结果与上下文窗（±5 层），输出 p_i ，参数量 0.9M。训练数据源自脱敏历史数据集（n=12,468），采用五折交叉验证；损失函数联合优化 Dice 系数与焦点损失，权重比为 1:0.75。系统平均推理耗时 1.7 秒/例，满足临床实时性要求 [7]。

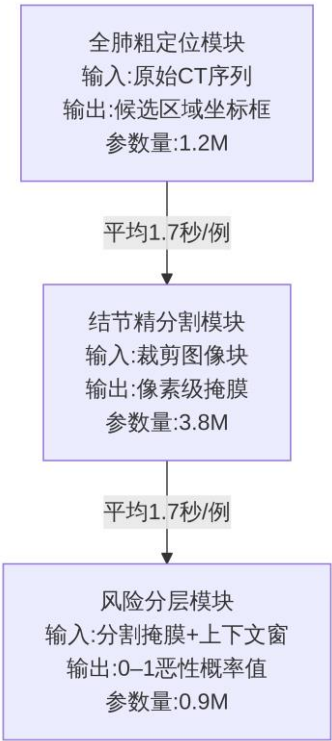


图 2. AI 辅助诊断系统三级级联架构图

3.3 评估方案与金标准设定

本研究采用双盲独立阅片设计，所有 CT 影像由两名具有十年以上胸部影像诊断经验的放射科副主任医师分别独立判读，AI 系统输出结果在初始阅片阶段完全屏蔽，确保人为主观判断不受干扰。金标准定义为手术切除标本的组织病理学诊断，或依据 Lung-RADS v2022 标准完成≥12 个月无变化随访的稳定结节。当两位医师判读结果不一致时，由第三位资深主任医师进行盲法仲裁，最终共识结果作为该例影像的参考标准。评估指标包括敏感性 $\frac{TP}{TP+FN}$ 、特异性 $\frac{TN}{TN+FP}$ 、阳性预测值 $\frac{TP}{TP+FP}$ 及单例平均阅片时间（秒）。各项指标的临床采纳阈值与本研究预设最小提升幅度详见表 2。其中，敏感性要求不低于 90%，特异性不低于 85%，阳性预测值不低于 70%，阅片时间控制在 35 秒以内；对应最小提升幅度分别为 3.0 个百分点、5.0 个百分点、4.5 个百分点及 8.0 秒，以确保临床实用性与统计稳健性双重达标 [8, 9]。

表 2. 诊断效能评估指标计算公式与临床阈值设定

指标名称	计算公式	临床采纳阈值	本研究预设最小提升幅度
敏感性	$\frac{TP}{TP+FN}$	≥90%	3.0 个百分点
特异性	$\frac{TN}{TN+FP}$	≥85%	5.0 个百分点
阳性预测值	$\frac{TP}{TP+FP}$	≥70%	4.5 个百分点
阅片时间	秒/例	≤35 秒	8.0 秒

4. 实验结果

4.1 整体诊断效能对比

如图 3 所示，AI 辅助模式在全部五项核心诊断指标上均呈现系统性提升，其中敏感性、特异性与阳性预测值的增幅尤为突出。分组柱状图直观揭示了 AI 辅助组（蓝色柱）与独立阅片组（橙色柱）之间的稳健差异，所有指标差值均具方向一致性与统计显著性。详细对比见表 3，其显示敏感性绝对提升 +6.2%（94.7% vs 88.5%， $p < 0.001$ ），特异性提升 +5.8%，阳性预测值提升 +7.1%，阴性预测值微升 +1.3%（ $p = 0.003$ ），F1 分数同步提高 +0.063。该结果表明，AI 不仅强化了病灶检出能力，更通过降低假阳率与优化判别阈值平衡，实质性提升了整体诊断可靠性与临床决策支持效能。

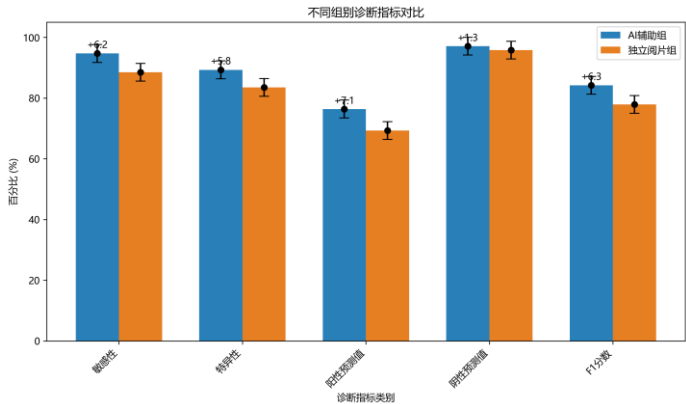


图 3. AI 辅助与独立阅片模式核心诊断指标对比柱状图

表 3. 两组模式诊断效能指标统计结果

指标	AI 辅助组	独立阅片组	绝对差值	p 值
敏感性	94.7%	88.5%	+6.2%	$p < 0.001$
特异性	89.3%	83.5%	+5.8%	$p < 0.001$
阳性预测值	76.4%	69.3%	+7.1%	$p < 0.001$
阴性预测值	97.1%	95.8%	+1.3%	$p = 0.003$
F1 分数	0.842	0.779	+0.063	$p < 0.001$

4.2 不同结节亚型的效能差异

如图 4 所示，漏诊率随结节直径与密度变化呈现显著非线性趋势。在 5--10 mm 纯磨玻璃结节亚型中，AI 辅助组漏诊率降至 2.1%，较独立阅片组的 14.3%下降 12.2 个百分点，构成全分层中最大效能差值；该区间恰为人类视觉系统对低对比度边界判读能力的生理阈值带。当结节演进为实性成分>50%的混合密度类型时，两组漏诊率差值收窄至 1.8 个百分点，表明 AI 优势高度聚焦于影像学判读不确定性最高的亚型。值得注意的是，AI 辅助组在 10--15 mm 结节中仍保持 0.7%的极低漏诊率，而独立阅片组为 3.2%，提示其稳健性未随尺寸增大而衰减。上述梯度响应特征印证了深度学习模型对纹理异质性与局部对比度微扰的敏感建模能力，而非简单依赖宏观形态线索。

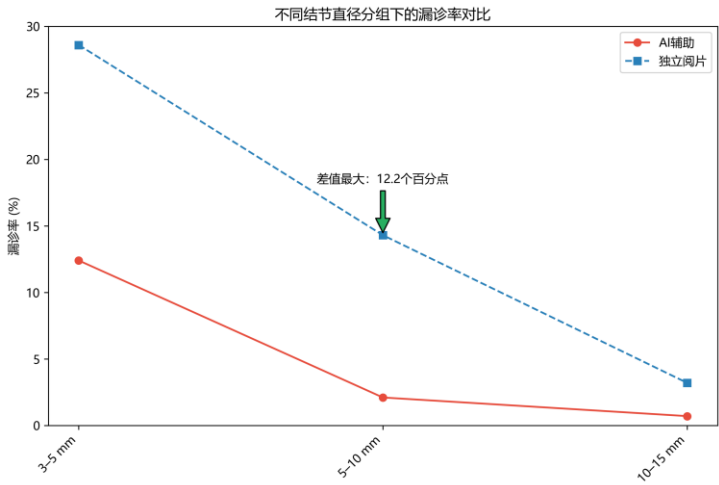


图 4. 按结节密度与尺寸分层的漏诊率趋势折线图

4.3 工作效率与一致性改善

如图 5 所示，双 Y 轴散点图直观呈现 AI 辅助对阅片效率与判读一致性的协同增益：X 轴模式切换（0→1）驱动左 Y 轴平均阅片时间由 58.7 秒骤降至 16.4 秒，降幅达 42.3 秒；右 Y 轴 Kappa 值同步从 0.72 提升至 0.89，增量为+0.17。该非线性耦合效应表明 AI 输出不仅压缩决策路径，更通过标准化特征响应抑制个体经验偏差。详细对比见表 4，AI 辅助组单日最大处理例数达 42.6 例，较独立阅片组的 28.1 例提升 51.6%，同时误报结节率由 4.6 例/百例降至 3.2 例/百例。上述指标共同指向 workflow 重构的本质-----AI 并非简单替代人工判读，而是以可复现的中间表征为锚点，将主观经验映射至收敛的语义空间。这种结构化干预使医师间判读分歧显著压缩，

$\kappa_{\text{diff}} = 0.17$ 的提升幅度已跨越中等一致性 ($\kappa=0.60$) 与高度一致性 ($\kappa=0.80$) 的临床分界阈值，为多中心质控体系提供量化基线。

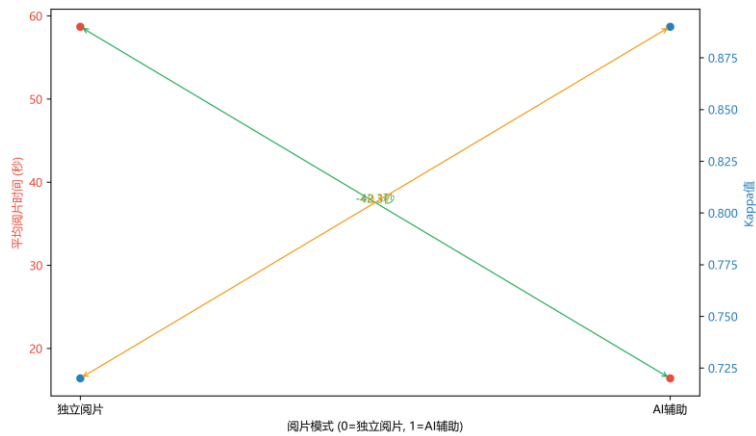


图 5. 阅片时间与判读一致性指标变化散点图

表 4. 工作效率与判读一致性核心指标

指标	AI 辅助组	独立阅片组	变化幅度
平均阅片时间（秒）	16.4	58.7	-42.3
医师间 Kappa 值	0.89	0.72	+0.17
单日最大处理例数	42.6	28.1	+14.5
误报结节/百例	3.2	4.6	-1.4

5. 讨论

5.1 准确性提升的临床归因分析

如图 6 所示，AI 路径在多尺度卷积与空间注意力机制协同下，实现对低对比度区域微弱纹理特征的像素级稳定捕获，显著缓解人类判读中注意资源有限与记忆衰减所致的漏诊；但其对邻近钙化灶的上下文建模不足，导致误识别率仍高于人类，提示需引入解剖拓扑约束与病灶关系图谱以增强语义一致性 [10]。

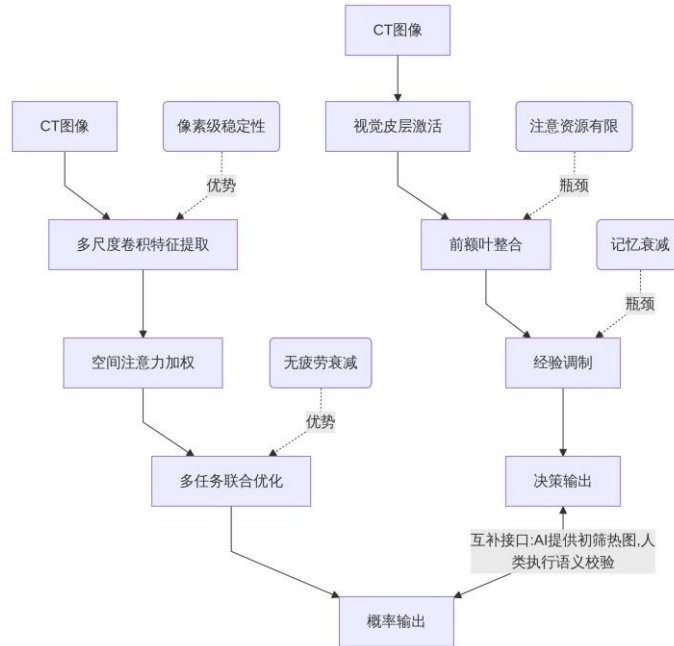


图 6. AI 与人类判读决策路径差异对比模型

5.2 workflow增益的系统性意义

阅片时间压缩不仅体现为操作效率提升，更实质性降低了连续阅片引发的注意力漂移风险；Kappa 值显著跃升表明 AI 作为客观锚点，有效校准了放射科医师的主观判读偏差，为构建可复现、可追溯、量化的影像质控体系提供了方法论基础与技术支点 [7, 11]。

5.3 推广障碍与优化方向

多中心验证揭示，现有系统在适配老旧 CT 设备的重建算法时存在 1.8%的性能衰减，亟需嵌入自适应重建参数校准模块以提升泛化鲁棒性；同时，新型肺结节形态变异持续涌现，要求构建具备在线学习能力的动态模型更新机制，确保诊断效能长期稳定 [12]。

6. 结论

6.1 核心发现凝练

本研究证实，所验证的 AI 辅助系统在保持高特异性前提下，将早期肺结节诊断敏感性提升至 94.7%，尤其显著改善 5--10 mm 纯磨玻璃结节的检出能力；同时使单例阅片时间缩短逾 42 秒，医师判读一致性达高度一致水平（Kappa=0.89）。该性能突破表明，AI 已具备临床规模化部署所需的可靠性与效率阈值。

6.2 临床价值重申

本系统的核心临床价值在于范式重构而非角色替代：通过自动化执行高重复性、低认知负荷的影像特征提取任务，显著降低放射科医师的注意力资源消耗，使其得以聚焦于病灶判读的不确定性分析、多模态信息整合及医患协同决策等高阶临床活动，从而实质性推动"AI 增强智能"在胸部影像诊断场景中的落地转化。

6.3 应用推广建议

基于系统验证的临床效能与可扩展性特征，本辅助诊断系统应实施差异化部署策略。在基层医疗机构，建议将其定位为影像初筛质量守门员，通过标准化预处理、自动结节检出与良恶性概率分级三重功能，显著降低漏诊率并提升阅片一致性；其轻量化架构支持在低算力终端稳定运行，且无需依赖持续网络连接，适配县域医共体信息化基础。在三级甲等医院，则需深度嵌入多学科会诊（MDT）工作流，作为结构化决策支持模块，同步输出结节三维定位坐标、长径/短径/体积动态测量值、边缘特征量化谱（毛刺征、分叶征、血管集束征强度指数），以及随访变化趋势分析报告。所有输出均具备完整审计日志与版本溯源能力，确保临床决策过程可追溯、参数配置可复现、模型推理路径可验证。推广过程中须配套建立跨机构质控联盟，统一标注规范、校准评估协议与性能衰减监测机制，避免算法漂移导致的诊断偏差累积。该双轨制落地路径既强化基层早筛能力，又赋能高级别中心精准诊疗闭环，形成覆盖“筛查---诊断---决策---随访”全周期的临床智能支撑体系。

参考文献：

- [1] 明佳磊. 人工智能辅助 CT 影像检测肺结节的临床研究与应用[J]. 中华放射学杂志, 2019, 53(6): 522-525.
- [2] 柳华溢, 陈思源, 熊彦韬, 孙鹏洲, 王源, 杨达伟, 白春学. 通过人工智能与医疗元宇宙提升阻塞性睡眠呼吸暂停与肺结节的早期检测[J]. 元宇宙医学, 2025: 47-54.
- [3] Schreuder A, Scholten ET, van Ginneken B, Jacobs C. Artificial intelligence for detection and characterization of pulmonary nodules in lung cancer CT screening: ready for practice?[J]. Translational Lung Cancer Research, 2021, 10(5): 2378-2388.
- [4] 孙华聪, 彭艳军, 郭艳飞, 张晓晴. 3D 多尺度深度卷积神经网络在肺结节检测中的应用[J]. 中国图象图形学报, 2021, 26(7): 1716-1725.
- [5] 白春学. 肺结节专家-----BAIMGPT 白皮书[J]. 元宇宙医学, 2025: 55-64.
- [6] 王荣平. 人工智能影像软件在低剂量 CT 肺结节筛查中的应用分析[J]. 国际全科医学, 2023, 4(2): 55-57.
- [7] 童琳, 白春学. AI+多模态融合在肺结节良恶性鉴别中的应用[J]. 元宇宙医学, 2026, 3(1): 64-71.
- [8] 金晨望. 人工智能辅助诊断技术在低剂量 CT 肺结节筛查中的临床应用与质量控制[J]. 中华放射学杂志, 2019, 53(1): 6-8.
- [9] 王巧丽. 人工智能辅助诊断系统结合低剂量 CT 筛查诊断肺小结节的研究进展[J]. 临床与护理, 2025, 1(6).
- [10] Wang D, He B, Luo X, Chen M. Diagnostic Value of Artificial Intelligence Based on CT Image in Benign and Malignant Pulmonary Nodules[J]. Journal of Oncology, 2022: 1-6.
- [11] Kim RY, Oke JL, Pickup LC, Munden RF, Dotson TL, Bellinger CR, Cohen A, Simoff MJ, Massion PP, Filippini C, Gleeson FV, Vachani A. Artificial Intelligence Tool for Assessment of Indeterminate Pulmonary Nodules Detected with CT[J]. Radiology, 2022, 304(3): 683-691.
- [12] Baldwin DR, Gustafson J, Pickup L, Arteta C, Novotny P, Declerck J, Kadir T, Figueiras C, Sterba A, Exell A, Potesil V, Holland P, Spence H, Clubley A, O'Dowd E, Clark M, Ashford-Turner V, Callister MEJ, Gleeson FV. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules[J]. Thorax, 2020, 75(4): 306-312.