

# AI 辅助诊断系统在法医临床影像学证据分析中的前瞻性 随机对照试验

翟伟博<sup>1</sup> 郑宣颖<sup>2</sup>

广东龙城司法鉴定所，广东省深圳市，518172；

**摘要：**法医临床影像学证据分析精准度对司法裁判事实认定具有直接影响，本研究以某三甲医院 2023—2024 年 300 例法医鉴定影像学病例为对象，开展前瞻性随机对照试验，将其分为 AI 辅助联合医师分析的实验组与单纯医师分析的对照组，对诊断准确率、一致性及效率进行对比。数据显示，实验组在复杂病例诊断准确率、观察者间一致性及分析耗时方面较对照组更具优势，在伤病关系认定等难点问题上表现尤为突出。研究显示，AI 辅助系统借助标准化流程及多维度特征提取，可提升分析客观性与效率，为司法影像学证据科学运用提供技术支持，助力司法鉴定向精准化方向发展。

**关键词：**法医临床影像学；AI 辅助诊断系统；诊断准确率；诊断一致性

**DOI：**10.69979/3029-2808.25.07.022

## 引言

医学影像学专业本科教学中的实习阶段是学生角色转换的一个过渡阶段，是理论知识与实践技能有机结合的重要时期，也是培养合格医学影像人才的重要环节。

因此如何提高医学影像学专业实践教学质量，医学教育界也在不断探索。近年来人工智能技术发展迅速，在医学影像学领域的应用研究逐渐深入，其中 AI 辅助诊断系统已逐渐嵌入临床工作流程之中，并取得良好的效果。传统医学分析常基于医师主观经验判断，复杂病例中易出现诊断分歧或漏诊情况，深度学习驱动的 AI 辅助诊断系统虽展现一定应用潜力，但其在法医临床领域的实际应用价值尚需验证。本研究通过前瞻性试验对该系统的有效性 with 可靠性展开评估，旨在为司法实践提供科学依据。

## 1 研究对象与方法

### 1.1 研究设计

本研究以某中心前瞻性随机对照试验为设计方案，依据 CONSORT 声明标准规划试验流程并撰写报告，试验方案提交医院伦理委员会，围绕研究目的医学价值、受试者风险收益权衡及数据安全防护等要点，开展科学性与伦理审查，顺利获取伦理批件。经伦理培训的研究人员，向所有入组受试者详尽阐释试验目标、操作流程、潜在风险与权益，确保受试者充分理解后，签署符合《涉及人的生物医学研究伦理审查办法》规范的书面知情同

意书。

分组借助计算机生成随机数字表完成，按四肢骨折、颅脑损伤、脊柱损伤、胸腹部脏器损伤等损伤类型实施分层随机化，保障两组病例构成比例均衡。选取此类分层因素，源于法医临床鉴定常见损伤分类体系，以此规避混杂变量干扰试验结果。试验全程依托定制化数据管理系统监控，涵盖病历录入、随机分配、干预开展、结果采集等流程，各环节配备电子签名认证与操作日志留存，实现数据可追溯。主要观察指标参照国际通用诊断试验评价准则界定，诊断准确率运用四格表法测算，一致性分析结合 Cohen's Kappa 系数与 Bland-Altman 分析，确保评价方法科学且可重复。

### 1.2 样本选择

病例筛选构建双盲独立审核模式，安排 2 位副高级职称以上的法医影像学专家各自开展候选病例初筛工作。遇到纳入标准存疑的病例，直接提交专家委员会裁定。

“具有法医鉴定需求”这一纳入标准，严格依据《法医临床检验规范》判定，具体涵盖人身伤害案件损伤程度评定、交通事故伤残等级鉴定、劳动能力鉴定等法定鉴定范畴。影像学资料完整性审查，涉及检查序列完备性、图像层厚以及窗宽窗位设定情况。

排除标准内“严重基础疾病”的医学界定参照 ICD-11 分类体系，具体包含恶性肿瘤骨转移、先天性脊柱裂、慢性阻塞性肺疾病引发的肺大疱等可能对损伤判断

产生干扰的基础性病变<sup>[1]</sup>。运用 PASS15.0 软件开展样本量测算,依据预试验数据将检验效能(1-β)设定为0.90,采用双侧检验且α=0.05,预计实验组与对照组在诊断准确率上存在10%的差异,经计算得出最小样本量为280例,最终纳入300例以缓冲脱落风险。样本人口学特征统计显示,男性病例占比达62%,与法医临床鉴定中男性损伤案件高发的流行病学特征相符;年龄分布以18-65岁劳动年龄人口为主体,这与该年龄段人群社会活动频繁、损伤风险较高的实际状况一致。

## 2 干预措施与评价指标

### 2.1 干预措施

AI 辅助诊断系统以多任务深度学习模型搭建技术架构,集成损伤定位、类型分类、时间推断三大核心模块<sup>[2]</sup>。损伤定位模块运用 Faster R-CNN 算法,借助锚框机制完成对骨折线、出血灶、脏器挫裂伤等感兴趣区域的像素级定位,定位精度可达亚毫米级别;损伤类型分类模块采用 ResNet-50 卷积神经网络,输入层涵盖 X 线、CT、MRI 多模态图像特征,输出层对应《人体损伤程度鉴定标准》中界定的 23 种损伤类型,分类模型在验证集上的平均精确率均值 mAP 为 91.2%;损伤时间推断模块构建时序卷积网络(TCN),通过分析血肿密度演变曲线、骨痂形成灰度值变化等时间依赖性特征,实现损伤时间的量化估算,误差范围控制在±72小时(针对伤后14天内的新鲜损伤)。

系统交互界面基于医学影像专用工作站设计,支持 DICOM3.0 标准图像格式导入,集成三维容积重建(VR)、多平面重组(MPR)、曲面重建(CPR)等可视化功能。损伤特征热力图以梯度颜色映射方式叠加至原始图像,概率诊断建议按置信度排序呈现于界面右侧,医师可利用触控笔对感兴趣区域进行标注及参数测量。对照组医师配置相同品牌型号的影像工作站,但关闭 AI 辅助功能模块,仅提供基础图像浏览与测量工具,以保障两组硬件环境的一致性。

参与研究的医师均具备《司法鉴定人执业证》,专业方向为法医临床学,平均从业时长 8.6 年(5-15 年)。标准化培训内容涵盖:①法医影像学诊断标准的深入学习,重点阐释《人体损伤致残程度分级》中影像学评定条款;②AI 系统操作规范(针对实验组),包括界面功能分区、结果解读原则及人机决策冲突处理流程;③一致性培训,借助 50 例历史标注病例开展预试验,要求

医师间诊断一致性 Kappa 值≥0.80 方可参与正式试验,最终所有医师均通过培训考核,平均 Kappa 值达 0.85。

### 2.2 评价指标

金标准确立实施三轮盲法会诊机制,首轮由 3 名专家独立阅片并出具诊断意见;次轮针对意见分歧病例,结合临床表现与影像学动态资料同步研讨以形成初步共识;第三轮将争议病例提交由 5 名外院专家组成的仲裁委员会进行最终裁定,以此保障金标准的权威性与可靠性<sup>[3]</sup>。诊断准确率核算从损伤维度细化考量:①损伤存在性判断,以金标准中“明确损伤”或“无损伤”作为二元变量;②损伤部位定位,要求解剖学定位精准至亚区域,如颅脑损伤具体到额叶、颞叶等脑区,脊柱损伤精准至椎体节段;③损伤类型鉴别,区分新鲜骨折与陈旧性骨折、挫裂伤与脑出血等易混淆类型;④伤病关系认定,依据《法医临床伤病关系判定指南》,分为完全由损伤引起、伤病共同作用、与损伤无关三类。

诊断一致性评估运用双向 Kappa 检验,首先测算各组与金标准的一致性,以反映诊断结果与权威意见的契合程度;其次随机抽取配对样本计算组内两名医师的观察者间一致性,用于评估诊断过程的可重复性。分析效率通过秒表计时法测量,将“影像学资料获取”界定为打开 DICOM 文件的时间节点,“诊断结论形成”界定为完成电子版鉴定意见书撰写并保存的时间节点,该时间段涵盖图像浏览、病灶分析、文献查阅等全部操作环节。医师主观满意度问卷包含 10 个条目,划分为功能维度(信息辅助效果、多模态融合能力、时间推断准确性)、操作维度(界面友好度、交互便捷性、数据加载速度)、心理维度(诊断信心提升、决策压力缓解、职业成就感),采用 Likert5 级评分结合开放式问题,后续对开放式问答开展内容分析,提取主题词进行质性研究。

## 3 结果与讨论

### 3.1 结果

亚组分析将复杂病例定义为涉及多部位损伤、跨系统损伤或伤病关系存疑的情形,该类病例共 120 例。数据显示,实验组诊断准确率为 89.2%,显著高于对照组的 71.7% ( $\chi^2=18.56, P<0.001$ );而在单纯四肢骨折等简单病例中,两组准确率分别为 95.3%和 93.8%,差异无统计学意义 ( $P=0.27$ ),表明 AI 系统在复杂临床场景中优势更为突出<sup>[4]</sup>。影像学模态分层结果显示,MR

I 检查的诊断准确率提升最为显著（实验组 91.5% vs 对照组 78.9%， $P=0.003$ ），这可能与 MRI 序列复杂、人工判读易受主观因素影响相关；CT 检查中，实验组在颅内微小出血灶检出率（82.4% vs 65.3%）、脊柱附件骨折识别率（78.9% vs 61.2%）上显著优于对照组；X 线检查因图像信息相对单一，两组差异较小（94.1% vs 92.3%， $P=0.19$ ）。

诊断一致性分析结果显示，实验组与金标准在伤病关系认定维度的 Kappa 值为 0.81，显著高于对照组的 0.63（ $Z=3.27$ ， $P=0.001$ ），表明 AI 系统在处理因果关系判定这类主观依赖性较强的问题时，可提供更趋近专家共识的标准化结论。观察者间一致性方面，实验组医师配对的平均 Kappa 值为 0.89（95%CI:0.85-0.92），对照组为 0.78（95%CI:0.72-0.83），经 Fleiss' Kappa 检验显示组间差异具有统计学意义（ $P=0.008$ ），说明 AI 系统的介入有效减少了个体认知差异引发的诊断分歧。

分析效率亚组数据表明，复杂病例平均分析时间实验组较对照组缩短 42.7%（18.6 分钟 vs 32.5 分钟， $P<0.001$ ），简单病例缩短 28.4%（9.2 分钟 vs 12.8 分钟， $P=0.012$ ），显示 AI 系统自动化处理功能对不同难度病例均有效率提升，且复杂场景中效益更突出。医师主观满意度调查质性分析显示，实验组医师高频提及“三维重建功能提升空间定位准确性”“时间推断模型减少伤病关系争议”“热力图辅助发现隐匿性病灶”等优势，同时提出“增加致伤机制模拟功能”“优化多模态图像切换流畅度”等改进建议；对照组医师反馈集中于“复杂病例需反复查阅标准指南”“陈旧性损伤鉴别依赖个人经验”等痛点，与量化数据形成互补验证。

### 3.2 讨论

传统人工分析依托单模态图像，AI 则可同步处理 CT、MRI、X 线等多模态数据，经跨模态特征融合获取更全面的损伤表征。以鉴别椎体骨折新旧程度为例，AI 既分析形态学改变，又结合 MRI 骨髓水肿信号强度与时间序列模型，为伤病关系认定提供量化依据，AI 还能将抽象诊断标准转化为可计算的特征权重模型，在四肢长骨骨折判定中自动测量参数并生成量化评分，降低医师因标准理解差异引发的判断分歧，实现司法鉴定的同质化<sup>[5]</sup>。不同模态应用里，AI 展现差异化优势。针对 MRI

多序列成像的复杂性，AI 借助注意力机制聚焦关键特征区域，以“智能加权”策略降低解读难度，提升诊断效能；而在 X 线简单病例判读中，因人工效率已处较高水平，AI 优势未显著体现，凸显人机协作的“互补性原则”。未来可引入动态阈值调节机制，根据病例难度智能切换辅助策略。现阶段 AI 缺乏力学模型整合能力，难以反推致伤机制，后续可结合有限元分析构建闭环，考虑到医师对 AI 建议采纳率较高，需建立“医师主导—系统辅助”双签核制度及系统误差实时监测机制，保障技术应用的安全性及可靠性。

### 4 结语

本研究表明，AI 辅助诊断系统能够提升法医临床影像学证据分析的诊断准确率、一致性及效率，增强医师诊断信心，该技术借助自动化特征提取与标准化流程，弥补了人工分析的主观性不足，为司法影像学证据的科学运用提供技术支撑。研究虽存在样本单一及技术适应性问题，但其应用潜力清晰，伴随算法优化，AI 系统有望成为该领域的关键工具，推动司法鉴定向精准化、智能化方向发展。实际应用时需构建质量控制体系，明晰人机协作的责任边界，确保技术符合法医伦理与法律规范，为司法公正提供科学保障。

### 参考文献

- [1] 蔡丽萍, 张姣. 超声 AI 影像诊断系统辅助诊断甲状腺结节性质的价值[J]. 影像研究与医学应用, 2025, 9(05): 35-38+42.
- [2] 姜繁. AI 辅助诊断系统精确定位对肺结节的类型和分级的判断价值[J]. 现代医用影像学, 2024, 33(05): 965-967+969.
- [3] 周应媛, 许茂盛. 基于 AI 视觉技术的医疗影像辅助诊断系统在放射科质量管理中的应用效果[J]. 中医药管理杂志, 2021, 29(12): 136-137.
- [4] 赵太良, 王冰冰, 梁威, 等. 基于 18F-FDGPET/CT 与结构 MRI 的人工智能辅助诊断系统在阿尔茨海默病中诊断准确性比较的 Meta 分析[J]. 中国循证医学杂志, 2024, 24(12): 1411-1418.
- [5] 卞佳, 王红霞, 李泉, 等. 人工智能辅助诊断系统在医学影像诊断学实践教学中的应用研究[J]. 中国高等医学教育, 2024, (11): 103-104.