

基于真实世界数据的儿童手部 X 线影像骨龄辅助评估系统评价研究*

吴崇数¹ 竺智荣¹ 王立新¹ 缪崇² 饶艳莺¹

(¹福建省儿童医院 福州 350014 ²福建省妇幼保健院 福州 350001)

[摘要] 目的/意义 基于医院真实世界数据,对儿童手部 X 线影像骨龄人工智能辅助评估系统进行评价。方法/过程 收集 2023 年 1—7 月来院接受手部 X 线影像骨龄评估的 1 101 例患者数据,采用平均绝对误差、均方根误差、组内相关性分析等统计方法,从有效性和适宜性两方面对人工智能系统的临床应用价值进行评价。结果/结论 人工智能系统具有较高的有效性和适宜性,对影像科医生日常临床骨龄评估具有一定辅助作用。

[关键词] X 线影像;骨龄评估;人工智能辅助诊断;真实世界数据

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2025.04.013

Study on the Evaluation of an Assisted Assessment System for Bone Age in Children's Hand X-ray Images Based on Real-world Data

WU Chongshu¹, ZHU Zhirong¹, WANG Lixin¹, MIAO Chong², RAO Yanying¹

¹Fujian Children's Hospital, Fuzhou 350014, China; ²Fujian Maternal and Child Health Hospital, Fuzhou 350001, China

[Abstract] **Purpose/Significance** To evaluate an artificial intelligence (AI) assisted assessment system for bone age in children's hand X-ray images based on real-world data from the hospital. **Method/Process** Data are collected from 1 101 patients who came to the hospital from January to July 2023 for bone age assessment by hand X-ray images. Statistical methods such as mean absolute error, root mean square error, and intraclass correlation analysis are used to evaluate the clinical application value of the AI system from the aspects of effectiveness and appropriateness. **Result/Conclusion** The AI system has high validity and appropriateness, which is helpful for daily clinical bone age assessment by radiologists.

[Keywords] X-ray image; bone age assessment; artificial intelligence (AI) assisted diagnosis; real-world data

[修回日期] 2025-01-30

[作者简介] 吴崇数,助理工程师,发表论文 4 篇;通信作者:饶艳莺,博士。

[基金项目] 工业和信息化部办公厅、国家药品监督管理局综合和规划财务司人工智能医疗器械创新任务揭榜项目(项目编号:2024-AIZSSJ-08004)。

1 引言

随着人工智能(artificial intelligence, AI)辅助诊断技术的发展,越来越多的 AI 系统被应用于临床领域,辅助医生开展疾病预测、诊断和分析等工作,在提升诊断效率的同时,减少误诊和漏诊情况的发生^[1]。AI 系统在进入临床市场前必须进行严格的临床试验验证其有效性,并取得国家药品监督管

理局二类或三类医疗器械注册证。然而，临床试验一般在理想条件下进行，受时间、样本量等限制，难以全面掌握 AI 系统在不同地域、不同时间条件下的有效性^[2]。

真实世界数据源自日常诊疗环境，能够反映患者真实情况，可使研究结果具有更高的外部真实性和实践价值^[3]。近年来，AI 相关真实世界研究日益增多。张涛等^[4]收集 2019 年 2 月—2020 年 1 月 222 例肺部结节且接受手术治疗患者的真实世界数据，证实 AI 肺部结节影像分析系统的诊断可信度较低。邓杨等^[5]利用 100 例临床乳腺浸润性导管癌诊断数据，验证全自动及半自动 AI 辅助诊断系统均可作为有效的临床辅助诊断工具。沈惠文等^[6]构建缺血性脑卒中患者溶栓效果预测模型，并使用 3 925 例患者的临床数据对模型进行外部验证。宋鹏杰等^[7]收集在院接受治疗的 72 例疑似乳腺癌患者数据，验证超声 AI 辅助诊断系统具有较高的灵敏度、特异度和准确率。

骨龄评估是衡量儿童生长发育水平的关键指标，在儿科内分泌学、生长发育研究以及法医学等多个领域扮演着重要角色^[8]。儿童手部 X 线影像骨龄评估结果，对临床医生制定个性化诊疗方案具有重要指导作用^[9]。AI 辅助评估系统在提高该领域评估效率和准确性方面展现出巨大潜力。然而在实际临床应用中，即使对于已获得二类或三类医疗器械注册证的 AI 骨龄评估系统，临床医生仍对其在不同地域、不同时间条件下的泛化能力存在顾虑，整体临床参与度较低^[10-12]。

本文基于福建省儿童医院儿童手部 X 线影像数据，建立真实世界研究数据集，设立产品的有效性和适宜性评价指标，对一款已获得三类医疗器械注册证的人工智能辅助儿童手部 X 线影像骨龄评估系统进行真实世界应用评价，判断其临床应用价值。

2 方法

基于真实世界数据，对 AI 骨龄评估系统的有效性和适宜性进行评价研究，见图 1。评价过程主要包括数据采集、数据治理、统计分析 3 部分。研

究为配对设计，通过比较试验组 (AI) 和参考标准 (专家组) 骨龄评估结果的误差及一致性评价 AI 系统结果的有效性，通过量表打分评价 AI 系统的适宜性。

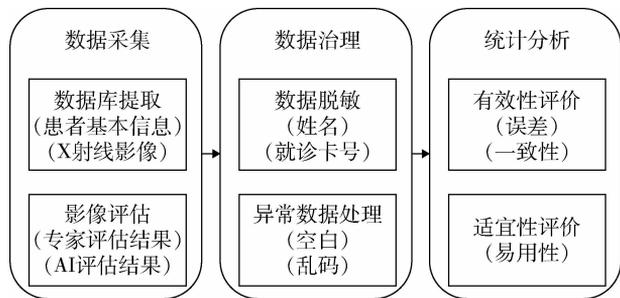


图 1 基于真实世界数据的 AI 骨龄评估系统评价流程

2.1 数据采集

本文所用数据为回顾性数据，所涉及的试验数据已获得福建省儿童医院伦理委员会许可，且免于签署患者知情同意书 (审批编号: 2023ETKLR09006)。

2.1.1 字段说明 用于骨龄 AI 评价的真实世界数据可分为 5 类，见表 1。患者基本信息中的登记号字段为全局唯一标识，可定位单一患者的单次手部 X 线检查。系统适宜性评价为影像科医生使用 AI 系统进行骨龄评估时的满意度。

表 1 真实世界数据字段概况

类别	具体字段
患者基本信息	登记号、姓名、年龄、性别
专家评估结果	专家 1 评估结果 (R 系列)、专家 1 评估结果 (C 系列)、专家 1ID、专家 2 评估结果 (R 系列)、专家 2 评估结果 (C 系列)、专家 2ID、复核专家评估结果 (R 系列)、复核专家评估结果 (C 系列)、复核专家 ID
AI 评估结果	AI 评估结果 (R 系列)、AI 评估结果 (C 系列)、AI 系统厂家
系统适宜性评价结果	系统功能评价、系统易用性评价
X 线影像	影像号、影像采集时间

2.1.2 采集范围 2023 年 1—7 月福建省儿童医院接受手腕部 X 线影像骨龄评估患者 3 000 余例。

依据纳入和排除标准，以及研究所需数据量，从中均匀随机选取 1 101 例。入选标准为：于福建省儿童医院进行过手腕部 X 线影像骨龄评估的患者；年龄 2~18 周岁；男女不限。排除标准为：影像显示手腕部有畸形、外伤史及骨折病史；影像显示手部有金属植入物。

2.1.3 数据采集 (1) X 线检查影像数据采集。X 线影像设备两台，信息如下。设备名称：Ysio Max。序列号：28033、28036。系统版本：VF10F。生产厂家：Siemens Healthineers。采集技术参数：Kvp 为 45-55、mAs 为 2.5-3.5。影像采集流程依据《医学影像检查技术学》^[13]教材中手部 X 线检查规范（手后前位（正位）），即被检者侧身坐摄影台一端，被检侧腕关节及指伸展，手掌向下，五指稍分开，平放于成像板上，第 3 掌骨远端置于成像板中心。射线中心线对准第 3 掌骨远端，垂直射入成像板。此外，在摄片时进行初步图像质量控制，保持手部静止 2~3 秒，以获取清晰图像；若图像模糊，则重拍一张。本文为回顾性研究，从影像数据库获取患者 X 线影像。(2) 结构化数据获取。患者基本信息已于患者就诊开具 X 线检查申请单时记录，可通过数据库获取结构化数据。(3) 专家评估结果获取。邀请两位骨龄评估专家独立阅片，遵循《中国青少年儿童手腕骨成熟度及评价方法》^[14]中 RUS-CHN (R 系列) 和 TW3-C Carpal (C 系列) 评价方法对手腕部正位 X 线影像进行阅片，并确定骨龄结果。若两位专家判断的骨龄值之间最大绝对误差大于 6 个月，则由第 3 位专家独立复核，并将其结果作为标准参考骨龄评估结果。骨龄评估专家职称均为主治医师及以上，是具有 5 年以上骨龄 X 线影像相关评估经验的影像科医生。骨龄评估复核专家职称为主任医师，是具有 8 年以上骨龄 X 线影像相关诊断经验的影像科医生。(4) AI 评估结果及适宜性评价结果获取。影像科医生日常使用 AI 辅助评估的流程，见图 2。影像科医生（不包括参与人工评估的医师）模拟真实的软件辅助阅片场景，对入组影像阅片并出具影像报告。依据评价指标对软件辅助阅片过程进行评价。

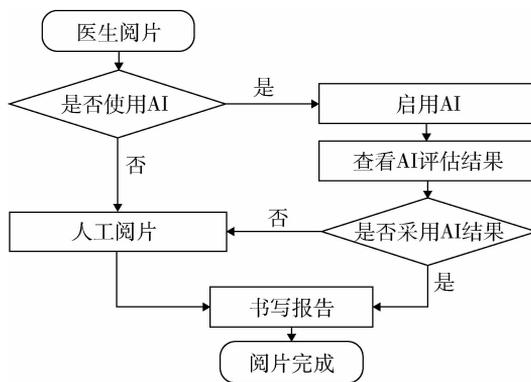


图 2 骨龄 AI 辅助评估系统的临床应用流程

2.2 数据治理

2.2.1 信息安全保护 患者隐私信息主要在患者基本信息和 X 线 Dicom 格式影像的头文件中。患者基本信息来源于医院信息系统 (hospital information system, HIS)。在访问 HIS 数据库提取敏感字段信息时，在 SQL 代码中加入脱敏语句，对姓名字段进行字符替代，只保留姓，名用“*”字符替换，如“张三”替换为“张*”，“张三三”替换为“张**”，依此类推。对于 Dicom 格式影像头文件中的信息，先利用患者主索引从影像存储与传输系统 (picture archiving and communication system, PACS) 数据库中将入组患者的 Dicom 格式影像取出，再用 Python 对头文件中的敏感信息进行批量匿名，将姓名、卡号改为“Anonymous”。

2.2.2 异常数据处理 在确保数据完整性的基础上对数据进行初步清理，去除重复数据。患者基本信息所包含的数据字段为临床诊疗常用字段，出现异常时能够及时纠正，因此回顾性数据中极少出现异常情况。本文实验中的异常数据主要包括两类：专家评估结果、AI 评估结果。处理原则为：当关键字段出现异常且无法通过其他关键字段进行核验修正时，删除整条数据。最终获得 1 060 例数据。

2.3 评价指标

2.3.1 有效性评价指标 比较 AI 评估结果与专

家评估结果间儿童骨龄的平均绝对误差 (mean absolute error, MAE) 和均方根误差 (root mean squared error, RSME), 并利用组内相关系数 (intraclass correlation coefficient, ICC) 评价 AI 评估结果与专家评估结果的一致性。其中, \hat{y}_i 为 AI 评估结果, y_i 为专家评估结果, n 为样本总数。MS_{between} 为组间均方误差, MS_{within} 为组内均方误差, k 为评估专家数量。

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

$$ICC = \frac{MS_{between} - MS_{within}}{MS_{between} + (k - 1) \cdot MS_{within}} \quad (3)$$

2.3.2 适宜性评价指标 适宜性评价包括系统功能评价和系统易用性评价, 采用 1—3 分制量表, 见表 2—表 3。将系统操作人员返回的量表结果取均值作为适宜性评价结果。

表 2 AI 骨龄评估系统功能评价

评价项目	未发生 (3 分)	发生 (2 分)	常发生 (1 分)
系统无法启动	无	有, 但能恢复, 不影响工作	有, 影响工作
操作与检出过程中意外死机	无	有, 但能恢复, 不影响工作	有, 影响工作
系统意外关闭或退出	无	有, 但能恢复, 不影响工作	有, 影响工作

表 3 AI 骨龄评估系统易用性评价

评价项目	满意 (3 分)	一般 (2 分)	不满意 (1 分)
系统功能界面设置	界面功能区域明确、操作流程清晰	界面功能区域欠明确、操作流程欠清晰	界面功能区域不明确、操作流程不清晰
图像浏览及处理功能	操作便捷, 响应灵敏	操作欠便捷, 响应一般	操作烦琐, 响应延迟
打印功能	操作便捷, 响应灵敏	操作欠便捷, 响应一般	操作烦琐, 响应延迟

3 结果

3.1 有效性

经过异常数据处理后得到的 1 060 例数据样本中, 对于评估结果为“骨骺发育成熟”的数据, 参照《中国人手腕部骨龄标准——中华 05 及其应用》^[15]一书中骨骺发育成熟的一般年龄进行量化, 其描述为: 根据成熟度分值第 50 百分位数曲线 RUS-CHN 法手腕骨发育成熟年龄为男 18 岁, 女 17 岁; TW3-C Carpal 法手腕骨发育成熟年龄为男 13.5 岁, 女 11.5 岁。因此将样本中评估结果为“骨骺发育成熟”的用上列年龄进行近似表示。此外, 由于 AI 系统的评估范围 R 系列在 2~18 周岁, C 系列在 3~18 周岁, 对于不在此范围的结果以字符描述, 如“骨龄小于 3 岁”, 经专家组讨论, 该类数据无法量化评估。因此, 在排除上述情况后, R 系列纳入分析的骨龄数据有 1 058 例 (男 331 例, 女 727 例), C 系列纳入分析的骨龄数据有 1 040 例 (男 319 例, 女 721 例)。采

用中位数 (四分位数间距) 描述所收集数据的基本特征, 见表 4。对于同一例样本, 若两位专家评估的骨龄差值小于等于 0.5 岁, 取平均值作为标准参考骨龄; 若差值大于 0.5 岁, 则取复核专家结果作为标准参考骨龄。对于 C 系列和 R 系列的评估结果, AI 均具有与标准参考骨龄相同的中位数与四分位间距, 表明二者在集中趋势及群体水平上的变异程度无明显差异。两位专家人工评估结果骨龄岁差分析, 见表 5。

表 4 骨龄评估结果的基本统计特征

变量	中位数 (四分位间距)
C 系列 (1 040 人)	
实际年龄	9.42 (3.08)
专家 1 骨龄	9.83 (2.92)
专家 2 骨龄	9.83 (2.92)
标准参考骨龄	9.83 (2.92)
AI 骨龄	9.83 (2.92)
R 系列 (1 058 人)	
实际年龄	9.42 (3.17)
专家 1 骨龄	10.08 (2.75)
专家 2 骨龄	10.08 (2.83)
标准参考骨龄	10.08 (2.83)
AI 骨龄	10.08 (2.83)

表 5 专家评估结果年龄岁差

岁差	R 系列		C 系列	
	数量 (例)	占比 (%)	数量 (例)	占比 (%)
[0, 0.1]	993	93.86	984	94.62
(0.1, 0.2]	10	0.95	4	0.38
(0.2, 0.3]	8	0.76	3	0.29
(0.3, 0.4]	14	1.32	8	0.77
(0.4, 0.5]	11	1.04	10	0.96
(0.5, +∞)	22	2.08	31	2.98

R 系列岁差 0.1 岁以内占比达到 93.86%，C 系列达到 94.62%。R 系列结果中需要复核的有 22 例，C 系列有 31 例。

AI 与标准参考骨龄的对比结果显示：C 系列平均绝对误差为 0.050、均方根误差为 0.282，R 系列平均绝对误差为 0.044、均方根误差为 0.234，组内相关系数基本大于 0.99，结合一致性检验结果可知，AI 与专家 1、专家 2 及标准参考骨龄具有较高的组间一致性 ($P < 0.001$)，见表 6。

表 6 AI 评估骨龄对比标准参考骨龄和专家评估骨龄的评价一致性分析

组别	平均绝对误差	均方根误差	组内相关系数	P
C 系列 (1 040 人)				
AI vs 标准参考骨龄	0.050	0.282	0.996 (0.995, 0.996)	<0.001
AI vs 专家 1	0.042	0.277	0.996 (0.995, 0.996)	<0.001
AI vs 专家 2	0.078	0.341	0.994 (0.993, 0.994)	<0.001
专家 1 vs 专家 2	0.040	0.205	0.998 (0.997, 0.998)	<0.001
R 系列 (1 058 人)				
AI vs 标准参考骨龄	0.044	0.234	0.992 (0.991, 0.993)	<0.001
AI vs 专家 1	0.037	0.230	0.993 (0.992, 0.994)	<0.001
AI vs 专家 2	0.058	0.278	0.989 (0.987, 0.990)	<0.001
专家 1 vs 专家 2	0.033	0.170	0.996 (0.995, 0.996)	<0.001

3.2 适宜性

系统功能和系统易用性各项量化评分结果均值均高于 2.9 分，见表 7。说明该 AI 系统具有较高的运行稳定性以及友好的系统功能界面，其中图像浏览及处理功能为临床较常用功能，能够取得评价高分也表明该 AI 系统的整体运行情况较优。

表 7 系统适宜性评价

评价维度	评价项目	均值 (标准差)
系统功能	系统无法启动	2.93 (0.34)
	操作与检出过程中意外死机	2.98 (0.14)
	系统意外关闭或退出	2.99 (0.11)
系统易用性	系统功能界面设置	2.99 (0.07)
	图像浏览及处理功能	2.97 (0.26)
	打印功能	2.99 (0.13)

展现出一定优势，对提升临床诊断准确性与效率，推动医疗服务同质化具有积极意义。在系统有效性方面，平均绝对误差、均方根误差及组内相关系数 3 个指标均处于理想区间，证明该系统具备较高的有效性。在系统适宜性方面，系统操作人员对其功能及易用性给予了较高评价，表明该 AI 系统具有良好的运行稳定性以及友好的功能界面。真实世界评价研究结果证实，该系统能够为医生的日常阅片工作提供辅助。通过 AI 自动化骨龄辅助评估，可有效减少人为判断差异导致的诊断偏差，使医生更加专注于病例分析与治疗决策。本研究有助于增强临床对于骨龄 AI 评估系统的信心，提高临床参与度。未来将纳入成本与效益评价指标，并进一步推广应用于其他 AI 辅助诊断产品评价。

作者贡献：吴崇数负责数据处理、论文撰写；竺智荣负责数据收集与处理；王立新负责统计分析；缪崇负责项目统筹协调；饶艳莺负责提供指导、数据标注、论文审核。

利益声明：所有作者均声明不存在利益冲突。

4 结语

在临床实际应用场景下，该骨龄 AI 评估系统

参考文献

- 1 MOHAMED K, MONA A. AI in diagnostic imaging: revolutionising accuracy and efficiency [J]. Computer methods and programs in biomedicine update, 2024, 5: 100146.
- 2 ZISIS K, PAVI E, GEITONA M, et al. Real - world data: a comprehensive literature review on the barriers, challenges, and opportunities associated with their inclusion in the health technology assessment process [J]. Journal of pharmacy & pharmaceutical sciences, 2024, 27 (2): 12302.
- 3 LIU F, PANAGIOTAKOS D. Real - world data: a brief review of the methods, applications, challenges and opportunities [J]. BMC medical research methodology, 2022, 22 (11): 287.
- 4 张涛, 张登国, 李建, 等. 人工智能影像系统在肺部结节诊断中的真实世界数据分析 [J]. 四川医学, 2021, 42 (2): 193 - 196.
- 5 邓杨, 李凤玲, 秦航宇, 等. 乳腺癌 Ki - 67 人工智能自动诊断系统的真实世界应用测试 [J]. 四川大学学报 (医学版), 2021, 52 (4): 693 - 697.
- 6 沈惠文, 林永忠, 陈淑良, 等. 基于人工智能算法的脑卒中溶栓药物精准治疗: 真实世界研究 [J]. 中国全科医学, 2023, 26 (17): 2070 - 2077, 2088.
- 7 宋鹏杰, 王艳蕾, 孙晨, 等. 基于真实世界数据的乳腺超声人工智能诊断和 X 线钼靶摄影检查在早期乳腺癌筛查中的对比研究 [J]. 中国医学装备, 2023, 20 (9): 20 - 24.
- 8 彭丽琴, 万雷, 汪茂文, 等. 机器学习在骨龄评估中的研究进展及展望 [J]. 法医学杂志, 2020, 36 (1): 91 - 98.
- 9 李亚璞, 刘戈力. 儿童骨龄评估和临床研究进展 [J]. 国际儿科学杂志, 2018, 45 (5): 365 - 368.
- 10 NADEEM W N W, GOH H G, ALI A, et al. Bone age assessment empowered with deep learning: a survey, open research challenges and future directions [J]. Diagnostics, 2020, 10 (10): 1 - 22.
- 11 吴莉婷, 吴瑞, 李书香, 等. 基于深度学习的人工智能技术与骨龄评估的研究现状 [J]. 中华实用儿科临床杂志, 2023, 38 (5): 394 - 397.
- 12 王杜春, 包勇, 何世秀, 等. 人工智能在儿童骨龄检测中的临床应用现状与展望 [J]. 影像研究与医学应用, 2024, 8 (13): 4 - 6.
- 13 张云亭, 于兹喜. 医学影像检查技术学 (第 3 版) [M]. 北京: 人民卫生出版社, 2010.
- 14 中国青少年儿童手腕骨成熟度及评价方法: TY/T 3001—2006 [S]. 北京: 中国标准出版社, 2006.
- 15 张绍岩. 中国人手腕部骨龄标准——中华 05 及其应用 [M]. 北京: 科学出版社, 2015.

《医学信息学杂志》编辑部严正声明

近期, 有不法人员冒充《医学信息学杂志》编辑部工作人员, 以核对收录信息、审核数据、发送录用通知等名义, 微信要求添加好友或发邮件, 进而收取稿件处理费等。以上行为严重侵害了广大作者、读者及本刊的权益, 编辑部保留追究其法律责任的权利。本刊特此严正声明: (1)《医学信息学杂志》暂不通过微信进行正式沟通, 未委托任何个人或机构代理收稿、征稿等业务, 唯一投稿渠道为杂志官网 <http://www.yxxxx.ac.cn> 在线投稿, 点击“作者投稿”按钮注册后即可投稿。杂志官方邮箱为 yxxxx@imicams.ac.cn 和 yxxxxzz01@163.com, 不会通过其他邮箱发送通知或对外联系。(2)《医学信息学杂志》是中国知网、万方数据、维普网全文收录期刊。期刊出版后 1 个月左右可在上述数据库查阅论文。(3)《医学信息学杂志》编辑部目前收取版面费仅有对公转账一种方式。开户行: 中国建设银行北京雅宝路支行; 开户名称: 中国医学科学院医学信息研究所; 银行账号: 11001028400059856368。

敬请广大作者、读者提高警惕, 请勿向任何个人账户支付任何费用, 以免造成不必要的损失。必要时建议通过报警等方式维护正当权益。遇到任何问题可拨打联系电话 010 - 52328686, 52328687 与编辑部取得联系。

《医学信息学杂志》编辑部