

## 临床研究

DOI: 10.13406/j.cnki.cyx.003402

## AI 辅助下医生对骨龄评估的效能提升

武 鹏<sup>1</sup>, 刘新顶<sup>1</sup>, 赵德利<sup>1</sup>, 靳翠翠<sup>1</sup>, 孙 蕊<sup>2</sup>

(1. 哈尔滨医科大学附属第六医院江北院区影像中心, 哈尔滨 150025;

2. 哈尔滨医科大学附属第六医院江南院区新生儿内科, 哈尔滨 150025)

**【摘要】目的:**比较放射科医师在人工智能(artificial intelligence, AI)骨龄评价系统辅助前后对儿童左手 X 线摄影的骨龄评估效能。**方法:**回顾性分析在我院就诊的 300 例患儿左手 X 线平片。骨龄评测采用中华-05 骨龄评定标准, 两位低年资医师(医师 1 及医师 2)分别在有无 AI 辅助下分别记录左手各骨质的骨龄发育等级, 并记录时间。以两位高年资放射医师分别在有无 AI 系统辅助下评估结果的均值为参考标准, 计算骨龄测评的准确率、均方根误差(root mean square error, RMSE)、测评时间。**结果:**无 AI 辅助下, 医师 1 及医师 2 分析差值在 6 个月及 12 个月诊断准确率分别为 77.3% 和 83%、88.7% 和 93.7%, RMSE 值分别为 9 和 8。在 AI 辅助下, 医师 1 及医师 2 分析差值在 6 个月及 12 个月诊断准确率分别为 88.7% 和 90.3%、97% 和 97.3%, RMSE 值分别为 6 和 6; 差异均具有统计学意义。无 AI 辅助下, 实验组医师和标准组医师, 平均评测耗时分别为 124.79 s 和 89.13 s; 有 AI 辅助下实验组医师和标准组医师, 平均评测耗时分别为 86.10 s 和 63.87 s, 在 AI 辅助下平均评测耗时均有较大幅度减少( $P=0.000$ )。**结论:**AI 辅助骨龄评价系统可显著提高医师工作效率, 减少阅片时间。

**【关键词】**影像诊断; 骨龄评估; 人工智能**【中图分类号】**J320.1140**【文献标志码】**A**【收稿日期】**2023-10-10

## Improvement in bone age assessment efficiency of physicians based on the artificial intelligence-assisted bone age assessment system

Wu Peng<sup>1</sup>, Liu Xinding<sup>1</sup>, Zhao Deli<sup>1</sup>, Jin Cuicui<sup>1</sup>, Sun Rui<sup>2</sup>

(1. Imaging Center, The Sixth Affiliated Hospital of Harbin Medical University [Jiangbei Hospital Area];

2. Department of Neonatal Medicine, The Sixth Affiliated Hospital of Harbin Medical University  
[Jiangnan Hospital Area])

**【Abstract】Objective:** To compare the bone age assessment efficiency of radiologists for children by left hand radiography before and after the implementation of the artificial intelligence (AI)-assisted bone age assessment system. **Methods:** We conducted a retrospective analysis of left hand X-ray plain films of 300 children treated in our hospital. The China-05 standards were used to assess bone age. The bone age development grade of each bone in the left hand was assessed by two junior physicians (physician 1 and physician 2, experimental group) with and without the assistance of the AI system, and the time was recorded. The accuracy, root mean square error (RMSE), and time of bone age assessment were calculated with the mean values of assessment results of two senior radiologists (control group) with and without the assistance of the AI system as the reference standards. **Results:** Without the assistance of the AI system, the diagnostic accuracy rates of physician 1 and physician 2 were 77.3%/83% and 88.7%/93.7% at month 6 and month 12, respectively, and the RMSE values were 9 and 8, respectively. With the assistance of the AI system, the diagnostic accuracy rates of physician 1 and physician 2 were 88.7%/90.3% and 97%/97.3% at month 6 and month 12, respectively, and the RMSE values were 6 and 6, respectively, showing significant differences. Without the assistance of the AI system, the mean assessment time of physicians in the experimental and control groups was 124.79 s and 89.13 s, respectively. With the assistance of the AI system, the mean assessment time of physicians in the experimental and control groups was 86.10 s and 63.87 s, respectively. By utilizing AI, the mean assessment time was significantly reduced ( $P<0.001$ ). **Conclusion:** The AI-assisted bone age assessment system can significantly improve physicians'

work efficiency and reduce the film reading time.

**作者介绍:**武 鹏, Email: 343187917@qq.com,

研究方向: 儿童行为发育预测评估。

**通信作者:**刘新顶, Email: liuxinding123@163.com。**优先出版:** <https://link.cnki.net/urlid/50.1046.R.20240119.1436.006>

(2024-01-22)

**【Key words】**imaging diagnosis; bone age assessment; artificial intelligence

随着科技飞速发展,我国的普通家庭的平均生活水平得到明显提高,绝大多数家长也越来越关心孩子的成长发育。骨龄(boneage, BA)是目前应用于评价儿童及青少年生物年龄的最主流方法,在体育及司法领域都有着重要的应用价值。全面且详细的骨龄评估,能准确地反映出儿童及青少年的生长发育水平<sup>[1]</sup>。医师采用图谱法(G-P 图谱、CHN 图谱),计分法(TW3 法、中华 05 法及 CHN 法)进行人工阅片,是目前主要的骨龄评价方法,此方法读片耗时较长,并且主观性强、稳定性差,导致骨龄的评估效率较低,评估结果一致性较差<sup>[2]</sup>。随着医疗领域的智能化发展,如何有效地利用人工智能(artificial intelligence, AI)图像处理能力,提高骨龄评估的数字化发展,缓解医疗资源紧张的局面,逐渐引起相关领域科研人员及临床医师的重视。深度卷积神经网络(Deep Convolutional Neural Network, DCNN)等一系列深度 AI 智能学习的不断更新发展,使其具有快速处理大批量的图像数据的能力。本研究通过 AI 与影像医师组成不同实验组的组间对比,探讨 AI 与医师在何种模式配合下对儿童及青少年骨龄评估的临床应用价值更高。

## 1 资料与方法

### 1.1 一般资料

回顾性分析因矮小、性早熟及预测身高等生长发育问题就诊哈尔滨医科大学附属第六医院的患儿 300 例,其中男童 97 例,女童 203 例;年龄 1~18 岁,平均年龄:8 岁 9 月,其中 7~

12 岁占比最多,为 71.67%;纳入标准:①儿童年龄 1~18 岁;②左手(包含左手掌骨、指骨、腕骨以及尺桡骨远侧骨干 3~4 cm)正位骨龄数字化 X 线摄影片。③手部位摆位及投射点正确,无骨骺缺失。排除标准:①严重骨畸形;②手腕部结构包含不全。

### 1.2 仪器与方法

仪器采用安健 Angell-DR-D-3 数字化 X 线摄影机,入组病例均行左手后前位 X 线平片检查,为保证检查结果的一致性,所有研究对象均使用同一型号 DR 仪器进行检查,并且技术人员均通过相应技术操作培训。具体方法如下:①左手腕部正位 X 线片,图像包含左手及腕部骨关节全部,也包括尺桡骨远端(桡骨和尺骨远端暴露 $\geq 3$  cm)。②左手掌面向下紧贴床面,中指长轴与前臂长轴呈一条直线,五指自然分开,拇指与手掌约呈 30°角。摄影焦片距控制在 75 cm 左右,并使用遮光栅。③球管十字定位中心线与手掌垂直,正对第三掌骨头进行投照,焦片距保持在合适范围,避免骨骺重叠<sup>[3]</sup>。④将骨龄片通过 DICOM 格式导入 PACS 系统中。

### 1.3 AI 支持系统

采用医准医疗的人工智能系统软件,掌指骨检测采用目标检测模型 Faster-RCNN,腕骨检测采用语义分割模型 Deeplab v3,各关节分类采用 backbone 为 Densenet 的序回归模型。图像展示关键点位分级评级标准,针对手掌 20 块骨头(13 块掌指骨,7 块腕骨)进行识别定位,并对每一块骨头进行分级。采用医准医疗的人工智能系统软件分析,内嵌中华 05 法 TW3-C-Rus 和 TW3-C-Carpal 双重计算结果,该系统可以 AI 单独测评,也可以将图像上关键点位分级评级信息提供给医师,由医师完成骨龄最终评测。所有 X 线平片以 DCIOM 格式导入 AI 支持系统,系统进行自动骨龄评估(图 1)。



图1 AI与医师相互配合下的骨龄评估分析

#### 1.4 骨龄测评

标准组:由两位高年资具有独立审核资质的影像医师组成,首先共同对入组 300 张骨龄片进行骨龄评估,评估方法采用适合中国儿童发育规律的中华 05 法骨龄标准,采用平均值作为参考标准<sup>[4]</sup>。同时从打开图像开始计时,直至完成报告关闭图像结束,记录期间的评测时间。

试验组:2 位低年资仅具备报告初写能力的年轻医师共同参与本次骨龄评估。骨龄评测使用中华 05 标准对入组影像进行判读,医师 1 及医师 2 分别在有无 AI 辅助下进行骨龄评测,审阅所有左手 X 线平片;分别记录左手各骨质的骨龄发育等级,并以相同方法记录时间。再由 2 位高年资医师联合 AI 辅助,对同一入组数据进行骨龄评测,并以相同方法记录时间。考虑到所有的医生都对入组数据进行了 2 次骨龄评估,为减少第 1 次评估对第 2 次评估的影响,实验将两次评估间隔设定在 2~3 月左右。

#### 1.5 统计学方法

采用 SPSS 26.0 统计分析软件。本实验将实验组与标准组之间的误差范围分别定为 6 个月组以及 12 个月组,即二者骨龄评测结果在 6 个月以内或 12 个月以内定义为结果一致;分别观察实验组与标准组之间误差在 6 个月及 12 个月的准确率及均方根误差 (root mean squared error, RMSE)。采用组内相关系数 (intraclass correlation coefficient, ICC) 评价两位医师分别在有无 AI 辅助下评测骨龄结果的一致性,采用 Bland-Altman 法比较两位医师分别在有无 AI 辅助下评测骨龄结果之间的平均差异。计数资料用率表示,分类变量比较采用卡方检验,检验水准  $\alpha=0.05$ 。

## 2 结 果

#### 2.1 不同实验组之间预测差异分析

2.1.1 两位医师有无 AI 辅助下误差在 6 个月及 12 个月内的准确率 组间判读结果比较分析后,显示在 AI 辅助下,2 位医师误差在 6 个月及 12 个月内的准确率均有所提高,医师 1 及医师 2 无 AI 辅助,6 个月及 12 个月诊断正确率分别为 77.3% 和 83%、88.7% 和 93.7%;医师 1 及医师 2 在 AI 辅助下,6 个月及 12 个月诊断正确率分别为 88.7% 和 90.3%、97% 和 97.3%;且差异均有统计学意义 ( $P<0.05$ );医师 1 及医师 2 无

AI 辅助下, RMSE 值分别为 9 和 8, 医师 1 及医师 2 在 AI 辅助下, RMSE 值分别为 6 和 6; 2 位医师在 AI 辅助下评估骨龄的 RMSE 值均减低 (表 1)。

2.1.2 各组间 Bland-Altman 分析 Bland-Altman 分析发现, 医师 1 在 AI 辅助下, 95% 误差分布区间为 (-10.964~12.904), 无 AI 辅助的 95% 误差分布区间为 (16.674~18.060)。医师 2 在 AI 辅助下, 95% 误差分布区间为 (-9.297~12.197), 无 AI 辅助的 95% 误差分布区间为 (-12.982~16.336), 说明在 AI 辅助下, 两位年轻医师的评价结果与金标准的一致性显著提高 (图 2)。

#### 2.2 2 组评测左手骨龄的一致性比较

以标准组评估的月数为结局, 观察 2 位医师在有无 AI 支持下的 ICC 值。医师 1、2+AI 组的 ICC 值为 0.981, 医师 1、2 独立组 ICC 值为 0.958, 医师 1、2+AI 组具有更高的一致性 (表 2)。

#### 2.3 试验组与标准组评测时间比较

AI 辅助下实验组医师和标准组医师, 平均评测耗时分别为 86.10 s 和 63.87 s; 无 AI 辅助下耗时分别为 124.79 s 和 89.13 s; 在 AI 辅助下平均评测耗时均有较大幅度减少 (表 3)。

## 3 讨 论

本次研究通过以中华-05 为理论基础, 结合卷积神经网络模型叠加建立网络智能模型, 对各个试验组对比研究, 观察试验组间骨龄评估的准确性, 最终目的是探索出 AI 与医师在实际临床工作中, 处于何种模式配合下, 对于儿童及青少年骨龄评估的临床应用价值更高, 进而更有助于临床对儿童内分泌问题的诊断和治疗。现阶段, 大多数医院依然使用传统骨龄评价方法, 即计数法、图谱法和计分法 3 种方法, 计数法通过观察和计算腕部骨化中心的数量, 推算骨龄, 由于误差较大目前应用较少。图谱法 (Greulich-Pyle 图谱法为代表) 是 20 世纪初以西方儿童为样本, 通过大样本追踪和整理后得出标准

表 1 两位医师有无 AI 辅助下误差在 6 个月及 12 个月内的准确率 (n, %)

医师	时间	分组	诊断正确 n (%)	$\chi^2$ 值	P 值
医师 1	6 个月	无 AI 辅助	232 (77.3)	111.681	0.000
		有 AI 辅助	266 (88.6)		
医师 2	12 个月	无 AI 辅助	274 (91.3)	75.436	0.000
		有 AI 辅助	291 (97.0)		
医师 1	6 个月	无 AI 辅助	249 (83.0)	143.987	0.000
		有 AI 辅助	271 (90.3)		
医师 2	12 个月	无 AI 辅助	281 (93.6)	43.709	0.000
		有 AI 辅助	292 (97.3)		



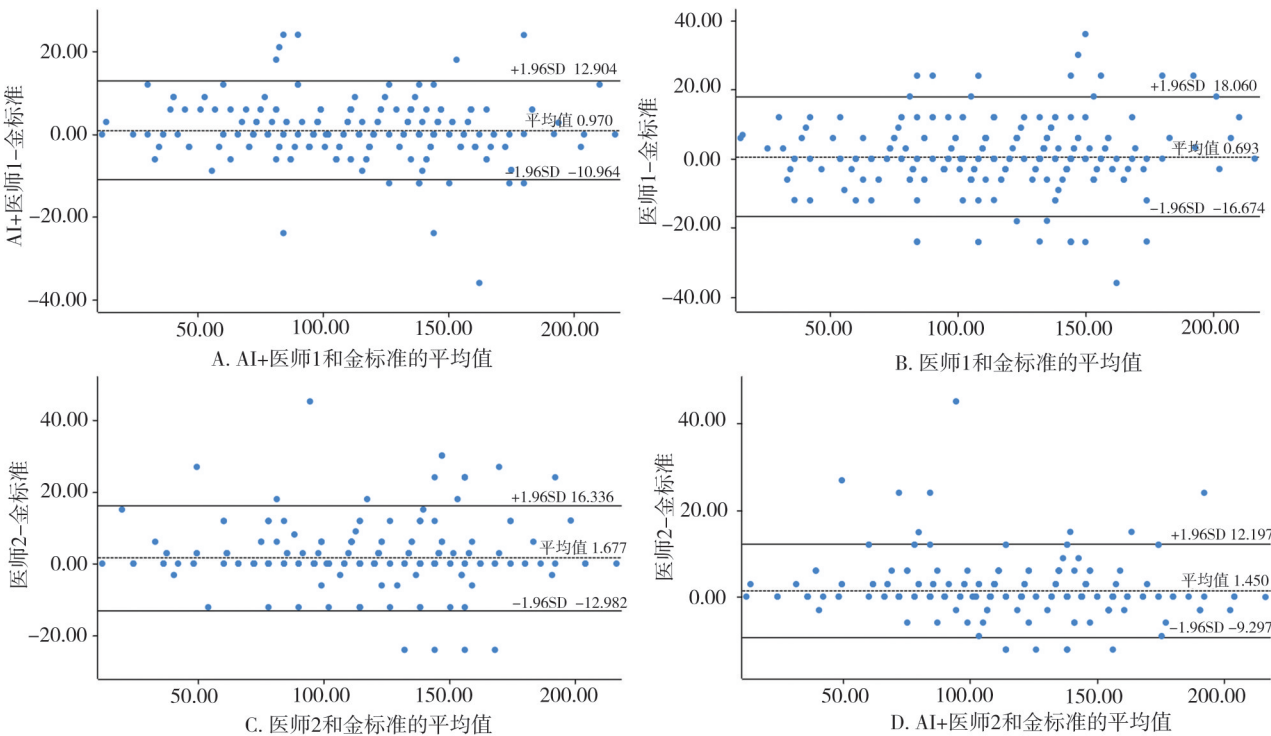


图2 各组间Bland-Altman分析

表2 两位医师在有无AI支持下的ICC值

项目	ICC	95%可信区间	P值
医师1 vs. 医师2	0.958	0.948~0.967	0.000
医师1+AI vs. 医师2+AI	0.981	0.976~0.984	0.000

表3 各组医师不同情况下评测时间比较

组别	评测时间(s)	t值	P值
实验组医师单独评测时间	124.793 ± 17.616	33.596	0.000
AI+实验组医师单独评测时间	86.103 ± 9.834		0.000
审核医师单独评测时间	89.127 ± 18.512	22.654	0.000
AI+审核医师单独评测时间	63.867 ± 5.121		0.000

骨龄图谱,再将被检者的骨龄片与标准骨龄图谱(即不同年龄层的平均水平)比较<sup>[5]</sup>,该方法目前使用较广泛,但除了受不同地域饮食及生活习惯影响、人种影响,也受医师主观因素影响,同一医师或不同经验医师,评估的结论均有不同程度差异,可重复性差。有研究显示,对同一位医师进行测试,方法是采用G-P图谱法对被检者的骨龄片进行2次骨龄评估,2次骨龄评估结果差异明显,平均为0.89岁,不同医师之间的骨龄评估也具有差异性,差异平均为1.25岁<sup>[6-7]</sup>;计分法(中华05法为代表)是我国专家在TW计分法基础上于2006年修订制定,该方法符合我国儿童的发育特点,评价骨龄更客观、准确,但中华05法在实际的评价过程中耗时、费力,因

此在实际工作中应用受到一定限制。

近些年来,AI在不同医学领域中的应用逐渐上升,AI的深度学习在视觉分析和自然语言处理等自动计算领域,取得了显著突破,在自动骨龄评估中已成为目前主流方法。基于深度学习的AI骨龄评估模型,属于一门基于多学科的新兴技术,可智能检测模型包括影像体位自动检测、关键骨化中心识别、关键骨化中心评级、生长发育测评等;根据我国最新制定的中华-05骨龄评分法,针对本地区不同人群间掌指骨、腕骨及尺桡骨对应特征进行训练,设计适应我国人群新的空间注意力机制,应用通道注意力、轻量级卷积块注意力、挤压和激发注意力及空间注意力<sup>[8]</sup>等多种机制处理自然语言、分类图像及分割语义等任务的表现出众,与原始卷积网络比较,可显著减弱处理高维输入数据时的计算负荷,降低数据维度,使网络更集中于训练ROI(Region of Interest),使ROI空间注意力的权重更高,模型训练更加集中;清晰显示手腕X线片中的关键部位局部信息,提取更强的特征,为进一步骨龄评估处理提供基础。早期研究<sup>[9]</sup>卷积神经网络模型的相关测试,结果显示通过一系列深度学习算法,可较好地自动评价骨龄。2017年Spampinato C等科研人员<sup>[10]</sup>通过搭建“Bonet”网络,对输入图像进行点对点

的深度骨龄回归学习,首次利用卷积神经网络(Convolutional Neural Network, CNN)模型进行自动骨龄评估,结果显示平均评估误差低于 10 个月。随着模型不断发展,其他研究<sup>[11-12]</sup>也得到比较大的收获,诸如 AI 系统对于 5~18 岁骨龄评价一岁以内的误差<92.2%。有研究通过分析基于 G-P 标准 AI 系统的阅片效率和结果准确性,其与金标准相差 1 岁以内的平均比例为 84.60%,平均判读耗时 AI 系统仅需要 1~2 s,阅片效率和结果准确性均有进一步的提高<sup>[13]</sup>。

本研究对入组数据的对比研究后显示:实验组两位医师诊断正确率由最初的 77.3% 和 83%、88.7% 和 93.7%;在联合 AI 辅助后都有明显改变,上升至 88.7% 和 90.3%、97% 和 97.3%;RMSE 值均有不同程度降低;在本研究中通过 Bland-Altman 分析,观察两位医师在有无联合 AI 情况下的 ICC 值,医师 1、2+AI 组的 ICC 值为 0.981,医师 1、2 独立组 ICC 值为 0.958,显示医师+AI 的组合在骨龄评估环节,在结果间比较中具有更高的一致性,以上研究数据对于日后儿童及青少年的临床复查有着积极的影响。在时间方面,不同组别的医师在联合 AI 后,平均评测耗时均有较大幅度减少,阅片时间上相比于人工有明显优势。本研究不足之处是样本量比较小,在年龄层面上大部分局限在 7~12 岁之间,导致没有更进一步细化讨论各个年龄层之间的骨龄评估效能。

AI 辅助骨龄评价系统可提高医师工作效率,显著减少阅片时间,医生间结果一致性得到提升,利于患儿复查。具有良好的临床应用前景,成为骨龄评估的主要发展方向。

## 参 考 文 献

- [1] 何晓芬,唐桂波,李国峰,等. 青海省西宁市青少年腕部骨龄评价[J]. 实用放射学杂志,2011,27(9):1396-1398.  
He XF, Tang GB, Li GF, et al. Investigation of wrist bone age of adolescents in Xining area of Qinghai Province[J]. J Pract Radiol, 2011, 27(9): 1396-1398.
- [2] Wicks EC, Menezes LJ, Barnes A, et al. Diagnostic accuracy and prognostic value of simultaneous hybrid 18F-fluorodeoxyglucose positron emission tomography/magnetic resonance imaging in cardiac sarcoidosis[J]. Eur Heart J Cardiovasc Imaging, 2018, 19(7):757-767.
- [3] Kagioka Y, Yasuda M, Okune M, et al. Right ventricular involvement is an important prognostic factor and risk stratification tool in suspected cardiac sarcoidosis: analysis by cardiac magnetic resonance imaging[J]. Clin Res Cardiol, 2020, 109(8):988-998.
- [4] Tajmir SH, Lee H, Shailam R, et al. Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability[J]. Skeletal Radiol, 2019, 48(2):275-283.
- [5] Berst MJ, Dolan L, Bogdanowicz MM, et al. Effect of knowledge of chronologic age on the variability of pediatric bone age determined using the Greulich and Pyle standards[J]. AJR Am J Roentgenol, 2001, 176(2):507-510.
- [6] 国家卫生健康委员会 2019 年 1 月 25 日例行新闻发布会文字实录[EB/OL].[2019-01-25].<http://www.nhc.gov.cn/xcs/s7847/201901/af1b200319ab4d4c9a8281a39a8be37a.shtml>.  
Transcript of the regular press conference of the National Health Commission on January 25, 2019[EB/OL].[2019-01-25].<http://www.nhc.gov.cn/xcs/s7847/201901/af1b200319ab4d4c9a8281a39a8be37a.shtml>.
- [7] 张鹏飞,李 辉. 三种骨龄评价方法在 3~17 岁儿童临床应用中的 consistency 比较研究[J]. 中国循证儿科杂志,2017,12(4):263-267.  
Zhang PF, Li H. Comparison of the consistency of three skeletal age methods in 3 to 17 years old children[J]. Chin J Evid Based Pediatr, 2017, 12(4):263-267.
- [8] Chen L, Zhang H, Xiao J, et al. SCA-CNN:Spatial and channel-wise attention in convolutional networks for image captioning[EB/OL].[2017-11-19]. <https://ieeexplore.ieee.org/document/8100150>.
- [9] Spampinato C, Palazzo S, Giordano D, et al. Deep learning for automated skeletal bone age assessment in X-ray images[J]. Med Image Anal, 2017, 36:41-51.
- [10] Spampinato C, Palazzo S, Giordano D, et al. Deep learning for automated skeletal bone age assessment in X-ray images[J]. Med Image Anal, 2017, 36:41-51.
- [11] Kim JR, Shim WH, Yoon HM, et al. Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency[J]. AJR Am J Roentgenol, 2017, 209(6):1374-1380.
- [12] Lee H, Tajmir S, Lee J, et al. Fully automated deep learning system for bone age assessment[J]. J Digit Imaging, 2017, 30(4):427-441.
- [13] Wang FD, Gu X, Chen S, et al. Artificial intelligence system can achieve comparable results to experts for bone age assessment of Chinese children with abnormal growth and development[J]. PeerJ, 2020, 8:e8854.

(责任编辑:李青颖)