

· 临床医学 ·

## 人工智能软件辅助诊断新鲜肋骨骨折的效能评估

朱雅茹, 祁良, 徐磊, 王梦悦, 徐蕴潮, 纪执琳, 邹月芬\*

南京医科大学第一附属医院放射科, 江苏 南京 210029

**[摘要]** 目的: 评估人工智能软件(artificial intelligence system, AI)辅助放射科医师在胸部计算机断层扫描(computed tomography, CT)新鲜肋骨骨折诊断中的应用效能。方法: 收集行胸部CT检查的新鲜肋骨骨折病例共508例。6名影像医师分为低年资组( $\leq 5$ 年)和高年资组( $> 5$ 年), 使用PACS系统对CT图像进行独立阅片; 在间隔为期4周的洗脱期后, 影像医师在结合AI辅助诊断结果的前提下对胸部CT图像进行第2轮阅片; 记录骨折类型、部位及诊断时间。采用配对卡方检验比较AI辅助前后诊断骨折的灵敏度、特异度有无统计学差异。绘制受试者操作特征(receiver operating characteristic, ROC)曲线, 并计算曲线下面积(area under curve, AUC), 应用Medcalc软件对AUC进行显著性检验。采用Cohen's kappa系数分析低年资医师和高年资医师诊断肋骨骨折的一致性。应用配对样本 $t$ 检验比较人工智能辅助前后医师的诊断时间是否具有差异。结果: 508例胸部CT共包含骨折2 883处, AI辅助后低年资医师和高年资医师的诊断灵敏度由77.95%、83.96%提升至88.52%、90.98% ( $P < 0.001$ ); 平均AUC从0.902增加到0.948 ( $P < 0.001$ ); 独立阅片和AI辅助诊断后低年资医师与高年资医师之间的Cohen's kappa系数由0.832提升至0.900; 诊断每个肋骨骨折病例的时间平均减少了28.43 s ( $P < 0.001$ )。结论: AI辅助医师诊断新鲜肋骨骨折能够提高检出效能, 并减少诊断时间。

**[关键词]** 人工智能; 新鲜肋骨骨折; 计算机断层扫描; 效能; 阅片时间

**[中图分类号]** R814.42

**[文献标志码]** A

**[文章编号]** 1007-4368(2022)02-227-06

**doi:** 10.7655/NYDXBNS20220214

## Effectiveness evaluation of artificial intelligence system in the diagnosis of fresh rib fractures

ZHU Yaru, QI Liang, XU Lei, WANG Mengyue, XU Yunchao, JI Zhilin, ZOU Yuefen\*

Department of Radiology, the First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China

**[Abstract]** **Objective:** This study aims to evaluate the effectiveness of using an artificial intelligence system to assist radiologists in the diagnosis of fresh rib fractures with computed tomography (CT). **Methods:** A dataset of 508 cases with fresh rib fracture CT images were collected and analyzed by 6 radiologists independently with the PACS system, the radiologists were divided into two groups: low seniority group ( $\leq 5$  years) and high seniority group ( $> 5$  years). After a washout period of four weeks, the CT images were evaluated again by the 6 radiologists with the assistance of AI system. The fracture type, site, and time of diagnosis were recorded for further analysis. The paired chi-square test was used to compare the sensitivity and specificity in diagnosis of fresh fracture with and without AI assistance. The receiver operating characteristic curve (ROC) and the area under curve (AUC) was calculated. Cohen's kappa coefficient was used to analyze the consistency between low and high seniority radiologists in diagnosing rib fractures. Paired sample  $t$  test was used to compare the difference in diagnosis time. **Results:** A total of 2 883 fresh rib fractures were found in the 508 CT images. With the help of AI-assisted diagnosis, the diagnostic sensitivity of low and high seniority radiologists increased significantly from 77.95%, 83.96% to 88.52%, 90.98% ( $P < 0.001$ ); The average AUC increased from 0.902 to 0.948 ( $P < 0.001$ ); The Cohen's kappa coefficient between low and high seniority radiologists increased from 0.832 to 0.900. The average diagnosis time of each case decreased 28.43s,  $P < 0.001$ . **Conclusion:** With the AI assistance in the diagnosis of fresh rib fracture, the detection efficiency of rib fracture is improved and the diagnosis time is reduced.

**[Key words]** artificial intelligence; fresh rib fracture; computed tomography; efficiency; diagnosis time

[J Nanjing Med Univ, 2022, 42(02):227-232]

**[基金项目]** 国家自然科学基金(81701652)

\*通信作者(Corresponding author), E-mail: zou\_yf@163.com

胸部创伤导致的直接死亡占创伤相关病死率的25%<sup>[1]</sup>,胸部创伤可分为胸壁、肺及心血管损伤,而肋骨在胸部创伤中最易受累。钝性胸部损伤由减速性、挤压性、撞击性或冲击性暴力所致;约40%的钝性胸部损伤会引起肋骨骨折<sup>[2]</sup>。肋骨骨折一定程度上代表着损伤严重程度,患者病死率随着肋骨骨折数量增多而增高,80%两根肋骨以上的肋骨骨折会引起血胸或气胸。肋骨骨折患者的住院率和病死率高于无肋骨骨折的患者<sup>[3-4]</sup>。传统诊断肋骨骨折的常用方法是胸部X线平片,但由于照射体位、肋骨骨折位置、隐匿性骨折等原因易导致骨折的误诊或漏诊。CT扫描显示外伤性肋骨骨折较X线平片能提供更加丰富、准确的影像学信息<sup>[5-6]</sup>。

由于肋骨数多而形态不规则,临床工作中需要从上百张CT图像中观察肋骨骨折,工作量巨大,有研究报道,仅有约2/3的肋骨骨折在胸部CT中被诊断<sup>[7]</sup>。肋骨骨折的漏诊及误诊经常会引起医疗纠纷,特别是在打击或车祸等病例中,骨折的数量、对位情况有时在伤害的程度及伤残等级的认定上起着决定性作用<sup>[8]</sup>。

目前已有研究报道基于卷积神经网络(convolutional neural network, CNN)深度学习的人工智能软件可以对海量CT图像进行初步筛查并标记可疑肋骨骨折,可以帮助医师减少工作量并提高诊断准确率<sup>[9-10]</sup>。

本研究旨在评估基于深度学习的人工智能软件(artificial intelligence, AI)辅助放射科医师在胸部CT新鲜肋骨骨折诊断中的应用价值及骨折形态和骨折位置对骨折检出率的影响。

## 1 对象和方法

### 1.1 对象

收集南京医科大学第一附属医院放射科2018年

6月—2020年5月期间共637例7 d内有外伤病史并行胸部CT检查发现肋骨有新鲜骨折的病例。图像纳入标准:①有薄层胸部CT图像(层厚≤1.25 mm);②有临床胸部损伤史。排除标准:①图像有引流管或金属内固定物;②肿瘤肋骨转移的病理性骨折;③严重运动伪影;④肋骨肿瘤及转移瘤;⑤有陈旧性骨折病史的病例;⑥肋骨先天发育异常。最终共508例患者纳入本研究,平均年龄(56.19 ± 14.86)岁,年龄范围12~97岁;其中,男316例,平均年龄(54.11 ± 14.09)岁,年龄范围12~87岁;女192例,平均年龄(59.60 ± 15.49)岁,年龄范围18~97岁。

### 1.2 方法

#### 1.2.1 CT检查

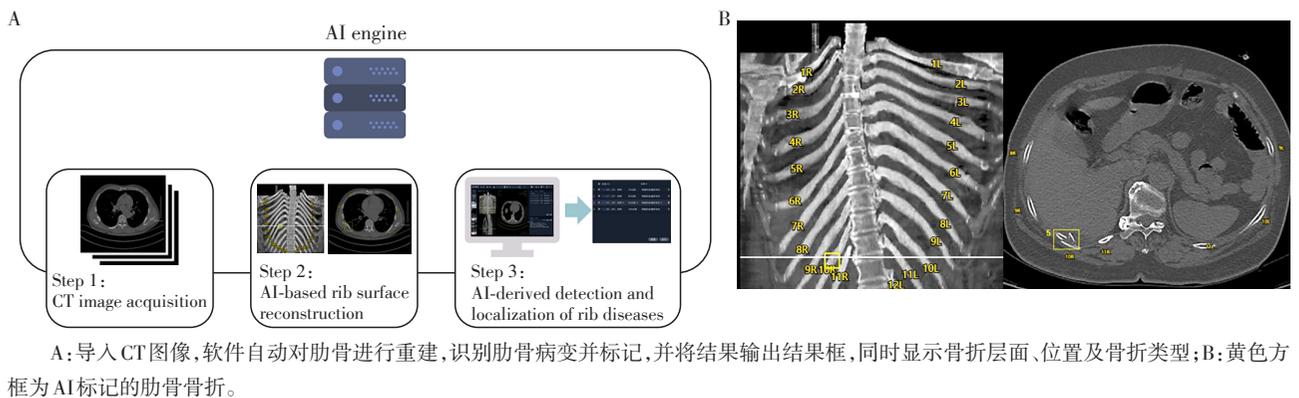
患者取仰卧位,上臂上举(部分患者因肩部或上肢外伤不能上举则自然置于身体两侧),扫描范围自胸廓入口至肺底部或全部肋骨,患者1次吸气后屏气完成扫描。扫描方式:螺旋扫描;管电压110 kV或120 kV;层厚:5.0 mm,使用骨算法重建1.25 mm或1.5 mm层厚轴位图像,图像矩阵:512×512,视野:360 mm。

#### 1.2.2 AI

本研究使用的AI是由杭州依图医疗技术有限公司开发,该软件基于CNN对左右肋骨进行分割及分段,同时三维重建,并检出骨折,标记其骨折位置和骨折类型(图1)。将508例胸部CT原始数据传输至工作站,软件系统自动批量进行肋骨骨折识别和标记。

#### 1.2.3 骨折标准认定

2名高级职称影像诊断医师在胸部CT薄层图像上对肋骨骨折进行判读,通过多平面重建、三维重建等方法并结合复查胸部CT观察有无骨痂形成来判定有无骨折;如无复查CT,以两者的一致性意见为骨折诊断标准,同时记录骨折的数目及形态。



A: 导入CT图像,软件自动对肋骨进行重建,识别肋骨病变并标记,并将结果输出结果框,同时显示骨折层面、位置及骨折类型;B: 黄色方框为AI标记的肋骨骨折。

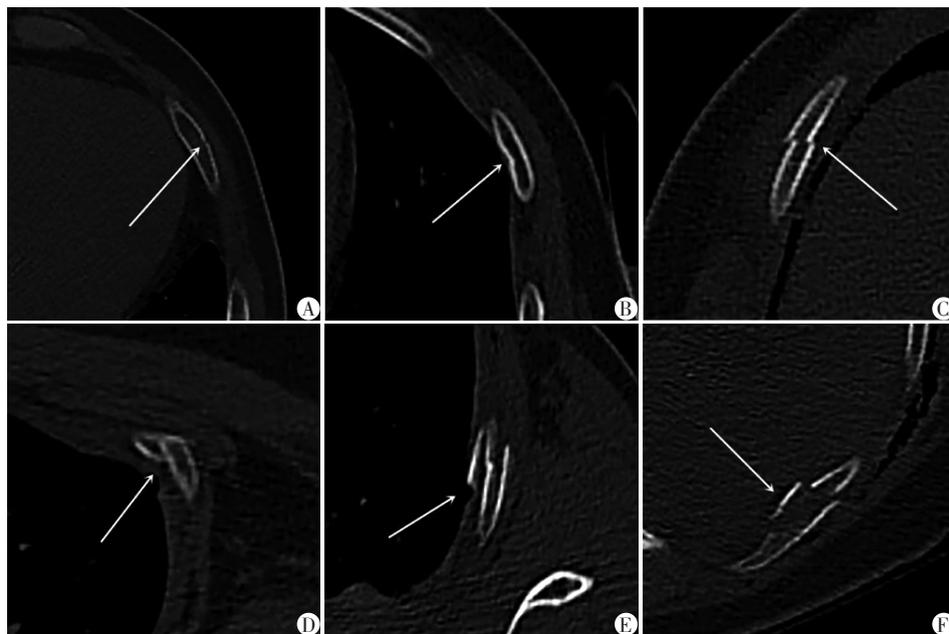
图1 人工智能软件检测肋骨骨折流程及结果界面

Figure 1 The process of AI system detecting rib fracture and display of results

骨折形态按照骨折断端是否移位分为3类,即无错位骨折[包括裂隙状骨折(图2A)及骨皮质扭曲(图2B)]、部分错位骨折[骨折断端不完全错位(图2C、D)]及完全错位骨折[包括分离骨折(图2E)及粉碎性骨折(图2F)]<sup>[11]</sup>。

#### 1.2.4 阅片

6名影像医师对508例患者胸部CT图像进行两轮判读,6名医师中3名为工作年限≤5年的低年资医师,3名为工作年限>5年的高年资医师。第一轮独立阅片时,6名影像医师使用医疗影像存储与传



A: 裂隙状骨折; B: 骨皮质扭曲型骨折; C、D: 部分错位骨折; E: 分离骨折; F: 粉碎性骨折。

图2 不同肋骨骨折形态示意图

Figure 2 Different patterns of rib fractures

输系统(picture archiving and communication system, PACS)对CT图像进行独立阅片,阅片过程中医师可根据需要调整窗宽、窗位、放大和缩小、使用多平面重建(MPR)、容积重现(VR)等。在间隔为期4周的洗脱期后,6名影像医师在结合AI辅助诊断结果的前提下对胸部CT图像进行第二轮判读。诊断过程中记录骨折位置、类型及阅片用时。

#### 1.3 统计学方法

应用SPSS 26.0统计学软件进行资料录入、整理及统计学分析, $P < 0.05$ 为差异有统计学意义。采用配对卡方检验(McNemar 检验)比较AI辅助前后影像医师检出肋骨骨折的能力差异有无统计学意义,并计算灵敏度与特异度。绘制受试者操作特征(receiver operating characteristic, ROC)曲线,并计算曲线下面积(area under curve, AUC),应用 Medcalc 软件对AUC进行显著性检验。采用Cohen's kappa系数分析低年资医师和高年资医师诊断肋骨骨折的一致性。AI辅助前后影像医师诊断每个肋骨骨折病例的时间使用配对t检验进行检验。

## 2 结果

### 2.1 肋骨骨折分类特征

508例CT图像经2位影像医师判读共包含2 883处骨折,共累及2 368根肋骨。其中,无错位骨折1 401处、部分错位骨折973处、完全错位骨折509处(表1)。

表1 骨折分类特征

Table 1 Characteristics of rib fracture		
骨折类型	发生骨折(处)	比例(%)
无错位骨折	1 401	48.60
部分错位骨折	973	33.75
完全错位骨折	509	17.66

### 2.2 肋骨骨折诊断灵敏度和特异度

低年资医师在独立阅片时,诊断肋骨骨折的灵敏度和特异度分别为77.95%、99.53%,对无错位骨折的诊断灵敏度和特异度分别为64.41%、99.54%,部分错位骨折的灵敏度和特异度分别为88.59%、99.99%,完全错位骨折的灵敏度和特异度分别为

94.90%、100.00%；在AI辅助诊断后，诊断肋骨骨折的灵敏度和特异度分别为88.52%、99.85%，对无错位骨折诊断的灵敏度和特异度分别为78.49%、99.86%，部分错位骨折的灵敏度和特异度分别为97.43%、99.99%，完全错位骨折的灵敏度和特异度分别为99.08%、100.00%，不同类型肋骨骨折的诊断灵敏度均显著提升，差异有统计学意义( $P < 0.001$ )，无错位骨折的诊断特异度差异有统计学意义( $P < 0.001$ )；部分错位及完全错位骨折的诊断特异度差异无统计学意义( $P=1.000$ )。

高年资医师在独立阅片时，诊断肋骨骨折的灵敏度和特异度分别为83.96%、99.83%，对无错位骨折的诊断灵敏度和特异度分别为71.38%、99.84%，部分错位骨折的灵敏度和特异度分别为94.76%、99.99%，完全错位骨折的灵敏度和特异度分别为97.97%、100.00%；在AI辅助诊断后，诊断肋骨骨折的灵敏度和特异度分别为90.98%、99.96%，对无错位骨折的诊断灵敏度和特异度分别为82.27%、99.95%，部分错位骨折的灵敏度和特异度分别为98.87%、100.00%，完全错位骨折的灵敏度和特异度分别为99.87%、100.00%，不同类型肋骨骨折的诊断灵敏度均显著提升，差异有统计学意义( $P < 0.001$ )，无错位骨折的诊断特异度差异具有统计学意义( $P < 0.001$ )；部分错位及完全错位骨折的诊断特异度差异无统计学意义( $P=1.000$ )。

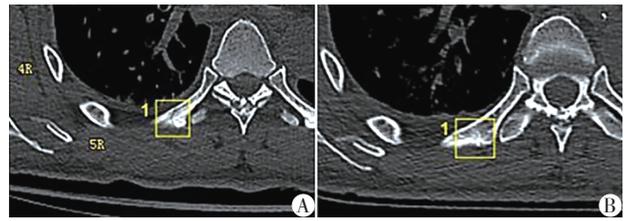
典型CT图像见图3。

### 2.3 ROC曲线及AUC

应用AI辅助诊断后，影像医生对肋骨骨折的检测性能提高，平均AUC从0.902增加到0.948(图4，表2)。每个阅片者的AUC变化范围为0.034~0.080，6名诊断医师AI支持下的AUC均升高， $P$ 值均 $<0.001$ 。

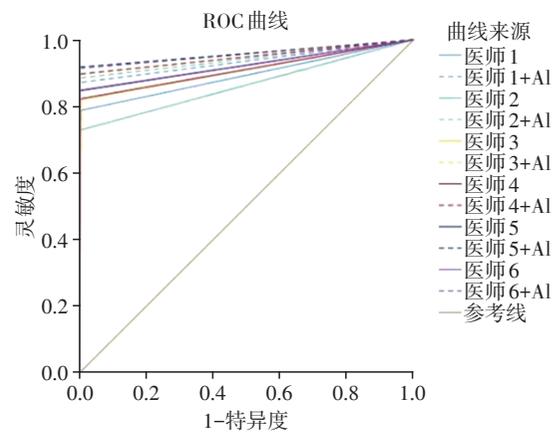
### 2.4 一致性检验

采用Cohen's kappa系数分析不同年资医师对



A:男,58岁,车祸伤1h,右侧第6后肋线状低密度影,为无错位骨折(黄色方框),低年资医师及高年资医师在独立阅片时均漏诊, AI辅助诊断后准确诊断;B:1个月后复查胸部CT示对应的右侧第6后肋骨骨折处骨痂形成。

图3 右侧第6后肋无错位骨折及1个月后复查CT图像  
Figure 3 Undisplaced fracture of the right 6th rib and re-view images after one month



AI辅助后的ROC曲线(虚线)均高于诊断医师独自阅片时的曲线(实线)。

图4 6名诊断医师在独立阅片和AI辅助下的ROC曲线  
Figure 4 ROC of 6 radiologists when diagnosing independently and with the assistance of AI

肋骨骨折诊断的一致性。结果显示在独立阅片时，低年资和高年资医师诊断肋骨骨折的Cohen's kappa系数为 $0.832 \pm 0.004$ ；AI辅助诊断后低年资和高年资医师诊断肋骨骨折的Cohen's kappa系数为 $0.900 \pm 0.003$ 。

### 2.5 阅片时间

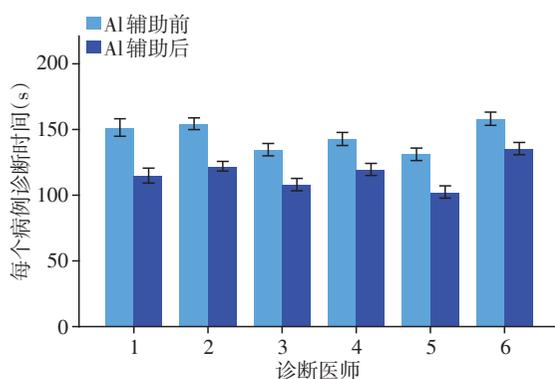
在阅片时间上，AI辅助后诊断医师诊断每个肋

表2 AI辅助前后的AUC

Table 2 AUC before and after AI assistance

诊断医师	工作年资	独立阅片	AI辅助	差值	$P$ 值
1	1	0.892	0.935	0.043	$< 0.001$
2	3	0.863	0.942	0.079	$< 0.001$
3	3	0.907	0.949	0.041	$< 0.001$
4	7	0.911	0.948	0.038	$< 0.001$
5	12	0.923	0.959	0.036	$< 0.001$
6	8	0.923	0.957	0.033	$< 0.001$
平均值	-	$0.902 \pm 0.023$	$0.948 \pm 0.009$	0.046	$< 0.001$

骨折病例的时间 $[(117.28 \pm 53.23) \text{ s}]$ 比AI辅助前 $[(145.71 \pm 58.78) \text{ s}]$ 的时间减少了28.43 s (95CI: 27.73~29.13 s), 差异具有统计学意义,  $t=79.697, P < 0.001$ 。低年资医师(图5, 诊断医师1和2)AI辅助前后每个病例平均阅片时间缩短最为显著, 分别为 $(36.35 \pm 21.70) \text{ s}$  (95%CI: 34.46~38.24 s;  $t=37.756, P < 0.001$ )、 $(32.56 \pm 16.64) \text{ s}$  (95%CI: 31.11~34.01 s;  $t=44.105, P < 0.001$ )。



使用AI辅助诊断肋骨骨折后, 6名放射科医生的诊断时间均减少, 平均缩短时间28.43 s,  $P$ 值均 $< 0.001$ 。

图5 使用AI诊断肋骨骨折前后时间

Figure 5 Time to diagnosis of rib fractures before and after using AI assistance

### 3 讨论

CT扫描显示外伤性肋骨骨折较X线平片能提供更加丰富、准确的影像信息<sup>[5-6]</sup>。而临床上肋骨骨折误诊及漏诊率较高<sup>[7]</sup>。CNN是深度学习(deep learning)的代表算法之一, 在骨骼肌肉放射学中已有多种应用<sup>[11]</sup>, 本研究采用基于CNN的人工智能软件, 对肋骨骨折实现检测。目前已有基于CNN的人工智能软件在提高肺结节及乳腺癌检出能力的相关研究表明, AI辅助后检出能力明显提高<sup>[13-15]</sup>。基于CNN的深度学习技术辅助肋骨骨折诊断的研究多集中于算法和模型<sup>[9-10, 16-17]</sup>。

在诊断效能方面, 本研究结果显示, AI辅助诊断后不同年资的影像医师诊断肋骨骨折的能力均升高。独立阅片时低年资医师的诊断灵敏度及AUC均低于高年资医师, 差异有统计学意义; 而AI辅助诊断后低年资医师与高年资医师的诊断灵敏度及AUC均提升。一致性分析结果显示AI辅助阅片后, 低年资医师与高年资医师之间的Cohen's kappa系数明显高于独立阅片时的系数(0.900 vs. 0.832), 说明AI辅助诊断后可以使不同年资医师在诊断肋骨骨折的一致性上趋于相同。这是因为基

于CNN深度学习的人工智能软件可以对肋骨骨折进行准确标记, 从而降低了诊断经验对影像医师的影响。谭辉等<sup>[18]</sup>使用深度学习的计算机辅助诊断系统诊断急性肋骨骨折的检出效能, 结果显示相比独立阅片, 辅助诊断后的检出效能明显提升, 本研究结果与其一致。

在临床工作中, 完全错位骨折容易诊断, 在本研究中, 无论是低年制医生还是高年资医生, 在没有AI的辅助下, 几乎均能对完全错位的骨折得出准确的诊断; 对于部分错位的骨折, 低年制医生诊断的敏感性稍低于高年制医生诊断的敏感性, 但运用AI辅助诊断后, 低年制医生诊断敏感性显著提高( $P < 0.001$ ), 几乎和高年制医生相同, 达到98%左右; 对于无错位肋骨骨折, 因为骨折处没有明显形态改变, 不同年资的诊断医师均容易误诊或漏诊。在AI辅助下, 不同年资的诊断医师对无错位骨折灵敏度均明显提高( $P < 0.001$ ), 其中低年资医师与高年资医师对无错位骨折的诊断效能提升更加显著(64.41% vs. 78.49%,  $P < 0.001$ ; 71.38% vs. 82.27%,  $P < 0.001$ )。虽然AI辅助后无错位骨折的诊断灵敏度提升显著, 但其诊断灵敏度仍较低。漏诊的骨折常是裂隙状骨折及骨皮质屈曲型骨折, 误诊原因常包括血管沟、呼吸伪影、肋软骨移行处等。

在诊断时间上, 本研究显示, AI辅助后诊断医师的阅片时间显著下降 $[(145.71 \pm 58.78) \text{ s} \text{ vs. } (117.28 \pm 53.23) \text{ s}, P < 0.001]$ , 这是因为AI软件提供了准确的骨折所在层面及位置, 能使诊断医师更快地做出诊断并定位骨折。Zhou等<sup>[10]</sup>建立了肋骨骨折检测模型, 对比独立阅片与AI辅助后诊断时间显示, 阅片时间缩短了约73.9 s, 本研究显示AI辅助后诊断医师的阅片时间缩短约28.43 s, 缩短时间的差异可能和软件选用的算法模型及阅片者的诊断经验有关。

本研究有一定的局限性。首先, 数据集病例为回顾性收集, 病例均为已被诊断的病例, 无法获得临床上漏诊的病例, 存在一定的选择性偏倚; 第二, 本研究是单中心研究, 不是多中心研究, 但样本量较大, 包括各种类型骨折, 研究结果具有一定可靠性。其次, 本研究中肋骨骨折的诊断基于资深放射科医师的诊断, 而不是病理结果, 但此情况也无法避免, 因为临床上很多肋骨骨折的患者无需手术; 最后, 在胸部CT图像上出现憋气不良和严重呼吸伪影的患者被排除在研究之外。然而, 此类患者在日常工作中并不少见, 在以后的进一步研究中, 应该

考虑继续优化模型,把此类患者纳入研究。

综上所述,AI辅助医师诊断肋骨骨折能够提高肋骨骨折检出的灵敏度,并减少诊断时间,使不同年资医师的诊断效能趋于一致。

#### [参考文献]

- [1] 刘云,都定元,胡旭,等.严重胸部创伤病死率的相关因素分析[J].中国医学科学院学报,2013,35(1):74-79
- [2] ZREIK N H, FRANCIS I, RAY A, et al. Blunt chest trauma: bony injury in the thorax [J]. Br J Hosp Med (Lond), 2016, 77(2): 72-77
- [3] SIRMALI M, TÜRÜT H, TOPÇU S, et al. A comprehensive analysis of traumatic rib fractures: morbidity, mortality and management [J]. Eur J Cardiothorac Surg, 2003, 24(1): 133-138
- [4] PEEK J, BEKS R B, HIETBRINK F, et al. Epidemiology and outcome of rib fractures: a nationwide study in the Netherlands [J]. Eur J Trauma Emerg Surg, 2020, 10.1007/s00068-020-01412-2
- [5] CHAPMAN B C, OVERBEY D M, TESFALIDET F, et al. Clinical utility of chest computed tomography in patients with rib fractures CT chest and rib fractures [J]. Arch Trauma Res, 2016, 5(4): e37070
- [6] 郝懿,杨海平,韩洋,等. DR平片、肌骨超声及多层螺旋CT诊断肋骨及肋软骨骨折的临床分析[J].医学影像学杂志,2019,29(4):697-699
- [7] MURPHY C E 4TH, RAJA A S, BAUMANN B M, et al. Rib fracture diagnosis in the panscan era [J]. Ann Emerg Med, 2017, 70(6): 904-909
- [8] 陈飞. 85例肋骨骨折的法医学鉴定分析[J]. 中国校医, 2015, 29(7): 529-530
- [9] ZHOU Q Q, TANG W, WANG J, et al. Automatic detection and classification of rib fractures based on patients' CT images and clinical information via convolutional neural network [J]. Eur Radiol, 2021, 31(6): 3815-3825
- [10] ZHOU Q Q, WANG J, TANG W, et al. Automatic detection and classification of rib fractures on thoracic CT using convolutional neural network: accuracy and feasibility [J]. Korean J Radiol, 2020, 21(7): 869-879
- [11] TALBOT B S, GANGE C P JR, CHATURVEDI A, et al. Traumatic rib injury: patterns, imaging pitfalls, complications, and treatment [J]. Radiographics, 2017, 37(2): 628-651
- [12] 王春杰,袁慧书.卷积神经网络在骨骼肌肉放射学中的研究进展[J].中国医学影像技术,2020,36(9):1375-1378
- [13] RODRÍGUEZ-RUIZ A, KRUPINSKI E, MORDANG J J, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system [J]. Radiology, 2019, 290(2): 305-314
- [14] LIU C, HU S C, WANG C, et al. Automatic detection of pulmonary nodules on CT images with YOLOv3: development and evaluation using simulated and patient data [J]. Quant Imaging Med Surg, 2020, 10(10): 1917-1929
- [15] ONISHI Y, TERAMOTO A, TSUJIMOTO M, et al. Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks [J]. Biomed Res Int, 2019, 2019: 6051939
- [16] WEIKERT T, NOORDTZIJ L A, BREMERICH J, et al. Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography [J]. Korean J Radiol, 2020, 21(7): 891-899
- [17] MENG X H, WU D J, WANG Z, et al. A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance [J]. Skeletal Radiol, 2021, 10.1007/s00256-021-03709-8
- [18] 谭辉,田占雨,潘宁,等.基于深度学习的计算机辅助诊断系统 in 提高急性肋骨骨折诊断效能上的价值 [J]. 临床放射学杂志, 2020, 39(12): 2493-2497

[收稿日期] 2021-11-27

(本文编辑:唐震)