

论 著

人工智能在早期胸部CT肋骨骨折的临床应用价值及局限性分析

李秀涛 谭理连* 邱绍清

文武成 林建真

广州医科大学第二附属医院放射科
(广东广州 510260)

【摘要】目的 探讨人工智能(artificial intelligence, AI)辅助下医师在早期胸部CT肋骨骨折的检测能力、肋骨骨折的分布特点及隐匿性骨折的比例。方法 203例患者接受早期胸部CT和随访CT检查。早期胸部CT进行常规检测法和三种检测方法检测肋骨骨折,以两次CT影像结果作为参照金标准,分析三个年龄段、四个部位、三种骨折类型的检出率,分析漏诊肋骨骨折的特征。结果 共检出781处骨折。常规检测法肋骨骨折的诊断准确度92.3%,三种检测方法中AI+医师的诊断准确度96.0%,AI+医师在年轻组检出率最高(100%),医师+审核对中年组检出率最高(96.1%),AI+医师在肋骨各段具有非常高的检测能力。常规检测法和三种检测方法对移位骨折有较高检出率。肋骨骨折漏诊大部分为隐匿性骨折,且在明确骨折邻近区域。结论 医师在AI辅助下能提高肋骨骨折诊断准确度,漏诊骨折多为隐匿性骨折,常在明确骨折邻近区域。

【关键词】人工智能;早期CT;肋骨;骨折

【中图分类号】R683.1; R814.42

【文献标识码】A

DOI:10.3969/j.issn.1672-5131.2023.10.035

Clinical Application Value and Limitation Analysis of Artificial Intelligence in Early Thoracic CT for Rib Fractures

LI Xiu-tao, TAN Li-Lian*, QIU Shao-qing, WEN Wu-cheng, LIN Jian-zhen.

Department of Radiology, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, Guangdong Province, China

ABSTRACT

Objective To investigate the ability of physicians assisted by Artificial intelligence (AI) to detect rib fractures in early chest CT, the distribution characteristics of rib fractures and the proportion of occult fractures. **Methods** 203 patients underwent both early chest CT and follow-up CT were enrolled. For early chest CT examination, the detection of rib fractures was divided into routine detection method and three detection methods. The detection rates of three age groups, four rib parts and three fracture types were calculated, and the characteristics of missed rib fractures were analyzed, based on the imaging results of the twice CT examinations as the reference standard. **Results** A total of 781 fractures were confirmed. The overall detection rate of rib fractures in the routine detection was 92.3%. The diagnostic accuracy of AI + radiologists was 96.0%. The detection rate of AI + radiologists was the highest in the young group (100%), and the detection rate of physician + reviewer was the highest in the middle-aged group (96.1%). The AI+ radiologists has a very high detection ability in all segments of the rib. Two assessments of detection rate in displaced fracture were higher. Most missed rib fractures are occult fractures in the immediate vicinity of the defined fracture. **Conclusion** Most missed rib fractures are occult on early CT, which is in the vicinity of definite fractures. AI-assisted radiologists can improve the diagnostic accuracy of rib fractures.

Keywords: Artificial Intelligence; The Early CT; Ribs; Fracture

人工智能(artificial intelligence, AI)和计算机视觉的最新进展导致了深度学习技术^[1-2]在医学图像分析和数字医学^[3-5]中的快速发展,为提高放射科医师工作中图像解读和诊断的速度、准确度和质量提供了机会。肋骨骨折占胸部创伤的10%,合并中重度脑损伤的总死亡率为20.2%^[6],及时诊断肋骨骨折在后续临床治疗极具意义。肋骨骨折最常用的诊断方法是计算机断层扫描(computed tomography, CT),胸部CT已经成为诊断肋骨骨折最常用的影像学检测手段^[7],但因主客观因素常存在漏诊。AI自动检测技术为提高肋骨骨折检出率提供可选择机会,有研究显示AI对肋骨骨折的检出率达95%^[8]。

诊断肋骨骨折的参考标准是评价肋骨骨折准确度的关键。肋骨骨折后几天内的早期胸部CT不能清楚显示轻微或隐匿性骨折,2~3周后因骨折端骨质吸收、骨痂形成致骨折征象更加清晰,能准确确定骨折的存在。本研究拟建立一个基于72小时内的早期胸部CT和2周至3个月内的随访CT进行对比的参考标准,旨在准确的探讨肋骨骨折的分布特点、医师在AI辅助下于早期胸部CT检测肋骨骨折的能力以及隐匿性骨折的比例。

1 材料与方法

1.1 研究对象 收集2018年1月至2022年5月因临床怀疑肋骨骨折并在72小时内行早期胸部CT检查,2周后至3个月内行胸部随访CT的203例患者。男101例,女102例,年龄22~91岁,平均年龄54.4±17.9岁。男性中位年龄49岁,女性中位年龄61岁,男女年龄差异明显($P=0.00005425<0.05$)。

纳入标准:年满18岁;临床表现胸痛,怀疑肋骨骨折1-3天内行早期胸部CT检查,2周后3个月内进行随访CT检查,影像资料及临床资料完整能满足影像诊断要求;早期胸部CT检查前无肋骨骨折手术治疗,无恶性肿瘤病史。

1.2 CT设备及参数 使用飞利浦Brilliance 64排128层螺旋CT扫描,采用iDose 4迭代扫描技术。取仰卧位,双臂上举过头,如不能上举者,将双臂置于身体两侧,吸气末屏住呼吸。扫描范围为颈7椎体上缘至腰3椎体下缘。扫描参数:管电压120kV,管电流35~500mA,管球旋转0.42秒/圈,准直为128×0.625mm,重建矩阵512×512,重建层厚1.0mm。

1.3 图像处理和阅片 早期胸部CT对肋骨骨折检测分为常规检测法和三种不同检测方法,所有阅片医师被指示只关注肋骨骨折。常规检测法模拟日常工作状态,由两名放射科医师(分别有7年和5年放射工作经验的初级医师)进行,两名医师采取双盲法在规定10分钟内完成阅片,根据需要调整窗宽、窗位,必要时调用MPR、VR图像。三种不同检测方法涉及两名放射科医师(分别有14年和20年放射诊断工作经验的中级职称医师和高级职称医师),采用三种不同检测方法对早期胸部CT进行肋骨骨折检测。方法A:医师独立阅片,中级职称医师独立调阅横断位图像,必要时调用MPR图像、VR图像。方法B:AI+医师阅片,方法A医师在AI辅助下,依据北京数坤科技(SK-Viewer)AI自动肋骨骨折检测软件对骨折检测情况提示进行全面阅片。方法C:医师+审核,方法A医师阅片后,由高年资放射医师进行审核,审核结果作为此方法的诊断结果。为了避免短期记忆对诊断结果造成影响,三种不同检测方法之间间隔1个月,依据患者IP号进行随机排序。结果

【第一作者】李秀涛,男,副主任医师,主要研究方向:人工智能在医学影像的应用。E-mail: yx080617@163.com

【通讯作者】谭理连,男,主任医师、教授,主要研究方向: CT/MRI诊断。E-mail: liliantan@163.com

均由一名不参与本课题研究的初级医师记录于Excel电子表格，记录内容包括：骨折部位、骨折数量、骨折类型。

为了研究AI在辅助诊断肋骨骨折上的作用，将肋骨骨折分为移位骨折、非移位骨折和陈旧性骨折。移位骨折指骨折断端移位分离2mm及以上者。非移位骨折指有明确的损伤部位，肋骨骨质的外缘或内缘一侧或两侧的骨皮质中断、皱褶、密度局部增高等影像征象^[9]。陈旧性骨折定义为骨折愈合或骨痂形成。每个肋骨骨折位置记录参照图1^[9-10]：将第1肋骨分为肋软骨交界处、前段、后段，第2~10肋骨分为肋软骨交界处、前段、侧段、后段，将第11肋骨分为肋软骨交界处、后段，将第12肋骨定为后段。每例患者共84个肋段，共对17052个肋段进行检测。

1.4 参照标准建立 所有早期胸部CT和随访CT检查图像，由两名放射科诊断医生(分别有22年和16年的经验)系统阅读，使用横断位图像、MPR图像和VR图像并记录所有明显和可疑的肋骨骨折。满足以下三种条件之一者，则确认存在肋骨骨折：(1)有明确的骨折线，骨皮质不连续，或骨折移位；(2)早期胸部CT仅显示疑似骨折征象，随访CT则显示肋骨局部密度增加或骨痂形成；(3)早期胸

部CT未见骨折，随访CT显示肋骨密度局部增加或骨痂形成。

1.5 统计学方法 使用Python编程语言Pandas库和NumPy函数进行统计学处理，符合正太分布计量资料以($\bar{x} \pm s$)表示，反之记为中位数，分类资料采用百分比(%)表示。骨折检出率使用四格表卡方检验， $P < 0.05$ 为差异具有统计学意义。采用Python matplotlib库(2D绘图库)对必要统计学进行数据可视化。

2 结果

2.1 病例基本特征 203例患者，前三位肋骨骨折原因为跌倒、车祸伤、撞击伤。跌倒98例(48.3%)、车祸伤32例(15.8%)、撞击伤30例(14.8%)。经两位放射科医师比较早期胸部CT和随访CT，最终确定631条肋骨781处骨折。发现肋软骨交界处18处骨折(2.3%，18/781)，前段371处骨折(47.5%，371/781)，侧段223处骨折(28.6%，223/781)，后段169处骨折(21.6%，169/781)。移位骨折193处(24.7%，193/781)，非移位骨折511处(65.4%，511/781)，陈旧性骨折77处(9.9%，77/781)。781处骨折分布见图2。

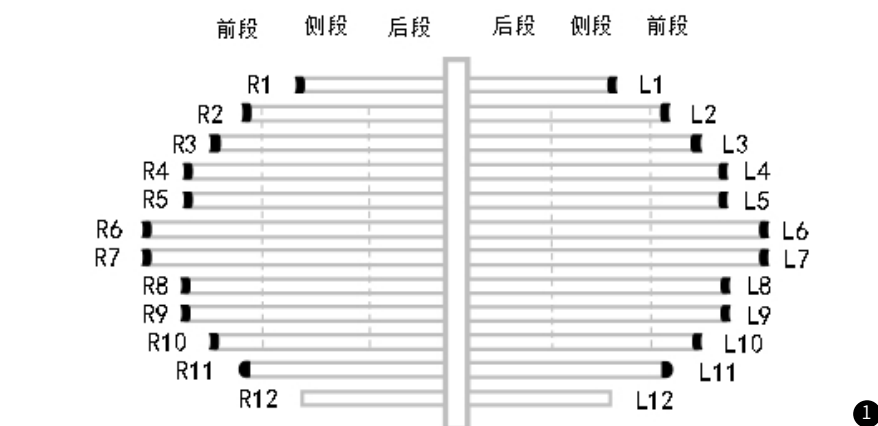


图1 阅片医师参照肋骨分段示意图记录肋骨骨折位置。黑色部分为第1至第11肋软骨交界处，灰色虚线将第2至第10肋骨分为肋骨前段、侧段和后段。

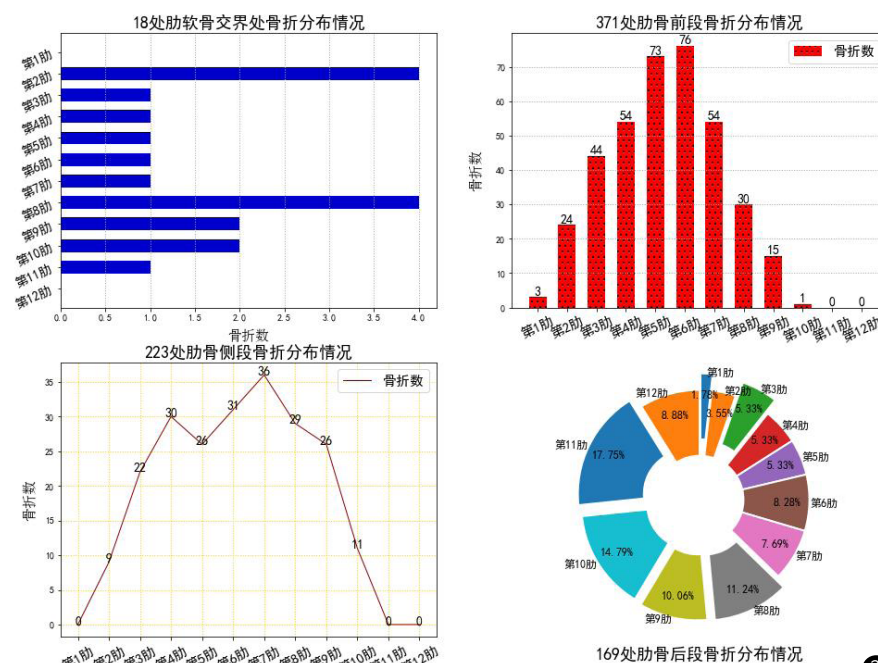


图2 781处骨折分布情况图，肋软骨交界处骨折以第2和第8肋骨最多，肋骨前段骨折以第4~7肋骨居多，侧段以第4~9肋骨居多，后段以第8~11肋骨居多。

2.2 常规检测法和三种不同检测方法的诊断准确性比较 本组常规检测法灵敏度55.5%，特异度93.8%，漏诊率43.4%，准确度92.3%。医师独立阅片灵敏度57.5%，特异度98.8%，漏诊率42.5%，准确度94.7%；AI+医师灵敏度77.6%，特异度99.3%，漏诊率22.4%，准确度96.0%；医师+审核灵敏度62.6%，特异度98.4%，漏诊率37.4%，准确度94.5%。AI+医师的诊断特异度和准确度明显提高，漏诊率明显降低，差异具有统计学意义($P<0.05$)。

2.3 不同年龄组肋骨骨折检出率 本组分为年轻组(≤ 39 岁)48例，中年组(40-59岁)82例和老年组(≥ 60 岁)73例。AI+医师及医师+审核对各年龄组检出率均高于常规检测法和医师独立阅片，医师独立阅片在不同年龄组间肋骨骨折检出率差异无统计学意义($P=0.212>0.05$ ，表1)。

2.4 不同部位肋骨骨折检出率 常规检测法检出率最高的是肋骨侧段(90.1%)，四个肋骨部位之间肋骨骨折检出率差异有统计学意义($P<0.05$ ，表2)。医师独立阅片在侧段检出率(96.0%)高于其它三处；AI+医师在肋骨后段的检出率(95.9%)均高于其它三处，

在肋软骨交界处与前段之间、在肋骨侧段与后段之间检出率无统计学差异($P>0.05$)；医师+审核在肋骨前段的检出率(94.9%)均高于其它三处。常规检测法和三种检测方法在肋骨水平位置组间骨折检出率差异均具统计学意义($P<0.05$)。

2.5 不同骨折类型检出率 常规检测法在移位骨折检出率93.8%(181/193)，非移位骨折检出率83.4%(426/511)，陈旧性骨折检出率83.1%(64/77)，非移位骨折和陈旧性骨折检出率差异无统计学意义($P>0.05$)。三种不同检测方法对肋骨移位骨折检出率分别是95.3%、97.9%、100%，对非移位骨折的检出率分别是83.8%、86.7%、97.5%，对陈旧性骨折检出率分别是41.6%、83.5%、97.4%。医师+审核在三种骨折类型的诊断差异无统计学意义($P=0.081>0.05$ ，表3)。

2.6 肋骨骨折漏诊特点 常规检测法共漏诊340处(43.5%)骨折。在医师独立阅片漏诊332处(42.5%)，AI+医师漏诊175处(22.4%)，医师+审核漏诊292处(37.4%)。与金标准相比，漏诊骨折常位于明确骨折邻近1~4条肋骨(图3-4)。

表1 不同年龄组间肋骨骨折数检出率的比较n(%)

年龄(岁)	常规检测法	三种检测方法			总骨折数
		医师独立阅片	AI+医师	医师+审核	
≤ 39	84(79.2)	93(87.7)	106(100.0)	120(89.6)	106
40-59	275(89.0)	248(80.3)	293(94.8)	297(96.1)	309
≥ 60	312(85.2)	303(82.8)	329(89.9)	349(95.4)	366
P值	0.040	0.212	0.000	0.014	

表2 常规检测法和检测方法中不同部位肋骨骨折的检出率n(%)

骨折位置	常规检测法	三种检测方法		
		医师独立阅片	AI+医师	医师+审核
肋软骨交界处	15(83.3)	22(84.6)	20(90.9)	16(88.9)
肋骨前段	322(86.8)	283(80.6)	334(90.0)	352(94.9)
肋骨侧段	201(90.1)	214(96.0)	212(95.1)	264(86.6)
肋骨后段	133(78.7)	125(74.0)	162(95.9)	134(79.3)
P值	0.012	0.000	0.038	0.000

表3 常规检测法和检测方法在不同类型肋骨骨折的检出率n(%)

骨折类型	常规检测法	三种检测方法		
		医师独立阅片	AI+医师	医师+审核
移位骨折	181(93.8)	184(95.3)	189(97.9)	193(100.0)
非移位骨折	426(83.4)	428(83.8)	443(86.7)	498(97.5)
陈旧性骨折	64(83.1)	32(41.6)	96(83.5)	75(97.4)
P值	0.001	0.000	0.000	0.081

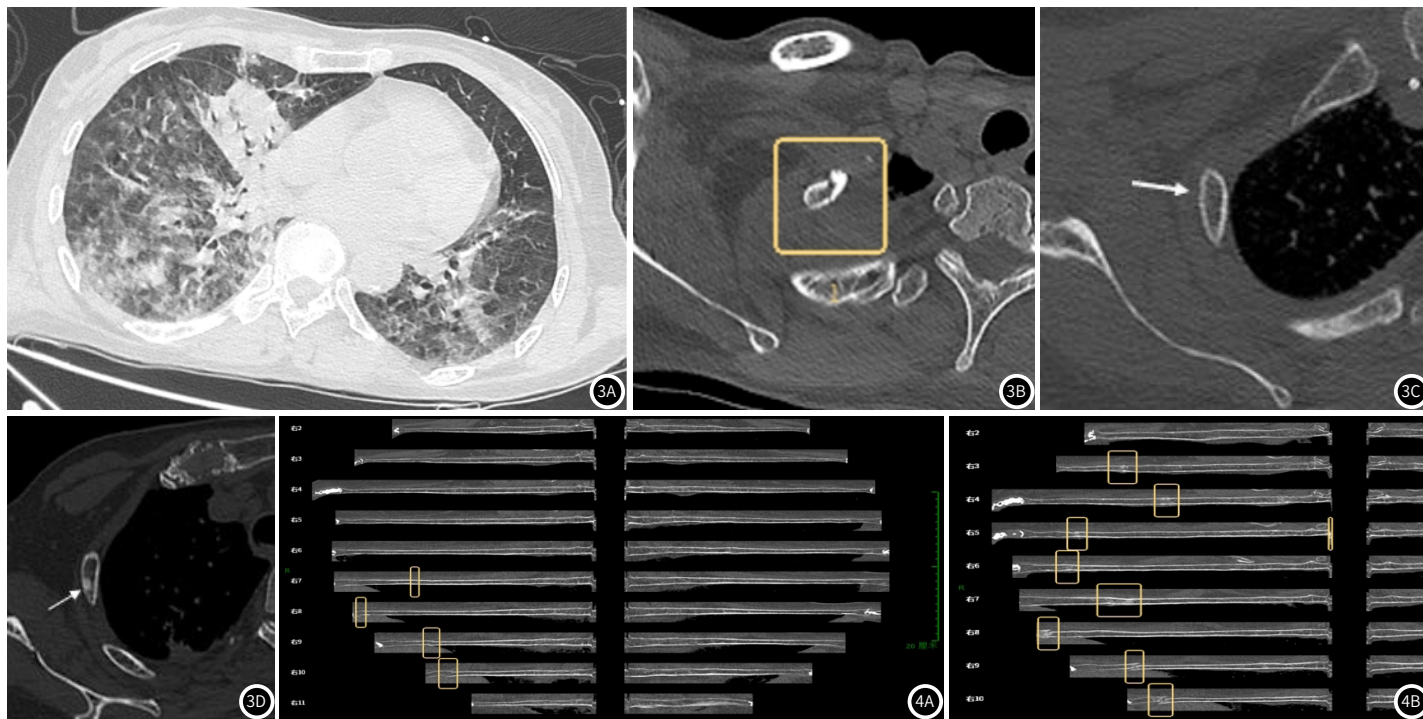


图3A-图3D 患者女，67岁，咳嗽4天，右胸痛1天。图3A 显示双肺炎症；图3B AI检测右侧第一肋前段骨折；图3C 右侧第2肋侧段外侧骨皮质不光整(白箭头)，图3D 第37天复查右侧第2肋骨局部密度增高(白箭头)。

图4A-图4B 患者女，87岁，跌倒致右侧胸痛半小时。图4A AI拉直CPR图，AI检测出右侧4处肋骨骨折；图4B AI拉直CPR图，外伤26天后复查CT，AI检测出右侧9处骨折。

3 讨论

肋骨骨折是评价胸部严重创伤的一个重要指标,有学者报道连枷胸肋骨骨折手术内固定的重要性是非连枷胸患者手术内固定的15倍以上,在专科医院甚至达107倍^[11],CT在肋骨骨折内固定术后能多角度、多方位观察固定情况^[12]。在老年患者中,与保守治疗相比,手术患者死亡率较低(4%)^[13]。所以正确诊断肋骨骨折对临床工作决策及对患者护理起作非常重要的意义^[14]。

以往大多数研究以早期胸部CT结果为参考标准,常因个人的主观因素导致诊断准确度存在一定的误差。肋骨骨折后通常在10天内CT检查较难发现隐匿性骨折^[15],在2~3周后骨折断端骨质开始吸收,骨密度增高、骨痂形成使骨折征象更加明显。因此,本研究建立的肋骨骨折参考标准,以获得医师在AI辅助下诊断肋骨骨折更高的准确度。在本研究中,常规检测法的诊断准确度为92.3%,三种不同检测方法对肋骨骨折诊断准确率达94.5%~96.0%,高于以往研究肋骨骨折的诊断准确度(70%~90%)。医师在AI辅助下能明显提高肋骨骨折的诊断准确度(96.0%),分析原因是AI对肋骨骨折定位准确,使假阳性及假阴性骨折数减少。

为了更好了解不同年龄段肋骨骨折的检出率,将203例患者分为年轻组(≤ 39 岁)、中年组(40~59岁)和老年组(≥ 60 岁)。研究结果显示,常规检测法在中年组的检出率最高(89.0%),而在年轻组检出率最低(79.2%),在三组不同年龄的检出率差异具有统计学意义($P < 0.05$)。AI+医师在年轻组的检出率最高(100%),在老年组的检出率为89.9%。虽然医师在AI辅助下在老年组中近10%漏检,但已经达到满意结果。老年组骨质疏松导致骨折显示不佳,AI在老年组骨折检出困难,这是在老年组检出率有所下降的原因;另外,医师在AI辅助下也可能过度参照AI检测结果,忽视细微肋骨骨折征象,从而漏诊部分非移位骨折,从医师+审核的结果也可以证明此观点。

在常规检测法中检出率非常高的是肋骨前段和后段,检出率为86.9%和89.7%;在肋软骨交界处检出率为83.3%,明显高于文献的60%~65%^[10];在肋骨后段检出率79.2%,可能是肋骨后段特殊部位骨折易漏诊(如肋骨头、肋骨颈)。在三种检测方法中,虽然医师独立阅片和医师+审核具有较高检出率,但只在肋骨侧段和前段较为显著;AI+医师对不同位置骨折检出率均明显高于常规检测法和医师独立阅片和医师+审核,且在肋骨四个部位间检出率差异没有医师独立阅片和AI+医师显著。结果表明,医师在AI辅助下在四个肋骨部位具有非常高的骨折检出率,在放射科诊断肋骨骨折中临床意义重大。

在常规检测法和三种不同检测方法中,早期胸部CT移位骨折的检出率达到93.8%和100%(医师+审核),AI+医师对移位骨折的检出率97.9%,已经达到非常高的检出率;非移位骨折的检出率分别是83.4%和97.5%(医师+审核),医师在AI辅助下对非移位骨折的检出率约86.7%,结果表明医师在AI辅助下对非移位骨折的检测能力存在不足,可能与医师对轻微骨折的辨别缺乏足够的认识,且存在过度依赖AI检测结果的情况。在陈旧性骨折检出率83.1%和97.4%(医师+审核),医师独立阅片则在陈旧性骨折的检出率仅41.6%,存在近58%的漏诊。医师在AI辅助下陈旧性骨折存在较多的假阳性,研究表明医师在诊断陈旧性骨折时存在更多依赖AI而造成误判情况,应该在工作中尽可能避免。本研究结果显示,AI对陈旧性骨折的检测效能较移位骨折和非移位骨折差,与既往研究观点一致^[16]。研究还发现,常规检测法和三种不同检测方法对移位骨折的检出率均高于非移位骨折和陈旧性骨折,且在移位骨折检测稳定性较高。如果在影像诊断工作中,能准确检出移位骨折,对临床医师治疗方案的选择更为重要^[10]。

在常规检测法中,203例患者漏诊340处骨折(43.5%)。在三种不同检测方法中,AI+医师漏诊率最低(22.4%),明显低于常规检测法和另外两种检测方法,分析主要原因是AI在诊断过程中对肋骨骨折定位准确。漏诊骨折主要表现为轻微骨折、骨皮质欠光滑、局部解剖形态改变等,最常见者骨折线模糊,骨皮质欠光整(61.8%,230/340),与对侧相比骨皮质光滑度减低,甚至出现局部解剖形态改变,这些征象的观察有助于提高细微骨折的检出率。研究还发现早期胸部CT漏诊骨折常位于明确骨折相邻1~4条

肋骨,发生率达61.8%。Liu,C等提示肋骨骨折漏诊常位于明确骨折相邻的1~3条肋骨^[10],本组则在明确骨折相邻1~4条肋骨可能存在骨折漏诊,分析原因可能与跌倒患者受力面积较大,增加轻微骨折的可能。常规检测法和医师独立阅片和医师+审核存在较多骨折漏诊,部分原因是肋骨水平定位不够准确导致;医师在AI辅助下诊断骨折定位准确,从而减少了漏诊。因此,医师在AI辅助下细致阅片是有效减少肋骨骨折漏诊的有效办法。

本研究是回顾性研究,病例的选择可能存在偏倚,研究结果的可靠性需进一步扩大样本量。另外,本研究随访CT的时间是2周~3个月,这当中部分骨折可能在随访CT图像中显示不明显,可能低估骨折的总数。

综上所述,受CT技术主客观上的限制,肋骨骨折存在一定的漏诊,主要是一些细微或隐匿性骨折。AI技术能提高肋骨各个位置骨折的检出率,医师在AI技术辅助下能尽最大可能减少肋骨各个位置骨折的漏诊,提高肋骨骨折诊断准确性。

参考文献

- [1] Blum A, Gillet R, Urbaneja A, et al. Automatic detection of rib fractures: Are we there yet[J]. EBioMedicine, 2021, 63: 103158.
- [2] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [3] 谭辉, 田占雨, 潘宁, 等. 基于深度学习的计算机辅助诊断系统在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.
- [4] 吕文晖, 夏菲, 周长圣, 等. 深度学习胸部CT辅助诊断系统在急诊创伤人群中的应用[J]. 中华医学杂志, 2021, 101(7): 481-486.
- [5] Hongbiao S, Shaochun X, Xiang W, et al. Comparison and verification of two deep learning models for the detection of chest CT rib fractures[J]. Acta Radiol, 2022: 2841851221083519.
- [6] Marini C P, Petrone P, Soto-Sánchez A, et al. Predictors of mortality in patients with rib fractures[J]. Eur J Trauma Emerg Surg, 2021, 47(5): 1527-1534.
- [7] 李刚, 冯延冰. 多层螺旋CT三维技术与X线数字断层融合技术在肋骨骨折诊断中的价值比较[J]. 中国CT和MRI杂志, 2021, 19(07): 167-169.
- [8] Zhang J, Li Z, Yan S, et al. An Algorithm for Automatic Rib Fracture Recognition Combined with nnU-Net and DenseNet[J]. Evid Based Complement Alternat Med, 2022, 2022: 5841451.
- [9] Zhang B, Jia C, Wu R, et al. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation[J]. Br J Radiol, 2021, 94(1118): 20200870.
- [10] Liu C, Chen Z, Xu J, et al. Diagnostic value and limitations of CT in detecting rib fractures and analysis of missed rib fractures: a study based on early CT and follow-up CT as the reference standard[J]. Clin Radiol, 2022, 77(4): 283-290.
- [11] Ingoo H M, Eardley W, Mcdaid C, et al. Epidemiology of adult rib fracture and factors associated with surgical fixation: Analysis of a chest wall injury dataset from England and Wales[J]. Injury, 2020, 51(2): 218-223.
- [12] 全志永, 孟庆凯, 郭艳鸽. MSCT对外伤性肋骨骨折的诊断及固定术后复查的应用价值分析[J]. 中国CT和MRI杂志, 2022, 20(02): 175-177.
- [13] Hoepelman R J, Beeres F J P, Heng M, et al. Rib fractures in the elderly population: a systematic review[J]. Arch Orthop Trauma Surg, 2022, 143: 887-893.
- [14] Weikert T, Noordtzi 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.
- [15] 李立, 钱伟军, 于敏, 等. 伤后不同多层螺旋CT检查时间诊断隐匿性肋骨骨折的效果分析[J]. 医学影像学杂志, 2020, 30(07): 1284-1287.
- [16] 周清清, 王佳硕, 唐雯, 等. 基于卷积神经网络成人肋骨骨折CT自动检测和分类的应用研究[J]. 影像诊断与介入放射学, 2020, 029(1): 27-31.

(收稿日期: 2023-01-01)

(校对编辑: 姚丽娜)