# 影像技术学・

# 利用深度学习构建自动分类模型以辅助髌骨轴位 X 线片的图像 质控

额·图娅,王岑,黄嘉豪,张耀峰,张晓东,王霄英

【摘要】目的:利用深度学习方法训练髌骨轴位 X 线片图像质量控制的自动分类模型。方法:回顾性收集髌骨轴位 X 线片,由两位专家将髌骨轴位 X 线片分为不同数据组以训练模型,分别为:术后/ 非术后共 175 例(术后 96 例,非术后 79 例),侧别共 735 例(左侧 419 例,右侧 316 例),图像质量共 453 例(图像质量不合格 246 例,图像质量合格 207 例)。上述每组数据均按 8:1:1 的比例随机分为训练集、 调优集和测试集,即:术后/非术后为 136 例、21 例、18 例,侧别 586 例、75 例、74 例,图像质量为 362 例、 46 例、45 例。训练 HRNet 模型对上述三组图像进行自动分类,应用混淆矩阵评价模型分类预测效能, 以符合率为评价指标。结果:测试集中,三组图像分类模型的预测符合率依次为:术后/非术后 94.4% (17/18)、侧别 98.6%(73/74)、图像质量 91.1%(41/45)。结论:基于深度学习训练的分类模型对髌骨轴 位 X 线片进行图像质量控制效能良好,有利于工作流程的优化及后续对接 AI 诊断模型。

【关键词】 髌骨;轴位 X 线; 深度学习; 质量控制; 结构化报告

【中图分类号】R814.41;R684 【文献标识码】A 【文章编号】1000-0313(2022)07-0884-05 DOI:10.13609/j.cnki.1000-0313.2022.07.016 开放科学(资源服务)标识码(OSID):



**[Abstract]** Objective: To develop classification models based on deep learning to automatically classify the axial patella X-ray images in order to assist in image quality control. Methods: Axial patellar X-ray images were retrospectively collected. All images were reviewed by two radiologists and classified into three groups of binary classification, which included postoperative and nonoperative group (n = 175, postoperative: n = 96, nonoperative: n = 79), laterality group (n = 735, left: n = 419, right: n = 316), image quality group (n = 453, poor image quality: n = 246, good image quality: n = 207). Each group data was randomly divided as train set, validation set and test set with the ration of 8:1:1. The classification models were trained using HRNet network, and the effectiveness of the models were e-valuated using confusion matrix. Results: The accuracy of classification models of the axial patella X-ray images in the test set were as following: postoperative and nonoperative group 94.4% (17/18), laterality group 98.6% (73/74), and image quality group 91.1% (41/45). Conclusion: The models of automatic classification of axial patella X-ray images have a good performance, which was beneficial to the quality control and the subsequent implementation of AI diagnosis models.

**(Key words)** Patella; Axial radiography; Deep learning; Quality control; Structured report

髌骨轴位 X 线片在髌股关节不稳、髌股关节骨关 节炎、髌骨纵行骨折等疾病诊断与髌股关节术后疗效 评估中具有重要地位。由于髌股关节的解剖和生物力 学特点,膝关节屈膝 30°时股四头肌松弛,髌股关节相 对处于不稳定状态,关节间隙较大,髌骨正好进入股骨 滑车,此时拍摄髌骨轴位 X 线片有助于更好地显示髌 股关节情况,是检查髌骨不稳、髌骨骨折等疾病的临床 首选影像诊断技术<sup>[1]</sup>。

近年来,人工智能(artificial intelligence,AI)在骨 关节系统方面有较多研究,结果显示 AI 可获取半定 量、定量数据或定性数据提供自动化及标准化的诊断

作者单位:100034 北京,北京大学第一医院医学影像科 (额·图娅、张晓东、王霄英);100045 北京,北京核工业医院 (王岑);100011 北京,北京赛迈特锐医疗科技有限公司(黄嘉 豪、张耀峰)

作者简介:额·图娅(1992-),女,新疆博乐人,博士研究 生,主要从事骨骼肌肉及腹部影像诊断的人工智能临床应用研 究。

通讯作者:王霄英,E-mail:wangxiaoying@bjmu.edu.cn

信息<sup>[2]</sup>,如对膝关节骨关节炎进展具有良好的预测效 能<sup>[3]</sup>,辅助医师对膝关节骨关节炎进行诊断并分级<sup>[4]</sup> 等。然而图像质量对 AI 诊断或预测模型效能具有较 大影响,尤其是在研究初始阶段数据量相对较少时更 为突出,因此完成图像性质识别工作是 AI 诊断或预 测模型广泛应用于临床的前提。

本研究旨在利用深度学习方法训练髌骨轴位 X 线片图像质量控制的自动分类模型,挑选出合格图像 输入到后续的 AI 诊断髌股关节骨关节炎及髌股关节 不稳模型中,以保证 AI 诊断模型的准确性。

#### 材料与方法

本研究获得伦理审查委员会的批准(批件号: [2019(168)]),按照本院人工智能 AI 模型训练规范 执行研究方案。

1.用例定义

根据本院 AI 训练管理方法,首先定义研发髌骨 轴位 X 线片图像性质分类模型的用户样例(use case)。包括:AI 模型的名称(identity document,ID)、 临床问题、场景描述、模型在实际工作中的调用流程、 模型输入输出数据结构等。模型输出结果自动输入到 结构化报告中,激活相应的选项,包括:术后/非术后、 右侧/左侧、图像质量合格/不合格。

依据实际临床问题,此图像质控分类模型包括术 后/非术后(图1)、侧别(图2)以及图像质量(图3)。 侧别模型用以区分右侧及左侧,评判标准为观察内、外 侧髌股关节间隙处于图像中的位置,即在右侧图像中 内侧髌股关节间隙位于图像的左侧,而在左侧图像中 内侧髌股关节间隙位于图像的右侧。图像质量模型以 图像对比度及图像显示度进行质量合格与不合格的评 估。图像对比度关注拍摄曝光情况,图像对比度差为 曝光过度或曝光不足,图像对比度良好则为曝光度适 宜。图像显示度关注髌骨与股骨的显示情况,具体包 括髌骨后缘关节面、股骨前缘关节面以及髌骨与股骨 重合度。若髌骨与股骨有明显重叠、髌骨后缘有多于 两条重叠线或髌骨与股骨关节面模糊则为显示度差, 反之若髌骨与股骨无重叠、髌骨后缘无明显双边线、髌 股关节间隙显示清晰、髌骨与股骨关节面线条清晰则 为显示度良好<sup>[5]</sup>。

2. 研究队列建立

本项研究为回顾性研究,收集本院 PACS 中符合 相应要求的髌骨轴位 X 线片,由两位专家依据相应标 准阅片并分组(表 1),当两名专家意见不一致时则进 行协商达成一致。术后/非术后共 175 例(男 76 例,女 99 例,年龄 17~88 岁,平均 65.50±10.47 岁),侧别共 735 例(男 270 例,女 465 例,年龄 10~91 岁,平均 54.40±16.56 岁),图像质量共 453 例(男 171 例,女 282 例,年龄 21~87 岁,平均 58.40±13.87 岁)。三个 模型的数据有共用。本单位 X 线机为 GE Discovery XR650 和 Carestream DRX-Evolution,髌骨轴位 X 线 片采取 Laurin 坐位法拍摄,患者被检测膝部屈膝 30°, 使中心线对准髌骨下缘由足侧向头侧进行投照。

3. 数据处理与模型训练

模型训练硬件环境为 GPU NVIDIA Tesla P100 16G,软件包括 Python3.6、Pytorch 0.4.1、Opencv、 Numpy、SimpleITK 等。

训练过程分四步。第一步,训练术后分类模型,深 度学习网络为 High Resolution Network(HRNet)<sup>[6]</sup>, 输入为髌骨轴位 X 线片,输出数据为"术后、非术后" 的分类预测结果。第二步,对预测结果为"非术后"



图 1 a)非术后髌骨轴位 X 线片; b)术后髌骨轴位 X 线片。 图 2 a)右侧髌骨轴位 X 线片; b)左侧髌骨轴 位 X 线片。 图 3 图像质量分类模型举例。a、b、c 为图像质量不合格的髌骨轴位 X 线片。a)曝光不足,图 像对比度差; b)曝光过度,图像对比度差; c)图像显示度差,即髌骨与股骨有明显的重叠、髌骨与股骨关节面 模糊; d)图像质量合格的髌骨轴位 X 线片,即曝光度适宜,且髌骨后缘无双边线、髌股关节间隙显示清晰、髌 骨与股骨关节面清晰。

表1 髌骨轴位X线图像分类模型情况

模型	入组时间	筛选方法	筛选例数	排除标准和排除例数	最终入组例数
术后/非术后	2019 年 11 月 14 日 -2020 年 9 月 8 日	报告中提示为:术后,按 1:1匹配入组非术后数据	253	图像导出失败(n=78)	175(术后 96 例,非术后 79 例)
侧别	2017年1月3日- 2020年7月6日	非术后病例	1489	术后(n=406),骨折或二 分髌骨(n=54),导出图 像失败(n=294)	735(左侧 419 例,右侧 316 例)
图像质量	2017年2月3日— 2020年12月30日	非术后病例,再次读片, 挑选出图像质量不合格 者,按1:1匹配图像质量 合格数据	1754	术后或骨折(n=540)	453(不合格 246 例,合格 207 例)

者,训练分割模型,深度学习网络为 U-Shaped Fully Convolutional Network (U-Net)<sup>[7]</sup>,输入为髌骨轴位 X线片,输出为髌骨和股骨区域的标签。第三步,训练 侧别、图像质量模型,输入是以髌骨和股骨区域的标签 为掩膜(mask)裁切后的图像,输出是"左侧、右侧"和 "图像质量合格、不合格"的分类预测结果。第四步,所 有分类模型预测的结果自动填写入结构化报告中 (图 4)。

全部模型训练过程包括:图像预处理、分数据集、 图像扩增。图像预处理方法包括:图像大小设置、图像 裁剪;图像扩增方法包括:随机噪声、平移、旋转、缩放 等。数据按8:1:1随机分为训练集(train set)、调优集 (validation set)、测试集(test set)。以 Adam 作为梯 度下降优化器。

表 2 HRNet 模型自动分类不同类别髌骨 轴位 X 线片在测试集中的结果

真实值	预注	预测值		
术后模型	术后	非术后		
术后	9	1	10	
非术后	0	8	8	
测试集合计	9	9	18	
侧别模型	左侧	右侧		
左侧	42	0	42	
右侧	1	31	32	
测试集合计	43	31	74	
图像质量模型	合格	不合格		
合格	20	1	21	
不合格	3	21	24	
测试集合计	23	22	45	

测试集中,分别有 3 例、1 例右侧髌骨轴位图像被预测 为左侧(图 5a)。在图像质量分类模型中,在训练集、



4. 模型评价

以两位专家阅片结果 为金标准,应用混淆矩阵 (confusion matrix)评价分 类模型效能,以符合率(accuracy)为评价指标<sup>[8]</sup>。

# 结果

应用 HRNet 建立的髌 骨轴位 X 线片图像质控的 自动分类模型结果如表 2 所示。在术后/非术后分类 模型中,在训练集及调优集 中符合率均为 99.3%(135/ 136)、100%(21/21),在测 试集中符合率为 94.4% (17/18),有 1 例术后的髌 骨轴位图像被预测为非术 后。在侧别分类模型中,在 训练集、调优集及测试集中 符合率分别为 99.5%(583/ 586)、100%(75/75)、98.6% (73/74);在训练集及

图4 模型训练过程。

调优集及测试集中符合率分别为 98.6%(357/362)、 93.5%(43/46)、91.1%(41/45),在上述三个数据集 中,依次有 4、3、3 例图像质量不合格的被预测图像质 量合格,1、0、1 例图像质量合格的被预测图像质量不 合格(图 5b、c)。

## 讨 论

目前 AI 在全身各系统影像诊断工作中得到了快 速发展<sup>[2-4,9-11]</sup>,但图像质量仍为 AI 诊断模型效能的重 要影响因素之一<sup>[12]</sup>。因此本研究着重利用 AI 技术识 别不同性质的髌骨轴位 X 线图像,挑选图像质量合格 的数据用于提高 AI 诊断模型效能。本研究结果显 示,HRNet 在髌骨轴位 X 线片不同图像性质甄别上 具有良好效能,测试集中预测符合率达到 91.1%~98. 6%,与其他图像序列甄别的研究结果相似<sup>[13-15]</sup>。由此 可见,此模型直接对接后续 AI 诊断模型具有可行性。

膝关节假体置换术后的影像评价重点与骨关节炎 的评价重点不完全一致。假体置换术后的影像评价要 点是:假体位置及力线对位、关节线高度、髌骨位置、 骨-假体交界面形态、有无骨溶解等。而骨关节炎的评 价重点则是观察髌股关节间隙狭窄、髌骨周缘骨赘形 成和软骨下骨硬化情况,明确是否存在髌骨倾斜、髌骨 脱位或半脱位和滑车发育不良等情况<sup>[11]</sup>。本单位髌骨 轴位X线检查患者中,髌股关节术后者和骨关节炎者 均很多,模型在临床应用的场景是检查后图像由 AI 判断是否为术后,如为术后则自动调用术后报告模板, 如非术后则自动调用骨关节炎和关节不稳报告模板。 本研究构建的术后与非术后模型在测试集中符合率为 94.4%(17/18),因此有望在科室范围内开展应用,从 而辅助影像医师快速分诊。

在实际临床工作中,可能受多种因素影响而出现 影像检查的左/右侧别与申请登记侧别不一致的情况。 本研究提出的侧别分类模型在测试集中的符合率为 98.6%(73/74),在临床应用的场景是检查后图像由 AI判断图像的侧别,如与 RIS 中登记的侧别一致,则 自动执行后续诊断流程,如不一致可提醒医生给予关注。分析本研究中模型误断的情况,首先是误断率很低,其次误断的图像均有其特殊性,主要表现为有明显内侧不稳合并内侧关节间隙狭窄且间隙窄于外侧(图5a),从而使得模型将右侧误判为左侧。这种情况下,即使是有经验的医生也不易仅根据图像来判断侧别,需要结合申请单、双侧对照以及其他体位图像来判断。因此,对于侧别分类模型的进一步迭代方向是增加更多有明显关节不稳和骨关节炎的数据,以提高模型在这种特定情况下的符合率。

图像曝光欠佳影响诊断结果,如曝光不足会影响 髌骨缘是否有骨赘及髌股关节间隙显示,而曝光过度 则会影响髌股关节细节结构显示,这类情况均会影响 AI 诊断模型特征提取从而降低符合率。髌骨轴位 X 线片为重叠影像,各组织结构重叠区域小、显示清晰是 诊断的基本要求。本单位研发髌股关节骨关节炎模型 时发现图像曝光度不良、显示度欠佳是导致诊断或分 级误判的主要原因:在髌股关节不稳方面,显示度不佳 不仅不适宜后续诸多评价髌股关节不稳指标的测量 (如股沟角、适合角等),也会导致模型学习寻找关节解 剖关键点时误判。本研究基于此类问题研发了图像质 量甄别的模型,模型在临床应用的场景是先用图像质 量甄别模型筛洗图像,将筛洗出的合格图像输入后续 诊断髌股关节炎及髌股关节不稳模型,而筛选出的不 合格图像则由人工处理。将来还应研发提高图像质量 的模型,对筛选出的不合格图像进行校正使其达到合 格图像,可进一步用于诊断。

本研究的局限性主要在于:①基于小样本对不同 图像性质进行甄别,后期需进一步扩充数据,整合模 型,提高效能,简化流程。只有在真正的临床工作场景 中进行研究,才能客观、可靠地评价 AI 的应用价 值<sup>[16]</sup>。②本研究以医生的主观判断作为图像对比度 和显示度的参考标准,当图像特征处于中间状态时判 断有一定困难,未来应尽可能将参考标准转换为可定 量的指标,以提高数据标注的准确性和一致性。



图5 自动分类模型预测错误的髌骨轴位X线片。a)侧别分类模型在训练集中将右侧预测为左侧;b)图像质 量分类模型在测试集中将图像质量不合格预测为图像质量合格;c)图像质量分类模型在测试集中将图像质 量合格预测为图像质量不合格。 总之,本研究基于 AI 技术训练了能够自动区分 髌骨轴位 X 线图像性质的分类模型,有利于工作流程 的优化,为后续对接疾病诊断 AI 模型奠定了坚实基 础。

### 参考文献:

- [1] 国家老年疾病临床医学研究中心(湘雅医院),中华医学会骨科学 分会关节外科学组.中国髌股关节骨关节炎诊疗指南(2020 年版)
  [J].中华骨科杂志,2020,40(18):1227-1234.
- [2] 钟京谕,姚伟武.人工智能在骨关节炎影像诊断中的研究现状与进 展[J].中华放射学杂志,2019,53(9):790-793.
- [3] Leung K, Zhang B, Tan J, et al. Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: data from the osteoarthritis initiative[J].Radiology.2020.296(3):584-593.
- [4] Norman B, Pedoia V, Noworolski A, et al. Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs[J]. J Digit Imaging, 2019, 32(3): 471-477.
- [5] 于兹喜,崔崔.髌股关节切线位新摄影术的研究[J].中华放射学杂 志,2001,35(8):613-616.
- [6] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation [C/OL]. arXiv e-prints. [Accessed 25 Feb 2019].https://arxiv.org/abs/1902.09212.
- [7] Ronneberger O, Fischer P, Brox T. U-Net; convolutional networks for biomedical image segmentation [C/OL]. arXiv e-prints. [Ac-

cessed 18 May 2019].https://arxiv.org/abs/1505.04597v1.

- [8] Tharwat A. Classification assessment methods [J]. Appl Comput Inform, 2021, 17(1):168-192.
- [9] Liu F,Zhou Z,Samsonov A, et al.Deep learning approach for evaluating knee MR images:achieving high diagnostic performance for cartilage lesion detection[J].Radiology,2018,289(1):160-169.
- [10] 朱丽娜,韩超,刘想,等.U-Net 深度学习网络对 MRI 表观扩散系 数图中前列腺癌自动分割及定位的可行性研究[J].中华放射学 杂志,2020,54(10):974-979.
- [11] 崔应谱,孙兆男,刘想,等.用深度学习和阈值算法自动检出 CT 平扫图像中肾结石的可行性研究[J].中华放射学杂志,2020,54 (9):869-873.
- [12] 张晓东,孙兆男,任昕,等.基于深度残差网络研发辅助诊断软件 用于X线胸片分类诊断[J].放射学实践,2019,34(9):952-957.
- [13] 方俊华,Qiubai L,余成新,等.人工智能深度学习对前列腺多序 列 MR 图像分类的可行性研究[J].中华放射学杂志,2019,53 (10):839-843.
- [14] Kim TK, Yi PH, Wei J, et al. Deep learning method for automated classification of anteroposterior and posteroanterior chest radiographs[J].J Digit Imaging, 2019, 32(6):925-930.
- [15] 孙兆男,崔应谱,刘想,等.利用深度学习实现腹盆部 CT 图像范 围及期相分类:临床验证研究[J].2021,36(4):551-554.
- [16] 朱丽娜,高歌,刘义,等.CAD 整合人前列腺多参数 MRI 结构化 报告:低经验读片者诊断效能研究[J].2020,35(10):1282-1287. (收稿日期:2021-06-28 修回日期:2021-08-31)