

doi: 10.3969/j.issn.1674-1242.2026.04.052

3D 卷积神经网络模型辅助诊断膝关节多韧带损伤的临床价值

郑多多^{1,2}, 陈界荣^{1,2}

(1. 蚌埠医科大学第二附属医院 影像科, 安徽蚌埠 233000;

2. 联勤保障部队第九〇三医院 健康医学科, 浙江杭州 310000)

【摘要】目的 探讨 3D 卷积神经网络(convolutional neural network, CNN)模型辅助磁共振成像(magnetic resonance imaging, MRI)诊断膝关节多韧带损伤(multiligament knee injuries, MLKIs)的临床价值。**方法** 回顾性纳入 2022 年 1 月至 2025 年 12 月蚌埠医学院第二附属医院影像科 100 例经关节镜证实的 MLKIs 患者,按 7:3 分为训练集(70 例)与测试集(30 例)。基于 3D CNN 架构,采用 3D U-Net 模型完成膝关节 MRI 数据的自动分割与 MLKIs 辅助诊断。比较模型、低年资医师与高年资医师对 4 条主要韧带损伤的诊断效能,评估模型辅助对高年资医师阅片时间、诊断一致性及美国航空航天局任务负荷指数量表(National Aeronautics and Space Administration-Task Load Index, NASA-TLX)评分的影响,并分析误判病例图像质量。**结果** 模型敏感度(92.5%)、特异度(89.7%)和准确率(91.5%)均显著优于低年资医师(81.3%、78.2%、80.0%)(均 $P < 0.01$);曲线下面积(area under the curve, AUC)(0.94)与高年资医师(0.96)比较差异无统计学意义($P > 0.05$)。4 条韧带诊断准确率差异无统计学意义($\chi^2 = 2.15, P = 0.54$)。模型辅助后,高年资医师阅片时间由 (12.5 ± 3.2) min 降至 (7.8 ± 2.1) min,诊断一致性由 70.8% 升至 91.7%,NASA-TLX 评分由 42.5 ± 10.2 降至 28.3 ± 8.5 (均 $P < 0.05$)。假阳性主因水肿信号或部分容积效应(46.67%),假阴性主因微小撕裂(70.00%);误判病例信噪比(signal-to-noise ratio, SNR)、对比噪声比(contrast-to-noise ratio, CNR)均显著低于正确诊断病例(均 $P < 0.05$)。**结论** 基于 3D CNN 架构的 3D U-Net 模型对 MLKIs 的诊断效能与高年资医师相当,可显著缩短阅片时间、降低认知负荷并提升诊断一致性。

【关键词】 磁共振成像; 膝关节多韧带损伤; 卷积神经网络; 辅助诊断**【中图分类号】** R445.2**【文献标志码】** A

文章编号: 1674-1242 (2026) 04-0242-04

The clinical value of the 3D convolutional neural network model in assisting diagnosis of multiligament knee injuries

ZHENG Duoduo^{1,2}, CHEN Jierong^{1,2}

(1. Department of Imaging, The Second Affiliated Hospital of Bengbu Medical University, Bengbu 233000, Anhui, China;

2. Department of Health Medicine, 903 Hospital of Joint Logistics Support Force, Hangzhou 31000, Zhejiang, China)

【Abstract】Objective To explore the clinical value of the 3D convolutional neural network (CNN) model in assisting magnetic resonance imaging (MRI) for the diagnosis of multiligament knee injuries (MLKIs). **Methods** A total of 100 patients with MLKIs who were confirmed by arthroscopy at the Department of Imaging of the Second Affiliated Hospital of Bengbu Medical University from January 2022 to December 2025 were retrospectively enrolled and divided into a training set (70 cases) and a test set (30 cases) at a 7 vs. 3 ratio. Based on the 3D CNN architecture, the 3D U-Net model was used to achieve automatic segmentation of knee joint MRI data and assist in the diagnosis of MLKIs. The diagnostic efficiency of the model, junior physicians, and senior physicians for injuries of the four major ligaments was compared. The impact of model assistance on senior physicians' image interpretation time, diagnostic consistency, and National Aeronautics and Space Administration-Task Load Index (NASA-TLX) scores was evaluated. Image quality of misclassified cases were also analyzed. **Results** The model's sensitivity (92.5%), specificity (89.7%), and accuracy (91.5%) were significantly superior to those of junior physicians (81.3%, 78.2%, 80.0%) (all $P < 0.01$), and its AUC (0.94) showed no statistically significant difference from that of senior physicians (0.96) ($P > 0.05$). There was no statistically significant difference in the diagnostic accuracy of

收稿日期: 2026-02-18。

作者简介: 郑多多, 本科, 初级技师, 研究方向: MRI 与关节重建。E-mail: 3440769492@qq.com。

the 4 ligaments ($\chi^2=2.15, P=0.54$). With model assistance, senior physicians' image interpretation time decreased from (12.5 ± 3.2) min to (7.8 ± 2.1) min, diagnostic consistency improved from 70.8% to 91.7%, and NASA-TLX scores decreased from (42.5 ± 10.2) to (28.3 ± 8.5) (all $P<0.05$). False positives were mainly caused by edema signals or partial volume effects (46.67%), while false negatives were mainly due to micro-tears (70.00%). The signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) of the misdiagnosed cases were significantly lower than those of the correctly diagnosed cases (all $P<0.05$). **Conclusion** The diagnostic efficiency of the 3D U-Net model based on the 3D CNN architecture for MLKIs is comparable to that of senior physicians. It can significantly reduce the image interpretation time, alleviate the cognitive load, and improve the diagnostic consistency.

【Key words】Magnetic resonance imaging; Multiligament knee injuries; Convolutional neural network; Assisted diagnosis

膝关节多韧带损伤(multiligament knee injuries, MLKIs)常由高能创伤导致,可累及2条及以上主要韧带,术前准确判断损伤范围是制订手术方案的关键^[1-2]。磁共振成像(magnetic resonance imaging, MRI)是该类损伤的首选影像学检查方法,但 MLKIs 因解剖结构重叠、信号混杂及水肿干扰,诊断难度大,低年资医师易漏诊或误诊。深度学习(deep learning, DL)技术,尤其是3D卷积神经网络(convolutional neural network, CNN),可充分挖掘MRI的三维空间信息并自动提取特征,实现病灶的精准检测与分割^[3]。U-Net等经典网络架构已在半月板撕裂等膝关节疾病的MRI诊断中表现出良好性能^[4]。然而,针对MLKIs这一复杂场景的3D U-Net研究仍相对不足,尤其缺乏模型与不同年资医师诊断效能的系统对比,以及辅助诊断效率的相关评估^[5]。为此,本研究基于3D CNN架构,构建3D U-Net模型,评估其辅助MRI诊断MLKIs的临床价值。

1 资料与方法

1.1 一般资料

回顾性选取2022年1月至2025年12月蚌埠医学院第二附属医院影像科经关节镜证实的MLKIs患者100例。纳入标准:①有明确膝关节外伤史,临床高度怀疑多韧带损伤;②术前完成标准膝关节MRI检查;③检查后2周内行关节镜探查并确诊。排除标准:①有同侧膝关节手术史;②合并周围骨折或完全性关节脱位;③MRI存在严重运动伪影,影响判读;④临床资料不完整。连续纳入MLKIs患者287例,排除187例(合并膝关节手术史42例、合并骨折/脱位58例、MRI伪影干扰39例、临床资料不全48例),最终纳入100例。本研究已通过蚌埠医学院第二附属医院伦理委员会批准(批件号:BYER2022-018)。

1.2 影像采集与处理方法

1.2.1 MRI扫描方案

采用美国GE医疗Discovery 750w 3.0T超导型磁

共振扫描仪,使用专用膝关节线圈^[6]。所有受检者取仰卧、足先进体位,膝关节伸直并外旋 $10^{\circ}\sim 15^{\circ}$ 。扫描序列及参数如下。①矢状位质子密度加权成像-脂肪抑制(proton density weighted imaging-fat suppression, PDWI-FS):重复时间(repetition time, TR)/回波时间(echo time, TE)=2 800/40 ms;②冠状位 T_2 加权成像-脂肪抑制(T_2 weighted imaging-fat saturation, T_2 WI-FS):TR/TE=2 800/45 ms;③横轴位 T_2 WI-FS:TR/TE=3 000/45 ms。设置层厚4.0 mm,层间距0.5 mm,矩阵 256×256 ,视野 $180\text{ mm}\times 180\text{ mm}$ 。

1.2.2 数据预处理与模型构建

所有DICOM图像完成脱敏后,由2名分别具有8年、12年肌骨诊断经验的高年资影像科医师双盲独立对前交叉韧带(anterior cruciate ligament, ACL)、后交叉韧带(posterior cruciate ligament, PCL)、内侧副韧带(medial collateral ligament, MCL)和外侧副韧带(lateral collateral ligament, LCL)的损伤区域行像素级标注,以关节镜结果为金标准进行校正;标注一致性良好(Dice系数0.92, Kappa系数0.89),存在分歧的区域经2名医师共同回顾关节镜影像、协商达成共识后确定。100例数据按7:3比例随机分为训练集($n=70$ 例)与测试集($n=30$ 例),两组年龄、性别、韧带损伤类型等基线资料均衡可比(均 $P>0.05$)。

基于3D CNN架构,构建3D U-Net模型,利用训练集对模型进行训练与参数优化。模型训练采用Adam优化器,初始学习率0.001,批大小(batch size)为4,以Dice Loss作为损失函数,共训练200个轮次;训练过程采用早停策略,当验证集损失连续10个轮次不再下降时终止训练。模型总参数量为5.8 M,单次前向推理浮点运算量为12.4 G FLOPs。在NVIDIA GeForce RTX 3090 GPU上,单例患者完整MRI序列的平均推理时间为0.8 s;在Intel Xeon Gold 6226R CPU上,同例推理时间为12.5 s。

1.3 观察指标

1.3.1 模型诊断效能

以关节镜为“金标准”,在测试集评估模型对 4 条韧带损伤的检测效能。由未参与标注的 3 名低年资医师(3 年经验)和 2 名高年资医师(15 年、18 年肌骨经验)独立阅片,计算 3D U-Net 模型、低年资组与高年资组的敏感度、特异度、阳性预测值、阴性预测值、准确率及曲线下面积(area under the curve, AUC)。

1.3.2 辅助诊断价值

统计高年资医师有无模型辅助下的平均阅片时间及诊断一致性,阅片后采用美国航空航天局任务负荷指数量表(National Aeronautics and Space Administration-Task Load Index, NASA-TLX)评估六维度主观负荷^[7],总分越高认知负荷越重。

1.3.3 假阳性/假阴性分析

回溯分析模型误判原因,比较误判病例与诊断正确病例的信噪比(signal-to-noise ratio, SNR)、对比噪声比(contrast-to-noise ratio, CNR)。

1.4 统计学方法

采用 SPSS 26.0 软件进行数据分析。符合正态分布的计量资料以均数±标准差($\bar{x}\pm s$)表示,医师阅片时间比较行配对 t 检验;计数资料以例(%)表示,组间比较行 χ^2 检验。3D U-Net 模型、低年资医师和高年资医师三者间敏感度、特异度、准确率的比较行 χ^2 检验,并行两两比较;模型对不同韧带诊断准确率的差异比较行 χ^2 检验。AUC 比较行 DeLong-Z 检验。认知负荷评分比较采用配对 t 检验。 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 3D U-Net 模型与医师诊断效能比较

以关节镜诊断结果为“金标准”,评估 3D U-Net 模型、低年资医师和高年资医师对 100 例患者 4 条主要韧带损伤的综合诊断效能。模型的敏感度、特异度及准确率均显著高于低年资医师,差异有统计学意义(均 $P<0.01$);与高年资医师相比,各指标差异均无统计学意义(均 $P>0.05$)。模型 AUC 显著高于低年资医师($Z=3.21, P<0.001$)(表 1)。

2.2 3D U-Net 模型对不同类型韧带损伤的诊断效能

结果显示,3D U-Net 模型对 ACL 损伤的诊断敏感度最高(95.0%),而对 PCL 损伤的特异度最高

表 1 不同诊断方法对 MLKIs 的诊断效能比较(%)

诊断方法	敏感度	特异度	准确率	AUC
3D U-Net 模型	92.5	89.7	91.5	0.94
低年资医师	81.3	78.2	80.0	0.80
高年资医师	94.8	92.1	93.7	0.96
χ^2	7.852	9.334	10.125	—
$P_{3D\ U-Net\ 模型:低年资}$	0.006	0.002	<0.001	<0.001

注: MLKIs: 膝关节多韧带损伤; AUC: 曲线下面积。

(97.5%)。模型对 MCL 损伤的诊断准确率相对较低。4 条韧带诊断准确率经 χ^2 检验,差异无统计学意义($\chi^2=2.15, P=0.54$)(表 2)。

表 2 3D U-Net 模型对不同韧带损伤的诊断效能(%)

韧带类型	敏感度	特异度	阳性预测值	阴性预测值	准确率
ACL ($n=72$)	95.0	92.9	97.1	88.5	94.0
PCL ($n=35$)	90.9	97.5	93.8	96.4	95.0
MCL ($n=58$)	88.5	86.7	89.8	85.1	87.6
LCL ($n=28$)	92.3	95.8	85.7	98.1	95.2

注: ACL: 前交叉韧带; PCL: 后交叉韧带; MCL: 内侧副韧带; LCL: 外侧副韧带。

2.3 3D U-Net 模型辅助对高年资医师诊断效能与效率的影响

结果显示,模型辅助后高年资医师的平均阅片时间显著缩短,诊断一致性明显提升,差异均有统计学意义($P<0.05$)。模型辅助后,高年资医师的 NASA-TLX 评分由独立阅片时的(42.5 ± 10.2)分降至(28.3 ± 8.5)分,差异有统计学意义($P<0.001$)(表 3)。

表 3 3D U-Net 模型辅助前后高年资医师诊断效能、效率及认知负荷比较

指标	无模型辅助	有模型辅助	t/χ^2	P
例数(例)	30	30	—	—
平均阅片时间($\bar{x}\pm s$, min)	12.5±3.2	7.8±2.1	-11.481	<0.001
诊断一致性[例(%)]	17(70.8)	22(91.7)	4.214	0.040
敏感度(%)	93.1	96.5	1.159	0.282
准确率(%)	92.0	96.0	0.748	0.387
NASA-TLX 评分($\bar{x}\pm s$, 分)	42.5±10.2	28.3±8.5	-5.669	<0.001

注: NASA-TLX: 美国航空航天局任务负荷指数量表。

2.4 3D U-Net 模型假阳性和假阴性结果的回溯性分析

30 例测试集共 120 个韧带区域中,3D U-Net 模型产生了 15 个假阳性判断和 10 个假阴性判断(表 4)。误判病例 SNR 和 CNR 均显著低于正确诊断病例($12.3\pm4.1: 18.5\pm5.2, t=-5.379, P<0.05; 8.9\pm$

3.5:14.2±4.8, $t=-4.903$, $P<0.05$)。

表4 3D U-Net模型假阳性和假阴性结果原因分析

误判类型	原因分类	例数 (个)	构成比 (%)
假阳性 ($n=15$)	韧带周围水肿信号/部分容积效应	7	46.67
	正常血管结构误判	4	26.67
	韧带内黏液样变性	2	13.33
	术后瘢痕组织干扰	2	13.33
假阴性 ($n=10$)	微小撕裂 (<2 mm)	7	70.00
	扫描层面与损伤走行平行	2	20.00
	患者运动伪影	1	10.00

3 讨论

本研究构建了基于3D CNN架构的像素级3D U-Net模型,在MLKIs场景下系统评估其对4条主要韧带的综合诊断效能、辅助价值及误判模式;同时引入NASA-TLX认知负荷量表,从医师主观体验维度拓展评估框架^[7-8]。结果显示,模型综合敏感度(92.5%)、特异度(89.7%)和准确率(91.5%)均显著优于低年资医师,与高年资医师诊断水平相近($AUC:0.94:0.96$, $P>0.05$);其核心优势为3D U-Net可充分捕捉MRI层间空间连续性,弥补传统2D CNN丢失层间关联的不足。分韧带评估结果显示,模型对ACL的敏感度最高(95.0%)、对PCL的特异度最高(97.5%),对MCL的准确率相对较低(87.6%),考虑与MCL紧邻内侧半月板及关节囊、水肿信号易蔓延干扰本体信号有关;4条韧带的评估效能差异无统计学意义,综合表现基本均衡。

辅助诊断层面,模型辅助后高年资医师阅片时间由(12.5±3.2)min缩短至(7.8±2.1)min,诊断一致性由70.8%提升至91.7%,NASA-TLX总分由(42.5±10.2)分降至(28.3±8.5)分。提示模型不仅可提升阅片效率,还可实质性降低医师的主观工作负荷。该模型参数量仅5.8 M,单例推理耗时0.8 s、单例计算成本约0.1元,具备嵌入图片归档与通信系统(picture archiving and communication system, PACS)、构建“人工智能初筛-医师审核”闭环流程的技术可行性与费效比优势。误判分析进一步显示,模型假阳性多源于韧带周围水肿信号或部分容积效应(占比46.67%),假阴性多见于微小撕裂(<2 mm)及扫描层面不佳(占比70.00%);误判病例SNR和CNR均

显著低于正确诊断病例,明确图像质量是模型失效的独立危险因素^[9]。后续可从“外部验证-PACS嵌入式前瞻试验-真实临床工作流评估”三阶段路径入手,推进临床转化。

综上,基于3D CNN架构的3D U-Net模型可高效辅助诊断MLKIs,在缩短阅片时间、降低认知负荷、提升诊断一致性方面展现出良好前景,其轻量化架构与高推理速度为临床转化奠定基础。经多中心验证与模型迭代优化后,该模型有望推动骨肌影像诊断向智能化方向发展。本研究存在单中心回顾性设计、样本量有限、缺乏外部验证及健康/单韧带损伤对照等局限,“金标准”标注亦可能存在像素级偏差,模型适用的最低图像质量阈值尚待进一步界定。

参考文献

- [1] 楼超,戴雪松,陆志凯. 膝关节多韧带损伤的研究进展[J]. 现代实用医学, 2023, 35 (4): 425-428.
- [2] 韩亚洲,刘伟健,董敬. 观察关节镜辅助分期手术治疗膝关节多韧带损伤合并胫骨平台骨折的临床效果[J]. 中国伤残医学, 2023, 31 (18): 47-50.
- [3] 蒋可欣,谢雨含,李勉文,等. 基于半月板MRI的3D卷积神经网络模型预测膝关节炎发生的研究[J]. 磁共振成像, 2024, 15 (2): 103-107, 121.
- [4] 袁典,杜昱峥,魏德健,等. 卷积神经网络基于MRI在半月板损伤诊断中的研究进展[J]. 磁共振成像, 2024, 15 (3): 223-229.
- [5] FRITZ B, FRITZ J. Artificial intelligence for MRI diagnosis of joints: a scoping review of the current state-of-the-art of deep learning-based approaches[J]. Skeletal Radiol, 2022, 51(2):315-329.
- [6] 李海军,胡友珍,冯世波. 核磁共振在评价超短波治疗对兔膝关节损伤模型关节功能恢复中的价值[J]. 临床和实验医学杂志, 2021, 20 (10): 1058-1061.
- [7] BELL S W, KONG J C H, CLARK D A, *et al.* The National Aeronautics and Space Administration-task load index: NASA-TLX: evaluation of its use in surgery[J]. ANZ J Surg, 2022, 92(11):3022-3028.
- [8] XIE Z, QIU Z, LI Y, *et al.* Development of a multi-task deep learning system for classification of nine common knee abnormalities on MRI: a large-scale, multicentre, stepwise validation study[J]. EClinicalMedicine, 2025, 89:103534.
- [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.