

• 方法学 •

人工智能医学影像研究报告规范： CLAIM 检查清单解读



诸宇佳^{1,2}, 韩慧¹, 卫建华³, 沈建通², 华雨婷², 莫航沅², 蔡婷婷²

1. 湖州师范学院附属第一医院(浙江湖州 313000)

2. 湖州师范学院医学院护理学院(浙江湖州 313000)

3. 浙江大学医学院附属第一医院(杭州 310000)

【摘要】 目前,基于人工智能的医学影像方法快速发展,相关的文献报告逐年增多,但未有专用的报告标准,结果报告不规范。为提高该类研究的报告质量,帮助读者和评价人员更科学地评估该类研究质量,国外提出了人工智能医学影像研究检查清单(CLAIM)。本文介绍 CLAIM 的内容,并对其条目进行解释说明。

【关键词】 人工智能;医学影像;CLAIM;报告质量

Interpretation of checklist for artificial intelligence in medical imaging (CLAIM)

ZHU Yujia^{1,2}, HAN Hui¹, WEI Jianhua³, SHEN Jiantong², HUA Yuting², MO Hangfeng², CAI Tingting²

1. The First Affiliated Hospital of Huzhou University, Huzhou 313000, P. R. China

2. Department of Huzhou University, School of Nursing Science, Huzhou 313000, P. R. China

3. The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310000, P. R. China

Corresponding author: SHEN Jiantong, Email: sjiantong@163.com

【Abstract】 Currently, the medical imaging methods based on artificial intelligence are developing rapidly, and the related literature reports are increasing year by year. However, there is no special reporting standard, and the reporting of the results is not standardized. In order to improve the report quality of this kind of research and help readers and evaluators evaluate the quality of this kind of research more scientifically, a checklist for artificial intelligence in medical imaging (CLAIM) was put forward abroad. This paper introduces the content of CLAIM and explains its items.

【Key words】 Artificial intelligence; Medical imaging; CLAIM; Report quality

基于人工智能(artificial intelligence, AI)的医学影像方法突破了传统方法的技术壁垒,是近年发展最快的领域之一^[1]。在医学影像领域,AI技术的介入结合大数据挖掘,使得医学影像大数据在AI的筛选、梳理和提取后,具有转换成有效的临床决策的潜力^[2]。有研究发现,2011—2020年AI+医学影像研究逐年增多,中国发表的数量最多,美国、韩国次之^[3]。AI在医学成像领域的应用应基于科学的原则,研究报告应描述具体的研究过程,其研究成果应具备可重复性^[4]。AI研究过程包括数据获取、数据处理、机器学习及模型评估,现有的报告指南并未完全涵盖AI系统特有的偏倚潜在来源^[5]。如诊断准确性研究报告规范旨在通过加强研究报

告的透明度及完整性,以提高诊断性试验的报告质量^[6]。加强流行病学观察性研究报告主要目的是为观察性流行病学研究论文提供报告规范^[7]。临床试验报告的统一标准则是为了改进随机对照试验的报告质量,方便读者了解研究实施情况并判断其结果的真实性^[8]。2020年加州大学旧金山分校放射学和生物医学成像系Mogan等学者提出了AI+医学影像研究报告的检查清单(checklist for artificial intelligence in medical imaging, CLAIM)^[9],本文对其内容进行介绍,以供国内学者参考。

1 CLAIM 的概述与内容

CLAIM参照诊断准确性研究报告规范,并结合已有的机器学习模型评估指南^[10-13]和一般文献报告标准制订,内容注重AI在医学影像中的应用,包括分类、图像重建、文本分析和 workflow 优化。

DOI: 10.7507/1672-2531.202303152

通信作者: 沈建通, Email: sjiantong@163.com



<http://www.cjebm.com>

(C)1994-2024 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

表 1 CLAIM 条目清单

栏目/主题	序号	条目
标题与摘要		
标题	1	报告人工智能方法的研究, 明确使用的技术类别 (例如: 深度学习)
摘要	2	包括研究设计、方法、结果和结论的结构化摘要
引言		
	3	说明临床背景, 包括人工智能方法的预期用途和临床作用
	4	说明研究目标和假设
方法		
研究设计	5	说明为前瞻性或回顾性研究
	6	报告研究目标, 如: 模型创建、探索性研究、可行性研究、非劣效性试验等
	7	说明数据来源
	8	报告纳入标准: 研究对象的招募方式、合格标准和源人群, 报告样本量及是否进行过样本统计效力和样本量的估算
	9	报告数据预处理步骤
	10	如果适用, 说明选择数据子集的方法
数据	11	参照公共数据元素, 定义数据元素
	12	报告数据去识别化方法
	13	报告如何处理缺失数据
	14	详细报告参考标准的定义, 使其能被重复
	15	选择参考标准的理由 (如果存在其他标准)
参考标准	16	说明参考标准注释的来源; 注释者的资格和准备工作
	17	说明注释工具
	18	报告测量研究者之间和研究者内部的差异性, 减少差异性和/或解决差异的方法
	19	说明预定的样本量以及如何确定
数据分区	20	说明如何将数据分配给分区, 并说明比例
	21	说明分区不相交的级别 (例如: 图像、研究、患者、机构)
	22	详细报告模型, 包括输入、输出、所有中间层和连接
模型	23	说明软件库、框架和软件包
	24	报告模型参数的初始化 (例如: 随机化、迁移学习)
训练	25	说明训练方法的细节, 包括数据增量、超参数、训练的模型数量
	26	说明选择最终模型的方法
	27	说明组建技术
	28	说明模型性能的度量
评价	29	说明显著性和不确定性的统计方法 (例如: 置信区间)
	30	报告稳定性或敏感性分析
	31	说明解释性方法 (例如: 显著性地图), 以及如何使用这些方法进行验证
	32	使用外部数据进行验证或测试
结果		
数据	33	说明受试者流向, 用流程图说明纳入和排除情况
	34	报告各分区病例的人口学和临床特征
	35	所有数据分区上的最佳模型的性能指标
模型性能	36	诊断准确性的估计及其精确度 (如: 95%的置信区间)
	37	对错误分类个体进行失效分析
讨论		
	38	说明研究的局限性, 包括潜在的偏倚、统计的不确定性和可推广性
	39	描述对实践的影响, 包括预期用途和 (或) 临床作用
其他信息		
	40	说明注册号和注册机构名称
	41	说明获取完整研究方案的路径
	42	说明资金和其他支持的来源, 资助者的作用

CLAIM 分为标题与摘要、引言、方法、结果、讨论和其他信息 6 个部分,共包含 42 个条目,条目清单见表 1。

2 CLAIM 工具条目的解读

以下是对 CLAIM 清单各条目的解读,部分条目将结合已发表的相关研究^[14]进行解读,以促进读者对报告规范的理解与应用。

2.1 标题和摘要

条目 1: 在文章的标题或摘要中指出 AI 技术的使用,如“深度学习”或“随机森林”。

实例: 例文^[14]在摘要中即表明该研究通过深度学习算法来预测高危儿童确诊自闭症谱系病的情况。

条目 2: 摘要应对研究的设计、方法、结果和结论进行总结。提供研究人群的概况(如患者或检查的数量、图像的数量、年龄和性别分布),说明该研究为前瞻性研究或回顾性研究,并总结所选择的统计分析方法。在介绍结果时,要包括比较的 *P* 值,并说明软件、数据和(或)结果模型是否可以公开获得。

实例: 例文^[14]在摘要部分已对研究的设计、方法、结果和结论进行了简要报告,并表明研究是对具有自闭症谱系病(autism spectrum disorder, ASD)高家族性风险的婴儿进行的前瞻性脑成像研究。在结果部分,未总结所选统计分析方法,也未对关于软件、数据和(或)结果模型是否可以公开获得进行说明。

2.2 引言

条目 3: 描述该研究的基本原理、研究目标和预期影响。总结相关文献,指出该研究建立的研究基础,以及与其他研究的不同之处。使读者了解研究的背景、基础科学、方法的基本假设。

实例: 例文^[14]中说明了是通过对前期相关研究的总结,从而建立了通过对具有 ASD 高家族性风险的婴儿进行的前瞻性脑成像研究能够预测 ASD 发生的假设。

条目 4: 定义要回答的临床或科学问题,清楚描述研究过程。

2.3 方法

清晰完整地描述研究的方法,使读者能够复制所描述的步骤。如果该部分描述超过了期刊的字数限制,在方法部分进行总结,在文章附件中提供完整的细节。

实例: 例文^[14]中说明了机器学习程序和有效

性测试的细节可从补充信息中获得。

2.3.1 研究设计 条目 5: 说明该研究为回顾性研究或前瞻性研究,如为前瞻性研究需评估预测模型。

条目 6: 界定研究的目标,如建立模型、探索性研究、可行性研究或非劣效性试验。对于分类系统则说明预期用途,如诊断、筛查、分期、监测、监视、预测或预后^[15]。说明 AI 算法相比于其他方法的优势,如分类、替代或补充^[15]。描述选取预测模型的类型、预测的目标以及将如何解决临床或科学问题。

2.3.2 数据 条目 7: 说明研究数据的来源,并指出数据与预测模型的拟合程度。描述预测模型的目标应用,使读者理解报告准确性估计的含义。参考以前使用相同数据集的研究,并说明与当前研究的不同。作者应遵守伦理准则,描述伦理审查和知情同意^[16]。如果可以的话,应提供数据来源和(或)图片的链接,鼓励作者将建模或分析的数据和(或)软件存放在一个可公开访问的资料库中。

实例: 例文^[14]中说明了该研究包括来自美国国立卫生研究院资助的自闭症卓越中心网络研究的数据,被称为“婴儿脑成像研究”。该网络包括 4 个临床数据收集站点(北卡罗来纳大学教堂山分校、华盛顿大学、费城儿童医院、华盛顿大学)。数据收集地点获得了其机构审查委员会的研究方案批准,所有入组受试者都得到了父母或者监护人的知情同意。

条目 8: 报告纳入标准:研究对象的招募方式、合格标准和源人群,报告样本量及是否进行过样本统计效力和样本量的估算。

实例: 例文^[14]中说明如果受试者的哥哥姐姐临床确诊为 ASD,则被登记为高危组,如果受试者的哥哥姐姐没有 ASD 的证据,且没有一级或二级亲属患有 ASD 的家族史,则被纳入低危组。排除标准包括:① 诊断或体征强烈提示与 ASD 相关的遗传性疾病;② 患有相关疾病(例如:中枢神经感染、癫痫、先天性心脏病);③ 感觉障碍(例如:视力或听力损失);④ 低出生体重(<2 000 克)或早产(妊娠<36 周);⑤ 母亲孕期曾暴露于酒精、致畸处方药等化学因素;⑥ 语言为非英语;⑦ MRI 禁忌症(例如:金属植入物);⑧ 被收养的受试者;⑨ 有智力残疾家族史、一级亲属有精神病、精神分裂症或双相情感障碍病史。

条目 9: 将不同来源的原始数据进行预处理,使其转换为定义明确、机器可读的格式以供分

析^[17-18]。详细描述数据预处理的步骤,以便其他研究者可以重复。具体说明如何对图像重新取样、改变色彩深度及调整窗口的设置。说明数据是否经过重新缩放、阈值限制(“二值化”)和(或)标准化处理。同时说明如何处理以下问题:区域格式、手工输入、不一致数据、缺失数据、错误的数据类型、文件操作和缺失的匿名化。定义删除异常值的标准^[10],说明软件库、软件(包括制造商名称和位置)和版本号,以及所有采用的选项和配置设置。

条目 10: 在一些研究中,研究者选择原始数据的一些子集作为预处理的一个步骤。例如:选择图像的一个子集,裁剪图像的一部分,或提取报告的一部分。如果这个过程是自动化的,需要说明所使用的工具和参数。如果是人工操作,说明人员的培训以及操作标准化。

条目 11: 定义预测变量和结果变量。如果适用的话,将其映射到公共数据,如放射学界^[19-21]或美国国立卫生研究院^[22-23]所维护的数据。

实例: 例文^[14]中说明支持这项研究结果的原始数据可从美国国立卫生研究院国家自闭症研究数据库公开获得。经合理要求时,可向通讯作者提供任何额外数据。

条目 12: 说明对数据进行去标识化的方法,以及如何删除受保护的健康信息以符合相关法律。由于面部轮廓可以进行识别,应具体说明如何删除此类信息或使之无法识别的方法^[17]。

条目 13: 说明缺失数据的处理。例如:用近似值或预测值来代替。描述估算数据引起的偏倚。

2.3.3 参考标准 **条目 14:** 对参考标准的注释最好是参考公用数据元素。避免模糊的描述,如“肝脏病变的大小”,应使用更精确的定义,如“在厚度为 2.5 毫米的轴向对比增强 CT 图像上测量的完全通过病变的最大线性测量值(毫米)”。向注释者提供实例图集,以说明主观的分级方案(例如:轻度/中度/重度),并提供该信息供审查。

条目 15: 说明选择参考标准的理由以及该参考标准的潜在误差、偏差和局限性。

条目 16: 具体说明注释者的数量和其资格,以及如何对注释者进行培训。如果可能的话,包括培训材料作为附件。说明注释是否是独立完成的,以及如何解决注释者之间的差异。

实例: 例文^[14]中说明所有的扫描结果都由一位儿科神经放射学家进行了审核,如果发现有任何放射学上的异常情况,就会告知参与者。此外,一名经委员会认证的儿科神经放射学家对“婴儿脑

成像研究”中所有 MRI 扫描结果进行盲审,并对其他发现进行了评级。第三位神经放射学家进行第二次盲评,如果前两次审查结果之间存在差异,通过他们来促成最终评级。

条目 17: 说明用于手动、半自动或自动注释的软件,包括版本号。说明是否使用自然语言处理或递归神经网络,以及如何从自由文本成像报告或电子健康记录中提取成像标签^[17,24-25]。

条目 18: 说明衡量评分者之间和评分者内部差异的方法,以及为减少或减轻这种差异和解决不一致而采取的措施。

2.3.4 数据分区 **条目 19:** 描述样本量及其估算方法。可使用把握度计算的方法来估计所需的样本量^[26]。

条目 20: 具体说明如何将数据分配到训练集、验证(调整)集和测试集等分区,指出每个分区的数据比例并说明选择的理由。如:每个分区的数据之间存在系统性差异,说明原因。

条目 21: 描述分区不相交的层级。医学图像集一般应在病人级别或更高的级别上分开,使同一患者的图像不会出现在每个分区中。

2.3.5 模型 **条目 22:** 提供完整而详细的模型结构,包括输入、输出和所有的中间层,以便其他研究人员可以准确地重新构建模型。对于神经网络模型,在层的描述中包括池化、归一化、正则化和激活的所有细节。模型输入必须与预处理数据的形式相匹配,输出必须符合所述临床问题的要求,对于有监督的学习来说,应该与参考标准注释的形式相匹配。如果采用已发表的模型结构,应引用相应参考文献,并充分描述对模型的每次修改。在某些情况下,可用代码提供模型的结构作为补充数据。

条目 23: 说明所有软件库、框架和软件包的名称和版本号。避免对硬件进行详细描述,除非基准计算性能是工作的重点。

条目 24: 说明模型的参数初始化的方法。描述随机初始化参数的随机值分布。如果采用迁移学习来初始化参数,说明起始权重的来源。当随机初始化和迁移学习相结合时,要清楚地说明模型的哪些部分是以何种策略初始化的。

2.3.6 训练 **条目 25:** 完整详细地描述所有的训练过程和超参数,以便其他研究者可以完全重复训练过程。通常情况下,为了完整地记录训练过程,稿件中应描述训练数据是如何被增加的(例如:对于图像来说,转换的类型和范围)。说明如何监测每个模型的训练收敛情况,以及停止训练的标准。

说明每个超参数所使用的值,其中哪些是在模型之间变化的,在什么范围内,以及使用什么搜索策略。对于神经网络,超参数的描述至少包括学习计划、优化算法、最小批量大小、辍学率和正则化参数。讨论采用的目标函数、选择理由和性能要求的符合程度。定义最佳性能模型的选择标准,如果某些模型参数被冻结或限制修改,如转移学习中经常出现的情况,应明确指出涉及哪些参数,限制它们的方法,以及限制所适用的训练部分。在代码中以训练脚本的形式描述这些细节可能更简洁,尤其是使用标准框架的神经网络模型。

条目 26: 说明用于选择最佳模型的方法和性能参数,以便对保留的测试集进行评估。如果选择了一个以上的模型,说明原因。

条目 27: 如果最终的算法涉及到模型的组合,完整详细地描述组合的每个模型,说明各组成模型的输出是如何加权和组合的。

2.3.7 评价 条目 28: 描述用于衡量模型性能的指标,并说明该指标对于临床和科学研究问题的重要性。将所提交的模型与以前发表的模型进行比较。

条目 29: 表明性能指标值的不确定性,如标准差和(或)置信区间。对指标进行统计学比较,指定统计软件。

条目 30: 分析模型的稳定性或敏感性。

实例: 例文^[14]中说明该研究的预测模型采用标准的十倍交叉验证方法进行评估,总体准确性为 94%,敏感性为 88%,特异性为 95%,阳性预测值为 81%,阴性预测值为 97%。

条目 31: 描述能够排除或解释模型结果的方法,并提供用于产生这些结果的参数^[13]。描述此类方法是如何在目前的研究中得到了验证。

条目 32: 描述用于评估算法性能的数据。当这些数据与训练数据是同一来源时,说明其存在的局限。如果训练集和评价集的注释或数据的结构有差异,解释这些差异,并描述和说明为适应这些差异而采取的方法。

2.4 结果

详细介绍实验的结果。如果对结果的描述超过了字数或期刊的其他要求,可以在稿件的附件部分提供数据。

实例: 例文^[14]对部分实验数据和结果以扩展表的形式展示,并以补充信息的形式将详细的机器学习程序和有效性测试的细节呈现。

2.4.1 数据 条目 33: 具体说明患者、检查或信息的纳入与排除标准,并记录符合每个标准的病例

数。建议在结果中以流程图或表的方式说明最初的患者群体以及被排除患者,并说明排除原因。描述数据集的技术特征(例如:对于图像应描述程序供应商或模型、采集参数、重新格式化参数。对于报告应描述实践环境,报告作者的数量和培训,报告结构化程度)。

条目 34: 说明每个分区中病例的人口学和临床特征,并说明所有数据分区的性能指标。

实例: 例文^[14]的扩展数据表 1 呈现了受试者人口统计数据(包括组间差异的测试)。

2.4.2 模型性能 条目 35: 说明最终模型在测试部分的表现,将 AI 模型的性能与当前的诊断标准相比较。

条目 36: 对于分类任务,诊断准确度及其精确度的估计(95% 置信区间)^[15],可用 ROC 曲线或校准曲线。当不能直接计算置信区间时,报告重复抽样的非参数估计^[10]。说明哪些变量被证明可以预测应变变量,指出预测模型最有效或最无效的亚组人群^[10]。

条目 37: 提供信息以帮助理解错误的结果。如果任务被分为两个或更多的类别,提供一个混淆矩阵,显示预测类别与实际类别的统计。考虑提供错误分类的案例,以帮助读者更好地理解算法的优势和局限性。

2.5 讨论

该部分提供总结、局限性、影响和未来方向 4 个方面的信息。

条目 38: 简明扼要地总结结果,并解释目前的研究如何推动下一阶段的研究工作。确定研究的局限性,包括研究方法、材料、偏倚、统计不确定性、意外结果和可推广性。

实例: 例文^[14]中说明虽然这项研究的结果没有直接应用于更大规模的 ASD 儿童群体,但也表明使用大脑生物标志物的早期前驱症状检测来预测 ASD 诊断的可能性。未来的分析将结合来自其他相关模式(例如:行为、分子遗传学、电生理学和其他成像模式)的补充数据,可能会提高我们观察到的预测的准确性。

条目 39: 描述对实践的影响,包括 AI 模型的预期用途和可能的临床作用。基于现有研究,说明下一阶段的研究计划,并讨论阻碍该模型实践转化的相关因素

实例: 例文^[14]中说明本研究中描述的算法将需要进行复制,才能被认为是预测高家族风险婴儿 ASD 的可能临床工具,因为错误的诊断预测有

可能对个体和家庭产生不利影响。

2.6 其他信息

条目 40: 遵守国际医学期刊编辑委员会的临床试验注册声明。国际医学期刊编辑委员会要求在招募第一个病人前在公共试验注册机构注册临床试验, 以此作为发表的条件^[27]。在临床试验注册机构(如 ClinicalTrials.gov 或 WHO 一级注册机构)注册研究方案, 有助于避免重复或多余的研究, 并方便感兴趣的人与研究协调者联系^[28]。

条目 41: 如果研究方案超过了期刊的字数限制, 则向读者提供可获得完整研究方案的途径, 以便帮助读者评估研究的有效性, 也可以帮助其他研究人员重复该研究^[28]。详细描述软件和算法, 以便重复研究。将所有建模和数据分析的计算机代码存入一个可公开访问的存储库。

实例: 例文^[14] 说明了支持这项研究结果的原始数据可从美国国立卫生研究院国家自闭症研究数据库公开获得, 任何附加的数据都可以从相应的权威机构获得。并且提供了研究分析中使用工具(AutoSeg、HeadCirc 和 ITK-SNAP)的下载途径, 以及用于构建、训练和测试预测管道的 Matlab 源代码的网址。

条目 42: 说明资金和其他支持的来源、资助者在研究过程中的作用以及作者在研究的每个阶段是否具有独立性^[28]。

实例: 例文^[14] 在最后说明了这项研究这项工作得到了美国国立卫生研究院自闭症卓越中心的拨款和支持, 并且国家医学图像计算联盟提供了进一步的支持。

3 讨论

随着 AI 在医学成像领域的应用越发广泛, 基于 AI+医学影像的文献报告会越来越多。建议国内期刊将 CLAIM 作为报告规范纳入其稿约或作者须知, 并要求投稿的论文提供 CLAIM 核查清单中规定的各个项目的内容。部分文献报告可能无法完全符合 CLAIM 的每项标准, 但 CLAIM 可为作者和审稿人提供了一个框架, 其最终目标是促进关于 AI 应用的清晰、透明和可重复的科学交流。但 CLAIM 只能用于评价和审查相关研究报告, 而不能作为评价研究质量的工具。

参考文献

- 1 吴绯红, 赵煌旋, 杨帆, 等. 医学影像+人工智能的发展、现状与未来. *临床放射学杂志*, 2022, 41(4): 764-767.

- 2 金征宇. 人工智能医学影像应用: 现实与挑战. *放射学实践*, 2018, 33(10): 989-991.
- 3 付姣慧, 常晓丹, 沙俏丽, 等. 2011 年—2020 年深度学习用于医学影像学研究文献分析. *中国介入影像与治疗学*, 2022, 19(1): 53-57.
- 4 Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*, 2020, 2(2): e200029.
- 5 李子孝, 熊云云, 丁玲玲, 等. 人工智能干预性临床试验报告指南: CONSORT-AI 扩展. *中国卒中杂志*, 2020, 15(12): 1327-1336.
- 6 朱一丹, 李会娟, 武阳丰. 诊断准确性研究报告规范 (STARD) 2015 介绍与解读. *中国循证医学杂志*, 2016, 16(6): 730-735.
- 7 王波, 詹思延. 如何撰写高质量的流行病学研究论文第一讲观察性流行病学研究报告规范—STROBE 介绍. *中华流行病学杂志*, 2006, (6): 547-549.
- 8 David M, Kenneth FS, Douglas GA, 等. CONSORT 声明: 提高平行随机试验报告质量的修订建议. *中国循证医学杂志*, 2005, 5(9): 702-707.
- 9 Si L, Zhong J, Huo J, et al. Deep learning in knee imaging: a systematic review utilizing a checklist for artificial intelligence in medical imaging (CLAIM). *Eur Radiol*, 2022, 32(2): 1353-1361.
- 10 Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res*, 2016, 18(12): e323.
- 11 Handelman GS, Kok HK, Chandra RV, et al. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *AJR Am J Roentgenol*, 2019, 212(1): 38-43.
- 12 Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*, 2018, 286(3): 800-809.
- 13 Bluemke DA, Moy L, Bredella MA, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers-from the Radiology Editorial Board. *Radiology*, 2020, 294(3): 487-489.
- 14 Hazlett HC, Gu H, Munsell BC, et al. Early brain development in infants at high risk for autism spectrum disorder. *Nature*, 2017, 542(7641): 348-351.
- 15 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*, 2015, 277(3): 826-832.
- 16 Geis JR, Brady AP, Wu CC, et al. Ethics of artificial intelligence in radiology: summary of the Joint European and North American Multisociety Statement. *Radiology*, 2019, 293(2): 436-440.
- 17 Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology*, 2020, 295(1): 4-15.
- 18 Harvey H, Glocker B. A standardised approach for preparing imaging data for machine learning tasks in radiology. In: Ranschaert ER, Morozov S, Algra PR, et al. Artificial Intelligence in medical imaging: opportunities, applications and risks. New York: Springer International, 2019.
- 19 Rubin DL, Kahn CE. Common data elements in radiology. *Radiology*, 2017, 283(3): 837-844.
- 20 Kohli M, Alkasab T, Wang K, et al. Bending the artificial intelligence curve for radiology: informatics tools from ACR and RSNA. *J Am Coll Radiol*, 2019, 16(10): 1464-1470.
- 21 Radiological Society of North America, American College of

- Radiology. RadElement: common data elements. 2020.
- 22 Sheehan J, Hirschfeld S, Foster E, *et al.* Improving the value of clinical research through the use of Common Data Elements. *Clin Trials*, 2016, 13(6): 671-676.
- 23 National Institutes of Health. NIH common data elements (CDE) repository. 2020.
- 24 Lakhani P, Kim W, Langlotz CP. Automated extraction of critical test values and communications from unstructured radiology reports: an analysis of 9.3 million reports from 1990 to 2011. *Radiology*, 2012, 265(3): 809-818.
- 25 Lipton ZC, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning. 2020.
- 26 Eng J. Sample size estimation: how many individuals should be studied. *Radiology*, 2003, 227(2): 309-313.
- 27 International Committee of Medical Journal Editors. Clinical Trials. 2020.
- 28 Cohen JF, Korevaar DA, Altman DG, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*, 2016, 6(11): e012799.

收稿日期: 2023-03-24 修回日期: 2023-09-07

本文编辑: 张洋

