

行业标准《人工智能医疗器械 质量要求和评价 第3部分： 数据标注通用要求》解析

郝 烨, 王 浩, 李佳戈

中国食品药品检定研究院, 北京 100050

通信作者: 李佳戈, E-mail: lijia@nifdc.org.cn

【摘要】伴随机器学习技术的迭代升级, 人工智能已成为医疗器械领域的新兴方向, 尤其近年来发展迅速。国家药品监督管理局于2022年发布了行业标准YY/T 1833《人工智能医疗器械 质量要求和评价》系列标准, 标志着我国已初步建立了人工智能医疗器械的标准体系。数据标注是人工智能医疗器械质控的重要环节, 数据标注的规范化、标准化不仅是标准体系的重要方面, 也决定了参考标准的准确性和可靠性, 从而对数据集质量和产品质量产生重要影响。《人工智能医疗器械 质量要求和评价》第3部分为数据标注通用要求, 其对人工智能医疗器械在数据标注环节的质量进行了要求, 并对评价方法进行了规范。本文将对其重要内容进行解析, 以期更好地开展产品质控, 促进产品及行业高质量发展。

【关键词】数据标注; 人工智能医疗器械; 产品质控; 数据集

【中图分类号】R-05; R608; TP3

【文献标志码】A

【文章编号】1674-9081(2023)06-1185-04

DOI: 10.12290/xhyxzz.2023-0463

Interpretation On Sectoral Standard Artificial Intelligence Medical Device-Quality Requirements and Evaluation-Part 3: General Requirements for Data Annotation

HAO Ye, WANG Hao, LI Jiage

National Institutes for Food and Drug Control, Beijing 100050, China

Corresponding author: LI Jiage, E-mail: lijia@nifdc.org.cn

【Abstract】 With the upgrading of machine learning technology, artificial intelligence has become an emerging direction in the field of medical devices. Especially in recent years, it has developed rapidly. In 2022, the National Medical Products Administration released the industry standard YY/T 1833 “Artificial Intelligence Medical Device-Quality Requirements and Evaluation” series, marking the initial establishment of a standard system for artificial intelligence medical devices in China. The standardization of data annotation is not only an important part of the standard system, but also determines the accuracy and reliability of reference standards, which has a significant impact on the quality of the dataset and product quality. Part 3 of “Artificial Intelligence Medical Device-Quality Requirements and Evaluation”, a general requirement for data annotation, proposes the quality requirements for artificial intelligence medical devices in the data annotation process and standardizes the evaluation methods. This paper highlights its important contents, in order to better carry out product quality control and promote high-quality development of products and industries.

基金项目: 中国食品药品检定研究院关键技术基金 (GJJS-2022-3-1)

引用本文: 郝烨, 王浩, 李佳戈. 行业标准《人工智能医疗器械 质量要求和评价 第3部分: 数据标注通用要求》解析 [J]. 协和医学杂志, 2023, 14 (6): 1185-1188. doi: 10.12290/xhyxzz.2023-0463.

【Key words】 data annotation; artificial intelligence medical devices; product quality control; datasets

Funding: Key Technology Fund of National Institutes for Food and Drug Control (GJS-2022-3-1)

Med J PUMCH, 2023,14(6):1185-1188

数据标注是基于监督学习的人工智能医疗器械在研发、测试阶段常用的一种技术服务,数据标注的准确性决定了数据集的使用价值^[1-6],故数据标注不仅需要完整的流程、规则和工具,还需制定相应的质量控制体系。YY/T 1833.3-2022《人工智能医疗器械 质量要求和评价 第3部分:数据标注通用要求》标准^[7]是医疗器械行业内首个侧重数据标注环节进行人工智能医疗器械质控管理分析的标准。该标准主要以人工智能医疗器械的数据标注环节为质控对象展开分析,不仅列举了标注工作质量控制方面的一系列特性和要求,并针对如何进行质控评价提出了专业性方法和建议,全文涵盖数据标注环节相关术语、标注任务说明文档、数据标注的质量特性、标注与质控流程、标注工具以及评价方法等多部分内容,对数据标注的质控分析客观、全面,对于人工智能医疗器械产品的数据标注工作具有重要指导意义,对于行业的规范化发展将发挥显著促进作用。本文笔者基于工作中积累的医疗器械检验检测经验及对该标准的理解,将对其重点内容进行解析。

1 指导标注任务文档编写

2007年斯坦福大学教授李飞飞等开启 ImageNet 图像数据集项目,完成了图片分类和标注,为后续机器学习提供了带标注的数据集,开启了数据标注的发展序幕^[8-10]。但数据标注行业经过十余年发展,针对数据标注的定义仍不明确,此种模糊的认知不利于整个标注行业特别是人工智能医疗器械行业的规范化发展。为此,YY/T 1833.1-2022《人工智能医疗器械 质量要求和评价 第1部分:术语》标准中,明确了数据标注的定义,其是指对数据进行分析,添加外部知识的过程^[11]。在此基础上,YY/T 1833.3-2022 标准根据数据标注的不同分类形式、关键要素等特点明确了包括标注任务、标注对象、结构化标注等在内的行业常用术语,为数据标注工作的规范和统一提供了必要的基础。

标注任务不同于传统的医疗器械产品,标注任务的说明文档在数据标注质量评价中具有十分重要的地位。在一项具体的标注任务开始前,标注责任方除了需明确标注任务的类型外,还需明确标注规则、标注

人员角色构成与分工、决策机制、标注工具、标注环境以及数据方面的要求。从结构角度来看,YY/T 1833.3-2022 标准是基于数据标注对完成任务的人、机、物、法、环等几大质控方面提出的要求。

制定清晰、合理的标注规则是整个标注任务顺利开展的基础。因此,在标注任务文档中,首先应明确标注规则。YY/T 1833.3-2022 标准针对此方面,提出的要求主要集中于对标注对象的定义需唯一且确定,同时指出标注对象的特征应该是可以验证或测量的,此为确保后续准确标注具有可衡量性的基础;标注对象及标注规则应具有充分的依从性文件作为支撑,如法规文件、技术标准、专家评议、文献分析等;此外,针对标注规则可能导致的偏倚风险,还需进行深入的风险分析。

标注人员作为标注工作的实际执行者,在标注工作中的表现直接决定了标注质量。实际标注中,根据任务的不同,可能需要多名人员承担不同的标注任务,包括初级标注人员、审核人员、仲裁人员等。YY/T 1833.3-2022 标准要求标注责任方应在文档中明确描述对标注人员的要求,包括人员资质、选拔依据、培训内容及评估指标等。标注工具是实现标注工作不可或缺的重要手段。标注工具的质控可参考医疗器械中软件的质控方法开展,YY/T 1833.3-2022 标准主要强调了对标注过程中所使用的软件、硬件和平台等应进行描述。标注环境的条件亦是标注工作质量管理的重要影响因素之一。环境条件对标注工作的影响主要表现在温湿度、照明条件及噪声干扰等方面。为保证标注任务的高质量,标注责任方有责任从标注环境对标注人员、标注过程以及标注工具的影响等多角度进行分析,并针对关键的环境要求进行说明。

2 明确标注任务质量特性

考虑到人工智能医疗器械产品数据标注工作面对的主要对象为医疗数据,由于数据标注的准确性决定了人工智能算法的有效性,医疗数据标注质量参差不齐可能导致临床误诊、漏诊等风险,甚至威胁患者的生命安全,故医疗数据标注的质量至关重要。

根据医疗数据集管理的整体思路,数据标注的质量特性应包含如下5个方面:(1)准确性。标注工

作中，为衡量标注结果的准确性，责任方应提供标注结果的适用指标，如召回率、准确率、灵敏度等。这些指标根据场景不同可能略有差异，可根据实际应用场景选择合适的指标。(2) 一致性。确保标注结果的一致性数据标注过程的重点，此为标注结果准确客观的必要条件，也是数据集管理中的重要环节，其包含两方面内容，即一方面应保证标注结果与参考标准的一致性；另一方面，针对标注过程中各个环节的输入与输出数据及其他信息，应保证内部一致性，如人员信息、标注结果及原始数据。(3) 精度。针对定量描述的标注结果，数据精度是评价标注结果的必要指标，也是相关责任方在标注工作中需明确的数据特性之一。(4) 可理解性、可访问性及可移植性。考虑到数据标注本质上更接近于医用软件类型的医疗器械产品^[12-15]，因此该部分的质控特性与软件有相似之处，需要考虑含标注结果的数据信息被授权用户理解和可访问的程度，以及此类数据能否被安装、移动或者替换到不同系统的能力。(5) 保密性及可追溯性。数据标注结果尤其其医疗数据的标注结果必须是安全保密且可追溯的。考虑到数据标注过程中可能涉及到复杂的人员结构和数据流转，很可能产生数据误操作、数据泄露、网络安全方面的风险，因此数据安全及其保密性不容忽视。可追溯性是用以表征标注任务可被追溯和记录到何种程度的特性。数据标注工作是数据集建设过程中的重要环节，为保证整个数据集的质量，在数据标注阶段需特别注重该环节的可追溯性，其涵盖的要素不仅包括标注任务的时间节点和标注对象，还应包括标注工作质控流程涉及的人员信息、标注任务的操作信息、标注工具的具体信息等。

3 指导标注质控流程设计

根据前节所述的标注任务质量特性，标注工作应建立完善的质控流程和管理机制，以保证质控工作高效进行。YY/T 1833.3-2022 标准分别从业务架构和过程组织两个方面明确了质控流程图的设计要点和相应要求。

从业务架构角度而言，质控流程的设计首先需根据数据流向及人员分工设计出质控业务架构图。数据标注的业务架构图根据标注人员的分工可分为初级标注、审核及仲裁环节。对于典型的医学影像类标注，可将标注任务分解为检出、分类、分割、测量等多项大任务。需注意的是，业务架构图中应明确关键的输入与输出节点，保存相应的标注结果和人员操作记

录。从过程组织的角度来讲，标注的质控流程应从任务生成、任务分配、任务实施、质量控制及安全管理5个方面进行质控架构图设计，以完善整个质控流程的实施。任务生成阶段可理解为数据获取和整理阶段，以及标注工具、标注环境和标注规则的确立阶段。为保证标注工作的可追溯性，该阶段需做好任务生成工作的记录和保存。任务分配及其实施是标注任务的重点，为确保标注工作可控与可靠，在标注分配过程中，需明确操作权限的设置，在实施阶段也应关注标注进度和标注人员的监控和协调。考虑到标注人员的标注质量是动态变化的，标注工作中应评估标注人员标注结果的可重复性和准确性。

安全管理措施设计是标注质控流程设计中另一个必不可少的重要部分。其需要重点关注标注前的数据脱敏、数据备份，标注过程中的网络安全，以及退出任务前的数据删除等，以保证标注环节中的数据安全性和隐私保护，避免数据泄露。

4 完善标注工具要求、明确标注任务评价方法

标注工具的选择主要基于标注任务的需求以及标注对象的数据格式等特点。适宜的标注工具不仅可提高标注任务效率，减少标注人员工作量，同时是数据标注过程中数据安全和高质量数据标注结果的重要保障。YY/T 1833.3-2022 标准中采用独立的一个章节明确了标注工具的质控方法，强调了标注工具在标注任务中的重要性，其明确标注工具的质控主要体现在性能和安全性两大方面：(1) 性能方面，需要重点关注标注过程中的一些标志性功能，如标注工具对于标注对象、数据标注以及标注结果在导入与导出功能、进度显示、任务调度以及审核与仲裁等方面的功能；(2) 安全性方面，根据数据标注结果安全性的规定，YY/T 1833.3-2022 标准对标注工具提出了身份鉴别以及访问控制方面的要求。考虑到标注工具的本质属于医用软件，安全特性与临床熟知的医用软件基本一致，因此，其安全质控与常规医用软件的质控方式大致相同。需要关注的是，传统人工标注的低效性与数据需求快速增长之间的矛盾逐渐凸显，因此，半自动甚至全自动标注将是数据标注的发展趋势^[16-17]。随着人工智能技术的高速发展，人工智能辅助标注已成为目前标注领域的研究热点。人工智能算法辅助的标注工具是基于人工智能技术赋能标注工作，对于提高标注效率大有裨益，故针对此类工具的质控和评价也需引起重视。辅助标注算法应经过确认

后交付使用，算法辅助标注的结果亦需要人工审核。YY/T 1833. 3-2022标准在附录中对相关内容的评价进行了原则性和功能性分析，为此类标注工具的应用提供了技术指导。

YY/T 1833. 3-2022 标准针对数据标注的说明文档、质量特性、质控流程和标注工具等方面不仅提出了具体要求且给出了相应的评价方法。参照 YY/T 1833. 2-2022 标准^[18]中对数据集质量评价的思路，YY/T 1833. 3-2022 标准对于标注结果的一致性和重复性评价亦以抽样检验的方法开展，通过专家论证、专家比对及定量计算等方式进行评价，为实际开展标注质控工作的相关人员提供了明确且全面的操作指导。为了更完整展示人工智能医疗器械领域数据标注工作的实际内容，YY/T 1833. 3-2022 标准在附录部分提供了典型的数据标注示例，包括医学影像、心电图特征、电子病历、心肺音等多类型临床数据，附录中的示例来自于人工智能医疗器械领域数据标注在临床中的实际应用，从每个标注任务的分类、标注规则、标注人员、标注工具、标注环境以及数据类型等角度分别归类汇总，为相关人员今后开展实际质控工作提供了重要技术参考。

5 小结

《人工智能医疗器械 质量要求和评价 第3部分：数据标注通用要求》从质量控制的角度出发，阐述了人工智能医疗器械产品在数据标注工作中可能产生的质量风险，从说明文档、质量特性、质控流程及评价方法等方面分别提出了关键性的建议和要求，对于人工智能医疗器械产品的标注工作具有重要指导和规范作用。该标准作为人工智能医疗器械标注体系内的一项关键基础性标准，必将有助于产品的规范化设计，为行业的快速有序发展提供关键技术指导。

作者贡献：郝烨负责文献调研、初稿撰写；王浩负责文献整理、论文修订；李佳戈负责论文修订、终稿审核。

利益冲突：所有作者均声明不存在利益冲突

参 考 文 献

[1] 王浩, 孟祥峰, 林晓兰, 等. 人工智能医疗器械数据集质控解决方案 [J]. 医疗卫生装备, 2023, 44: 12-15.

[2] 唐娜, 王浩, 钟代笛. 人工智能医疗器械监管现状分析 [J]. 医疗卫生装备, 2022, 43: 54-58, 68.

[3] 王浩, 李澍, 王晨希, 等. 人工智能医疗器械质量管理标准化趋势分析 [J]. 中国医疗设备, 2021, 36: 20-23.

[4] 刘毅, 王浩, 李澍, 等. 人工智能医疗器械企业质量管理体系构建关键指标筛选研究 [J]. 中国医疗设备, 2021, 36: 24-27, 43.

[5] 王浩, 唐桥虹, 郝烨, 等. 人工智能医疗器械标准体系设计探索 [J]. 中国医疗设备, 2021, 36: 15-18, 28.

[6] 彭亮, 刘泉寅, 张宇晶, 等. 人工智能医疗器械国际监管比较研究 [J]. 中国数字医学, 2023, 18: 1-7.

[7] 国家药品监督管理局. 人工智能医疗器械 质量要求和评价 第3部分：数据标注通用要求 [S]. YY/T 1833. 3-2022.

[8] 王浩, 孟祥峰, 王权, 等. 人工智能医疗器械用数据集管理与评价方法研究 [J]. 中国医疗设备, 2018, 33: 1-5.

[9] 蔡莉, 王淑婷, 刘俊晖, 等. 数据标注研究综述 [J]. 软件学报, 2020, 31: 302-320.

[10] 王权, 王浩, 孟祥峰, 等. 人员管理对人工智能医疗器械用数据集质量的影响分析 [J]. 中国医疗设备, 2018, 33: 6-9.

[11] 孟祥峰, 王浩, 李佳戈. 行业标准《人工智能医疗器械质量要求和评价 第1部分：术语》解析 [J]. 协和医学杂志, 2023, 14: 1175-1179.

[12] 孟祥峰, 王浩, 张超, 等. 医学人工智能产品检测平台的设计 [J]. 医疗卫生装备, 2023, 44: 1-7.

[13] 李澍, 王浩, 王晨希, 等. 人工智能医疗器械软件生产质量管理体系特殊要求研究 [J]. 中国医疗设备, 2021, 36: 15-18, 22.

[14] 曾雪, 王浩, 李佳戈. 分析人工智能医疗器械环境特殊要求 [J]. 中国医疗设备, 2021, 36: 19-22.

[15] 王浩, 孟祥峰, 李澍, 等. 数据集在人工智能医疗器械质控中的角色与要求 [J]. 中国医疗器械杂志, 2019, 43: 54-57.

[16] 王海林, 冯瑞, 张晓波. 融合深度主动学习的医学图像半自动标注系统 [J]. 计算机系统应用, 2023, 32: 75-82.

[17] 尹兆杰. 基于人机交互的深度学习训练数据标注系统 [J]. 铁路通信信号工程技术, 2021, 18: 24-30.

[18] 王浩, 孟祥峰, 郝烨, 等. 行业标准《人工智能医疗器械 质量要求和评价 第2部分：数据集通用要求》解析 [J]. 协和医学杂志, 2023, 14: 1180-1184.

(收稿：2023-10-07 录用：2023-10-30)
(本文编辑：董 哲)