



中华人民共和国医药行业标准

YY/T 1833.2—2022

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

Artificial intelligence medical device—Quality requirements and evaluation—
Part 2: General requirements for datasets

2022-07-01 发布

2023-07-01 实施

国家药品监督管理局 发布

目 次

前言 I

引言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 数据集说明要求 3

5 数据集质量要求 7

6 数据集质量符合性评价..... 10

附录 A（规范性） 数据集类型说明 14

附录 B（资料性） 数据筛选与清洗说明 15

参考文献 17

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

本文件是 YY/T 1833《人工智能医疗器械 质量要求和评价》的第 2 部分。YY/T 1833 已经发布了以下部分：

——第 1 部分：术语；

——第 2 部分：数据集通用要求。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由国家药品监督管理局提出。

本文件由全国人工智能医疗器械标准化技术归口单位归口。

本文件起草单位：中国食品药品检定研究院、解放军总医院、上海长征医院、中国科学院深圳先进技术研究院、华为技术有限公司、飞利浦（中国）投资有限公司、上海西门子医疗器械有限公司、杭州依图医疗技术有限公司、腾讯医疗健康（深圳）有限公司、上海联影智能医疗科技有限公司、广州柏视医疗科技有限公司、美中互利医疗有限公司、慧影医疗科技（北京）有限公司、北京安德医智科技有限公司、广州市妇女儿童医疗中心、中山大学中山眼科中心。

本文件主要起草人：王浩、刘士远、何昆仑、郑海荣、李佳戈、詹翊强、孟祥峰、萧毅、葛鑫、刘东泉、颜子夜、钱天翼、丁子建、陆遥、任海萍、柴象飞、周娟、王珊珊、张培芳、梁会营、林浩添。

引 言

近年来,人工智能医疗器械不断发展,成为医疗器械标准化领域的一个新兴方向。我国已初步建立人工智能医疗器械标准体系。在该标准体系中,YY/T 1833《人工智能医疗器械 质量要求和评价》是基础通用标准,为开展细分领域的标准化活动提供指导,拟由八个部分组成。

- 第1部分:术语。目的在于为人工智能医疗器械的质量评价活动提供术语。
- 第2部分:数据集通用要求。目的在于提出数据集的通用质量要求与评价方法。
- 第3部分:数据标注通用要求。目的在于提出数据标注环节的质量要求与评价方法。
- 第4部分:可追溯性通用要求。目的在于提出人工智能医疗器械可追溯性的通用要求及评价方法。
- 第5部分:算法安全要求。目的在于规范人工智能医疗器械采用的人工智能算法的安全要求与评价方法。
- 第6部分:环境要求。目的在于规范人工智能医疗器械的运行环境条件要求与评价方法。
- 第7部分:隐私保护要求。目的在于加强人工智能医疗器械保护受试者隐私的能力。
- 第8部分:伦理要求。目的在于从技术层面实现人工智能伦理的要求,保护人的权益。

本文件为其他部分提供数据集质量评价相关的工作思路。

数据集是人工智能医疗器械研发、训练、测试、质控等环节常用的资源,对产品质量有重要影响。本文件作为 YY/T 1833《人工智能医疗器械 质量要求和评价》的第2部分,把数据集纳入产品质量评价体系的一部分,也为后续制定细分领域的数据集专用要求提供依据。

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

1 范围

本文件规定了人工智能医疗器械全生命周期使用的数据集的通用质量要求和评价方法。

本文件适用于人工智能医疗器械研发、生产、测试、质控等环节使用的数据集的开发与评价。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 2828.4 计数抽样检验程序 第4部分:声称质量水平的评定程序

GB/T 2828.11 计数抽样检验程序 第11部分:小总体声称质量水平的评定程序

GB/T 6378.4 计量抽样检验程序 第4部分:对均值的声称质量水平的评定程序

YY/T 1833.1 人工智能医疗器械 质量要求和评价 第1部分:术语

3 术语和定义

YY/T 1833.1 界定的以及下列术语和定义适用于本文件。

3.1

计数检验 inspection by attributes

关于规定的一个或一组要求,或者仅将单位产品划分为合格或不合格,或者仅计算单位产品中不合格数的检验。

[来源:GB/T 2828.1—2012,3.1.3]

3.2

计量质量特性 variables quality characteristic

被检的单位产品特性能用连续尺度进行度量的质量特性。

[来源:GB/T 8054—2008,3.1.3]

3.3

计量抽样检验 sampling inspection by variables

按规定的抽样方案从批中随机抽取一定数量的单位产品。用测量、试验或其他方法取得它们的质量特性值,与质量要求进行对比,并判断该批产品能否接收的过程。

[来源:GB/T 8054—2008,3.1.4]

3.4

批 lot

按抽样目的,在基本相同条件下组成的总体的一个确定部分。

[来源:GB/T 10111—2008,3.1.4]

3.5

准确度 accuracy

对数据内容正确、形式有效的一种度量。

[来源:GB/T 11457—2006,2.22,有修改]

3.6

精度 precision

对于说明的量的精确或差异的程度。例如,2 位十进制数字对 5 位十进制数字。

[来源:GB/T 11457—2006,2.1160]

注:含义有别于算法性能指标的精确度。

3.7

一致性 consistency

在数据集的各阶段、部分之间,一致、标准化、无矛盾的程度。

[来源:GB/T 11457—2006,2.320,有修改]

3.8

可得性 availability

数据集在投入使用时可操作或可利用的程度。

[来源:GB/T 11457—2006,2.115,有修改]

3.9

信息安全[性] information security

对数据进行的保护,以防止其受到意外的或蓄意的存取、使用、修改、毁坏或泄密。

安全性也涉及对受试者隐私、元数据、通信以及计算机安装的物理保护。

[来源:GB/T 11457—2006,2.1420,有修改]

3.10

可移植性 portability

数据集能被安装、替换或从一个系统移动到另一个系统中,并保持已有质量的属性的程度。

3.11

数据集制造责任方 dataset manufacture responsible organization

对某数据集的设计、制造负有责任的实体。

[来源:GB 9706.1—2020,3.101,有修改]

3.12

离群值 outlier

一组数据中被认为与该组其他数据不一致的观测值。

注:离群值可能来源于不同的总体,或由于不正确的记录或其他误差的结果。

[来源:CNAS GL-002—2018,3.6,有修改]

3.13

数据集说明 dataset description

陈述数据集各种性质的文档。

4 数据集说明要求

4.1 数据集描述

4.1.1 数据集整体描述

4.1.1.1 数据集类型

数据集说明应按照附录 A 的要求,注明数据集的类型。

4.1.1.2 数据集形态

数据集说明应描述数据集的数据模态、数据格式、数据量、存储方式。

4.1.1.3 数据集层次

数据集说明应描述数据集的数据层次、子集分解和组合变化。

4.1.1.4 应用场景

数据集说明应描述数据集预期模拟的应用场景。

4.1.1.5 版本控制

数据集说明应描述数据集的版本号、存储信息、说明文档编制时间、使用状态。

4.1.1.6 数据集更新

对于动态数据集,数据集说明应描述新增数据的入组规则和数据退役规则。

4.1.1.7 文档管理

数据集说明宜通过计算机技术(如软硬件、普通网络、云服务),对文档内容进行发布、交换、管理和查询。

4.1.1.8 元数据属性

数据集说明宜描述数据集元数据属性,如数据集名称、标识符、发布方、语种、分类类目名称、摘要等。

4.1.1.9 数据元属性

数据集说明宜描述数据元的共用属性、专用属性,如数据集版本、注册机构、分类模式、主管机构等。

注:注册机构、主管机构定义见 WS/T 303—2009。

4.1.1.10 标注信息

数据集说明应描述数据集是否具有标注信息。

4.1.2 数据采集信息

4.1.2.1 合规性陈述

数据集说明应提供数据来源的合规性陈述。

4.1.2.2 隐私保护

数据集说明应描述用于保护受试者隐私的技术手段,如数据去标识化、数据匿名化等。适当时,数据集说明文档应描述数据去标识化或者数据匿名化的规则。

4.1.2.3 多样性

数据集说明应提供数据来源多样性的描述,如人群、采集场所、采集设备、参数设置、操作人员资质、采集流程、采集时间等。

4.1.2.4 数据采集依从原则

数据集说明应提供数据采集依据的法规、技术标准、临床规范、专家共识或其他参考文献。

4.1.2.5 数据筛选

数据集说明应描述数据的录入、排除标准,以及进行数据筛选的方法,如人工清洗、自动清洗。

注:示例见附录 B。

4.1.3 数据预处理

适当时,数据集说明应描述数据预处理的操作步骤和内容。

4.1.4 数据集标注

4.1.4.1 数据集标注依从原则

数据集如具有标注信息,数据集说明应描述数据集标注依据的法规、技术标准、临床规范、专家共识或其他参考文献。

4.1.4.2 参考标准

数据集如具有标注信息,数据集说明应描述数据集参考标准的制定规则、范围、存储格式与数据规范。如果参考标准是可验证的,应描述参考标准的验证方式。

4.1.4.3 标注流程

数据集如具有标注信息,数据集说明应描述数据标注与质控流程,明确决策机制。在多人标注、多重标注的情况下,应描述标注分歧的仲裁机制。

4.1.4.4 其他标注信息

数据集如具有标注信息,数据集说明应描述除参考标准外的其他标注信息的范围、数据规范和存储格式。

4.1.5 数据集存储信息

数据集说明宜描述数据存储信息,如数据集存储方式与存储路径、安全控制、备份、恢复的说明。在数据集使用云服务存储的情况下,应提供云服务的提供商名称和资质、访问路径、使用权限说明等。

4.1.6 数据集用户访问

4.1.6.1 访问控制

数据集说明应描述用户访问控制机制,如用户类型、权限分配、授权机制。

4.1.6.2 访问条件

数据集说明应描述访问数据集需要的条件,如软硬件配置、访问方式、数据接口、协议、工具等。

4.1.6.3 可视化

数据集说明宜描述数据集信息的可视化呈现方式。

4.1.7 开发管理

数据集说明宜描述数据集开发过程依从的管理标准。

4.2 数据集标识

4.2.1 标识

数据集应显示唯一的标识,包含数据集名称、版本号、数据集制造责任方信息,可以随附文件的形式给出,可在数据集说明文档中进行详细描述。若数据集制造责任方向用户提供数据集管理软件/平台,应支持数据集标识的预览和查询。

数据集如发生变更,标识应同步变更。

当数据集作为医疗器械功能实现的一部分时,数据集标识可沿用产品标识。

4.2.2 数据集制造责任方信息

数据集说明宜包含数据集制造责任方的单位名称和联系方式。适当时,应陈述数据集制造责任方是否对数据集的使用提供技术支持和维护,以及技术支持和维护的具体内容。

4.2.3 文件命名与编码

数据集说明应陈述数据集中的文件命名和编码规则。

4.3 数据集质量特性描述

4.3.1 完整性

4.3.1.1 准确性

数据集说明应陈述数据与真实值一致的程度,可按照数据采集、信息记录、数据标注等环节展开,并以书面形式提供可验证的指标。

4.3.1.2 完备性

数据集说明应陈述数据集包含的信息能覆盖数据集的预期用途的程度,并以书面形式展示可验证的证据。

4.3.2 唯一性

数据集说明应陈述数据集、子集和数据在何种程度上是唯一的,并以书面形式提供可验证的指标。

4.3.3 一致性

数据集说明应陈述数据之间一致的程度,考虑内部一致性和外部一致性,并以书面形式提供可验证的指标。

4.3.4 确实性

数据集说明应陈述数据集在何种程度上是真实和可信的,包括数据、元数据的获取和处理,并以书面形式展示可验证的证据。

4.3.5 时效性

数据集说明应陈述数据集开发阶段的各个环节所需时限符合预期的程度,考虑预处理、清洗、标注等,并以书面形式提供可验证的指标。

4.3.6 可访问性

数据集说明应陈述数据集可被访问的程度,并以书面形式展示可验证的证据。

4.3.7 依从性

数据集说明应陈述数据集依从的标准规范、专家共识、操作规程或其他参考文献。

4.3.8 保密性

数据集说明应描述信息安全、数据保密相关的措施,并以书面形式展示可验证的证据。

4.3.9 资源利用性

数据集说明应陈述执行数据集相关任务需要的资源消耗,并以书面形式展示可验证的依从性证据,如访问、读取数据、预览、检索等任务需要的软件、硬件、网络配置。

4.3.10 精度

数据集说明应描述数据定量信息与真实值的接近程度,考虑数据元、元数据、数据标注结果,并以书面形式提供可验证的指标,如空间/时间分辨率、有效数字、最小测量单位。

4.3.11 可追溯性

数据集说明应描述数据可被追溯的程度,考虑数据采集历史、数据标注历史、数据访问踪迹和数据变更踪迹,并以书面形式展示可验证的证据。

4.3.12 可理解性

数据集说明应使用数据集的用户可理解的术语,对数据元、元数据和标注结果的含义提供解释,并以书面形式展示可验证的证据。

4.3.13 可得性

数据集说明应陈述数据集能被授权用户使用和检索的程度,并以书面形式展示可验证的证据。

4.3.14 可移植性

数据集说明应陈述数据集能被安装、替换或从一个系统移动到另一个系统中,并保持已有质量的属性的能力,考虑数据安装、替换、移动的效率,并以书面形式展示可验证的证据。

4.3.15 可恢复性

数据集说明应陈述数据集可被恢复的程度,并以书面形式展示可验证的证据。数据集说明可提供

用于数据恢复的措施。数据集说明可提供防止数据集使用过程中出现中断、失效的措施。

4.3.16 代表性

数据集说明应对样本的组成、比例、人群分布特征、数据的多样性和接近应用场景的程度开展分析，并以书面形式提供可验证的指标。

5 数据集质量要求

5.1 概述

本文件所述内容关注数据集的质量特性、整体风险等要素，宜根据数据集的预期用途、应用场景对数据集开展质量评价，形成技术报告，作为对数据集质量的验证。

5.2 质量特性

5.2.1 完整性

5.2.1.1 准确性

数据集应符合数据集说明有关准确性的陈述，如：

- a) 记录信息的准确性；
- b) 文字描述准确、清晰、无歧义；
- c) 数据标注的准确性；
- d) 采集设备的有效性；
- e) 人员操作的有效性；
- f) 数据形式的合理性。

5.2.1.2 完备性

数据集应包含实现数据集预期用途、满足应用场景需要的信息，如：

- a) 数据结构与格式应支持数据元、标注信息、元数据的保存、传输与访问，避免信息空白、遗漏或丢失；
- b) 对不可避免的缺失数据、离群值，应给出明确的处理方式，证明应用场景的合理性，对相关数据元加以标识；
- c) 元数据应描述数据集标识信息、内容信息、数据质量特性等相关信息；
- d) 数据包含的信息应支持数据标注与参考标准溯源的需要。

5.2.2 唯一性

同一数据集或子集内的数据元应是唯一的。同一数据集的各个子集应是可区分的。同一数据集内的数据唯一性宜以受试者个体为识别特征，对同一个体的不同时间、空间的数据应明确描述处理方法。

5.2.3 一致性

数据集应符合数据集说明有关一致性的陈述，包括内部一致的程度和外部一致的程度。

内部一致性是指相同来源数据之间的相关性，如：

- a) 同一数据元在数据集生存周期不同阶段应保持一致；
- b) 数据的特征应保持一致；
- c) 数据的采集、预处理、标注方法应保持一致；

d) 原始记录、中间记录与最终记录应保持一致。

外部一致性是指不同来源数据之间的相关性,如:

- a) 不同来源的数据应在数据特征方面保持一致;
- b) 离群值应是可解释的;
- c) 不同来源的数据在采集、标注环节应依从相同的法规、技术标准、医学规范、其他文献要求。

5.2.4 确实性

数据集应符合数据集说明有关确实性的陈述,如:

- a) 数据应来自真实的临床数据采集流程;适当时,数据采集涉及的设备、人员、方法应符合技术标准、临床规范或专家共识;
- b) 数据扩增、数据合成活动和结果应是可溯源和可解释的;
- c) 元数据应如实对数据进行描述。

5.2.5 时效性

数据的采集、标注、流转、归档、变更等活动的时限应符合数据集说明有关时效性的陈述。动态更新的数据集应明确数据的更新周期、更新方法和更新比例。若数据涉及临床诊断、治疗中的时序过程,应证明数据在临床时效上的合理性。

5.2.6 可访问性

数据集应满足数据集的预期用途和应用场景范围内的访问需要。

5.2.7 依从性

数据集应符合数据集说明有关依从性的陈述。

5.2.8 保密性

数据集应符合数据集说明有关保密性的陈述,采取措施确保其只能被授权用户访问。

隔离使用的数据集应具有数据集授权访问机制、隔离保护机制,应具有防止数据泄露、数据篡改、数据丢失的措施,如数据匿名化、物理隔离、数据审计等。

5.2.9 资源利用性

数据集的处理和使用应符合数据集说明有关资源利用性的陈述。

5.2.10 精度

数据集应符合数据集说明文档有关精度的陈述。

5.2.11 可追溯性

数据集应符合数据集说明有关可追溯性的陈述,具有相关记录,如:

- a) 原始数据来源、元数据来源、合规性证明;
- b) 数据采集活动记录;
- c) 人员管理记录;
- d) 数据标注流程记录;
- e) 盲态管理记录;
- f) 数据流通记录;

- g) 数据质疑、审计、停用、纠正记录；
- h) 标注工具、平台使用记录；
- i) 数据集标注结果的统计信息查询,包括标注进展、标签统计、标注者进展统计、难例集等；
- j) 数据服务异常与故障记录；
- k) 数据维护与备份记录；
- l) 数据更新记录；
- m) 云服务提供商名称、联系方式、云服务类型等。

5.2.12 可理解性

数据集应符合数据集说明有关可理解性的陈述。

5.2.13 可得性

数据集应符合数据集说明文档有关可得性的陈述。

5.2.14 可移植性

数据集应符合数据集说明有关可移植性的陈述。如果数据集允许在不同的平台和系统下使用,数据质量不应随平台和系统发生变化。

5.2.15 可恢复性

用于保持数据集质量并抵御失效事件的措施应符合数据集说明有关可恢复性的陈述。

5.2.16 代表性

数据集的数据特征层次、流行病学统计、样本来源多样性、数据多样性等应符合数据集说明有关代表性的陈述。

5.3 数据集风险分析

5.3.1 选择偏倚

数据集制造责任方应对数据集的代表性(受试者人群、采集场所、设备选择、参数设置等能否代表临床实际)进行分析。

5.3.2 覆盖偏倚

数据集制造责任方应对数据集病种组成能否覆盖应用场景下的目标人群全部情形(单发、多发、并发、复杂病理等)进行分析。

5.3.3 参考标准偏倚

数据集制造责任方应验证数据标注结果或参考标准的确立过程,分析参考标准偏倚。

5.3.4 验证偏倚

数据集制造责任方应描述参考标准的验证方法,分析数据标注过程与临床诊断过程的关联、差异和影响。

5.3.5 标注顺序偏倚

数据集制造责任方应描述数据标注流程设计,分析标注者分工、标注者工作顺序、记忆效应对标注

结果的影响。标注过程如使用计算机辅助识别(CAD)工具,应评估由 CAD 引起的偏倚情况,给出预防措施,避免 CAD 对人的干扰。

5.3.6 情景偏倚

数据集制造责任方应描述数据集患病率与临床目标人群患病率的差异。

5.3.7 其他偏倚

数据集制造责任方宜对其他偏倚进行分析,如分配偏倚、混杂偏倚等。

6 数据集质量符合性评价

6.1 通则

数据集质量符合性评价包括对数据集说明、质量特性和数据集风险分析文档的评价,评价流程如图 1 所示。测试前,应提供数据集、原始数据、元数据、标注工具、存储介质和其他工具的访问权限。数据集制造责任方可提供一个或多个文件,作为数据集说明;如提供多个文件,应建立索引。测试环境应符合数据集说明的要求。对数据集的各质量特性,可采用全体验证或抽样检验的方式,形成抽样样本集,作为评价对象。评价过程编写的测试用例宜包含测试目标、测试用例唯一标识、测试对象、测试步骤、测试环境、测试边界、预期的响应或产出、测试结果解释、用于判断测试用例是否通过的准则。数据集制造责任方应提交数据集风险分析文档,可参照 YY/T 0316—2016 进行编写。

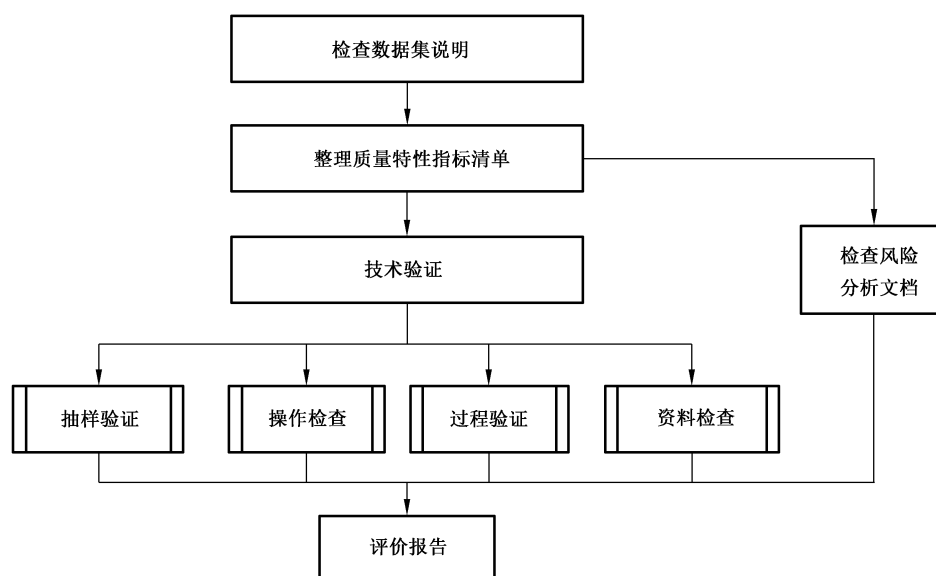


图 1 数据集质量评价流程图

表 1 给出了各质量特性推荐使用的评价方式。抽样检验方案宜根据具体质量特性的类型制定。属性数据宜采用计数型抽样检验,数据集制造责任方应以书面形式提供声称质量水平、使用方风险、生产方风险等指标,具体抽样方法按照 GB/T 2828.4、GB/T 2828.11 执行;变量数据宜采用计量型抽样检验,数据集制造责任方应以书面形式提供声称质量水平、接收常数、使用方风险、生产方风险等指标,具体抽样方法按照 GB/T 6378.4 执行。

表 1 质量特性分类与评价方式

质量特性	评价方式	抽样方式
准确性	试验验证、过程验证	计量型抽样、计数型抽样
完备性	试验验证+资料检查	计数型抽样
唯一性	查重验证	计数型抽样
一致性	试验验证	计量型抽样
确实性	操作检查	计数型抽样
时效性	试验验证	计数型抽样
可访问性	操作检查	—
依从性	资料检查	计数型抽样
保密性	资料检查	计数型抽样
资源利用性	试验验证	—
精度	试验验证	计数型抽样
可追溯性	资料检查	计数型抽样
可理解性	操作检查	—
可得性	操作检查	—
可移植性	操作检查	—
可恢复性	操作检查	—
代表性	统计分析	—

6.2 数据集说明评价

对数据集说明中的数据集描述内容进行检查,应完整、准确、无歧义,应符合 4.1 的要求;

对数据集说明中的数据集标识进行检查,适当时对相关的管理工具、可视化工具进行操作验证,应符合 4.2 的要求;

对数据集说明中的数据集质量描述进行检查,技术指标的定义和书面证据的展示应清晰、可验证,应符合 4.3 的要求。

6.3 质量特性评价

6.3.1 准确性

根据数据集说明有关准确性的陈述,开展抽样检验、过程验证,应符合 5.2.1.1 的要求。

6.3.2 完备性

根据数据集说明有关完备性的陈述,编写测试用例,对抽样样本集的数据、元数据和标注内容进行检查,应符合 5.2.1.2 的要求。

6.3.3 唯一性

对抽样样本集在总体样本中进行查重验证,结果应符合 5.2.2 的要求。

6.3.4 一致性

根据数据集说明声称的一致性指标,计算抽样样本集中符合要求的样本比例,应符合 5.2.3 的要求。

6.3.5 确实性

对照数据集说明有关确实性的陈述,对抽样样本集的数据、元数据、头文件和标注内容进行检查,应符合 5.2.4 的要求。

6.3.6 时效性

提取和检查抽样样本集的时间信息,应符合 5.2.5 的要求。

6.3.7 可访问性

根据数据集说明有关可访问性的陈述,编写测试用例,进行实际操作验证,应符合 5.2.6 的要求。

6.3.8 依从性

编写测试用例,检查抽样样本集与相关法规、技术标准、技术规范的符合性,应符合 5.2.7 的要求。

6.3.9 保密性

编写测试用例,检查抽样样本集的授权访问机制、数据的隔离保护机制等,应符合 5.2.8 的要求。

6.3.10 资源利用性

编写测试用例,在数据集制造责任方规定的运行环境下,对数据集进行操作验证,应符合 5.2.9 的要求。

6.3.11 精度

如适用,对数据集标注工具进行过程验证;检查抽样样本集的数据、元数据、标注结果包含的数据定量特征、数据集定量描述的精度,应当满足 5.2.10 的要求。

6.3.12 可追溯性

对抽样样本集可追溯的记录进行检查,应符合 5.2.11 的要求。

6.3.13 可理解性

对照数据集说明,实际操作,对数据进行预览和解释,应符合 5.2.12 的要求。

6.3.14 可得性

对照数据集说明,实际操作,对数据进行使用和检索,应符合 5.2.13 的要求。

6.3.15 可移植性

对照数据集说明,对数据集的安装、替换、转移进行实际操作,应符合 5.2.14 的要求。

6.3.16 可恢复性

对照数据集说明,编写测试用例,模拟使用过程中的失效事件,应符合 5.2.15 的要求。

6.3.17 代表性

对照数据集说明,对整个数据集开展统计分析,检查统计分层、样本组成、数据多样性等,应符合 5.2.16 的要求。

6.4 数据集风险分析评价

检查数据集制造责任方提供的风险分析文档,内容应考虑选择偏倚、覆盖偏倚、参考标准偏倚、验证偏倚、标注顺序偏倚、情景偏倚和其他偏倚;可采用专家评议法对数据集的整体质量、偏倚问题进行第三方问卷评估,作为补充。结果应符合 5.3 的要求。

附 录 A
(规范性)
数据集类型说明

数据集类型按照以下五个维度进行划分：

- 预期用途：模型训练与调优集、模型验证集、第三方性能测试集、临床评价数据集、产品质控集等；
- 数据来源：公有数据集、私有数据集、混合数据集等；
- 用户类型：自用数据集、他用数据集等；
- 访问管理方式：开放数据集、封闭数据集等；
- 更新形式：静态数据集、动态数据集等。

模型训练与调优集是指人工智能算法模型训练、调优阶段使用的数据集。该算法模型由人工智能医疗器械直接调用。

模型验证集是指医疗器械生产厂家的测试集，应独立于模型训练与调优数据集，避免交叉。

产品质控集是指人工智能医疗器械产品在临床使用阶段开展期间核查、日常质控等活动使用的数据集。产品质控集宜使用专用介质进行存储。

第三方性能测试集是指用于性能测试的第三方测试集。第三方性能测试集应确保测试数据与被测产品所用的训练、内部验证数据（包含被测产品通过预训练模型等方式间接使用的数据）不存在交叉；此类数据集仅用于封闭测试，在测试完成后应确保从被测产品和第三方系统中删除，避免针对性调优。第三方性能测试集应独立存储，与外界物理隔离。第三方性能测试集应具有防止未经授权访问的措施，并对访问活动形成记录。第三方性能测试集的数据来源应追溯至采集机构和标注者。

临床评价数据集是指人工智能医疗器械临床评价使用的数据集。临床评价数据集同样独立于医疗器械生产厂家，与被测产品所用的训练、内部验证数据（包含被测产品通过预训练模型等方式间接使用的数据）不存在交叉。临床评价数据集的存储、访问授权要求不低于性能独立测试集。临床评价数据集的数据来源应至少追溯至采集机构，在不违反伦理的前提下可追溯至受试者。

自用数据集是指数据集制造责任方为本机构建立的数据集，用于模型训练与调优、模型验证、产品质控等目的。自用数据集在数据自身不提供给其他机构的前提下，可由其他机构间接使用（如联邦学习场景）。

他用数据集是指数据集制造责任方直接提供给其他机构使用的数据集。他用数据集可能涉及供应商资质审核，其管理要求不在本文件讨论范围。

访问管理方式指的是数据集制造责任方是否允许外部用户访问。数据集制造责任方内部的访问管理可参考本文件进行设计。

由于数据集的数据停用后可能改做其他用途，为保障数据的有序使用，应定义数据流向的控制等级。本文件推荐按照模型训练与调优、模型验证、产品质控、性能独立测试、临床评价的顺序，由低到高设置控制等级。高等级数据集的数据停用后可流入低等级数据集，不宜从低等级流入高等级。

仅在研发早期阶段用于概念验证、对产品和算法模型无直接影响的数据集、非辅助决策应用相关的数据集属于特殊情形，可参照执行本文件的要求。

附 录 B
(资料性)
数据筛选与清洗说明

数据筛选考虑数据入选、排除和剔除等方面的要求，涉及数据合规性、受试者个体条件、数据质量、数据标签等维度。表 B.1 和表 B.2 分别给出了数据入选标准、排除标准的部分举例。

表 B.1 数据入选标准的举例

考虑维度	举例
数据合规性	数据的采集具有伦理审批相关手续； 来自前瞻性临床试验的数据具有知情同意书
受试者个体条件	年龄范围为 A 岁~B 岁； 身高为:C m~D m； 体重为:E kg~F kg； 所在地区为东北/华北/华东/华中/华南等； 所在社区为 G 疫情低/中/高风险地区； 是否具有某疾病病史:是/否
数据质量	图像覆盖 I、J 等生理结构； 眼底图像的视场角不低于 45°； CT 图像的层厚/层间距要求不大于 K mm； 生理信号采样率不低于 L Hz； 医学影像的对比度、分辨率、拍摄位置满足标注医生的要求
数据标签	疾病诊断分期或筛查分级:0/1/2/……； 病灶数量:M~N 个/人； 病灶尺寸范围:X mm~Y mm

表 B.2 数据排除标准的举例

考虑维度	举例
数据合规性	数据来源缺乏合规性证明
受试者个体条件	受试者超龄、超重等不符合应用场景的情形
数据质量	DICOM 头文件的字段缺乏必要的信息,例如 CT 图像的层厚； 医学影像不能覆盖需标注的区域； CT 影像出现漏层、错层； 文本信息具有语法错误,难以理解
数据标签	具有复杂的并发症,难以给出分类或分期； 受试者体内的植入物对数据采集带来不可接受的干扰,比如图像上的伪影或生理信号上的杂波

数据清洗是数据筛选的技术手段。图 B.1 给出了胸部 CT 肺结节影像数据清洗的部分步骤,作为示例。

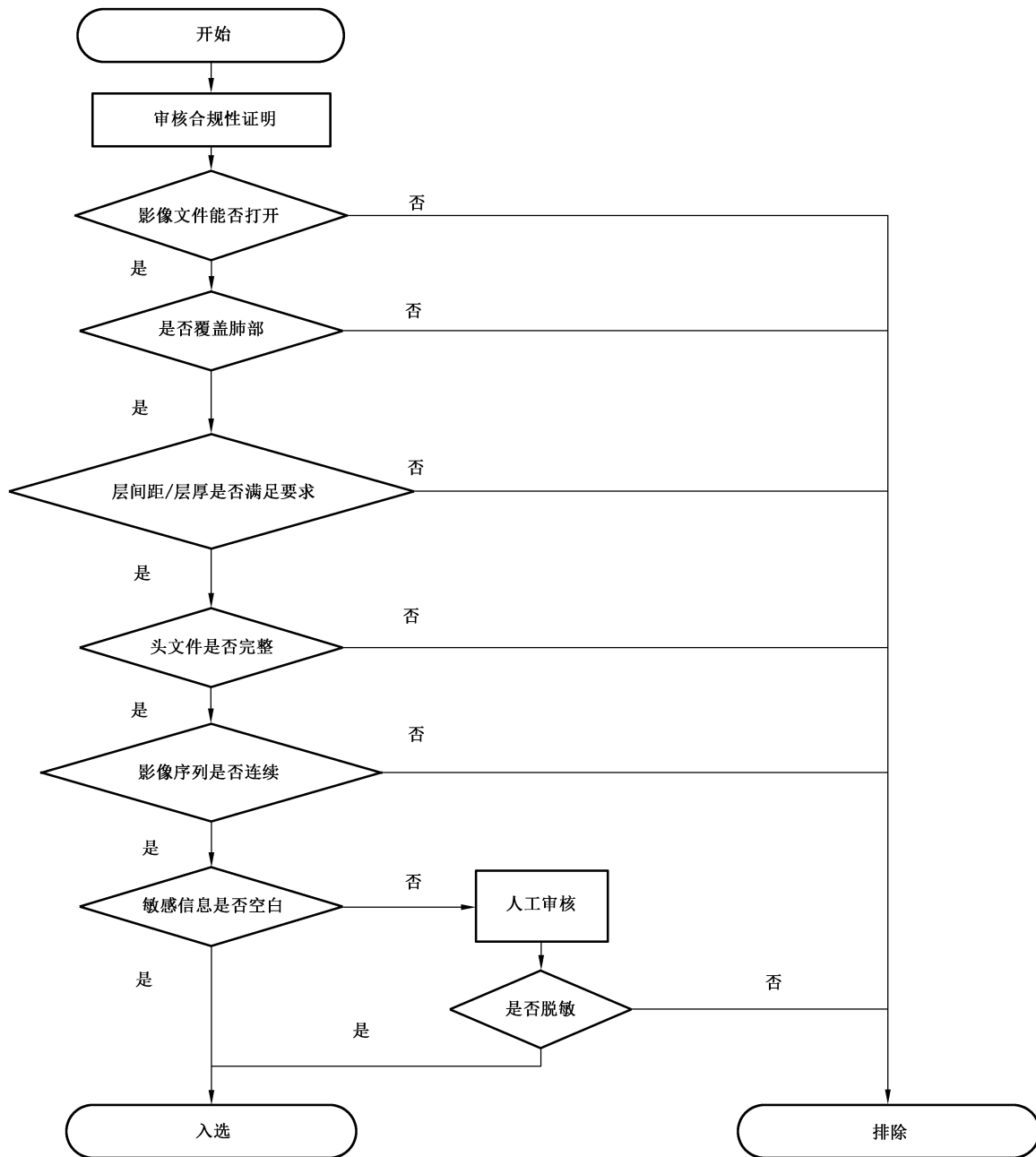


图 B.1 胸部 CT 肺结节数据清洗示例

参 考 文 献

- [1] GB/T 5271.1—2000 信息技术 词汇 第1部分:基本术语
- [2] GB/T 8054—2008 计量标准型一次抽样检验程序及表
- [3] GB 9706.1—2020 医用电气设备 第1部分:基本安全和基本性能的通用要求
- [4] GB/T 10111—2008 随机数的产生及其在产品质量抽样检验中的应用程序
- [5] GB/T 11457—2006 信息技术 软件工程术语
- [6] GB/T 18391.1—2009 信息技术 元数据注册系统(MDR) 第1部分:框架
- [7] GB/T 25000.12—2017 系统与软件工程 系统与软件质量要求和评价(SQure) 第12部分:数据质量模型
- [8] GB/T 34960.5—2018 信息技术服务 治理 第5部分:数据治理规范
- [9] GB/T 35295—2017 信息技术 大数据 术语
- [10] GB/T 36344—2018 信息技术 数据质量评价指标
- [11] WS/T 303—2009 卫生信息数据元标准化规则
- [12] WS/T 305—2009 卫生信息数据集元数据规范
- [13] YY/T 0287—2017 医疗器械 质量管理体系 用于法规的要求
- [14] YY/T 0316—2016 医疗器械风险管理对医疗器械的应用
- [15] ISO 3951-1:2013 Sampling procedures for inspection by variables—Part 1: Specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection for a single quality characteristic and a single AQL
- [16] ISO 19157:2013 Geographic Information—Data quality
- [17] CNAS GL-002:2018 能力验证结果的统计处理和能力评价指南
- [18] 国家食品药品监督管理总局. 总局关于发布临床试验数据管理工作技术指南的通告(2016年第112号)[Z]. 2016
- [19] 国家药品监督管理局医疗器械技术审评中心. 关于发布深度学习辅助决策医疗器械软件审评要点的通告(2019年第7号)[Z], 2019
- [20] 国家食品药品监督管理总局, 国家卫生和计划生育委员会. 《医疗器械临床试验质量管理规范》(第25号令)[Z], 2016
- [21] Mats Bergdahl, Manfred Ehling, Eva Elvers, et al. Handbook on Data Quality Assessment Methods and Tools[M]. European Commission, 2007
- [22] DAMA UK. The six primary dimensions for data quality assessment[R]. DAMA UK, 2013
- [23] Michael G. Kahn, Marsha A. Raebel, Jason M. Glanz, et al. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research[J]. Med Care. 2012 July;50 Suppl(0):S21-9. doi:10.1097/MLR.0b013e318257dd67
- [24] Nicole Gray Weiskopf, Chunhua Weng. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research[J]. J Am Med Inform Assoc 2013; 20:144-151. doi:10.1136/amiajnl-2011-000681
- [25] Lee K, Weiskopf N G, Pathak J, et al. A Framework for data quality assessment in clinical research datasets [C]. American medical informatics association annual symposium, 2017: 1080-1089

[26] David Boone, Suzanne Cloutier, Sergio Lins. Measuring the quality of HIV/AIDS client-level data using lot quality assurance sampling (LQAS) [R]. MEASURE Evaluation, University of North Carolina at Chapel Hill, 2019

[27] Organization for Economic Co-operation and Development. OECD Recommendation of the council on health data governance [R], OECD/LEGAL/0433, 2017

[28] World Health Organization. Data quality review: a toolkit for facility data quality assessment [M]. World Health Organization, 2017
