

《医疗行业专科高质量数据集建设 指南 第1部分：呼吸内科》 标准编制说明

中国信息通信研究院标准起草组

2026 年 7 月 29 日

1、标准范围

本文件规定了医疗行业呼吸内科高质量数据集建设的数据需求、数据规划、数据采集、数据预处理和数据标注等内容。

本文件适用于医疗机构、科研院所、医疗健康数据运营机构以及人工智能企业开展呼吸内科高质量数据集的设计、采集、处理、标注及维护等工作。

2、工作简况

本项目计划名称为“医疗行业专科高质量数据集建设指南 第1部分：呼吸内科”，由中国互联网协会归口。由中国信息通信研究院、广州医科大学附属第一医院牵头研制。

本文件起草单位包括广州医科大学附属第一医院、广州呼吸健康研究院、中国信息通信研究院、上海明品医学数据科技有限公司、北京大学第一医院、内蒙古自治区人民医院、中国医科大学附属第一医院、吉林大学第二医院、复旦大学附属中山医院、浙江大学医学院附属第二医院、郑州大学第一附属医院、新疆维吾尔自治区人民医院、兰州大学第一医院、四川大学华西医院、云南省第一人民医院、复旦大学附属儿科医院、上海人工智能实验室，熵极涌现数据（天津）有限公司。

本文件主要起草人钟南山、李时悦、黄超仪、陈智、陆殷、苏越明、简文华、王文熙、王柱锋、钟长镐、韩林月、桑新雨、武雅文、冯天宜、闵栋、姚娟娟、高柳村、王少伟、梁志龙、潘昊天、马靖、廖纪萍、高笑宇、马尚寅、李文扬、张捷、宋元林、李雯、王华启、魏雪梅、包海荣、刘丹、张云辉、张晓波，葛小玲，马傲、徐捷，曲通。

本文件于2026年7月在中国互联网协会通过立项申请。

起草组于2026年5月召开线上讨论会，结合医疗AI大模型训练的实际需求与呼吸内科重点疾病临床路径的特点进行了深入讨论，充分吸收了“医工交叉”各方专家的合理建议，最终形成了符合行业应用要求的标准文件。

3、标准编制原则和确定标准主要内容的依据

- 标准性要求，充分借鉴国内外医学指南与健康数据安全规范，结合我国“国家呼吸医学中心”建设实际情况，力求标准的可执行性；

- 实用性要求，建立在对医疗AI大模型应用场景充分调研的基础上，确保核心数据元采集契合临床实际；
- 可行性要求，数据需求根据预期人工智能应用确定，应用场景包括疾病诊断、治疗决策、检查解读、预后与随访、紧急情况识别；
- 有效性要求，兼容多模态数据（文本、影像CT/MRI、心电波形等），满足未来大模型技术发展的扩充需求。

4、主要试验（或验证）的分析、综述报告

本标准的技术内容从呼吸内科临床真实业务流与AI大模型训练需求出发。经过对医疗机构EMRS/PACS等信息系统以及AI算法企业数据处理方案的调研与分析，在核心数据元提取、多模态数据标注规则等方面，得到了临床医学专家和AI技术专家的共同认可。标准可为呼吸内科AI数据集的建设与评估提供可靠的依据。

5、标准在起草过程中遇到的问题及解决办法；重大分歧意见的处理经过和依据；有无重要技术问题需要说明

在标准起草过程中遇到的问题通过标准编制组内部协调和讨论已经解决。无重大分歧意见。没有重要技术问题需要说明。

6、与国外标准的关系：包括：采用国际标准和国外先进标准的程度，与国外标准主要技术内容的差异（可引用标准前言的内容）

本标准无对应的国际标准或国外标准。

7、修订标准时，说明与标准前一版本的重大技术变化，并列岀所涉及的新、旧版本的有关章条（可引用标准前言的内容）；废止/代替现行有关标准的建议

本标准为第一版制定标准。

8、说明标准与其他标准或文件的关系（可引用标准前言的内容），特别是与有关的现行法律、法规和强制性国家标准的关系

标准向上承接《GB/T 39725-2020 信息安全技术 健康医疗数据安全指南》、《WS 445-2014 电子病历基本数据集》等现有国家及行业基础标准，向下细化呼吸内科数据需求、架构、采集、预处理与标注要求。

9、标准作为强制性标准或推荐性标准的建议

建议本文件作为推荐性团体标准发布实施。

10、贯彻国家标准的要求和措施建议（包括组织措施、技术措施、过渡办法等内容）；标准发布后，对国内外业界可能产生的影响

本标准的发布与实施可协调产业共识，形成呼吸内科高质量数据集的统一建设规范。不仅能极大加速国内医疗行业专科智能体的研发周期，还将促进优质医疗资源以数字化的形式下沉，助力基层医院急救水平的提升。

11、标准是否涉及知识产权的情况说明；如标准中含有自主知识产权，说明产品研发程度、产业化基础及进程

本文件不涉及知识产权的问题。

12、其他应予说明的事项

无。