DOI:10.3969/j.issn.1004-3845.2024.07.001

· 专家共识 ·

人工智能囊胚形态评估数据集构建与质控专家共识

王浩，张孝东，孙莹璞，孙海翔，邓成艳，黄学锋，刘平，周灿权，冯云，郝桂敏，卢文红，沈浣，师娟子，张松英，滕晓明，王晓红，王秀霞，伍琼芳，全松，曾勇，钟影，邵小光，柯林楠，毛歆，韩倩倩*，黄国宁*

（中华医学会生殖医学分会第五届委员会；中国食品药品检定研究院）

【摘要】 囊胚形态人工智能（AI）评估是 AI 医疗器械发展的新兴方向，也是 AI 在辅助生殖领域的重要应用。AI 在新领域应用的起步阶段，数据集的构建与质控对产品质量有重要影响。目前，囊胚形态学 AI 评估在数据采集、标注、质控等方面尚未形成统一的规范。在参考 AI 医疗器械、辅助生殖医疗器械现有国家行业标准的基础上，本文以囊胚形态 AI 评估数据集为主题，对数据集构建与质控要求进行了探讨，对数据集质量特性进行了解析，旨在指导数据集制造责任方加强数据集全生命周期管理，更好地为产品研发、测试、临床试验等环节提供质量保障，助力产业发展。

【关键词】 人工智能（AI）； 囊胚形态评估； 数据集构建； 数据集标注； 数据集质量控制

【中图分类号】 R321.2；TP391 **【文献标识码】** A

Expert consensus on the construction and quality control of datasets for artificial intelligence（AI） assisted blastocyst morphometric assessment

WANG Hao，ZHANG Xiao-dong，SUN Ying-pu，SUN Hai-xiang，DENG Cheng-yan，HUANG Xue-feng，LIU Ping，ZHOU Can-quan，FENG Yun，HAO Gui-min，LU Wen-hong，SHEN Huan，SHI Juan-zi，ZHANG Song-ying，TENG Xiao-ming，WANG Xiao-hong，WANG Xiu-xia，WU Qiong-fang，QUAN Song，ZENG Yong，ZHONG Ying，SHAO Xiao-guang，KE Lin-nan，MAO Xin，HAN Qian-qian，HUANG Guo-ning**

The Fifth Session of the Committee of Chinese Society of Reproductive Medicine；National Institutes for Food & Drug Control

【Abstract】 Computer assisted assessment of blastocyst morphology is an emerging direction in the artificial intelligence（AI） medical devices and an important application of AI in the field of assisted reproduction. In the initial stage of the application of AI in new fields，the construction and quality control of data sets have an important impact on product quality. At present，AI-assisted blastocyst morphology assessment has not yet formed a unified specification in terms of data collection，labeling，and quality control. Based on the existing national industry standards for AI medical devices and assisted reproduction medical devices，this paper discusses the requirements for data set construction and quality control and analyzes the quality characteristics of data sets with the theme of blastocyst morphology assessment datasets，with the aim of guiding data set manufacturers to strengthen the management of datasets in the whole life cycle，and to better provide quality assurance for the product research and development，testing，and clinical trials in order to help the development of the industry.

【收稿日期】 2024-04-02

【基金项目】 重庆市技术创新与应用发展专项项目（CSTB2022TIAD-KPX0146）；国家自然科学基金面上项目（82371728）

【通讯作者】 韩倩倩，中国食品药品检定研究院（Email：hanqianqian@nifdc.org.cn）；黄国宁，重庆市妇幼保健院、重庆医科大学附属妇女儿童医院（Email：gnhuang217@sina.com）

【Key words】 Artificial intelligence (AI); Blastocyst morphology assessment; Data set construction; Data set annotation; Data set quality control

(*J Reprod Med* 2024,33(7):843-851)

随着深度学习等新一代人工智能(AI)算法的发展,用于囊胚形态分析的智能医疗器械(独立软件、软件组件等)的研发活动日益活跃,其在辅助生殖医学领域的应用也越来越广泛,产品预期用途包括胚胎图像分割、测量、评级、临床结局预测等[1-5]。为了促进产品的研发、测试与临床评价,国外相关机构以囊胚形态 AI 评估为主题,积极开展数据集建设,为行业发展提供支撑[6]。目前,囊胚形态 AI 评估数据集的开发建设还没有形成专用的、系统性的标准规范;在数据标注方面,Gardener 评分[7]的应用比较广泛,但标注人员的培训、分工等细节也缺乏统一规则。目前,国内囊胚形态 AI 评估数据集的发展刚刚起步,在执行层面容易出现差异,影响数据质量,进而制约算法性能和产品质量。数据集的建设需要与标准规范同步发展,在行业共识基础上推进。

近年来,我国的辅助生殖医疗器械行业标准、AI 医疗器械行业标准体系初具规模,现有标准涵盖了囊胚染色与计数[8]、数据集通用质量评价[9]、数据标注通用质量评价[10]等主题,为建立囊胚形态 AI 评估数据集的专用规范提供了参考。AI 产品算法性能测试、生产研发可追溯性标准也明确了数据集使用与管理的定位[13-14]。国内的 AI 医疗器械注册审评相关技术文件也多次强调了数据集的重要性[15-16]。在此背景下,本文以囊胚形态 AI 评估数据集构建、质量控制及评价的具体问题为导向,对数据集质量的表现形式进行解析,对囊胚形态 AI 评估智能产品相关的数据集构建过程给予示范及引导,旨在引导本领域数据集的科学有序发展。

数据集说明文档要求

根据 YY/T 1833.2(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)[9]的要求,数据集制造责任方应当建立说明文档,供监管和用户了解数据集。依据 YY/T 1833.2 的定义,数据集制造责任方指的是对某个数据集的设计、制造负有责任的实体(中国境内收集数据的责任方应当为获批开展辅助生殖技术的医疗机构,而科研机构、生产企业提供需求和技术支持)。

一、数据集基本信息

参考现有卫生行业标准[15]的定义方式,本文将囊胚形态 AI 评估数据集定义为:以体外培养的胚胎显微图像为主题、可以标识并可以被计算机化处理的数据集合。根据医疗器械行业标准(YY/T 1833.2)[9],囊胚形态 AI 评估数据集的说明文档应声明数据集的类型,按照预期用途、数据来源、用户类型、访问管理方式、更新形式等维度进行划分。

根据目前相关 AI 产品的研发现状,构建囊胚形态 AI 评估的数据集的影像数据格式可能包括 avi、rm、rmvb、flv、mpg、mov、mkv 或二次视频分解的 jpg、tiff、bmp、gif、ufo、exif、raw,数据来源于辅助生殖中心胚胎实验室获取的胚胎体外培养真实图像。根据产品研发的需要,需采集患者的临床和胚胎实验室信息,以配合 AI 诊断等功能的实现。采集图像的胚胎时差培养箱及显微镜成像系统应具有典型性,能够代表不同地区、不同临床机构的装备水平。构建的数据集适用于囊胚形态 AI 分析产品的训练、测试等,产品预期用途包括胚胎发育辅助分析、囊胚形态 AI 评级等。

囊胚形态 AI 分析数据集的标注对象包括透明带、卵周隙、内细胞团(inner cell mass,ICM)、滋养外胚层细胞(trophectoderm,TE)、囊胚腔(blastocyst cavity,BC)。同时,数据集制造责任方可根据产品预期使用的人群特征,对其临床阶段性[囊胚分级、胚胎种植前遗传学检查、囊胚种植、早期胚胎丢失及最终治疗结局(活产)]进行分类,作为算法训练的标签或测试的基准。

囊胚形态 AI 分析数据集的内容除影像数据、标注结果外,还包含 YY/T 1833.2 要求的元数据。数据集制造责任方需要为数据集分配名称、版本号,与数据的更新保持同步。

二、数据采集

(一)伦理批准与患者隐私保护

尽管囊胚形态影像数据来自体外培养,隐私保护的要求仍然适用,要求数据集制造责任方开展伦理审查,审查范围包括体外胚胎各阶段的原始图像、视频、患者自身的年龄、病史、临床干预等流行病学

信息，以及其他相关的临床数据、信息资料等。患者的知情权、同意书、补偿等应当满足法规的要求。数据集制造责任方在启动数据收集之前，应提请伦理委员会审批，或通过同等效力的批准程序保证数据脱敏，保障患者隐私安全和患者利益（表 1A-数据合规性及入选标准）。

表 1A 数据合规性及入选标准

维度	示例
数据合规性	数据的采集具有伦理审批相关手续； 来自前瞻性临床实验数据具有知情同意书； 每个培养孔按照溯源性原则，只能培养 1 个胚胎，即每样本只能出现 1 枚胚胎。
人群代表性	女性年龄范围为 20～37 周岁（20～35 岁抽样 75%，35～37 岁抽样 15%，≥38 岁抽样 10%）；男性年龄范围为 22～48 周岁； 体质量指数（BMI）为 18.5～27.0 kg/m^2； 女性来源地区应具有多样性：单一地区人群＜70%； 疾病种类：盆腔输卵管因素、排卵障碍、子宫内膜异位症、男性因素、不明原因及免疫性不孕； 获卵数 5～15 枚； 技术治疗类型：IVF、ICSI、PGT。
数据质量	数据来源于真实体外胚胎图像/视频； 光学放大率不低于 400×，图像矩阵不低于 512×512 像素，成像视野不低于 A μm×B μm，三维成像的层数不低于 7； 图像覆盖－75 μm～＋75 μm 的胚胎层面； 数据分析集图像为胚胎图像最清晰层面； 数据记录间隔为 5～15 min/次； 数据记录时间为 IVF 受精后 4～6 h 或 ICSI 受精后的 2～4 h。
数据标签	囊胚形态学特征：透明带、卵周隙、极体、卵膜、原核、核仁、卵裂球、胚胎碎片； 胚胎发育结果：正常卵裂模式胚胎、异常卵裂模式胚胎、可移植囊胚、不可移植囊胚、塌缩囊胚、染色体正常胚胎、染色体异常胚胎； 胚胎临床结果：种植、未种植、HCG 阳性、生化妊娠、临床妊娠、继续妊娠、活产、死产、流产。

（二）数据脱敏、清洗、查重要求

1. 数据脱敏：体外胚胎的原始图像、视频、数据集的元数据均不应包含与患者隐私有关的任何信息。敏感信息的判定可参照 GB/T 35273—2020 标准的规定，常见情形包括患者身份信息、临床病史、社会经济状况、家庭情况、财务信息等。数据集制造责任方在解析体外胚胎图像/视频时，应明确脱敏范围和具体字段。

数据脱敏的过程不应改变图像、视频的灰度信息，除非敏感信息直接存储于灰度矩阵，例如图像水印或图像周围显示的患者名称、年龄等。数据的采集、传输、保存和使用必须符合《中华人民共和国网络安全法》、《科技部人类遗传资源管理办法》和《医疗器械网络安全注册技术审查指导原则》等法律法规的要求。

胚胎数据集标签应按照“Istanbul”胚胎共识[17]或 Gardener 评分[7]进行分类。基础分类为优质胚胎、可移植胚胎（亚标签可为“移植”或“冷冻”）及丢弃胚胎；对附加临床决策的数据集可追加“胚胎种植”、“生化妊娠”、“临床妊娠”、“继续妊娠”“累计妊娠”“累计继续妊娠”及“流产”等（表 1A-数据标签）。

2. 数据清洗：数据集制造责任方应预先规定数据纳入/排除要求，作为数据清洗的依据。图像的质量要求一般包括格式的有效性、单个文件的完整性、视频的连续性、图像内容的合理性等。例如，排除培养条件异常导致的非正常培养条件下获得的数据资料、视野不完整的图像、破损或无法读取的图像文件、缺失关键帧的视频、有遮挡或污损的图像，确保每个囊胚的图像、视频保持连续完整，排除出现缺层、错层等情况的三维图像。数据集制造责任方可根据需要，制定更具体的数据清洗规程。未通过数

据清洗的数据应在受控条件下存储，避免泄露或误用。如采用 AI 算法进行辅助清洗，应对结果进行人工审核（表 1B）。

表 1B　胚胎数据排除标准

维度	示例
数据合规性	数据来源缺乏合规性证明； 每个培养孔出现超过 1 枚胚胎的图像数据。
人群代表性	受试夫妇年龄及 BMI 不符合应用场景的情况。
数据质量	数据文件缺乏必要的描述信息，如患者来源信息； 胚胎数据影像出现漏层、无清晰层面等； 患者胚胎数（影像资料）与实际胚胎数量差异超过＞50％； 文本信息存在语法错误，难以理解。
数据标签	无正常发育胚胎，及全部胚胎冷冻未有相关随访结局的； 患者所有胚胎图像均不清晰的。

3. 数据查重：为保证数据的唯一性，数据集制造责任方应开展查重验证，避免与外部的数据集发生数据重合，避免同一病例的数据重复出现，并剔除重复样本。

4. 数据储存与传输：采用安全可靠的数据存储设备，如服务器和云存储，以确保大量数据的安全保存和备份。同时，应当规定数据传输的加密标准，保障数据在采集和传输过程中的隐私和安全。

（三）数据采集与多样性要求

数据集应当使用视频或二次视频提取的真实体外胚胎培养全过程的视频，光学放大率不低于 400×，成像视野覆盖整个胚胎；如进行三维成像，图像层数足以支持对关键帧的选取。数据集制造责任方应确保数据采集前后的完整性，避免有损压缩、图像滤镜等情形。每个胚胎的采集时间建议从受精后 4～6 h 内开始收集至 120～144 h（采集层数至少＞7，每小时收集＞4 次）（表 1A-数据质量）。

为确保数据集的多样性，在数据采集阶段需要尽可能地覆盖到更多具有通用性的统计维度，以降低数据集的覆盖偏倚。这些维度包括以下：

1. 患者维度：主要应考虑患者的年龄、性别、BMI、生育史、疾病史、地区、职业等因素，这些因素与不孕症存在联系，因此对于数据集的临床代表性有重要影响，有助于确保模型对不同群体的囊胚形态具有较好的泛化能力。患者人群分布应参考流行病学统计进行均匀随机抽样（按临床剩余需求年龄分布，20～35 岁抽样 75％，35～37 岁抽样 15％，≥38 岁抽样 10％）的形式进行分层抽样（表 1A-人群代表性）。

根据患者不孕症病因，设置数据集的样本量和比例。作为对真实临床数据的抽样，数据集的数据容量决定了抽样误差。抽样误差越小，数据集越有代表意义。关于抽样误差的计算，可以参照原国家食品药品监督管理总局发布的《医疗器械临床试验设计指导原则》给出的方法进行计算，在条件允许的情况下尽量提高样本量。根据《医疗器械临床试验设计指导原则》，按分层抽样的方式计算样本。

样本抽样的估算公式为：$N=\frac{\mu_{\alpha}^{2}\times\rho(1-\rho)}{\sigma^{2}}\times deff$，其中 μ_{α} 为设定置信区间的统计量（如 α 设定为 0.05 时，$\mu_{\alpha}=1.96$）；ρ 为目标总期望值（百分率值，如可移植胚胎率 45％）；σ 为允许绝对误差（等于允许相对误差乘以目标总期望值，一般允许相对误差为 15％，$\sigma=\rho\times15\%$）；$deff$ 为分层抽样的层数。应对同一患者的多个试管婴儿周期进行数据采集，以考虑患者在不同周期之间的变化和囊胚质量的差异。这有助于建立更全面的囊胚形态模型。

临床结果的多样性：在数据集中引入不同的临床结果，包括成功的试管婴儿案例和失败的案例。这有助于模型对成功和失败案例的区分，提高其临床应用的准确性。

实验室多样性：数据集应覆盖不同生殖医学实验室的数据，考虑到实验室环境可能对囊胚形态评估有影响。不同实验室的技术水平和实验室操作流程的变化都应考虑在内。

2. 设备及数据采集技术标准维度：设备方面主要应考虑胚胎图像采集设备（例如时差培养箱）制造厂家、设备型号、成像参数设置（培养箱工作环境设置、成像层间距、记录像素等）的合理性与多样性。这些因素影响图像的对比度、分辨率、信噪比、细节丰富程度等基本参数，同时也会影响数据标注结论，如对胚胎的评级、分类及后续量化测量等。从操作层面来说，可以依据表 2A、2B 的参数范围进行选择。

表 2A 胚胎图像采集设备工作环境

数据元标识符(DE)	数据元名称	定义	数据元允许值
TL01. 01. 001	温度控制范围	TL 工作时的温控系统控制的温度范围	+36. 5℃～+37. 5℃
TL01. 01. 002	温度控制精度	TL 工作时的温控系统的温度控制精度	±0. 1℃
TL01. 01. 003	温度波动范围	TL 工作时的温控系统温度控制波动范围	±0. 1℃
TL01. 01. 004	恢复温度时间	TL 开仓后的温控系统控制温度恢复到设定的时间	60 s
TL01. 01. 005	温度设定值	TL 工作时的温控系统控制的培养舱室温度	37℃
TL01. 02. 001	CO_2浓度控制范围	TL 工作时的气控系统控制的 CO_2浓度范围	0%～15. 0%
TL01. 02. 002	CO_2浓度控制精度	TL 工作时的气控系统 CO_2浓度控制精度	±0. 1%
TL01. 02. 003	CO_2浓度波动范围	TL 工作时的气孔系统 CO_2浓度控制波动范围	±0. 2%
TL01. 02. 004	恢复 CO_2浓度时间	TL 开仓后的气控系统控制的 CO_2浓度恢复到设定的时间	＜90 s
TL01. 02. 005	CO_2浓度设定值	TL 工作时的气控系统控制的培养舱室 O_2浓度	6%
TL01. 03. 001	O_2浓度控制范围	TL 工作时的气控系统控制的 O_2浓度范围	2. 0%～17. 0%
TL01. 03. 002	O_2浓度控制精度	TL 工作时的气控系统 O_2浓度控制精度	±0. 1%
TL01. 03. 003	O_2浓度波动范围	TL 工作时的气孔系统 O_2浓度控制波动范围	±0. 1%
TL01. 03. 004	恢复 O_2浓度时间	TL 开仓后的气控系统控制的 O_2浓度恢复到设定的时间	＜90 s
TL01. 03. 005	O_2浓度设定值	TL 工作时的气控系统控制的培养舱室 O_2浓度	5%
TL01. 04. 001	光源颜色	TLI 工作时的光源颜色	红光
TL01. 04. 002	光源波长	TLI 工作时的光源波长	635 μm
TL01. 04. 003	曝光量	每天(24 h)每个胚胎总曝光时长	＜4 s
TL01. 05. 001	胚胎成像分辨率	胚胎成像后的图形分辨率	＞500 pixel
TL01. 05. 002	胚胎成像时长	胚胎单次成像时长	单位 s
TL01. 05. 003	胚胎成像焦距	胚胎成像探头使用的焦距	15 μm
TL01. 05. 004	胚胎成像工作距离	胚胎成像探头的工作距离	15 μm×7
TL01. 05. 005	胚胎图像分辨率	预处理的胚胎图像分辨率	800×800

表 2B 胚胎图像采集设备运行采集数据

数据元标识符(DE)	数据元名称	定义	数据元允许值
TL02. 01. 001	采集温度	特定胚胎所在 TL 舱室温度	＜ 24 h
TL02. 01. 002	采集温度频次	特定胚胎所在 TL 舱室采集温度的频次	＜30 s
TL02. 01. 003	采集 CO_2浓度	特定胚胎所在 TL 舱室 CO_2浓度	＜ 24 h
TL02. 01. 004	采集 CO_2浓度频次	特定胚胎所在 TL 舱室采集 CO_2浓度的频次	＜30 s
TL02. 01. 005	采集 O_2浓度	特定胚胎所在 TL 舱室 O_2浓度	＜ 24 h
TL02. 01. 006	采集 O_2浓度频次	特定胚胎所在 TL 舱室采集 O_2浓度的频次	＜ 24 h
TL02. 02. 001	采集图像	特定胚胎所在皿培养成像图像	JPEG
TL02. 02. 002	采集图像次数	特定胚胎所在皿培养成像次数	＜ 20 min
TL02. 02. 003	采集图像焦距	特定胚胎所在皿培养成像探头焦距	＜ 20 min
TL02. 02. 004	采集图像成像时间	特定胚胎所在皿培养成像时间	时间标准格式,精度到 s
TL02. 02. 005	采集图像胚胎龄	特定胚胎所在皿培养成像时胚胎龄	0～168 h
TL02. 02. 006	采集图像文件名	特定胚胎所在皿培养成像后存储文件名	每更换一次培养胚胎

（四）数据标注

数据标注人员的能力、数据标注结果的质量、标注流程的一致性、标注过程的质量控制体系都会直接影响到囊胚形态数据集参考标准的准确性，从而影响 AI 模型的临床可靠性。因此，囊胚图像或视频应由从事胚胎工作 5 年以上的资深胚胎学家标注，再交由高级职称专家审核、分类描述数据集内囊胚实际的形态特征与分布，并为数据标注相应的特征值及数据标签（表 3），确保标注的一致性和可追溯性。另，不同发育天数（D5、D6、D7）形成的囊胚可能存在临床结局的差异，因此囊胚评级应对应发育天数的标签。

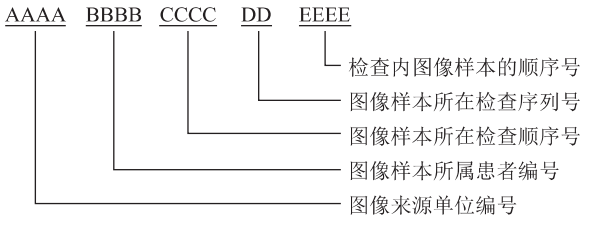
表 3　囊胚的标签分类

指标	分类	描述
囊胚腔大小	1	早期囊胚，囊胚腔室小于胚胎总体积的 1/2
	2	囊胚腔体积大于或等于胚胎总体积的 1/2
	3	完全扩展囊胚，囊胚腔完全占据了胚胎的总体积
	4	扩展囊胚，囊胚腔完全充满胚胎，胚胎总体积增大，透明带变薄
	5	正在孵出的囊胚，囊胚的一部分从透明带中逸出
	6	孵出的囊胚，囊胚全部从透明带中逸出
内细胞团	A	细胞数目多，排列紧密
	B	细胞数目少，排列松散
	C	细胞数目很少
滋养外胚层细胞	A	上皮细胞层由较多的细胞组成，结构致密
	B	上皮细胞层由不多的细胞组成，结构松散
	C	上皮细胞层由稀疏的细胞组成

三、数据样本溯源信息记录

为了提升数据管理的规范性、确保数据集在 AI 医疗器械生产质量管理体系中的有序流转，数据样本需要建立唯一标识；编码时可考虑地域、临床机构、患者、胚胎编号、采集时间等要素。具体原则可参照医疗器械行业标准《医疗器械唯一标识基本要求》[18]。

对单个影像文件，唯一标识可采用的字段举例如图 1，其信息应与原始文件形成映射关系，便于检索和预览。当数据发生更新时，二者保持同步。唯一标识的编码方式应当进行校验，以保证整体或组成部分的正确性和唯一性。



该示例从大到小排列，可以识别的信息包括图像采集的单位、患者编号（脱敏后）、文件所在序列的检查顺序、文件所在的检查序列、文件在序列中的顺序，从而实现对单一影像文件的溯源。数据集制造责任方可以参照补充其他字段，丰富标识信息。各字段之间可用“—”作为数据分隔符，帮助正确识读和解析各个字段。

图 1　唯一标识字段命名字段示例

数据集质量评价

在数据集的质量验收阶段，数据集制造责任方宜采用医疗器械行业标准 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》[9]的框架，开展相关试验。

一、质量特性与评价方式

1. 准确性：关注数据集的信息与“真值”的接近程度，包含数据采集、标注层层面，例如对时差培养设备的有效性、人员操作的有效性进行检查，对影像报告、标注结果的正确性、数据形式的合理性进行抽查[9]。可能情况下，对囊胚影像标注准确性的抽查建议由专业的第三方医学专家团队进行，第三方团队的资质、从业年限、检查流程和分歧处理应有明确的要求。囊胚数据集的抽样检查可以把单个囊胚发育序列作为基本单元，例如先计算每个序列的准确率，进而对整个数据集的准确率进行统计估计，因而

适合采用计量型抽样检验方法。标注人员的客观表现宜列入准确性的考量范围，例如以仲裁人员作为参考标准，计算标注人员的分类准确率，应符合数据集制造责任方的声称。

2. 完备性：囊胚训练数据集应包含支持产品训练、满足临床适用场景需要的信息，例如囊胚期级别，时差培养箱设备型号、设备厂家、图像及视频采集参数等均需要接入医院信息系统（Hospital Information System，HIS）或辅助生殖技术专病系统；以及伦理批准使用的非敏感信息，例如受试者年龄、不孕症病因临床干预、预后等。数据集制造责任方可制定具体的信息列表，对信息完备性进行抽查。由于囊胚序列都可以明确其是否具有完备性，因此可使用计量型抽样检验方法。

3. 唯一性：用于判定同一数据集内的数据元是否唯一，相当于对数据清洗中的查重进行验证。本部分适合计量型抽样检验。

4. 一致性：YY/T 1833.2 对于内部一致性、外部一致性的要求适用于囊胚形态数据集，意味着对于同一数据集而言，来自不同培养体系的数据在采集、预处理、标注等环节应依从相同的法规、标准、规则。对一致性的符合性判定一般采用计量型抽样检验。

5. 确实性：囊胚形态数据集应采用临床真实数据，对可疑样本进行排除，如错误引入动物胚胎实验数据、数据污染等情形。对确实性的评价可采用计量型抽样检验。

6. 时效性：囊胚图像考虑到临床的实际操作，为完整观察到合子阶段的发育特征，建议从受精后 4～6 h 内开始收集至 120～144 h（采集层数至少＞7，每小时收集＞4 次）；数据集的开发建设应在声称的时限内完成，以保证数据集符合当前的医学认知和产品开发需求。时效性的评价需要从数据集的过程记录中提取时间信息，计算实际时限，其符合性属于计数型抽样检验范畴。

7. 可访问性：YY/T 1833.2 对于可访问性的要求适用于囊胚形态数据集，客观上要求数据集制造责任方具有数据访问控制的措施，例如用户权限、数据访问授权机制。本质量特性一般通过操作检查进行判定。

8. 依从性：囊胚数据集的标注活动应依从 Istanbul 共识[17]、Gardener 评分[7]，以开展胚胎分类；数据集元数据字段设置应符合囊胚观察和评价的各项定义[20]。这些文献应体现在数据集的文档描述中，因此对依从性的评价需要对数据集文档、标注结果、过程记录进行检查。

9. 保密性：由于囊胚数据集可能包含受试者的信息，数据集制造责任方应防止囊胚影像、标注结果、元数据等信息的泄露，避免数据被篡改、盗用等问题的发生，形成相关记录。对保密性的评价可采用过程验证、文件记录审核等方式进行。

10. 效率：关注数据集的用户调用数据集的速度，体现了数据集作为一种“产品”对使用环境的要求。效率的评价可以采用在数据集制造责任方规定的软硬件与网络环境下，实际读取、传输数据集，验证操作的时间。

11. 精度：关注囊胚影像数据定量特征、数据集总体定量特征、囊胚标注结果等误差大小的程度，例如囊胚径线测量的精度可用微米表述。对精度的验证可以采用比对试验、工具验证等方式实现。

12. 可追溯性：可追溯性关注囊胚数据集的全生命周期中，质量管理活动是否形成记录。可追溯性的评价主要通过对文档和记录进行检查，要求医院建立数据采集活动记录、标注人员选拔与培训记录、数据标注流程记录、标注工具使用记录等。

13. 可理解性：关注囊胚数据集能被授权用户预览和解释的程度，例如能否将囊胚标注结果直观地呈现在原始图像上，供用户了解细胞分裂情况。对可理解性的评价主要通过实际操作进行，可能需要数据集制造责任方提供相关工具。

14. 可得性：关注囊胚数据集能被授权用户访问和检索的程度，例如数据能否复制粘贴、建立索引、由算法模型调用。对可得性的评价通过实际操作进行。为了确保可得性，囊胚发育图像序列可在元数据或文件名中进行特殊编码，将受精卵编号、细胞分裂时间等信息进行融合，以帮助建立索引。

15. 可移植性：YY/T 1833.2 对于可移植性的一般要求适用于囊胚数据集，一般对数据集进行操作验证，判断数据集能否在不同的操作系统、软硬件配置下被调用。

16. 可恢复性：YY/T 1833.2 对于可恢复性的一般要求适用于囊胚数据集，客观上要求数据集制造责任方提供。对可恢复性的评价可通过模拟失效事件、实际操作验证。

17. 代表性：关注数据集的数据特征层次、流行病学统计、样本来源多样性、数据多样性等能否代表

辅助生殖领域的受试者人群。数据集制造责任方需要对这些维度进行统计分析，适当时与流行病学统计进行比较，以论证数据集的代表性。

二、质量风险评估

建议数据集制造责任方评估数据集的整体质量风险，例如各种统计偏倚情况，可借鉴行业标准 GB/T 42062[21] 的要求开展风险管理活动，尤其是把数据集的偏倚列入风险分析的对象[21]。此外，也可采用专家评议法，设计问卷，由第三方医学专家组对数据集质量进行评议，对偏倚进行分析，形成研究资料。

综上所述，囊胚数据集的质量评价应包括对数据集文档、质量特性和数据集风险分析文档的评价。参照医疗器械行业标准 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》[9]，评价流程如图 2 所示。适当时，数据集制造责任方应提供数据集、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。

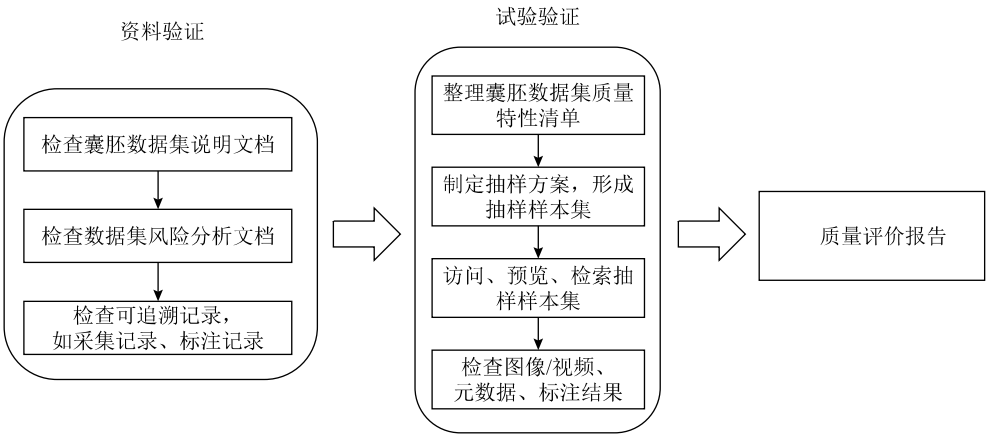


图 2　数据集质量评价流程图

小　结

近年来，随着相关技术的快速发展，AI 在医疗领域的应用也在快速的推广，在包括医学影像、临床决策支持、病例分析、语言识别、药物挖掘、健康管理、病理学等众多场景。医学影像数据的数量和质量决定了 AI 模型学习的结果。高质量的数据库必须同时满足多个要求：数量巨大、来源多样、质量优异、标注规范、标注标准统一等。囊胚数据集作为胚胎 AI 产品训练和测试不可或缺的重要组成部分，扮演着举足轻重的角色，也是产品的重要保障。

本共识会根据技术升级和临床实际情况不断迭代更新，逐步达成该领域数据集建设的广泛共识。

执笔者：王浩（中国食品药品检定研究院）；张孝东（重庆市妇幼保健院、重庆医科大学附属妇女儿童医院）

参与共识制定专家组（排名不分先后）：韩倩倩（中国食品药品检定研究院）、黄国宁（重庆市妇幼保健院、重庆医科大学附属妇女儿童医院）、孙莹璞（郑州大学第一附属医院）、孙海翔（南京大学医学院附属鼓楼医院）、邓成艳（中国医学科学院北京协和医院）、黄学锋（温州医科大学附属第一医院）、刘平（北京大学第三医院）、周灿权（中山大学附属第一医院）、冯云（上海交通大学附属瑞金医院）、郝桂敏（河北医科大学第二医院）、卢文红（国家卫健委科研所）、沈浣（北京大学人民医院）、师娟子（西北妇女儿童医院）、张松英（浙江大学医学院附属邵逸夫医院）、滕晓明（上海市第一妇婴保健院）、王晓红（空军军医大学第二附属医院）、王秀霞（中国医科大学附属盛京医院）、伍琼芳（江西省妇幼保健院）、全松（南方医科大学南方医院）、曾勇（深圳中山泌尿外科医院）、钟影（成都市锦江区妇幼保健院）、邵小光（大连大学附属中山医院）、柯林楠（中国食品药品检定研究院）、毛歆（中国食品药品检定研究院）

利益冲突　所有作者均声明不存在利益冲突。

【参　考　文　献】

[1] Dimitriadis I, Zaninovic N, Badiola AC, et al. Artificial intelligence in the embryology laboratory: a review[J/OL]. Reprod Biomed Online, 2022, 44: 435-448.

[2] Khosravi P, Kazemi E, Zhan Q, et al. Deep learning enables robust assessment and selection of human blastocysts after in

vitro fertilization[J]. NPJ Digit Med,2019,2:21.
[3] Fruchter-Goldmeier Y, Kantor B, Ben-Meir A, et al. An artificial intelligence algorithm for automated blastocyst morphometric parameters demonstrates a positive association with implantation potential[J]. Sci Rep,2023,13:14617.
[4] Zhan Q, Sierra ET, Malmsten J, et al. Blastocyst score, a blastocyst quality ranking tool, is a predictor of blastocyst ploidy and implantation potential[J]. F S Rep, 2020, 1: 133-141.
[5] Cimadomo D, Chiappetta V, Innocenti F, et al. Towards automation in IVF: pre-clinical validation of a deep learning-based embryo grading system during PGT-A cycles[J]. J Clin Med,2023,12:1806.
[6] Kromp F, Wagner R, Balaban B, et al. An annotated human blastocyst dataset to benchmark deep learning architectures for in vitro fertilization[J]. Sci Data,2023,10:271.
[7] Gardner DK, Sakkas D. Assessment of embryo viability: the ability to select a single embryo for transfer-a review[J]. Placenta,2003,24(Suppl B):S5-12.
[8] 国家药品监督管理局. YY/T 1688—2021 人类辅助生殖技术用医疗器械 囊胚细胞染色和计数方法[S]. 北京:中国标准出版社,2021.
[9] 国家药品监督管理局. 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[S]. 北京:中国标准出版社,2022.
[10] 国家药品监督管理局. 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[S]. 北京:中国标准出版社,2022.
[11] 国家药品监督管理局医疗器械技术审评中心. 深度学习辅助决策医疗器械软件审评要点及相关说明[EB/OL].[2019-06-28]. https://www.cmde.org.cn/CL0004/19342.html.
[12] 国家药品监督管理局医疗器械技术审评中心. 人工智能医疗器械注册审查指导原则[EB/OL].[2022-03-09]. https://www.cmde.org.cn//flfg/zdyz/zdyzwbk/20220309091014461.html.
[13] 国家药品监督管理局. YY/T 1858—2022 人工智能医疗器械 肺部影像辅助分析软件 算法性能测试方法[S]. 北京:中国标准出版社,2022.
[14] 国家药品监督管理局. YY/T1833.4—2023 人工智能医疗器械 质量要求和评价 第 4 部分:可追溯性[S]. 北京:中国标准出版社,2023.
[15] 卫生部政策法规司. WS/T 305—2009 卫生信息数据集元数据规范[S]. 北京:中国标准出版社,2009.
[16] 中国食品药品检定研究院,中华医学会放射学分会心胸学组. 胸部 CT 肺结节数据标注与质量控制专家共识(2018)[J]. 中华放射学杂志,2019,53:9-15.
[17] Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting[J]. Hum Reprod,2011,26:1270-1283.
[18] 国家药品监督管理局. YY/T 1630—2018 医疗器械唯一标识基本要求[S]. 北京:中国标准出版社,2018.
[19] Reimer AP, Milinovich A, Madigan EA. Data quality assessment framework to assess electronic medical record data for use in research[J]. Int J Med Inform,2016,90:40-47.
[20] European Society of Human Reproduction and Embryology. Atlas of Human Embryology: from oocytes to preimplantation embryos[EB/OL]. Available at: https://atlas.eshre.eu/es/
[21] 国家市场管理总局 国家标准化委员会. GB/T 42062—2022 医疗器械风险管理对医疗器械的应用[S]. 北京:中国标准出版社,2022.

[编辑:罗宏志]