



团 体 标 准

T/CAME 79—2025

人工智能龋病辅助检测软件临床评价指南

Clinical evaluation guidelines for artificial intelligence-assisted dental
caries detection software

2025-03-15 发布

2025-03-15 实施

中国医学装备协会 发布
中国标准出版社 出版

目 次

前言	Ⅲ
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 临床有效性评价方法	1
4.1 概述	1
4.2 临床有效性评价步骤	1
4.3 临床有效性评价指标	2
5 临床有效性评价	2
5.1 阅片医师培训	2
5.2 筛选影像样本	3
5.3 随机分组	4
5.4 医师阅片	4
5.5 结果分析	5
6 结果判定	6
7 临床有效性评价过程中的口腔影像样本管理	6
参考文献	7

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。

本文件由中国医学装备协会口腔装备与技术专委会提出。

本文件由中国医学装备协会归口。

本文件起草单位：北京大学口腔医院、北京大学临床研究所、上海交通大学医学院附属第九人民医院、中山大学附属口腔医院、四川大学华西口腔医院、空军军医大学(第四军医大学)第三附属医院(口腔医院)、重庆医科大学附属口腔医院、山东大学口腔医学院(口腔医院)、吉林大学口腔医院、南方医科大学口腔医院、湖南中南大学湘雅口腔医院、天津医科大学口腔医院、华中科技大学同济医学院口腔医学院、北京羽医甘蓝信息技术有限公司、北京朗视仪器股份有限公司。

本文件主要起草人：邓旭亮、江泳、华红、王晓燕、徐明明、林红、汪海波、刘妍、刘木清、刘思毅、吴宇佳、蒋欣泉、陈莉莉、林云锋、陈发明、宋锦璘、葛少华、孙宏晨、夏娟、邵龙泉、刘欧胜、韩建民、单艳华、徐骏、袁正林、丁鹏、白海龙、王亚杰。

人工智能龋病辅助检测软件临床评价指南

1 范围

本文件提供了人工智能龋病辅助检测软件临床有效性评价方法的建议,包括评价目的、评价步骤、评价指标、评价要求、结果判定和临床有效性评价过程中的口腔影像样本管理。

本文件适用于基于口腔影像样本采用人工智能技术辅助检测龋病的软件产品的临床有效性评价。

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。

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

3 术语和定义

YY/T 1833.1—2022界定的以及下列术语和定义适用于本文件。

3.1

人工智能龋病辅助检测软件 **artificial intelligence-assisted dental caries detection software**

基于计算机人工智能算法辅助临床医师做出患者是否患龋病的诊疗决策的软件产品。

注:包含模式识别和数据分析等功能,通过识别、标记、突出等方式提示医师关注可能的异常/病变区域,为独立软件或嵌入式软件。

3.2

口腔影像样本 **oral imaging samples**

通过医学影像技术获取的用于口腔疾病诊断、治疗、规划或研究的图像数据或记录。

注:通常包括根尖片、骀翼片、曲面体层片和口腔颌面锥形束CT等。

3.3

参考标准 **reference standard**

专家组对口腔影像样本的阅片判定结果。

[来源:YY/T 1833.1—2022,3.2.16,有修改]

4 临床有效性评价方法

4.1 概述

评价人工智能龋病辅助检测软件在预期适用范围下使用时辅助检测龋病的临床有效性,亦可一并观察软件对辅助检测龋病的效率影响。

4.2 临床有效性评价步骤

临床有效性评价的步骤包括阅片医师阅片、筛选影像学样本、随机分组、医师阅片、结果分析、结果判

定,见图1。

医师阅片宜设置医师联合辅助检测软件阅片为试验组、医师独立阅片为对照组,专家组医师对口腔影像样本的阅片判定为参考标准。

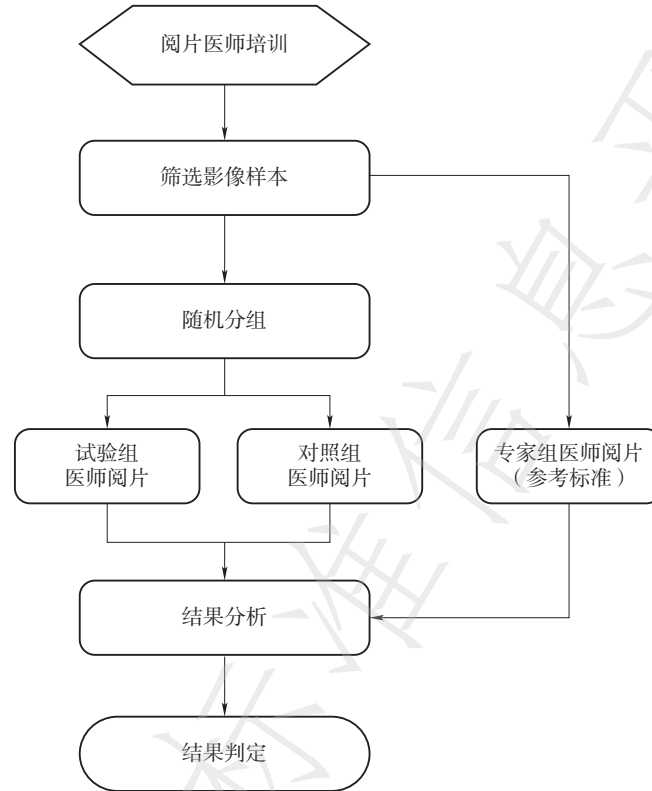


图1 临床有效性评价步骤

4.3 临床有效性评价指标

4.3.1 主要评价指标

4.3.1.1 灵敏度(S_n):在参考标准结果为龋病阳性的牙齿中,阅片医师判断结果为龋病阳性的牙齿数量所占的百分比。

4.3.1.2 特异度(S_p):在参考标准结果为龋病阴性的牙齿中,阅片医师判断结果为龋病阴性的牙齿数量所占的百分比。

4.3.2 次要评价指标

阅片时间(TI):阅片医师对单个病例阅片诊断龋病所需的平均时间,阅片时间定义为阅片医师从打开口腔影像样本到提交阅片结果所需的总时间。

5 临床有效性评价

5.1 阅片医师培训

临床有效性评价前,宜对所有阅片医师培训软件功能和操作、评价流程、影像样本质量要求、评价方法、阅片标准等。

宜选取匿名化影像样本进行培训。培训样本宜包括充足的龋病阳性和阴性病例,独立于临床有效性评价影像学样本。计算评价者内部信度和参考标准组评价者间信度。

5.2 筛选影像样本

5.2.1 通则

筛选影像样本宜由1名~3名从事相关工作5年及以上的高年资口腔医师或口腔影像技师负责,独立于试验组、对照组和参考标准阅片医师。

负责筛选影像样本的医师如超过1名,需进行一致性检验。

5.2.2 样本量的计算

临床有效性评价的影像样本数量试验组和对照组原则上每组至少100例,最终宜根据临床有效性评价标准和统计学要求确定。

根据灵敏度优效性原则计算样本量,见公式(1)。

$$n_T = n_C = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 [P_C(1 - P_C) + P_T(1 - P_T)]}{(D - \Delta)^2} \dots\dots\dots (1)$$

根据特异度非劣效性原则计算样本量,见公式(2)。

$$n_T = n_C = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 [P_C(1 - P_C) + P_T(1 - P_T)]}{(D - \Delta)^2} \dots\dots\dots (2)$$

式中:

- n_T —— 试验组样本量;
- n_C —— 对照组样本量;
- $Z_{1-\alpha/2}$ 、 $Z_{1-\beta}$ —— 标准正态分布的分位数;
- P_T —— 软件辅助医师检测龋病的灵敏度/特异度;
- P_C —— 医师独立检测龋病的灵敏度/特异度;
- D —— 两组预期率的差异;
- Δ —— 界值。

公式(1)中为优效性界值,公式(2)中为非劣效界值。

5.2.3 筛选样本的原则

5.2.3.1 入选原则

5.2.3.1.1 口腔影像样本包括根尖片、殆翼片、曲面体层片和口腔颌面锥形束CT。

5.2.3.1.2 口腔影像样本拍摄规范,感兴趣区显示完整、影像清晰、无干扰诊断的伪影。

5.2.3.2 排除原则

5.2.3.2.1 在软件开发过程中已用于模型训练、优化和测试的口腔影像样本。

5.2.3.2.2 影像拍摄质量差,影响医师阅片的口腔影像样本。

5.2.4 筛选样本的要求

各研究机构宜从某一固定时间点之前拍摄的口腔影像样本中根据拍摄时间依次向前连续性纳入符合要求的口腔影像样本。

5.3 随机分组

筛选后的口腔影像样本宜由专门数据管理人员随机分成两组(A组和B组)。

5.4 医师阅片

5.4.1 通则

临床有效性评价宜尽可能选取来自不同地域、不同层级的多家有代表性的临床试验机构实施。

5.4.2 试验组和对照组医师阅片

5.4.2.1 试验组和对照组阅片医师宜从各研究机构选择具有口腔从业资质的医师。医师阅片可从多阅片者多病例研究设计、交叉自身对照设计中任选一种设计类型设置阅片顺序。

5.4.2.2 自身交叉对照研究设计方法的阅片顺序:阅片医师宜分成两组,A组口腔影像样本先由第一组阅片医师独立阅片,洗脱期(≥ 4 周)后由同一医师联合辅助检测软件阅片;B组口腔影像样本先由第二组阅片医师联合辅助检测软件阅片,洗脱期(≥ 4 周)后由同一医师独立阅片,见图2。

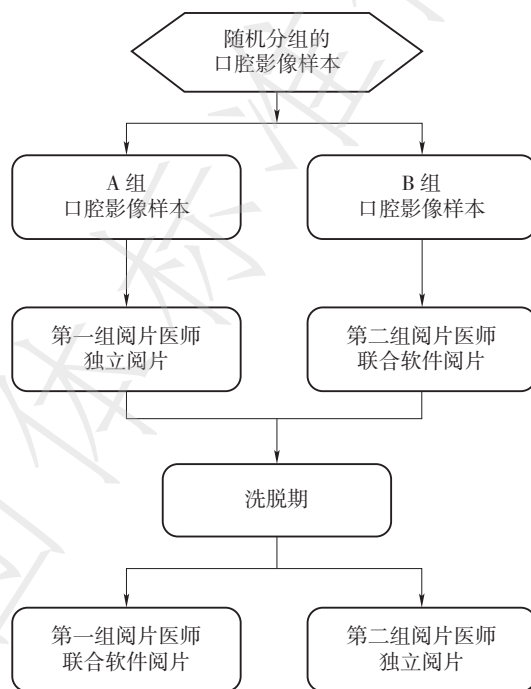


图2 医师阅片顺序(自身交叉对照设计)

5.4.2.3 多阅片者多病例研究设计的阅片顺序:通常采用完全交叉设计,多名阅片医师将评估多种阅片模式下的影像样本,所有阅片医师在每一种阅片模式下,对A组和B组口腔影像样本给出对应的诊断结果,见图3。

阅片医师通常为5名及以上。阅片模式包括阅片医师独立阅片与阅片医师联合辅助检测软件阅片。

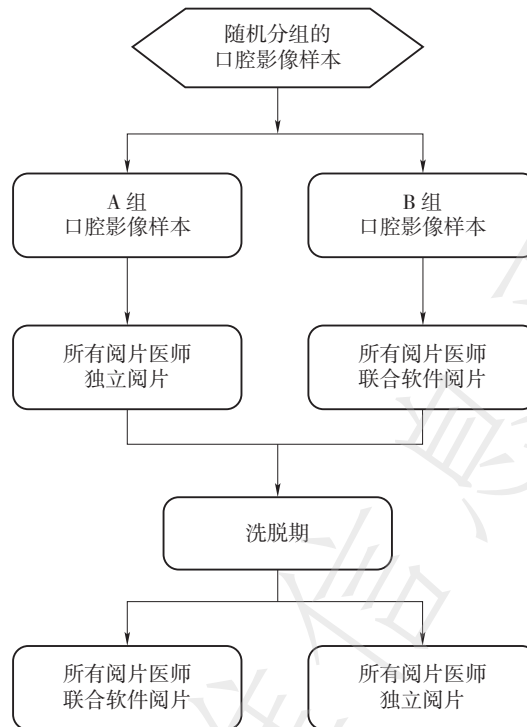


图3 医师阅片顺序(多阅片者多病例研究设计)

5.4.3 参考标准制定

参考标准可基于影像学制定。宜由至少3名具有10年以上龋病诊断经验的口腔临床专家及口腔影像学专家等阅片制定,参考标准制定医师宜独立于试验组和对照组阅片医师。

5.5 结果分析

完成阅片后,宜由专门数据管理人员对试验组、对照组的检测结果与参考标准的一致性进行评价。灵敏度的计算见公式(3)。

$$S_n = \frac{T_p}{Case} \times 100\% \quad \dots\dots\dots(3)$$

式中:

- S_n —— 灵敏度;
- T_p —— 阅片医师判断为患龋病的牙齿数;
- $Case$ —— 参考标准确诊患龋病的牙齿数。

特异度的计算见公式(4)。

$$S_p = \frac{T_N}{Control} \times 100\% \quad \dots\dots\dots(4)$$

式中:

- S_p —— 特异度;
- T_N —— 阅片医师判断为未患龋病的牙齿数;
- $Control$ —— 参考标准确诊未患龋病的牙齿数。

阅片时间的计算见公式(5)。

$$TI = T_1 - T_2 \quad \dots\dots\dots(5)$$

式中：

TI ——阅片时间；

T_1 ——阅片医师打开影像样本时间；

T_2 ——阅片医师提交阅片结果时间。

6 结果判定

6.1 按照4.3临床有效性评价指标,通过统计学方法计算。

6.2 当试验组和对照组两组的灵敏度差异有统计学意义或两组灵敏度差值的95%可信区间下限大于0,即可判定试验组的灵敏度优于对照组。当两组的特异度差值的95%可信区间下限大于非劣效界值时,即可判定试验组的特异度非劣于对照组。当同时满足灵敏度优效性和特异度非劣效时,可以判定人工智能龋病辅助检测软件辅助检测龋病具有临床有效性。

6.3 在满足临床有效性的前提下,当两组的阅片时间差异有统计学意义或两组灵敏度差值的95%可信区间上限小于0时,即满足试验组的阅片时间短于对照组,可以判定人工智能龋病辅助检测软件辅助检测龋病亦可提高检测效率。

6.4 仅满足试验组的阅片时间短于对照组,不可以判定人工智能龋病辅助检测软件辅助检测龋病具有临床有效性。

7 临床有效性评价过程中的口腔影像样本管理

7.1 临床有效性评价使用的口腔影像样本宜由各研究机构医学影像部门指定专人负责,存储于单独的存储介质中,存放于安全的物理环境中。

7.2 口腔影像样本的管理不宜由阅片评价小组成员承担。

7.3 口腔影像样本的脱敏、交接、使用与备份、处置、销毁宜根据各临床研究机构的管理要求规范处理,并保存记录。

参 考 文 献

- [1] 国家药品监督管理局医疗器械技术审评中心. 人工智能辅助检测医疗器械(软件)临床评价注册审查指导原则(2023年第38号)[EB/OL]. (2024-12-16). <https://www.cmde.org.cn/flfg/zdyz/zdyz-wbk/20231116154535183.html>.
- [2] 张祖燕. 口腔颌面医学影像诊断学(第7版)[M]. 北京:人民卫生出版社,2020.
- [3] 陈新林,胡月,莫传伟,等. 诊断准确性研究报告指南——STARD 2015简介[J]. 中国循证医学杂志,2016. 16(10):1227-1230. DOI:10.7507/1672-2531.20160186.
- [4] 侯艳,李康,宇传华,周晓华. 诊断医学中的统计学方法(第二版)[M]. 北京:北京高等教育出版社,2016.
- [5] 国家食品药品监督管理总局. 医疗器械临床试验设计指导原则(2018年第6号)[EB/OL]. (2024-12-16). <https://www.nmpa.gov.cn/xxgk/ggtg/ylqxggtg/ylqxqtggtg/20180108183301635.html>.
- [6] 孙振球. 医学统计学(第3版)[M]. 北京:人民卫生出版社,2010.
- [7] FDA. Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data in Premarket Notification [510(k)] Submissions [EB/OL]. (2024-4-25). <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-performance-assessment-considerations-computer-assisted-detection-devices-applied-radiology>
- [8] 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(2019年第7号)[EB/OL]. (2024-12-16). <https://www.cmde.org.cn/xwdt/shpgzgg/gztg/20190703141714991.html>.
- [9] 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(2022年第8号)[EB/OL]. (2024-12-16). <https://www.cmde.org.cn/xwdt/shpgzgg/gztg/20220309090800158.html>.
- [10] 郭秀花,李卫等. 医学影像人工智能医疗器械临床试验统计学设计要点专家共识[J]. 北京:中国卫生统计,2024(41)3:474-480.
-