

附件 3

推荐性国家标准项目申报书

项 目 名 称 : 人工智能医疗器械 心电辅助分析软件算法
性能测试方法

技术归口单
位 (或技术
委员会) : 国家药品监督管理局

提 出 日 期 : 2025-12-2

一、基本信息

中文名称	人工智能医疗器械 心电辅助分析软件算法性能测试方法		
英文名称	Artificial Intelligence Medical Devices - Computer assisted analysis software for Electrocardiogram - Algorithm performance test methods		
标准性质	<input checked="" type="checkbox"/> 推荐性国家标准 <input type="checkbox"/> 指导性技术文件		
制定/修订	<input checked="" type="checkbox"/> 制定 <input type="checkbox"/> 修订	被修订标准号	/
是否采标	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	采标类型	/
采标号	/	采标中文名称	/
项目周期	<input type="checkbox"/> 12个月 <input type="checkbox"/> 16个月 <input checked="" type="checkbox"/> 18个月		
上报单位	国家药品监督管理局		
技术归口单位 (或技术委员会)	国家药品监督管理局		
主管部门	国家药品监督管理局		

二、论证评估报告

（一）制修订推荐性国家标准的必要性、可行性

1. 必要性

根据中国远程心电诊断学组数据，我国每年心电图检查量约3亿人次（含动态心电图），而我国精通心电图诊断的医生不足4万人，心电诊断工作负荷远超合理范围。部分地区由于认知不足和设备限制，无症状房颤常规体检漏诊率高达30%。

随着人工智能技术在医疗领域的深度渗透，人工智能心电辅助分析软件应运而生，已广泛应用于房颤筛查、心肌缺血识别、心律失常分类等临床场景，成为提升基层医疗诊断效率、降低漏诊误诊率的关键工具。然而，心电辅助分析诊断在实际应用中面临如下挑战：

1.1 测试标准缺失。不同企业采用的测试数据集质量参差不齐，部分数据集未覆盖儿童、老年患者及特殊疾病（如起搏心电图、多病变复杂心电图）样本，且现行有效标准涉及心电算法测试的内容均是针对传统的心电波形分析，未对辅助决策的性能评价方法作出规范，缺乏心电人工智能辅助诊断算法的测试标准。

1.2 性能评价指标不统一。部分企业仅关注灵敏度、阳性预测值等基础指标，忽视泛化性、鲁棒性、多标签分类准确性等关键特性，难以全面反映算法临床适用性。

1.3 测试流程不规范。部分测试未进行数据封样与可追溯管理，存在测试集被重复使用、算法“记忆”测试数据的风险，影响测试结果真实性。

因此，建立人工智能心电辅助分析软件算法性能的测试方法，实现该类软件在产品研发、注册检验、临床应用、上市后评价等阶段的算法性能测试，对推动人工智能心电辅助分析软件行业的规范化发展及人工智能技术与心电诊疗深度融合，提升心血管疾病诊断效率、降低诊疗风险提供技术支撑有着重要意义。

2. 可行性论证

2.1 技术研究积累丰富。国内外在人工智能算法、心电信号处理、临床数据库构建等方面已有多年的深入研究。国内外学术机构、领先企业及检测机构已积累了大量关于心电波形识别、分类算法性能评估的经验，发表了诸多相关研究成果，为标准化测试方法的核心技术要素（如测试数据集要求、金标准定义、性能指标体系等）提供了坚实的技术理论基础

2.2 国际经验可供借鉴。国际标准化组织（如 ISO、IEC）、国际电工委员会医用电器设备标委会（IEC/TC62）以及美国电气电子工程师学会（IEEE）等已发布或正在研制多项医疗器械软件、人工智能医疗应用相关的通用安全与性能标准。我国可充分借鉴其框架、理念与方法，结合我国心电数据特征、临床实践和监管要求，进行本土化创新与深化，形成具

有中国特色的专用测试方法标准。

2.3 国内标准化工作基础扎实。我国已初步构建医疗器械软件、人工智能医疗器械相关的标准体系框架，并已发布《人工智能医疗器械 质量要求和评价 第1部分：术语》、《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》等基础通用标准。在同类标准中，针对肺部影像辅助分析软件算法性能测试方法（YY/T 1858-2022）已于2023年9月实施。本标准可在上述基础上，聚焦“心电辅助分析软件”、“算法性能测试”这一行业及核心环节进行专项细化，与现有标准协同配套，共同完善人工智能医疗器械标准体系。

2.4 验证基础条件完善。随着人工智能技术在医疗领域的深度应用，我国心电辅助分析软件市场迎来了高速发展期。众多人工智能企业与传统医疗器械制造商纷纷布局，已有多款人工智能心电分析软件通过医疗器械审批程序获批上市，产品功能覆盖心律失常分析、房颤筛查、心肌梗死辅助诊断、长时程心电事件监测等核心场景，并在医疗行业普遍应用，因此具有完善的验证基础。

综上所述，制定《人工智能医疗器械 心电辅助分析软件算法性能测试方法》国家推荐性行业标准，产业需求迫切，技术基础成熟，标准基础扎实，实施条件具备，项目立项具备充分的必要性和可行性。

（二）主要技术要求

本标准主要技术要求围绕心电辅助分析软件算法性能测试的全流程展开，重点覆盖以下核心技术要求：

1. 测试流程与质量控制

1.1 测试环境。明确测试的环境要求，包含计算设备、心电设备、外网设备等硬件环境，操作系统、被测试软件、安全软件等软件环境及被测试软件的网络条件，确保产品测试的基础设施的科学性、合理性。

1.2 测试资源。明确测试数据集的构建规范，包括样本多样性、标注流程、质量控制，同时规定数据集元数据需包含患者年龄、性别、采集设备型号、导联数量等信息，确保可追溯性。并对样本量及分布进行规范，以便代表软件在目标患者人群和预期使用环境中的真实性能。通过对数据扩增规范要求，以便评估算法在面对预期范围内的信号变化和干扰时的性能保持能力。

1.3 测试流程及报告。本标准制定了测试全流程的管理规范，包括测试前（需制定测试计划，明确测试项、风险对策、执行人员）、测试中（需记录测试设备型号、软件版本、环境参数）、测试后（需对测试数据集、结果报告进行物理封存与哈希值固化，确保可追溯），同时要求开展错误分析（如假阴性结果中“未达匹配阈值”与“完全未匹配”的比例，多分类场景下各类别的错误分布），并形成完整测试报告。

2. 针对不同应用场景制定专项测试方法

2.1 特征波形检测与分类。规定 P 波 / T 波（位置距离 $\leq 120\text{ms}$ 视为匹配）、QRS 波（位置距离 $\leq 100\text{ms}$ 视为匹配）的匹配规则，计算真阳性（TP）、假阳性（FP）、假阴性（FN），并采用 “波形级 + 样本级” 双维度评估灵敏度（Sen_Gross、Sen_Average）、阳性预测值（PPV_Gross、PPV_Average），同时引入 F1 分数、马修斯相关系数（MCC）评估分类平衡性；

2.2 心电波形参数自动测量。针对平均心率、P/QRS/T波的幅度、时限、间期等核心参数，要求对测量误差进行统计，以评估算法的自动测量准确性；

2.3 心电图自动诊断。针对标签分类场景，采用单标签度量（灵敏度、特异度、PPV、MCC）与多标签整体度量（Kappa 系数、汉明损失、宏观 F1 分数）结合的方式对指标进行评估；

2.4 长时程心电事件检测。从 “片段级”（统计事件片段的灵敏度、PPV）与 “时间级”（统计事件持续时间的灵敏度、PPV）双维度评估，确保对阵发性房颤、ST 段改变等事件的检测准确率。

3. 算法质量特性测试

针对人工智能医疗器械特殊质量需求，规定不确定度、泛化性、鲁棒性、一致性、重复性等特性的测量\测试方法：

3.1 不确定度测量。要求对测试集\算法准确度开展不确定

度测量，以量化标准测试集因数据采集、标注、扩增引入的误差，避免因测试集不可靠导致算法性能误判，评估算法准确性指标（如准确率、召回率）的统计波动范围，确保测试结果的置信水平符合医疗器械监管要求，包含不确定度来源与分类识别、不确定度量化评估、不确定度可接受阈值与控制。

3.2 泛化性测试。要求对产品研发使用的训练集与真实世界陌生样本之间的差异进行分析，形成文档，作为配置测试集的依据。在实际测试中应通过测试集的多样性与变化性，对算法的泛化能力进行验证；

3.3 鲁棒性测试。要求根据心电产品风险分析和临床部署环境特征，评估临床使用阶段各种可能干扰算法性能的因素，获取或模拟相关数据，组成专用测试集，对算法性能进行对抗测试，分析各指标的变化情况，形成鲁棒性研究资料，包含面向硬件变化、软件前处理、欺骗攻击的对抗测试等内容；

3.4 重复性测试。在完全相同的测试条件下，软件对同一心电数据重复执行相同辅助分析任务时，输出结果的一致性程度。核心体现为“无随机波动”，避免因算法不稳定导致临床诊断困惑（如同一心电图两次分析结果分别为“正常心律”和“房颤”）。测试人员对同一数据集合应测试三次及以上，避免单次结果的偶发性；

3.5 一致性测试。要求在测试平台开展测试时，需验证测试

平台的输出与产品的最终输出结果是否一致，确保被测试对象的行为、接口、性能等维度与预先定义的“基准规则”保持一致，避免因不符合标准导致的兼容性问题、功能失效或互操作障碍。

3.6 效率测试。在临床实际使用场景下，完成心电辅助分析任务的“时间性能”与“资源占用性能”的综合表现，核心体现为时间效率和资源效率。其中时间效率为软件从“完成心电数据导入”到“输出完整分析结果”（含分类结论、数值特征、异常提示）的总时间，不包含“数据传输时间”；资源效率为软件分析过程中对硬件资源的消耗，包括CPU使用率、内存占用量、显卡（GPU）使用率。

4. 测试流程与质量控制

明确测试全流程的管理规范，包括测试前（需制定测试计划，明确测试项、风险对策、执行人员）、测试中（需记录测试设备型号、软件版本、环境参数）、测试后（需对测试数据集、结果报告进行物理封存与哈希值固化，确保可追溯），同时要求开展错误分析（如假阴性结果中“未达匹配阈值”与“完全未匹配”的比例，多分类场景下各类别的错误分布），并形成完整测试报告。

本标准通过构建“流程-指标-控制”三位一体的测试体系，通过测试样本量计算+多维度数据分布控制，保障统计学合理性；首次引入不确定度、定量证据等算法和指标，确

保测评结果的可信性；针对不同应用场景设计多标签综合评测指标，提升应用场景的覆盖度；通过引入白盒对抗测试及极端压力样本，提升测试的安全与鲁棒评估；通过在全流程测试引入平台化、网络安全、测试后封样等机制，确保测试结果的安全、可溯源，为企业产品研发、监管机构审评提供统一技术依据，推动行业产品质量提升及良性发展。

本标准制定的过程中，未涉及有争议指标。

（三）国内外标准情况、与国际标准一致性程度情况

在标准化方面，国际心电相关标准主要有 IEC 60601-2-25: 2011、IEC 60601-2-47: 2012、IEC 60601-2-27: 2011，国内已将标准进行等同转化，但现行有效标准涉及心电算法测试的内容均是针对传统的心电波形分析，未对辅助决策的性能评价方法作出规范。国内现行行业标准 YY/T 1833 系列（如 YY/T 1833.2-2022《医疗器械 人工智能 数据集 第 2 部分：通用要求》）仅规范数据集通用要求，未针对心电算法的“临床场景化测试、实时性、泛化性”提出专项方法，无法满足行业实际需求。因此本标准在国内外尚属空白。

（四）与相关强制性标准、法律法规配套情况

与有关法律、法规无冲突，协调一致。

（五）标准所涉及的产品、过程或者服务目录

本标准为人工智能医疗器械方法标准，涉及所有注册为第三类医疗器械的人工智能医疗器械产品，当前共计 7 个。

（六）可能涉及的相关知识产权情况

本标准项目为自主制定项目，未识别到相关的专利情况。

（七）征求国务院有关部门或关联 TC 意见的情况

标准化对象主要为医疗器械，由国家药品监督管理局负责实施监督。

（八）经费预算

本标准项目主要依托自筹经费，预算估计如下：

1. 出版费 0.5 万元
2. 资料费 1.02 万元
3. 起草费 1.85 万元
4. 试验费 0.6 万元
5. 差旅费 1.3 万元
6. 咨询费 0.85 万元
7. 验证费 1.1 万元
8. 会议费 1.08 万元
9. 审查费 0.2 万元

总计：8.5 万元

（九）项目进度安排

本项目预计在 12 个月内完成，工作进度规划如下：

1. 项目预研：75 天
2. 调研起草：75 天
3. 征求意见：60 天

4. 修订送稿：30 天

5. 技术审查：30 天

6. 整理与报批：30 天

(十) 需要申报的其他事项
无。