

“人工智能医学信息系统软件 审评指导体系构建”课题

—病理诊断人工智能分析软件技术审评指导原则（征求意见稿）

本指导原则是人工智能医学信息系统软件审评指导体系构建的组成部分，基于人工智能医疗器械审评指导原则的通用要求，细化了病理图像人工智能软件的一般要求。

本指导原则是供病理图像人工智能软件的研发、生产和技术审评使用的指导性文件，不涉及相关行政审批事项，亦不作为法规强制执行，应在遵循相关法规的前提下使用本指导原则。

本指导原则是在现行法规和标准体系以及当前认知水平下制定的，随着法规和标准的不断完善，以及科学技术的不断发展，在使用过程中应对相关内容适时进行调整。

一、适用范围

本指导原则适用于病理图像人工智能分析软件的产品注册，该类软件主要采用人工智能技术对来源于乳腺、前列腺、宫颈等组织或细胞层面的病理样本的图像（如采用显微设备、扫描仪等方式获取）进行处理、分析，临床上用于疾病的辅助诊断、检测、筛查或通过数据处理和测量等为临床提供参考信息等。

本指导原则适用于病理人工智能独立软件，软件组件可采用适用部分。

二、注册审查要点

25 （一）监管信息

26 明确申请表中产品名称、管理类别、分类编码、型号规
27 格、产品组成等信息。

28 1. 产品名称

29 产品名称应符合独立软件产品名称应符合《医疗器械通
30 用名称命名规则》、《医疗器械通用名称命名指导原则》、
31 《医用软件通用名称命名指导原则》的要求和国家标准、行
32 业标准中的通用名称要求，主要由“特征词+核心词”结构编制，
33 特征词可描述使用部位（如乳腺、前列腺、宫颈等）、作用
34 对象（如数字病理图像、细胞学病理图像等）；核心词可描
35 述为辅助诊断/检测/辅助筛查软件。

36 2. 分类编码

37 依据《医疗器械分类目录》和《人工智能医用软件产品
38 分类界定指导原则》，原则上产品用于辅助决策，则分类编
39 码为 21-04-02；按第三类医疗器械管理。

40 3. 注册单元划分

41 根据产品的技术原理、结构组成、性能指标、适用范围
42 进行注册单元划分。

43 （二）综述资料

44 1. 产品描述

45 1.1 工作原理

46 注册申请人应在综述资料中明确产品工作原理，包括逻
47 辑结构和物理结构。结构示意图中根据需要可预留接口。

48 1.1.1 逻辑结构

49 逻辑结构应描述产品的体系结构，可以按照功能模块或
50 组成模块进行描述，也可采用其他方式进行描述，提供示意

51 图并依据示意图进行描述。需进一步阐述核心算法的任务，
52 如对图像进行测量、检测、分割或分类等，分类的标准可基
53 于形态学等。

54 1.1.2 物理结构

55 物理结构应描述产品的物理拓扑和技术规格，物理拓扑
56 可以结合产品架构或产品规模进行描述，也可采用其他方式
57 进行描述，提供示意图并依据示意图进行描述。技术规格应
58 描述产品运行所需的技术要求和硬件要求，如通信标准或协
59 议（如 HL7）、存储模式 [如三级存储模式（在线、近线和
60 离线）或两级存储模式（在线和备份）]、存储格式、网络
61 类型（如局域网、广域网、无线网、遥测网）、传输内容（如
62 文字、图像、视频）、存储介质（如磁盘、光盘、移动存储
63 器）、显示器（如分辨率、亮度）、辅助设备（如图像扫描
64 设备、IC 卡读写设备）等。

65 1.1.3 结构组成

66 明确交付内容和功能模块，其中交付内容包括软件安装
67 程序、授权文件、外部软件环境安装程序等软件程序文件，
68 功能模块包括客户端、服务器端（若适用）、云端（若适用），
69 若适用注明选装、模块版本。基于人工智能算法（如深度学
70 习）的功能模块，需在模块名称中予以注明。

71 物理组成+逻辑组成+软件发布版本（如 V1.0），其中物
72 理组成如光盘、U 盘或预装于计算机等；逻辑组成应描述临
73 床功能模块，包括客户端、服务器端（若适用）、云端（若
74 适用），如适用应注明选装和模块版本。

75 1.2 型号规格

76 需明确申报产品的型号规格及发布版本。产品型号规

77 格及其划分，如同一个注册单元包含多个型号规格，需提供
78 产品型号规格区分列表或配置表。

79 2. 适用范围和禁忌证

80 (1) 适用范围

81 适用范围需明确处理对象、核心功能、适用人群、目标
82 用户、临床用途。

83 如具体描述诊断部位的基本构成（如乳腺、前列腺、宫颈
84 颈等组织或细胞），算法的任务（如测量、分类（二分类、
85 多分类）、检测和分割等）和核心功能（如分级、预测、疗
86 效评估、辅助筛查等），目标用户须经培训合格的医师使用，
87 不能单独用作临床诊疗决策依据。

88 (2) 预期使用环境

89 需明确设备使用场所和使用环境要求。设备使用场所包
90 括：医疗机构机房等。使用环境要求需至少包括：推荐的硬
91 件使用的适宜温度、湿度、大气压、光照条件。

92 (3) 适用人群

93 需详述产品的适用人群、感兴趣器官/疾病/病灶/异常以
94 及预期使用该产品的目标用户。如可供 18 岁及以上门诊及
95 体检人群使用，须经培训合格的病理医师、专科医师使用。

96 (4) 禁忌证

97 需明确产品临床应用的禁忌证以及器械限制，包括导致
98 器械无效且不应使用的疾病、病症、异常。

99 (三) 非临床资料

100 1. 产品风险管理资料

101 依据 YY/T 0316《医疗器械风险管理对医疗器械的应用》，
102 提供产品风险管理报告。

103 申请人需重点说明：申报产品的研制阶段已对有关可能的
104 的危害及产生的风险进行了估计和评价，针对性地实施了降
105 低风险的技术和管理方面的措施。产品性能测试对上述措施
106 的有效性进行了验证，达到了通用和专用标准的要求。申请
107 人对所有剩余风险进行了评价，全部达到可接受的水平。产
108 品风险分析资料需为申请人关于产品安全性的承诺提供支持。
109 持。

110 1.1 风险管理报告一般包括以下内容：

- 111 (1) 申报产品的风险管理组织。
- 112 (2) 申报产品的组成。
- 113 (3) 申报产品符合的安全标准。
- 114 (4) 申报产品的预期用途，与安全性有关的特征的判定。
- 115 (5) 对申报产品的可能危害作出判定。
- 116 (6) 对所判定的危害采取的降低风险的控制措施。
- 117 (7) 对采取控制措施后的剩余风险进行估计和评价。

118 1.2 本产品的主要风险点：

- 119 (1) 假阳性或假阴性诊断结果引起的风险；
- 120 (2) 较频繁的更新引起的风险；
- 121 (3) 因使用环境的变化引起软件失效，未达预期功能和
122 性能要求等。

123 1.3 降低风险的措施

- 124 (1) 在产品说明书或标识中明确要培训操作和告知患者
125 的信息；
- 126 (2) 补充其他确诊手段对诊断结果进行进一步确认以降
127 低风险；
- 128 (3) 确保用户知悉软件更新方法、更新内容和相关注意

129 事项，提供必要的培训，加强软件更新后的跟踪监督；

130 （4）加强软件运维和异常监督措施等

131 2.产品技术要求及检验报告

132 2.1 产品技术要求

133 依据《医疗器械产品技术要求编写指导原则》（2022 年
134 第 8 号）和《医疗器械软件注册审查指导原则》（2022 年修
135 订版）的附录：独立软件产品技术要求模板进行编制。

136 2.1.1 规格信息

137 明确软件发布版本和版本命名规则。软件版本命名规则
138 原则上需涵盖算法驱动型更新、数据驱动型更新和网络安全的
139 的更新，明确并区分重大软件更新和轻微软件更新。

140 明确不同型号间产品的差异。

141 2.1.2 性能指标

142 2.1.2.1 通用要求

143 应制定产品的处理对象、使用限制、输入输出、接口、
144 必备软硬件、运行环境、性能效率、最大并发数、用户界面、
145 消息、用户差错防御、访问控制、版权保护、可靠性、维护
146 性。

147 2.1.2.2 临床功能要求

148 依据说明书明确软件全部临床功能纲要（注明可选）。

149 其他辅助功能包括：诊断报告输出，辅助业务质控，远
150 程会诊，云计算等。

151 若含有基于测评数据库测试的性能指标，其要求参考
152 《人工智能医疗器械注册审查指导原则》。

153 运行环境，运行在不同计算机系统的产品模块（如客户
154 端和云端）需分别描述其运行环境、服务器部署（局域网、

155 云端）信息。

156 附录中明确体系结构图、用户界面关系图与主界面图示、
157 物理拓扑图、测评数据库、数据库/集信息。

158 3.软件研究

159 3.1 基本要求

160 生产企业需依据《医疗器械软件注册审查指导原则
161 （2022 年修订版）》提交相应软件研究资料。其中，核心算
162 法所述人工智能算法需依据《人工智能医疗器械注册审查指
163 导原则》提交相应算法研究资料。

164 生产企业需依据《医疗器械网络安全注册审查指导原则
165 （2022 年修订版）》提交网络安全描述文档。按照《人工智
166 能医疗器械注册审查指导原则》提交数据安全资料。

167 若使用云计算服务，生产企业需依据《人工智能医疗器
168 械注册审查指导原则》提交相应研究资料。使用云计算服务
169 需明确服务模式、部署模式、核心功能、数据接口、网络安
170 全能力、服务（质量）协议等要求。

171 若使用移动计算，相关要求详见《移动医疗器械注册技
172 术审查指导原则》。

173 3.2 软件安全级别

174 软件风险程度采用软件安全性级别进行表述，软件安全
175 性级别越高，生存周期质控要求越严格，注册申报资料也越
176 详尽。软件安全性级别基于软件风险程度分为轻微、中等、
177 严重三个级别（轻微级别、中等级别、严重级别分别与 YY/T
178 0664 所定义的 A 级、B 级、C 级相对应），其中轻微级别即
179 软件不可能产生伤害，中等级别即软件可能直接或间接产生
180 轻微（不严重）伤害，严重级别即软件可能直接或间接产生

181 严重伤害或导致死亡。

182 软件安全性级别可结合软件的预期用途、使用场景、核
183 心功能进行综合判定。通常用于辅助决策，如提供病灶特征
184 识别、病变性质判定等临床诊疗建议或全新核心算法时，按
185 照 C 级进行风险控制。

186 3.3 核心功能

187 列明软件核心功能的名称、所用核心算法、预期用途并
188 注明类型，全新的核心功能、核心算法、预期用途均需提供
189 安全有效性研究资料。

190 表 1.核心算法示例

算法名称	用途	功能
如基于 YOLO 的细胞检出算法	对病变细胞进行检测	辅助医生进行病变细胞检测
如基于 Xception 的细胞分类算法等	对病变细胞进行分类	辅助医生进行病变细胞分类
如基于 Densenet 的细胞分类算法等	对较大视野的 ROI 进行分类	辅助医生进行病变细胞分类
如基于 U-NET 的细胞分割算法等	对目标对象进行分割，提取感兴趣区域	辅助医生进行病变细胞分割
如基于图像像素的测量算法等	对目标对象的形态、颜色、纹理、分布等可量化指标进行测	辅助医生进行量化分析。

	量计算。	
--	------	--

191 3.4 软件更新要求

192 常见更新类型包括算法驱动型、数据驱动型和其他类型。
193 各软件更新类型的考量如下：

194 (1) 算法驱动型软件更新，属于重大软件更新，应提交
195 许可事项变更申请。

196 (2) 数据驱动型软件更新，是否属于重大软件更新，原
197 则上以算法性能评估结果为准，若算法性能评估结果发生显
198 著性改变（即与前次注册所批准的算法性能评估结果相比存
199 在统计学显著差异）则属于重大软件更新。

200 (3) 对于其他类型的软件更新，若发生重大软件更新，
201 生产企业应及时提交许可事项变更申请；若发生轻微软件更
202 新，生产企业通过质量管理体系进行软件更新质量控制，无
203 需提交许可事项变更申请。

204 (4) 若同时发生多种类型的软件更新，注册申报要求遵
205 循风险从高原则。例如，同时发生数据驱动型和算法驱动型
206 软件更新，按照算法驱动型软件更新处理。

207 4. 算法研究

208 4.1 算法基本信息

209 需提供算法设计和功能的信息，提供流程图及注释，明
210 确算法名称、版本、软件平台的相关特性。

211 需描述算法每个阶段的设计和功，明确所用算法的名
212 称、结构（如层数、参数规模）、流程图、现成框架（Tensorflow、
213 Caffe、PyTorch、Keras 等）、输入与输出、运行环境、算法
214 来源依据（或注明原创）等信息。同时应当明确算法选择与
215 设计的原则、方法和风险考量等。

216 4.2 算法需求规范

217 算法功能需求建议关注数据库需求、算法性能评价指标
218 及制定依据。

219 4.2.1 数据库需求

220 算法性能评价需要基于训练和测试数据库。

221 4.2.2 算法性能评价指标

222 在指定检出算法评估指标之前，需明确标记匹配的方式，
223 即算法标记目标与参考标准目标的匹配方式；明确算法任务，
224 明确算法训练和调优过程中不同任务的的评估指标及定义。

225 明确不同任务的算法评价指标的确定依据、分层影响因
226 素选择依据，建议参考相关疾病的中国专家共识、国内外指
227 南、相关标准等。

228 检出指标包括不限于召回率、精确度、特异度。

229 分类指标包括不限于灵敏度、特异度、总体的 Kappa 系
230 数、准确率。

231 分割指标包括不限于：Dice 系数、交并比。

232 注：指标定义可参考 YY/T 1833.1-2022《人工智能医疗
233 器械 质量要求和评价 第 1 部分:术语》。

234 4.3 数据质控

235 建议参考《人工智能医疗器械注册审查指导原则》(2022
236 年第 8 号)及 YY/T 1833.2-2022《人工智能医疗器械质量要
237 求和评价 第 2 部分:数据集通用要求》。

238 用于病理 AI(人工智能)软件算法训练的直接数据为数
239 字图像文件；原始数据为病理切片，切片经数字化、预处理
240 和标注后使用。

241 4.3.1 数据采集

242 数据采集首先应定义采集的目标和用途，并明确数据采
243 集源和采集范围；样本尽可能来自多家、不同地域、不同层
244 级的代表性临床机构，尽可能来自多种、不同采集参数的采
245 集设备。明确采集设备要求和采集的数据的处理方法。

246 4.3.2 数据整理

247 明确数据转移保存的方法。明确数据纳入排除标准，以
248 及进行数据筛选的方法（人工清洗、自动清洗），提供数据
249 整理软件工具的研究资料。

250 提供数据清洗流程图，明确清洗规则，例如确定数据合
251 规性、连续性、完整性、已完成脱敏等并加以筛选。清洗结
252 果明确弃用数据的数量和原因。明确预处理的操作步骤和内
253 容。

254 4.3.3 数据标注

255 数据标注建议参考 YY/T 1833.3-2022《人工智能医疗器
256 械质量要求和评价 第 3 部分：数据标注通用要求》。

257 数据标注应当考虑标注资源管理、标注过程质控、标注
258 质量评估等要求。需明确标注任务分类（包括数据模态（如
259 数字病理图像）、执行主体（人工标注、自动标注等）、标
260 注结果格式（如 jpeg、tiff、jpeg2000）、标注结果性质（给
261 出病理图像的分类）、标注结果形式等维度），提供标注任
262 务描述文档（标注规则（专家组根据文献、专家共识等共同
263 给出）、标注人员、标注工具（商用现成软件 COTS/开源软
264 件/自制软件）、标注环境、数据）。

265 如采用的标注工具、标注平台使用了人工智能算法进行
266 辅助标注，需提交标注工具、标注平台算法性能研究资料。

267 4.3.4 数据集构建

268 依据《人工智能医疗器械注册审查指导原则》指南明确
269 训练集、调优集、测试集的划分方法、划分依据、数据分配
270 比例。

271 明确训练集、调优集、测试集的划分方法、划分依据、
272 数据分配比例。训练集应当保证样本分布具有均衡性，测试
273 集、调优集应当保证样本分布符合临床实际情况，训练集、
274 调优集、测试集的样本应当两两无交集。

275 明确数据扩增需明确扩增的对象、范围、方式（离线、
276 在线）、方法（如翻转、旋转、镜像、平移、缩放、滤波、
277 生成对抗网络等）、倍数、在线数据扩增记录。

278 提供扩增数据库与标注数据库样本量、样本分布（注明
279 扩增倍数）对比表，以证实扩增数据库样本量的充分性以及
280 样本分布的合理性。

281 如果采用生成对抗网络进行数据扩增，应提供生成对抗
282 网络的算法基本信息以及算法选用依据资料。数据集的扩增
283 应当考虑扩增倍数过大对软件的影响及其风险。

284 4.4 算法训练

285 算法训练需明确训练过程所采用的优化器及其相关参
286 数，如 Adam, SGD；在算法训练阶段，需明确验证集的划分
287 方式，如留出法，交叉验证法；需明确训练目标，即判断何
288 时停止训练，如设定最大的训练 epoch 数目，依据损失函数
289 判断损失稳定且不继续下降，根据训练 epoch 数量-评估指标
290 曲线判断评估指标不继续提升等。当训练停止后，明确训练
291 模型最佳 epoch 的选择方法。同时，结合临床需求（如灵敏
292 度、精准度），明确算法出厂阈值的选择与方法，并论证训
293 练所得模型是否满足产品既定目标。算法出厂阈值的选择需

294 提供制定依据。

295 4.5 算法验证与确认

296 明确算法任务，明确不同任务的算法评价指标的阈值及
297 确定依据。明确算法标记目标与参考标准的匹配方式和匹配
298 阈值。明确病例水平和目标对象水平计算方法及定义一个目
299 标对象的处理方法及病例阳性的方法。区分算法任务制定具
300 体测试方法和结果计算方法。算法确认通常包括灵敏度、特
301 异性、准确度、阳性预测值、阴性预测值、ROC 曲线和 AUC
302 值。

303 4.6 算法性能综合分析

304 结合算法训练、算法性能评估、临床评价等结果开展算
305 法性能综合评价，针对训练样本量和测试样本量过少、测试
306 结果明显低于算法设计目标、算法性能变异度过大等情况，
307 对产品的适用范围、使用场景、核心功能进行必要限制。

308 需对算法测试产生的错误结果进行分析，评估算法在检
309 出、分类等任务中出现的假阳性、假阴性结果的含义与危险
310 程度，提供算法性能影响因素分析过程，如采集设备、采集
311 参数、疾病构成、病变特征等因素影响，并提供特殊案例及
312 数据分析过程，以提升算法可解释性，以上内容形成定量报
313 告。

314 5. 病理数字图像质量要求

315 数字图像的质量与图像生成设备密切相关，应明确配套
316 设备的选择、参数、工作流程和验证、确认过程等，同时采
317 集数据时应对图像清晰度、有效视野范围、图像噪声等做出
318 明确规定。

319 5.1 制片方式与设备

320 数据采集前需关注制片的流程是否符合要求，标准化的
321 制片流程可减少制片环节带入的偏倚。制片设备尽可能选择
322 全自动化、高效率和稳定性的设备，同时应统一制片方式，
323 如宫颈液基细胞目前多采用膜式和沉降式两种，此两种方式的
324 的图像质量存在差异，对软件性能影响较大。

325 5.2 成像设备

326 成像设备根据采集方式可分为实时采集设备和扫描采
327 集设备，一般实时采集设备要控制放大倍数、像素、分辨率、
328 帧速率、色彩等参数；扫描采集设备则要控制扫描倍数、扫
329 描范围、像素、分辨率、色彩等参数。

330 6. 互操作性

331 这里互操作性（又称互用性）是指软件与其他医疗器械
332 或通用设备通过电子接口交换并使用信息的能力。其中电子
333 接口包含硬件接口和软件接口，信息包括但不限于医学图像、
334 体外诊断等数据和控制指令。

335 重点关注医疗器械软件的接口设计及其风险，接口包括
336 内部接口和外部接口，注册申请人需考虑软件接口的需求分
337 析、风险管理、验证与确认、维护计划等活动要求，以及说
338 明书与标签的设计要求。

339 申请人需提交一份互操作性研究资料，内容包括基本信
340 息、需求规范、风险管理、验证与确认、维护计划；亦可在
341 自研软件研究报告相关条款中体现软件接口要求，具体要求
342 详见《医疗器械软件注册审查指导原则（2022 年修订版）》。

343 （四）临床评价

344 该产品不属于《免于临床评价医疗器械目录》（国家药
345 监局关于发布免于临床评价医疗器械目录的通告（2021 年第

346 71 号)) 产品, 应按照《医疗器械临床评价技术指导原则》
347 中同品种医疗器械比对或临床试验途径提交临床评价资料,
348 应基于核心功能或核心算法, 结合预期用途和成熟度予以综
349 合考虑。

350 1. 通过同品种医疗器械对比路径

351 非辅助决策类功能基于核心功能开展同品种医疗器械
352 比对, 全新的功能、算法和用途原则上均需开展临床评价;
353 辅助决策类功能基于核心算法开展同品种医疗器械比对, 所
354 选同品种医疗器械的临床证据原则上需基于临床试验 (含回
355 顾性研究)。

356 2. 通过临床试验途径提交临床评价

357 2.1 试验目的

358 应根据产品适用范围确定临床试验目的。适用范围考虑
359 因素包括但不限于: 目标疾病 (如宫颈病变)、临床用途 (如
360 辅助筛查、辅助诊断、辅助筛查/诊断等)、禁忌症、预期适
361 用人群 (如流行病学史, 临床表现的患者)、预期使用场所
362 (如医院、第三方检验机构)、预期目标用户 (如病理医师)、
363 预计兼容的采集设备 (如电子显微设备/显微扫描设备) 和处
364 理对象 (文件格式) 等。

365 2.2 试验设计

366 临床试验设计应优先考虑对照设计。当产品输出的异常
367 结果为某个疾病类型, 或对目标疾病的病变病灶进行辅助检
368 测, 可用于指导下一步诊疗意见时, 建议采用对照设计。当
369 产品输出结果不能给出某种疾病判读建议, 只能提供辅助决
370 策建议时, 可以考虑单组目标值设计。

371 (1) 采用单组目标值试验设计时，以诊断的敏感性、特
372 异性等固有诊断准确度指标作为主要评价指标。生产企业应
373 明确目标值的确定依据。

374 (2) 当采用对照设计时，对照组通常为采用疗效/安全性
375 公认的已上市器械或标准诊断方法。各个指标的优劣效设计
376 应当根据产品的预期目标进行设定。如若将敏感性、特异性
377 和时间效率作为次要指标，并非每个指标都需要进行优效对
378 照，例如预期以提高辅助诊断时间效率为主要目标的软件，
379 其中敏感性、特异性可为非劣性对照，时间效率指标应当为
380 优效对照。

381 2.3 受试对象

382 应基于目标疾病流行病学特征和产品应用场景，考虑人
383 种、年龄范围、地域等因素的影响，尽可能具有代表性。入
384 选样本应包含阳性样本和阴性样本，同时考虑病例组和对照
385 组样本的确立方法。需综合考虑以下因素：

386 (1) 临床试验纳入病例应独立于产品开发所用病例；

387 (2) 临床试验纳入病例需来源于不同地域医疗机构；

388 (3) 选择性入组病例可能导致研究人群与预期适用人群
389 存在较大差异，建议连续入组符合入排标准的病例，以消除
390 或减小选择性偏倚；

391 (4) 需考虑阳性病例中病种分型、分类等情况的分布合
392 理性。

393 2.4 评价指标

394 根据产品预期用途适用人群、病症（部位）和临床功能
395 （定量、定性、分类等）选择包括灵敏度、特异性、阳性预
396 测值、阴性预测值、ROC 曲线和 AUC 值作为主要观察指标。

397 次要评价指标可包括时间效率（如诊断平均时间增益
398 率）、数据有效使用率、软件易用性（可采用主观感受评价，
399 如李克特量表等）和安全性等。若同时观察量化分析等功能
400 的临床安全有效性，可设立相应次要评价指标。

401 注：诊断平均时间增益率定义：单位时间、单位人员条
402 件下，人工智能辅助诊断平均时间和人工诊断平均时间差值
403 与人工智能辅助诊断平均时间和人工诊断平均时间中高值
404 的比例。计算方法：

$$\begin{aligned} & \text{单位时间、单位人员条件下,} \\ 405 \quad \text{诊断平均时间} &= \frac{\text{人工智能辅助诊断平均时间和人工诊断平均时间差值}}{\text{人工智能辅助诊断平均时间和人工诊断平均时间中的高值}} \times 100\% \\ & \text{增益率} \end{aligned}$$

406 2.5 标准方法构建

407 生产企业应详述金标准的选择、构建方法及理由。可供
408 选择的金标准构建方法包括但不限于：一是以临床确定结果
409 为金标准，即流行病学史、临床表现（含影像学特征）的综
410 合诊断结果；二是通过专家阅片小组构建金标准。

411 2.6 样本量

412 样本量估算需综合考虑临床试验设计、主要评价指标和
413 统计学要求。需明确计算公式、相应参数及确定理由，以及

414 所用的统计软件。同时应充分考虑可能的脱落剔除率等情况，
415 合理设定样本量要求。

416 对于复合终点，样本量估算需考虑满足所有单项指标的
417 假设检验的样本量需求。

418 （五）产品说明书和标签标识

419 说明书、标签和包装标识需符合《医疗器械说明书和标签
420 管理规定》和《医疗器械软件注册审查指导原则（2022 年修
421 订版）》、《医疗器械网络安全注册审查指导原则（2022 年
422 修订版）》、《人工智能医疗器械注册审查指导原则》和相
423 关标准的规定。

424 说明书内容需重点关注：

425 1.用户说明

426 对预期用户和推荐用户培训的详细说明。如，预期用户
427 工作年限或执业资格要求，且需经培训合格。

428 2.使用限制

429 若产品采用人工智能黑盒算法，需根据算法影响因素分
430 析报告，在说明书中明确产品使用限制和必要警示提示信息。

431 3.软件

432 需明确软件发布版本、提供网络安全说明和使用指导，
433 明确用户访问控制机制、电子接口（含网络接口、电子数据
434 交换接口）及其数据类型和技术特征、网络安全特征配置、

435 数据备份与灾难恢复、运行环境（含硬件配置、外部软件环
436 境、网络环境，若适用）、安全软件兼容性列表（若适用）、
437 外部软件环境与安全软件更新（若适用）、现成软件清单等
438 要求。

439 4.产品接口和联合使用设备

440 需明确对配合使用的图像扫描设备、制片设备型号、参
441 数和厂家信息等。

442 5.算法训练总结

443 训练方法概述、开发数据库基本信息、训练指标与结果、
444 调优指标及结果。

445 6.算法性能评估总结

446 -算法输入与输出

447 -测试集基本信息

448 -用于确定器械标记的每个区域的性质的评分标准

449 -每个可用器械操作点的总体敏感度和假阳性率指标

450 -分层分析（如，根据病变大小、病变类型、采集参数、
451 成像或数据特征）

452 -独立 FROC 性能（如适用），需和操作特性曲线一起说
453 明。

454 -测试结果

455 7.临床试验总结

456 需包括临床试验设计基本类型、研究对象（受试者及阅
457 片者情况）、评价指标，金标准、对收集临床信息方法的描
458 述、统计方法描述、样本量，临床试验结果。

459 8.软件生存周期管理的注意事项；

460 9.明确处理对象（图像）的质量要求；

461 10.明确兼容的数据采集设备和数据采集过程相关要求。

462 11.明确采样、制片方式、染色、扫描成像的标准操作流
463 程。

464

465 三、编写单位

466 广东省药品监督管理局审评认证中心、广州锑元方青医
467 疗科技有限公司、深圳视见医疗科技有限公司、腾讯医疗健
468 康（深圳）有限公司

469 四、参考文献：

- 470 [1] 医疗器械监督管理条例[Z].
- 471 [2] 医疗器械注册与备案管理办法[Z].
- 472 [3] 医疗器械说明书和标签管理规定[Z].
- 473 [4] 医疗器械注册申报资料要求和批准证明文件格式[Z].
- 474 [5] 医疗器械通用名称命名指导原则[Z].
- 475 [6] 医疗器械产品技术要求编写指导原则[Z].
- 476 [7] 医疗器械软件注册审查指导原则[Z].
- 477 [8] 人工智能医疗器械注册审查指导原则[Z].
- 478 [9] 人工智能类医用软件产品分类界定指导原则[Z].
- 479 [10] 移动医疗器械注册技术审查指导原则[Z].
- 480 [11] 医疗器械安全和性能的基本原则[Z].
- 481 [12] 医疗器械临床评价技术指导原则[Z].
- 482 [13] 肺结节 CT 图像辅助检测软件注册审查指导原则 [Z].
- 483 [14] YY/T 0316, 医疗器械风险管理对医疗器械的应用[S].
- 484 [15] YY/T 1833.1, 人工智能医疗器械质量要求和评价 第 1
485 部分: 术语[S].
- 486 [16] YY/T 1833.2, 人工智能医疗器械质量要求和评价 第 2

- 487 部分：数据集通用要求[S].
- 488 [17]YY/T 1833.3，人工智能医疗器械 质量要求和评价 第 3
- 489 部分：数据标注通用要求[S].
- 490 [18] T/CSBME 050，宫颈液基细胞人工智能医疗器械质量要
- 491 求和评价第 1 部分：数据集要求[S].
- 492 [19]T/GZBC 8，病理人工智能辅助诊断系统技术要求（宫颈
- 493 液基细胞）[S].
- 494 [20]GB/T 25000.51,软件工程 软件产品质量要求与评价
- 495 （SQuaRE）商业现货（COTS）软件产品的质量要求和测试
- 496 细则[S].
- 497 [21]AIOSS-01，人工智能深度学习算法评估规范[S].
- 498 [22]宫颈液基细胞学的数字病理图像采集与图像质量控制中
- 499 国专家共识[S].
- 500 [23]基于病理图像的宫颈癌人工智能辅助系统临床试验规范
- 501 （草稿）[S].
- 502

附录

宫颈细胞学图像人工智能医用软件技术审评要点

一、适用范围

本附录适用于宫颈细胞学病理人工智能医用软件的注册技术评审，软件组件可采用适用部分。

二、注册审查要点

（一）产品名称

主要由“特征词+核心词”结构编制，特征词可描述使用部位构成（宫颈）、作用对象（细胞学病理图像、数字病理图像、显微图像等）；核心词可描述为辅助诊断/辅助检测/辅助筛查软件，如宫颈细胞学病理图像辅助筛查软件。

（二）综述资料

1. 产品工作原理

目前宫颈细胞学病理人工智能软件主要基于以下基础医学技术来实现临床功能：

基于 TBS（宫颈细胞学报告系统）标准的细胞形态分析

（1）对病变细胞进行检测、分类和统计，提供临床参考信息，辅助医生对样本的病变性质作出诊断。

（2）对病变细胞进行检测、分类和统计，并给出样本病变性质的判读参考，供医生进行复核。

524 2. 产品的结构和组成

525 物理组成可描述为软件以光盘、U 盘、网络等形式交付。
526 逻辑组成可描述为功能模块包括数据采集模块、数据处理模
527 块、基于深度学习的疾病评估模块、临床辅助诊断及输出模
528 块、人工干预审核及修正模块、用户访问入口及报告接口模
529 块、数据安全模块等组成。

530 3. 产品的适用范围

531 运用人工智能算法对宫颈细胞学图像中的细胞进行测量/
532 检测/分割/分类，从而为医生对宫颈异常细胞/疾病的诊断提
533 供辅助信息/初步诊断信息。

534 (三) 非临床资料

535 1. 产品技术要求

536 1.1 通用要求

537 包括产品处理对象、软件的使用限制、输入输出、接口、
538 必备软硬件、运行环境、性能效率、最大并发数、用户界面、
539 消息、用户差错防御、访问控制、版权保护、可靠性、维护
540 性等。

541 1.2 临床功能要求

542 明确所有临床功能纲要。例如，支持多功能阅片（如浏
543 览阅片、标注、测距等），阅片时间（如平均每例 AI 阅片处
544 理时间应该小于 10 秒），宫颈细胞病变部位的识别检测，如
545 包括特征计算（如细胞及细胞核大小、分布、数量、胞浆比
546 等）、分类（一般能进行 15 分类（ASC-US（非典型性鳞状

547 细胞,意义不明确)、LSIL(低级别鳞状上皮内病变)、ASC-
548 H(非典型鳞状细胞,不除外高级别鳞状上皮内病变)、HSIL
549 (高级别鳞状上皮内病变)、SCC(鳞状细胞癌)、AGC-
550 NOS(非典型腺细胞,意义不明确)、AGC-FN(非典型腺
551 细胞,倾向肿瘤性)、AIS(子宫颈管原位腺癌)、ADC(宫
552 颈腺癌)、子宫内膜细胞、念珠菌、滴虫、单纯疱疹病毒、
553 放线菌、线索细胞),系统自动审核(如适用,即能对阴性
554 (正常类)诊断自动审核,且自动审核结果的符合率应符合
555 临床要求),量化统计分析,数据对比(如手动、自动均可),
556 切片图像质控,人工审核及修正,图像放大缩小等。

557 其他辅助功能包括:患者信息查询,图像检索,诊断报
558 告输出,辅助业务质控,远程会诊,云计算等。

559 若含有基于测评数据库测试的性能指标,其要求参考
560 《人工智能医疗器械注册审查指导原则》。

561 2.软件研究

562 要求同正文部分。

563 3.算法研究

564 3.1 算法基本信息

565 需提供算法设计和功能的信息,提供流程图及注释,明
566 确算法名称、版本、软件平台的相关特性。

567 需描述算法每个阶段的设计和功,明确所用算法的名
568 称(如基于 U-net 算法)、结构(如层数 20 层、参数规模:
569 如 700 万个浮点数)、流程图、现成框架(一般使用
570 Tensorflow)、输入与输出、运行环境(如 Linux)、算法来
571 源依据(如基于 U-net 原生算法)等信息。

572 3.2 算法需求规范

573 算法功能需求建议关注数据库需求、算法性能评价指标
574 及制定依据。

575 3.2.1 数据库需求

576 算法性能评价需要基于训练和测试数据库。

577 3.2.2 算法性能评价指标

578 在指定检出算法评估指标之前，需明确标记匹配的方式，
579 即算法标记目标与参考标准目标的匹配方式；明确算法任务，
580 明确算法训练和调优过程中不同任务的的评估指标及定义。

581 明确不同任务的算法评价指标的确定依据、分层影响因
582 素选择依据，建议参考相关疾病的中国专家共识、国内外指
583 南、相关标准等。

584 检出指标包括不限于：召回率、精确度、特异度。

585 分类指标包括不限于：灵敏度、特异度、总体的 Kappa
586 系数、准确率。

587 分割指标包括不限于：Dice 系数、交并比。

588 注：指标定义可参考 YY/T 1833.1-2022《人工智能医疗
589 器械 质量要求和评价 第 1 部分:术语》。

590 3.3 数据质控

591 建议参考《人工智能医疗器械注册审查指导原则》(2022
592 年第 8 号)及 YY/T 1833.2-2022《人工智能医疗器械质量要
593 求和评价 第 2 部分：数据集通用要求》。

594 用于病理 AI 软件算法训练的直接数据为数字图像文件；
595 原始数据为病理切片，切片经数字化、预处理和标注后使用。

596 3.3.1 数据采集

597 数据采集首先应定义采集的目标和用途，并明确数据采
598 集源和采集范围；数据来源于与产品预期用途有代表性的地

599 域、不同级别（如基层、三甲）的多家医疗机构，选择与产
600 品预期使用相适应的制片方式、制片设备、成像设备。如能
601 适用多厂家、多型号设备时，应考虑数据来源于代表性设备
602 （涵盖不同参数）。

603 3.3.2 数据整理

604 明确数据转移保存的方法。明确数据纳入排除标准，以
605 及进行数据筛选的方法（人工清洗、自动清洗），提供数据
606 整理软件工具的研究资料。

607 提供数据清洗流程图，明确清洗规则，例如确定数据合
608 规性、连续性、完整性、已完成脱敏等并加以筛选。清洗结
609 果明确弃用数据的数量和原因。明确预处理的操作步骤和内
610 容。

611 3.3.3 数据标注

612 数据标注建议参考 YY/T 1833.3-2022《人工智能医疗器械
613 质量要求和评价 第 3 部分：数据标注通用要求》。

614 数据标注应当考虑标注资源管理、标注过程质控、标注
615 质量评估等要求。需明确标注任务分类（包括数据模态（宫
616 颈液基细胞图片的数字病理图像）、执行主体（人工标注、
617 自动标注等）、标注结果格式（如 jpeg、tiff、jpeg2000）、
618 标注结果性质（给出病理图像的分类）、标注结果形式等维
619 度），提供标注任务描述文档（标注规则（专家组根据文献、
620 专家共识等共同给出）、标注人员、标注工具（商用现成软
621 件 COTS/开源软件/自制软件）、标注环境、数据）。

622 标注与质控流程建议提供业务架构、过程组织（任务生
623 成、任务分配、任务实施、质量控制、验收准则及验收报告）。
624 其中业务架构宜采用流程图介绍单例数据的标注、审核、仲

625 裁过程。

626 如采用的标注工具、标注平台使用了人工智能算法进行
627 辅助标注，需提交标注工具、标注平台算法性能研究资料。

628 目前宫颈液基数字图像标注形式可分为矩形边界框标
629 注（粗略标注）和多边形轮廓标注（精细标注）。标注方式
630 可依据数据集的预期用途设定。在采用矩形边界框标注时，
631 可参考如下要求：

632 （1）单个细胞标注：以包括细胞浆外轮廓的最小外接矩
633 形进行标注；

634 （2）成片细胞标注：细胞片指出现一些细胞（两个及以
635 上）聚集在一起无法分割而形成的团簇。细胞片以包括片状
636 外轮廓的最小外接矩形进行标注。若两个细胞可以分离，应
637 分开单独标注；

638 （3）成团细胞标注：细胞团指多数细胞（三个及以上，
639 多层）细胞混杂在一起无法分割。细胞团以包括团状外轮廓
640 的最小外接矩形进行标注；

641 （4）病原微生物标注：

642 a.滴虫：以包括滴虫外轮廓的矩形进行标注；

643 b.念珠菌：以包括所有菌丝的矩形进行标注；

644 c.放线菌：以包括放线菌外轮廓的矩形进行标注；

645 d.线索细胞：以包括菌落覆盖的细胞外轮廓的矩形进行
646 标注；

647 e.疱疹：以包括病毒引起形态学改变的细胞的矩形进行
648 标注。

649 （5）当标注区域中含有多个类别的细胞时，宜以最严重的
650 类别进行标注。例如标注区域中既存在 **LSIL**，又存在 **HSIL**，

651 则标注类别应是 **HSIL**。也可以依据数据集的预期用途制定不
652 同标注规则，如当标注区域中含有多个类别细胞时，需在该
653 区域内标注所有类别的细胞。

654 3.3.4 数据集构建

655 依据《人工智能医疗器械注册审查指导原则》指南明确
656 训练集、调优集、测试集的划分方法、划分依据、数据分配
657 比例。

658 明确训练集、调优集、测试集的划分方法、划分依据、
659 数据分配比例。训练集应当保证样本分布具有均衡性，测试
660 集、调优集应当保证样本分布符合临床实际情况，训练集、
661 调优集、测试集的样本应当两两无交集。

662 训练集可基于训练 **epoch** 数量-评估指标曲线评估指标
663 不再继续提升来确定训练数据量，即增加数据量性能不会持
664 续提升，再通过召回率等指标确定算法模型是否达到预期目
665 标来评估目前数据量的合理性。

666 测试集根据每种分类及定义（如二分类、15 分类等）分
667 别计算每个分类的样本量，同时应考虑按照 **TBS** 分类中各病
668 变的流行病学分布，并覆盖全分类，可采用灵敏度和特异度
669 按以下公式计算样本量：

$$670 \quad n = \frac{Z^2 P(1-P)}{d^2}$$

671 其中：

672 n =所需样本量；

673 Z =置信水平 Z 的统计量，如 95%置信水平 Z 的统计量
674 为 1.96；

675 P =灵敏度或特异度的预期值；

676 d = P 的允许误差大小，一般取 P 的 95%置信区间的半宽，

677 常用取值为 0.05~0.1

678 3.4 算法训练

679 算法训练需明确训练过程所采用的优化器及其相关参
680 数，如 Adam, SGD；在算法训练阶段，需明确验证集的划分
681 方式，如留出法，交叉验证法；需明确训练目标，即判断何
682 时停止训练，如设定最大的训练 epoch 数目，依据损失函数
683 判断损失稳定且不继续下降，根据训练 epoch 数量-评估指标
684 曲线判断评估指标不继续提升等。当训练停止后，明确训练
685 模型最佳 epoch 的选择方法。同时，结合临床需求（如灵敏
686 度、精准度），明确算法出厂阈值的选择与方法，并论证训
687 练所得模型是否满足产品既定目标。算法出厂阈值的选择需
688 提供制定依据。

689 3.5 算法验证与确认

690 明确算法任务，明确不同任务的算法评价指标的阈值及
691 确定依据。明确算法标记目标与参考标准的匹配方式和匹配
692 阈值。明确病例水平和目标对象水平计算方法及定义一个目
693 标对象的处理方法及病例阳性的方法。区分算法任务制定具
694 体测试方法和结果计算方法。算法确认通常包括灵敏度、特
695 异性、准确度、阳性预测值、阴性预测值、ROC 曲线和 AUC
696 值。

697 3.6 算法性能综合分析

698 结合算法训练、算法性能评估、临床评价等结果开展算
699 法性能综合评价，针对训练样本量和测试样本量过少、测试
700 结果明显低于算法设计目标、算法性能变异度过大等情况，
701 对产品的适用范围、使用场景、核心功能进行必要限制。

702 需对算法测试产生的错误结果进行分析，评估算法在检

703 出、分类等任务中出现的假阳性、假阴性结果的含义与危险
704 程度，提供算法性能影响因素分析过程，如采集设备、采集
705 参数、疾病构成、病变特征等因素影响，并提供特殊案例及
706 数据分析过程，以提升算法可解释性，以上内容形成定量报
707 告。

708 4. 病理数字图像质量要求

709 参考《宫颈液基细胞学的数字病理图像采集与图像质量
710 控制中国专家共识》，宫颈液基细胞数字图像质量需综合图
711 像清晰度、图像扫描范围、图像噪声等多方面因素进行综合
712 判断：

713 (1) 图像的清晰度：

714 达标：细胞质及细胞核边界清晰，细胞核内部细节可辨；

715 未达标：细胞质及细胞核边界模糊不清，细胞核内部细
716 节难以辨认。

717 (2) 图像有效视野范围要求：

718 达标：图像中细胞分布的范围均在 WSI（全视野的数字
719 切片）视野范围内

720 未达标：图像中细胞分布的范围超出 WSI 视野范围，不
721 论比例。

722 (3) 图像噪声要求：

723 达标：扫描成像未引入玻片不包含的图像噪声，无外部
724 噪声影响阅片诊断；或引入了玻片不包含的图像噪声，如拼
725 接不全时，产生横向/纵向条纹，但不影响细胞判读阅片诊断。

726 未达标：扫描成像引入了玻片不包含的图像噪声，导致
727 细胞不可辨认，或影响细胞判读阅片诊断。

728 (4) 图像整体质量的可信度评估：

729 对宫颈液基细胞图像的整体质量进行评估，需综合上述
730 成像评估原则。图像的上述三项指标均达到“达标”，评定
731 为整体“可信度好”；三项指标中任何一项为“未达标”，
732 评定为“可信度差”。

733 对于可信度良好的图像，可用于辅助医生判读项目的研
734 究测试或应用，并予以存储。对于可信度差的图像，无法提
735 供可靠的判读结果，建议重新扫描并存储。

736 4.1 制片方式与设备

737 4.1.1 制片设备

738 明确制片设备参数要求，制片质量应符合以下控制要求；

739 4.1.2 制片通用要求：

740 (1) 细胞制片满意；

741 (2) 细胞玻片完好或无明显破损，不影响图像完整采集；

742 (3) 图像无明显褪色，细胞染色鲜亮，胞浆和核的结构
743 特征清晰可见；

744 (4) 玻片封片剂不宜过多。

745 4.1.3 制片特殊要求：

746 (1) 对制片方式是否有限制：如膜式、离心式、自然沉
747 降式。

748 (2) 对染色试剂是否有限制：如巴氏为主。

749 4.2 成像设备

750 成像设备根据采集方式可分为实时采集设备和扫描采
751 集设备，一般实时采集设备要控制放大倍数、像素、分辨率、
752 帧速率、色彩等参数；扫描采集设备则要控制扫描倍数、扫
753 描范围、像素、分辨率、色彩等参数。

754 数字切片扫描仪建议满足以下参数要求：

755 (1) 扫描倍数：数字化扫描仪应采用 20 倍或以上倍数的
756 物镜；在满足原始像素密度小于 $0.25\mu\text{m}/\text{pixel}$ 条件时宜采用
757 20 倍物镜；

758 (2) 图像扫描应保证视野成像清晰完整，视野直径大于
759 13mm；

760 (3) 图像色彩模式为 RGB 彩色；

761 (4) 扫描分辨率：分辨率 $\leq 0.25\mu\text{m}/\text{pixel}$ ；

762 5. 互操作性

763 互操作性（又称互用性）是指软件与其他医疗器械或通
764 用设备通过电子接口交换并使用信息的能力。其中电子接口
765 包含硬件接口和软件接口，信息包括但不限于医学图像、生
766 理参数、体外诊断等数据和控制指令。

767 重点关注医疗器械软件的接口设计及其风险，接口包括
768 内部接口和外部接口，注册申请人需考虑软件接口的需求分
769 析、风险管理、验证与确认、维护计划等活动要求，以及说
770 明书与标签的设计要求。

771 申请人需提交一份互操作性研究资料，内容包括基本信
772 息、需求规范、风险管理、验证与确认、维护计划；亦可在
773 自研软件研究报告相关条款中体现软件接口要求，具体要求
774 详见《医疗器械软件注册审查指导原则（2022 年修订版）》。

775 (四) 临床评价

776 1. 通过同品种医疗器械对比路径

777 非辅助决策类功能基于核心功能开展同品种医疗器械
778 比对，应基于核心算法和临床功能开展同品种医疗器械比对，
779 所选同品种医疗器械的临床证据原则上需基于临床试验（含
780 回顾性研究）

781 2.通过临床试验途径评价

782 2.1 试验目的

783 应根据产品适用范围确定临床试验目的。适用范围考虑
784 因素包括但不限于：目标疾病（宫颈细胞异常识别、癌变分
785 类）、临床用途（辅助筛查、辅助诊断、辅助筛查/诊断）、
786 预期适用人群（如流行病学史、临床表现的患者）、预期使
787 用场所（如三甲医院、第三方检验机构）、预期目标用户（如
788 病理医师）、预计兼容的采集设备（电子显微设备/显微扫描
789 设备)和处理对象(数字病理图像,格式为:通用格式如 dcm、
790 tif、tiff、svs 等；私有格式如 SDPC、DYQ 等）。

791 2.2 试验设计

792 该产品一般采用对照设计，对照组通常为采用通过专
793 家阅片小组构建标准方法作为对照组。

794 示例：当产品预期用途为辅助人工阅片时，临床设计可
795 考虑病理图像在经 AI 判读结果后，人工 A 单独进行诊断，
796 与金标准进行比较，人工 B 进行再次诊断，其诊断结果与金
797 标准进行比较，最终评价 AI+人工判断结果的敏感性和特异
798 性优于人工直接诊断（A、B 交互阅片）。临床评价的主要评
799 价指标为：

$$800 \quad \text{敏感性} = \frac{\text{真阴}}{\text{真阴} + \text{假阴}}$$

$$801 \quad \text{LSIL 及以上敏感性} = \frac{\text{LSIL 及以上敏感性}}{\text{LSIL 及以上敏感性} + \text{LSIL 及以上假阴}}$$

$$802 \quad \text{特异性} = \frac{\text{真阴}}{\text{真阴} + \text{假阴}}$$

803 其中:

804 阴性: 图像结果为 NILM(无上皮内病变或恶性病变)。

805 阳性: 图像结果为 ASC-US、LSIL、ASC-H、HSIL、SCC、
806 AGC-NOS、AGC-FN、AIS、ADC。

807 LSIL 及以上: 图像结果为 LSIL、ASC-H、HSIL、SCC、
808 AGC-NOS、AGC-FN、AIS、ADC。

809 真阳: 诊断结果与金标准均为阳性的图像。

810 真阴: 诊断结果与金标准均为阴性的图像。

811 假阳: 诊断结果为阳性, 但金标准结果为阴性的图像。

812 假阴: 诊断结果为阴性, 但金标准结果为阳性的图像。

813 LSIL 及以上真阳: 诊断结果为 LSIL 以上的图像中的真
814 阳。

815 LSIL 及以上假阴: 诊断结果为 LSIL 以上的图像中的假
816 阴。

817 金标准: 是指由三位高年资病理医生组成的专家组(独
818 立于 A、B), 均具备十年以上宫颈液基细胞片诊断经验, 每
819 个中心各挑一位。对于需要查看的片子, 各个医生分别在显

820 显微镜下阅片，并给出各自的诊断结果，诊断结果达成一致（三
821 位医生结果需一致）后，记录并保存，作为金标准。若三位
822 医生结果不一致，需共同在显微镜下阅片，讨论确定最终结
823 果。

824 注：当人工阅片需要多次审阅结果作为最终结果时，考
825 虑合理设定一定的洗脱周期，以消除结果的偏倚。

826 2.3 受试对象

827 应基于目标疾病流行病学特征和产品应用场景，考虑
828 人种、年龄范围、地域等因素的影响，尽可能具有代表
829 性。入选样本应包含阳性样本和阴性样本，同时考虑病例
830 组和对照组样本的确立方法。

831 （1）年龄分布：如 25-60 岁宫颈疾病高发的年龄段。

832 （2）病变种类：根据产品临床功能而定，阳性样本应
833 符合宫颈病变流行病学分布，如覆盖 TBS 标准分类。

834 2.4 评价指标

835 根据产品预期用途适用人群、病症（宫颈病变）和临床
836 功能（定量、定性、分类等）选择敏感性（结合临床设定，
837 如不低于 85%）、特异性（如不低于 85%）、ROC/AUC 作
838 为主要观察指标。以诊断平均时间增益率为次要指标。

839 2.5 标准方法构建

840 生产企业应详述金标准的选择、构建方法及理由。可供
841 选择的金标准构建方法包括但不限于：一是以临床确定结果

842 为金标准，即流行病学史、临床表现（含影像学特征）的综
843 合诊断结果；二是通过专家阅片小组构建金标准。

844 2.6 样本量

845 样本量估算需综合考虑临床试验设计、主要评价指标和
846 统计学要求。需明确计算公式、相应参数及确定理由，以及
847 所用的统计软件。同时应充分考虑可能的脱落剔除率等情况，
848 合理设定样本量要求。

849 对于复合终点，样本量估算需考虑满足所有单项指标的
850 假设检验的样本量需求。

851 宫颈细胞学辅助筛查类软件一般为定性产品，即产品主
852 要是根据图像分析结果对宫颈病变进行分类。在临床试验设
853 计中，试验组的样本量首先应覆盖疾病各阶段，样本量估算
854 需要按照阳性受试者和阴性受试者两部分，分别按照以下公
855 式计算：

$$856 \quad n = \frac{\left[Z_{1-\alpha/2} \sqrt{P_0(1-P_0)} + Z_{1-\beta} \sqrt{P_T(1-P_T)} \right]^2}{(P_T - P_0)^2}$$

857 以上公式中，公式中， n 为样本量 $Z_{1-\alpha/2}$ 、 $Z_{1-\beta}$ 为显著性
858 水平和把握度的标准正态分布的分数位， P_0 为评价指标的临
859 床可接受标准， P_T 为试验体外诊断试剂评价指标预期值。当
860 $\alpha=0.05$ 时， $Z_{1-\alpha/2}=1.96$ ，当 $\beta=0.2$ 时， $Z_{1-\beta}=0.842$ ， P =目标总体
861 的比例期望值； d =置信区间的半宽。

862 示例：以宫颈细胞学辅助筛查类软件为例，假设采用与
863 专家组阅片进行比较研究的方法（非劣效、等效），根据临
864 床需求，低度鳞状细胞病变（LSIL 类）样本的灵敏度性、特
865 异度可分别达到 85% 和 90%，根据探索性试验结果，某产品
866 与对比方法的阳性、阴性符合率预期分别可达到 90% 和 95%。
867 计算如下，

$$868 \quad n_{+} = \frac{[1.96\sqrt{0.85(1-0.85)} + 0.84\sqrt{0.90(1-0.90)}]^2}{(0.90-0.85)^2} = 363$$
$$869 \quad n_{-} = \frac{[1.96\sqrt{0.90(1-0.90)} + 0.84\sqrt{0.95(1-0.95)}]^2}{(0.95-0.90)^2} = 158$$

870 则可知当置信度为 95% (1- α) 情况下，低度鳞状细胞病
871 变（LSIL 类）样本阳性组（n₊）和阴性组（n₋）最低样本量
872 估计分别为 363 和 158, 总体样本需另外考虑脱落剔除率 10%
873 进行计算。