

# “人工智能医学信息系统软件 审评指导体系构建”课题

## ——DR 影像人工智能医疗器械技术审评指导原则 (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 DR 影像人工智能软件的一般要求。

本指导原则是供 DR 影像人工智能软件研发、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 DR 影像人工智能医疗器械的产品注册，包括使用 DR 影像做主要处理模态的人工智能独立软件和含有人工智能软件组件的医疗器械。该产品以 DR 影像为主要处理模态，可以结合其他数据如问诊、临床检验、其他影像模态等，采用以机器学习为主的人工智能技术进行处理分析，临床用于疾病的辅助诊断或其他医学用途。

### 二、注册审查要点

#### （一）监管信息

##### 1. 独立软件

26 产品名称应符合《医用软件通用名称命名指导原则》  
27 《人工智能技术审评通用原则》的要求，通常体现输入数据  
28 模态（DR 影像等）、目标征象（例如肺炎、骨折、骨龄等）、  
29 预期用途（辅助分诊、辅助诊断、辅助检测、辅助测量）等  
30 特征词。例如“骨折 X 射线图像辅助检测软件”、“儿童手  
31 部 X 射线影像骨龄辅助评估软件”。

32 DR 影像人工智能医疗器械从用途角度可分为辅助决策类  
33 和非辅助决策类，按照 2017 版《医疗器械分类目录》和《人  
34 工智能医用软件产品分类界定指导原则》进行分类管理。

35 型号规格应给出产品的型号和软件发布版本。

36 结构组成应明确软件交付方式（如光盘/U 盘安装、网络  
37 下载），结合网络部署情况给出软件功能模块，应体现出包  
38 括人工智能技术相关功能的所有模块（如深度学习）。

39 按照国家药监局发布的注册申报资料要求提供产品的适  
40 用范围、预期使用环境、适用人群、禁忌症，需与申报产品  
41 的性能、功能相符，并需与临床评价资料结论一致。适用范  
42 围应体现出预期用途（目标疾病、临床用途、重要程度、紧  
43 迫程度）、使用场景（目标用户、使用场所、临床流程）、  
44 核心功能（处理对象、数据兼容性、功能类型）、所支持  
45 DR 设备类型要求（内容较多的可在产品技术要求中体现）、  
46 主要使用限制。

## 47 2.软件组件

48 DR 影像的人工智能处理功能可作为软件组件集成于其  
49 他医疗器械（如 DR 机）中时，产品名称、分类编码、型号  
50 规格、结构组成与所集成医疗器械保持一致，其分类应当与

51 系统内风险程度最高的医疗器械一致。在适用范围和产品技  
52 术要求中体现 DR 影像人工智能处理功能的相关信息，软件  
53 功能名称可参照独立软件要求。

## 54 （二）综述资料

### 55 1. 器械原理与架构描述

56 按照《人工智能医疗器械注册审查指导原则》对产品进  
57 行软件架构和算法研究报告。

58 DR 影像人工智能后处理软件的基本框架一般包括：X 射  
59 线影像加载、数据预处理、数据分析、数据可视化等部分。  
60 对于采用监督学习的辅助决策功能，算法研究中应包括学习  
61 模型构建、数据预处理、数据集划分、模型迭代训练、微调  
62 /特征选取、结果验证等内容描述。

### 63 2. 产品的适用范围、适用人群、禁忌症

64 按照国家药监局发布的注册申报资料要求提供产品的适  
65 用范围、预期使用环境、适用人群、禁忌症，需与申报产品  
66 的性能、功能相符，并需与临床评价资料结论一致。

### 67 3. 风险管理

68 DR 影像人工智能医疗器械的软件安全分级按照《YY/T  
69 0664—2020 医疗器械软件 软件生存周期过程》，一般具有  
70 辅助决策功能的应按照 C 类进行管理。

## 71 （三）非临床资料

### 72 1. 产品技术要求

#### 73 1.1 规格信息

74 明确软件发布版本和版本命名规则。软件版本命名规则  
75 原则上需涵盖算法驱动型更新和数据驱动型更新，明确并区

76 分重大软件更新和轻微软件更新，其中重大软件更新列举常  
77 见典型情况。

78 应给出 DR 设备的兼容性要求，DR 设备应满足行业标准  
79 要求（《YY/T 0106-2020 医用诊断 X 射线机通用技术条件》、  
80 《YY/T 0741-2018 数字化摄影 X 射线机专用技术条件》、  
81 《YY/T 0742-2021 胃肠 X 射线机专用技术条件》）等，明  
82 确列出经测试可支持的 DR 型号、厂家等。

83 拍摄要求至少应明确曝光方式（静态、动态）、剂量条  
84 件（管电压、管电流、时间、AEC 等）、拍摄部位、患者体  
85 位、拍摄距离等要求，并提供相关信息。

## 86 1.2 性能指标

87 应制定软件的所有临床功能。对进行人工智能处理功能  
88 应给出预期的性能描述指标（例如 AUC 曲线、特异性、敏  
89 感性、信噪比、DICE 指标等），若含有基于测评数据库测  
90 试的性能指标，按照《YY/T 1833.2 人工智能医疗器械质量  
91 要求和评价 第 2 部分：数据集通用要求》和《人工智能医  
92 疗器械注册审查指导原则》对数据集和测试条件进行说明。

93 对辅助决策的功能，如进行病灶检出、分类、分期等应  
94 给出参考的医学证据（标准、共识、指南或者文献等），或  
95 者所用规则的依据、原因和方法等。对决策建议功能明确生  
96 成的所有诊断结果、转诊建议、图像质量判定的措辞等。

97 明确运行效率（如时间、周期等），以及主要的步骤（如  
98 上传、处理、返回、显示）。运行环境，运行在不同计算机  
99 系统的产品模块（如客户端和本地服务器、云端服务器）应  
100 分别描述其运行环境。

101 给出潜在的性能风险的分析和处理措施（例如：

102 1) 存在对图像泛化能力不足的风险，DR 型号、图像采  
103 集模式、摆位、图像质量不足（成像面积、分辨率、对比度、  
104 噪声、伪影等）会出现软件处理失败或者错误，应明确不合  
105 格图像的处理策略和风险提示；2) 人工智能处理自身的性  
106 能存在设计上特异性和敏感性的预期，发生一定的误诊和漏  
107 诊；3) 人工智能医疗器械的临床评价只在一定程度评价器械  
108 的性能，不能覆盖在实际应用场景中全部情形；4) 进口软  
109 件因中外人种、流行病学特征、临床诊疗规范等方面差异的  
110 影响，软件准确度存在偏倚。）

## 111 2. 软件研究资料

112 注册申请人应依据《医疗器械软件技术审查指导原则》、  
113 《医疗器械网络安全技术审查指导原则》、《移动医疗器械  
114 注册技术审查指导原则》、《人工智能医疗器械注册审查指  
115 导原则》等指导原则，提交相应注册申报资料。

## 116 3. 算法研究资料

### 117 3.1 算法基本信息

118 算法选择应当明确所用模型的名称（如决策树、随机森  
119 林、支持向量机，或多种模型的结合）、结构（如层数、参  
120 数规模、数学基础）、流程图、开发语言、学习平台（如  
121 Tensorflow、Pytorch）、输入与输出（输入的数据格式数量、  
122 输出的结果类型参数）、运行环境、算法来源依据（参考文  
123 献、相关专利、注明原创）等信息。同时应当明确算法选择  
124 与设计的原则、方法和风险考量，如量化误差、梯度消失、  
125 过拟合、白盒化等。

126 若使用迁移学习技术，除上述内容外还应当补充预训练  
127 模型的数据集构建、验证与确认等总结信息。

### 128 3.2 算法需求规范

129 产品设计应充分考虑 DR 相关诊疗指南、方案（跟踪最  
130 新版本）。明确产品的使用场景、使用者、适用人群，如允  
131 许操作者不同于报告签署人，应分别描述。明确图像质量控  
132 制的方式，即达到何种要求的图像可供算法分析，尤其应描  
133 述算法在图像质量不佳情形下的处理方法（如：不进行分析，  
134 或在给出分析结果时明确提示图像质量不佳的负面影响）。  
135 明确图像质控过程中使用者和算法（若有）的职责。

136 明确算法性能设计目标，应给出人工智能处理功能的性  
137 能指标设定依据，应能够满足预期临床使用需求，其他辅助  
138 功能参考《人工智能医疗器械注册审查指导原则》的建议设  
139 定性能设计目标。

### 140 3.3 算法训练

141 算法训练需要基于训练集、调优集进行训练和调优，应  
142 当明确评估指标、训练方法、训练目标、调优方法、训练数  
143 据量-评估指标曲线等要求，应当明确测试集的数量和分布情  
144 况。

145 评估指标建议根据临床需求进行选择，如敏感性、特异  
146 性等。训练方法包括但不限于留出法和交叉验证法。训练目  
147 标应当满足临床要求，提供 ROC 曲线等证据予以证实。调  
148 优方法应当明确算法优化策略和实现方法。训练数据量-评估  
149 指标曲线应当能够证实算法训练的充分性和有效性。

150 应基于《人工智能医疗器械审查指导原则》第四章生存

151 周期过程、第五章技术考量、第六章算法研究资料的所有相  
152 关要求提供算法研究报告或算法更新研究报告，以及人机交  
153 互研究资料，认为不适用的应详述理由。

### 154 3.4 数据集

155 按照《YY/T 1833.2 人工智能医疗器械质量要求和评价  
156 第 2 部分：数据集通用要求》和《人工智能医疗器械审查指  
157 导原则》的要求对 DR 影像数据集进行管理和质量评价。

158 应当明确数据脱敏的类型（静态、动态）、规则、程度、  
159 方法，选用软件工具的名称、型号规格、完整版本、供应商、  
160 运行环境、确认等要求，同时考虑数据脱敏选用方法对软件  
161 的影响及其风险。自主开发的软件需要验证脱敏效果有效  
162 性。

163 应当明确数据处理的规则、方法（如滤波、增强、重采  
164 样、尺寸裁剪、均一化等），选用软件工具的名称、型号规  
165 格、完整版本、供应商、运行环境、确认等要求，同时考虑  
166 数据处理选用方法对软件的影响及其风险。自主开发的软件  
167 需要验证脱敏效果有效性。

168 明确经数据清洗弃用数据的原因和数量。数据预处理环  
169 节应采用典型图片和流程图说明处理过程和每一步处理的  
170 结果。原始数据库的预处理过程若与产品中算法运行的预处  
171 理过程存在差异，应通过比较来说明。

172 标注过程应按照《YY/T 1833.3-2022 人工智能医疗器械  
173 质量要求和评价 第 3 部分：数据标注通用要求》的要求。  
174 质控应当建立数据标注操作规范，明确标注人员（如资质、  
175 数量、职责）、标注流程（如标注对象、标注形式、标注轮

176 次、标注步骤、操作要求）、临床诊疗规范（如临床指南、  
177 专家共识）、分歧处理（如仲裁人员、仲裁方式）、可追溯  
178 性（如数据、操作）等要求。

179 对于标注前的基础数据库，标注后划分的训练集、调优  
180 集、测试集，应给出样本量和分布情况及其确定依据，以及  
181 集合划分的方法、依据。

### 182 3.5 算法验证与确认

#### 183 (1) 算法性能评估

184 基于测试集提供假阴性与假阳性、重复性与再现性、鲁  
185 棒性/健壮性等算法性能指标评估结果，以证明算法性能满足  
186 设计目标。亦可基于第三方数据库开展算法性能评估。

187 若有图像质量判定模块，应选择合适的评价指标验证图  
188 像质量分级算法性能。

#### 189 (2) 算法性能综合评价

190 按照《人工智能医疗器械注册审查指导原则》开展算法  
191 性能比较分析和算法性能综合评价。自建测试数据集测试、  
192 临床试验、真实世界数据测试、第三方数据库测试、压力测  
193 试、对抗测试等均可视为算法性能评估场景。按照《YY/T  
194 1833.2 人工智能医疗器械质量要求和评价 第 2 部分：数据  
195 集通用要求》对数据集进行描述。

196 “分析算法性能影响因素及其影响程度，如采集设备、采  
197 集参数、训练数据分布、疾病构成、病变特征等因素对假阴  
198 性与假阳性、重复性与再现性、鲁棒性/健壮性影响，以提升  
199 算法可解释性，并列举对算法性能质量控制措施，并作为软  
200 件验证、软件确认的基础。

### 201 3.6 人机交互研究资料

202 应充分考虑人机交互，对人因工程和人机协同的设计给  
203 出说明。

### 204 4. 网络安全研究资料

205 网络安全方面除遵循《医疗器械网络安全注册技术审查  
206 指导原则》之外，由于深度学习算法的特殊性，还需要遵循  
207 《人工智能医疗器械注册审查指导原则》中关于网络安全和  
208 数据安全过程控制的要求。

### 209 5. 系统方案实现与测试

210 系统根据医疗器械质量管理规范，进行系统立项、需求  
211 调研、设计方案评审、结果验证和确认，并按照风险管理相  
212 关规范进行风险分析评估等。

213 软件根据全生命周期管理需进行软件验证与软件确认。

214 软件验证测试（单元测试、集成测试、系统测试）、设  
215 计评审等系列活动。

216 软件确认是指通过提供客观证据认定软件满足用户需求  
217 和预期目的，包括软件确认测试（用户测试）、临床评价、  
218 设计评审等系列活动，其中软件确认测试应当基于软件需求  
219 在真实或模拟使用场景下予以实施。

### 220 6. 云计算服务、移动计算终端要求

221 关于运行环境兼容性，依据《医疗器械软件注册审查指  
222 导原则》：每个互不兼容的运行环境（含云计算）需作为一  
223 个检测单元分别检测；进行兼容性验证，对于 B/S 架构的软  
224 件，还应提供所有声称可兼容的客户端浏览器的兼容性测试  
225 报告。

226 软件若采用云计算服务，支持云端部署，需考虑云计算  
227 服务要求。若企业自建云计算平台，注册人应遵循云服务商  
228 相关规定，参照自研软件要求提交相应研究资料。若企业外  
229 部采购云服务，提供云服务提供商出具的安全等级测评报  
230 告、云服务的验证确认报告、云服务器数据存储、管理、利  
231 用的权限和相关协议。具体要求参考《医疗器械软件技术审  
232 查指导原则（2022 年版）》中云计算要求。

233 软件若运行于移动计算终端，应综合考虑移动计算终端  
234 的技术特征及其风险。提供可配用移动设备及操作系统与产  
235 品兼容性研究资料。具体要求参考《移动医疗器械注册技术  
236 审查指导原则》。

#### 237 （四）临床评价

238 对未列入《免于临床评价医疗器械目录》的产品，需按  
239 照《医疗器械临床评价技术指导原则》开展临床评价。临床  
240 评价的输入主要是来源于临床试验报告、临床文献和临床经  
241 验的临床数据。也可按照《真实世界数据用于医疗器械临床  
242 评价技术指导原则（试行）》将真实世界证据作为补充。

##### 243 1. 试验目的

244 注册人应根据产品适用范围确定临床试验目的。适用范  
245 围考虑因素包括但不限于：目标疾病（如肺炎、肺结节、肺  
246 气肿、肺结核、骨龄）、临床用途（如分诊提示、异常识别、  
247 疾病类型分析、病程阶段分析、提供辅助诊断结果、提供医  
248 学建议）、禁忌症（如急性疾病、孕妇、婴儿、哺乳期妇女  
249 等对 X 射线高度敏感人员、如再生障碍性贫血不宜接触 X  
250 射线的病症、合并其他病症）、预期适用人群（如具有目标

251 疾病临床表现的患者）、预期使用场所（体检机构、医院、  
252 诊所）、预期目标用户（如放射科医师）、预期兼容的 DR  
253 设备等。

254 临床试验目的主要对临床用途（分诊提示、异常识别、  
255 疾病类型分析、病程阶段分析、提供辅助诊断结果、提供医  
256 学建议）的诊断准确度进行确认，确认试验器械是否优于/  
257 等效于/非劣于已上市同类产品或单组目标值。

## 258 2. 试验设计

259 对采用临床试验方式进行临床评价的产品，辅助决策功  
260 能临床试验建议采用对照设计，试验组为医师与申报产品  
261 共同检测，对照组为传统检测诊断方法（如临床医师的阅片  
262 /综合诊断）。临床试验比较类型应能够体现产品受益风险的  
263 可接受性，建议考虑优效性设计。

264 实施机构应当具备代表性和广泛性，不同于训练数据主  
265 要来源机构，地域分布尽可能广泛，机构数量尽可能多，以  
266 确认算法泛化能力。

## 267 3. 受试对象

268 受试对象入排标准应当基于目标疾病（如肺气肿、肺结  
269 核、气胸、关节炎）临床特征，保证阳性病例和阴性病例选  
270 取的合理性和充分性。在此基础上，病例的选择还需要综合  
271 考虑以下因素：

- 272 1) 临床试验纳入病例应独立于产品开发所用病例；
- 273 2) 临床试验纳入病例需来源于代表地域医疗机构。
- 274 3) 临床试验纳入阳性病例中目标疾病的类型、良恶性、  
275 病程阶段分布合理性。

#### 276 4.评价指标

277 针对采用人工智能算法设计的临床功能原则上选择敏感  
278 性、特异性、受试者工作特性曲线（ROC）曲线下面积（AUC）  
279 为主要评价指标，亦可在此基础上根据软件特点选择敏感性  
280 /特异性衍生指标、受试者工作特性曲线（ROC）曲线下面积  
281 （AUC）衍生指标如诊断准确率、精确率、召回率）、F1  
282 系数、组内相关系数、Kappa 系数、时间效率（如辅助诊断  
283 平均时间、日人均诊断量增益率和诊断平均时间增益率）、  
284 数据有效使用率等指标作为次要评价指标。

#### 285 5.金标准构建

286 注册人应详述金标准的选择、构建方法及理由。可供选  
287 择的金标准构建方法包括：一是以临床确定结果为金标准，  
288 即流行病学史、临床表现（含影像学特征）的综合诊断结果；  
289 二是通过专家阅片小组构建金标准。适用的临床诊疗规范有  
290 非影像学证据作为金标准，且相关证据获得符合伦理要求，  
291 优先选用方法一。

292 若选用方法二，需明确专家数量、来源科室、专家资质  
293 要求（如职称）、专家相关培训要求、培训效果评测（判定  
294 一致性）、结果判定标准、结果判定规则（如少数服从多数，  
295 双人背靠背评判，高水平医师仲裁）、专家抽取标准（随机  
296 抽取或顺序轮转）等。建议由来自影像科、呼吸科等多个科  
297 室的医师联合阅片，以减少不完美金标准的偏倚。

#### 298 6.样本量

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

301 以及所用的统计软件。

302 对于复合终点，样本量估算需考虑满足所有单项指标的  
303 假设检验的样本量需求。假设软件的敏感性为 95%，临床认  
304 可的目标值为 90%，则在双侧显著性水平 0.05、把握度 80%，  
305 至少需 231 例阳性病例。假设软件的特异性为 85%，临床认  
306 可的目标值为 80%，则在双侧显著性水平 0.05、把握度 80%，  
307 至少需 466 例阴性病例。在此基础上考虑 5%-10%脱落率确  
308 定最终样本量。

#### 309 （五）说明书

310 说明书应符合《医疗器械说明书和标签管理规定》相关  
311 要求。明确适用范围、禁忌症、使用限制、注意事项、DR  
312 设备兼容性与扫描参数要求、DR 影像质量要求、算法输入  
313 与输出、算法训练总结、算法性能评估总结、软件临床评价  
314 总结、软件发布版本、软件运行环境、使用期限等内容。

#### 315 （六）质量管理体系文件

316 应当明确产品生产研发过程，可采用流程图的形式，并  
317 说明其过程控制点。按照《YY/T 1833.2-2022 人工智能医疗  
318 器械质量要求和评价 第 2 部分：数据集通用要求》提供数  
319 据集资料。

### 320 三、其他说明

321 （一）软件需根据《人工智能医疗器械技术审查指导原  
322 则》提交软件描述文档，明确软件发布版本号和完整版本号，  
323 如不同模块独立编制版本号，需列出各模块版本号信息。

324 （二）采用深度学习技术算法应提供算法研究资料，包  
325 括数据多样性分析、数据标注质控、训练数据量-评估指标曲

326 线、算法性能评估、算法性能影像因素分析、重复性和再现  
327 性等研究资料，证实算法性能满足设计要求。

328 （三）通过专家阅片构建金标准，应对培训、考核方式、  
329 使用设备的可行性进行研究并提交相应研究结果。

330 （四）若软件可以配用多种型号 DR 影像设备影像，需  
331 对每种可配用的影像设备兼容性进行研究（通常进行接口、  
332 图像、算法测试，确认结果是否符合要求），并提供研究结  
333 果。

334 （五）软件支持在移动设备上运行，还对移动设备的硬  
335 件及操作系统进行兼容性研究。软件若具备电子数据交互、  
336 远程控制或用户访问功能，需根据《医疗器械网络安全注册  
337 技术审查指导原则》的要求提交自研软件网络安全研究报  
338 告。软件若采用云计算技术（部署第三方云服务器），需根  
339 据《医疗器械软件技术审查指导原则（2022 年修订版）》中  
340 云计算的相关要求提供相关研究资料。

341 （六）考虑人机协同问题，通过对用户操作进行培训，  
342 并进行考核减少偏差。

### 343 三、编写单位

344 广东省药品监督管理局审评认证中心、广州柏视医疗科  
345 技有限公司、深圳迈瑞生物医疗电子股份有限公司

### 346 四、参考文献

347 [1] 《医疗器械注册与备案管理办法》 [Z].

348 [2] 《医疗器械说明书和标签管理规定》 [Z].

349 [3] 《医疗器械注册申报资料要求和批准证明文件格式》  
350 [Z].

- 351 [4] 《医疗器械通用名称命名指导原则》 [Z].
- 352 [5] 《医疗器械安全和性能基本原则》 [Z].
- 353 [6] 《医疗器械软件注册审查指导原则（2022 年修订版）》
- 354 [Z].
- 355 [7] 《医疗器械网络安全注册审查指导原则（2022 年修订
- 356 版）》 [Z].
- 357 [8] 人工智能医疗器械注册审查指导原则 [Z].
- 358 [9] 《人工智能类医用软件产品分类界定指导原则》 [Z].
- 359 [10] YY/T 0316-2016 医疗器械 风险管理对医疗器械的
- 360 应用[S]
- 361 [11] YY/T 1833.1-2022 人工智能医疗器械 质量要求和评
- 362 价 第 1 部分：术语[S].
- 363 [12] YY/T 1833.2-2022 人工智能医疗器械 质量要求和评
- 364 价 第 2 部分：数据集通用要求 [S].
- 365 [13] YY/T 1833.3-2022 人工智能医疗器械 质量要求和评
- 366 价 第 3 部分：数据标注通用要求 [S].
- 367 [14] 《医疗器械产品技术要求编写指导原则（2022 年修
- 368 订版）》 [Z].