

# “人工智能医学信息系统软件 审评指导体系构建”课题

## ——超声影像人工智能诊断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”课题的组成部分，内容上基于《人工智能医疗器械审评指导原则》的通用要求，细化了超声影像人工智能诊断软件的一般要求。

本指导原则是供超声影像人工智能诊断软件（以下简称“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

本指导原则适用于超声影像人工智能诊断软件的技术审评。该产品通常作为软件组件集成在影像型超声诊断设备中。若作为独立软件运行通用计算机平台，可以参考本指导原则的要求。

该类产品的核心功能是采用深度学习等人工智能算法对超声影像数据进行处理、测量、分析等，包括非辅助决策类功能和辅助决策类功能，前者如生理结构自动识别及生理参数自动测量等，后者如病灶特征识别、病灶性质判定等。

## 二、技术审评要点

### （一）监管信息

#### 1. 软件组件

超声影像人工智能诊断软件集成于超声诊断设备中时，产品名称、分类编码、型号规格、结构组成与所集成产品保持一致。产品技术要求中应体现基于人工智能算法的软件功能的相关信息。若有辅助决策类软件功能，结构组成（若适用）和适用范围应予以体现。

#### 2. 独立软件

产品名称应符合通用名称命名规范要求。通常体现超声检查项目（如产前超声、乳腺超声）、预期用途（如影像处理、乳腺结节辅助检测）等特征词。例如：产前超声影像处理软件、乳腺结节超声影像辅助检测软件等。

依据《医疗器械分类目录》，分类编码为21-04-02。

型号规格应给出产品的型号和软件发布版本。

结构组成应明确软件交付方式（如光盘/U 盘安装、网络下载），结合网络部署情况给出软件功能模块，应体现出应用人工智能算法的所有模块。

适用范围需明确处理对象、核心功能、适用人群、目标用户、临床用途。例如，可供适龄女性乳腺肿瘤筛查及体检人群使用，供经培训合格的超声科、妇科医生或技师使用。

### （二）综述资料

#### 1. 产品描述

基于人工智能算法的功能，应重点描述以下内容：

51 详述工作原理，明确具体的功能模块（如胎儿  
52 切面自动识别功能）、使用的超声硬件（包括超声  
53 诊断设备及探头）要求、以及成像模式（如B模式）。

54 明确临床工作流程，工作流程中使用申报产品  
55 的时间（如在超声医师检查过程中使用，或在检查  
56 结束后使用）。

57 基于参考文献及已上市同类产品说明该功能当  
58 前的临床实践，包括传统算法及人工智能算法的应  
59 用。

60 详细介绍产品技术要求及说明书中对应的临床  
61 功能、量化分析（如自动测量）的具体内容、操作  
62 设置、运行界面和报告界面的细节。明确算法自动  
63 生成的标注、输出结果以及哪些可以修改。说明功  
64 能模块失效时对超声诊断的影响等信息。

## 65 2. 适用范围和禁忌证

66 基于人工智能算法的功能，应重点描述以下内容：

67 需详述产品的适用人群、感兴趣区域、以及预期使用该  
68 产品的目标用户。

69 需明确产品临床应用的禁忌证以及器械使用限制，包括  
70 导致器械无效且不应使用的疾病、病症、异常。

### 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 1.2 性能指标

79 基于人工智能算法实现的功能模块，性能指标应注意以  
80 下内容：

81 软件功能明确所有临床功能纲要。例如，支持神经自动  
82 识别功能（需明确必要的信息，如支持的探头、可自动识  
83 别的神经、识别后的表现如目标增强等）；乳腺结节检出、  
84 BI-RADS 分类结果、乳腺结节测量功能（如体积、面积、前  
85 后径、上下径、横径、相位、距离等）；胎儿标准切面自  
86 动识别（需明确具体切面）等。

87 使用限制明确应用场景（如筛查、辅助诊断等）、患者  
88 情况（如年龄、生理结构或疾病类型、可检查的器官/疾病  
89 /病灶/异常限制）、超声诊断及探头、成像模式、超声  
90 影像质量要求（如乳腺扫查时的灰度对比、深度及穿透力  
91 适当的要求等）。

92 接口明确输入数据模式/类型（如灰度图、三维容积图）、  
93 特定软硬件名称和型号（如适用）。

94 若含有基于测评数据库测试的性能指标，可参考《人工  
95 智能医疗器械注册审查指导原则》。

## 96 1.3 附录

97 附录中明确测评数据库信息。

## 98 2. 软件研究

### 99 （1）基本要求

100 软件研究相关资料包括软件研究资料、网络安全研究资

101 料、算法研究资料三部分。

102 依据已发布的《医疗器械软件注册审查指导原则》、《医  
103 疗器械网络安全注册审查指导原则》、《移动医疗器械注册  
104 技术审查指导原则》、《人工智能医疗器械注册审查指导原  
105 则》等指导原则提交相关资料。网络安全方面还可参考本课  
106 题《人工智能医学软件产品网络安全技术审评指导原则》的  
107 有关内容。

108 测量功能提供测量准确性的研究资料。数据资源（如参  
109 考数据库）明确数据种类以及每类数据的样本量、数据分布  
110 等情况。

## 111 （2）软件安全级别

112 该类产品的软件安全性级别为中等（B）级及以上，其  
113 中辅助决策类软件为严重（C）级。

## 114 3. 算法研究资料

115 应基于《人工智能医疗器械审查指导原则》提交算法研  
116 究资料。根据超声影像人工智能产品的特点需要细化之处如  
117 下。

### 118 3.1 算法基本信息

119 （1）输入应明确检查的器官/疾病/病灶/异常及应用（如  
120 心脏结构自动识别）、扫查条件、成像模式、临床使用限制、  
121 超声诊断设备及探头的型号、软件运行环境，输出应明确算  
122 法的输出和自动生成的诊断结论；

123 （2）如具有通过检测操作者扫查手法来引导标准切面  
124 的获取算法，应明确扫查手法标准、标准切面判定标准。

### 125 3.2 算法需求规范

126 产品设计应充分考虑国内和国际公认的相关领域的超  
127 声诊疗指南，跟踪最新版本，如 2020 甲状腺结节超声恶性  
128 分层中国指南：C-TIRADS、美国放射学会推荐的“乳腺超  
129 声影像报告和数据系统（BI-RADS）”等。

130 明确产品的使用场景、操作者、适用人群，如允许操作  
131 者为非超声医师（如技术员、护士），应重点描述。

132 明确超声图像质量控制的方式。

133 明确算法训练数据集要求，明确预期采集各分型/分期/  
134 分级样本、阴性样本、各特征样本的数量和分布要求，明确  
135 采集的超声诊断及探头信息、扫查部位、成像模式、图  
136 像分辨率/探测深度等、以及人群分布（性别、年龄、地域、  
137 基础疾病等）、医院级别、医院数量等要求。根据产品的  
138 预期用途和使用场景尽可能多采集数据，例如来源于多家、  
139 多地域、多层级的代表性临床机构，以及多家、多种、多参  
140 数的代表性采集设备（如适用）。

141 用于疾病/病灶/异常检测的人工智能算法，应结合流行  
142 病学特征纳入一定量阳性样本。用于正常生理结构识别的人  
143 工智能算法，应明确保证样本多样性的方法以及异常样本的  
144 排除情况。

145 受超声系统硬件（探头、数模芯片、处理器性能等）、  
146 成像参数（信号采集滤波、波束合成、重建方法、增益补偿、  
147 图像调试等）设置、系统平台参数等因素的影响，不同厂家  
148 的超声系统平台的图像在风格、颗粒度等方面往往存在较大  
149 差异。因此，难以做到一个模型兼容所有厂家超声系统平台  
150 的数据进行部署。如果采用多种超声系统平台的数据，需明

151 确数据分布，包括样本的组成、比例、分群分布特征、数据  
152 的多样性和接近应用场景的程度开展分析，并以书面形式提  
153 供可验证的指标，以确定数据的代表性。

154 明确算法性能评价目标。如标准切面识别指标的准确  
155 性、实时性等，测量指标的准确性、重复性等，病灶检出指  
156 标的召回率、精确度等。需提供文献综述论证评价指标的定  
157 义、计算公式及确定依据。应能够满足预期临床使用需求。

### 158 3.3 数据质控

159 建议参考《人工智能医疗器械注册审查指导原则》及  
160 YY/T 1833.2《人工智能医疗器械 质量要求和评价 第 2 部  
161 分：数据集通用要求》。同时注意以下方面的内容。

#### 162 （1）数据采集

163 详述采集过程的采集设备（含超声诊断设备及探头）、采  
164 集特征、采集人员管理、采集流程、采集质量评估要求、数  
165 据脱敏要求。其中采集特征应当考虑采集设备的采集方式  
166 （如静态/动态影像、成像模式）、采集参数（如超声探头  
167 频率、探头阵元数、侧向和轴向分辨率、深度、图像放大倍  
168 率等要求）。采集人员管理详述采集人员数量、资质、培训  
169 方式和考核方式等。采集流程详述入排标准、扫查手法、留  
170 图要求等。若使用历史数据，至少应明确采集设备、采集参  
171 数、图像质量的要求等。

#### 172 （2）数据整理

173 采集的图像数据不应包含患者隐私信息，需说明数据脱  
174 敏的技术手段。

175 数据预处理需明确数据转移保存的方法。明确预处理的

176 操作步骤和内容，如数据格式转化、尺寸裁剪、归一化、去  
177 除无用信息等。原始数据库的预处理过程若与算法运行的预  
178 处理过程存在差异，应加以说明并纳入风险分析。

179 明确数据清洗的方法（如：人工、自动或两者结合）、  
180 工具（如有）、规则等，例如是否已完成数据脱敏、是否满  
181 足图像唯一性、图像标识或图像内容是否与检查部位相关、  
182 常规切面是否遗漏、动态图像是否连续、探头工作频率和探  
183 头深度是否满足要求、图像质量是否满足要求等并加以筛  
184 选。清洗结果明确弃用的数据量和原因。

### 185 （3）数据标注

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

188 建议列表给出标注、审核、仲裁人员的基本信息，如科  
189 室/专业、职称、工作年限、所在机构、培训考核情况、工  
190 作量。明确标注对象（如超声静态或动态图像中某解剖结构  
191 区域、某病灶及其位置等）、标注形式（根据任务的不同可  
192 有不同的标注形式，如切面类型标注、病灶位置标注、病灶  
193 性质标注等）、标注工具及平台、标注环境、标注规则等。

194 标注工具及平台应支持超声数据格式或其转化后数据  
195 格式的显示功能，如 Dicom 格式数据、视频格式数据、图片  
196 格式数据、文本格式数据等。标注工具等使用人工智能算法  
197 进行辅助标注的，需提交算法确认资料。

198 标注规则需明确参考依据，如指南、专家共识等。

199 明确标注过程的分歧处理方法、控制偏倚方法、标注质  
200 量评估方法、以及追溯方法等，辅以流程图说明标注流程。

#### 201 (4) 数据集构建

202 对于标注前的基础数据库，标注后划分的训练集、调优  
203 集、测试集，应给出样本量和分布情况及其确定依据，以及  
204 集合划分的方法、依据。为确保测试结果的合理性，在划分  
205 数据集时，还应确保训练集、调优集、测试集所采样的病例  
206 两两无交集。注意使各数据正负样本、各特征类别样本分布  
207 均衡。样本分布应主要考虑超声诊断设备及探头种类/型号、  
208 成像模式、扫查参数、扫查部位/切面、疾病/病灶/异常分级  
209 /分期/分型等，以及人群分布（性别、年龄、地域、基础疾  
210 病等）、来源医院等因素。

211 数据扩增应符合《人工智能医疗器械注册审查指导原  
212 则》的要求。

### 213 3.4 算法验证与确认

#### 214 (1) 算法性能评估

215 应根据功能模块的具体任务制定算法性能评估指标、目  
216 标值，提供制定依据并论述其合理性。例如，图像分割任务  
217 可采用 DICE 系数、mIOU 等指标来评价；图像分类任务可  
218 采用准确率、灵敏度、特异性、ROC 曲线等指标来评价；目  
219 标检测任务可以采用 mAP 等进行评价。以乳腺结节 BI-RADS  
220 特征分类为例，通常临床上认为当 BI-RADS 分类  $< 3$  时，可  
221 不进行临床干预，当 BI-RADS 分类  $\geq 3$  则需要进行临床干预。  
222 那么，除使用灵敏度、特异性等对算法在各特征上的整体性  
223 能进行评价外，从临床实际意义出发，分层级对各特征性能  
224 进行统计分析就显得尤为必要。同时，还应考虑算法的重复  
225 性和鲁棒性。

226 明确测试样本量估计的公式、参数及制定依据。基于测  
227 试集提供算法性能评估结果，以证明算法性能满足设计目  
228 标。

229 若有扫查手法检测模块，应选择合适的评价指标验证其  
230 算法性能。

### 231 （2）算法性能影响因素分析

232 应进行算法性能影响因素分析。详述影响算法性能的主  
233 要因素及其影响程度，基于分析结果明确产品使用限制和必  
234 要警示提示信息。

235 主要性能影响因素包括但不限于超声诊断设备及探头、  
236 成像模式、超声影像质量要求（如乳腺扫查时的灰度对比、  
237 深度及穿透力适当的要求等）、患者情况（如年龄、生理  
238 结构或疾病类型、可检查的器官/疾病/病灶/异常限制）等，  
239 应论述其合理性。

### 240 （3）算法确认

241 人工智能算法功能模块可在多个不同平台/超声诊断系  
242 统（含探头）运行的，应论述算法确认所用平台/超声诊断系  
243 统（含探头）的代表性。明确相关安全性、有效性、实时性  
244 评价指标，如标准切面识别的准确性、图像增强效果的有效  
245 性、实时性等。明确样本量及估算依据。

246 目前尚无超声相关测评数据库，原则上应基于临床评价  
247 方式进行算法确认。

## 248 4. 用户培训方案

249 用户培训需要重点说明的内容：

250 对于软件安全性级别为严重级别或在基层医疗机构使

251 用/非超声医生使用的产品，原则上需单独提供一份用户培训  
252 方案，包括用户培训的计划、材料、方式、师资等。

253 用户培训需关注以下内容：预期用户要求，如工作年限  
254 或执业资格；医生必须对软件结果进行确认，软件只用于辅  
255 助检测，不能替代医生；明确图像质量要求（如有扫描质量  
256 问题的超声图像慎用）；禁忌情况，如基于临床试验验证，  
257 不宜使用该软件的疾病。

#### 258 （四）临床评价

259 临床评价应基于核心功能或核心算法，结合产品的预期  
260 用途和成熟度予以综合考虑。

261 针对成像和后处理的非辅助决策类软件功能，如心脏标  
262 准切面识别，原则上可基于核心功能进行同品种医疗器械比  
263 对，并开展一定数量的人体图像样本研究（区别于训练集、  
264 调优集、测试集的样本数据）。但全新的核心算法、核心功  
265 能、预期用途原则上需基于临床试验数据开展临床评价。

266 辅助决策类软件功能，如乳腺结节良恶性判断，应基于  
267 核心算法进行同品种医疗器械比对，所选同品种医疗器械的  
268 临床证据原则上需基于临床试验。全新的核心算法、核心功  
269 能、预期用途原则上均需开展临床试验。临床试验一般选择  
270 同品种产品或临床金标准进行非劣效对照设计，非劣效界值  
271 的确定应当有充分的临床依据。原则上选择灵敏度、特异性、  
272 **ROC/AUC** 作为主要评价指标，亦可在此基础上根据软件特  
273 点选择敏感性/特异性衍生指标、**ROC/AUC** 衍生指标、组内  
274 相关系数、**Kappa** 系数、时间效率等指标作为次要评价指标。  
275 临床试验的机构应避免训练数据的主要来源机构。

## 276 （五）产品说明书

277 说明书应符合《医疗器械说明书和标签管理规定》和《医  
278 疗器械软件注册技术审查指导原则》、《医疗器械网络安全  
279 技术审查指导原则》、《人工智能医疗器械注册审查指导原  
280 则》和相关标准的规定。

281 说明书内容需重点提醒的内容：

### 282 1.用户说明

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

### 285 2.使用限制

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

289 示例：不应仅仅依靠本软件所标识的输出，应由专业医  
290 师对结果进行解释。

291 已发现该器械对于XX的受检者无效。具有这种疾病/病  
292 症/异常的受检者不应使用该器械。

293 对训练数据、测试数据与临床试验的算法性能评估结果  
294 不佳，数据量偏少的，此类受检者使用该器械，应由专业医  
295 师结合受检者的病史、症状、体征、其他检查结果情况综合  
296 给出最终的检查结论，核实是否需要进行下一步诊疗的决  
297 策，并对临床诊断结果负责。

### 298 3.注意事项

299 明确图像质量要求（如有扫描质量问题的超声图像慎  
300 用）；

301 医生必须对软件结果进行确认，软件只用于辅助检测/  
302 诊断，不能替代医生。原始的软件结果应保留，确保软件结  
303 果的可追溯性与可责性。

#### 304 4. 预防措施

305 需明确与器械使用相关的不良事件，并提供缓解措施建  
306 议。对辅助决策类，不良事件讨论需至少包括对假阳性事件  
307 和假阴性事件的不良事件的讨论。

#### 308 5. 器械描述

309 需提供以下内容：

310 -算法设计和功能的概述，如有特殊声明，如用于继发性  
311 结节或更小结节检测，可以特别说明。

312 -培训范例和培训或开发数据库的概述

313 -研发和调整算法中所用的受检者数据的参考标准的描  
314 述

315 -与本器械兼容的采集技术

316 -适当显示器械标记的要求

#### 317 6. 软件

318 需明确软件发布版本、运行环境（含硬件配置、外部软  
319 件环境、网络环境，若适用）、安全软件兼容性列表（若适  
320 用）、外部软件环境与安全软件更新（若适用）、现成软件  
321 清单（SBOM，若适用）等要求。

#### 322 7. 产品接口和联合使用设备

323 需明确对配合使用的图像工作站（如PACS）适当显示  
324 器械标记的要求（如适用）。

325 若该软件部署于超声设备外接计算介质内，需明确与其

326 兼容的超声设备厂家、设备型号、探头型号等。

327 8. 辅助决策类产品的算法训练总结

328 训练集基本信息、训练指标与结果

329 9. 辅助决策类产品的算法性能评估总结

330 -算法输入与输出

331 -测试集基本信息

332 -用于确定器械标记的每个区域的性质的评分标准

333 -每个可用器械操作点的总体敏感度和假阳性率指标

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

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

338 -测试结果

339 10. 临床试验总结（如有）

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

343 11. 公开数据库及测试结果（如有）

344 12. 第三方测评数据库及测试结果（如有）

345 13. 决策指标定义（或提供决策指标定义所依据的临  
346 床指南、专家共识等参考文献）等信息。（如有）

347 （六）质量管理体系

348 除了符合质量管理体系法规及相关规范要求外，还可关  
349 注本课题《人工智能医学软件产品现场检查指导原则》的相  
350 关要求。

351           **三、编写单位**

352           广东省药品监督管理局审评认证中心、深圳开立生物医  
353 疗科技股份有限公司。

354           **四、参考文献**

355           [1]医疗器械注册与备案管理办法[Z].

356           [2]医疗器械说明书和标签管理规定[Z].

357           [3]医疗器械注册申报资料要求和批准证明文件格式[Z].

358           [4]医疗器械通用名称命名指导原则[Z].

359           [5]医疗器械安全和性能基本原则[Z].

360           [6] 医疗器械产品技术要求编写指导原则[Z].

361           [7]医疗器械软件注册审查指导原则（2022 年修订版）

362 [Z].

363           [8]医疗器械网络安全注册审查指导原则（2022 年修订

364 版） [Z].

365           [9]人工智能医疗器械注册审查指导原则[Z].

366           [10]人工智能类医用软件产品分类界定指导原则[Z].

367           [11]YY/T 0316,医疗器械 风险管理对医疗器械的应用

368 [S].

369           [12] YY/T 1833.1,人工智能医疗器械 质量要求和评价

370 第 1 部分：术语[S].

371           [13] YY/T 1833.2,人工智能医疗器械 质量要求和评价

372 第 2 部分：数据集通用要求[S].

- 373 [14] YY/T 1833.3,人工智能医疗器械 质量要求和评价  
374 第 3 部分：数据标注通用要求[S].
- 375 [15] T/AII 004,面向人工智能的乳腺超声数据采集与标  
376 注技术规程[S].

“人工智能医学信息系统软件测试审评指导体系构建”课题