

# “人工智能医学信息系统软件 审评指导体系构建” 课题

## ——CT 影像人工智能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CT影像人工智能软件的一般要求。

本指导原则是供CT影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CT影像人工智能软件的产品注册。该产品以CT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。按现行《医疗器械分类目录》和《人工智能医用软件产品分类界定指导原则》，该类软件的管理类别应结合产品的预期用途、算法成熟度等因素综合判定。

CT影像人工智能软件包括CT影像处理软件、CT影像辅助决策软件，前者主要用于非辅助决策，如进行数据的显示、

26 标注、处理和测量等提供临床参考信息，按照第二类医疗器  
27 械管理；后者主要用于辅助决策，如提供目标对象特征识别、  
28 病变性质判定、用药指导、治疗计划制定等临床诊疗建议，  
29 按照第三类医疗器械管理。

30 CT设备中采用人工智能技术的图像后处理功能也可参  
31 考本指导原则。

32

## 33 二、注册审查要点

### 34 （一）监管信息

35 明确产品名称的确定依据、管理类别、分类编码、规格  
36 型号、产品组成等信息。

#### 37 1. 产品名称

38 独立软件产品命名应符合《医疗器械通用名称命名规  
39 则》的要求，通常体现输入数据（CT 影像）、目标疾病（例  
40 如肺炎、骨折、肺结节等）、预期用途（如辅助分诊、辅助  
41 诊断、辅助测量、三维重建等）等特征词。例如“肺炎 CT  
42 影像辅助分诊与评估软件”、“肺结节 CT 影像辅助检测软  
43 件”。

#### 44 2. 分类编码

45 依据《医疗器械分类目录》和《人工智能医用软件产品  
46 分类界定指导原则》，产品如主要用于非辅助决策，则分类  
47 编码为21-02-02，按第二类医疗器械管理。如主要用于辅助  
48 决策，则分类编码为21-04-02，按第三类医疗器械管理。

#### 49 3. 注册单元划分

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

#### 52 4.适用范围

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

### 55 （二）综述资料

#### 56 1. 产品描述

##### 57 1.1器械及操作原理描述

###### 58 1.1.1工作原理

59 需详述产品的工作原理，例如：基于深度学习技术对医  
60 学影像进行图像重建，对感兴趣区域进行分割、三维重建、  
61 目标检测、图像配准等。

###### 62 1.1.2结构组成

63 结构组成明确交付内容和功能模块，其中交付内容包括  
64 软件安装程序、授权文件、外部软件环境安装程序等软件程  
65 序文件，功能模块包括客户端、服务器端（若适用）、云端  
66 （若适用），若适用注明选装、模块版本。基于人工智能算  
67 法（如深度学习）的功能模块，需在模块名称中注明人工智  
68 能（如深度学习）。

#### 69 1.2 型号规格

70 需明确申报产品的型号规格及发布版本。产品型号/规格  
71 及其划分，如同一个注册单元包含多个型号规格，需提供产  
72 品型号规格区分列表或配置表。

#### 73 1.3研发历程

74 阐述申请注册产品的研发背景和目的。如有参考的同类

75 产品或前代产品，需提供同类产品或前代产品的信息，并说  
76 明选择其作为研发参考的原因。

## 77 2.适用范围和禁忌证

### 78 （1）适用范围

79 适用范围需明确处理对象、核心功能、适用人群、目标  
80 用户、临床用途、软件与医生阅片顺序。

### 81 （2）预期使用环境

82 需明确设备使用场所和使用环境要求。

83 设备使用场所包括：医疗机构机房等。

84 使用环境要求需至少包括：推荐的硬件使用的适宜温  
85 度、湿度、大气压、光照条件。

### 86 （3）适用人群

87 需详述产品的适用人群。目标人群信息（如用于筛查、  
88 诊断）或无预期治疗特定人群的声明、感兴趣器官/疾病/目  
89 标对象/异常、以及预期使用该产品的目标用户（如经培训合  
90 格的放射科医师、技师、专科医师等）

### 91 （4）禁忌证

92 明确产品临床应用的禁忌证以及不推荐使用该产品的  
93 情况。

## 94 3.产品功能

95 明确软件与医生阅片顺序。结合用户界面图示详细介绍  
96 技术要求中对应的临床功能，如有特殊声明，可以在功能中  
97 体现。明确自动、手动、半自动输出的数据、量化分析的具  
98 体内容、服务器部署（局域网、内外网互联、云端）等。

### 99 （三）非临床资料

100 1.产品风险管理资料

101 依据YY/T 0316-2016《医疗器械风险管理对医疗器械的  
102 应用》，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  
109 持。

110 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  
115 定。

116 （5）对申报产品的可能危害作出判定（见附录1）。

117 （6）对所判定的危害采取的降低风险的控制措施。

118 （7）对采取控制措施后的剩余风险进行估计和评价。

119 2.产品技术要求及检测报告

120 2.1产品技术要求

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

124 2.1.1规格信息

125 明确软件发布版本和版本命名规则。软件版本命名规则  
126 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  
127 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  
128 见典型情况。

129 明确不同型号间产品差异。

### 130 2.1.2 性能指标

131 明确产品处理对象以及数据接口信息，包括成像模态  
132 （如计算机体层摄影）和特定软硬件名称及型号（如适用）。

133 明确软件所有的临床功能，如目标对象检出的尺寸范  
134 围、目标对象分类类别、目标对象位置、目标对象标记方式  
135 （紧密包裹目标对象的边界框、不紧密包裹目标对象的边界  
136 框）、分割方式（自动分割、半自动分割、手动分割）、测  
137 量功能（如体积、最大轴向平面最长直径、短径、平均直径、  
138 最大三维直径、有效直径、平均密度、面积等）、随访评估  
139 功能（如倍增时间、数值参数的百分比(%)和绝对变化、目  
140 标对象长轴、短轴、平均直径、最大三维直径、有效直径、  
141 体积、平均密度等）、三维重建功能、分割精度（如 Dice 系  
142 数、Hausdorff 距离等）、测量精度（如相对误差绝对值的平  
143 均值、绝对误差等）、目标对象的 3D 或 MIP 可视化等。

144 使用限制明确应用场景（如门诊、手术、住院、急救、  
145 家庭、转运、公共场所等）、患者（年龄、地域、疾病类型）、  
146 CT设备厂家、探测器排数、CT扫描参数(如管电压、管电流、  
147 层厚/层间距、剂量、窗宽窗位、重建方式、显示方式)、放  
148 射学检查条件(CT增强扫描、CT平扫)、CT图像质量要求(如  
149 分辨率、伪影)、目标对象检出的尺寸范围、分类类别。如软

150 件包含图像质量判定功能，明确软件使用限制。

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》（2022 年第 8 号）。

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  
157 示、物理拓扑图、测评数据库（如有）、数据库/集信息。

### 158 3.软件研究

#### 159 （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）》（2022 年第 9 号）提交相应软件研究  
162 资料。其中，核心算法所述人工智能算法需依据《人工智  
163 能医疗器械注册审查指导原则》（2022 年第 8 号）提交相  
164 应算法研究资料。

165 生产企业需依据《医疗器械网络安全注册审查指导原则  
166 （2022 年修订版）》（2022 年第 7 号）提交网络安全描述  
167 文档。按照《人工智能医疗器械注册审查指导原则》（2022  
168 年第 8 号）提交数据安全资料。

169 若使用云计算服务，生产企业需依据《人工智能医疗器  
170 械注册审查指导原则》（2022 年第 8 号）提交相应研究资料。  
171 使用云计算服务需明确服务模式、部署模式、核心功能、数  
172 据接口、网络安全能力、服务（质量）协议等要求。

173 软件版本命名规则涵盖算法驱动型软件更新和数据驱  
174 动型软件更新；区分重大软件更新和轻微软件更新，其中重

175 大软件更新需列举常见典型情况。

## 176 （2）软件安全级别

177 软件风险程度采用软件安全性级别进行表述，软件安全  
178 性级别越高，生存周期质控要求越严格，注册申报资料也越  
179 详尽。软件安全性级别基于软件风险程度分为轻微、中等、  
180 严重三个级别（轻微级别、中等级别、严重级别分别与 YY/T  
181 0664 所定义的 A 级、B 级、C 级相对应），其中轻微级别即  
182 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  
183 轻微（不严重）伤害，严重级别即软件可能直接或间接产生  
184 严重伤害或导致死亡。

185 软件安全性级别可结合软件的预期用途、使用场景、核  
186 心功能进行综合判定（详见 IMDRF/SaMD WG/N12 FINAL:  
187 2014）。其中，预期用途主要考虑软件的用途类型（如治疗、  
188 诊断、监护、筛查）、重要程度（如重要作用、参考作用、  
189 补充作用）、紧迫程度（如危重情形、严重情形、普通情形）、  
190 成熟程度（成熟、全新）等因素，使用场景主要考虑软件的  
191 使用场所（如门诊、手术、住院、急救、家庭、转运、公共  
192 场所）、疾病特征（如严重性、紧迫性、传染性）、适用人  
193 群（如成人、儿童、老人、孕妇）、目标用户（如医务人员、  
194 患者）等因素，核心功能主要考虑软件的功能类型（如重要  
195 程度、技术特征、复杂程度、成熟程度）、核心算法（如重  
196 要程度、复杂程度、可解释性、成熟程度）、输入输出（输  
197 入数据如医学图像、生理参数、体外诊断等数据，输出结果  
198 如处理、测量、分析等结果）、接口（如应用程序接口（API）、  
199 数据接口、产品接口）等因素。

200 软件安全性级别也可根据风险管理所确定的风险等级  
 201 进行判定，软件安全性级别与风险等级的分级可以不同，但  
 202 二者存在对应关系，因此可根据风险等级来判定软件安全性  
 203 级别，但应在采取风险控制措施之前进行判定，后续可通过  
 204 外部风险控制措施（含软件措施、硬件措施）降低初始软件  
 205 安全性级别。。

### 206 (3) 核心算法

207 需指出核心算法的选择，明确算法名称、类型、用途  
 208 以及功能，可以参考核心算法示例如下表，其中采用全新  
 209 算法（如深度学习）需提供算法研究资料。

210 表1. 核心算法示例

算法名称	类型	用途	功能
基于深度学习的 检出算法(例如: Faster RCNN, YOLO 等)	全新	对目标对象进行检出	辅助医生检测目标对象
基于深度学习的 分割算法(例如: U-NET, V-NET 等)	全新	对目标对象进行分割, 提 取感兴趣区域	检出目标对象的感兴 趣区域
基于深度学习的 分类算法(例如: RESNET 等)	全新	对目标对象的类型进行 分类	辅助医生判断目标对 象的类型
基于深度学习的 测量算法(例如: VGG 等)	全新	对目标对象进行测量	辅助医生对目标对象 进行测量
灰度拉伸图像增 强算法	公认成熟 算法	影像增强, 用于增强影像 对比度, 提升算法精度	目标对象分割及检测
NMS 非极大值抑 制	公认成熟 算法	目标选择	以预测概率对检测框 进行优先级排序, 消 除冗余检测框, 抑制 假阳性检出
... ..			

## 211 4. 算法研究资料

## 212 4.1 算法基本信息

213 CT 影像人工智能软件所涉及的算法一般包括检出算法、  
214 分割算法、分类算法和测量算法等。检出算法用于自动检出  
215 CT 影像中的目标对象，辅助医生筛查；分割算法用于自动  
216 提取 CT 影像中的感兴趣区域，辅助医生获取感兴趣影像信  
217 息；分类算法用于对 CT 影像中的目标对象进行自动分类，  
218 按照类别不同，可根据实际分类情况设计不同分类算法；测  
219 量算法用于对 CT 影像中的目标对象进行量化测量，区别于  
220 分类算法，测量算法输出的是连续的量化值，而分类算法输  
221 出的是离散的类别。按照测量属性不同，目标对象测量算法  
222 包括但不限于体积测量算法，长短径测量算法，病变严重程度  
223 测量等。

### 224 4.1.1 目标对象检出算法

225 检出算法需明确算法的输入，比如算法采用的是 2D，  
226 2.5D 还是 3D 的 CT 影像作为神经网络的输入，并明确算法的  
227 输出。

228 明确算法所采用的神经网络结构，如 Faster RCNN；明  
229 确算法训练采用的损失函数，如交叉熵，L1 范数；明确算法  
230 设计过程中网络结构、损失函数等核心组件选择和设计的原  
231 则、方法与风险考量、目标对象背景正负样本的不均衡性、  
232 假阳性、过拟合等；若检出采用多个模型融合，需明确不同  
233 模型训练与推理的方式以及模型融合的策略，如级联，多数  
234 投票。

235 明确算法的流程图，需包含算法运行前所进行的预处理  
236 （图像缩放、图像像素值归一化、图像重采样、降噪等）以

237 及算法运行后所进行的后处理操作（图像重采样、连通域处  
238 理等）。明确训练与部署所采用的框架（如 Tensorflow,  
239 Pytorch）、算法运行环境（如内存、显存的需求）。

#### 240 4.1.2 感兴趣区域分割算法

241 分割算法对 CT 影像感兴趣区域进行提取，辅助医生获  
242 取感兴趣区域影像信息。

243 分割算法需明确算法的输入，比如算法采用的是 2D，  
244 2.5D 还是 3D 的图像作为神经网络的输入，并明确算法的输  
245 出。

246 明确算法所采用的神经网络结构，如 U-Net, V-Net 等  
247 网络；明确算法训练采用的损失函数，如交叉熵, Dice 损失；  
248 损失函数等核心组件选择和设计的原则、方法与风险考量、  
249 不同类别分割样本的不均衡性、过拟合等；最后，明确算法  
250 的流程图、训练与部署所采用的框架（如 Tensorflow,  
251 Pytorch）、算法运行环境（如内存、显存的需求）。

252 明确算法的流程图，需包含算法运行前所进行的预处理  
253 （图像缩放、图像像素值归一化、图像重采样、降噪等）以  
254 及算法运行后所进行的后处理操作（图像重采样、连通域处  
255 理等）。明确训练与部署所采用的框架（如 Tensorflow,  
256 Pytorch）、算法运行环境（如内存、显存的需求）。

#### 257 4.1.3 分类算法

258 分类算法需明确输出类别信息以及临床意义，类别制  
259 定依据以及临床适用性。

260 分类算法需明确算法的输入，比如算法采用的是 2D，  
261 2.5D 还是 3D 的图像作为神经网络的输入，并明确算法的输

262 出。

263 明确算法所采用的神经网络结构，如 ResNet，  
264 Dense-Net；明确算法训练采用的损失函数，如交叉熵，Focal  
265 损失函数；损失函数等核心组件选择和设计的原则、方法与  
266 风险考量、不同类别样本的不均衡性、过拟合等；若分类采  
267 用多个模型融合，需明确不同模型训练与推理的方式，以及  
268 模型融合的策略，如多数投票；最后，明确算法的流程图，  
269 需包含算法运行前所进行的预处理（图像缩放、图像像素值  
270 归一化、图像重采样、降噪等）以及算法运行后所进行的后  
271 处理操作（图像重采样、连通域处理等）。明确训练与部署  
272 所采用的框架（如 Tensorflow, Pytorch）、算法运行环境  
273 （如内存、显存的需求）。

#### 274 4.1.4 目标对象测量算法

275 针对测量算法，明确测量算法实现的方式。如体积测量  
276 算法、长短径测量算法、病变严重程度测量算法等。

277 可以通过分割技术来实现，也可以通过机器学习中的回  
278 归方法来进行预测。

279 若申报产品涉及上述多个算法，需提供算法整体流程  
280 图，明确各个算法调用先后关系，以及输入输出依赖关系。

281 若使用迁移学习技术，除上述内容外还需补充预训练模  
282 型的数据集构建、验证与确认等总结信息，并论证迁移学习  
283 的适用性。

#### 284 4.2 算法需求规范

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

#### 287 4.2.1 数据库需求

288 算法性能评价需要基于训练和测试数据库，数据库具  
289 体要求详见附录2。

#### 290 4.2.2 算法性能评价指标

291 在指定检出算法评估指标之前，需明确标记匹配的方式  
292 式，即算法标记目标与参考标准目标的匹配方式，常见的标  
293 记匹配方式有按照区域重叠的比例、中心点的距离、中心是  
294 否落入来判断标记是否匹配；明确算法任务，明确算法训练  
295 和调优过程中不同任务的评估指标及定义。

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

299 检出指标包括不限于召回率、精确度、特异度。

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

302 分割指标包括不限于体积交并比、Dice 系数、Jaccard 系  
303 数、Hausdorff 距离。

304 尺寸测量指标包括不限于体积相对误差率、平均径相  
305 对误差率、长径绝对误差均值 MAE、短径绝对误差均值  
306 MAE。

307 提供文献综述论证评价指标选择及分层影响因素选择的  
308 合理性。分层分析的影响因素包括年龄、病变类型、大  
309 小、层厚、采集协议等重要变量。

310 算法质量特性包括泛化能力、鲁棒性（面向硬件变化的  
311 对抗测试、面向软件预处理的对抗测试、压力测试）、重复

312 性、一致性、效率。

### 313 4.2.3 算法性能测试基本要求

#### 314 4.2.3.1 目标对象检出

315 软件检出目标对象的召回率和精确度不低于 X% 等。

316 召回率和精确度的计算一般针对全体目标对象进行。在  
317 设置筛选条件后，可以使用筛选后的参考标准与筛选后的自  
318 动检出结果进行匹配。

#### 319 4.2.3.2 分类准确度

320 明确分类场景，如二分类场景、多分类场景等。对于二  
321 分类场景下的指标，如软件能够对 A 类型进行分类，准确率  
322 不低于 X%，灵敏度不低于 Y%，特异性不低于 Z%。总体的 Kappa  
323 系数不低于 N%；多分类问题可以转化为二分类问题，按每一  
324 类进行描述。

#### 325 4.2.3.3 分割

326 体积交并比不小于 X%，或 DICE 系数不小于 Y% 等。

#### 327 4.2.3.4 测量

328 体积测量偏差  $\pm X\%$ ；长径测量偏差  $\pm Y\%$ ；短径测量偏差  
329  $\pm Z\%$  等。

#### 330 4.2.4 样本量

331 明确样本量估计的公式、参数及制定依据。研究样本应  
332 足够大，以使声称的性能具有统计学显著性，若有分层性能  
333 的声称，应确定子集的样本量以检测声称的统计学显著性。

### 334 4.3 数据质控

335 建议参考《人工智能医疗器械注册审查指导原则》（2022  
336 年第 8 号）及 YY/T 1833.2-2022《人工智能医疗器械质量要

337 求和评价 第 2 部分：数据集通用要求》。

#### 338 4.3.1 数据采集

339 需提供数据采集来源合规性证明文件（如数据授权使用  
340 证明性文件、伦理批件、采集协议、科研合作协议均可），  
341 需要考虑明确偏倚控制的方法，如通过各亚组设置进行偏移  
342 控制，不应为提高算法结果刻意筛选数据。采集数据的地区  
343 分布建议覆盖具有临床特征的代表性区域。

344 明确数据库信息（要求详见附件 2）

#### 345 4.3.2 数据整理

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

349 提供数据清洗流程图，明确清洗规则，例如确定数据合  
350 规性、图像唯一性、层厚、层间距等信息满足要求、连续性、  
351 完整性、已完成脱敏等并加以筛选。清洗结果明确弃用数据  
352 的数量和原因。明确预处理的操作步骤和内容。

#### 353 4.3.3 数据标注

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

356 需明确标注任务分类（包括数据模态、执行主体、标注  
357 结果格式、标注结果性质、标注结果形式等维度），提供标  
358 注任务描述文档（标准规则、标注人员、标注工具、标注环  
359 境、数据）。其中标注规则需明确制定依据并提供参考文献。  
360 标注人员建议列表给出标注、审核、仲裁人员的基本信息，

361 如数量、医疗机构、科室、工作年限、职称、培训、培训考  
362 核情况、工作量、标注任务、参考的所有受检者临床信息(如  
363 病理检测结果)。

364 标注与质控流程建议提供业务架构、过程组织（任务生  
365 成、任务分配、任务实施、质量控制、验收准则及验收报告。  
366 其中业务架构宜采用流程图介绍单张图片的标注、审核、仲  
367 裁过程。

368 如标注工具、标注平台使用人工智能算法进行辅助标  
369 注，需提交标注工具、标注平台算法性能研究资料。

#### 370 4.3.4 数据集构建

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

374 提供查重验证结果，以证实训练集、调优集、测试集的  
375 样本两两无交集。

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

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

382 如果采用生成对抗网络进行数据扩增，应提供生成对抗  
383 网络的算法基本信息以及算法选用依据资料。

#### 384 4.4 算法训练

385 算法训练需明确训练过程所采用的优化器及其相关参数，

386 如 Adam, SGD; 在算法训练阶段, 需明确验证集的划分方式,  
387 如留出法, 交叉验证法; 需明确训练目标, 即判断何时停  
388 止训练, 如设定最大的训练 epoch 数目, 依据损失函数判断  
389 损失稳定且不继续下降, 根据训练 epoch 数量-评估指标曲  
390 线判断评估指标不继续提升等。当训练停止后, 明确训练模  
391 型最佳 epoch 的选择方法, 如在留出法中, 根据留出部分的  
392 验证集选取评价指标最优的 epoch; 在交叉验证法中, 计算  
393 多次随机划分验证集的评价指标平均值, 选择最优的 epoch。  
394 同时, 结合临床需求 (如灵敏度、精准度), 明确算法出厂  
395 阈值的选择与方法, 并论证训练所得模型是否满足产品既定  
396 目标。算法出厂阈值的选择需提供制定依据。

397 算法训练阶段需结合训练数据量-评估指标曲线验证算  
398 法训练数据量的充分性, 常用的方式有以最小训练数据量为  
399 基础, 逐步增加数据量。对于不同的训练数据量, 计算对应  
400 的评价指标。当评价指标为单一标量时 (如灵敏度), 训练  
401 数据量-评估指标曲线的 x 轴为训练数据的样本量 (如 CT 序  
402 列个数), y 轴为在使用特定训练量时, 算法在测试集的评  
403 价指标。结合训练数据量-评估指标曲线, 判断当训练数据  
404 量有限时, 评价指标是否随数据量的增加而增加, 并在数据  
405 量达到一定程度后, 评价指标趋于平稳。当评价指标为曲线  
406 时, 可考虑根据不同训练数据量, 绘制对应的评价指标曲线,  
407 判断曲线下面积 (AUC) 是否先随数据量的增加而增加, 最  
408 后当数据量达到一定程度后趋于平稳。以 fROC 曲线为例,  
409 在评估训练数据量的充分性时, 当训练数据量有限时, fROC  
410 曲线下面积需随数据量增加而增加, fROC 曲线逐步逼近坐标

411 左上角；当训练数据量达到一定程度后，fROC 曲线下面积慢慢  
412 趋于稳定。

413 若训练过程中采用了数据扩增的方式，需明确扩增方  
414 式，如离线数据扩增，在线数据扩增。同时，需明确数据扩  
415 增的方法以及相应的参数设置，如图像平移，旋转，缩放，  
416 弹性形变等。若训练过程未采用数据扩增，需论证未进行数  
417 据扩增的理由。

#### 418 4.5 算法验证与确认

419 明确算法任务，明确不同任务的算法评价指标的阈值及  
420 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  
421 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  
422 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  
423 体测试方法和结果计算方法。

##### 424 4.5.1 检出算法性能评估

425 基于算法的出厂阈值，明确算法在训练集，调优集和测  
426 试集的假阴性（召回率）与假阳性（精确度），通过比较三  
427 个数据集的性能来评估算法的泛化性。

428 若在训练过程中使用数据扩增，需在调优集上比较使用  
429 与不使用数据扩增对目标对象检出性能的影响。

430 对于目标对象检出算法，需分层统计算法对于目标对象  
431 的检出效能，可以通过 fROC 曲线，召回率以及精确度等指  
432 标来评价，同时结合临床需求论述结果的合理性。

433 除此之外，还需考虑性别、年龄、设备厂家、重建方式、  
434 层厚/层间距、管电流、管电压等参数对目标对象检出算法  
435 效能的影响。需明确设备厂家，且必须符合 DICOM 3.0 协议

436 标准数据；管电压、管电流、层厚与层间距参数需根据相应的  
437 的 CT 影像数据集构建及质量控制专家共识和标准来确定。

438 在亚组的分层统计中，目标对象的检出效能需在不同因  
439 素下均能取得较好性能。若在某些影响因素下，性能结果存  
440 在差异，需进行合理地论证，并在说明书给出使用限制。

441 需提供算法性能测试报告，至少包括软件环境、硬件环  
442 境、测试平台描述（如适用）、测试集描述、算法性能指  
443 标的符合性分析（性能指标的定义、测试通过准则、统计  
444 分析）、算法错误统计。需包括算法性能及算法质量特性、  
445 随访功能测试结果。

#### 446 4.6 算法性能综合分析

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

#### 451 5. 用户培训方案

452 对于软件安全性级别为严重级别的产品，原则上需单独  
453 提供一份用户培训方案，包括用户培训的计划、材料、方式、  
454 师资等。

455 用户培训需关注以下内容：预期用户要求，如工作年限  
456 或执业资格；医生必须对软件结果进行确认，软件只用于辅  
457 助检测，不能替代医生。CT 图像要求，如严重呼吸、金属  
458 伪影或有扫描质量问题的 CT 影像慎用；基于临床试验验证，  
459 不宜使用该软件的疾病。

#### 460 （四）临床评价

461 临床评价基本原则详见医疗器械软件指导原则，基于  
462 核心功能或核心算法，结合预期用途和成熟度予以综合考  
463 虑：非辅助决策类功能基于核心功能开展同品种医疗器械  
464 比对，全新的功能、算法和用途原则上均需开展临床评价；  
465 辅助决策类功能基于核心算法开展同品种医疗器械比对，  
466 所选同品种医疗器械的临床证据原则上需基于临床试验  
467 （含回顾性研究，下同），全新的功能、算法和用途原则  
468 上均需开展临床试验。

469 CT 影像人工智能软件常见异常识别、疾病类型分析、  
470 良恶性风险提示、病程阶段分析功能的临床试验要求具体如  
471 下：

##### 472 1. 试验目的

473 生产企业应根据产品适用范围确定临床试验目的。适用  
474 范围考虑因素包括但不限于：目标疾病（如肺炎、肺结节、  
475 冠脉血流分析、骨折等）、临床用途（如分诊提示、异常识  
476 别、疾病类型分析、病程阶段分析、提供辅助诊断结果、提  
477 供医学建议）、禁忌症（如急性疾病、孕妇、婴儿、哺乳期  
478 妇女等对 X 射线高度敏感人员、如再生障碍性贫血不宜接触  
479 X 射线的病症、合并其他病症）、预期适用人群（如具有目  
480 标疾病临床表现的患者）、预期使用场所（如门诊、手术、  
481 住院、急救、家庭、转运、公共场所等）、预期目标用户（如  
482 放射科医师等）、预期兼容的 CT 设备等。

483 临床试验目的主要对临床用途（工作效率提升、分诊提  
484 示、异常识别、疾病类型分析、病程阶段分析、提供辅助诊  
485 断结果、提供医学建议等）的诊断准确度进行确认，确认试  
486 验器械是否优于/等效于/非劣于已上市同类产品或传统检测  
487 诊断方法（如临床医师的阅片/综合诊断方法）或单组目标值。

## 488 2. 试验设计

489 对患者是否患有目标疾病，从而对患者的分诊转诊提供  
490 辅助决策建议的产品，该产品不给出具体病变情况，且无  
491 论辅助分诊结果为阴性、阳性，均需专业医师再一次对患者  
492 影像进行评阅，如肺炎辅助分诊、脑出血辅助分诊等各类目  
493 标疾病患者的计算机辅助分诊、转诊产品等，可以考虑采用  
494 单组目标值设计。

495 对目标疾病的病变病灶进行辅助检测的产品，如肺结节  
496 辅助检测产品、骨折 CT 影像辅助检测产品等，临床试验建  
497 议采用对照设计，试验组为医师与申报产品共同检测，对照  
498 组为传统检测诊断方法（如临床医师的阅片/综合诊断）。临  
499 床试验比较类型应能够体现产品受益风险的可接受性，建议  
500 优先考虑优效性设计；根据临床需求，必要时考虑等效/非劣  
501 效设计。如针对 4mm 以上肺结节 CT 影像辅助检测软件可考  
502 虑患者水平的特异度优效和病灶水平的敏感度非劣效。

503 临床机构应当具备代表性和广泛性，建议不同于训练数

504 据主要来源机构，地域分布尽可能广泛，机构数量尽可能多，  
505 以确认算法泛化能力。

### 506 3.受试对象

507 受试对象入排标准应当基于目标疾病临床特征，保证阳  
508 性病例和阴性病例选取的合理性和充分性。在此基础上，病  
509 例的选择还需要综合考虑以下因素：

510 3.1 临床试验纳入病例应独立于产品开发所用病例；

511 3.2 临床试验纳入病例需来源于不同地域医疗机构。

512 3.3 临床试验纳入阳性病例中目标疾病的类型、良恶性、  
513 病程阶段分布合理性。

### 514 4.评价指标

515 针对采用人工智能算法设计的临床功能原则上选择敏  
516 感性、特异性、受试者工作特性曲线（ROC）曲线下面积  
517 （AUC）、检出率等为主要评价指标，亦可在此基础上根据  
518 软件特点选择敏感性/特异性衍生指标、受试者工作特性曲线  
519 （ROC）曲线下面积（AUC）衍生指标如诊断准确率、精确  
520 率、召回率、F1 系数、组内相关系数、Kappa 系数、时间效  
521 率（如辅助诊断平均时间、日人均诊断量增益率和诊断平均  
522 时间增益率）、数据有效使用率等指标作为次要评价指标。  
523 对目标疾病的病变病灶进行辅助检测的产品，一般灵敏度考  
524 虑病灶病变水平，特异度考虑患者水平。辅助分诊结果的诊

525 断准确度指标（如敏感度、特异度等），通常为患者水平。

## 526 5.金标准构建

527 生产企业应详述金标准的选择、构建方法及理由。可供  
528 选择的金标准构建方法包括：一是以临床确定结果为金标  
529 准，即流行病学史、临床表现（含影像学特征）的综合诊断  
530 结果；二是通过专家阅片小组构建金标准。适用的临床诊疗  
531 规范有非影像学证据作为金标准，且相关证据获得符合伦理  
532 要求，优先选用方法一。

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

## 539 6.样本量

540 样本量估算需综合考虑临床试验设计、主要评价指标和  
541 统计学要求。生产企业需明确计算公式、相应参数及确定理  
542 由，以及所用的统计软件。

543 对于复合终点，样本量估算需考虑满足所有单项指标的  
544 假设检验的样本量需求。例如：采用单组目标值法，假设软  
545 件的敏感性为 95%，临床认可的目标值为 90%，则在双侧显

546 著性水平 0.05、把握度 80%，至少需 231 例阳性病例。假设  
547 软件的特异性为 85%，临床认可的目标值为 80%，则在双侧  
548 显著性水平 0.05、把握度 80%，至少需 466 例阴性病例。在  
549 此基础上考虑 5%-10%脱落率确定最终样本量。

## 550 7.其他

551 临床试验资料需明确以下信息：

552 7.1 病例基线情况统计学描述，包括年龄、性别、流行  
553 病学特征、病例数（阳性病例数、阴性病例数）、疾病分型、  
554 影像学分期等。

555 7.2 CT 影像采集情况统计学描述，包括数据来源机构、  
556 CT 设备厂商及型号规格、CT 设备扫描参数（如排数、层数、  
557 层厚、管电压、管电流、加载时间）等。

## 558 （五）产品说明书和标签样稿

559 说明书、标签和包装标识需符合《医疗器械说明书和标  
560 签管理规定》（国家食品药品监督管理总局令第6号）和《医  
561 疗器械软件注册审查指导原则》（2022年修订版）（2022年  
562 第9号）、《医疗器械网络安全注册审查指导原则》（2022  
563 年修订版）（2022年第7号）、《人工智能医疗器械注册审  
564 查指导原则》（2022年第8号）和相关标准的规定。

565 说明书内容需重点关注：

### 566 1.用户说明

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

## 569 2.使用限制

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

573 示例：不应仅仅依靠本器械所标识的输出，应由专业医  
574 师对结果进行解释。

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

577 对训练数据、测试数据与临床试验的算法性能评估结果  
578 不佳，数据量偏少的，此类受检者使用该器械，应由专业医  
579 师结合受检者的病史、症状、体征、其他检查结果情况综合  
580 给出最终结论，核实是否需要进一步诊疗的决策，并对临  
581 床诊断结果负责。

## 582 3.注意事项

583 CT影像人工智能软件的分析结果仅供医生参考。医生必  
584 须对软件结果进行确认，软件只用于辅助检测等，不能替代  
585 医生进行诊断。原始的分析结果应保留，确保软件结果的可  
586 追溯性与可责性。

587 CT影像质量要求，如严重呼吸、金属伪影或有扫描质量  
588 问题的CT影像慎用；图像序列未完整包含目标对象全部组织  
589 的图像数据禁用。

590 数据/图像的标注要求符合YY/T 1833.3-2022《人工智能  
591 医疗器械质量要求和评价 第3部分：数据标注通用要求》的  
592 要求。

## 593 4.预防措施

594 需明确与器械使用相关的不良事件，并提供缓解措施建  
595 议。不良事件讨论需至少包括对假阳性事件和假阴性事件的  
596 不良事件的讨论。

### 597 5.器械描述

598 需提供以下内容：

599 -算法设计和功能的概述，如有特殊声明，可以特别说明。

600 -培训范例和培训或开发数据库的概述

601 -研发和调整算法中所用的受检者数据的参考标准的描  
602 述

603 -与本器械兼容的采集技术

604 -适当显示器械标记的要求

605 -软件输出报告及界面数据图示

### 606 6.软件

607 需明确软件发布版本、提供网络安全说明和使用指导，  
608 明确用户访问控制机制、电子接口（含网络接口、电子数据  
609 交换接口）及其数据类型和技术特征、网络安全特征配置、  
610 数据备份与灾难恢复、运行环境（含硬件配置、外部软件环  
611 境、网络环境，若适用）、安全软件兼容性列表（若适用）、  
612 外部软件环境与安全软件更新（若适用）、现成软件清单  
613 （SBOM，若适用）等要求。

### 614 7.产品接口和联合使用设备

615 需明确对配合使用的图像工作站、CT设备和PACS适当  
616 显示器械标记的要求（如适用）。

617 需明确CT设备兼容（厂家、型号（如适用）、层数）与  
618 扫描参数要求（如层厚、层间距、重建方式、管电压、管电

619 流、剂量（常规剂量、低剂量））。

## 620 8.算法训练总结

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

## 623 9.算法性能评估总结

624 -算法输入与输出

625 -测试集基本信息

626 -确定算法性能评估标准

627 -确定算法性能目标指标

628 -测试结果，含亚组分析结果

## 629 10.临床试验总结

630 需包括临床试验设计基本类型、研究对象（受试者及阅  
631 片者情况）、评价指标，金标准、对收集临床信息方法的描  
632 述、统计方法描述、样本量，临床试验结果（如各结节尺寸、  
633 各密度类型结节情况等）。

634 11.公开数据库及测试结果（如有）

635 12.第三方测评数据库及测试结果（如有）

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

## 638 三、编写单位

639 广东省药品监督管理局审评认证中心、深圳睿心智能医  
640 疗科技有限公司、广州柏视医疗科技有限公司

## 641 四、参考文献

642 [1]医疗器械注册与备案管理办法 [Z].

643 [2]医疗器械说明书和标签管理规定 [Z].

- 644 [3] 医疗器械注册申报资料要求和批准证明文件格式  
645 [Z].
- 646 [4] 医疗器械通用名称命名指导原则 [Z].
- 647 [5] 医疗器械安全和性能基本原则 [Z].
- 648 [6] 医疗器械产品技术要求编写指导原则（2022 年修订  
649 版） [Z].
- 650 [7] 医疗器械软件注册审查指导原则（2022 年修订版）  
651 [Z].
- 652 [8] 医疗器械网络安全注册审查指导原则（2022 年修订  
653 版） [Z].
- 654 [9] 人工智能医疗器械注册审查指导原则 [Z].
- 655 [10] 人工智能类医用软件产品分类界定指导原则 [Z].
- 656 [11] YY/T 0316 医疗器械 风险管理对医疗器械的应用  
657 [S]
- 658 [12] YY/T 1833.1 人工智能医疗器械 质量要求和评价  
659 第 1 部分：术语 [S].
- 660 [13] YY/T 1833.2 人工智能医疗器械 质量要求和评价  
661 第 2 部分：数据集通用要求 [S].
- 662 [14] YY/T 1833.3 人工智能医疗器械 质量要求和评价  
663 第 3 部分：数据标注通用要求 [S].

664

665

- 666       附件： 1.风险管理文档  
667               2.数据库（集）  
668

669 附录1

670 风险管理文档

671 下表所列为常见可预见事件及事件序列 / 可能的损害  
672 示例，需关注：

危害类型	可预见事件及事件序列	可能的损害
信息危害	算法对目标对象检出存在错误判断	导致用户对目标对象的主观判断产生误导
	算法对目标对象类型存在分类错误	导致用户对目标对象的主观判断产生误导
	算法对目标对象分割给出过大或过小结果	导致用户对目标对象的主观判断产生误导
	算法对目标对象测量给出过高或过低结果	导致用户对目标对象的主观判断产生误导
	未按规定编制产品说明书，使用说明不明确	导致软件无法正常使用并对用户造成不便
	界面实用性差	导致误操作
	无法正常显示医学影像，或影像失真	导致软件无法正常使用并对用户造成不便
	菜单设置不实用	导致误操作
	影像信息丢失或损坏	延误治疗

	软件不兼容导致产品无法正常运行	对用户造成不便
	算法模型失效导致无法进行影像识别	对用户造成不便
	服务器和客户端之间无法正常传输数据	导致软件无法正常使用并对用户造成不便
操作危害	算法无法对接收到的 CT 影像进行处理分析	导致软件无法正常使用并对用户造成不便
	无法与相关设备或系统进行通讯连接	导致软件无法正常使用并对用户造成不便
	软件设计缺陷，可能软件死机或软件自动退出	导致软件无法正常使用并对用户造成不便
	算法被未经培训的用户、使用不熟练的用户、或非专业用户使用并完成主观诊断	导致过度诊断或延误诊断
	算法对接收到的 CT 影像进行处理分析的时间过长	导致软件无法正常使用并对用户造成不便

673

674

675

676

677

678

采取风险控制措施前后的风险矩阵表示例

679

680

表 采取风险控制措施前的风险分布图

严重度 发生概率	1 可忽略的	2 轻微的	3 严重的	4 危急的	5 灾难性的
5 频繁					
4 很可能					
3 偶尔			4		
2 极少	1	2			
1 不可能		1			

681

682

683

表 采取风险控制措施后的风险分布图

严重度 发生概率	1 可忽略的	2 轻微的	3 严重的	4 危急的	5 灾难性的
5 频繁					
4 很可能					
3 偶尔					
2 极少					
1 不可能	4	4			

684

685

686

687

688

689 附录2

690 **数据库(集)**

691 按照产品开发时间顺序明确使用的数据库/集情况，需  
692 明确数据量变化原因，抽样及合并方法及合理性。数据库/  
693 集包括不限于基础数据库、标注数据库、训练集、调优集、  
694 测试集、对抗测试集、公开数据库，测评数据库、临床试验  
695 数据库。

696 每个数据库建议提供以下信息：

697 -数据库名称及版本、类型（如外部、内部；训练、测  
698 调优、测试）

699 -受检者数据收集的纳入和排除标准

700 -人口统计学数据（年龄、种族、人种）

701 -采集方式（前瞻性、回顾性）

702 -放射学检查条件，如 CT 增强扫描，CT 平扫

703 -采集数据的方法

704 数据来源（体检中心、医学影像诊断中心、门诊、病房  
705 等）

706 设备厂家、型号、探测器排数

707 扫描参数：管电压范围、管电流范围、层厚范围、辐射  
708 剂量（低剂量平扫 CT、常规剂量平扫或增强扫描）、图像像  
709 素、矩阵、扫描视野

710 -重建方式（标准算法或其他）

711           -样本量（正常案例数、疾病案例数、确定疾病状态、  
712 位置、范围的方法）

713           -按影像因素分层的病例分布，如年龄、性别、地区、  
714 病变类型、图像重建层厚/层间距、疾病类型、重建方式、  
715 扫描协议、成像厂家、医疗机构。

716           需明确总体数据范围及分层数量及占比，并需列表对比  
717 分析各数据库信息。

718           采用公开数据数据库的需明确用途，不应作为测评数据  
719 库使用。

720

721