

中华人民共和国医药行业标准

YY/T 1907—2023

人工智能医疗器械 冠状动脉 CT 影像 处理软件 算法性能测试方法

Artificial intelligence medical device—Coronary CT image processing software—
Algorithm performance test methods

2023-09-05 发布

2024-09-15 实施

国家药品监督管理局 发布

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由国家药品监督管理局提出。

本文件由人工智能医疗器械标准化技术归口单位归口。

本文件起草单位：中国食品药品检定研究院、上海长征医院、中国医学科学院阜外医院、首都医科大学附属朝阳医院、中国人民解放军东部战区总医院、国家药品监督管理局医疗器械技术审评中心、中国科学技术大学苏州高等研究院、清华大学、辽宁省医疗器械检验检测院、中国人民解放军总医院、上海联影智能医疗科技有限公司、上海西门子医疗器械有限公司、科亚医疗科技股份有限公司、深圳睿心智能医疗科技有限公司、数坤(北京)网络科技股份有限公司、推想医疗科技股份有限公司、杭州深睿博联科技有限公司、飞利浦(中国)投资有限公司。

本文件主要起草人：李佳戈、刘士远、吕滨、萧毅、杨旗、张龙江、彭亮、王晨希、王晶、王浩、郝烨、孟祥峰、周少华、冯建江、刘凯、张龙达、何昆仑、李澍、吴迪嘉、费震宇、李广、马骏、郑超、王少康、乔昕、葛鑫、刘盼。

人工智能医疗器械 冠状动脉 CT 影像 处理软件 算法性能测试方法

1 范围

本文件描述了采用人工智能技术的冠状动脉 CT 影像处理软件的算法性能测试方法。

本文件适用于采用人工智能技术对冠状动脉 CT 影像进行后处理的软件产品。

本文件不适用于影像前处理和过程优化。

注：冠状动脉 CT 影像的采集方式包括冠状动脉 CT 血管成像 (coronary computed tomography angiography, CCTA) 和 CT 平扫成像。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

YY/T 1833.1 人工智能医疗器械 质量要求和评价 第 1 部分：术语

YY/T 1833.2 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

YY/T 1833.3 人工智能医疗器械 质量要求和评价 第 3 部分：数据标注通用要求

3 术语和定义

YY/T 1833.1、YY/T 1833.2、YY/T 1833.3 界定的以及下列术语和定义适用于本文件。

3.1

冠状动脉中心线 coronary centerline

沿着血管的轴向、连接冠状动脉各个横断面图像上的管腔中心点形成的曲线。该曲线的起点是冠状动脉开口，终点是各支冠状动脉清晰可辨的最远端。

注：冠状动脉中心线一般呈现在曲面影像上。

3.2

CT 血流储备分数 CT fractional flow reserve; CT FFR

根据冠状动脉 CT 血管影像计算得到的血流储备分数。

注：附录 A 给出了血流储备分数的补充说明。

3.3

冠状动脉钙化积分 coronary artery calcium score

计算冠状动脉钙化程度的一项量化指标，一般在冠状动脉 CT 血管成像前使用心电门控进行 CT 平扫（轴位扫描），结合数理公式对冠状动脉钙化进行量化。

注 1：目前常用的计算方法包括 Agatston 积分、体积积分和质量积分。

注 2：Agatston 积分是冠状动脉钙化积分的常用指标，附录 B 补充说明其计算方式。

注 3：冠状动脉钙化是根据病变大小超过某一面积或体积阈值、密度超过规定的 CT 值阈值来定性，不同计算方法不尽相同，指南推荐采用 Agatston 积分。

注4：钙化积分包括左主干(LCA)、前降支(LAD)、回旋支(LCX)、右冠状动脉(RCA)各自积分,以及相加后总积分(即患者维度的钙化积分)。

4 算法性能测试要求

4.1 概述

冠状动脉 CT 影像处理软件的算法性能测试过程宜参照 YY/T 1858—2022 中 4.1 的要求,建立测试文档;如测试过程需要复测,对复测次数进行限制,避免算法对参考标准进行推测或针对性调优。

4.2 测试环境

算法性能的测试环境要求宜参照 YY/T 1858—2022 中 4.2。

4.3 测试资源

4.3.1 测试集要求

测试集的质量应满足 YY/T 1833.2 的要求;测试集应独立于算法训练、调优过程,保证封闭性和安全性;制造商可根据产品预期用途和临床应用场景,对测试数据进行限定。

4.3.2 数据采集要求

4.3.2.1 影像质量要求

冠状动脉 CT 影像宜满足如下要求:

- 测试集使用影像设备采集的原始图像,进行脱敏,去除个人敏感信息;脱敏后不对图像内容进行修改;
 - 每个病例的影像序列保持连续、完整;
 - 来自非能谱 CT 的冠状动脉 CT 影像明确冠状动脉左主干和 3 支冠状动脉近中段(RCA、LAD、LCX)的 CT 值范围;
- 注1: CT 值范围例如不低于 350HU。
- 明确图像的信噪比,可使用主动脉根部管腔内的 CT 值标准差数值;
 - 血管和器官成像扫描范围宜完全覆盖冠状动脉;
 - 避免运动、重组、金属、噪声、对比剂显影不足等情形带来的伪影;
 - 避免不合适的扫描参数或过低管电压导致的信噪比降低。

数据集制造责任方应记录影像采集的技术参数。

注2: 示例见附录 B 中 B.2。

4.3.2.2 测试集样本量要求

对于辅助诊断场景下的算法准确性测试,测试集宜参照 YY/T 1858—2022 中 4.3.2 计算。其他场景下的算法测试可参考的样本量计算公式见附录 B。

4.3.2.3 测试集多样性要求

测试数据集宜体现产品适用范围和临床使用场景内的样本多样性,包括成像设备维度、患者维度。

成像设备维度考虑以下内容:

- 图像信噪比,例如依据左主干和 3 支冠状动脉近中段管腔内 CT 值,考虑增强图像的信噪比区间;

——CT 设备、型号、成像参数的多样性。

患者维度考虑以下内容。

——性别、年龄等因素的多样性。

——产品适用范围,考虑生理结构、病变类型或分级,如下:

- 1) 冠状动脉供血类型(右优势型、左优势型、均衡型);
- 2) 有无解剖变异;
- 3) 不同分级的管腔狭窄;
- 4) 斑块类型;
- 5) 斑块尺寸、分布范围;
- 6) 高危斑块类型特征;
- 7) 慢性完全闭塞、冠状动脉瘘;
- 8) 单支病变、多支病变(两支病变、三支病变)、多节段病变(含弥漫病变)、分叉病变;
- 9) 钙化积分的数值、分级;
- 10) 心肌异常密度影(如心肌不均匀密度影,包括区域心肌低密度影、脂肪密度影、少见高密度钙化影)。

4.3.2.4 扩增数据要求

在算法鲁棒性测试中,可参照 YY/T 1858—2022 中 4.3.4 对测试集进行扩增;扩增数据的参考标准应通过人工确认后交付使用。

4.3.3 测试工具要求

算法测试使用的体模、标准器宜参照 YY/T 1858—2022 中 4.3.5 的要求。

用于采集有创 FFR 的压力导丝应具有医疗器械注册证。

4.4 测试平台

如使用测试平台进行算法测试,测试平台宜满足 YY/T 1858—2022 中 4.4 的要求。

4.5 测试指标与通过准则

测试指标与通过准则的选取宜参照执行 YY/T 1858—2022 中 4.5 的一般要求;根据产品的适用功能选择测试指标,一般按照病灶、节段、血管和患者等不同的统计维度进行计算。

测试计划应根据产品预期用途制定,描述病灶的定义、测量方式、分类/分级方式、冠状动脉分段方式、病例总体结论的确立规则等适用的信息,作为开展测试的依据。

4.6 测试流程

测试人员宜参照 YY/T 1858—2022 中 4.6 的一般要求,根据测试计划开展测试活动,建立测试记录。

4.7 测试结果

测试人员宜参照 YY/T 1858—2022 中 4.7 的一般要求,描述测试结果。

5 算法性能测试方法

5.1 算法应用场景的测试方法

5.1.1 区域分割

如待测产品预测管腔区域,宜使用 Dice 系数计算预测区域和参考标准区域的重合度,计算公式见公式(1)。

$$Dice = 2 \times \frac{|A \cap B|}{|A| + |B|} \dots\dots\dots(1)$$

式中:

- Dice —— Dice 系数;
- A —— 预测区域;
- B —— 参考标准区域。

注: 预测区域、参考标准区域可能是连续的区域,也可能是断续区域的累加。

如待测产品预测管腔边界,可使用单向 Hausdorff 距离表征预测边界和参考边界之间的相似程度,可使用双向 Hausdorff 距离表征预测边界和参考边界之间的最大不匹配程度,两者的计算公式见公式(2)。

$$d_H(X, Y) = \max\{d_{XY}, d_{YX}\} = \max\{\max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y)\} \dots\dots\dots(2)$$

式中:

- $d_H(X, Y)$ —— 双向 Hausdorff 距离;
- X —— 预测的分割边界;
- Y —— 人工标注的分割边界;
- d_{XY} —— 从 X 到 Y 的单向 Hausdorff 距离,用于表征预测边界和参考边界之间的相似程度;
- d_{YX} —— 从 Y 到 X 的单向 Hausdorff 距离,不常使用;
- $d(x, y)$ —— 点 x 到点 y 的距离, x 与 y 分别为 X 和 Y 中的任意点。

在血管维度计算上述指标,然后在整个测试集上求平均值和标准差,作为最终结果。

5.1.2 冠状动脉中心线提取与测量

如中心线提取与测量功能不作为产品的主要临床性能指标,本文件不对测试指标与方法进行限定。制造商可参考附录 B 的方法,开展过程验证。

5.1.3 冠状动脉分段识别

第一步,根据制造商声称的分段方式和依据,在同一病例的冠状动脉各个分段上分别比较算法的输出结果和参考标准,构造混淆矩阵,计算同一病例的总体准确率、各分段的灵敏度、特异度、准确率。

第二步,对测试集各病例的混淆矩阵进行合并,计算测试集整体的准确率、各分段的灵敏度、特异度、准确率。

计算时,参考标准管腔直径 < 1.5 mm 的分段应从混淆矩阵中剔除。

混淆矩阵的构造方法和相关指标的计算公式见 YY/T 1858—2022 中 5.1.3,后同。

注 1: 常见的分段方式包括 15、17 或 18 分段,由制造商根据临床文献进行选择。

注 2: 分段依据是产品判定某段冠状动脉的具体分段名称的依据。例如,某产品对某段冠状动脉的管腔区域进行分割,当分割结果与参考标准中的某分段的 Dice 系数大于制造商规定的阈值时,认为产品分割的区域对应该分段。

5.1.4 钙化积分评估

5.1.4.1 一致性分析

当待测产品输出钙化积分的具体数值(连续性变量)时,宜使用算法预测的钙化积分、参考标准给出的钙化积分,开展两组数据之间的一致性分析(测试集整体)。

注:这里一般指患者水平的总钙化积分;当产品输出单支血管的钙化积分时,则对血管水平的钙化积分进行计算。一致性分析可选用以下指标和计算方法。

- Bland-Altman 分析:以算法预测结果减去参考标准结果的差值为纵轴,以算法预测结果和参考标准的算数平均值为横轴,绘制散点图,并标示 95% 一致性界限。
- 组内相关系数:两组钙化积分数值之间的一致性可使用公式(3)进行计算。

$$ICC = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_\epsilon^2} \quad \dots\dots\dots (3)$$

式中:

ICC —— 组内相关系数;

σ_s^2 —— 钙化积分预测值与参考标准之间的方差;

σ_ϵ^2 —— 钙化积分预测值、参考标准内部的方差。

- Pearson 相关系数:算法预测的钙化积分和参考标准钙化积分的协方差除以它们标准差的乘积,其计算方式见公式(4)。

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad \dots\dots\dots (4)$$

式中:

$\rho(X, Y)$ —— Pearson 相关系数;

X —— 算法预测钙化积分的平均值;

Y —— 参考标准钙化积分的平均值;

μ_X —— X 的平均值;

μ_Y —— Y 的平均值;

σ_X —— X 的标准差;

σ_Y —— Y 的标准差。

当待测算法输出钙化积分的分级时,可对算法预测钙化积分、参考标准钙化积分同时进行升序或降序排序后,计算 Spearman Rank(斯皮尔曼等级)相关系数,计算方式见公式(5)。

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad \dots\dots\dots (5)$$

式中:

r_s —— 斯皮尔曼等级相关系数;

n —— 钙化积分数据点总量;

d_i —— 两个数组第 i 个点的等级差。

注:斯皮尔曼等级相关系数仅作为衡量一致性的方式之一,不能取代基于混淆矩阵的分析。

5.1.4.2 基于混淆矩阵的分析

当待测算法输出钙化积分分级时,宜根据算法输出结果和参考标准的分级构造混淆矩阵,计算分级的准确率、灵敏度、特异度和 Cohen's Kappa 系数。

5.1.5 冠状动脉狭窄评估分级

首先,在发生冠状动脉狭窄的各分段比较算法输出的狭窄评估分级和参考标准,构造混淆矩阵;其次,根据混淆矩阵,计算测试集整体的准确率、灵敏度、特异度、阳性预测值、阴性预测值;最后,构造受试者操作特征曲线,计算曲线下面积。

注:以上方法分别以血管、患者为个体展开计算。

5.1.6 CT 血流储备分数计算

首先,使用有创压力导丝测得的 FFR 作为金标准,在 FFR 的测量位置(一般为冠状动脉狭窄远端 2 cm~3 cm)比较算法输出的 CT 血流储备分数和金标准的分类结果,构造混淆矩阵;其次,根据混淆矩阵,计算测试集整体的准确率、灵敏度、特异度、阳性预测值、阴性预测值;最后,构造受试者操作特征曲线,计算曲线下面积。

5.1.7 斑块评估

5.1.7.1 斑块检出

在同一幅 CT 图像上(推荐使用横断面),计算算法预测斑块区域和参考标准斑块区域之间的交并比,如交并比大于制造商规定的阈值,则认为匹配成功,记为真阳性结果(TP);如果算法预测的某斑块区域无对应的参考标准区域,则记为假阳性结果(FP);如果参考标准斑块区域未被算法检出,则记为假阴性结果(FN)。计算整个测试集的 TP、FP、FN 数量,进而计算斑块检出的召回率、精确度。

当斑块跨越多幅 CT 图像时,其中任何一幅图像满足匹配关系,均认为该斑块被正确检出,否则属于假阳性结果。

注:本文件以血管分段作为斑块检出、分类的最小单元;测试结果一般体现病例、血管层面。

5.1.7.2 斑块分类

对于被正确检出的斑块,比较算法给出的分类结果与参考标准,基于测试集整体构造混淆矩阵。混淆矩阵的维度推荐考虑以下情形。

——混淆矩阵常见包括 3 种分类:钙化斑块、非钙化斑块、混合斑块。

——如产品声称可进行针对易损斑块/高危斑块的分类评估,混淆矩阵亦可包括此分类结果。易损斑块的影像学表征如低密度斑块、正性重构、点状钙化、餐巾环征。根据混淆矩阵,可计算准确率、灵敏度、特异度、阳性预测值、阴性预测值。

——可根据患者有无斑块,构造混淆矩阵,以患者作为测试集的单元,计算灵敏度、特异度、准确率。构造受试者操作特征曲线,计算曲线下面积。

5.2 算法质量特性与测试方法

5.2.1 泛化能力

宜参照 YY/T 1858—2022 中 5.2.1,结合 4.3.2.3 的要求和制造商规定的产品适用范围,对测试样本进行抽取和组合,观测算法在测试集的不同子集上的性能差异,子集一般体现患者个体差异(如年龄、性别)、设备差异、成像条件差异、图像质量、病变多样性。

5.2.2 鲁棒性

宜参照 YY/T 1858—2022 中 5.2.2,基于制造商规定的产品适用范围开展对抗测试、压力测试,形成鲁棒性研究资料。

注：对抗测试、压力测试场景下，根据需要引入噪声、伪影数据，用于数据扩增。

对抗测试宜考虑 CT 设备差异和数据采集条件差异导致的图像变化，可使用模拟仿真产生的数据，在人工确认后用于测试。

在不影响医生阅片的前提下，对抗测试可在冠状动脉 CT 影像上叠加轻微或局部的噪声、伪影，例如：

- a) 运动伪影；
- b) 错层伪影；
- c) 晕状伪影；
- d) 螺旋内插伪影。

压力测试宜引入临床常见干扰项、罕见样本、复杂样本（但仍符合数据质量要求），举例如下：

- a) 心肌病灶；
- b) 心腔病灶；
- c) 冠状动脉搭桥术后、植入冠状动脉支架后；
- d) 冠状动脉-肺动脉瘘；
- e) 慢性完全闭塞。

5.2.3 重复性

宜参照 YY/T 1858—2022 中 5.2.3，对同一版本的算法至少进行 3 次重复测试，观察测试结果是否变化。

5.2.4 一致性

如算法输出的中间结论具有参考标准，宜参照 YY/T 1858—2022 中 5.2.4，使用 5.1 的适用方法对中间结论进行验证，对算法输出的中间结论和产品输出的最终结论进行比较。

注：例如当产品判断某病例的冠状动脉狭窄程度为重度时，算法判断的冠状动脉各分段狭窄程度是否与该结论一致。

5.2.5 效率

宜参照 YY/T 1858—2022 中 5.2.5，计算算法运行的效率。

宜以测试数据开始导入的时刻作为起点，以算法输出结果的时刻作为终点。

如适用，对于 CT FFR 应分别记录完成图像分割、完成 FFR 计算的效率。

5.2.6 算法错误统计

宜参照 YY/T 1858—2022 中 5.2.6，对算法测试错误结果进行统计，纳入测试结果，如：

- 冠状动脉狭窄程度的误判；
- 斑块检出的假阴、假阳；
- 中心线提取与测量误差；
- FFR 假阴、假阳结果；
- 图像分割误差。

附录 A

(资料性)

对相关医学概念的补充说明

A.1 血流储备分数(FFR)

本文件讨论的 FFR,指的是冠状动脉存在狭窄病变的情况下,该血管所供心肌区域能获得的最大血流与同一区域正常情况下所能获得的最大血流之比,计算方式为狭窄远端冠状动脉内平均压/冠状动脉口部主动脉平均压(血管最大扩张时)。临床常见一支血管上多处病变,测量点位于拟测量病变远端,根据临床规范确定测量点与病变的距离;由于冠状动脉血管细小,直径 $<2\text{ mm}$ 的血管不易用于 CT FFR 值的评估,会出现假阳性。临床一般认为 $\text{FFR}>0.8$ 的病变不会导致心肌缺血, $\text{FFR}<0.75$ 的病变存在心肌缺血,FFR 在 $0.75\sim 0.8$ 之间为灰区,需要结合临床和病变的狭窄程度综合判断。

临床上,测定 FFR 的金标准是有创压力导丝测量方法。不使用压力导丝,根据冠状动脉图像通过建模计算得到的 FFR 结果称为虚拟血流储备分数或计算血流储备分数。CT FFR 是其中的一种形式。

A.2 冠状动脉狭窄

冠状动脉狭窄,临床上通用的诊断“金标准”是经导管有创的冠状动脉造影(ICA)。虽然目前有软件定量计算管腔狭窄率,但实际临床工作中,仍然采用“目测直径法”诊断冠状动脉狭窄程度。观察者提倡依据心脏舒张期图像进行诊断,因为收缩期图像容易夸大狭窄程度。观察者选择展示狭窄的切线位图像,即选择狭窄最重的体位和角度,目测狭窄处管腔的直径,并与其近心端和远心端正常管腔直径(二者的平均值称为参考管腔直径)进行比对,定义为参考管腔直径-狭窄处直径/参考管腔直径。狭窄处管腔直径的取值若为参考管腔直径的 50% ,则称为 50% 狭窄;取值若为参考管腔直径的 30% ,则称为 70% 狭窄(即 $100\%\sim 30\%$)。根据此规则类推,完成冠状动脉狭窄程度的评估。

注: CCTA 舒张期图像的狭窄程度较 ICA 图像的狭窄程度平均夸大 $5\%\sim 10\%$,收缩期图像对狭窄程度的夸大更明显,钙化病变处夸大 $20\%\sim 30\%$;直径 $<1.5\text{ mm}$ 的血管不用于狭窄程度的诊断。

附 录 B

(资料性)

特定条款的指南和方法说明

B.1 通用指南

对于采用人工智能技术的冠状动脉 CT 影像处理软件的算法性能质量评价,测试要求贯穿整个测试过程,第 4 章对测试要求进行了详细阐述,包含概述、测试环境、测试资源、测试平台、测试指标与通过准则、测试流程和测试结果。第 5 章对算法性能测试方法进行了描述,按照场景测试和质量特性两个角度展开。

B.2 特定条款的指南或方法说明

以下内容与正文的条款相对应,提供推荐方法或举例。

对于 4.3.2 数据采集要求,建议结合临床文献、指南、专家共识等,考虑成像设备和成像过程中技术参数的要求,在执行过程中参考以下信息:

- a) CT 设备的硬件参数,如探测器数目、球管旋转时间、空间分辨率。建议采用 64 排及以上、球管旋转时间 ≤ 350 ms、具有亚毫米级空间分辨率的 CT 设备;
- b) 条件具备时,提供同一患者 3 个月以内的冠状动脉造影图像;
- c) 对于传统 64 排 CT,建议将心率控制在 70 次/min 以下;对于 64 排及以上的单源 CT,建议将心率控制在 80 次/min 以下;对于双源 CT,根据设备性能要求推荐心率低于 90 次/min;
- d) 建议心电门控电极按照美国标准(表 B.1)或欧洲标准(表 B.2)进行放置;R 波清晰可见,波形规律,无杂波干扰;
- e) 如使用压力导丝进行有创 FFR 测量,在均衡(EQ)时狭窄远端冠状动脉内平均压(Pd)、冠状动脉口部主动脉平均压(Pa)的平均压差不大于 ± 5 mmHg,在校验(verify)时的平均压差不大于 ± 3 mmHg。

表 B.1 安放心电门控电极的美国标准

导联名称	导联位置
白色导联	右锁骨中线、锁骨下
黑色导联	左锁骨中线、锁骨下
红色导联	左锁骨中线、第六或第七肋间
绿色导联	右锁骨中线、第六或第七肋间

表 B.2 安放心电门控电极的欧洲标准

导联名称	导联位置
红色导联	右锁骨中线、锁骨下
黄色导联	左锁骨中线、锁骨下
黑色导联	右锁骨中线、第六或第七肋间
绿色导联	左锁骨中线、第六或第七肋间

对于冠状动脉 CT 平扫及钙化积分扫描,建议考虑以下信息。

- a) 采用各 CT 设备推荐的默认参数。
- b) 扫描范围一般仅覆盖心脏。记录扫描范围时,一般包含上界和下界具体数值,建议扫描范围的上界在气管隆嵴下 1 cm~2 cm 水平(根据患者体型调整);下界在心脏膈面左右各大于心缘两侧 1 cm~2 cm。
- c) 根据设备条件和心率条件首选前瞻性心电门控扫描;辐射剂量宜按照尽可能的最低原则。
- d) 一般进行冠状动脉 CT 平扫,除非冠状动脉及旁路有金属植入物干扰(如金属支架)等情形导致不可接受的伪影。

对于冠状动脉 CT 血管成像,建议考虑以下信息。

- a) 对于患者体重≤60 kg,管电压宜采用 70 kV 或 80 kV;患者体重 60 kg~90 kg,管电压宜采用 100 kV 或 120 kV;患者体重≥90 kg,管电压宜采用 120 kV 及以上。
- b) 推荐层厚 0.500 mm~0.625 mm、使用(17 cm×17 cm)~(20 cm×20 cm)、像素 0.330 mm~0.390 mm 的重建图像;一般选择平滑的卷积核;PCI 术后选择平滑、锐利两种卷积核。
- c) 图像重建时间窗宜根据患者心率进行选择,例如:心率<70 次/min 的患者,重建时间窗建议为舒张中期(大致位于 70%~75%的 R-R 间期);心率>70 次/min 时,重建时间窗建议为收缩末期(35%~45%的 R-R 间期)。
- d) 扫描模式可参照表 B.3 的形式进行设置。

表 B.3 冠状动脉 CT 血管成像的扫描模式选择

患者心律/心率		前瞻性心电门控扫描		回顾性心电门控螺旋扫描
		前瞻性大螺距螺旋扫描	前瞻性步进式轴位扫描	
心律齐	心率<70 次/min	心率变化<3 次/min 时推荐; 心电脉冲窗宽:70% R-R 间期	推荐;心电脉冲窗宽 70% R-R 间期	不推荐
	心率 70~90 次/min	心率变化<3 次/min 时推荐; 心电脉冲窗宽:30%~70% R-R 间期	推荐;心电脉冲窗宽:30%~70% R-R 间期	不推荐
	心率>90 次/min	不推荐	推荐;心电脉冲窗宽:30%~45% R-R 间期	不推荐
心律不齐		不推荐	可尝试;心电脉冲窗宽:30%~70% R-R 间期或 R-R 绝对时相	可尝试

对于图像质量要求,建议考虑以下信息:

- a) 来自非能谱 CT 的冠状动脉 CT 血管图像建议 CT 值处于 300 HU~450 HU 范围内,不小于 200 HU。
- b) 主动脉根部管腔内的 CT 值标准差宜低于 30 HU,不超过 40 HU。

对于测试集多样性要求中的冠状动脉狭窄分级,建议参考表 B.4。

表 B.4 冠状动脉狭窄程度的分级描述

狭窄程度	描述
正常冠状动脉	无粥样硬化斑块,管腔狭窄 0
轻微狭窄	有粥样硬化斑块,管腔狭窄 1%~24%
轻度狭窄	有粥样硬化斑块,管腔狭窄 25%~49%
中度狭窄	有粥样硬化斑块,管腔狭窄 50%~69%
重度狭窄	有粥样硬化斑块,管腔狭窄 70%~99%
完全闭塞	有粥样硬化斑块,管腔狭窄 100%
危重病变(如适用)	左主干>50%狭窄或 3 支病变(均≥70%狭窄)

对于图像质量的多样性,可依据冠状动脉 CT 图像的 CT 值进行划分,例如优秀(351 HU~450 HU)、良好(300 HU~350 HU)、一般(200 HU~300 HU)3 个区间;可依据图像噪声进行划分,例如按照主动脉根部 CT 图像的标准差取值范围分为优秀(<20 HU)、良好(20 HU~30 HU)、质量差(>30 HU)、不可接受(>40 HU)。

对于 4.3.2.2,算法测试的单次试验样本量与具体算法性能指标有关。当算法性能指标为 Pearson 相关系数时,根据制造商期望的相关系数 r 和允差 d ,依据公式(B.1)~公式(B.6)计算单次测试样本量 N 的最低要求。

$$Z_r = \frac{1}{2} \ln \frac{(1+r)}{(1-r)} \quad \dots\dots\dots (B.1)$$

$$Z_L = Z_r - Z_{1-\alpha/2} \sqrt{\frac{1}{N-3}} \quad \dots\dots\dots (B.2)$$

$$Z_U = Z_r + Z_{1-\alpha/2} \sqrt{\frac{1}{N-3}} \quad \dots\dots\dots (B.3)$$

$$r_U = \frac{\exp(2 Z_U) - 1}{\exp(2 Z_U) + 1} \quad \dots\dots\dots (B.4)$$

$$r_L = \frac{\exp(2 Z_L) - 1}{\exp(2 Z_L) + 1} \quad \dots\dots\dots (B.5)$$

$$d = r_U - r_L \quad \dots\dots\dots (B.6)$$

式中:

Z_r ——相关系数 r 经过转换后获得的标准正态分布分位数;

r ——估计的相关系数;

Z_L ——标准正态分布分位数的置信区间下限;

$Z_{1-\alpha/2}$ ——标准正态分布的 $1-\alpha/2$ 分位数;

α ——显著性水平,常用取值为 0.05;

N ——总体样本量;

Z_U ——标准正态分布分位数的置信区间上限;

r_U ——Pearson 相关系数的置信区间上限;

r_L ——Pearson 相关系数的置信区间下限;

d ——Pearson 相关系数的允差。

当算法性能指标为组间相关系数 ICC 时,根据制造商期望的相关系数 r 和允差 d ,依据公式(B.7)~公式(B.12)计算单次测试样本量 N 的最低要求。

$$r = \frac{MS_B - MS_E}{MS_B + (K - 1)MS_E} \dots\dots\dots (B.7)$$

$$r_L = \frac{\frac{MS_B}{MS_E} - 1}{\frac{MS_B}{MS_E} + K - 1} \dots\dots\dots (B.8)$$

$$r_U = \frac{\frac{MS_B}{MS_E} F_{1-\frac{\alpha}{2}, V_1, V_2} - 1}{\frac{MS_B}{MS_E} F_{1-\frac{\alpha}{2}, V_1, V_2} + K - 1} \dots\dots\dots (B.9)$$

$$V_1 = (N - b)(K - 1) \dots\dots\dots (B.10)$$

$$V_2 = N - 1 \dots\dots\dots (B.11)$$

$$d = r_U - r_L \dots\dots\dots (B.12)$$

式中：

- r —— 估计的相关系数；
- MS_B —— ANOVA 的组间均方；
- MS_E —— ANOVA 的组内均方(单向模型)或残差均方(双向模型)；
- K —— 观察者数量；
- r_L —— 相关系数的置信区间下限；
- $F_{1-\frac{\alpha}{2}, V_1, V_2}$ —— 自由度为 V_1, V_2 的 F 分布上的 $1-\frac{\alpha}{2}$ 分位数；
- α —— 显著性水平,常用取值为 0.05；
- r_U —— 相关系数的置信区间上限；
- V_1 —— 自由度 1；
- N —— 总体样本量；
- b —— 0(单向模型)或 1(双向模型)；
- V_2 —— 自由度 2；
- d —— 相关系数的允差。

对于 5.1.2,制造商对算法的中心线提取与测量结果进行量化评估时,以下方法可供参考。

首先,将算法预测中心线、参考标准中心线视为点集,由制造商在测试计划中说明中心点的采样率设置,对中心线提取的准确性进行测试,步骤如下。

- a) 中心点匹配:在同一幅 CT 图像上(推荐使用横断面图像),算法预测的管腔中心点与参考标准管腔中心点的距离不大于制造商声称的阈值(如 1 mm),则认定为匹配成功,算法预测的管腔中心点记入 TP,否则记入 FP;匹配失败的参考标准管腔中心点记入 FN。沿着冠状动脉的轴向遍历匹配关系,统计 TP、FP、FN 的总数。

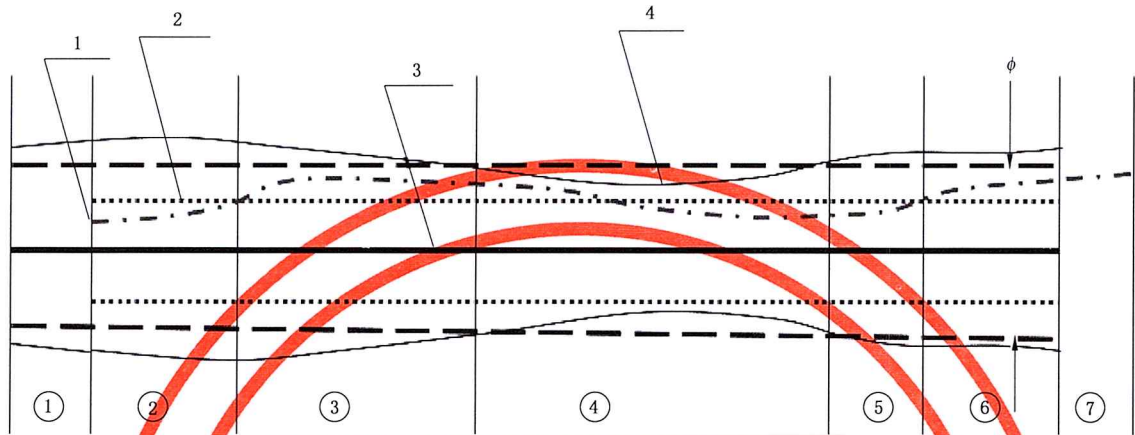
制造商可使用其他方式建立匹配关系(例如在空间三维坐标系中按照最近点的方式进行匹配),在测试计划中提供详细说明。

- b) 中心点提取评价:根据上述的 TP、FP、FN,计算测试集整体的精确度、召回率、 F_1 度量,具体计算公式见 YY/T 1858—2022 中 5.1.1。
- c) 中心线重合度:宜计算预测中心线与参考标准中心线的 Dice 系数,公式同公式(1)。

图 B.1 给出 TP、FP、FN 的判定示例,根据 TP、FP、FN 的总数,可计算中心线的重合度指标。此时,TP 的总数对应 5.1.1 公式(1)中的分子;TP 总数与 FP 总数之和对应 5.1.1 公式(1)中的 A,TP 总数

与 FN 总数之和对应 5.1.1 公式(1)中的 B。

依据上述结果,制造商可使用其他指标进一步评估 AI 预测中心线对参考标准中心线的跟踪能力,如文献[8]给出的 Overlap until first error (OF)、Overlap with the clinically relevant part of the vessel (OT)等,本文件不做详细展开。



标引序号说明:

- 1 —— AI 预测中心线;
- 2 —— 制造商自定义的距离阈值;
- 3 —— 参考标准中心线;
- 4 —— 实际的管腔边界;
- ϕ —— 1.5 mm;
- ① —— 假阴情形,该段的参考标准中心点数量计入 FN;
- ② —— AI 预测中心线与参考标准中心线匹配成功的情形,该段的中心点数量计入 TP;
- ③ —— AI 预测中心点与参考标准中心点的距离超出制造商自定义阈值的情形,该段的 AI 预测中心点数量计入 FP,参考标准中心点数量计入 FN;
- ④ —— 冠状动脉管腔直径 < 1.5mm 的情形,该段不计入统计;
- ⑤ —— 与②相同的情形,中心点数量计入 TP;
- ⑥ —— 与③相同的情形,该段的预测中心点数量计入 FP,参考标准中心点数量计入 FN;
- ⑦ —— 假阳情形(例如 AI 预测中心线长度超过参考标准时),该段的 AI 预测中心点数量计入 FP。

图 B.1 中心线提取的情形示例

其次,对中心线的测量结果进行测试,可选用以下指标和方法。

- a) 中心线长度误差:在物理坐标下,可依托中心线提取时的 TP,计算预测中心线、参考标准中心线长度的相对误差,计算方式见公式(B.13)。

$$\text{error} = \frac{|L_p - L_r|}{L_r} \times 100\% \dots\dots\dots (B.13)$$

式中:

- error —— 中心线长度的相对误差;
- L_p —— 基于预测值的中心线长度;
- L_r —— 参考标准中心线长度。

- b) 欧氏距离误差:在物理坐标下,可依托中心线提取时的 TP,(基于横断面)计算预测中心点、参考标准中心点之间的欧氏距离,计算公式见公式(B.14)。

$$E_i = \sqrt{(x_{p,i} - x_{r,i})^2 + (y_{p,i} - y_{r,i})^2 + (z_{p,i} - z_{r,i})^2} \dots\dots\dots (B.14)$$

式中：

- E_i ——第 i 个中心点对应的欧氏距离；
- $x_{p,i}$ ——第 i 个预测中心点的横坐标；
- $x_{r,i}$ ——第 i 个参考标准中心点的 x 坐标；
- $y_{p,i}$ ——第 i 个预测中心点的纵坐标；
- $y_{r,i}$ ——第 i 个参考标准中心点的 y 坐标；
- $z_{p,i}$ ——第 i 个预测中心点的纵坐标；
- $z_{r,i}$ ——第 i 个参考标准中心点的 z 坐标。

对全体中心点，计算均方误差(MSE)，计算公式见公式(B.15)。

$$MSE = \frac{\sum_{i=1}^N \{ (x_{p,i} - x_{r,i})^2 + (y_{p,i} - y_{r,i})^2 \}}{N} \dots\dots\dots (B.15)$$

式中：

- MSE ——均方误差；
- N ——中心点的总数；
- $x_{p,i}$ ——第 i 个预测中心点的横坐标；
- $x_{r,i}$ ——第 i 个参考标准中心点的横坐标；
- $y_{p,i}$ ——第 i 个预测中心点的纵坐标；
- $y_{r,i}$ ——第 i 个参考标准中心点的纵坐标。

上述计算针对测试集整体进行。

对于 5.1.3 的冠状动脉分段识别要求，可使用临床常用的 15、17 或 18 分段，由制造商明确分段方式的文献出处、分段含义及图示；如冠状动脉的分支直径 < 1.5 mm 或难以清晰识别，建议聚焦冠状动脉主干各个分段的识别准确性。

对于 5.1.4，钙化积分的标注方法补充介绍如下，供本文件的用户参考：

在冠状动脉感兴趣区域范围内寻找 CT 值 ≥ 130 HU、钙化面积 $\geq 1 \text{ mm}^2$ 的斑块，作为钙化区域。根据各钙化区域的最大密度值选择密度加权指数(最大密度值 130 HU~199 HU，取值为 1；最大密度值 200 HU~299 HU，取值为 2；最大密度值 300 HU~399 HU，取值为 3；最大密度值 ≥ 400 HU，取值为 4)，将该区域面积与密度加权指数相乘即为该钙化区域的钙化积分。左主干(LCA)、前降支(LAD)、回旋支(LCX)、右冠状动脉(RCA)钙化积分之和为该患者冠状动脉总钙化积分。

当进行基于混淆矩阵的分析时，钙化积分可依据其数值区间分为不同的风险等级，可使用以下 5 个区间：无钙化(0)，微小钙化(1~10)，轻度钙化(11~100)，中度钙化(101~399)，广泛钙化(≥ 400)；亦可使用另一种划分方式：无钙化(0)、轻度钙化(1~99)、中度钙化(100~399)、重度钙化(≥ 400)、极重度钙化(≥ 1000)。

对于 5.1.5，开展冠状动脉狭窄评估分级时，混淆矩阵推荐至少考虑表 A.4 所示的前 6 种分类；如制造商使用其他分级方式，提供依从的文献。

对于 5.1.6，进行血流储备分数计算时，根据 FFR 的取值范围，混淆矩阵建议包括 3 种分类：FFR > 0.8、FFR < 0.75、 $0.75 \leq \text{FFR} \leq 0.8$ (灰区)。

参 考 文 献

- [1] YY/T 1858—2022 人工智能医疗器械 肺部影像辅助分析软件 算法性能测试方法
- [2] 深度学习辅助决策医疗器械软件审评要点[国家药品监督管理局医疗器械技术审评中心(2019年第7号)]
- [3] 人工智能医疗器械注册审查指导原则[国家药品监督管理局医疗器械技术审评中心(2022年第8号)]
- [4] 医疗器械软件注册审查指导原则[国家药品监督管理局医疗器械技术审评中心(2022年第9号)]
- [5] Ricardo C. C, Suhny A, Stephan A, et al. CAD-RADS™ Coronary Artery Disease - Reporting and Data System: An expert consensus document of the Society of Cardiovascular Computed Tomography (SCCT), the American College of Radiology (ACR) and the North American Society for Cardiovascular Imaging (NASCI). Endorsed by the American College of Cardiology [J]. Journal of Cardiovascular Computed Tomography. 2016.10:269-281.
- [6] Suhny A, Philipp B, Christopher D. M., et al. SCCT guidelines for the performance and acquisition of coronary computed tomographic angiography: A report of the Society of Cardiovascular Computed Tomography Guidelines Committee Endorsed by the North American Society for Cardiovascular Imaging (NASCI)[J]. Journal of Cardiovascular Computed Tomography. 2016.10:435-449.
- [7] Jagat N, Y.C, Amir A., et al. SCCT 2021 Expert Consensus Document on Coronary Computed Tomographic Angiography: A Report of the Society of Cardiovascular Computed Tomography [J]. Journal of Cardiovascular Computed Tomography. 2021.15:192-217.
- [8] H Greenspan, Laifentfeld M., Einav S., et al. Evaluation of Center-Line Extraction Algorithms in Quantitative Coronary Angiography [J]. IEEE TRANSACTIONS ON MEDICAL IMAGING MI, 2001, 20(9):928-941.
- [9] Schaap M, Metz CT, van Walsum T, van der Giessen AG, et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms[J]. Med Image Anal. 2009 Oct;13(5):701-14. doi: 10.1016/j.media.2009.06.003. Epub 2009 Jun 30. PMID: 19632885; PMCID: PMC3843509.
- [10] 国家心血管病专业质控中心心血管影像质控专家工作组,中华医学会放射学分会心胸学组,《中华放射学杂志》心脏冠状动脉多排CT临床应用指南写作专家组. 冠状动脉CT血管成像的适用标准及诊断报告书写规范[J]. 中华放射学杂志, 2020, 54(11): 1044-1055. DOI: 10.3760/cma.j.cn112149—20200413-00544.
- [11] 国家心血管病专业质控中心专家委员会心血管影像质控专家工作组. 冠状动脉CT血管成像扫描与报告书写专家共识[J]. 协和医学杂志, 2019, 10(1): 23-30. DOI: 10.3969/j.issn.1674-9081.2019.01.005.
- [12] 中华医学会放射学分会质量控制与安全管理专业委员会,江苏省医学会放射学分会智能影像与质量安全学组. 冠状动脉CT血流储备分数应用中国专家建议[J]. 中华放射学杂志, 2020, 54(10): 925-933. DOI: 10.3760/cma.j.cn112149—20191108-00896.
- [13] 《冠状动脉钙化病变诊治中国专家共识》专家组. 冠状动脉钙化病变诊治中国专家共识(2021版)[J]. 中国介入心脏病学杂志, 2021, 29(5): 251-259. DOI: 10.3969/j.issn.1004-8812.2021.05.002.
- [14] 《中国冠状动脉血流储备分数测定技术临床路径专家共识》专家组. 中国冠状动脉血流储备

分数测定技术临床路径专家共识[J]. 中国介入心脏病学杂志, 2019, 27(3): 121-133. DOI: 10.3969/j.issn.1004-8812.2019.03.001.

[15] 冠状动脉血流储备分数临床应用专家共识专家组. 冠状动脉血流储备分数临床应用专家共识[J]. 中华心血管病杂志, 2016, 44(4): 292-297. DOI: 10.3760/cma.j.issn.0253-3758.2016.04.004.

[16] 中华医学会放射学分会心胸学组, 《中华放射学杂志》心脏冠状动脉多排 CT 临床应用指南写作专家组. 心脏冠状动脉 CT 血管成像技术规范应用中国指南[J]. 中华放射学杂志, 2017, 51(10): 732-743. DOI: 10.3760/j.issn.1005-1201.2017.10.004.

[17] 中华放射学杂志心脏冠状动脉多排 CT 临床应用协作组. 心脏冠状动脉多排 CT 临床应用专家共识[J]. 中华放射学杂志, 2011, 45(1): 9-17. DOI: 10.3760/cma.j.issn.1005-1201.2011.01.003.

[18] 徐橙, 易妍, 李颜玉, 等. 深度学习重建算法在提升冠状动脉 CT 血管成像钙化所致血流梗阻性病变诊断准确性中的价值[J]. 中华医学杂志, 2021, 101(39): 3202-3207. DOI: 10.3760/cma.j.cn112137—20210304-01391.

[19] 中国医师协会放射医师分会. 冠状动脉 CT 血管成像斑块分析和应用中国专家建议[J]. 中华放射学杂志, 2022, 56(6): 595-607. DOI: 10.3760/cma.j.cn112149—20211129-01055.

中华人民共和国医药
行业标准
人工智能医疗器械 冠状动脉 CT 影像
处理软件 算法性能测试方法
YY/T 1907—2023

*

中国标准出版社出版发行
北京市朝阳区和平里西街甲 2 号(100029)
北京市西城区三里河北街 16 号(100045)

网址 www.spc.net.cn

总编室:(010)68533533 发行中心:(010)51780238

读者服务部:(010)68523946

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

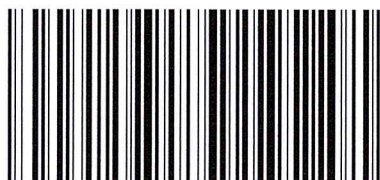
*

开本 880×1230 1/16 印张 1.5 字数 37 千字
2023 年 9 月第一版 2023 年 9 月第一次印刷

*

书号: 155066·2-37564 定价 36.00 元

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68510107



YY/T 1907-2023

