

YY

中华人民共和国医药行业标准

YY/T XXXXX—XXXX

人工智能医疗器械 脑卒中 CT 影像辅助分 析软件 算法性能测试方法

Artificial intelligence medical device—Computer assisted analysis software for
stroke CT images—Algorithm performance test methods

（征求意见稿）

本草案完成时间：2024 年 7 月 21 日

XXXX—XX—XX 发布

XXXX—XX—XX 实施

国家药品监督管理局 发布

目 次

前 言 II

引 言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 算法性能测试要求 2

5 算法性能测试方法 4

附 录 A （资料性） 对相关医学概念的补充说明..... 12

附 录 B （资料性） 特定条款的指南和方法说明..... 14

附 录 C （资料性） 头部 CT 测试集描述样例..... 22

参 考 文 献 25

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：

本文件主要起草人：

引 言

近年来,人工智能算法在脑卒中CT影像辅助决策软件当中的应用逐渐增多。已经获得国家药品监督管理局注册的第三类医疗器械已涵盖颅内出血图像辅助分诊软件、头颈血管造影图像辅助评估软件、缺血性卒中图像辅助评估软件、脑缺血灌注图像辅助评估软件等多个产品。

不同的人工智能模型和算法,对产品的有效性与安全性影响较大。算法性能测试是产品质量评价的重要环节。针对脑卒中常见的临床应用场景,本文件作为方法标准,针对现有产品,面向颅内出血检测、血肿风险评估、脑缺血程度评分、颅脑灌注分析、颅内血管病变检测等辅助分诊、辅助检测、辅助诊断类的辅助决策任务,对相应的算法性能指标的定义、计算方式、测试过程进行规范,旨在加强相关产品的质量评价。

人工智能医疗器械 脑卒中 CT 影像辅助分析软件 算法性能测试方法

1 范围

本标准规定了采用人工智能技术对脑卒中CT影像进行辅助分析的软件算法性能的测试方法。

本标准适用于采用人工智能技术对脑卒中CT影像进行后处理的辅助分析软件。

本标准不适用于影像前处理、过程优化软件。

注：头部CT影像的采集方式包括CT平扫成像(Non-Contrast Computed Tomography, NCCT)、CT血管成像(Computed Tomography Angiography, CTA)和CT灌注成像(Computed Tomography Perfusion, CTP)

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 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

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

3 术语和定义

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

3.1

CT 平扫成像 Non-Contrast Computed Tomography, NCCT

不使用任何对比剂或增强剂，使用计算机断层扫描（CT）技术获得的影像。

3.2

CT 血管成像 Computed Tomography Angiography, CTA

通过静脉内注射对比剂，并在目标血管中对比剂浓度达到最高峰的时间点进行CT扫描，获得的血管系统增强影像。

注：CTA原始影像经计算机图像后处理，形成三维立体血管图像。

3.3

CT 灌注成像 Computed Tomography Perfusion, CTP

通过静脉快速团注对比剂后，对感兴趣脑区进行连续多次CT扫描，获得的带时间序列信息的脑灌注影像。

注1：灌注影像在感兴趣脑区的时间-密度曲线反映对比剂在脑血管和脑组织中的扩散和清除情况。

注2：对时间-密度曲线的后处理计算，才可获取量化的组织血流灌注参数。

3.4

梗死核心 infarct core

缺血性卒中所致的脑组织死亡的区域。

3.5

缺血半暗带 penumbra

梗死核心周围存在的因缺血发生细胞功能障碍，但尚未死亡的脑组织区域。

注：及时恢复血流灌注能使缺血半暗带部分的脑组织恢复正常，但梗死核心的脑组织不可逆转。

3.6

脑中线 brain midline

两侧大脑半球之间的大脑镰，以及侧脑室之间的透明隔。

注1：正常情况下，脑中线位于脑组织横断面影像的正中间，并与额顶骨正中骨嵴连线重合。

注2：脑组织的异常病变，如血肿、水肿或萎缩等，导致中线结构受压或受牵拉，偏向健侧或病灶侧。

3.7

前循环 Alberta 卒中项目早期 CT 评分 Alberta Stroke Program Early CT Score, ASPECTS

用于评价前循环急性缺血性卒中患者早期缺血性改变病情程度的评分方式。

注：附录A.1给出评分方式和细则。

3.8

后循环急性卒中预后早期 CT 评分 posterior circulation Acute Stroke Prognosis Early CT Score, pc-ASPECTS

用于评价后循环急性缺血性卒中患者早期缺血性改变病情程度的评分方式。

注：附录A.2给出评分方式和细则。

3.9

脑血流量 Cerebral Blood Flow, CBF

单位时间内血液通过脑组织截面的流量。具体来说，是指每100克脑组织在每分钟通过的血流量，单位为毫升每100克每分钟（ml/(100g·min)）。

3.10

脑血容量 Cerebral Blood Volume, CBV

每100克脑组织内所含血液的量，单位为毫升每100克（ml/100g）。

3.11

平均通过时间 Mean Transit Time, MTT

血液从进入脑组织区域到完全离开该区域所需的平均时间，单位为秒（s）。

3.12

残差函数达峰时间 time to maximum of the residual function, Tmax

脑灌注成像过程中，造影剂以理想瞬时弹丸方式注入供血动脉后，从造影剂开始进入脑组织到组织中造影剂浓度达到峰值所需的时间，单位为秒（s）。

4 算法性能测试要求

4.1 概述

脑卒中CT影像辅助分析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，应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

脑卒中CT影像测试集的构建考虑以下因素：

- 测试集宜使用影像设备采集的原始图像；
- 测试集影像应进行脱敏，去除个人敏感信息，应保留影像扫描参数信息；
- 脱敏后不得对影像进行任何修改，不得进行有损压缩等操作；
- 每个病例的影像序列应保持连续、完整；
- 影像序列的层厚、层间距、空间分辨率等，应符合制造商宣称的参数；

- f) NCCT、CTA、CTP 影像宜以听眶线为扫描起始基线向上扫描，直至颅顶最高点；
- g) CTA 影像扫描范围宜完全覆盖大脑前动脉、中动脉、后动脉、基底动脉，以及椎动脉和颈内动脉颅内段；
- h) CTA 影像应至少包含一个动脉期。
- i) CTP 影像宜包含动脉期和静脉期在内的至少 10 期影像，扫描总时长至少 25s。
- j) 对卒中患者常见的因躁动导致存在运动伪影的影像，或存在金属伪影的影像，宜以不影响测试集参考标准的确立为纳入依据；
- k) 对因对比剂显影不足或扫描参数不合适等原因导致的低信噪比影像，亦宜以不影响测试集参考标准的确立为纳入依据。

注：听眶线：外耳孔中点与同侧眼眶下缘间的连线。

4.3.2.2 测试集样本量要求

对预期用于分类或检出的产品，宜参照YY/T 1858-2022中4.3.2进行计算，明确单次测试样本量的最低要求。

对预期用于尺寸及体积计算的产品，当算法性能指标为Pearson相关系数或组间相关系数ICC时，宜参照YY/T 1907-2023中附录B进行计算，明确单次测试样本量的最低要求。

对预期用于图像分割及体积计算的辅助检测类产品，测试人员宜根据主要评价指标的目标值、最大允差、检验水准、把握度等，计算单次测试的样本量要求，计算公式见公式（1）。

$$n = \frac{(\mu_{\alpha} + \mu_{\beta})^2 \sigma_d^2}{\delta^2} \dots\dots\dots (1)$$

式中：

α 、 β ——分别表示可接受的检验水准（第Ⅰ类错误）和把握度（第Ⅱ类错误），通常 α 取值为0.05， β 取值为0.10；

δ ——主要评价指标的最大允差，可为单侧或双侧，优效性检验时为单侧；

μ_{α} 、 μ_{β} —— α 、 β 下的 μ 检验临界值，根据允差单双侧确定单侧检验或双侧检验的临界值；

σ_d ——预试验过程中主要评价指标在算法输出与参考标准间差值的标准差。

对其他预期用途的产品，制造商宜描述单次测试样本量选取的依据。

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。从成像设备维度宜考虑：

- CT 设备、型号、成像参数的多样性；
- CTP 时序扫描的多样性，考虑多种时序间隔。

从患者维度宜考虑：

- 人群分布多样性：如性别、年龄、地域等。
- 制造商可根据产品适用范围，考虑生理结构、病变类型或分级上的多样性，如：
 - 1) 不同类型的脑出血卒中（如脑实质出血、蛛网膜下腔出血、硬膜外出血、硬膜下出血、脑室积血）；
 - 2) 不同类型的脑缺血卒中（如前循环缺血、后循环缺血，大血管闭塞型梗死、穿支血管闭塞型梗死、分水岭梗死）；
 - 3) 不同位置的脑出血和脑缺血卒中（如脑干、小脑、额叶、颞叶、枕叶、顶叶、基底节）；
 - 4) 梗死核心体积的不同范围（如按 30mL、50mL、70mL、100mL 界值划分）；
 - 5) 缺血半暗带体积的不同范围（如<15mL、≥15mL）；
 - 6) ASPECTS 评分的不同范围（如<6、≥6）；
 - 7) 有无陈旧性脑梗死灶；
 - 8) 有无颅内血管解剖变异；
 - 9) 不同分级的血管管腔狭窄；
 - 10) 血管单支病变、多支病变、串联病变。

4.3.2.4 扩增数据

原则上不得对测试集进行数据扩增，对抗测试除外。在算法可靠性、鲁棒性等对抗测试中，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。

4.4 测试平台

如使用测试平台，测试平台宜满足YY/T 1858-2022中4.4的要求，在符合制造商声称的算法运行环境下开展测试，对算法输出结果和参考标准进行预览和比对。

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，一般按照患者维度进行统计计算。制造商可根据产品预期用途和临床应用场景，进一步按照病灶类型（如出血分类）、病灶体积大小（如梗死核心）、血管分段（如前后循环）、病变严重程度（如ASPECTS评分范围）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，建议描述病灶的定义、检测/分析/计算方式、分类/分级/分期方式、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 颅内出血

5.1.1.1 颅内出血分类识别

如待测产品作为辅助分诊软件以患者为单元识别是否存在颅内出血灶的结论，宜参考YY/T 1858-2022中5.1.3.1的表2构建二分类混淆矩阵，宜参考YY/T 1858-2022中5.1.3.2至5.1.3.10的公式计算各性能指标。

如待测产品分类识别颅内出血位置、出血类型或血肿扩大风险等级等多分类结论，宜参考YY/T 1858-2022中5.1.3.1的表1构建多分类混淆矩阵，宜参考YY/T 1858-2022中5.1.3.7和5.1.3.9的公式计算多分类性能指标。

注：此处构建的多分类混淆矩阵，适用于在某一统计维度下多分类仅单选的情况，如按患者维度下血肿扩大的风险等级。若某一统计维度下多分类为多选，则先将其统计维度转变为仅单选，然后再建立混淆矩阵进行性能评价。

如按患者维度，出血类型为多选，则先将统计维度转为病灶维度，再进行评价。

构建多分类混淆矩阵后，若需将多分类转为二分类，应使用如本文件表1所示方法进行转化。转化后宜参考YY/T 1858-2022中5.1.3.2至5.1.3.10的公式计算转化为二分类后的各性能指标。

表1 多分类混淆矩阵转化为二分类混淆矩阵

分类		人工智能分类	
		阳性	阴性
参考标准分类	阳性	$TP = \sum_{i \in P, j \in P} N_{i,j}$	$FN = \sum_{i \in P, j \in N} N_{i,j}$
	阴性	$FP = \sum_{i \in N, j \in P} N_{i,j}$	$TN = \sum_{i \in N, j \in N} N_{i,j}$

注：n为多分类类型个数，各类型集合记为 $C = \{1, 2, \dots, n\}$ 。 $N_{i,j}$ ($i = 1 \sim n, j = 1 \sim n$)为参考标准分类为i类，被人工智能分类为j类的个数。 P 为多分类类型中可转化为二分类中“阳性”类别的分类集合，记 $P = \{P_1, P_2, \dots, P_m\}$ ； N 为多分类类型中可转化为二分类中“阴性”类别的分类集合，记 $N = \{N_1, N_2, \dots, N_k\}$ 。满足 $m + k = n$ 、 $P \cup N = C$ 及 $P \cap N = \emptyset$ （空集）。 TP 为真阳性； FP 为假阳性； FN 为假阴性； TN 为真阴性。

5.1.1.2 颅内出血区域分割

5.1.1.2.1 区域定义

如待测产品识别输出血肿或积血区域的结论，宜使用图1所示定义区域，A为参考标准分割的目标区域，B为算法辅助分割的目标区域，C为二者的交集（ $C = A \cap B$ ），D为影像有效信息区域（不含背景区域）。A、B、C、D均为连续三维空间或非连续三维空间的并集累加。测试记录应描述其构成的闭合表面是否进行平滑处理，以及所采用平滑方法。

目标区域也可按影像切片逐层以二维平面空间形式给出，此时可以切片层序数*i*作为下标，以*A_i*、*B_i*、*C_i*、*D_i*表示各区域。

注1：如产品分割可给出多种类型的病灶，算法性能测试宜分病灶类别各自独立进行。

注2：对分割性能的评价宜在三维空间中以体素级别进行。如遇特殊应用场景，需在二维空间中逐层评价性能，则对每层测试算法性能后，建议给出各指标的平均值。

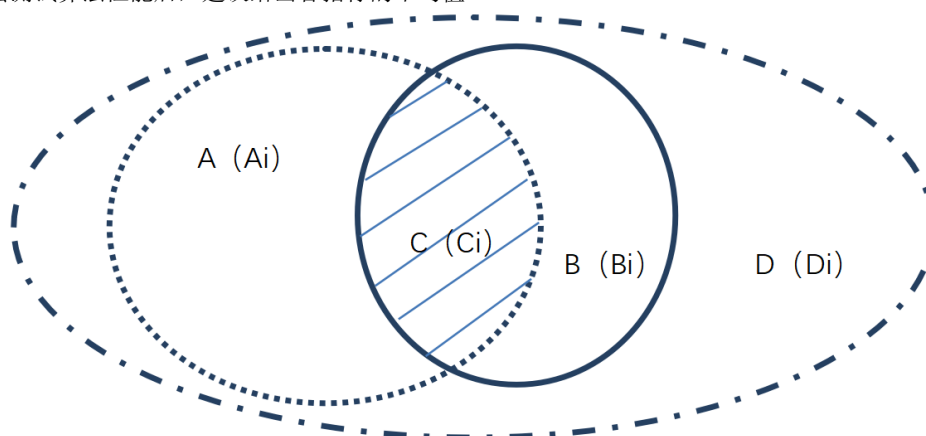


图1 区域分割定义

5.1.1.2.2 灵敏度

灵敏度为算法分割的目标区域与参考标准分割的目标区域的交集除以参考标准分割的目标区域，用公式（2）表示：

$$SEN = \frac{TP}{TP+FN} = \frac{|A \cap B|}{|A|} = \frac{|C|}{|A|} \dots\dots\dots (2)$$

式中：

SEN——灵敏度，也可称为召回率；

TP ——真阳性区域；

FN ——假阴性区域；

A ——参考标准分割的目标区域；

B ——算法分割的目标区域；

C ——算法分割的目标区域与参考标准分割的目标区域的交集。

5.1.1.2.3 特异度

特异度为算法分割的目标区域与参考标准分割的目标区域并集之外的有效信息区域除以参考标准分割的目标区域之外的有效信息区域，用公式（3）表示：

$$SPE = \frac{TN}{TN+FP} = \frac{|D-A \cup B|}{|D-A|} \dots\dots\dots (3)$$

式中：

SPE——特异度；

TN ——真阴性区域；

FP ——假阳性区域；

A ——参考标准分割的目标区域；

B ——算法分割的目标区域；

D ——影像中的有效信息区域。

5.1.1.2.4 阳性预测值

阳性预测值为算法分割的目标区域与参考标准分割的目标区域的交集除以算法分割的目标区域，用公式（4）表示：

$$PPV = \frac{TP}{TP+FP} = \frac{|A \cap B|}{|B|} = \frac{|C|}{|B|} \dots\dots\dots (4)$$

式中：

PPV ——阳性预测值，也可称为精确度；

TP ——真阳性区域；

FP ——假阳性区域；

A ——参考标准分割的目标区域；

B ——算法分割的目标区域；

C ——算法分割的目标区域与参考标准分割的目标区域的交集。

5.1.1.2.5 阴性预测值

阴性预测值为算法分割的目标区域与参考标准分割的目标区域并集之外的有效信息区域除以算法分割的目标区域之外的有效信息区域，用公式（5）表示：

$$NPV = \frac{TN}{TN+FN} = \frac{|D-A \cup B|}{|D-B|} \dots\dots\dots (5)$$

式中：

NPV ——阴性预测值，也可称为回溯精确度；

TN ——真阴性区域；

FN ——假阴性区域；

A ——参考标准分割的目标区域；

B ——算法分割的目标区域；

D ——影像中的有效信息区域。

5.1.1.2.6 漏检率

漏检率用*MR*表示，表达式见公式（6）：

$$MR = 1 - SEN \dots\dots\dots (6)$$

式中：

SEN ——灵敏度。

5.1.1.2.7 约登指数

约登指数用*Y*表示，表达式见公式（7）：

$$Y = SEN + SPE - 1 \dots\dots\dots (7)$$

式中：

SEN ——灵敏度；

SPE ——特异度。

5.1.1.2.8 Dice 系数

Dice系数为算法分割的目标区域与参考标准分割的目标区域交集的两倍除以两者之和。Dice系数用*DICE*表示，表达式见公式（8）：

$$DICE = 2 \times \frac{|A \cap B|}{|A| + |B|} \dots\dots\dots (8)$$

式中：

A ——参考标准分割的目标区域；

B ——算法分割的目标区域。

5.1.1.2.9 Jaccard 系数

Jaccard系数为算法分割的目标区域与参考标准分割的目标区域交集除以两者的并集。Jaccard系数用 JAC 表示，表达式见公式（9）：

$$JAC = \frac{|A \cap B|}{|A \cup B|} \dots\dots\dots (9)$$

式中：

- A ——参考标准分割的目标区域；
B ——算法分割的目标区域。

5.1.1.2.10 表面距离

表面距离为算法和参考标准分割的目标区域之间的距离度量，可用于评价轮廓或边界的分割效果。A、B区域的双向豪斯多夫距离（Hausdorff Distance）用 $H(A, B)$ 表示，表达式见公式（10）：

$$H(A, B) = \max \left\{ \max_{a \in A_x} \left\{ \min_{b \in B_x} \|a, b\| \right\}, \max_{b \in B_x} \left\{ \min_{a \in A_x} \|b, a\| \right\} \right\} \dots\dots\dots (10)$$

式中：

- A ——参考标准分割的目标区域；
B ——算法分割的目标区域；
 A_x ——参考标准分割的目标区域的边界；
 B_x ——算法分割的目标区域的边界；
 a —— A_x 中的任一点；
 b —— B_x 中的任一点；
 $\|a, b\|$ ——点 a 和点 b 之间的距离范式，如欧几里得距离。

5.1.1.2.11 尺寸测量

对分割目标区域，可进行尺寸测量，如长短径等。对这些尺寸参数，可评价算法分割区域结果与参考标准目标区域结果之间的差异，计算绝对误差或相对误差、绝对误差或相对误差的绝对值以及它们的平均值和标准差等参数。

5.1.1.2.12 体积测量

分别统计算法分割的目标区域和参考标准分割的目标区域内的体素数量，乘以每个体素的体积，可计算目标区域的总体积，评价算法分割区域与参考标准区域之间的差异，计算体积绝对误差或相对误差、绝对误差或相对误差的绝对值以及它们的平均值和标准差等参数。

5.1.1.2.13 一致性分析

一致性分析可选用以下指标和计算方法。

- a) 皮尔逊（Pearson）相关系数：以 R 表示，表达式见公式（11）：

$$R = \frac{\sum_n [(V_A^n - \bar{V}_A) * (V_B^n - \bar{V}_B)]}{\sqrt{\sum_n [(V_A^n - \bar{V}_A)^2] \sum_n [(V_B^n - \bar{V}_B)^2]}} \dots\dots\dots (11)$$

式中：

- V ——对分割目标区域的测量参数，如体积；
 V_A^n ——第 n 个病例的参考标准分割目标区域的参数；
 V_B^n ——第 n 个病例的算法辅助分割目标区域的参数；
 $\bar{V}_A$ ——所有病例的参考标准分割目标区域的参数的平均值；
 $\bar{V}_B$ ——所有病例的算法辅助分割目标区域的参数的平均值。
b) Bland-Altman 分析：以算法分割结果（如血肿体积）减去参考标准结果（如专家标注血肿体积）的差值为纵轴，以算法分割结果和参考标准的算术平均值为横轴，绘制散点图，并标示 95% 一致性界限（95% limit of agreement, 95% LOA），结合由制造商根据临床文献、指南、专家共识等确定的可接受最大差值，评价算法分割与参考标准是否具有较好的一致性。

5.1.2 脑中线提取与测量

5.1.2.1 脑中线提取

算法提取的脑中线与人工标注参考标准提取的脑中线，在脑CT各轴位影像上表现为线段，宜使用线段角度误差和线段点集之间的双向豪斯多夫距离评价重合度，如：

- a) 脑中线角度误差绝对值，用 MA 表示，表达式见公式（12）为：

$$MA_n = |AL_A^n - AL_B^n| \dots\dots\dots (12)$$

- b) 脑中线双向豪斯多夫距离，用 MH 表示，表达式见公式（13）为：

$$MH_n = \max \left\{ \max_{a \in L_A^n} \left\{ \min_{b \in L_B^n} \|a, b\| \right\}, \max_{b \in L_B^n} \left\{ \min_{a \in L_A^n} \|b, a\| \right\} \right\} \dots\dots\dots (13)$$

公式（12）至（13）中：

n ——轴位影像序号；

AL_A^n ——第 n 张轴位影像上，人工标注参考标准提取的脑中线与水平线的夹角；

AL_B^n ——第 n 张轴位影像上，算法提取的脑中线与水平线的夹角；

L_A^n ——第 n 张轴位影像上，人工标注参考标准提取的脑中线；

L_B^n ——第 n 张轴位影像上，人工标注参考标准提取的脑中线；

a —— L_A^n 上的任一点；

b —— L_B^n 上的任一点；

$\|a, b\|$ ——点 a 和点 b 之间的距离范式，如欧几里得距离。

在患者维度计算上述指标的平均值，然后在整个测试集上求平均值和标准差，作为最终结果。

5.1.2.2 脑中线偏移量

脑中线偏移量以脑中线为测量基线，以中脑、第三脑室、透明隔、大脑镰为标志，标志点与测量基线间的垂直距离为脑中线偏移量。图2示意了3层4个不同标记点处的中线偏移测量情况。

计算算法提取的脑中线偏移量与人工标注参考标准提取的脑中线偏移量之间的误差，如：

- a) 脑中线偏移绝对误差，用 AE 表示，表达式见公式（14）为：

$$AE_k = |ML_A^k - ML_B^k| \dots\dots\dots (14)$$

- b) 脑中线偏移相对误差，用 RE 表示，表达式见公式（15）为：

$$RE_k = \frac{|ML_A^k - ML_B^k|}{ML_A^k} \times 100\% \dots\dots\dots (15)$$

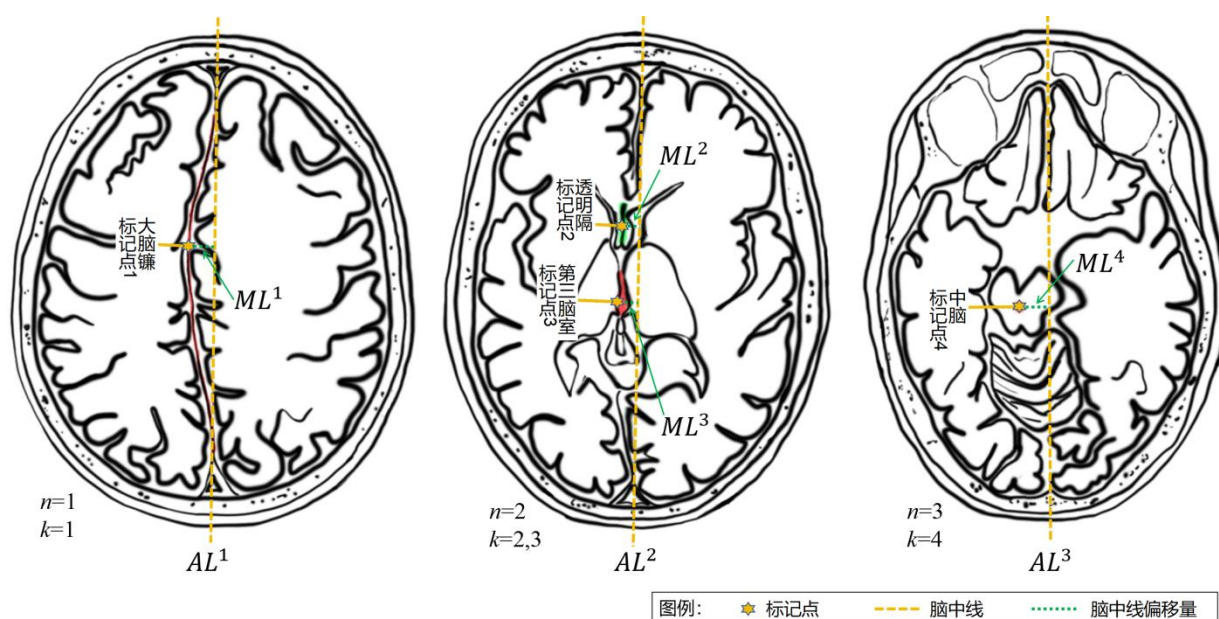
公式（14）至（15）中：

k ——标记点序号；

ML_A^k ——标记点 k 处，人工标注参考标准提取的脑中线偏移量；

ML_B^k ——标记点 k 处，算法提取的脑中线偏移量。

在患者维度计算上述指标的平均值，然后在整个测试集上求平均值和标准差，作为最终结果。



标引序号说明：

AL^n ——第 n 层的脑中线；

ML^k ——第 k 个标记点处的脑中线偏移量。

图2 脑中线偏移测量示意图

5.1.3 颅内血管评估

5.1.3.1 颅内血管重建

颅内血管重建需待测产品识别输出颅内血管管腔所在区域范围，宜参考5.1.1.2的要求，计算算法分割血管区域的各性能指标，以及血管管径和长度测量参数与参考标准间的差异，和一致性分析指标。

5.1.3.2 颅内血管中心线提取与测量

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YY/T 1907-2023中附录B，开展过程验证。

5.1.3.3 颅内动脉血管分段识别

第一步，根据制造商声称的分段方式和依据，在测试集各病例的颅内动脉血管各个分段上分别比较记录算法的输出结果和参考标准，构建混淆矩阵。

第二步，计算测试集整体的灵敏度、特异度、准确率；或计算某一分段的灵敏度、特异度、准确率。

注1：常见的分段方式见附录B.5，由制造商根据临床需求结合文献进行选择。对制造商宣称不适用的分段（如远端血管分段）需从混淆矩阵中删除。

注2：分段定位判定依据是产品判定某段颅内动脉的分段结果是否匹配参考标准的依据。例如，某产品对某段颅内动脉血管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，认为产品分割区域对应的分段与参考标准的分段匹配。

5.1.3.4 颅内血管病变检测

本文件所指颅内血管病变主要包括血管狭窄和血管闭塞。对血管狭窄可进一步按狭窄程度进行分级。分级方式和标准，由制造商根据临床需求结合文献进行选择。

注：颅内血管狭窄程度分级方式和标准，可参见附录B.6。

如待测产品以患者为单元识别是否存在血管病变的二分类结论，宜参考5.1.1.1的方法计算二分类性能指标。

如待测产品以血管分段为单元识别各分段是否存在血管病变，或输出血管病变分级结论，宜参考5.1.1.1的方法计算多分类性能指标。

5.1.4 ASPECTS 评分

5.1.4.1 评分一致性

当待测产品输出ASPECTS评分值时，可按算法输出的实际分值和参考标准给出的评分值，开展两组数据之间的一致性分析。

注：ASPECTS评分含义及评分标准见附录A.1和A.2。

一致性分析宜参考5.1.1.2.13的指标和计算方法，或参考YY/T 1907-2023中5.1.4.1的b)的公式计算组内相关系数。

5.1.4.2 评分分类

如制造商宣称按ASPECTS评分值进行临床常用亚组分类作为产品的主要性能指标，宜参考5.1.1.1的方法计算二分类或多分类性能指标。

注：ASPECTS评分在临床中常用的亚组分类方法见附录A.1和A.2，或由制造商根据临床文献、指南、专家共识等确定。

5.1.4.3 评分区域分割

当待测产品识别输出ASPECTS评分区域的结论，宜作为区域分割任务进行性能测试，参考5.1.1.2的要求，计算算法分割ASPECTS评分区域的各性能指标。

注：ASPECTS评分区域的定义见附录A.1和A.2。

5.1.5 脑灌注评估

5.1.5.1 分诊分类

如待测产品作为辅助分诊软件以患者为单元识别是否存在梗死核心，或是否存在低灌注区，或是否存在缺血半暗带等二分类结论，宜参考5.1.1.1的方法计算二分类性能指标。

5.1.5.2 区域分割

如待测产品识别输出梗死核心，或低灌注区，或缺血半暗带位置区域的结论，宜作为区域分割任务进行性能测试，参考5.1.1.2的要求，计算算法分割的梗死核心区域、低灌注区域、缺血半暗带区域的各性能指标，以及各区域尺寸和体积测量参数与参考标准间的差异及一致性分析指标。

5.1.5.3 一致性分析

当待测产品以体素为单元，或以特定脑区的体素空间集合为单元，计算输出如CBF、CBV、MTT、Tmax等灌注参数时，制造商可参考附录B.7和B.8的方法构建脑灌注数字体模，或采用脑灌注数字体模的公开数据集，并将产品应用于该数字体模，解算输出灌注参数。以数字体模体素内标称的灌注参数，或体素空间集合内各体素标称灌注参数的均值为金标准，参考5.1.1.2.11至5.1.1.2.13的要求，可计算各灌注参数的绝对误差或相对误差、绝对误差或相对误差的绝对值、以及它们的平均值和标准差、Pearson相关系数、Bland-Altman图等性能指标。

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，以通过不同子集的多样性和变化性，观测算法的性能差异。子集一般体现患者的个体差异（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，评估临床使用中可能干扰算法性能的因素，收集真实世界的数据或使用模拟仿真产生的数据，组成专用测试集，依据5.1对算法性能进行对抗测试、压力测试，形成鲁棒性研究资料。

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。

在不影响阅片的前提下，对抗测试可在头颅CT影像上叠加轻微或局部的噪声、运动伪影、金属伪影。

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) 合并血管病变，如动静脉畸形、动脉瘤；
- b) 合并脑小血管病影像学改变，如脑白质高信号、腔隙性梗死灶、微出血灶；
- c) 合并脑出血后手术治疗，如去骨瓣减压术后、钻孔引流术后、开颅血肿清除术后；
- d) 合并植入物，如脑血管支架植入术后、脑部电极片植入术后；
- e) 合并其他病变，如颅外伤、颅内肿瘤、脑炎。

5.2.3 重复性

宜参照YY/T 1858-2022中5.2.3，对同一版本的算法使用相同的样本，使用5.1的适用方法进行不低于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备参考标准，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与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颅内血管狭窄程度为重度时，算法判断的颅内各分段血管的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；

5.2.6 算法错误统计

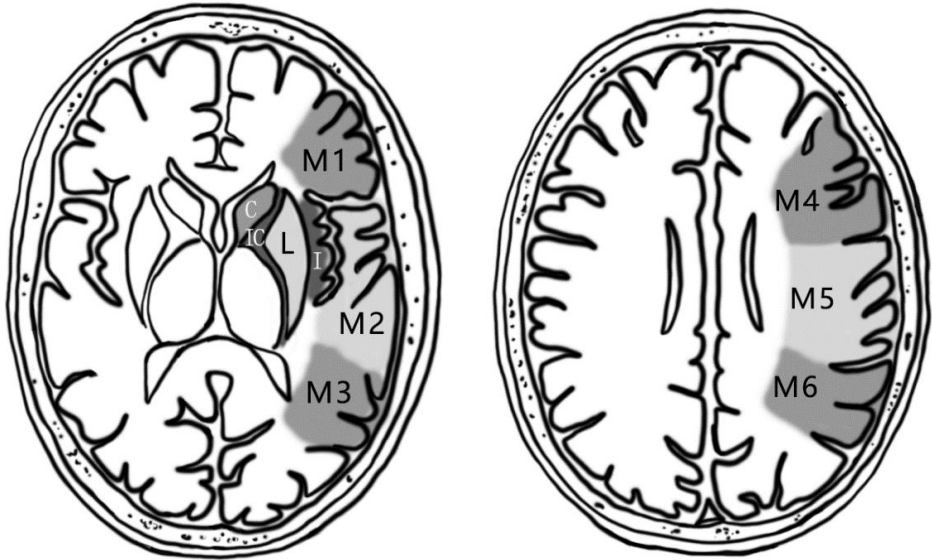
宜参照YY/T 1858-2022中5.2.6，对算法测试的错误结果按制造商规定的方法进行统计，纳入测试结果，如：

- a) 颅内出血部位的误判；
- b) 颅内出血类型的误判；
- c) 脑中线提取与测量误差；
- d) 脑血管狭窄程度的误判；
- e) 脑血管闭塞结论的误判；
- f) 颅内出血、梗死核心、缺血半暗带等区域分割误差与体积差异；
- g) ASPECTS 评分的误判。

附 录 A
(资料性)
对相关医学概念的补充说明

A. 1 前循环 ASPECTS

本文件讨论的前循环ASPECTS是《Lancet》杂志在2000年发表的一种基于颅脑NCCT平扫系统评价前循环动脉供血区早期脑梗死改变的评分方法，见参考文献[5]。基于颅脑NCCT平扫图像两个经典层面：核团层面（即丘脑和纹状体平面）和核团以上层面（在核团层面上方2cm）。ASPECTS将大脑中动脉供血区划分为10个区域，分别是尾状核（Caudate, C）、内囊（Internal Capsule, IC）、豆状核（Lentiform, L）、岛叶（Insular ribbon, I）和皮层区域M1-M6。M1、M2和M3分别为前部皮质、岛叶外侧皮质和后部皮质，位于核团层面。M4、M5和M6分别为M1、M2和M3相对应的上部，位于核团以上层面。前循环ASPECTS评分层面和评分区域示意图见图A. 1。



图A. 1 前循环 ASPECTS 评分层面及评分区域示意图

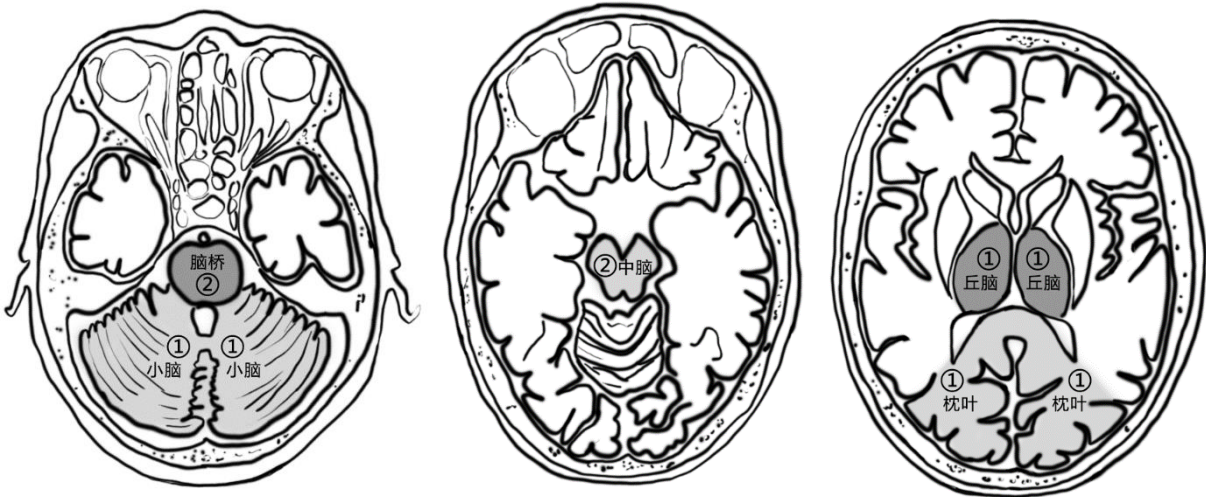
ASPECTS 10个评分区域体积不同，但权重相同，均为1分，详见表A. 1。具体的评分方法是从10分中减去存在早期梗死改变的区域所对应的分值，得到ASPECTS分值结果。ASPECTS分值范围位于0-10分，10分代表前循环动脉供血区无早期梗死改变，分值越低提示早期梗死范围越大，0分表示前循环动脉供血区广泛性早期梗死改变。临床使用ASPECTS评分，常按分值归并为亚组，如分为0-5分组和6-10分，或分为0-2分、3-5分、6-10分。

表A. 1 前循环 ASPECTS 评分分值表

区域	分值
尾状核/C	1
内囊/IC	1
豆状核/L	1
岛叶/I	1
M1	1
M2	1
M3	1
M4	1
M5	1
M6	1

A.2 后循环 ASPECTS (pc-ASPECTS)

本文件讨论的pc-ASPECTS是《Stroke》杂志在2008年发表的一种基于颅脑NCCT平扫和颅脑CTA图像系统评价后循环动脉供血区早期脑梗死改变的评分方法，见参考文献[6]。基于颅脑NCCT平扫图像的三个层面：小脑半球层面、中脑层面和丘脑层面。pc-ASPECTS将后循环动脉供血区划分为8个区域，分别是中脑、脑桥、双侧丘脑、双侧小脑半球和双侧大脑后动脉供血区。后循环ASPECTS评分层面和评分区域示意图见图A. 2。



图A. 2 后循环 ASPECTS 评分层面及评分区域示意图

pc-ASPECTS 8个评分区域体积不同，权重也不同，详见表A. 2。具体的评分方法是从10分中减去存在早期梗死改变的区域所对应的分值，得到pc-ASPECTS分值结果。pc-ASPECTS分值范围位于0-10分，10分代表后循环动脉供血区无早期梗死改变，分值越低提示早期梗死范围越大，0分表示后循环动脉供血区广泛性早期梗死改变。临床使用pc-ASPECTS评分，常按分值归并为亚组，如分为0-5分组和6-10分组。

表A. 2 后循环 ASPECTS 评分分值表

区域	分值
脑桥	2
中脑	2
左侧丘脑	1
右侧丘脑	1
左侧小脑半球	1
右侧小脑半球	1
左侧大脑后动脉供血区	1
右侧大脑后动脉供血区	1

附 录 B (资料性) 特定条款的指南和方法说明

B.1 CT 扫描前的准备

- a) 患者仰卧于检查床，摆好体位，必要时采用头部固定带制动。
- b) 射线防护：对患者非检查部位辐射敏感器官佩戴铅衣或铅围裙。禁止家属陪同，如病情需要必须家属陪同的，陪同者须佩戴铅衣。

需要注射对比剂患者，在执行过程中应参考以下信息：

- c) 停用二甲双胍：肾功能异常者，使用对比剂前需提前 48 小时以上暂停使用二甲双胍，使用对比剂后仍需停药 48~72 小时，并在复查肾功能正常后方可继续用药；肾功能正常者，使用对比剂前无需停用二甲双胍，但在在使用对比剂后应在医师指导下停用 48~72 小时，并在复查肾功能正常后方可继续用药。
- d) 碘过敏试验：如产品说明书无特别要求，原则上不建议采用碘过敏试验进行碘过敏反应预测。
- e) 对比剂选择：推荐采用亲水性较好的非离子型次高渗碘对比剂，或非离子型等渗碘对比剂，并依据相关指南选择合适碘流率。

B.2 CT 平扫的成像设备和成像过程中的技术参数要求

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。在执行过程中可参考以下信息：

- a) 定位：以听眶线为基准行平面扫描。
- b) 范围：原则上覆盖全脑，从后颅窝底部向上扫描至颅顶。根据临床需求不同，适当缩小或扩大扫描范围，但完整包含病变所在范围。
- c) 扫描参数：管电压 120kV；管电流 200~300mA；常规层厚 5~8mm；常规层距 5~8mm；1~2mm 层厚更好。

B.3 CT 血管成像的成像设备和成像过程中的技术参数要求

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) 扫描范围一般覆盖全脑，根据临床需求不同，可适当缩小或扩大扫描范围。
- b) 扫描参数：管电压 120kV，管电流 240~480mA 左右，采集层厚 0.50~1.25mm，重建层距 0.50~1.25mm，重建矩阵 512x512。

B.4 CT 灌注成像的成像设备和成像过程中的技术参数要求

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) 扫描范围及层厚：根据 CT 设备探测器层数，可选择全脑或者部分灌注成像。全脑灌注成像覆盖全脑，从颅底至脑突面最高点；部分灌注成像需根据 NCCT 结果，对病变区域进行扫描。幕上病推荐覆盖核团层面和核团以上层面。
- b) 扫描参数：管电压 80~120kV，管电流 120~150mA，扫描时间 40~50 秒。扫描层厚 0.5~5mm，重建层厚和层距 2.5~5mm。

B.5 CTA 头颈动脉分段识别要求

针对颈内动脉，目前临床使用最广的分段方法为顺血流的7段法——Bouthillier分段法。该分段是1996年Bouthillier等在参考文献[8]提出的，以数字（C1~C7）顺血流方向标志颈内动脉全程，并考虑到对神经外科具有重要意义的颈内动脉周围解剖，且各段的解剖分界明确。具体地，Bouthillier分段描述如表B.1。

表B.1 颈内动脉 Bouthillier 分段法

分段	描述
C1 颈段	颈段起于颈总动脉分叉水平，终止于颈动脉管颅外口。
C2 岩段	这段颈内动脉位于颈动脉管内，起于颈动脉管颅外口，终止于破裂孔后缘。
C3 破裂孔段	破裂段起于颈动脉管末端，动脉越过孔部，但不穿过这个孔，在破裂孔的垂直管内上升，向着海绵窦，止于岩舌韧带上缘。
C4 海绵窦段	此段始于岩舌韧带上缘，止于近侧硬膜环。
C5 床突段	此段起于近侧硬膜环，止于远侧硬膜环。
C6 眼段	该段起于远侧硬膜环，止于后交通动脉起点的紧近侧。
C7 交通段	交通段起于紧靠后交通动脉起点的近侧，止于颈内动脉分叉处。

椎动脉起自锁骨下动脉第一段发出后上升并经除第7颈椎外的横突孔，自寰椎横突孔穿出，绕经寰椎侧块后方的椎动脉沟，转向上方经枕骨大孔进入颅内。按照解剖行程，椎动脉分为4段，如表B.2所示。

表B.2 椎动脉分段描述

分段	描述
V1颈段	自锁骨下动脉发出后，至进入横突孔之前的部分。
V2椎骨段	为穿经横突孔的部分。
V3枕段	自寰椎横突孔穿出部至进入颅内的部分（枕下三角）。
V4颅内段	接枕段上行，穿过蛛网膜，在脑桥上缘左右汇合成基底动脉，供应小脑后下及C1~C2段脊髓。

大脑前动脉（ACA）是颈内动脉的主要分支之一，主要供应大脑半球内侧。ACA经胼胝体膝部至其背侧面，沿胼胝体沟向后穿行至胼胝体压部稍前方，并分出终末支。根据Fischer分类法（5段），具体描述如表B.3所示。

表B.3 大脑前动脉分段描述

分段	描述
A1 水平段	起始后向前内经视交叉背面折入大脑纵裂至前交通动脉。
A2 垂直段	为前交通动脉至胼胝体膝部下方的一段，侧位片上呈由后下向前上方走行，略成“S”形。
A3膝段	以额极动脉与A2段为界，与胼胝体的膝弯曲一致。
A4 胼周段	位于胼胝体沟内，也叫胼周动脉，分布于额叶。
A5 终段	为楔前动脉，分布于顶叶。

大脑中动脉是颈内动脉的直接延续，是其最大分支，供应区域最广，供应整个大脑半球的背外侧面：额中回以下，中央前后回下3/4，顶下小叶，枕外侧回，颞下回上缘，颞极内、外侧面，额叶眶部外侧半及岛叶各部皮质。具体分段描述如表B.4所示。

表B.4 大脑中动脉分段描述

分段	描述
M1水平段	自颈内动脉分叉部起点延伸至侧裂。
M2脑岛段	自M1段远端大脑中动脉主干转向后上形成的膝部处至侧裂顶部到达环状沟的终端。
M3侧裂段	由侧裂顶部转向外侧离开侧裂开始先在大脑半球表面行走。
M4和M5 合称为终末段或 皮层支	始于侧裂表面较靠前的分支出侧裂后陡然向上或向下转弯。

大脑后动脉可以分为4段，如表B.5所示。

表B.5 大脑后动脉分段描述

分段	描述
P1交通前段	起源于基底动脉的末端，终止于后交通动脉处。
P2环池段	由水平段转折向后上方走行，止于四叠体池。
P3四叠体段	走行于四叠体池的后内侧，终止于进入枕叶沟处。
P4距裂段	为大脑后动脉进入距状沟后的部分。

B. 6 CTA 头颈动脉狭窄程度的分级描述

头颈动脉血管分为颅内和颅外动脉，相应的狭窄率计算方法不同。颈内动脉和椎动脉跨越颅内和颅外两个区域。根据Bouthillier分段方法，颈内动脉被分为7段，其中C5段（即床突段）是进入蛛网膜下隙的起点，因此被视为颅内外段的分界；而椎动脉的颅内和颅外段则以枕骨大孔为分界。颅内动脉包涵了颈内动脉的C6-7段、大脑中动脉、大脑前动脉、大脑后动脉、椎动脉的V4段以及基底动脉；而颅外动脉则包括颈内动脉的C1-5段、颈外动脉、椎动脉的V1-3段、颈总动脉、锁骨下动脉起始段以及主动脉弓。

结合多个指南推荐[8][9]，评估颅外动脉的狭窄率推荐采用NASCET（North American Symptomatic Carotid Endarterectomy Trial，NASCET）法。NASCET法是通过比较血管最狭窄部分的残余管腔直径与狭窄远端正常的颈内动脉管腔直径来计算狭窄率，计算公式为：狭窄率=[(狭窄远端正常直径-狭窄段最窄直径)/狭窄远端正常直径]×100%。

针对颅外动脉，不同指南中狭窄分级的狭窄率范围有一定差异，主要是轻度狭窄和中度狭窄的分界不一致，如表B. 6所示。制造商需要明确颅外动脉的狭窄分级规则，以避免争议和混淆。

表B. 6 颅外动脉狭窄分级描述

狭窄分级	描述
未见狭窄	无动脉粥样硬化斑块，无狭窄
轻度狭窄	有动脉粥样硬化斑块，狭窄率<50%（参考文献[8]） 有动脉粥样硬化斑块，狭窄率<30%（参考文献[9]）
中度狭窄	有动脉粥样硬化斑块，狭窄率50%~69%（参考文献[8]） 有动脉粥样硬化斑块，狭窄率30%~69%（参考文献[9]）
重度狭窄	有动脉粥样硬化斑块，狭窄率70%~99%
闭塞	有动脉粥样硬化斑块，狭窄率100%

对于颅内动脉狭窄率的计算，目前广泛接受的是WASID（The Warfarin-Aspirin Symptomatic Intracranial Disease Study，WASID）所公布的方法。其计算公式涉及病变血管的最狭窄处直径（Ds）和病变血管正常处的直径（Dn）。制造商需要明确Dn的选择标准，以避免争议和混淆。

在评估头颈动脉血管狭窄程度时，通常采用未见狭窄、轻度狭窄、中度狭窄、重度狭窄和闭塞的五级分类方法。针对颅内动脉，不同狭窄分级的狭窄率范围一般没有争议，如表B. 7所示。

表B. 7 颅内动脉狭窄分级描述

狭窄分级	描述
未见狭窄	无动脉粥样硬化斑块，无狭窄
轻度狭窄	有动脉粥样硬化斑块，狭窄率<50%
中度狭窄	有动脉粥样硬化斑块，狭窄率50%~69%
重度狭窄	有动脉粥样硬化斑块，狭窄率70%~99%
闭塞	有动脉粥样硬化斑块，狭窄率100%

B. 7 CTP 参数计算方式

CTP下灌注参数的解算以下述公式为约束，以体素为单位，逐体素解算。

根据参考文献[10][11]，由CTP测得的信号强度—时间曲线计算对比剂浓度—时间曲线的公式如(B. 1)所示：

$$S(t) = S_0 + k \times C(t) \cdots \cdots \cdots (B. 1)$$

式中：

t ——CTP扫描时间；

$S(t)$ ——信号强度—时间曲线；

S_0 ——CTP基线信号强度；
 k ——线性乘子，常用取值为100；
 $C(t)$ ——对比剂浓度—时间曲线。

由于脑组织的对比剂浓度—时间曲线和供血动脉的对比剂浓度—时间曲线满足公式(B. 2)，

$$C_b(t) = C_a(t) \otimes H(t) \cdots \cdots \cdots (B. 2)$$

式中：

$C_b(t)$ ——脑组织的对比剂浓度—时间曲线；
 $C_a(t)$ ——供血动脉的对比剂浓度—时间曲线。

其中 $\otimes$ 为卷积运算， $H(t)$ 可表示为：

$$H(t) = P_{CBF} \cdot R(t) \cdots \cdots \cdots (B. 3)$$

式中：

P_{CBF} ——灌注参数CBF（脑血流量）；
 $R(t)$ ——残差函数。

公式(B. 2)的积分形式如公式(B. 4)所示：

$$C_b(t) = \int_0^t C_a(\tau) H(t - \tau) d\tau \cdots \cdots \cdots (B. 4)$$

公式(B. 4)可被离散化为公式(B. 5)：

$$C_b(t_j) = \Delta t \cdot \sum_{i=0}^j C_a(t_i) \cdot H(t_j - t_i) \cdots \cdots \cdots (B. 5)$$

或，

$$\begin{bmatrix} C_b(t_0) \\ C_b(t_1) \\ \vdots \\ C_b(t_{T-1}) \end{bmatrix}_{T \times 1} = \Delta t \cdot \begin{bmatrix} C_a(t_0) & 0 & \cdots & 0 \\ C_a(t_1) & C_a(t_0) & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ C_a(t_{T-1}) & C_a(t_{T-2}) & \cdots & C_a(t_0) \end{bmatrix}_{T \times T} \cdot \begin{bmatrix} H(t_0) \\ H(t_1) \\ \vdots \\ H(t_{T-1}) \end{bmatrix}_{T \times 1} \cdots \cdots \cdots (B. 6)$$

式中：

T ——时间点个数；
 Δt ——时间分辨率。

公式(B. 6)可被简化为：

$$c = A \cdot b \cdots \cdots \cdots (B. 7)$$

式中：

c ——从CTP中获得的 $C_b(t)$ ；
 A ——从CTP中获得的 $C_a(t)$ 填充的矩阵乘以 Δt ；
 b —— $H(t)$ 。

从公式(B. 7)中解算得到的 b 即为 $H(t)$ 。

基于SVD（奇异值分解）算法，根据公式(B. 8)表示矩阵 A ：

$$A = U \cdot S \cdot V^T \cdots \cdots \cdots (B. 8)$$

式中：

U ——正交矩阵；
 S ——对角矩阵；
 V ——正交矩阵。

根据公式(B. 9)表示矩阵 A^+ （矩阵 A 的伪逆）：

$$A^+ = V \cdot S^+ \cdot U^T \cdots \cdots \cdots (B. 9)$$

式中：

S^+ ——矩阵 S 的伪逆。

根据公式(B. 10)表示 $H(t)$ ：

$$H(t) = b = A^+ \cdot c = V \cdot S^+ \cdot U^T \cdot c \cdots \cdots \cdots (B. 10)$$

而灌注参数的计算可通过 $H(t)$ 表示为如下公式：

$$P_{CBF} = \max(H(t)) \dots\dots\dots (B. 11)$$

$$P_{CBV} = \int_0^t H(t) dt \dots\dots\dots (B. 12)$$

$$P_{MTT} = \frac{P_{CBV}}{P_{CBF}} \dots\dots\dots (B. 13)$$

$$P_{Tmax} = \operatorname{argmax}_t(H(t)) \dots\dots\dots (B. 14)$$

式中：

P_{CBV} ——灌注参数CBV（脑血容量）；

P_{MTT} ——平均通过时间；

P_{Tmax} ——残差函数达峰时间。

B.8 脑灌注数字体模的构建

数字体模是一个反应CTP动态变化的数字模型，基于药代动力学模型，通过设置血流相关参数和实验缺陷，模拟不同情况下的CTP图像。具体方法参见参考文献[12][13][14][15]。数字体模的参数取值范围由制造商选定并记录，如参考文献[16]所述。表B.8提供了各参数的参考范围。

表B.8 脑灌注数字体模参数参考取值

参数类别	参数名/(单位)	参数取值范围	
		常见取值范围	扩展取值范围
P_Z	T	30~60	
	$\Delta t/(s)$	1~3	
P_A	$t_0/(s)$	5~10或 $T/6 \sim T/3$	0.1~5或20~30
P_P	$P_{CBV}/(\frac{mL}{100g})$	0.001~5.5	5.5~10
	$P_{MTT}/(s)$	3.4~24	
	$P_{delay}/(s)$	-1~3	3~6
	$P_{CBF}/(\frac{mL}{100g}/minute)$	由 P_{CBV} 和 P_{MTT} 决定	
P_S	k	100	
	$S_{0T}/(HU)$	32~38	
	$S_{0A}/(HU)$	40	
P_N	n	50~230	30~50

注：表中参数含义如下：

P_Z ——数字体模数据集中，信号强度—时间曲线的整体参数，包括：

T ——曲线的时间点个数；

Δt ——曲线的时间分辨率。

P_A ——动脉输入函数（AIF，具体构建方式由制造商选定）曲线参数，其中：

t_0 ——对比剂到达时间。

P_P ——仿真的灌注参数，其中：

P_{CBV} ——脑血容量(CBV)；

P_{MTT} ——平均通过时间(MTT)；

P_{delay} ——对比剂到达延迟时间(delay)；

P_{CBF} ——脑血流量(CBF)。

P_S ——对比剂浓度—时间曲线转换信号强度—时间曲线的参数，其中：

k ——常数乘子；

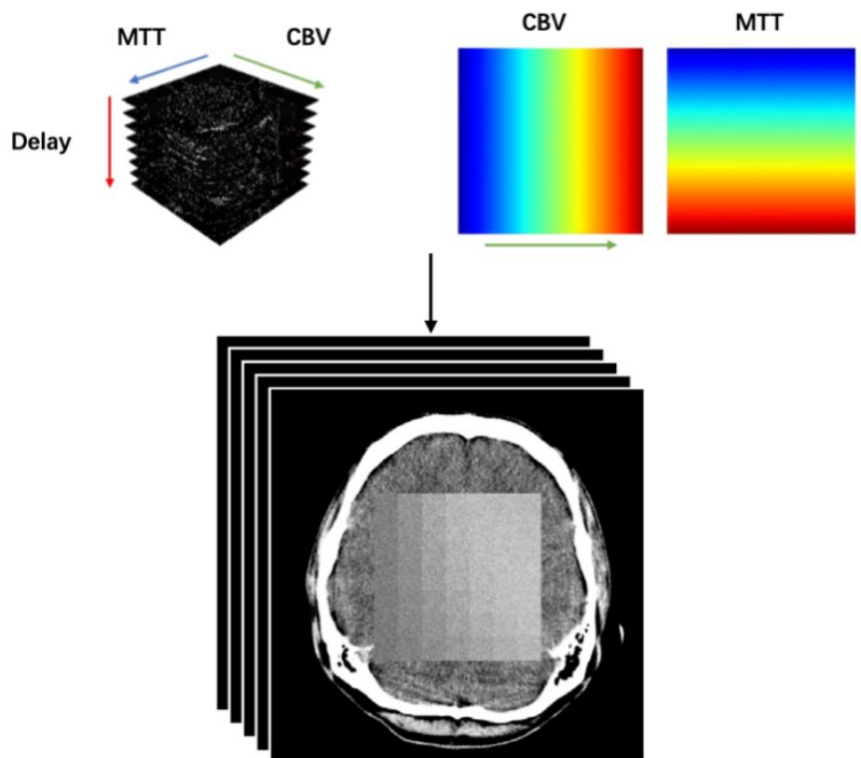
S_{0T} ——脑组织初始时刻的信号强度；

S_{0A} ——动脉初始时刻的信号强度。

P_N ——仿真的噪声参数，其中：

n ——信噪比调控因子，噪声标准差 $SD = S_{0T}/n$

数字体模与影像的嵌入构建方式，由制造商制定。一种推荐方式为：在 512×512 的CTP图像中，平面维度为 192×192 的数字体模被嵌入，并且在三维空间的不同x、y、z轴分别实现 P_{CBV} 、 P_{MTT} 、 P_{delay} 的梯度变化，如图B.1所示。AIF动脉点位于第一层图像左上角的 10×10 的区域，灌注参数梯度变化表现在每一层坐标点(160, 160)到(351, 351)的方形区域。

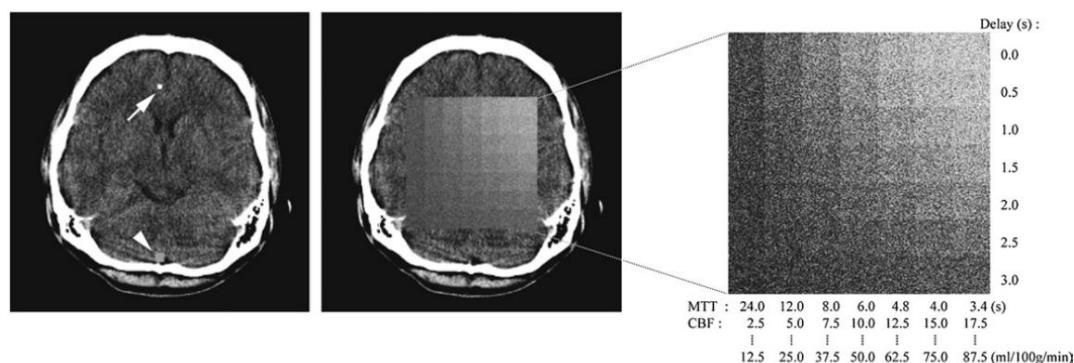


图B.1 脑灌注数字体模示意图

B.9 数字体模用于CTP灌注参数验证的方法

以参考文献[14]所公开的数字体模 (<https://amrc.iwate-med.ac.jp/en/project-2/download/>) 为例说明验证过程和方法。

该CT灌注成像数字模拟数据集包含了16个切面位置(如图B.2所示)。第1个切面包含了动脉输入函数和静脉输出函数，如图B.2最左图箭头所示。而组织曲线则嵌入在第2至第16个切面中。切面的矩阵尺寸为 512×512 ，共30期480层。每个组织切面都包含 7×7 个网格切片，每个网格切片尺寸为 32×32 。每个切片在列(从左到右)上编码了逐渐增加的平均通过时间(MTT)，也即编码了逐渐减少的脑血流量(CBF)，在行(从上到下)上则编码了逐渐增加的时间延迟。每个网格内的灌注参数(MTT和CBF)是相同的。



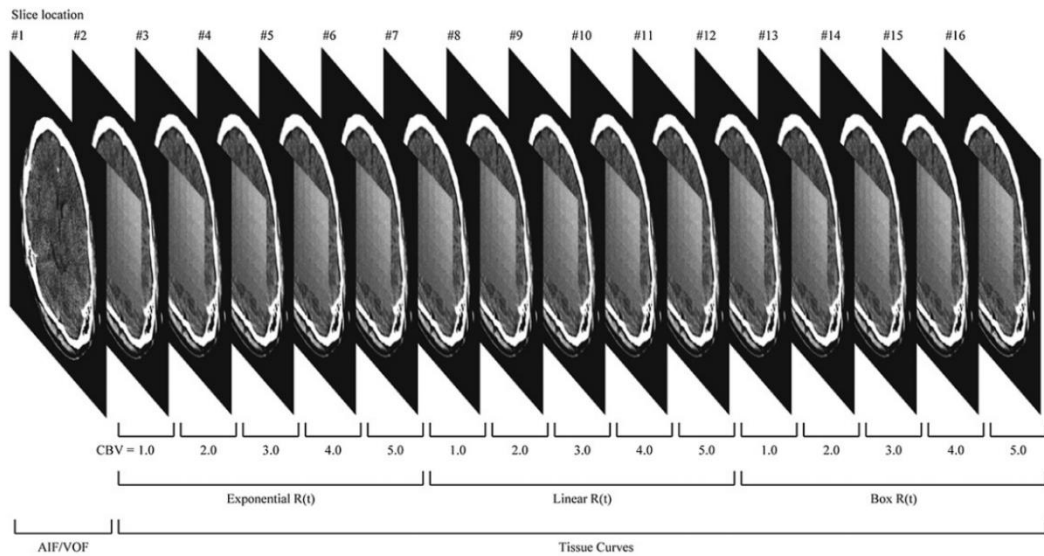
图B.2 数字体模参数示意图

为了模拟各种组织灌注曲线，该数字体模使用了七个平均通过时间(MTT)值(3.4、4.0、4.8、6.0、8.0、12.0和24.0秒)和五个脑血容量(CBV)值(1.0、2.0、3.0、4.0和5.0 mL/100g)。通过使用

心体积原理 ($CBF = CBV/MTT$) (参考文献[17])，从每对CBV和MTT值中确定了脑血流量 (CBF) 的值。例如，当CBV为1.0 mL/100 g时，CBF值分别为2.5、5.0、7.5、10.0、12.5、15.0和17.5 mL/100g/minute；而当CBV为5.0 mL/100 g时，CBF值分别为12.5、25.0、37.5、50.0、62.5、75.0和87.5 mL/100g/minute。

通过相对于动脉输入函数移动组织浓度-时间曲线 $C(t)$ ，实现了七种不同的示踪剂延迟(Delay)(0、0.5、1.0、1.5、2.0、2.5和3.0秒)。残留函数($R(t)$)则使用三种不同类型的曲线，分别是Exponential、Linear、Box。由此，从三种类型的残留函数、七个MTT值、五个CBV值和七个延迟中，共生成了735种不同的组织曲线组合，如图B.3所示。735个网格切片各自代表了一种特定的CBF、CBV、MTT和延迟的组合。对于T_{max}，数字体模中以延迟加MTT/2作为计算标准。同时，为模拟实际扫描噪声，在各切片中加入了随机高斯噪声，信噪比为5。

这种数据结构设计允许研究者通过模拟不同的脑血容量 (CBV)、平均通过时间 (MTT) 和示踪剂延迟，来全面评估CT灌注成像技术的准确性和性能，从而可以系统地研究这些参数对图像质量和灌注参数估计的影响。



图B.3 数字体模同期不同切片示意图

算法验证流程如下：

a) 获得动脉输入函数 ($C_a(t)$)

动脉点为图 B.2 左图中上部箭头所指的高亮点。其空间坐标为 $x=252$ 到 259 ， $y=124$ 到 131 ， $z=1+16i$ (其中 i 为整数，代表 CTP 扫描的期数，满足 $0 \leq i \leq 29$)。

从 CTP 图像中分别提取对应层的动脉点范围内像素的信号强度值，具体为 CTP 图像中的 $[252:259, 124:131]$ 。随后，我们计算这些区域的信号强度的均值，将其作为该期的动脉信号强度值。通过连续提取各期的动脉信号强度值，获得 $C_a(t)$ 曲线。

b) 获得组织信号强度-时间曲线 ($C_b(t)$)

组织曲线在图 B.2 右侧嵌入 CTP 切片中的网格块中获取。每个单层切片用于展示不同 Delay (纵轴) 和 CBF (横轴) 条件下的组织，不同层的切片用于展示不同 CBV 和 $R(t)$ 条件下的组织 (如图 B.3 所示)。其空间坐标为 $x=144$ 到 368 ， $y=144$ 到 368 ， $z=j+16i$ (其中 i, j 为整数，满足 $0 \leq i \leq 29$ ， $2 \leq j \leq 16$)。每张单层切片有包含 7×7 个不同的灌注参数，纵轴上有 7 个不同的 Delay，横轴上有 7 个不同的 CBF。每个相同的灌注参数由 32×32 个像素的网格矩形块表示，但 32×32 网格块内各点的组织信号强度-时间曲线是不相同的，因为数据中叠加了随机高斯噪声。

c) 计算灌注参数

输入动脉输入函数 $C_a(t)$ 与组织的信号强度-时间曲线 $C_b(t)$ 到灌注参数计算算法中，计算该组织的核心灌注参数 (CBF、CBV、MTT、T_{max})。

d) 评价灌注参数

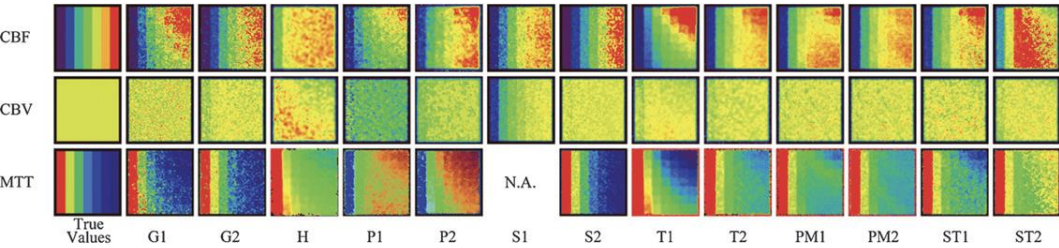
计算算法输出的灌注参数与数字体模所设定的灌注参数金标准间的绝对误差或相对误差、绝对误差或相对误差的绝对值以及它们的平均值和标准差、Pearson 相关系数、线性回归的斜率和截

距、Bland-Altman 图等性能指标，评估算法输出的参数值与真实值间的关系，如图 B.4 所示。通过可视化灌注参数彩图，直观评价算法性能，如图 B.5 所示。

Table 3									
Correlation and Linear Regression with True Value for CT Perfusion									
Software	Correlation Coefficient, <i>r</i>			Slope			Intercept		
	CBF	CBV	MTT	CBF	CBV	MTT	CBF	CBV	MTT
G1	0.930*	0.987*	0.882	0.858 (0.761, 0.956)	0.858 (0.847, 0.882)	0.560 (0.471, 0.650)	5.694 (4.512, 6.876)	0.085 (0.059, 0.117)	2.248 (1.260, 3.237)
G2	0.884	0.979*	0.867	0.601 (0.475, 0.727)	0.887 (0.877, 0.956)	0.578 (0.470, 0.685)	4.623 (3.252, 5.993)	0.111 (0.052, 0.133)	4.510 (3.254, 5.765)
H	0.841	0.964*	0.921*	1.033 (0.929, 1.137)*	2.290 (2.211, 2.414)	0.171 (0.153, 0.190)	63.678 (44.275, 83.081)	-0.046 (-0.072, 0.075)*	2.555 (2.310, 2.800)
P1	0.881	0.920*	0.849	0.982 (0.896, 1.069)*	2.266 (2.189, 2.525)	0.567 (0.465, 0.670)	18.619 (12.784, 24.454)	-0.080 (-0.288, 0.109)*	5.115 (4.188, 6.043)
P2	0.903*	0.966*	0.896	0.383 (0.364, 0.403)	0.849 (0.834, 0.879)	0.626 (0.556, 0.696)	7.971 (5.755, 10.188)	0.722 (0.700, 0.818)	6.634 (5.437, 7.831)
S1	0.938*	0.790	NA	1.227 (1.115, 1.338)	1.660 (1.606, 2.099)	NA	10.994 (6.164, 15.825)	-0.533 (-0.590, -0.297)	NA
S2	0.869	0.949*	0.870	1.745 (1.376, 2.113)	2.197 (2.131, 2.423)	0.651 (0.503, 0.799)	4.164 (1.675, 6.653)	0.001 (-0.008, 0.066)*	4.106 (2.418, 5.794)
T1	0.863	0.957*	0.922*	0.648 (0.569, 0.727)	1.018 (0.994, 1.113)*	0.677 (0.650, 0.704)	2.524 (1.381, 3.668)	-0.397 (-0.411, -0.359)	3.930 (3.167, 4.693)
T2	0.921*	0.974*	0.947*	0.491 (0.455, 0.528)	1.018 (0.997, 1.079)*	0.537 (0.496, 0.578)	7.629 (4.858, 10.401)	-0.331 (-0.346, -0.277)	4.091 (3.376, 4.807)
PM1	0.913*	0.983*	0.886	0.796 (0.734, 0.858)	0.914 (0.894, 0.953)	0.550 (0.433, 0.666)	14.256 (9.907, 18.604)	-0.240 (-0.247, -0.207)	4.850 (3.948, 5.751)
PM2	0.923*	0.983*	0.907*	0.849 (0.778, 0.920)	0.914 (0.894, 0.953)	0.575 (0.466, 0.684)	13.378 (9.378, 17.379)	-0.240 (-0.247, -0.207)	4.736 (3.826, 5.646)
ST1	0.917*	0.960*	0.802	0.607 (0.568, 0.646)	1.223 (1.178, 1.319)	0.557 (0.421, 0.693)	9.600 (6.285, 12.916)	-0.529 (-0.548, -0.503)	7.596 (6.236, 8.956)
ST2	0.844	0.940*	0.673	0.318 (0.244, 0.393)	1.107 (1.060, 1.214)	0.291 (0.192, 0.389)	14.236 (8.957, 19.514)	-0.229 (-0.240, -0.171)	11.753 (10.432, 13.073)
Note.—The software abbreviations are defined in Tables 1 and 2. Numbers in parentheses for slope and intercept are 95% confidence intervals. $y = ax + b$, x = true value, y = measured value, a = slope, b = intercept. NA = not available.									
* Values represent good correlations ($r > 0.9$).									
† The interval of the slope includes 1.0.									
‡ The interval of the intercept includes zero.									

注：图中不同行代表不同的CTP灌注参数计算软件

图B. 4 CTP 灌注参数计算算法性能定量评价结果示意图



注：图中不同列代表不同的灌注参数计算软件

图B. 5 CTP 灌注参数计算算法性能可视化评价示意图

附录 C
(资料性)
头部 CT 测试集描述样例

C.1 概述

本样例以头部NCCT用于脑出血及脑缺血ASPECTS的测试集、头部或头颈部CTA用于头颈动脉狭窄的测试集、头部CTP用于脑缺血灌注分析的测试集为具体案例，给出测试集描述的举例，仅作为参考信息。

C.2 数据集适用范围

数据集适用于声称能对头部NCCT进行脑出血和脑缺血ASPECTS辅助分析的人工智能医疗器械软件产品，如脑出血辅助检出、分类、分割、测量等，或ASPECTS评分的分类和分割等。数据集还适用于声称能对头部或头颈部CTA进行头颈动脉狭窄程度辅助分析的人工智能医疗器械软件产品，如狭窄的辅助检出、分类等。数据集还适用于声称能对头部CTP进行灌注状态分析的人工智能医疗器械软件产品。

C.3 数据采集

数据采集考虑患者人群、采集设备、数据格式、扫描部位等方面的多样性，具有合规性证明，如伦理审批。表C.1给出了数据来源的多样性统计举例，可进一步细化。

表C.1 数据来源的多样性统计

统计维度	范围	病例数量	适用数据集
年龄	18~40岁	XX例	NCCT、CTA、CTP
	40~60岁	XX例	NCCT、CTA、CTP
	60~80岁	XX例	NCCT、CTA、CTP
	大于80岁	XX例	NCCT、CTA、CTP
性别	男	XX例	NCCT、CTA、CTP
	女	XX例	NCCT、CTA、CTP
地域	华东地区	XX例	NCCT、CTA、CTP
	华南地区	XX例	NCCT、CTA、CTP
	华中地区	XX例	NCCT、CTA、CTP
	华北地区	XX例	NCCT、CTA、CTP
	西北地区	XX例	NCCT、CTA、CTP
	西南地区	XX例	NCCT、CTA、CTP
	东北地区	XX例	NCCT、CTA、CTP
CT机型	XX公司 XX型号	XX例	NCCT、CTA、CTP
	XX公司 XX型号	XX例	NCCT、CTA、CTP
管电流	小于50mAs	XX例	NCCT、CTA、CTP
	大于等于50mAs	XX例	NCCT、CTA、CTP
管电压	小于110kV	XX例	NCCT、CTA、CTP
	大于等于110kV	XX例	NCCT、CTA、CTP
层厚	3mm及以下	XX例	NCCT、CTP
	3~6mm	XX例	NCCT、CTP
	6mm及以上	XX例	NCCT、CTP
	1mm及以下	XX例	CTA
	1mm以上	XX例	CTA
扫描范围	80mm以下	XX例	NCCT、CTA、CTP
	80mm~160mm	XX例	NCCT、CTA、CTP
	160mm~320mm	XX例	NCCT、CTA、CTP
	320mm以上	XX例	CTA
扫描部位	头部加颈部	XX例	CTA
	头部	XX例	CTA
	颈部	XX例	CTA

C.4 数据集分布构成

头部NCCT/CTA/CTP数据集的分布构成描述模板如表C.2所示。其中的分类划分方式根据具体临床指南调整。

表C.2 头部 NCCT/CTA/CTP 数据集分布统计

类别	分类	病例数量	分类
脑出血	脑实质出血（包含基底节区、脑叶、脑干、小脑等）	XX例	脑实质出血（包含基底节区、脑叶、脑干、小脑等）
	蛛网膜下腔出血	XX例	蛛网膜下腔出血
	硬膜外出血	XX例	硬膜外出血
	硬膜下出血	XX例	硬膜下出血
	脑室积血	XX例	脑室积血
ASPECTS（按缺血区域）	前循环梗死	XX例	前循环梗死
	后循环梗死	XX例	后循环梗死
ASPECTS（按缺血程度）	<3分	XX例	<3分
	3~5分	XX例	3~5分
	6~10分	XX例	6~10分
动脉狭窄（患者水平）	未见狭窄	XX例	未见狭窄
	轻度狭窄	XX例	轻度狭窄
	中度狭窄	XX例	中度狭窄
	重度狭窄	XX例	重度狭窄
	闭塞	XX例	闭塞
动脉狭窄（血管水平）	颈内动脉	未见狭窄	XX例
		轻度狭窄	XX例
		中度狭窄	XX例
		重度狭窄	XX例
		闭塞	XX例
	可包含如下血管水平：椎动脉、基底动脉、大脑中动脉、大脑前动脉、大脑后动脉、前交通动脉、后交通动脉等		同上
动脉狭窄（血管分段水平）	颈内动脉C1段	未见狭窄	XX例
		轻度狭窄	XX例
		中度狭窄	XX例
		重度狭窄	XX例
		闭塞	XX例
	血管分段水平可为：颈内动脉C1~C7、椎动脉V1~V4、大脑中动脉M1~M3、大脑前动脉A1~A3、大脑后动脉P1~P3等		同上

C.5 数据集标注规则

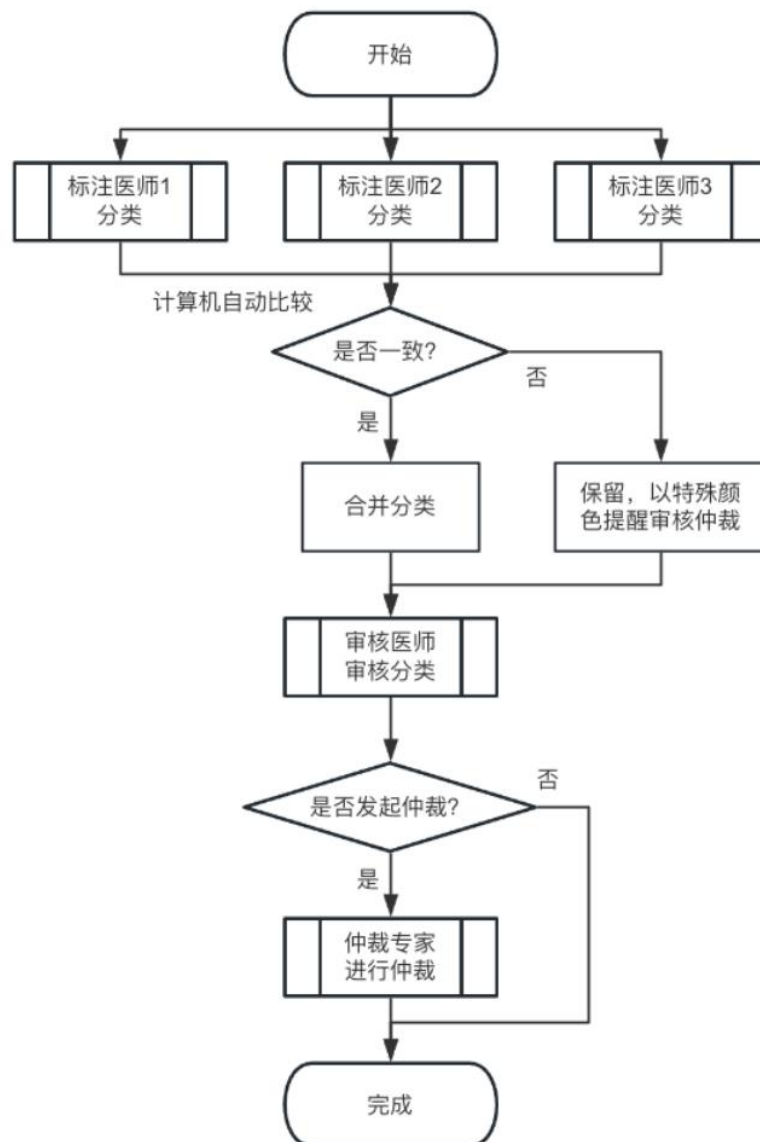
标注工作建议依据行业指南或关键参考文献进行，如参考文献[5][6][8][18][19][20]。

标注要求和标准：标注前，对标注工具、标注人员资质、标注内容等给出明确规定，并对标注方法进行充分讨论，形成共识性标注原则和方法步骤。

标注流程：为了提高标注的准确性，避免记忆偏倚，标注流程提倡多轮次分组交叉进行（见图C.1）。同时，考虑标注工作量，优化人力资源，标注工作需要标注医师、审核医师和仲裁专家三组医生参加。标注医师由头部NCCT和CTA阅片经验3年以上的医师担任，审核医师由阅片经验10年以上的副主任医师担任，仲裁专家由阅片经验15年以上的副主任医师或主任医师担任。标注工作分为三个阶段：

- a) 初标阶段：由3名（至少2名）标注医生背靠背独立标注，然后用计算机根据事先确定的判据自动判断标注的一致性，保留不同意见。

- b) 审核阶段：由 1 名审核医师对于标注结果进行审核，并对标注不一致的案例给出审核结果。如果审核医生仍然无法确定，则发起仲裁。
- c) 仲裁阶段：对于经过前两个阶段仍然无法确定的病例，由仲裁专家进行仲裁。如果遇到疑难问题，仲裁专家可以进行集体讨论和确认。



图C.1 头部 NCCT/CTA 标注流程图

C.6 样本量的估计

参考YY/T 1858-2022中附录A.6的要求对样本量进行估计计算。

C.7 测试集偏倚分析

参考YY/T 1858-2022中附录A.7的要求对测试集进行偏倚分析。

参 考 文 献

- [1] 国家药品监督管理局医疗器械技术审评中心. 深度学习辅助决策医疗器械软件审评要点[Z]. 北京:国家药品监督管理局医疗器械技术审评中心, 2019.
- [2] 国家药品监督管理局医疗器械技术审评中心. 人工智能医疗器械注册审查指导原则[Z]. 北京:国家药品监督管理局医疗器械技术审评中心, 2022.
- [3] 国家药品监督管理局医疗器械技术审评中心. 医疗器械软件注册审查指导原则[Z]. 北京:国家药品监督管理局医疗器械技术审评中心, 2022.
- [4] 王家良, 康德英. 临床流行病学——临床科研设计、测量与评价(第4版)[M]. 上海:上海科学技术出版社, 2014.
- [5] Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy[J]. Lancet. 2000;355(9216):1670-1674. doi:10.1016/S0140-6736(00)02237-6.
- [6] Puetz V, Sylaja PN, Coutts SB, et al. Extent of hypoattenuation on CT angiography source images predicts functional outcome in patients with basilar artery occlusion[J]. Stroke. 2008;39(9):2485-2490. doi:10.1161/STROKEAHA.107.511162.
- [7] 脑卒中防治工程委员会. 中国脑卒中防治指导规范[Z]. 北京: 国家卫生健康委员会, 2021.
- [8] Bouthillier A, van Loveren HR, Keller JT. Segments of the internal carotid artery: a new classification[J]. Neurosurgery. 1996;38(3):425-433. doi:10.1097/00006123-199603000-00001.
- [9] Orlický M, Sameš M, Hejčl A, Vachata P. Carotid-ophthalmic aneurysms—Our results and treatment strategy[J]. British journal of neurosurgery. 2015;29(2):237-242. doi:10.3109/02688697.2014.976176.
- [10] Arvidsson J, Starck G, Lagerstrand K, Ziegelitz D, Jalnefjord O. Effects of bolus injection duration on perfusion estimates in dynamic CT and dynamic susceptibility contrast MRI[J]. Magnetic Resonance Materials In Physics Biology And Medicine. 2023;36(1):95-106. doi:10.1007/s10334-022-01038-y.
- [11] Aycok RD, Westafer LM, Boxen JL, Majlesi N, Schoenfeld EM, Bannuru RR. Acute Kidney Injury After Computed Tomography: A Meta-analysis[J]. Annals of emergency medicine. 2018;71(1):44-53. e4. doi:10.1016/j.annemergmed.2017.06.041.
- [12] Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis[J]. Magnetic resonance in medicine. 1996;36(5):715-725. doi:10.1002/mrm.1910360510.
- [13] Wu O, Østergaard L, Weisskoff RM, Benner T, Rosen BR, Sorensen AG. Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix[J]. Magnetic resonance in medicine. 2003;50(1):164-174. doi:10.1002/mrm.10522.
- [14] Kudo K, Christensen S, Sasaki M, et al. Accuracy and reliability assessment of CT and MR perfusion analysis software using a digital phantom[J]. Radiology. 2013;267(1):201-211. doi:10.1148/radiol.12112618.
- [15] Pianykh OS. Digital perfusion phantoms for visual perfusion validation[J]. American journal of roentgenology. 2012;199(3):627-634. doi:10.2214/AJR.11.7061.
- [16] Wintermark M, Albers GW, Broderick JP, et al. Acute Stroke Imaging Research Roadmap II[J]. Stroke. 2013;44(9):2628-2639. doi:10.1161/STROKEAHA.113.002015.

- [17] Meier P, Zierler KL. On the theory of the indicator-dilution method for measurement of blood flow and volume[J]. Journal of applied physiology. 1954;6(12):731-744. doi:10.1152/jappl.1954.6.12.731.
- [18] Greenberg SM, Ziai WC, Cordonnier C, et al. 2022 Guideline for the Management of Patients With Spontaneous Intracerebral Hemorrhage: A Guideline From the American Heart Association/American Stroke Association[J]. Stroke. 2022;53(7):e282-e361. doi:10.1161/STR.0000000000000407.
- [19] 中国卒中学会. 症状性颅内动脉粥样硬化性大动脉狭窄管理规范——中国卒中学会科学声明[Z]. 北京:中国卒中学会, 2017.
- [20] 国家卫生健康委员会脑卒中防治工程委员会神经影像专业委员会, 中华医学会放射学分会神经学组. 脑血管病影像规范化应用中国指南Ⅶ[J]. 中华放射学杂志, 2019;53(11):916-940.
-