



中华人民共和国医药行业标准

XX/T XXXXX—XXXX

人工智能医疗器械 骨折 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
fracture CT images Algorithm performance testing methods

征求意见稿

XXXX - XX - XX 发布

XXXX - XX - XX 实施

发布

目 次

前 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	1
4.1 概述	1
4.2 测试环境	1
4.3 测试资源	1
4.3.1 测试集要求	1
4.3.2 数据采集要求	1
4.3.3 测试工具	2
4.4 测试平台	2
4.5 测试指标与通过准则	2
4.6 测试流程	2
4.7 测试结果	2
5 算法性能测试方法	2
5.1 算法应用场景的测试方法	2
5.1.1 骨折分割场景	3
5.1.2 骨折目标检测场景	3
5.1.3 骨折分类场景	6
5.2 算法质量特性与测试方法	7
5.2.1 泛化能力	7
5.2.2 鲁棒性	7
5.2.3 重复性	8
5.2.4 一致性	8
5.2.5 效率	8
5.2.6 算法错误统计	8
附录 A （资料性） 对相关医学概念补充说明	9
A.1 骨折	9
A.2 骨折的一般表现	9
A.3 骨折的特有特征	9
A.4 骨折的影像学检查	9
A.5 骨折分类标签	9
附录 B （资料性） 特定条款的指南和方法说明	11
B.1 通用指南	11
B.2 特定章条的指南或者方法说明	11
B.3 测试集标注要求	13
B.3.1 骨骼分割场景	13

B.3.2 骨折检出场景	13
B.4 标注流程	13
参 考 文 献	16

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：

人工智能医疗器械 骨折 CT 影像辅助分析软件 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骨折CT影像辅助分析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、对骨折CT影像进行辅助分析的医疗器械软件产品。
本文件不适用于影像前处理和过程优化。

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

骨折 fracture

骨结构的连续性完全或部分断裂。

3.2

通过准则 pass criteria

判断一个软件项或算法功能的测试是否通过的判别依据。

[来源：GB/T 9386-2008. 3.2有修改]

4 算法性能测试要求

4.1 概述

本文件宜参照YY/T 1858-2022中4.1的要求，建立测试文档。

4.2 测试环境

本文件宜参照YY/T 1858-2022中4.2的要求。

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；测试人员可根据产品预期用途和临床应用场景，对测试集进行限定。

4.3.2 数据采集要求

4.3.2.1 影像质量要求

骨折影像数据应具备以下要求：

——测试集采用的影像数据预先开展去标识化，宜参考GB/T 39725-2020条款10.2的方法，例如去除姓名、证件号、联系方式、医保账户、银行账号。

——图像分辨率不低于512×512像素；

——测试集中的影像数据满足待测产品的适用范围。测试数据应包含体现数据多样性的字段，包括但不限于考虑患者年龄、骨折类型、骨折部位、CT重建层厚、重建算法等信息；

——避免运动、金属、曝光不足、操作不当等情形带来的严重伪影；

——每个病例的图像应当保证完整、连续，不能出现缺失或断层等现象。

注：示例见附录B.2

4.3.2.2 测试集样本量要求

测试集样本量宜参照YY/T 1858-2022中4.3.2计算。

4.3.2.3 测试集多样性要求

测试数据集宜体现待测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：

——设备制造商；

——产品型号；

——成像参数（如层厚、重建内核、设备场强、序列、不同剂量条件等）。

患者维度考虑以下内容：

——性别；

——年龄；

——地区；

——医疗机构；

——来源场景（如急诊、门诊、病房、体检等）；

——骨折阶段（如新鲜、陈旧等）；

——骨折类型（如开放、闭合等）；

——骨折部位（如肱骨、股骨等）。

4.3.2.4 扩增数据要求

宜参照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

如适用，算法测试使用的体模与标准器宜参照YY/T 1858-2022中4.3.5的要求。

4.4 测试平台

如通过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。

测试计划应根据产品预期用途制定，按照产品对骨折的诊断方式（分割、分类、检测）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 骨折分割场景

5.1.1.1 骨折块重合度计算

当分割区域为一般实体（如骨折块），宜采用*Dice*系数、*Jaccard*系数评价骨折块的重合度。当对多个骨折块进行分割时（如适用），采用骨折块并集区域计算*Dice*系数和*Jaccard*系数。

*Dice*系数如公式（1）所示：

$$Dice = \frac{2*|Out \cap True|}{|Out|+|True|} \dots\dots\dots (1)$$

式中：

Dice —— *Dice*系数；

Out —— 算法分割的骨折区域；

True —— 参考标准的骨折区域。

*Jaccard*系数如公式（2）所示：

$$Jaccard = \frac{|Out \cap True|}{|Out \cup True|} \dots\dots\dots (2)$$

式中：

Jaccard —— *Jaccard*系数；

Out —— 算法分割的骨折区域；

True —— 参考标准的骨折区域。

5.1.1.2 骨折块轮廓线重合度

骨折块轮廓线是指骨折块的边缘轮廓线。如适用，宜采用Hausdorff距离，倒角距离进行重合度计算。

Hausdorff距离的计算方法宜参照YY/T 1858-2022中5.1.2.6。

倒角距离的计算公式如公式（3）所示：

$$D_{charf}(T, I) = \frac{1}{|T|} \sum_{t \in T} d_t(t) \dots\dots\dots (3)$$

式中：

$D_{charf}(T, I)$ —— 倒角距离；

I —— 算法输出骨折块的边缘轮廓线；

T —— 参考标准中骨折块的边缘轮廓线；

$d_t(t)$ —— *T*中任意点 *t*在 *I*上的最短距离。

5.1.1.3 测试步骤

在骨折块分割场景下，算法测试按照如下步骤进行：

- a) 向待测分割算法输入测试集，输出算法分割结果；算法输出结果格式应与参考标准兼容，内容应指明骨折块区域；
- b) 算法分割的骨折块区域与参考标准的骨折块区域的性能指标按照5.1.1.1~5.1.1.2中的性能指标进行计算；
- c) 以计算结果的平均值作为最终结果。

5.1.2 骨折目标检测场景

5.1.2.1 骨折目标二维检测标注与匹配

如待测算法利用二维标记框检测骨折部位，导出算法预测骨折部位的标记框，测试数据应包含对应的参考标准。测试人员应记录算法标记的检测框与参考标准中的检测框的重叠程度。利用交并比计算算法预测标记框与参考标准标记框的重叠度，

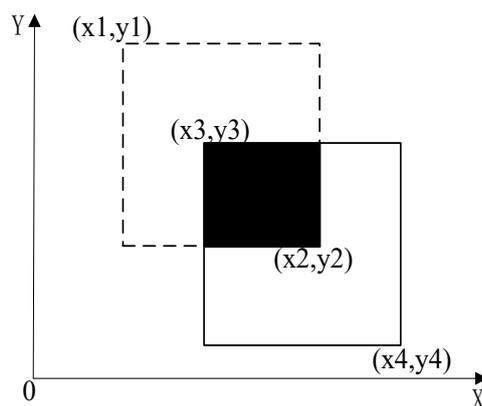


图1 二维检测框示意图

如图1所示，虚线框为算法计算出的标记框，左上角和右下角的坐标分别为 (x_1, y_1) ， (x_2, y_2) 。实线框为参考标准中的标记框，左上角和右下角的坐标分别为 (x_3, y_3) ， (x_4, y_4) 。则算法预测标记框的相关评价参数计算方式如下所示：

$$Intersect = (\min(x_2, x_4) - \max(x_1, x_3)) * (\min(y_2, y_4) - \max(y_1, y_3)) \dots \dots \dots (4)$$

$$Union = (|x_2 - x_1| * |y_2 - y_1|) + (|x_4 - x_3| * |y_4 - y_3|) - Intersect \dots \dots \dots (5)$$

$$IOU = \frac{Intersect}{Union} \dots \dots \dots (6)$$

式中：

$\max(a, b)$ —— a, b 中的最大值；

$\min(a, b)$ —— a, b 中的最小值；

$Intersect$ —— 算法预测标记框与参考标准标记框的重叠部分（图1中的黑色部分）；

$Union$ —— 算法预测标记框与参考标准标记框的并集；

IOU —— 交并比。

借助IOU与置信度（Confidence）对算法预测标记框和参考标准标记框进行重叠程度分析。重叠情况分为三种情况：

- a) 真阳性，正确检测骨折区域，总数记为TP（True Positives）；
- b) 假阳性，未匹配参考目标的算法标记目标，总数记为FP（False Positives）；
- c) 假阴性，未匹配算法标记目标的参考目标，总数记为FN（False Negatives）。

5.1.2.2 骨折目标三维检测标注与匹配

如待测算法利用三维包围盒检测骨折部位，导出算法预测骨折部位的包围盒，同时测试数据应包含对应的参考标准。测试人员应记录算法标记的包围盒与参考标准中包围盒的重叠程度。利用交并比计算算法预测包围盒与参考标准包围盒的重叠度。

如图2所示，假设虚线框为算法计算出的包围盒，左前下角和右后上角的坐标分别为 (x_1, y_1, z_1) ， (x_2, y_2, z_2) 。实线框为参考标准中的标记框，左前下角和右后上角的坐标分别为 (x_3, y_3, z_3) ， (x_4, y_4, z_4) 。则算法预测包围盒的相关评价参数计算方式如下所示

$$\Delta x = (\min(x_2, x_4) - \max(x_1, x_3)) \dots \dots \dots (7)$$

$$\Delta y = (\min(y_2, y_4) - \max(y_1, y_3)) \dots \dots \dots (8)$$

$$\Delta z = (\min(z_2, z_4) - \max(z_1, z_3)) \dots \dots \dots (9)$$

$$Intersect = \Delta x * \Delta y * \Delta z \dots\dots\dots (10)$$

$$Pre = (|x2 - x1| * |y2 - y1| * |z2 - z1|) \dots\dots\dots (11)$$

$$gt = (|x4 - x3| * |y4 - y3| * |z4 - z3|) \dots\dots\dots (12)$$

$$Union = Pre + gt - Intersect \dots\dots\dots (13)$$

$$IOU = \frac{Intersect}{Union} \dots\dots\dots (14)$$

式中：

- $max(a, b)$ —— a, b 中的最大值；
- $min(a, b)$ —— a, b 中的最小值；
- Δx —— 算法预测包围盒与参考标准包围盒的重叠部分 x 轴向长度；
- Δy —— 算法预测包围盒与参考标准包围盒的重叠部分 y 轴向长度；
- Δz —— 算法预测包围盒与参考标准包围盒的重叠部分 z 轴向长度；
- $Intersect$ —— 算法预测包围盒与参考标准包围盒的重叠部分；
- Pre —— 算法预测包围盒体积
- gt —— 参考标准包围盒体积；
- $Union$ —— 算法预测包围盒与参考标准包围盒的并集；
- IOU —— 交并比。

借助 IOU 与置信度 (Confidence) 对算法预测包围盒和参考标准包围盒进行重叠程度分析。重叠情况分为三种情况：

- d) 真阳性，正确检测骨折区域，总数记为 TP；
- e) 假阳性，未匹配参考目标的算法标记目标，总数记为 FP；
- f) 假阴性，未匹配算法标记目标的参考目标，总数记为 FN。

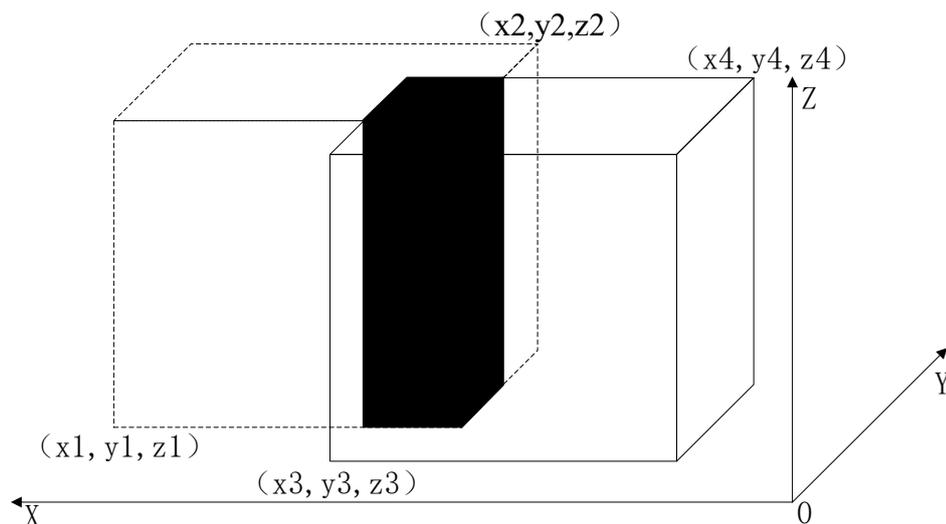


图2 三维包围盒示意图

5.1.2.3 精确度

按照YY/T 1858-2022中5.1.1.4要求进行计算。

5.1.2.4 召回率

按照YY/T 1858-2022中5.1.1.3要求进行计算。

5.1.2.5 F1 分数

按照YY/T 1858-2022中5.1.1.5要求进行计算。

5.1.2.6 平均精确度

按照YY/T 1858-2022中的5.1.1.6要求进行计算。

5.1.2.7 平均精确度均值

按照YY/T 1858-2022中的5.1.1.7要求进行计算。

5.1.2.8 非病变定位率

非病变定位率公式如公式（7）所示：

$$AveFP = \frac{N_{FP}}{N_{Case}} * 100\% \dots\dots\dots (15)$$

式中：

- $AveFP$ —— 非病变定位率；
 N_{FP} —— 假阳性总数；
 N_{Case} —— 参与测试的用例总数。

5.1.2.9 自由响应受试者操作特征曲线

特征曲线计算方式宜参照YY/T 1858-2022中5.1.1.8，横坐标为非病变定位率，取值一般为等比数列，即0.5, 1, 2, …, n。其中n应大于病例个体的平均病灶数量，以CT肋骨骨折检测为例，假设单个病例平均有5个骨折，则n的取值不低于5，非病变定位率详见5.1.2.8。

5.1.2.10 代替自由响应受试者操作特征曲线

改变算法阈值设置，计算各个阈值对应的召回率和病例层面的特异性。以召回率为纵坐标，病例层面的特异性为横坐标，构造的曲线为替代自由响应接收者操作特征曲线。通过曲线下面积，综合反映检出性能。曲线下面积越大，算法检出性能越佳。

5.1.2.11 测试步骤

宜参照YY/T 1858-2022中5.1.1.2测试步骤，对算法输出的结果按照5.1.2.1和5.1.2.2中的原则进行匹配度计算，得到TP、FP、FN、*Intersect*、*Union*，按照5.1.2.3~5.1.2.10计算各性能指标。

- a) 向待测算法输入测试集，输出算法检出的结果；该结果的格式宜与参考标准兼容，内容至少包括目标位置（标注框坐标）、检出概率；
- b) 输出结果按照 5.1.2.1（二维检测适用）与 5.1.2.2（三维检测适用）中的原则与参考标准进行匹配，计算 TP、FP 和 FN；
- c) 按照 5.1.2.3~5.1.2.10 描述的公式计算各性能指标，若是多处骨折，计算每一处骨折的性能指标，取平均值作为该测试用例的结果；若是单一骨折，计算对应部位的性能指标作为该测试用例的结果；
- d) 取所有测试用例结果的平均值作为最终结果。

5.1.3 骨折分类场景

5.1.3.1 灵敏度

按照YY/T 1858-2022中5.1.3.2要求进行计算。

5.1.3.2 特异度

按照YY/T 1858-2022中5.1.3.3要求进行计算。

5.1.3.3 阳性预测值

按照YY/T 1858-2022中5.1.3.5要求进行计算。

5.1.3.4 阴性预测值

按照YY/T 1858-2022中5.1.3.6要求进行计算。

5.1.3.5 准确率

按照YY/T 1858-2022中5.1.3.7要求进行计算。

5.1.3.6 Kappa 检验

按照YY/T 1858-2022中5.1.3.9要求进行计算。

5.1.3.7 受试者操作特征曲线

按照YY/T 1858-2022中5.1.3.10要求进行计算。

5.1.3.8 测试步骤

在骨折分类场景下，算法测试执行以下步骤：

- a) 向待测算法输入测试集，输出算法分类的结果；格式与参考标准兼容，内容包含分类结果、分类概率(如适用)；
- b) 比较算法分类与参考标准分类，计算真阳性、假阳性、真阴性、假阴性样本的数量，构造混淆矩阵；
- c) 按照 5.1.3.1~5.1.3.7 计算各性能指标，以计算结果的平均值作为最终结果。

注：常用标签见附录A.5

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中的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，进行分层分析。子集一般体现在患者个体差异，设备成像差异，骨折位置，骨折类型等方面，例如：

- 子集按照不同年龄段进行拆分；
- 子集按照性别进行拆分；
- 对某一部位骨折进行检测，子集按照该部位骨折类型进行拆分；
- 对多个部位进行检测（如软件适用），子集应按照不同部位进行拆分；
- 子集按照不同成像设备制造商进行拆分；
- 子集按照不同医院进行拆分。

5.2.2 鲁棒性

5.2.2.1 通用要求

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，收集真实世界数据或产生仿真数据，组成专用测试集，开展对抗测试、压力测试，形成鲁棒性研究资料。

5.2.2.2 对抗测试

对抗测试考虑CT设备差异和数据采集条件差异导致的图像变化，差异性包括：物理分辨率、像素分辨率、亮度、调焦、射线质量等。可使用模拟仿真产生的数据，在人工确认后用于测试。

测试人员宜考虑软件前处理的多样性，收集或模拟生成更多的图像数据，作为测试集的扩充，验证算法面对软件前处理的鲁棒性，软件前处理的多样性包括：背景裁切、图像压缩、背景填充、平滑预处理等。模拟生成的图像数据不应影响标注结论。

在不影响医师阅片的前提下，对抗测试可在影像上叠加轻微或局部的噪声、伪影。例如：

- 运动伪影；
- 金属伪影；
- 环状/晕状/条纹伪影；
- 错层伪影。

5.2.2.3 压力测试

压力测试宜引入真实临床数据，比如常见干扰项，罕见样本，复杂样本（但仍符合数据质量要求）。注意压力样本不应影响医师判断。压力样本的选取可考虑，但不限于以下特征：

- 有植入物（干扰项）的影像；
- 特定疾病的影像；
- 有伪影但满足数据质量要求的影像；
- 影像的层厚极大或者极小；
- 影像序列包含的图像数量极大；
- 多发骨折。

5.2.3 重复性

如适用，按照YY/T 1858-2022中5.2.3要求进行计算。

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：中间结论是算法运行的过程产物，为辅助决策提供线索；最终结论是算法辅助决策的结果。

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。

宜以测试数据导入时刻作为起点，以算法允许结束的时刻作为终点。

5.2.6 算法错误统计

宜参照YY/T 1858-2022中5.2.6，对算法错误测试结果进行统计，纳入测试结果，如：

- 骨折检出的假阴，假阳；
- 对骨折类型的误判；
- 骨骼分割误差；
- 骨骼标签错误等。

附录 A (资料性) 对相关医学概念补充说明

A.1 骨折

骨折即骨的完整性连续性中断。骨折可能由创伤和骨骼疾病所致，后者如骨髓炎、骨肿瘤所致骨质破坏，受轻微外力即发生的骨折，称为病理性骨折。

骨折形成的因素主要有以下类别：

- a) 直接暴力。暴力直接作用使受伤部位发生骨折，常伴有不同程度的软组织损伤。
- b) 间接暴力。暴力通过传导、杠杆、旋转和肌肉收缩使肢体受力部位的远处发生骨折。
- c) 疲劳性骨折。长期、反复、轻微的直接或间接外力可致使肢体某一特定部位骨折，也称为应力性骨折。

A.2 骨折的一般表现

骨折后一般表现为局部疼痛、肿胀和功能障碍。骨折时，骨髓、骨膜以及周围组织血管破裂出血，在骨折处形成血肿，以及软组织损伤所致水肿，使患肢严重肿胀，甚至出现张力性水泡和皮下瘀斑，由于血红蛋白的分解，可呈现紫色、青色或黄色。骨折局部出现剧烈疼痛，特别是移动患肢时加剧，伴明显压痛。局部肿胀或疼痛使患肢活动受限，若为完全性骨折，可使受伤肢体活动功能完全丧失。

A.3 骨折的特有特征

骨折后主要表现为以下特有特征：

- a) 畸形：骨折端移位可使患肢外形发生改变，主要表现为缩短、成角或旋转畸形。
- b) 异常活动：正常情况下肢体不能活动的部位，骨折后出现不正常的活动。
- c) 骨擦音或骨擦感：骨折后，两骨折端相互摩擦时，可产生骨擦音或骨擦感。

具有以上三个骨折特有体征之一者，即可诊断为骨折。值得注意的是，有些骨折如裂缝骨折、嵌插骨折、脊柱骨折或骨盆骨折，没有上述三个典型的骨折特有体征，应常规进行X线平片检查，必要时行CT或MRI检查，以便确诊。

A.4 骨折的影像学检查

骨折的影像学检查主要包括X线检查，CT检查和磁共振（Magnetic Resonance Imaging, MRI）检查。

X线检查对骨折的诊断和治疗具有重要价值。X线检查可以显示临床上难以发现的不完全性骨折、深部的骨折、关节内骨折和小的撕脱性骨折等。骨折的X线检查一般应拍摄包括邻近一个关节在内的正、侧位片，必要时拍摄特殊位置的X线平片。

对早期、非典型病例及复杂的解剖部位，CT能弥补X线检查的不足。对于骨和关节复杂解剖部位或常规X线难以检查的部位，CT能提供更多的诊断信息。

MRI是一种生物磁自旋成像技术，利用人体中遍布全身的氢原子在外加的强磁场内受到射频脉冲的激发，产生磁共振现象，用探测器检测并接收以电磁波形式放出的磁共振信号，经过空间编码和数据处理转换，将人体各组织形成图像。MRI所获得的图像异常清晰、精细，分辨率高，对比度好，信息量大，特别对软组织层次显示较好，可以发现隐匿性骨折并确定骨挫伤的范围。

A.5 骨折分类标签

常见骨折分类标签为：

- a) 裂缝骨折：骨质出现裂隙，无移位。
- b) 青枝骨折：多见于儿童，骨质和骨膜部分断裂，可有成角畸形。有时成角畸形不明显仅表现为骨皮质劈裂，与青嫩树枝被折断时相似而得名。
- c) 横形骨折：骨折线与骨干纵轴接近垂直。
- d) 斜形骨折：骨折线与骨干纵轴呈一定角度。

- e) 螺旋形骨折：骨折线呈螺旋状。
- f) 粉碎性骨折：骨质碎裂成三块以上。
- g) 嵌插骨折：骨折块相互嵌插，多见于干骺端骨折。即骨干的密质骨嵌插入干骺端的松质骨内。
- h) 压缩性骨折：骨质因压缩而变形，多见于松质骨。
- i) 骨骺损伤：经过骨骺的骨折，骨骺的断面可带有数量不等的骨组织。
- j) 其他骨折。

附录 B
(资料性)
特定条款的指南和方法说明

B.1 通用指南

对于人工智能医疗器械 骨折CT影像辅助分析软件 算法性能测试方法，测试要求贯穿整个测试过程，第4章对算法性能测试要求进行详细阐述，第5章对算法性能测试方法进行了描述。

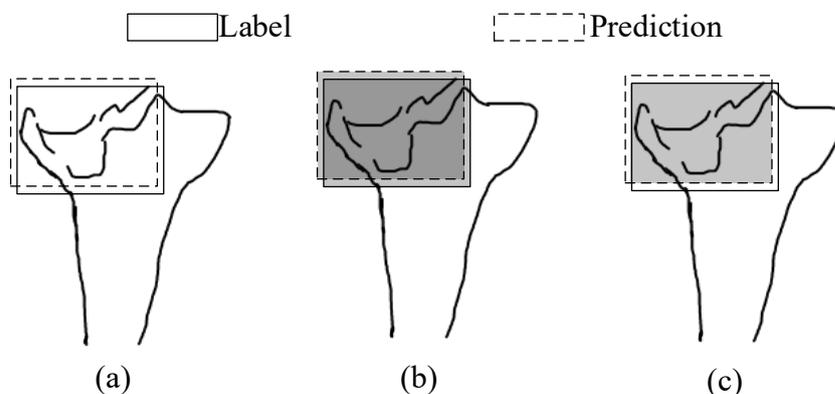
B.2 特定条款的指南或者方法说明

以下内容与正文的条款相对应，提供方法或举例。

对于4.3.2中的数据采集要求，考虑成像设备和成像过程中技术参数要求，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) 通常成人CT扫描的管电压在70-120 kV之间，具体数值需根据患者体型、部位及骨骼密度调整。较厚或密度高的部位可能需要较高的管电压以保证足够的穿透力。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。
- c) 管电流应根据管电压调整，以确保足够的图像亮度而不至于增加不必要的辐射剂量。自动曝光控制功能可以根据受检部位的厚度自动调节管电流，以达到最佳图像质量；
- d) 通过调整技术参数和使用高质量的影像接收器，确保骨折线清晰可见，同时维持软组织与骨骼之间的良好对比；

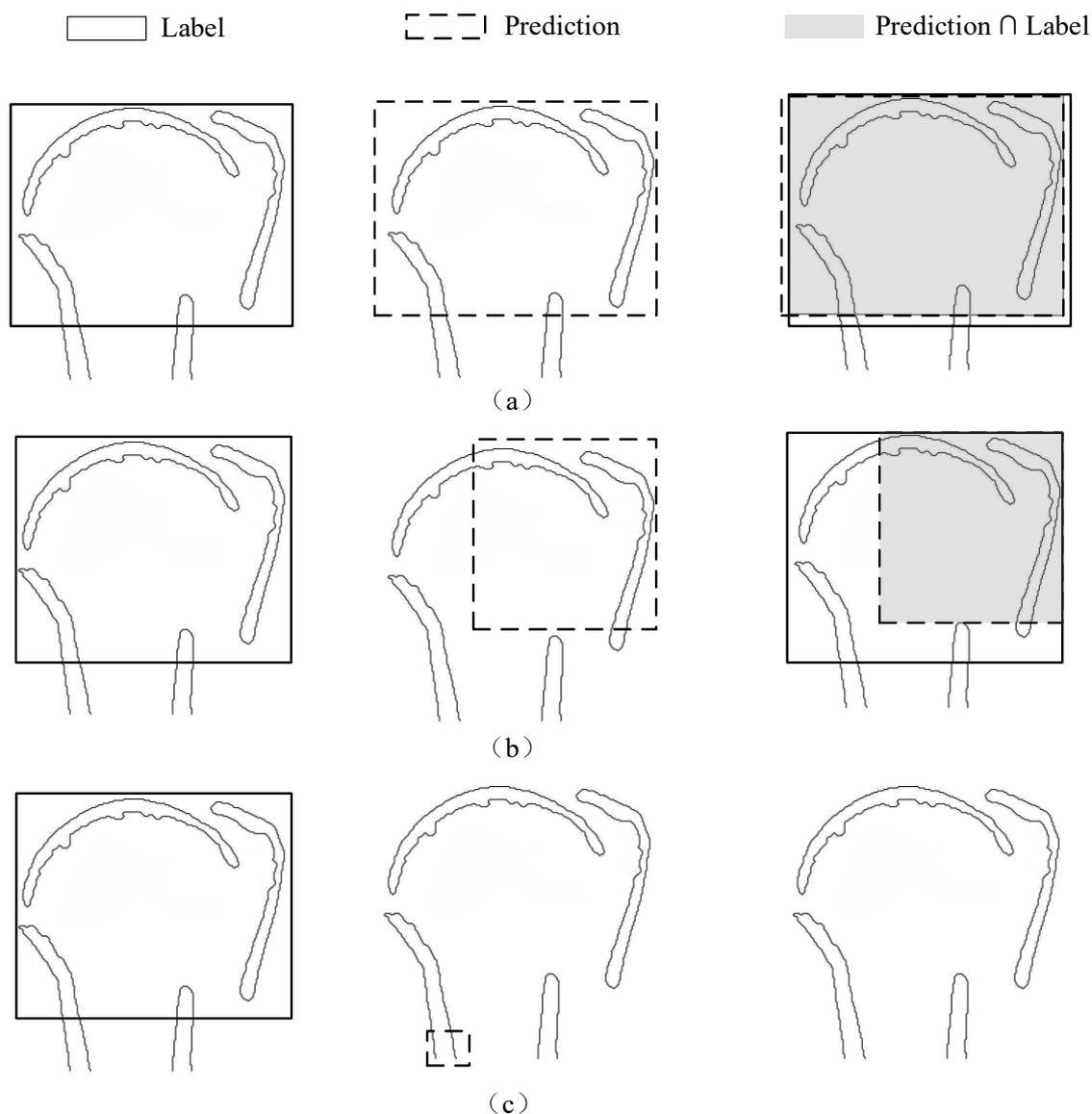
对于5.1.2中的骨折目标检测场景，在图B.1中，给出胫骨平台骨折CT影像检测结果，其中实线框为参考标准的标识框，虚线框为算法识别框。



图B.1 胫骨平台骨折 CT 影像检测示意图

图B.1 (b) 中描述了二维参考标准标记框与算法计算标记框的并集，计算方式见5.1.2.1中的公式(5)，图B.1 (c) 中描述了二维参考标准标记框与算法计算标记框的交集，计算方式见5.1.2.1中的公式(4)。

关于5.1.2.1中TP, FP, FN的相关描述，以肱骨近端骨折为例，图B.2给出在冠状面上检测结果示意图。实线框为参考标准的标识框，虚线框为算法识别框。



图B.2 肱骨近端骨折 CT 影像检测示意图

B.2 (a) 中，参考标准的标识框与算法标识框重叠度高，此时为真阳性（TP），即正确预测骨折，B.2 (b) 中，参考标准的标识框与算法标识框重叠度比较低，部分骨折区域未能正确识别，此时为假阴性（FN）。B.2 (c) 中，算法标识框的区域不是骨折区域，此时为假阳性（FP）。

对于5.1.1中的骨折分割场景，图B.3为肱骨近端骨折分割的示例，分别显示了输入的原始图像，参考标准以及算法输出的结果。

在图B.3中，黑色线条为肱骨近端骨折原始CT图像；实线灰色部分为肱骨近端骨折参考标准图像；虚线灰色部分为肱骨近端骨折分割结果图像。



图B.3 肱骨近端骨折 CT 影像分割示意图

B.3 测试集标注要求

B.3.1 骨骼分割场景

对于骨折分割场景，标注时具体要求如下：

- 采用多点标记法对骨折区域所有骨块进行解剖轮廓勾画；
- 沿骨折块皮质骨外缘解剖边界，按照骨折块解剖走行及结构进行逐层标注；
- 保持标记轮廓与骨性结构几何形态吻合，不能出现漏标、错标、多标的现象；
- 勾画完成后，生成二值化掩码，目标骨碎块像素区域赋值为1，背景区域赋值为0。

B.3.2 骨折检出场景

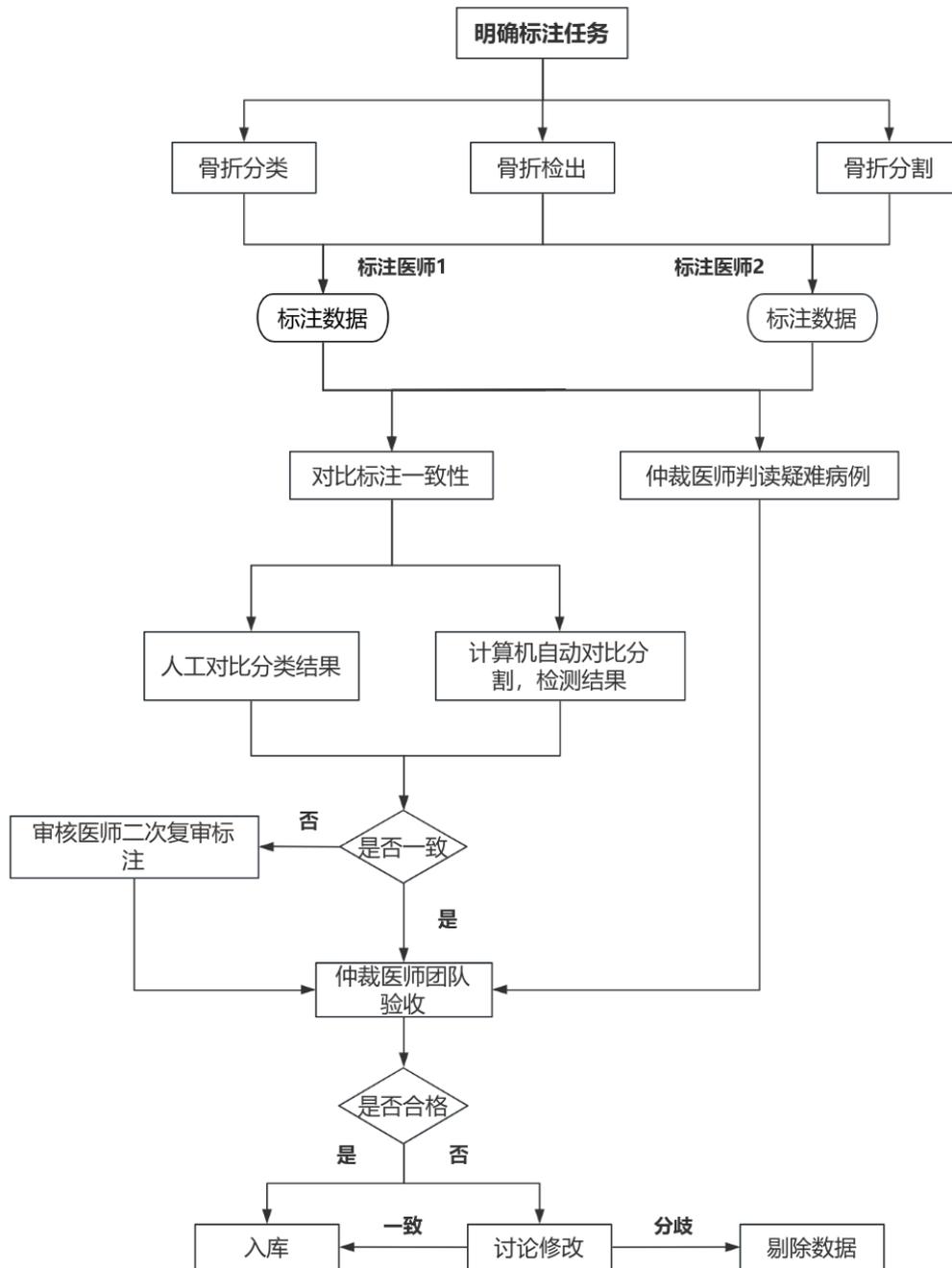
对于骨折检出标注场景，标注具体要求如下：

- 对于目标检测中存在多个骨折块的场景，若是多处骨折，对每一处进行单独标注；若是单一骨折，使用一个标注框标记整体骨折区域；
- 二维标注时，基于断层影像的骨折块形态学特征逐层生成闭合矩形框，确保矩形框完全包裹目标骨碎块；
- 三维标注时，通过 CT 影像的冠状位、矢状位和横断面优化三维包围盒的空间大小至最小有效容积（即完全包裹目标骨折块且排除周边非骨折区域）；
- 标注数据输出满足待测算法接口规范（至少包二维矩形框或三维包围盒坐标信息）。

B.4 标注流程

标注工作主要依据文献[18]进行。标注团队由标注医师，审核医师，仲裁医师组成，标注医师由两年及以上骨折临床工作经验的住院医师担任，审核医生由五年及以上骨折临床工作经验的主治医师

担任，仲裁医师由十年及以上骨折临床工作经验的副主任/主任医师担任。整体标注流程分为三个主要环节，如图B.4所示。



图B.4 骨折标注流程图

- a) 标注阶段：在确定标注任务后，由两名标注医师在双盲状态下独立完成标注。若在标注过程中遇到疑难病例，由仲裁医师进行判断并给出最终标注结果；仲裁医师仍存在分歧，将该数据剔除，不纳入数据库。当完成标注后，对于分类结果，由不少于两名审核医师判断标注是否一致，一致则判定双盲标注通过；对于骨折分割和骨折检出结果，由计算机自动计算 Dice 系数（分割结果）、Jaccard 系数（目标检测结果），当 Dice 系数 $\geq 95\%$ ，则认为骨折分割双盲标注结果一致；Jaccard $\geq 95\%$ ，则认为骨折检出双盲标注结果一致。

- b) 审核阶段：由审核医师对标注结果进行复审，标注结果一致的数据交由仲裁医师进行验收；标注结果不一致的数据，则由审核医师给出二次标注结果后进入验收阶段；存在分歧时交由仲裁医师给出最终结论。
- c) 验收阶段，由不少于两名仲裁医师对审核通过的结果进行独立检查与修改，纠正遗漏、错判和误判。若出现复杂情况，由仲裁医师集体讨论后给出最终结果，将达成一致的数据纳入数据库，存在分歧的数据进行剔除。

参 考 文 献

- [1] GB/T 9386-2008 计算机软件测试文档编制规范
- [2] YY/T1858-2022 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
- [3] 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)]
- [4] 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)]
- [5] 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)]
- [6] Bhatnagar A, Kekatpure AL, Velagala VR, Kekatpure A. A review on the use of artificial intelligence in fracture detection [J]. Cureus, 2024, 16(4): e58364. DOI: 10.7759/cureus.58364. PMID: 38756254; PMCID: PMC11097122.
- [7] Prijs J, Liao Z, To MS, et al. Development and external validation of automated detection, classification, and localization of ankle fractures: inside the black box of a convolutional neural network (CNN) [J]. European Journal of Trauma and Emergency Surgery, 2023, 49(2): 1057-1069. DOI: 10.1007/s00068-022-02136-1. PMID: 36374292; PMCID: PMC10175446.
- [8] Liu PR, Zhang JY, Xue MD, et al. Artificial intelligence to diagnose tibial plateau fractures: an intelligent assistant for orthopedic physicians [J]. Curr Med Sci, 2021, 41(6): 1158-1164. DOI: 10.1007/s11596-021-2501-4. Epub 2021 Dec 31. PMID: 34971441; PMCID: PMC8718992.
- [9] Hou Y, Zou X, Wang Y. Towards Cross-Scale Attention and Surface Supervision for Fractured Bone Segmentation in CT[J]. arXiv preprint arXiv: 2405.01204, 2024.
- [10] Yoon SJ, Hyong Kim T, Joo SB, Eel Oh S. Automatic multi-class intertrochanteric femur fracture detection from CT images based on AO/OTA classification using faster R-CNN-BO method. Journal of Applied Biomedicine. 2020 Dec ; 18(4) : 97-105. doi : 10.32725/jab.2020.013. PMID: 34907762.
- [11] Nattaphon Twinprai, Artit Boonrod, Arunnit Boonrod, et al. Artificial intelligence (AI) vs. human in hip fracture detection[J]. Heliyon, 2022, 8(11): e11266. doi: 10.1016/j.heliyon.2022.e11266. ISSN 2405-8440.
- [12] Cohen M, Puntinet J, Sanchez J, et al. Artificial intelligence vs. radiologist: accuracy of wrist fracture detection on radiographs[J]. European Radiology, 2023, 33(6): 3974-3983. doi: 10.1007/s00330-022-09349-3; PMID: 36515712.
- [13] 吕文晖, 夏菲, 周长圣, 等. 深度学习胸部CT辅助诊断系统在急诊创伤人群中的应用. 中华医学杂志, 2021, 101(07): 481-481. DOI: 10.3760/cma.j.cn112137-20201117-03123.
- [14] 张倩, 王大为, 隋时, 等. 深度学习计算机辅助诊断系统检测胸部DR肋骨骨折的验证性研究[J]. 中国临床医学影像杂志, 2022, 33(6): 5.
- [15] 钟守昌, 现代医学影像成像原理及其临床应用[M]. 北京: 清华大学出版社, 2019.
- [16] ISO 12052: 2017 Health informatics — Digital imaging and communication in medicine (DICOM) including workflow and data management
- [17] GB/T 35273-2020 信息安全技术 个人信息安全规范
- [18] 骨折影像数据标注与质量控制专家共识.