



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 images Algorithm performance testing methods

(点击此处添加与国际标准一致性程度的标识)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX—XX—XX 发布

XXXX—XX—XX 实施

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。

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的骨折影像辅助分析软件。
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 第一部分：术语

YY/T 1833.2 人工智能医疗器械 质量要求和评价 第二部分：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 第三部分：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 X射线检查（X ray examination）

用X射线诊断疾病的方法。分普通检查、特殊检查和造影检查。透视是一种简便而常用的检查方法，可从不同角度观察脏器的形态及功能改变。

3.3 计算机断层扫描（computed tomography, CT）

利用精确准直的X线束、 γ 射线、超声波等，与灵敏度极高的探测器一同围绕人体的某一部位作一个接一个的断面扫描，具有扫描时间快，图像清晰等特点，可用于多种疾病的检查。

3.4 医学数字成像和通信（digital imaging and communications in medicine, DICOM）

医学图像和相关信息的国际标准（ISO 12052）。它定义了质量能满足临床需要的可用于数据交换的医学图像格式。

3.5 三维（three dimensional）

用宽度、长度和深度来定义一个物体的计算机图形。

3.6 三维数据元素（3D data element）

三维数据的类型，通过定义、标识、表示和允许值等一系列属性描述的数据单元。是全息导航系统中不可再分的最小数据单元。

3.7 三维重建（3D reconstruction）

对三维物体建立适合计算机表示和处理的数学模型，是在计算机环境下对其进行处理、操作和分析其性质的基础，也是在计算机中建立表达客观世界的虚拟现实的关键技术。

3.8 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/T9386-2008. 3. 2有修改]

4 算法性能测试要求

4.1 概述

骨折影像辅助诊断软件 算法性能测试方法过程宜参照YY/T 1857-2022中4.1的要求，建立测试文档。

4.2 测试环境

算法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：

——影像设备采集的原始图像进行脱敏，去除敏感信息；脱敏后不对影像内容进行处理（包括但不限于修改，操作或者有损压缩等操作）；

——确保测试集中的影像数据严格满足产品适用范围，不满足适用范围的数据需要进行剔除。测试数据（CT, Xray）应包括但不限于考虑患者年龄，照射体位，骨折类型，骨折位置等信息。对于CT数据，还应考虑层厚，重建核等信息；

——X-ray影像应该是正位图像或者侧位图像；

——同一测试集内的数据元应是唯一的，不重复的。同一数据集的各个子集是可区分的；

——避免严重运动，金属，曝光不足，操作不当等情形带来的伪影数据。

注：示例见附录B.2

4.3.2.2 测试集金标准要求

宜参照YY/T 1833.3-2022 数据标注通用要求中第四章。根据厂商实际情况，可选择标注加审核或者多标注加仲裁的标注范式。

如为标注+审核的标注范式。标注医生对数据进行初步骨折检测/分割/分类任务。需要记录所有标注信息，包括骨折具体位置和范围（如左上角坐标和右下角坐标。或者通过标注掩膜位置间接提取左上角和右下角坐标），骨折的每一种类型（如有），骨折分割中的语义标签（如有），时间等信息。注意在骨折检测与骨折分割的标注中需要精确反映骨折位置，尽量避免多余的标注。在标注医师完成后，审核医师对所有标注结果进行审核和矫正，检测标注的准确性。如发生漏检，则需要对骨折病灶进行添加。如发生性质和位置的错判，则需要对此进行订正。如发生分割不准确，则需要对轮廓进行修改。经审核的数据为最终的标注结果。

如为多标注加仲裁的标注范式，多名标注医师背靠背对数据进行初步骨折检测/分割/分类任务需要记录所有标注信息，包括骨折具体位置和范围（如左上角坐标和右下角坐标。或者通过标注掩膜位置间接提取左上角和右下角坐标），骨折的每一种类型（如有），骨折分割中的语义标签（如有），时间等信息。注意在骨折检测与骨折分割的标注中需要精确反映骨折位置，尽量避免多余的标注。在几名标注医师完成后，由计算机自动判断结果的一致性，以所有人标注结果的并集作为结果。如标注结果完全一致或多数一致，则直接进入金标准。如结果不一致，则需要进行仲裁。在分割类任务中，仲裁医生择取更准确的标注结果再进行调整。经仲裁的数据为最终的标注结果。如果仲裁过程中遇到疑难问题，仲裁专家需进行集体讨论并得出最终结果。

注1：标注医师宜由医院从事阅片工作并具备住院医师及以上职称的医师担任。

注2：审阅医师宜由为医院高年资主治医师及以上职称的医师。

注3：仲裁医师宜由为医院高年资副主任医师或主任医师担任。

4.3.2.3 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。

4.3.2.4 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：

——设备厂商，型号的多样性；

——成像参数多样性包括但不限于层厚，重建内核（CT），设备场强和序列，不同剂量条件；

——照射体位多样性（如Xray中的AP，PA，斜位等，CT中的仰卧位，俯卧位）。

患者维度考虑以下内容：

——性别，年龄多样性；

——宜来自不同地区，不同医疗机构；

——来源场景多样性（如急诊，门诊，体检等场景）；

——骨折阶段多样性（如新鲜，陈旧，愈合期骨折等）；

——骨折类型以及程度多样性（如完全，不完全，开放性，闭合性，压缩性骨折等）；

——骨折发生位置的多样性（如肩胛骨，锁骨，肋骨，脊柱性骨折等）。

4.3.2.5 扩增数据要求

宜参照YY/T 1858-2022中4.2.3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，标准器官宜参照YY/T 1858-2022中4.3.5的要求。

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/T1858-2022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 测试步骤

在骨折块分割场景下，比如分割股骨骨折中的骨折碎块，测试步骤宜参照YY/T 1858-2022中5.1.2.1。

5.1.1.2 骨折碎块重合度计算

当分割区域为一般实体（如骨折碎块），宜采用Dice系数，Jaccard系数评价骨折碎块的重合度。当对多个骨折碎块进行分割时（如适用），需计算所有骨折碎块的平均Dice系数和平均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。

5.1.1.3 骨折碎块轮廓线重合度

骨折块轮廓线是指骨折碎块的边缘轮廓线。如适用，宜采用豪斯多夫距离，倒角距离进行重合度计算。

豪斯多夫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：

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

式中：

$D_{charf}(T, I)$ —— 倒角距离；

I —— 算法输出骨折碎块的边缘轮廓线；

T —— 参考标准中骨折碎块的边缘轮廓线；

$d_I(t)$ —— T 中任意点 t 在 I 上的最短距离。

5.1.2 骨折目标检测场景

5.1.2.1 骨折目标检测标注与匹配

如待测算法利用标记框检测骨折部位，导出算法预测骨折部位的标记框，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，计算方式如下所示：

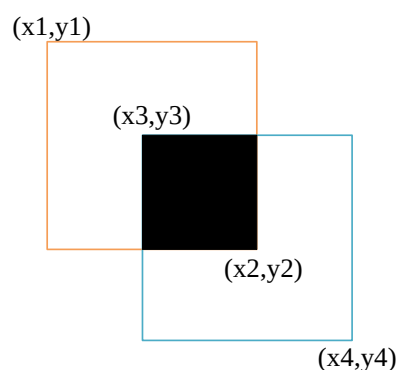


图1 检测框示意图

如图1所示，假设橘黄色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) 。蓝色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 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中的原则进行匹配度计算，得到TP、FP、FN、 $Intersect$ 、 $Union$ ，按照本文件条款5.1.2.3~条款5.1.2.10计算各性能指标。

5.1.2.3 准确度

准确度计算方法宜参照YY/T 1858-2022中5.1.1.4。

5.1.2.4 召回率

召回率计算方法宜参照YY/T 1858-2022中5.1.1.3。

5.1.2.5 F1 分数

F1分数计算方法宜参照YY/T 1858-2022中5.1.1.5。

5.1.2.6 平均准确度

平均准确度计算方法宜参照YY/T 1858-2022中的5.1.1.6。

5.1.2.7 平均准确度均值

平均准确度均值计算方法宜参照YY/T 1858-2022中的5.1.1.7。

5.1.2.8 非病变定位率

非病变定位率公式如公式（7）所示：

$$AveFP = \frac{N_{FP}}{N_{Case}} * 100\% \dots\dots\dots (7)$$

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。

5.1.3 骨折分类场景

5.1.3.1 测试步骤

在骨折分类场景下（如对股骨骨折分可为两型、三型、四型、五型、六型骨折），算法测试执行以下步骤：

- 1) 将测试集分别由住院医师、主治医师进行分析。结果格式宜与参考标准兼容。内容包括分类结果；
- 2) 将测试集输入算法、输出结果宜于参考标准兼容，内容包括分类结果；
- 3) 住院医师、主治医师结合算法分类结果对骨折病灶区域进行分类；
- 4) 将住院医师、主治医师、住院医师加算法、主治医师加算法的结果与参考标准进行对比，计算真阳性、假阳性、真阴性的样本数量，构造混淆矩阵，混淆矩阵宜参照YY/T 1858-2022中的5.1.3.1。计算如下性能指标。

5.1.3.2 灵敏度

灵敏度的计算方法参照YY/T 1858-2022中5.1.3.2。

5.1.3.3 特异度

特异度的计算方法宜参照YY/T 1858-2022中5.1.3.3。

5.1.3.4 阳性预测值

阳性预测值的计算方法宜参照YY/T 1858-2022中5.1.3.5。

5.1.3.5 阴性预测值

阴性预测值的计算方法宜参照YY/T 1858-2022中5.1.3.6。

5.1.3.6 准确率

准确率的计算方法宜参照YY/T 1858-2022中5.1.3.7。

5.1.3.7 Kappa 检验

Kappa检验的计算方法参照YY/T 1858-2022中5.1.3.9。

5.1.3.8 受试者操作特征曲线

受试者操作特征曲线宜参照YY/T 1858-2022中5.1.3.10。

5.1.3.9 Phi 系数

对每一个骨折类型，计算住院医师与主治医师的一致性。通过对比使用算法前后，住院医师与主治医师对每一个骨折类型的一致性证明算法是否能提高不同级别医师的诊断效能。Phi的计算公式如公式（8）所示：

$$\phi = \frac{(N_{TT} * N_{FF} - N_{FT} * N_{TF})}{\sqrt{(N_{TT} + N_{FT}) * (N_{TF} + N_{FF}) * (N_{TT} + N_{TF}) * (N_{FT} + N_{FF})}} \dots\dots\dots (8)$$

式中：

ϕ —— phi系数；

- N_{TT} —— 主治医师与住院医师均鉴定为骨折的数量；
 N_{FT} —— 主治医师鉴定为非骨折，住院医师鉴定为骨折的数量；
 N_{TF} —— 主治医师鉴定为骨折，住院医师鉴定为非骨折的数量；
 N_{FF} —— 主治医师与住院医师均鉴定非为骨折的数量。

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（性别，年龄），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（分为儿童，青年，中年，老年）；
- 子集按照性别进行拆分；
- 对某一部位骨折进行检测，子集按照该部位所有骨折类型（如闭合性骨折、开放性骨折、完全性骨折、不完全性骨折等）进行拆分；
- 对多个部位进行检测（如软件适用），子集应按照国家不同部位进行拆分；
- 子集按照不同成像设备厂家进行拆分；
- 子集按照不同医院、不同部门进行拆分。

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，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，收集压力测试，形成鲁棒性研究资料。

对抗测试适考虑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。

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。举例如下：

- 术后患者病例；
- 病理性骨折病例；
- 骨转移病例；
- 并发症病例；
- 病例中存在不完整骨结构（例如胸部CT病例中不包含所有肋骨）。

5.2.3 重复性测试

如适用，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。使用SIT(System Integration Test)、FAT(Feature Acceptance Test)、UAT(User Acceptance Test), 准生产等不同测试环境进行测试。观察结果是否变化。

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。

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。

如适用，对于CT肋骨骨折的检测和定位，应分别记录完成肋骨检出，肋骨定位，骨折类型-判断的计算效率。

5.2.6 算法错误统计

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

- 肋骨/其他骨折检出的假阴，假阳；
- 对肋骨/其他骨折发生位置的误判（如肋骨骨折发生的具体位置标准1-12肋，前/中/后肋，左/右肋等）；
- 对肋骨/其他位置骨折类型的误判（如新鲜骨折，陈旧性骨折等）；
- 骨骼分割误差；
- 骨骼标签错误，如肋骨标签，脊柱标签。锁骨标签，肩胛骨标签等。

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:

- 1) 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。如小腿受到撞击,于撞击处发生胫腓骨骨干骨折。
- 2) 间接暴力。暴力通过传导、杠杆、旋转和肌收缩使肢体受力部位的远处发生骨折。如跌倒时以手掌撑地,依其上肢与地面的角度不同,暴力向上传导,可致桡骨远端骨折或肱骨髁上骨折。骤然跪倒时,股四头肌猛烈收缩,可致髌骨骨折。
- 3) 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,如远距离行军易致第 2、3 跖骨及腓骨下 1/3 骨干骨折,称为疲劳性骨折,也可称为应力性骨折。

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:

- 1) 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。
- 2) 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。
- 3) 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。即使临床上已表现为明显骨折者,X线平片检查也是必要的,可以帮助了解骨折的类型和骨折端移位情况,对于骨折的治疗具有重要指导意义。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,必要时应拍摄特殊位置的X线平片。有时不易确定损伤情况时还需拍对侧肢体相应部位的X线平片,以便进行对比。

X线检查是骨折最常用的检查方法,但对早期、不典型病例及复杂的解剖部位,X线在确定病变部位和范围上受到限制。CT弥补了传统X线检查的不足。一般来讲,骨和关节解剖部位越复杂或常规X线越难以检查的部位,CT越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X线及CT未能发现的隐匿性骨折并确定骨挫伤的范围。

附录 B (资料性) 特定条款指南和方法说明

B.1 通用指南

对于骨折影像辅助诊断软件 算法性能评价，测试要求贯穿整个测试过程，第四章对测试要求进行详细阐述，包括概述、测试环境、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。

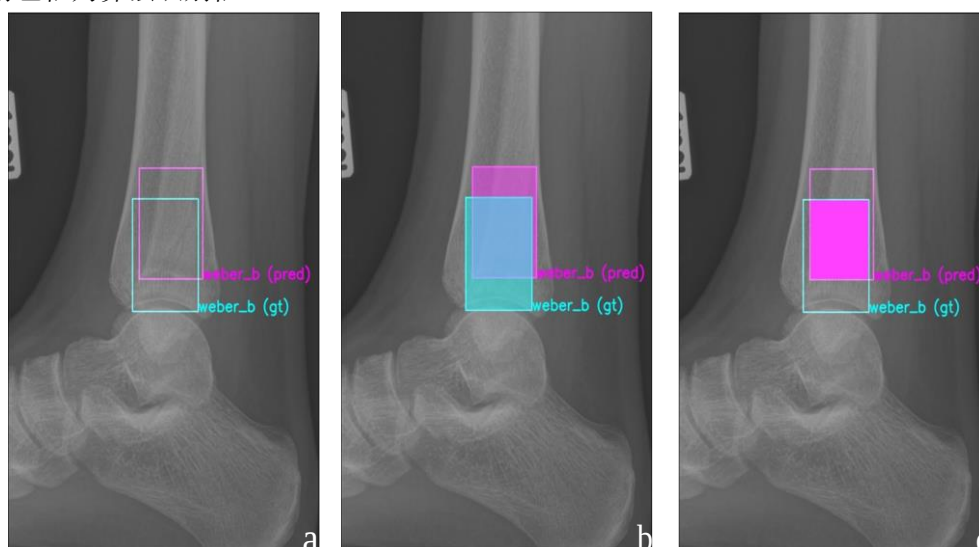
B.2 特定条款的指南或者方法说明

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) 管电压 (kVp)：通常，成人四肢和脊柱的 X 射线检查使用的管电压在 70–120 kVp 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；
- b) 管电流 (mA)：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 (AEC) 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；
- c) 曝光时间：较短的曝光时间可以减少因患者移动造成的图像模糊，通常通过提高管电流或管电压来实现。使用 AEC 可以帮助自动设定最合适的曝光时间；
- d) 图像锐利度与对比度：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；
- e) 定位与投照角度：选择正确的投照角度对于显示特定骨折类型至关重要，如前后位、侧位或特定角度的斜位拍摄，可以更清晰地显示骨折情况。

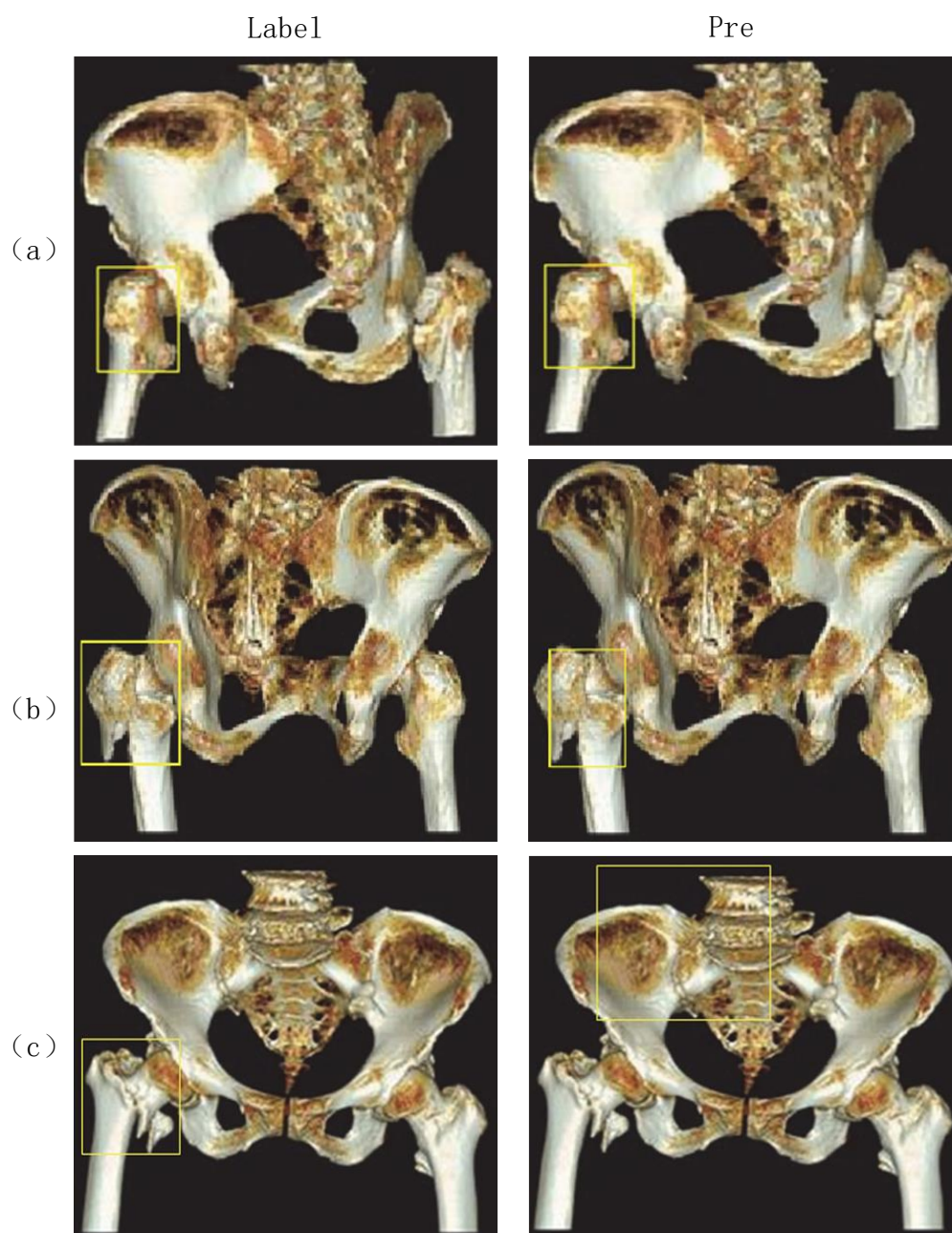
对于5.2.2中的骨折检测场景，在图B.1中，给出了踝关节X光骨折检测结果，其中蓝框为参考标准的标识框，粉色框为算法识别框。



图B.1 踝关节骨折检测示意图

图B.1 (b) 中描述了参考标准与算法计算的重合区域，计算方式见5.2.2中的公式 (2)，图B.1 (c) 中描述了参考标准与算法计算的相交区域，计算方式见5.2.2中的公式 (1)。

关于5.2.2中TP, FP, FN的相关描述, 以股骨转子间骨折为例, 图B. 2给出三维视图上检测结果意图。左侧为参考标准的标识框, 右侧为算法计算出的标识框。

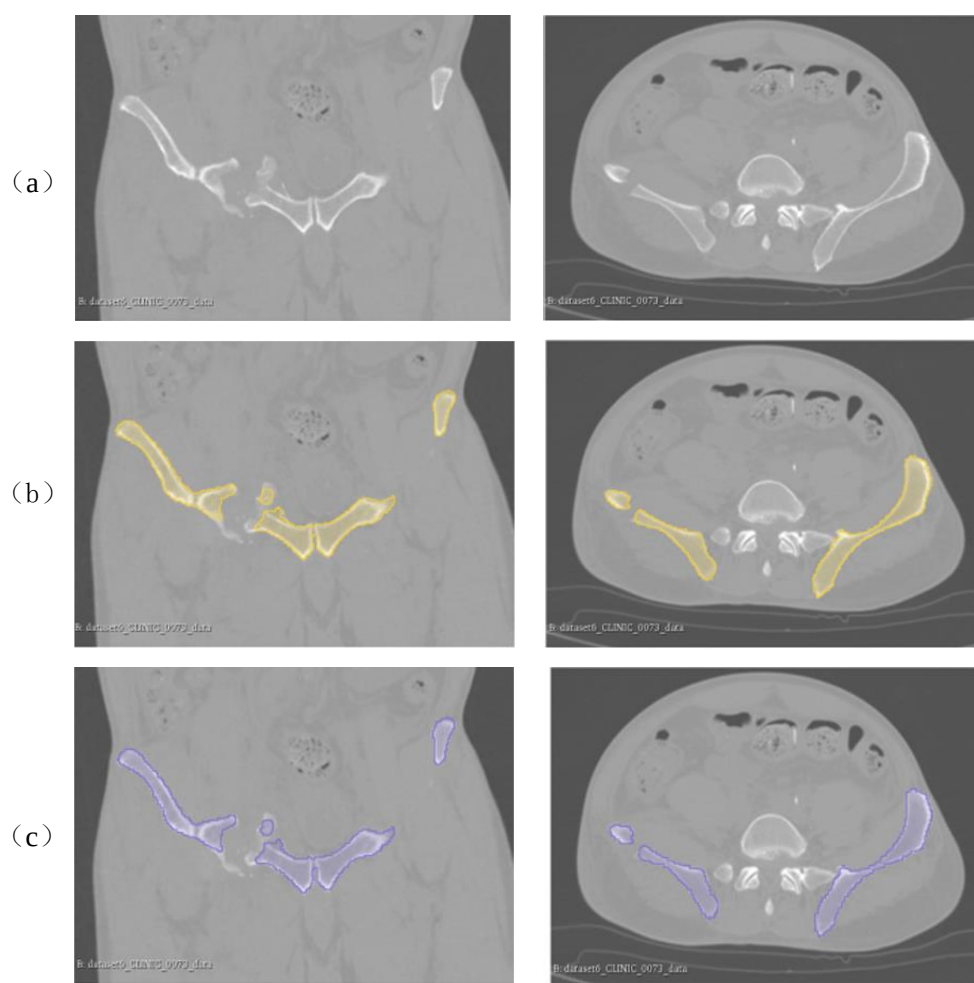


图B. 2 股骨转子间骨折检测示意图

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, 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5.2.1中的骨折分割场景, 图B. 1盆骨骨折分割的示例, 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, (a) 为待分割盆骨骨折的原始CT图像; (b) 为参考标准中盆骨骨折的标准图像 (黄色部分), 为公式(1)中的True; (c) 为算法输出的盆骨骨折分割的结果图像 (紫色部分)。



图B. 3 骨盆骨折分割示意

参 考 文 献

- [1] 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
- [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] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] Cohen M, Puntonet 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.
- [12] 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.
- [13] 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.

