

T/

团 标 准

T/XXX XXXX—XXXX

# 特种设备 人工智能 焊接接头射线检测图 像智能诊断系统性能评估规范

Special Equipment — Artificial Intelligence — Specification for Performance  
Evaluation of Intelligent Diagnostic Systems for Radiographic Testing Images of  
Welded Joints

(征求意见稿)

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

## 目 次

前 言 .....	II
1 范围 .....	3
2 规范性引用文件 .....	3
3 术语和定义 .....	3
4 缩略语 .....	6
5 一般要求 .....	6
6 系统结构与流程 .....	7
7 数据的要求 .....	8
8 算法模型设计 .....	10
9 系统功能要求 .....	11
10 系统智能诊断性能评估 .....	13

## 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。

本文件由中国特种设备检验协会提出并归口。

本文件起草单位：

本文件主要起草人：

# 特种设备 人工智能 焊接接头射线检测图像智能诊断系统性能评估规范

## 1 范围

1.1 本文件规定了基于人工智能的焊接接头射线检测图像智能诊断系统的性能规范, 所述技术规范内容包括一般要求、系统的结构与流程、数据的要求、算法模型设计要求、系统功能要求、系统智能诊断性能评估。

1.2 本文件适用于 NB/T 47013. 2、NB/T 47013. 11 和 NB/T 47013. 14 所涉及的承压设备射线检测范围, 适用的焊接接头射线检测图像包括: 射线照相法获得的底片数字化图像、射线数字成像检测获得的射线检测图像和射线计算机辅助成像检测获得的射线检测图像。

本文件适用于承压设备以外的其他金属材料、支承件和结构件的焊接接头射线检测。NB/T 47013. 2、NB/T 47013. 11 和 NB/T 47013. 14 标准未覆盖领域的射线检测智能诊断可参照本文件执行。

## 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。

GB/T 25000. 51	系统与软件工程 系统与软件质量要求和评价 (SQuaRE) 第51部分: 就绪可用软件产品 (RUSP) 的质量要求和测试细则
GB/T 26141. 1	无损检测 射线照相底片数字化系统的质量鉴定 第1部分: 定义、像质参数的定量测量、标准参考底片和定性控制
GB/T 26141. 2	无损检测 射线照相底片数字化系统的质量鉴定 第2部分: 最低要求
GB/T 28452	信息安全技术 应用软件系统通用安全技术要求
GB/T 30821	无损检测 数字图像处理与通讯
GB/T 41864	信息技术 计算机视觉 术语
GB/T 41867	信息技术 人工智能 术语
GB/T 43613	无损检测 数字射线检测图像处理与通信
GB/T 45225	人工智能 深度学习算法评估
NB/T 47013. 2	承压设备无损检测 第2部分: 射线检测
NB/T 47013. 11	承压设备无损检测 第11部分: X射线数字成像检测
NB/T 47013. 14	承压设备无损检测 第14部分: X射线计算机辅助成像检测
TSG Z8001	特种设备无损检测人员考核规则

## 3 术语和定义

NB/T 47013. 1、NB/T 47013. 2、NB/T 47013. 11 和 NB/T 47013. 14 界定的下列术语和定义适用于本文件。

### 3.1 智能诊断 Intelligent diagnostic

利用计算机分析焊接接头射线检测质量评定相关的信息，以对焊接接头中的缺陷进行定位、分类、尺寸测量及分级评定等。

### 3.2 长度 Length

用于量化两个点之间的实际物理距离，以毫米（mm）为基本单位。

### 3.3 置信度 confidence

总体参数值落在样本统计值某一区内的概率。

### 3.4 交并比 intersection over union(IoU)

智能诊断模型预测的缺陷区域与实际缺陷区域的交集面积与其并集面积的比值，其计算公式为：

$$IoU = \frac{Area(A \cap B)}{Area(A \cup B)} \quad (1)$$

式中， $IoU$  为两个几何区域面积的交并比， $Area(A \cap B)$  为智能诊断模型预测的缺陷区域与实际缺陷区域的交集的面积， $Area(A \cup B)$  为预测的缺陷区域与实际的缺陷区域的并集的面积。

### 3.5 平均交并比 Mean Intersection over Union (mIoU)

所有缺陷类别交并比的算术平均值。其计算公式为：

$$mIoU = \frac{1}{k} \sum_{i=1}^k IoU_i \quad (2)$$

式中， $k$  为缺陷类别总数， $IoU_i$  为第  $i$  类缺陷的交并比。

### 3.6 混淆矩阵

$n$  行  $n$  列的矩阵， $n$  为缺陷类型总数。矩阵的每一列代表了模型预测的缺陷类别，每一行代表了缺陷的真实归属类别。

### 3.7 检出样本 detected sample (DS)

智能诊断模型在对应位置区域检测出一个  $IoU$  不低于 0.3 的缺陷的焊接缺陷标注样本。

### 3.8 漏检样本 missed sample (MS)

智能诊断模型在对应位置区域没有检测出一个  $IoU$  不低于 0.3 的缺陷的焊接缺陷标注样本。

### 3.9 虚检样本 fabricated sample (FS)

智能诊断模型在没有焊接缺陷标注样本的位置检测出的缺陷样本。

### 3.10 正确分类样本 correctly classified sample (CCS)

智能诊断模型预测的缺陷类型与标注样本的类型一致的缺陷检出样本。

### 3.11 正确分级样本 correctly graded samples (CGS)

智能诊断模型评定的缺陷等级与对应标注样本一致的正确分类样本。

### 3.12 不可接受缺陷 unacceptable defect

依据评定标准被认为质量等级不合格的缺陷。

### 3.13 缺陷检出率 defect detection rate (DDR)

对于给定的测试数据集，某一类缺陷的检出样本占相应标注样本的数量比。其计算公式为：

$$DDR_i = \frac{DS_i}{DS_i + MS_i} \times 100\% \quad (3)$$

式中， $i \in \{\text{气孔, 夹渣, 未焊透, 未熔合, 内凹, 烧穿, ...}\}$ 为缺陷类别， $DDR_i$ 为第*i*类缺陷的检出率， $DS_i$ 为第*i*类缺陷的检出样本数量， $MS_i$ 为第*i*类缺陷的漏检样本数量。

### 3.14 缺陷虚检率 defect false alarm rate (DFAR)

对于给定的测试数据集，某一类缺陷的虚检样本占相应样本的数量比。其计算公式为：

$$DFAR_i = \frac{FS_i}{DS_i + MS_i + FS_i} \times 100\% \quad (4)$$

式中， $i \in \{\text{气孔, 夹渣, 未焊透, 未熔合, 内凹, 烧穿, ...}\}$ 表示缺陷类别， $DFAR_i$ 表示第*i*类缺陷的虚检率， $DS_i$ 为第*i*类缺陷的检出样本数量， $MS_i$ 为第*i*类缺陷的漏检样本数量， $FS_i$ 为第*i*类缺陷的虚检样本数量。

### 3.15 缺陷分类准确率 defect classification accuracy (DCA)

对于给定的测试数据集，某一类缺陷正确分类样本占相应检出样本的数量比，其计算公式为：

$$DCA_i = \frac{CCS_i}{DS_i} \times 100\% \quad (5)$$

式中， $i \in \{\text{气孔, 夹渣, 未焊透, 未熔合, 内凹, 烧穿, ...}\}$ 表示缺陷类别， $DCA_i$ 表示第*i*类缺陷的分类准确率， $DS_i$ 为第*i*类缺陷的检出样本数量， $CCS_i$ 为第*i*类缺陷的正确分类样本数量。

### 3.16 缺陷分级准确率 defect grading accuracy (DGA)

对于给定的测试数据集，缺陷分级准确率为某一类缺陷正确分级样本占相应正确分类样本的数量

比, 其计算公式为:

$$DGA_i = \frac{CGS_i}{CCS_i} \times 100\% \quad (6)$$

式中,  $i \in \{\text{气孔, 夹渣, 未焊透, 未熔合, 内凹, 烧穿, ...}\}$  表示缺陷类别,  $DGA_i$  表示第  $i$  类缺陷的分级准确率,  $CCS_i$  为第  $i$  类缺陷的正确分类样本数量,  $CGS_i$  为第  $i$  类缺陷的正确分级样本数量。

### 3.17 效率 efficiency

对于给定的软硬件环境条件和测试数据集, 焊接缺陷智能诊断系统评估每个样本耗费的平均响应时间。其计算公式为:

$$\tilde{T} = \frac{1}{N} \sum_{i=1}^N (T_{fi} - T_{si}) \quad (7)$$

3.18 式中,  $\tilde{T}$  表示平均响应时间,  $N$  表示样本数量,  $T_{fi}$  表示第  $i$  个图像样本结果输出的时间,  $T_{si}$  表示第  $i$  个图像样本结果输入的时间。

## 4 缩略语

BMP: 位图 (Bitmap)

CPU: 中央处理单元 (Central Processing Unit)

DICOM: 医学数字成像和通信 (Digital Imaging and Communication in Medicine)

DICONDE: 无损评价中数字成像和通信 (Standard Practice for Digital Imaging and Communication in Nondestructive Evaluation)

GPU: 图形处理单元 (Graphics Processing Unit)

JPEG: 联合图像专家组 (Joint Photographic Experts Group)

TIFF: 标签图像文件格式 (Tag Image File Format)

NMS: 非极大值抑制 (Non-Maximum Suppression)

## 5 一般要求

### 5.1 人员要求

5.1.1 从事本文件所用数据集制作的人员, 应取得特种设备无损检测 RT-II、RT-III 或 RT (D) -II 人员资格, 从事数据标注的人员应至少取得特种设备无损检测 RT-II 或 RT (D) -II 人员资格, 从事标注数据复合的人员应取得特种设备无损检测 RT-III 人员资格。

5.1.2 从事本文件所述系统性能评估的人员, 应取得特种设备无损检测 RT-II、RT-III 或 RT (D) -II 人员资格, 并充分理解本文件中的系统评估指标体系, 熟练掌握及应用性能评估流程与方法。

5.1.3 采用本文件所述系统开展射线检测智能诊断并出具正式检测报告的人员, 应经过专业培训, 熟悉系统的操作方法和性能指标, 并且应当取得特种设备无损检测 RT-II 或 RT (D) -II 及以上人员资格。

## 5.2 系统运行环境要求

### 5.2.1 端侧部署技术要求

端侧适用于本地化部署、实时性要求高、数据隐私敏感或网络受限场景，具体要求如下：

- a) 算力应支持低延迟推理；
- b) 存储容量设计应不低于单日焊接接头射线检测原始图像及简单推理结果的短期缓存需求。

### 5.2.2 边侧部署技术要求

边侧适用于区域化数据汇聚、中等实时性及局部协同场景，具体要求如下：

- a) 服务器算力应满足实时处理需求，支持多端侧设备并发数据处理与模型推理；
- b) 存储应保证多设备上传的射线图像及推理结果高效存取，支持存储容量弹性扩展。

### 5.2.3 云侧部署技术要求

云侧适用于大规模数据处理、非实时性分析、集中管理多源数据的场景，具体要求如下：

- a) 服务器算力应满足模型训练及多边侧并发调度需求，支持算力弹性扩展；
- b) 存储应采用分布式存储系统，支持海量历史数据长期存储，同时配置冗余与容灾机制以保障数据持久性，支持存储容量弹性扩展；
- c) 网络应采用高带宽、高稳定性、可扩展的网络架构，同时数据传输符合信息安全标准。

## 6 系统结构与流程

### 6.1 系统结构

焊接接头射线检测图像智能诊断系统结构如图1所示，系统应包含数据管理、训练管理、智能诊断、结果审核与报告生成及系统管理等功能模块，宜根据需要设置数据标注功能模块。各模块的主要功能如下：

- a) 数据管理用于系统内的所有数据进行管理，包括检测项目及工艺信息的获取与管理、数据质量管理、图像数据预处理、数据集划分、数据展示、报告模板配置、历史检测图像管理等功能；
- b) 数据标注用于对射线检测图像中的缺陷等对象进行标注，包括提供图形化标注工具、图像标注、标注结果管理等功能；
- c) 训练管理用于智能诊断的算法及检测模型进行管理，并配置其运行的环境，包括算法管理、模型训练、版本管理及环境配置等功能；
- d) 智能诊断用于对训练所得的检测模型进行部署及应用，包括缺陷识别与分级评定、像质计识别、铅质标识识别、结果可视化展示等功能；
- e) 结果审核与报告生成用于展现智能诊断结果并接受评定人员的评估审核结果，最终生成检测报告，包括显示交互、评估审核及生成报告等功能；
- f) 系统管理用于对系统后台数据和用户权限进行管理及维护，包括用户及权限管理、安全管理等功能。



图 1 焊接接头射线检测图像智能诊断系统结构

## 6.2 系统流程

焊接接头射线检测图像智能诊断系统流程如图2所示：

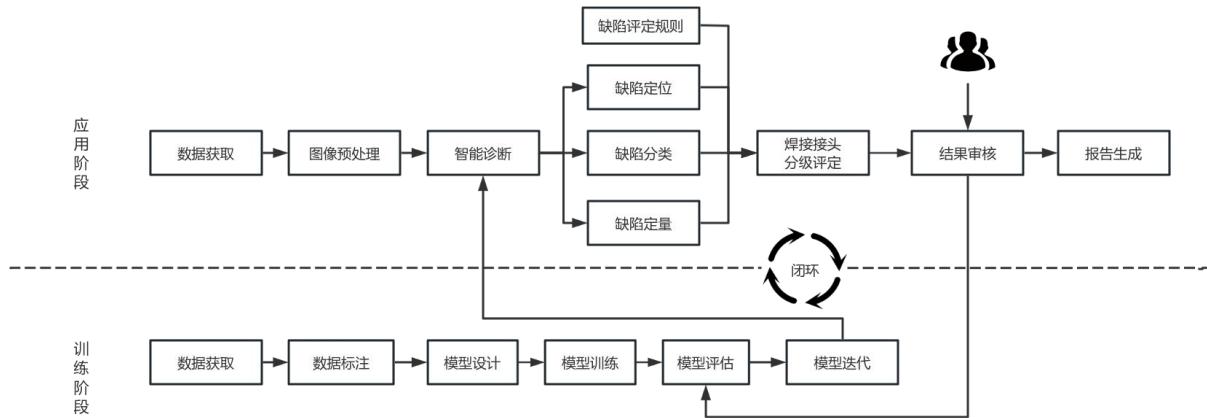


图 2 焊接接头射线检测图像智能诊断系统流程

系统流程整体包括训练阶段与应用阶段：

- 在训练阶段系统完成训练样本数据获取、数据标注、模型设计、模型训练、模型评估与模型迭代；
- 在应用阶段完成待测样本数据获取、图像预处理、模型应用、缺陷的定位、缺陷分类、缺陷定量、焊接接头分级评定、结果审核（由评定人员进行）、报告生成；
- 整个流程中的模型评估、模型迭代、智能诊断及结果审核形成闭环，实现系统的循环优化。

## 7 数据的要求

### 7.1 数据信息

7.1.1 完整的数据信息应包括射线检测图像和检测信息。

7.1.2 检测信息应满足影像诊断的参数需要。影像诊断需要的参数涉及检测对象规格和材质、像素大小、检测标准、检测方式（如 X 射线或  $\gamma$  射线胶片成像、X 射线数字成像检测、X 射线计算机辅助成像检测）、透照工艺、焊接方法、坡口形式等信息。

7.1.3 检测信息应能够从 DICONDE 格式文件、智能诊断系统数据库或其它关联系统的数据库中获得。

## 7.2 影像及图像质量

### 7.2.1 原始检测影像质量

原始检测影像的成像质量应符合 NB/T 47013 第 2 部分、第 11 部分和第 14 部分中的相关要求。

### 7.2.2 底片数字化质量要求

7.2.2.1 图像像素间距应不大于  $85 \mu\text{m}$ 。

7.2.2.2 图像 A/D 转换位数应不小于 12bit。

7.2.2.3 数字化图像中所能分辨的最细像质计丝应与底片一致，密度对比度灵敏度  $\leq 0.02$ 。

7.2.2.4 数字化过程中如产生伪缺陷，伪缺陷不应干扰质量评定。

7.2.2.5 单个焊接接头包括多张检测图像数据时，片序准确，影像正向。

## 7.3 格式与存储

7.3.1 图像文件宜采用 DICONDE 格式，DICONDE 文件应符合 GB/T 30821 和 GB/T 43613 标准要求。

7.3.2 DICONDE 格式文件的扩展名类型应为 “.dcm”、“.diconde”、“.dicom”、“.dcn”。

7.3.3 DICONDE 格式数字化图像可转换为 bmp、jpg、jpeg、tif、tiff、png 等格式，转换后的图像不能用于人工质量复审。

7.3.4 可采用移动设备（硬盘或 U 盘）或平台对数据进行存储。

## 7.4 数据预处理

7.4.1 数据预处理包括数据清洗及数据脱敏等过程。数据清洗应明确处理的方法，必要时，应当进行数据脱敏。

7.4.2 数据经预处理后形成的基础样本库，应当明确检测样本类型、样本量、样本分布等信息。检测样本类型应包含焊缝、缺陷、伪缺陷等。检测样本量应当明确样本规模及确定依据，需要考虑样本量不足对算法精度的影响及其风险。检测样本分布应当依据埋藏缺陷和表面缺陷的缺陷方向、缺陷位置、危害程度等因素明确数据分布情况，需要考虑样本分布不均衡对算法精度的影响及其风险。

## 7.5 数据标注

### 7.5.1 标注方法

7.5.1.1 应根据焊接工艺明确标注出焊接缺陷图像中的缺陷类型。

7.5.1.2 应采用矩形边界框或像素分割等方法进行缺陷区域精确标定，矩形框的边界应最小化框选缺陷，对于重叠缺陷，应逐个框选。像素分割标注的区域应完全框选或覆盖缺陷，边界应尽量贴合缺陷的真实轮廓，对于复杂形状的缺陷，分割边界应能清晰区分缺陷与背景。

7.5.1.3 宜对射线检测数字图像中的铅字、定位标记及像质计规格等进行标注。

7.5.1.4 宜对伪缺陷进行标注。

### 7.5.2 标注过程控制

7.5.2.1 应编写标注规程并对标注人员进行培训。标注规程应包括标注对象、标注方法、标注数量、标注工具及复核规定等。

7.5.2.2 标注过程应生成标注记录，包括标注的批次、人员、时间、内容及质量审核结果等。

### 7.5.3 标注结果复核

7.5.3.1 应根据标注规程对标注的结果进行复核。

7.5.3.2 应复核缺陷类型、缺陷位置及标注框边界。

7.5.3.3 应复核图像中所有可见的需标注对象。

7.5.3.4 标注复核人员应对有异议或错误的标注结果进行修改。

## 7.6 数据集构建

7.6.1 基于标注结果构建训练集、验证集和测试集。训练集和验证集样本应具有均衡性，测试集样本分布应符合系统应用要求，训练集和测试集的样本应当无交集。

7.6.2 可对训练集小样本量数据进行扩增。数据扩增应当明确扩增的方式、方法、倍数，并考虑扩增方法选用以及扩增倍数对算法精度的影响。训练集经扩增后，应记录扩增倍数并对比扩增前后样本量、样本分布等方面差异。

其中：

- 扩增的方式，包括：离线、在线等；
- 扩增的方法，包括：翻转、旋转、镜像、平移、缩放、滤波、合成等。

7.6.3 数据集包含图像与标注标签。图像应为无损处理的格式，标签宜为 xml 或 json 格式。

7.6.4 数据集应有版本管理。

## 8 算法模型设计

### 8.1 智能算法模型类型

#### 8.1.1 缺陷特征提取模型

缺陷特征提取模型应至少具备一种特征提取能力，如提取缺陷几何特征、灰度特征、位置特征，并具有特征值的量化分析功能。

#### 8.1.2 缺陷定位模型

缺陷定位模型应根据焊接接头射线检测整幅图像，基于像素跨度与长度对应关系，输出缺陷的像素位置坐标和实际位置坐标。

#### 8.1.3 缺陷类型识别模型

缺陷类型识别模型应根据焊接接头射线检测目标区域图像，输出缺陷的名称。

#### 8.1.4 缺陷分割模型

缺陷分割模型应根据焊接接头射线检测目标区域图像，输出缺陷轮廓或分割掩膜。模型应具备像素级分割以及基于像素跨度与长度对应关系计算缺陷实际物理尺寸的能力。

#### 8.1.5 像质计识别模型

像质计识别模型应具备识别像质计特征、定位像质计区域的能力，输出像质计的型号和可见丝数量。

#### 8.1.6 铅质标识识别模型

铅质标识识别模型应识别并输出焊接接头射线检测整幅图像上的铅字和标记信息。

## 8.2 基础架构和模型选择

宜根据智能算法模型类型及计算资源进行选择：

- a) 选择卷积神经网络或Transformer等深度神经网络架构，当计算资源不能满足时，可选择机器学习模型或轻量化模型；
- b) 缺陷分割模型选择具备像素级分割能力的架构；
- c) 少样本情况可选择预训练模型。

## 8.3 模型训练策略

### 8.3.1 超参数配置

模型训练时关键超参数的配置与优化宜符合下列要求：

- a) 根据模型大小及计算资源选择合适的优化器（如SGD、AdamW等）；
- b) 综合考虑计算资源、训练稳定性、泛化误差等，通过实验确定批次大小；
- c) 采用动态学习率调整策略，根据训练步数设置预热与衰减步数；
- d) 根据模型结构与大小合理配置权重衰减、Dropout比例等正则化参数。
- e) 通过实验确定其他超参数的设置，并予以记录。

### 8.3.2 模型持续学习

模型持续学习宜符合下列要求：

- a) 建立误检与漏检样本的收集与反馈机制，定期将经专家复核的困难样本纳入训练数据集；
- b) 依据新增样本数量选择全量微调或局部微调的持续学习策略。

## 8.4 模型后处理

### 8.4.1 过滤与优化

模型输出结果的过滤与优化宜符合下列要求：

- a) 采用非极大值抑制（NMS）、置信度阈值筛选或基于位置特征等技术，对检测结果进行过滤筛选；
- b) 采用连通域分析、形态学操作等技术，对缺陷的边界轮廓或分割掩膜进行优化处理。

### 8.4.2 融合

当使用多个模型时，宜采用加权平均、投票机制等技术，对分类概率、检测框、分割掩膜等输出信息进行整合。

## 9 系统功能要求

### 9.1 数据管理

数据管理模块应至少包括以下功能：

- a) 支持获取并管理焊接接头射线检测项目及工艺等业务数据，包括：检测项目信息、设备信息、焊接工艺信息、检测工艺信息等；
- b) 支持数据质量校验与验证，包括：底片数字化图像像素校验、位深校验、像质计与灵敏度验证以及伪缺陷筛查等；
- c) 支持对图像数据的预处理，包括：放大、缩小、亮度调节、对比度调节、旋转等；
- d) 支持训练集、验证集、测试集的划分，并对标注数据进行增加、删除、修改、查询；
- e) 支持评定后的数据和图像的查询、统计分析、展示、调取等；
- f) 支持配置智能诊断报告模板；
- g) 支持基于历史底片库创建智能诊断任务。

### 9.2 数据标注

数据标注模块宜包括以下功能：

- a) 支持对标注对象的增加、移除等辅助操作；
- b) 提供用于缺陷标注的图形化工具（如矩形框、多边形框等）；
- c) 支持对图像中缺陷类型、位置与区域进行标注；
- d) 支持对已完成的标注结果进行修改、复核、重置以及标注数据版本管理等；
- e) 支持标注结果的保存、查询及标注数据批量导出等。

### 9.3 训练管理

#### 9.3.1 算法管理

算法管理模块应至少包括以下功能：

- a) 对焊接接头射线检测智能诊断相关算法进行参数配置、部署、运行及管理；

- b) 根据不同场景配置相应算法, 场景特征包括焊接工艺、缺陷类型、检测图像来源(如底片数字化、射线数字成像、计算机辅助成像等)等, 支持按版本、场景、时间等维度进行算法检索与筛选。

### 9.3.2 模型训练

模型训练模块应至少包括以下功能:

- a) 支持自定义训练任务配置, 包括训练集/验证集比例设置、超参数(如优化器、学习率、迭代次数等)调整、正则化策略选择, 满足不同场景下的模型优化需求;
- b) 实时展示训练过程关键指标(如损失值、mAP、F1分数值、超参数等), 对异常情况(如数据异常、资源不足)实时报警。

模型训练模块宜具备迭代训练功能, 保障模型性能持续优化。

### 9.3.3 版本管理

版本管理模块应至少包括以下功能:

- a) 支持设置模型版本唯一标识, 关联训练数据集信息、超参数配置、性能评估结果及适用场景说明;
- b) 支持版本发布和回退功能。

### 9.3.4 环境配置

环境配置模块应至少包括以下功能:

- a) 支持按训练任务配置运行环境参数(如计算资源、内存、存储等);
- b) 支持根据环境资源监控优化配置;
- c) 兼容模型部署框架, 训练后的模型可直接用于智能诊断模块推理。

## 9.4 智能诊断

智能诊断模块应至少包括以下功能:

- a) 识别检测图像中的焊接接头目标区域;
- b) 识别缺陷位置、缺陷类别、缺陷实际尺寸;
- c) 识别的缺陷种类范围包括圆形、条形、内凹、未熔合、未焊透及裂纹缺陷;
- d) 支持焊接接头的质量评级, 并给出评级结果;
- e) 支持可视化显示识别结果, 可在图像上标记显示缺陷位置、类别和尺寸信息。

智能诊断模块宜包括以下功能:

- a) 自动识别像质计信息(如像质计编号、像质计可见丝丝号等);
- b) 自动识别检测图像中铅质标识信息;
- c) 支持批量检测图像。

## 9.5 结果审核与报告生成

### 9.5.1 结果审核

结果审核模块应至少包括以下功能:

- a) 支持对智能诊断模型输出的缺陷位置、类别、尺寸及评级等结果进行审核, 并记录所有过程日志;
- b) 支持将经确认的审核结果纳入标注数据库, 并反馈至模型训练流程, 以用于模型持续优化;
- c) 支持按时间段、缺陷类型等提取和分析反馈数据, 生成统计报告, 为模型优化提供依据。

### 9.5.2 报告生成

报告生成模块宜包括以下功能:

- a) 支持基于智能诊断结果自动生成初步诊断报告;
- b) 报告的内容包括项目信息、设备信息、焊接工艺信息、检测工艺信息、诊断信息等内容;
- c) 支持将诊断报告导出为WORD、PDF等标准格式文件;

- d) 完整记录报告的在线编辑过程，包括编辑人员、编辑时间及修改内容等信息。

## 9.6 系统管理

### 9.6.1 用户及权限管理

用户及权限管理模块应至少包括以下功能：

- a) 支持为用户配置相应角色，包括系统管理人员、评定人员、审核人员、普通用户等；
- b) 支持用户注册、权限设定、维护等；
- c) 支持用户通过浏览器或客户端进行图像分析、缺陷评级等操作，支持多终端接入，实现多终端异地协同工作。

### 9.6.2 安全管理

安全管理模块应至少包括以下功能：

- a) 具备保障数据安全的功能，如支持数据访问控制、数据备份与恢复、数据完整性保护等，必要时进行数据加密控制，适配端、边、云不同部署层次的安全需求；
- b) 具备日志记录与痕迹管理功能，记录用户操作行为，监控系统运行状态，提供可视化日志查询和导出；
- c) 具备网络入侵检测、网络安全防护与远程访问控制功能，应对接入设备进行认证；
- d) 具备应急响应与容灾功能，确保数据全生命周期管理及系统持续稳定运行。

## 10 系统智能诊断性能评估

### 10.1 性能评估流程

焊接接头射线检测智能诊断系统的性能评估流程包括评估准备、评估执行、评估分析和编制评估报告四个阶段，如图1所示。

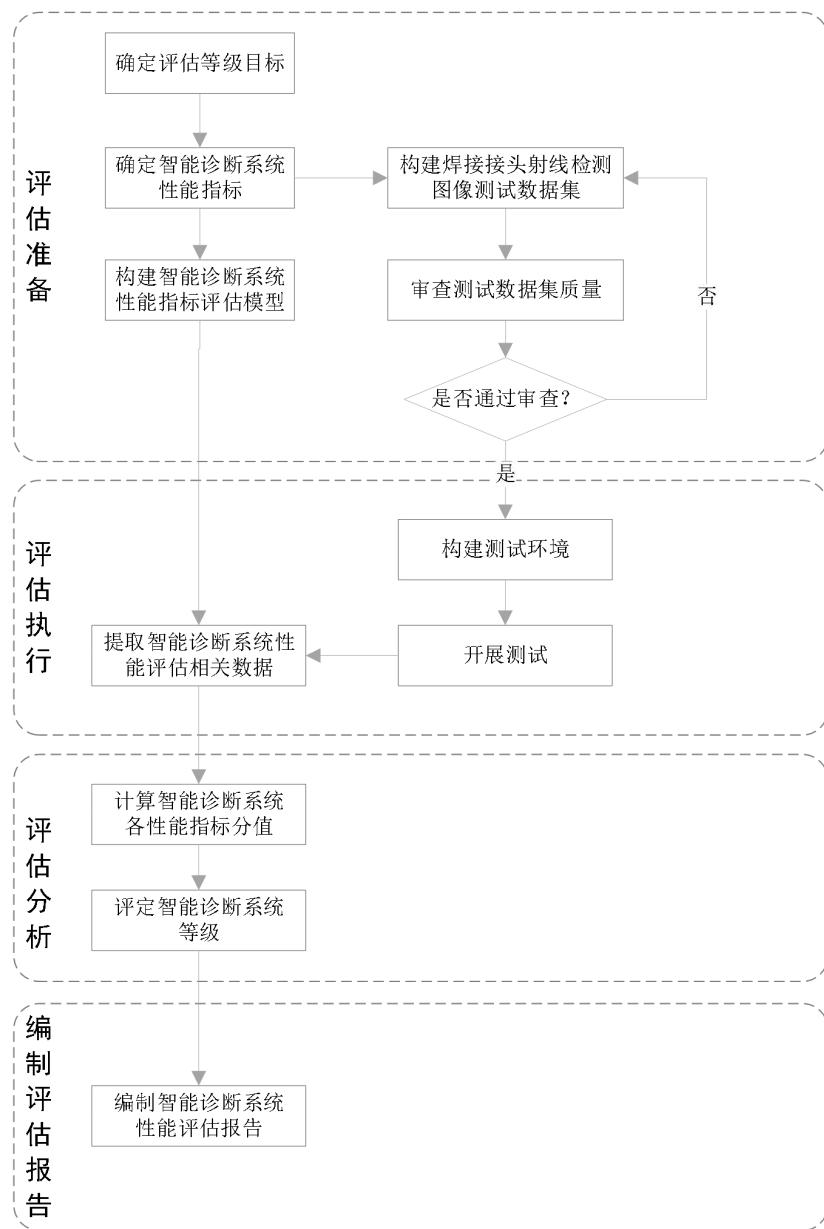


图3 焊接接头射线检测智能诊断系统性能评估流程

## 10.2 评估准备

### 10.2.1 确定评估等级目标

由焊接接头射线检测智能诊断系统的采购方、研发方或第三方测试机构等根据实际情况协商或独立确定评估等级目标。

### 10.2.2 确定性能指标

根据评估等级目标选取所需性能指标，应包括综合缺陷检出率、不可接受缺陷检出率、综合缺陷虚检率、不可接受缺陷虚检率、综合缺陷分类准确率、不可接受缺陷分类准确率、综合缺陷分级准确率、

不可接受缺陷分级准确率和效率。

### 10.2.3 构建性能评估模型

根据选取的性能指标，设置等级阈值。结合功能需求，构建性能评价的评估模型。

### 10.2.4 构建测试数据集

根据评估等级目标，构建测试数据集。测试数据集满足以下要求：

- a) 焊接缺陷类型覆盖评估等级目标，每类缺陷数量应不少于100个样本，同时还应包含不少于缺陷总数10%的无缺陷样本；
- b) 样本宜具有多样性，包括不同焊接工艺、射线检测工艺以及图像参数等要素；
- c) 不同类型的缺陷样本数量宜分布均衡，类型应无争议；
- d) 总样本量应不少于1000个。

### 10.2.5 审查测试数据集

测试数据集审查应包括：

- a) 样本是否存在异常或未标记缺陷等情况；
- b) 标注信息等内容是否存在错误；
- c) 测试数据格式是否规范；
- d) 测试集中样本是否重复。

测试数据集宜审查不同类型的缺陷样本数量分布是否均衡。

当测试数据集通过审查后才能进入“评估执行”阶段。否则，需重建测试数据集，直至通过审查。

## 10.3 评估执行

### 10.3.1 构建测试环境

根据智能诊断系统运行条件，构建与实际资源相同或相近的测试环境，包括搭建服务器、安装操作系统和数据库、配置网络、部署智能诊断模型等。

### 10.3.2 测试

根据选取的性能指标以及智能诊断系统开展测试：

- a) 加载系统，测试系统是否能够正常使用；
- b) 将测试数据集输入系统，获取智能诊断模型的预测结果；
- c) 记录测试运行时间与测试数据集的样本数量。

## 10.4 评估分析

### 10.4.1 计算性能指标分值

10.4.1.1 将智能诊断模型的预测结果与相应标注样本比较，统计检出样本、漏检样本、虚检样本、正确分类样本以及正确分级样本的数量，统计不可接受缺陷检出样本、漏检样本、虚检样本、正确分类样本以及正确分级样本得出数量；

10.4.1.2 计算每种类型的缺陷检出率、缺陷虚检率、缺陷分类正确率和缺陷分级正确率，计算综合缺陷检出率、综合缺陷虚检率、综合缺陷分类正确率和综合缺陷分级正确率，计算不可接受缺陷的缺陷检出率、缺陷虚检率、缺陷分类正确率和缺陷分级正确率；

10.4.1.3 根据测试运行时间和样本数量，计算智能诊断系统的平均响应时间，宜小于5秒。

#### 10.4.2 评定智能诊断系统等级

根据智能诊断模型测试出的性能指标分值及等级阈值将焊接接头射线检测智能诊断系统划分为了L1~L4四个等级，其中L1级为最低等级，L4级为最高等级。各等级划分阈值如表1所示。

表1 焊接接头射线检测智能诊断系统的等级划分阈值表

智能诊断系统能力	性能指标	L1 级	L2 级	L3 级	L4 级
缺陷检出能力	综合缺陷检出率	≥90%	≥93%	≥96%	≥99%
	不可接受缺陷检出率	≥94%	≥96%	≥98%	100%
	综合缺陷虚检率	≤15%	≤12%	≤8%	≤5%
	不可接受缺陷虚检率	≤10%	≤7%	≤4%	≤1%
缺陷分类能力	综合缺陷分类准确率	≥75%	≥84%	≥90%	≥95%
	不可接受缺陷分类准确率	≥83%	≥87%	≥91%	≥97%
缺陷分级能力	综合缺陷分级准确率	≥78%	≥83%	≥90%	≥96%
	不可接受缺陷分级准确率	≥83%	≥87%	≥91%	≥97%

每一项性能指标均应符合相应评估等级的阈值要求，按高级别定级。

#### 10.5 编制评估报告

根据评定级别编制评估报告。报告内容应包括但不限于：智能诊断系统的性能与技术、测试数据集、各项性能指标分值和评定级别说明。

# 《特种设备 人工智能 焊接接头射线检测图像智能诊断系统性能评估规范》（征求意见稿）

## 编制说明

### 1 工作简况

#### 1.1 任务来源

根据《中国特种设备检验协会团体标准工作委员会标准化工作组管理办法》，经过审核，《特种设备 人工智能 焊接接头射线检测图像智能诊断系统性能评估规范》列入中国特种设备检验协会团体标准制定计划，由中国特种设备检测研究院、天津渤海化工集团有限公司、北京远舢智能科技有限公司、北京理工大学、北京巴布科克·威尔科克斯有限公司、上海锅炉厂有限公司、广州帕理检测技术有限公司、上海电气集团股份有限公司中央研究院、新疆维吾尔自治区特种设备检测研究院、福建省特种设备检测研究院、中特检验集团有限公司等单位负责起草。

#### 1.2 制定背景

基于人工智能的焊接接头射线检测图像智能诊断是一种利用计算机视觉技术对焊接质量进行自动化分析与评定的无损检测手段。该技术主要应用于特种设备、石油化工等能源行业，涵盖压力容器、管道、船舶、电力设施、油气储运等设备的制造、安装及在役检测环节。近年来，为响应国家推动人工智能与产业融合的战略部署，提升射线检测的评定效率、规范性与准确性，解决传统人工评定效率低、主观性强、知识传承难等问题，该技术已在行业中得到快速推广，市场上已出现多种相关产品。然而，目前尚未建立针对此类AI系统的性能评估规范，导致以下几方面问题凸显：一是评价指标不统一，依赖厂家自主定义，可比性差；二是产品质量参差不齐，诊断效果难以保证；三是用户无法全面、客观评价系统性能，在实际应用中难以匹配需求，影响技术信任度；四是缺少标准支撑，用户对新技术的接受度受限；五是反过来制约了该技术的进一步推广与持续创新。

因此，本项目拟通过研制该系统的性能评估规范，在行业内建立统一标准，明确系统功能要求、性能指标、评估流程与方法，从而推动产品质量提升、促进市场健康发展、激励技术创新、助力行业智能化转型、支撑能源设备智慧管理，并在必要时为基于人工智能的质量评定方式提供立法保护基础，最终促进相关领域经济的高质量发展。

#### 1.3 起草过程

##### 1.3.1 标准启动会

2024年9月3日，由中国特种设备检测研究院组织的《特种设备 人工智能 焊接接头

射线检测图像智能诊断系统性能评估规范》标准启动会在线上召开。来自全国多个行业领域的代表，包括国内外主要的射线检测智能诊断技术研发团队、科研院所、行业协会、检测单位以及其他相关企业的 24 家单位共 33 位代表参加了会议。

依照会议议程安排，会议针对“焊接接头射线检测图像智能诊断系统技术规范”的内容框架进行了研讨和梳理，确定标准框架包括：1) 范围；2) 规范性引用文件；3) 术语和定义；4) 缩略语；5) 一般要求（5.1 人员要求，5.2 系统运行环境要求）；6) 系统结构与流程；7) 数据的要求；8) 算法模型设计要求；9) 系统功能要求；10) 系统智能诊断性能评估（10.1 系统评估指标体系，10.2 性能评估流程与方法）。并且明确标准的内容和关键技术指标制定需要综合考虑需求多样性、技术兼容性、行业通用性等多方面需求，逐条研讨草案内容及关键技术指标。

按照整体标准内容结构成立六个工作组：系统运行环境要求组、系统结构与流程组、数据要求组、算法/模型要求组、系统功能要求组和系统评估组。以工作组的方式对标准中的内容进行分章节撰写，最终将各个工作组的撰写内容合并，形成初稿。

### 1.3.2 征求意见稿

2024 年 10 月至 2025 年 12 月，标准工作组随标准内容编制同期在中国特种设备检测研究院、福建省特种设备检测研究院、天津渤海化工集团有限公司、广州帕理检测技术有限公司、机械工业上海蓝亚石化设备检测所有限公司等机构对标准中涉及的方法，对射线检测图像智能诊断系统性能评估流程、方法及指标进行逐条验证，排除实际操作中可能遇到的问题和歧义。保障了标准中内容的可操作性，形成征求意见稿初稿。

## 2 标准编制原则、主要内容及其确定依据

### 2.1 编制原则

保证标准的适用性；保持标准的先进性；注意标准的统一性和协调性；结合我国国情并考虑标准的经济性和社会效益。

### 2.2 编制要求

编写标准草案时要在充分调查研究的基础上，认真分析国内外同类技术标准的技术水平，在预期可达到的条件下，积极地把先进技术纳入标准，提高技术水平。编制过程中要注意符合法律法规的规定以及与相关标准协调，避免与法律法规、相关标准之间出现矛盾，给标准的实施造成困难。

### 2.3 主要内容

本标准规定了基于人工智能的焊接接头射线检测图像智能诊断系统的技术规范，包括：一般要求（人员要求，系统运行环境要求）、系统结构与流程、数据的要求、算法/模型设计要求、系统功能要求、系统智能诊断性能评估（系统评估指标体系，性能评估流程与方

法)等内容。

## 2.4 确定依据

本标准编制的主要依据是,全面涵盖焊接接头射线检测图像智能诊断系统在新建、二次开发与实际应用过程中所涉及的关键要素,包括:人员、运行环境、系统结构与流程、数据、算法与模型、系统功能、性能评估指标及评估流程与方法。以此构建的标准技术体系,力求逻辑清晰、内容完整、可操作性强,从而保障标准的有效贯彻与实施。

## 3 试验验证的分析、综述报告,技术经济论证,预期的经济效益、社会效益和生态效益:

### 3.1 试验验证的分析

在征求意见稿起草过程中,起草小组对标准中描述的方法,进行了试验验证,证实可以根据标准所描述的方法对涡流阵列检测系统进行测试和评价。

### 3.2 预期的经济效益、社会效益和生态效益

随着“十四五”规划的推进,加快发展现代产业体系、巩固壮大实体经济根基已成为国家战略。装备制造业是实体经济的重要组成部分,其高质量发展直接关系到我国经济高质量发展的全局。焊接接头射线检测图像智能诊断技术作为人工智能与无损检测深度融合的典型应用,在提升检测效率、保障设备安全、推动制造业智能化转型方面具有重要意义。

本标准的制定,将填补该领域行业标准空白,对我国能源及承压设备行业焊接接头射线检测图像智能诊断系统的研发、生产、应用与评价产生积极影响。经济效益方面,标准将统一系统性能评价指标,降低企业研发与采购成本,促进产品质量提升和技术迭代,推动该技术在制造业、能源等领域的规模化应用,提高检测效率与一致性,减少因漏检、误检导致的设备故障与停机损失,助力行业降本增效。社会效益方面,标准有助于规范市场秩序,增强用户对智能诊断系统的信任度,推动检测行业向智能化、标准化方向发展,提升特种设备安全管理水平,保障能源、化工等重要领域设备运行安全,促进高技能人才队伍建设与技术传承。生态效益方面,通过推动数字化、智能化检测技术普及,可逐步减少对传统胶片等耗材的依赖,降低检测过程中的物料消耗与废弃物产生,助力行业绿色低碳转型。

## 4 与国际、国外同类标准技术内容的对比情况,或者与测试的国外样品、样机的有关数据对比情况:

目前国际上没有针对焊接接头射线检测图像智能诊断系统的标准,仅有 ASTM E3327/E3327M-21 Standard Guide for the Qualification and Control of the Assisted Defect Recognition of Digital Radiographic Test Data 标准,该标准针对数字射线检测数据辅助缺陷识别计算机算法的通用系统鉴定方法,以过程为导向建立生命周期管理框架,提供从概念、性能定义、数据收集、鉴定、控制到维护的完整循环。然而,该标准不具备能源行业承压设

备焊接接头射线检测评价的专业属性，较难直接实现最终能源行业场景的应用。而 NB/T 专门针对承压设备焊接接头的射线检测智能诊断系统，行业和应用目标明确，给出了更加明确具象的系统要求，并且给出了缺陷检出率、虚检率、分类和分级准确率等具体指标以及详细的系统等级划分阈值表。同时，注重与 NB/T 47013 系列标准的协调与配合。

5 以国际标准为基础的起草情况，以及是否合规引用或者采用国际国外标准，并说明未采用国际标准的原因：

目前国际上没有针对焊接接头射线检测图像智能诊断系统的标准，仅有 ASTM E3327/E3327M-21 Standard Guide for the Qualification and Control of the Assisted Defect Recognition of Digital Radiographic Test Data 标准，该标准针对数字射线检测数据辅助缺陷识别计算机算法的通用系统鉴定方法，以过程为导向建立生命周期管理框架，提供从概念、性能定义、数据收集、鉴定、控制到维护的完整循环。然而，该标准不具备能源行业承压设备焊接接头射线检测评价的专业属性，较难直接实现最终能源行业场景的应用。故未采用该标准，对于标准内容根据行业需求，进行整体重构与适应性编制。

6 与有关法律、行政法规及相关标准的关系：

本标准与我国的现行法律、法规和强制性国家标准不存在冲突。

当前，国内在基于人工智能的焊接接头射线检测图像智能诊断系统领域，尚未发布专用的国家或行业标准。这一标准空白导致在系统开发、应用与评估等关键环节缺乏统一的技术依据，进而导致开发目标不明，应用实践无章可循，性能评估难以客观衡量。为有效引导技术发展、保障检测质量与可靠性，亟需制定本标准以填补该领域标准体系的空缺。

7 重大分歧意见的处理经过和依据：

无

8 专利的有关说明：

无

9 国家标准的要求，以及组织措施、技术措施、过渡期和实施日期的建议等措施建议：

本标准是新制定标准，应在行业内进行充分的征求意见，吸收行业内焊接接头射线检测图像智能诊断系统厂商、系统用户、检验检测机构、科研院所的经验反馈，推动产品质量提升、促进市场健康发展、激励技术创新、助力行业智能化转型、支撑能源设备智慧管理。

10 其他应当说明的事项。

本标准适用于 NB/T 47013.2、NB/T 47013.11 和 NB/T 47013.14 所涉及的承压设备射线检测范围，适用的焊接接头射线检测图像包括：射线照相法获得的底片数字化图像、射线数

字成像检测获得的射线检测图像和射线计算机辅助成像检测获得的射线检测图像。

本标准适用于承压设备以外的其他金属材料、支承件和结构件的焊接接头射线检测。

NB/T 47013.2、NB/T 47013.11 和 NB/T 47013.14 标准未覆盖领域的射线检测智能诊断可参考本标准执行。

特种设备 人工智能 焊接接头射线检测图像智能诊断系统性能评估规范

标准工作组

2026 年 02 月 03 日