

团 体 标 准

T/CESA 1109—2020

智能医疗影像辅助诊断系统技术要求和测试评价方法

Technical requirements and testing and evaluation methods for intelligent medical image aided diagnosis system

2020-07-30 发布

2020-08-01 实施



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

目 次

前 言.....	IV
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 缩略语.....	2
5 功能结构和处理流程.....	3
5.1 功能结构.....	3
5.2 处理流程.....	3
6 功能要求.....	4
6.1 数据管理.....	4
6.2 影像处理.....	5
6.3 模型管理.....	5
6.4 诊断评估.....	6
6.5 人机协同.....	7
6.6 系统管理.....	7
7 数据要求.....	8
7.1 超声影像.....	8
7.2 数字组织病理影像.....	9
7.3 MRI.....	10
7.4 CT 影像.....	10
7.5 X-ray 影像.....	11
7.6 核医学影像.....	12
7.7 心电图.....	13
7.8 脑电图.....	14
8 测试评价方法.....	16
8.1 测试评价流程.....	16
8.2 确定评估目标.....	17
8.3 建立评估模型.....	17
8.4 选取测试指标.....	17
8.5 构建基准测试数据集.....	19
8.6 构建测试环境.....	19
8.7 开展测试活动.....	19
8.8 收集测试指标数据.....	20
8.9 得出评估结论.....	20

附录 A	(资料性附录)	胸部 CT 辅助诊断系统.....	21
附录 B	(资料性附录)	智能医疗肺部 CT 辅助影像诊断系统.....	25

全国团体标准信息平台

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这些专利的责任。

本文件由中国电子技术标准化研究院提出。

本文件由中国电子技术标准化研究院、中国电子工业标准化技术协会归口。

本文件起草单位：中国电子技术标准化研究院、四川大学华西医院、杭州依图医疗技术有限公司、中国科学院软件研究所、中国医学科学院生物医学工程研究所、北京邮电大学、中国医学科学院北京协和医院、中国人民解放军总医院、中国医学科学院阜外医院、中南大学湘雅医院、上海市儿童医院、中国人民解放军空军特色医学中心、成都知识视觉科技有限公司、沈阳东软智能医疗科技研究院有限公司、成都索贝数码科技股份有限公司、四川久远银海软件股份有限公司、北京推想科技有限公司、电子科技大学、国家超算长沙中心（湖南大学）、华为技术有限公司、上海计算机软件技术开发中心、四川希氏异构医疗科技有限公司、云从科技集团股份有限公司、中国科学院计算技术研究所、北京航空航天大学、锦天城律师事务所、成都智信信息技术有限公司、北京安德医智科技有限公司、香港大学深圳医院、中科院软件研究所南京软件技术研究院。

本文件主要起草人：张伟、殷晋、李真林、周平、董建、张群、马珊珊、颜子夜、黄皓、孟令中、薛云志、邱甲军、蒲江波、王枞、茅枫、李立安、严欣欣、李文政、杨秀军、王新宴、彭玉兰、朱庆莉、李春漾、包骥、吴小艾、向飞、程万军、曾筱茜、温序铭、吴钊、王少康、秦科、徐圣普、罗妍、张唯一、刘俊、刘文广、李婷婷、雷震、曾欢、赵赫、王炜、郭炫志、邓甜、彭绍亮、宋捷、符海芳、陈敏刚、雷翔、李军、赵屹、刘祥龙、叶飞、宋井宽、陈曦、王俊人、蒋静文、武斌、李婷、王智宇、张燕娜、陈伟、黄寅、王树林、胡勇、吴玉美。

智能医疗影像辅助诊断系统技术要求和测试评价方法

1 范围

本文件规定了计算机视觉领域的智能医疗影像辅助诊断系统的基本功能结构和功能要求、影像数据要求、临床测试评价方法等。

本文件适用于医疗机构、研究机构、企业等对智能医疗影像辅助诊断系统的设计、研发和管理，其他相关领域可参考使用。

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 35273—2020 信息安全技术 个人信息安全规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

计算机辅助诊断 computer aided diagnosis

利用计算机分析疾病相关的信息，以对疾病进行识别、预后及对疗效进行评价的过程等。

3.2

医疗影像辅助诊断 medical image aided diagnosis

基于医学图像定量数据分析与挖掘，通常为基于影像组学相关知识与技术手段分析疾病相关的信息，以对疾病进行识别、预后、以及对疗效进行评价等。

注：基于医学影像进行定量数据分析与挖掘的相关技术、研究方法等的一个专业领域。

3.3

感兴趣区 region of interest

医学影像中与研究问题相关或被用户所关注的区域或图像序列范围。

3.4

分割 segmentation

从医学图像中截取所需要的区域。

注：通常是指截取感兴趣区，可包括全自动分割、半自动分割、手动分割等。

3.5

数据挖掘 data mining

对提取的特征进行数据定量分析的过程。

注：一般通过包括分类、回归、生存分析、风险模型评估等方法来实现。

3.6

准确率 accuracy

分类或预测等结果中被正确预报的样本数量占整个样本数量的比率。

3.7

敏感度 sensitivity

实际患病又被诊断标准正确地诊断出来的百分比，阳性且称正类。

3.8

特异度 specificity

与敏感度类似，阴性类中被准确诊断为阴性的样本所占阴性数量的百分比，阴性且称负类。

3.9

辨别力 discrimination power

评估方法区分正样本和负样本的程度。

3.10

过拟合 overfitting

为了得到一致假设而使假设变得过度严格。

4 缩略语

下列缩略语适用于本文件。

18F-FDG: 氟代脱氧葡萄糖 (Fluorodeoxyglucose)

AI: 人工智能 (Artificial Intelligence)

AUC: 接受者操作特性曲线下的面积 (Area Under ROC)

BMP: 图像文件格式 (Bitmap)

CAD: 计算机辅助诊断 (Computer Aided Diagnosis)

CT: 计算机断层成像 (Computed Tomography)

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

DTPA: 二乙基三胺五乙酸 (Diethylenetriaminepentaacetic acid)

EDF: 欧洲数据格式 (European Data Format)

EEG: 脑电图 (electroencephalograph)

FDA-XML: 美国食品及药物管理局制定的可扩展标记文件格式 (FDA-XML Data Format, 又称FDADF)

FLAIR: 磁共振成像液体衰减反转恢复序列 (Fluid Attenuated Inversion Recovery)

fNIRS: 近红外功能成像 (functional Near-Infrared Spectroscopy)

HL7: 美国食品及药物管理局制定的结构化产品标签格式 (GUDID Health Level 7)

HU: 亨氏单位 (Hounsfield Scale)

JPEG: 静止图像压缩标准/联合图像专家组 (Joint Photographic Experts Group)

mAs: 毫安秒 (Miliampere seconds)

MAT: Matlab数据文件格式 (MATlab data format)

MRI: 磁共振影像 (Magnetic Resonance Imaging)

MI: 机械参数 (Mechanical Index)

PDF: 可携式文件格式 (Portable Document Format)

PET: 正电子扫描 (Positron Emission Tomography)

PACS: 影像归档和通信系统 (Picture Archiving and Communication Systems)

ROI: 感兴趣区 (Region Of Interest)

ROC: 接受者操作特性曲线 (Receiver Operating Characteristic Curve)

SD: 标准差 (Standard Deviation)

SID: 摄影距离 (Source-Image Distance)

SPECT: 单光子发射计算机断层成像术 (Single-Photon Emission Computed Tomography);

TDMS: 技术文档管理系统 (Technical Document Management System)

TXT: 文本格式 (Text)

X-ray: X射线 (Roentgen ray)

5 功能结构和处理流程

5.1 功能结构

智能医疗影像辅助诊断系统应包括: 数据管理、影像处理、模型管理、诊断评估、人机协同子系统以及系统管理子系统。基本功能结构见图1。

- 系统完成智能医疗影像辅助诊断应包括(以数据处理的不同阶段为主要排序依据): 数据管理、影像处理、模型管理、诊断评估、人机协同子系统;
- 系统还应包括完成一般性管理的系统管理子系统。



图1 智能医疗影像辅助诊断系统基本功能结构

5.2 处理流程

智能医疗影像辅助诊断系统的完成辅助诊断的基本流程和处理步骤见图2。

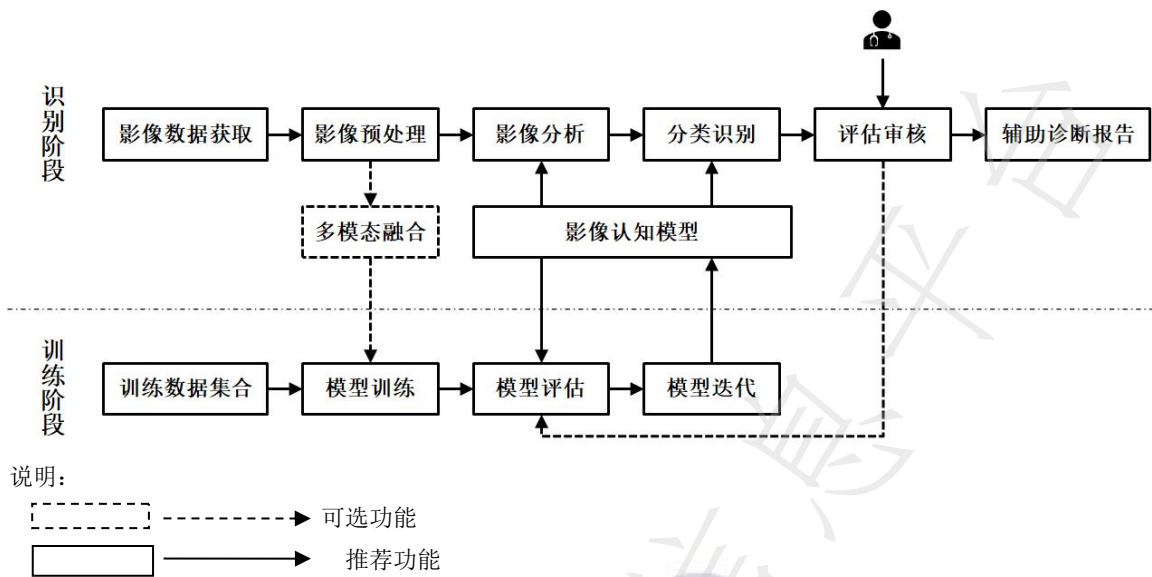


图2 智能医疗影像辅助诊断系统基本处理流程

系统的处理过程参照一般模式识别、机器学习系统的框架，应分为识别和训练两个阶段；

- 系统的识别阶段是进行智能辅助诊断的核心流程，应包括影像数据获取、影像预处理、影像分析、分类识别、评估审核、辅助诊断报告步骤，其中评估审核阶段应有使用者对系统输出的结果进行判读、确认或者修改；
- 系统的识别阶段中应包含影像认知模型处理步骤作为影像分析和分类识别的底层处理支撑；
- 系统的识别阶段中包含多模态融合处理步骤（虚线框表示为可选步骤，非必须），作为影像预处理后续的一个分支操作，并为训练阶段的模型训练提供输入；
- 系统训练阶段是利用各种数据和算法建立影像认知模型的过程，应包括训练数据集数据获取、模型训练、模型评估、模型迭代处理步骤；
- 系统训练阶段中模型训练处理可由识别阶段多模态融合处理的结果作为输入之一；
- 系统训练阶段中模型评估可由识别阶段的评估审核作为输入，根据评估审核结果决定是否进入模型评估（虚线表示为可选步骤，非必须）；
- 系统中影像认知模型、模型训练、模型评估、模型迭代应形成一个闭环，以持续的改进认知模型的性能，改进过程可在线自动完成、也可在系统版本控制下完成。

6 功能要求

6.1 数据管理

6.1.1 影像源管理

影像源管理符合下列要求：

- 系统应支持导入源自一种或多种影像源的影像数据，宜包括但不限于MRI、CT、X-Ray、超声、核医学影像，心电图、和脑电图；
- 系统可根据影像源类型和辅助诊断任务，实现影像质量的自动检测，输出影像数据质量报告，向用户提示并过滤不合格样本，提高数据可用性。

6.1.2 数据接口

数据接口符合下列要求：

- a) 系统应支持一种或多种影像文件格式的处理；
- b) 系统宜支持接入电子病历和实验室检查结果等结构化临床数据；
- c) 数据接口可通过系统直接支持对特定文件格式的读取和存储，也可通过调用第三方组件实现；
- d) 文件格式宜包括但不限于 DICOM、JPEG/JPG/JPE、TIFF/TIF、BMP、XML、TXT、CSV 等。

6.1.3 隐私保护

系统应支持对接入数据的脱敏处理，在数据管理的接入、存储、读取等各环节保护患者的知情权和隐私权，并防止数据篡改和个人敏感信息泄露。具体应符合GB/T 35273—2020的相关规定。

6.1.4 数据后处理

数据后处理符合下列要求：

- a) 系统宜支持对数据集划分、数据复用和数据销毁功能；
- b) 数据集划分功能宜支持对训练集、验证集、测试集的分类和标签管理。

6.2 影像处理

6.2.1 影像预处理

影像预处理符合下列要求：

- a) 系统可包含针对导入的影像数据预处理功能，可根据不同诊断方向，不同影像类型，由用户自行选择预处理功能；
- b) 预处理可包括灰度变换、色调调整、图像增强、边缘检测等。

6.2.2 影像分析

影像分析是智能医学影像辅助诊断的核心子系统之一，包含但不限于下列功能：

- a) 系统可具备一种或多种特征提取功能，提取影像中器官、病灶、ROI 区域的颜色、纹理、形状等特征；
- b) 系统可具备一种或多种对影像的分割功能，对器官、病灶、ROI 区域进行自动、半自动化分割，达到区分病灶或器官的目的，为定性或定量辅助分析提供基础；分割算法可包括基于图像处理的灰度或语义等方法，或基于机器学习的分割方法。

6.2.3 多模态融合

多模态融合符合下列要求：

- a) 系统宜具备影像配准、异构数据融合、可视化等功能，实现医学影像的融合，关联来自各模态的数据特征，实现必要的转换和误差评估，为增强影像认知模型的能力提供更全面的特征；
- b) 针对来自不同成像和记录设备的异构数据、影像的元数据、文本化的病历数据，系统宜具备结构化和非结构化数据融合功能，例如：可将 CT、核磁等影像数据与脑电等时间序列数据进行融合，对齐事件和特征序列等；
- c) 对于具备多模态数据可视化功能的系统，宜提供将来自多个来源或多个序列的影像和其他格式数据进行特征对齐和浏览的功能。

6.3 模型管理

6.3.1 模型训练

模型训练模块接受影像处理子系统输出的影像数据，选择和配置机器学习算法，进行学习训练和评估，生成认知模型，符合下列要求：

- a) 应提供在线、离线或其他多样的训练数据的获取方式，供模型训练者进行认知模型的迭代训练与升级；
- b) 模型训练宜定期获取新增数据集，对模型进行持续训练、验证、发布更新。

6.3.2 模型库

模型库模块对系统中各类影像认知模型及其版本进行管理，符合下列要求：

- a) 应对系统内影像认知模型的上下文信息进行管理与控制；
- b) 应提供审核机制和版本管理机制，支持认知模型的升级与回退。

6.3.3 版本管理

版本管理模块实现对某一特定影像认知模型发布、更新和终止等过程的管理，符合下列要求：

- a) 版本发布应包含版本标识，变更记录，及其他必要上下文信息；
- b) 版本发布应有发布权限控制，由有权限的发布者设置和管理认知模型的同步发布、异步发布及具体细节；
- c) 应提供必要且合规的审核流程，对不符合规定和要求的认知模型发布操作宜提供禁止机制，审核流程应参照医疗器械质量管理体系进行。

6.4 诊断评估

6.4.1 评估任务

系统评估任务主要为对智能医疗影像辅助诊断系统的功能实现程度、泛化能力、鲁棒性、以及安全性等进行自我评价。包含如下评估任务：

- a) 功能实现程度与泛化能力的评估应包括准确率、敏感度、特异度、错误率、辨别力、AUC、F值、过拟合要求等；
- b) 鲁棒性评估应包括异常数据情况下的准确率、敏感度、特异度、AUC等；
- c) 安全性评估是指系统保护信息和数据的程度，应包括评估系统是否具有与其授权类型和授权级别一致的数据访问控制、脱敏处理要求、加密要求、应急处置要求等；
- d) 其他评估应包括响应时间、稳定性（应包括系统崩溃、无逻辑错误等系统漏洞）要求。

6.4.2 分类识别

智能医疗影像辅助诊断系统包含分类、识别或分割的能力：

- a) 能够自动对病灶或器官进行标记，如肺结节、肺叶等；
- b) 标记能够达到像素级，对疑似病灶进行分类识别；
- c) 分割处理的输入为经过预处理后的影像或者影像序列，经过处理后输出为分割后图像或者记录分割点信息的数据结构，标记处理后保存原始的输入影像。

6.4.3 预测分析

智能医疗影像辅助诊断系统应具有评估预测分析的能力，宜包含以下预测分析结果：

- a) 标示疑似病灶的位置；
- b) 计算疑似病灶的大小；
- c) 分析疑似病灶的性质。

6.5 人机协同

6.5.1 显示交互

显示交互模块通过可视化界面进行人机交互和协作，符合下列要求：

- a) 交互模式自然高效，具备良好可用性；
- b) 系统应提供对医学影像感兴趣区突出显示；
- c) 系统对病灶影像应给出文字诊断说明。

6.5.2 评估审核

评估审核模块支持医疗专家对辅助诊断结果进行审核、修订和确认，符合下列要求：

- a) 评估审核指标应满足功能指标要求，见 6.4.1；
- b) 应支持医生对模型训练结果（感兴趣区）、文字诊断说明修改，确认，并保留修改日志；
- c) 系统宜将医疗专家的修订结果在经过质量评估审核后，纳入至标注数据库，并反馈至模型评估流程，支持按时间阶段提取；
- d) 审核过程应由至少一位其他具备该影像领域知识的专家对医生诊断结果进行审核，确认。

6.5.3 报告输出

报告输出模块支持在医疗专家的最终决策和确认下市场诊断报告，符合下列要求：

- a) 应支持自动生成诊断报告，经由医疗专家修改或确认后对外发布；
- b) 报告输出内容宜包括：门诊信息、患者信息、指标说明或影像描述、诊断结论、建议及相关医疗人员说明等；
- c) 诊断报告术语应符合标准医学术语规范，内容输出宜采用结构化模板。

6.6 系统管理

6.6.1 任务管理

任务管理功能包括但不限于：

- a) 系统应具备至少一种类型的影像分析任务调度功能；
- b) 系统应具备批量影像分析任务的执行驱动，执行状态收集和任务属性查看功能；
- c) 系统应具备一个以上的执行器实例接入和状态监控功能；
- d) 系统可具备影像分析任务的优先级调整、失败任务重做和历史任务查询功能；
- e) 系统可具备多种类型影像分析任务的流程编排功能；
- f) 影像分析任务触发方式宜支持定时触发和事件触发。

6.6.2 访问控制

访问控制功能包括但不限于：

- a) 系统应具备使用用户的认证管理功能，使用用户可实名认证，宜具备双因素认证、复杂密码策略和密码定期修改功能；
- b) 系统应具备多级功能权限设定功能，至少应分为管理员和一般用户两级授权机制，实现不同等级用户的功能使用权限控制，防止非授权人员使用非授权功能；
- c) 系统可具备以影像分析任务为对象的数据权限功能，实现用户间的数据可见性隔离；
- d) 系统功能权限类别可包括但不限于：数据操作类、任务管理类和安全审计类。

6.6.3 安全管理

安全管理功能包括但不限于：

- a) 系统应恪守医学科研伦理；
- b) 系统应保障数据的保密性、完整性、可用性。应采取数据分类、重要数据备份、加密认证等措施保证数据安全；
- c) 系统应具备日志功能，规范数据接入、使用和销毁过程进行痕迹管理，确保对数据访问行为可管、可控；对服务管理的全流程留痕，对安全、隐私、风险及事故可查询、可追溯；
- d) 系统宜具备数据生命周期管理功能，可针对不同类型数据设定备份或销毁策略；

6.6.4 运维管理

运维管理功能包括但不限于：

- a) 系统可具备可视化的系统运作状态监控功能，系统运行状态包括但不限于关键服务状态、计算资源状态和各类任务概况；
- b) 系统可具备便捷的核心服务启停和系统启停功能，提高系统运维效率；
- c) 系统核心服务宜具备高可用性；
- d) 系统宜具备系统日志可视化查询和导出下载功能。

7 数据要求

7.1 超声影像

7.1.1 应用场景

超声影像应用场景包括：

- a) 浅表器官及血管病变（临床较多为乳腺、甲状腺）；
- b) 腹部脏器实体肿瘤（临床较多为肝脏、胰腺、肾、子宫及卵巢）；
- c) 妇科检查（子宫双附件）。

7.1.2 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP等。如有RF信号数据，亦可保存为二进制RF格式。

7.1.3 仪器设置与成像操作

仪器设置要求应包括：

- a) 浅表器官检查：选用高频线阵宽频带探头，频率范围通常为 5 MHz~18 MHz，中心频率 10 MHz 或以上，根据病变部位及深度程度灵活调节成像频率，探头穿透深度达到 5 cm；
- b) 腹部器官检查：选用低频凸阵宽频带探头，频率范围通常为 3 MHz~5 MHz，穿透深度达到 18 cm；
- c) 妇科检查（子宫双附件）：选用频率范围为 5 MHz~9 MHz 凸阵腔内探头，穿透深度 10 cm。

成像操作要求应包括：

- a) 灰阶超声检查
 - 1) 当评估病灶时，聚焦区应置于病灶的中央；
 - 2) 增益水平应该能够显示灰阶渐变，以分别显示局部组织的各个解剖成分，先根据实际显示灰度调节总增益，然后观察整个图像，适当调节时间增益补偿，使图像的浅部、中部和深部灰度相似；
 - 3) 肿物测量：寻找病灶的最长轴，然后获得两幅正交的图像并测量，测量肿块时沿肿块的长轴测量；

- 4) 容积数据获取：对于浅表器官肿物，采用三维图像采集装置，沿肿物长轴，获取三维图像。
- b) 彩色/能量多普勒超声检查：浅表器官肿瘤调节速度标尺通常设置为 1 cm/s~5 cm/s；腹腔脏器肿瘤速度标尺可根据肿瘤血供情况进行调节，一般设置为 5 cm/s~18 cm/s；彩色增益和壁滤波的调节以显示最多的彩色血流信号，但同时不出现彩色噪声为基本原则。
- c) 脉冲多普勒超声：调节频谱高度处于图像显示高度的 2/3 左右水平。在测量甲状腺上动脉频谱时，需将声束-血流夹角（ θ 角）控制在 60 度以下。
- d) 超声弹性成像（浅表器官肿瘤）：探头轻触皮肤，尽量保持探头与皮肤垂直，调节弹性成像 ROI 大小及位置，将结节置于 ROI 中心位置，ROI 需包括肿物周围部分正常腺体；二维剪切波弹性成像时，无需压放探头，嘱患者屏住呼吸，待弹性图像稳定，无明显伪像时冻结图像；点式剪切波弹性测量时，也无需压放探头，将 ROI 置于结节相应位置，嘱患者屏住呼吸，然后触发控制键获取数据。
- e) 超声造影：调节机械指数 MI 至 0.1 以下，同时调节成像频率、焦点位置、总增益、TGC 曲线、成像深度和成像宽度等，以达到最佳成像效果。注意成像时探头不能加压。

7.1.4 影像质量

最低分辨率要求应为 72ppi，图像尺寸应为 1024 像素×768 像素，图像文件大小应不小于 768 kb。图像清晰、均匀性、超声切面标准、图无斑点，雪花细粒，网纹等干扰。

7.2 数字组织病理影像

7.2.1 应用场景

数字组织病理影像应用场景包括：常规组织学诊断、冰冻诊断、免疫组化、免疫荧光、细胞学诊断等。

7.2.2 数据格式与存储

数据文件格式应包括：Generic tiled TIFF、JPEG、BMP、JPEG2000 等。

7.2.3 仪器设置与成像操作

仪器设置与成像操作要求应包括：

- a) 图像采集倍率：常规情况使用 20× 或 40× 物镜进行 200 倍~400 倍放大扫描，特殊情况根据实际需要选择更大倍数物镜扫描；
- b) 扫描图像对焦要求：切片扫描所使用的数字扫描仪器需要具有自动对焦模式。扫描图像要求对焦准确，组织结构、细胞形态及边缘保留相对完整。扫描图像对焦不准的表现为组织或细胞结构虚化、重影等；
- c) 多层扫描要求：在扫描图像中，应具备多层扫描功能，以便在重叠现象出现时进行多层图像融合观察。

7.2.4 影像质量

影像质量要求应包括：

- a) 切片完整情况：切片完整、无破损情况；
- b) 切片染色情况：染色保持鲜亮，视觉感知明显，能清晰辨识细胞大小及形状；
- c) 细胞以及核边界清晰；
- d) 包含细胞分布区域 95% 以上；

- e) 未引入涂片不包含的图像噪声，无外部噪声影响阅片诊断的。

7.3 MRI

7.3.1 应用场景

MRI应用场景包括：

- a) 中枢及外周神经系统病变及肿瘤鉴别和分级；
- b) 心血管系统病变检出及心功能评估；
- c) 肝胆胰病变检出及肿瘤鉴别；
- d) 泌尿系统及前列腺病变检出及肿瘤分级；
- e) 胃，结直肠病变检出和鉴别；
- f) 骨组织病变及肿瘤鉴别；
- g) 关节、软骨损伤评估；
- h) 乳腺肿瘤鉴别与分级。

7.3.2 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP等。

7.3.3 仪器设置与成像操作

根据扫描部位和目的合理选择线圈、射频脉冲序列等参数，成像应完整包含靶器官和病变部位。对具体图像成像标准推荐扫描序列如下：

- a) 头部扫描应包含 T1，T2，FLAIR，增强 T1 图像；
- b) 心脏扫描应根据心脏结构从多个方向进行 T1，T2，Cine 序列扫描；
- c) 肝胆胰等器官扫描应包含 T1（压脂和非压脂），T2，增强 T1，弥散加权。有条件可加扫动态增强；
- d) 乳腺、前列腺扫描应包含 T1，T2，增强 T1 和弥散加权图像；
- e) 关节、软骨扫描应包含 T1（压脂和非压脂），T2，增强 T1 图像。

7.3.4 影像质量

影像质量要求应包括：

- a) 应使用原厂水模和规程定期对磁共振成像设备性能进行质控；
- b) 常规图像无明显畸变，弥散加权图像畸变在正常范围内；
- c) 图像信号均匀，无明显信噪比异常，无明显射频干扰及其他各种形式的伪影；
- d) 患者配合良好，图像无明显运动伪影；
- e) 图像平面内分辨率应满足临床诊断要求；3D 头部扫描层厚不超过 1 mm，头部普通扫描层厚不超过 3 mm；心脏、腹部扫描层厚不超过 5 mm；关节、软骨扫描层厚不超过 3 mm。磁共振检查设备建议说明 1.5 T 及以上；扫描层厚要求过薄，泛化推广欠佳，大量临床数据将被排除在外；建议增加层间距的参数；建议增加图像质量评分标准。

7.4 CT 影像

7.4.1 应用场景

CT影像应用场景包括：

- a) 全身各部位新生物的检出，定位、定性诊断，肿瘤分布范围，浸润和转移及 CT 引导下的活检；

- b) 全身各部位炎症检出及其范围大小的确定；
- c) 全身各部位大血管病变的检出和定性诊断，冠状动脉病变的范围、程度和斑块性质的确定；
- d) 重要脏器外伤出血的定量及定性，多种外伤异物的定位；
- e) 某些器官部位的钙化或结石检出；
- f) 脏器变性（如肝脂肪变性）或先天异常。

注：附录 A 是胸部 CT 辅助诊断系统应用实例，附录 B 是智能医疗肺部 CT 辅助影像诊断系统应用实例。

7.4.2 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP等。

7.4.3 仪器设置与成像操作

仪器设置与成像操作要求应包括：

- a) 扫描范围：依据检查部位和目的而定（头颈部、胸部、上腹部、腹盆腔、四肢及关节、脊柱）；
- b) 检查设备：16层以上 CT，冠状动脉检查要求 64 层以上 CT；
- c) 扫描参数：螺旋扫描方式，管电压为 100 KVp~120 KVp，毫安量 200 mAs~250 mAs 或自动调制。采集层厚为 0.5 mm~1.0 mm，准直宽度为 4 mm~40 mm，螺距为 0.6~1.0；
- d) 重建参数：重建层厚 1 mm~2 mm，重建间隔 0.5 mm~1 mm。用标准卷积核重建软组织窗图像，必要时以高分辨卷积核重建骨窗和肺窗图像。头颈部及脊柱重建视野为 150 mm~250 mm，胸腹部重建视野 300 mm~400 mm，重建矩阵 512×512 及以上；
- e) 增强扫描：碘对比剂浓度 300 mgI/ml~400 mgI/ml，总量 1.2 ml/Kg~1.5 ml/Kg，注射速率 3.0 ml/s~5.0 ml/s。延时时间采用团注智能跟踪技术；
- f) 窗宽窗位：依据检查部位和目的而定（头颈部、胸部、上腹部、腹盆腔、四肢及关节、脊柱）。

7.4.4 影像质量

影像质量要求应包括：

- a) 一般图像：每个病例的图像应当保持连续完整，不得出现缺层、错层等情况。不得进行任何修改、编辑，不得进行有损压缩。根据对解剖细节的显示情况，分为以下几类：
 - 1) 可见：器官和结构在检查范围内可观察到。但细节未显示；
 - 2) 显示：解剖结构细节可见，但不能清晰辨认，即细节显示；
 - 3) 清晰显示：解剖细节清晰辨认，即细节清晰。
- b) 冠脉图像：可诊断的冠状动脉（直径 ≥ 1.5 mm）节段中的 90%（15 个节段中的 13 个节段）没有伪影，能够诊断；确保冠状动脉和心脏扫描范围的完整性；冠状动脉 CT 值最佳范围 300 HU~450 HU，特别注意冠状动脉远端增强效果是否满意；图像噪声：测量主动脉根部图像的 CT 值标准差（SD 值）作为图像噪声， < 20 HU 为优秀， 20 HU~ 30 HU 为良好， > 30 HU 为图像质量差， > 40 HU 为检查失败（图像不能评估）。推荐目标控制在 20 HU~ 30 HU 以下，推荐使用迭代重建降低图像噪声。

7.5 X-ray 影像

7.5.1 应用场景

X-ray 影像应用场景包括：

- a) 数字 X 线普通摄影；

- b) 数字乳腺摄影；
- c) 数字减影血管造影；
- d) 数字胃肠摄影。

7.5.2 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP等。

7.5.3 仪器设置与成像操作

仪器设置与成像操作要求应包括：

- a) 体位：被检部位的中心位于图像的正中，图像分布比例适当；
- b) 中心线：对准被检部位的中心，使影像显示的被检部位解剖结构无变形与放大；
- c) 摄影参数：采用自动电离室曝光；
- d) 摄影距离（SID）：100 cm~120 cm 适用于四肢、骨关节、胸腹部仰卧位摄影；120 cm 左右适用于床旁摄影；150 cm 适用于立位颈椎腰椎摄影、负重位等；180cm 适用于站立位胸片；200 cm 适用于心脏摄影。

7.5.4 影像质量

影像质量要求应包括：

- a) 清晰可见：解剖学结构或/和病变特征细节可清晰辨认，即细节清晰；
- b) 可见：解剖学结构或/和病变特征等细节可见，但不能清晰辨认，即细节显示；
- c) 隐约可见：解剖学结构或/和病变特征等细节可观察到，但细节没有完全显示，只特征可见。

7.6 核医学影像

7.6.1 应用场景

核医学影像应用场景包括：

- a) 胸腹部脏器实体肿瘤（临床较多为肺、结直肠、前列腺等）；
- b) 神经系统显像，主要为大脑；
- c) 部分特殊器官显像，如心脏。

7.6.2 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP等。

7.6.3 仪器设置与成像操作

仪器设置要求应包括：

- a) CT 诊断使用的 mAs 参数：SPECT/CT 和 PET/CT 中的 CT 为临床诊断型 CT，此项参数可按疾病种类和检查部位参考本文件 CT 部分进行设置，一般为 150 mAs~250 mAs 或自动调制；
- b) CT 诊断使用 kV 参数：SPECT/CT 和 PET/CT 中的 CT 为临床诊断型 CT，此项参数可按疾病种类和检查部位参考本文件 CT 部分进行设置，一般为 100 kV~120 kV；
- c) 图像矩阵参数：图像矩阵大小应恰当。矩阵过小会导致图像分辨率损失，不利于诊断。太大不利于存储和传输。CT 参数可按疾病种类和检查部位参考本文件 CT 部分进行设置，一般为 256×256~1024×1024。SPECT 一般采用 128×128~256×256 的矩阵。PET 一般采用 256×256~512×512 的矩阵；

- d) SPECT 和 PET 采用迭代法能获得较好的断层图像, 不同的厂商和不同代次的机型往往有不同的迭代重建算法, 迭代次数依据迭代算法而定, 一般为 2 次~25 次;
- e) SPECT 和 PET 采用迭代法能获得较好的断层图像, 不同的厂商和不同代次的机型往往有不同的迭代重建算法, 迭代子集数依迭代重建算法而定, 一般 4 次~16 次;
- f) 一般 SPECT/CT 为 5 mm~10 mm, PET/CT 为 1 mm~4 mm;
- g) CT 部分可参考本文件 CT 部分。一般头部, 5 mm~8 mm 采用层面扫描, 其它部位采用螺旋扫描;
- h) SPECT 和 PET 的层厚由半高宽决定, 一般 SPECT 为 5 mm~10 mm, PET 为 1 mm~4 mm。

成像操作要求应包括:

- a) SPECT 全身骨显像。显像剂 $^{99m}\text{TcMDP}$, 成人 15 mCi~25 mCi。儿科患者 0.25 mCi/kg, 最小剂量不低于 2 mCi。特殊原因造成显像剂量低, 需延长采集时间。显像方法, 显像剂注射后 2 小时~5 小时内进行。肾功能正常的婴幼儿骨显像从软组织清除的时间比成年人快, 可在 1.5 小时后进行。仪器条件: 低能高分辨平行孔准直器的双探头或者单探头 SPECT 仪器, 能峰 140 keV, 能窗窗宽 20%。显像条件: 全身显像: 矩阵 256×1024 , 扫描速度: 10 cm/min~20 cm/min, 全身分段显像: 矩阵 256×256 , 每个部位采集 2 分钟以上, 采集完毕后拼接成全身图像。图像采集时, 探头尽量贴近患者或者使用体表轮廓跟踪技术, 以提高图像质量。体位: 仰卧位。全身平面采集前位和后位。病人平静放松, 左右肢体和躯干应保持对称, 双手五指分开平放。根据病情和需要, 可采用其他体位。检查过程中, 不能移动躯体;
- b) SPECT 肾动态显像。显像剂, $^{99m}\text{Tc-DTPA}$, 药物标记率 $>96\%$, 成人 5 mCi~10 mCi, 儿童 0.2 mCi/kg 体位常规仰卧位或者坐位, 后位采集。移植肾, 仰卧位, 前位采集。显像方法: 肘静脉弹丸注射显像剂, 同时启动采集开关, 行连续双肾动态采集。肾血流相 2 秒/帧, 连续采集 60 秒, 共 30 帧。肾功能相 60 秒/帧, 连续采集 20 分钟。可根据病情调整。仪器条件: 低能通用型准直器, 能峰 140 keV, 窗宽 20%, 矩阵 64×64 或 128×128 , 图像处理应用感兴趣区 (ROI) 技术分别勾画出双肾区及腹主动脉区, 获取肾血流灌注和功能曲线及相关定量参数;
- c) $^{18}\text{F-FDG}$ PET/CT 肿瘤显像。所有患者在注射 $^{18}\text{F-FDG}$ 之前至少禁食 6 小时, 6 小时之前可以食用易消化的食物, 4 小时内禁止静脉使用含碳水化合物的液体。注射 $^{18}\text{F-FDG}$ 之前要求血糖水平在 7.0 mmol/L~11.1 mmol/L 以下 $^{18}\text{F-FDG}$ 注射剂量建议为 0.10 mCi/Kg~0.15 mCi/Kg, 儿童最少不低于 1.0 mCi, 超重儿童不超过 10 mCi, 成人一般不超过 15 mCi。建议药物注射后 50 分钟~75 分钟内, 最好在 1 小时进行 PET 扫描;
- d) $^{18}\text{F-FDG}$ PET/CT 脑代谢显像。所有患者在注射 $^{18}\text{F-FDG}$ 之前至少禁食 6 小时, 6 小时之前可以食用易消化的食物, 4 小时内禁止静脉使用含碳水化合物的液体。注射 $^{18}\text{F-FDG}$ 之前要求血糖水平在 8.9 mmol/L 以下。 $^{18}\text{F-FDG}$ 注射剂量建议为: 成人 2D 扫描 8.1 mCi~16.2 mCi (一般 10 mCi), 成人 3D 扫描 3.4 mCi~6.8 mCi (一般 4.1 mCi); 儿童 2D 扫描最少 0.7 mCi, 最多 10 mCi, 儿童 3D 扫描最少 0.4 mCi, 最多 5.3 mCi。建议 $^{18}\text{F-FDG}$ 注射后 30 分钟~60 分钟时进行 PET 扫描。

7.6.4 影像质量

建议最低分辨率要求为 72 ppi, 图像尺寸 1024 像素 \times 768 像素, 图像文件大小不小于 768 kb。图像清晰、均匀性、超声切面标准、图无斑点, 雪花细粒, 网纹等干扰。

7.7 心电图

7.7.1 应用场景

心电图应用场景包括:

- a) 心律失常;
- b) 心肌缺血;
- c) 其他涉及心电图指标及其运算的应用场景等。

7.7.2 数据格式与存储

数据文件格式应包括：BMP、JPEG、PDF、FDA-XML（HL7）等。

7.7.3 仪器设置与成像操作

仪器设置要求应包括：

- a) 前置放大电路，从强的噪声背景中提取心电信号；
- b) 带通滤波电路，使频率为 0.05 Hz~100 Hz 的心电信号通过，该范围以外的信号将大幅度衰减掉；
- c) 50 Hz 拮波电路，用于滤掉 50 Hz 工频干扰；
- d) 主放大电路，将前级放大的心电信号进行再次放大；
- e) A/D 转换电路，将系统采集到的模拟信号转换为数字信号；
- f) 单片机及液晶显示器输出电路，处理采集到的数据并输出。

成像操作要求应包括：

- a) 心电图导联
 - 1) 躯干导联电极记录的心电图不等同于标准心电图，坐位和立位记录心电图与也不等同于标准卧位心电图，特殊位置记录的心电图应区别对待；
 - 2) 人工合成 12 导联心电图不能等同于标准 12 导联心电图，亦应区别对待。
- b) 成像系统
 - 1) 电子病历系统：原始电子病历系统中心电图以图像静态文件保存；
 - 2) 纸质病历系统：应含图像采集系统及图像输出系统。

7.7.4 影像质量

能分辨小波形与伪差之间的鉴别等细节；能识别P、QRS、T、U波形及各个间期，如PR间期、ST段、QT间期；扫描电子病历和纸质病历心电图图像应清晰体现心电图标尺，即每小格表示的振幅（纵坐标）及时长（横坐标）。

7.8 脑电图

7.8.1 应用场景

脑电图应用场景包括：

- a) 按照记录时间和方式还可分为常规脑电图、动态脑电图、视频脑电图等。脑电图常用于对脑功能的高时间分辨率成像场景，如：癫痫诊断等；
- b) 光电联合检测是指将脑电图谱术（EEG）和近红外光谱术（fNIRS）在空间上联合排布，在时间上同步采集的双模态脑功能检测技术，目前多应用于对脑功能活动的高时间和高空间分辨率成像场景。

7.8.2 数据格式与存储

数据格式与存储应符合下列要求：

- a) 脑电图数据文件格式应包括：EDF、TXT、MAT、XLS 等，数据内容应包括各通道电极所记录的

数据以及电极类型、数量、位置、固定方式、阻抗等附加信息；

- b) 采用脑电和近红外光同时记录时，近红外光部分数据格式为三维矩阵 $P_a \times Ch \times N_t$ ，其中 P_a 代表 fNIRS 检测的参数种类，至少包含含氧血红蛋白与脱氧血红蛋白两种参数， Ch 代表通道数， N_t 代表时间，建议大于 100。文件类型应包括 TDMS，XLS，MAT 和 TXT。

7.8.3 仪器设置与成像操作

仪器设置要求应包括：

- a) 普通电极可通过导电膏进行固定，同时提高导电能力。在无发皮肤处，可使用胶带或可粘着的胶圈对电极进行固定。当使用多电极阵列时，宜使用弹性帽或弹性网进行辅助固定。当受试对象为儿童时，电极固定前后都容易发生移动，宜使用弹性帽或弹性网进行辅助固定。弹性帽/网使用过程中应选择适应头部大小的对应型号，并保证电极安放位置准确；
- b) 应使用电极阻抗（即电极与头皮间阻抗）参数描述电极与头皮的连接、贴合程度。电极阻抗可用于参与评价放大器准确度、传感器信号拾取能力、差分放大器抑制工模信号的能力、电极下皮肤完整程度。为使放大器准确记录，电极阻抗应低于放大器输入阻抗至少 100 倍。可通过导电膏和磨砂膏等方式将脑电信号频率范围内的电极阻抗降至小于 $5 \text{ k}\Omega$ ；
- c) ERP 测量电极固定于头皮，位置误差在 $\pm 5 \text{ mm}$ ，电极间距离大致相等。具体位置应参照标准电极位置放置，其中美国脑电图协会提出的 10~20 标准使用最为广泛；
- d) 光电联合记录时可包含前额叶，颞叶，枕叶，顶叶等在内的大脑皮层区域。近红外光与脑电检测区域应部分或全部重叠；
- e) 时间同步可以是硬件同步触发或者软件同步标记，其时间同步精度应小于近红外光成像的时间分辨率 ($< 0.1 \text{ s}$)。

成像操作要求应包括：

- a) 理论上，参考电极应放置在身体的零电位点。实际中，只能将距离脑足够远的身体位置视为零电位点。若在躯干或四肢位置放置参考电极，脑电会受到心电的严重干扰，因此，脑电记录宜使用耳垂、鼻尖或乳突部作为参考电极。另外，脑电记录中的接地电极有助于去除工频干扰，可放置于头前部中点，也可放在头部和身体的任何部位；
- b) 在使用多电极阵列记录脑电时，接触不良或放大器故障等问题可能造成某个或多个通道信号存在大量噪声。这些电极拾取的信号应被视为被污染数据，这些受污染通道数量应低于电极总数的 5%。受污染数据对脑电记录分析具有负面影响，如果直接删除被污染通道数据，则所有样本都不应使用该通道数据，应详细记录受污染数据的处理方式；
- c) 刺激时间参数可包括：刺激持续时间和刺激间隔时间。如果实验过程中包括不同类型刺激，则还应准确测量与记录不同刺激的时间间隔。该刺激间隔时间可定义为从当前刺激开始到下一刺激开始的间隔时间，也可定义为从当前刺激结束到下一刺激开始的间隔时间，需要明确说明所选标注方式。如果受试者需要做出动作或语言反馈，则应记录反馈相对于刺激的出现时间。刺激序列结构是脑电图数据的重要组成部分。应记录刺激产生、出现规则（依据概率的完全随机生成、或者相同刺激不连续出现的随机生成）。使用显示器进行视觉刺激时，应使用光感受器检查刺激时间；
- d) 当需要记录运动诱发响应时，应对运动产生部位、速度、力量、幅度做出明确记录，并记录受试实验过程中的姿态（躺、坐、站等）和眼部状态（睁眼、闭眼）。

7.8.4 影像质量

影像质量要求应包括：

- a) 系统增益与分辨率：信号放大器应具有至少 100 dB 的共模抑制比以消除共模噪声。分辨率不

应低于 8bit；

- b) 滤波要求：应记录带通滤波器的上、下限截止频率（-3 dB）。转换器采样频率应高于 ERP 信号最高频率的两倍。通常模拟滤波器上限截止频率可设定为转换器采样频率 1/4，下限截止频率为扫描时间 4 倍的对应该频率。如果需要进一步滤波，可用数字滤波进行后续分析。滤波器并不能完全滤除截止频率之外的频率成分，超出滤波范围的强信号仍可能导致混叠现象。此外，截止频率 50 Hz~60 Hz 的陷波滤波器可能会使脑电事件相关电位波形严重失真，应谨慎使用；
- c) 采样率要求：模数转换器的采样频率会影响测量信号的拾取效果，应记录标明具体采样频率。采样频率不应低于所测量信号最高频率成分的 2 倍，以防止发生混叠现象，造成波形失真；
- d) 信号伪迹和污染数据剔除要求：从头皮记录的电信号并不是单纯的脑电，头部肌肉运动、舌动、眼动、皮电等噪声都对该信号有影响。信号伪迹可包括：生理性伪迹（如：心电伪迹、肌电伪迹、眼动伪迹等）、来自采集系统的伪迹、电磁干扰伪迹、运动伪迹等。为便于脑电后续分析去除伪迹，降低其他源信号对脑电信号的干扰，应设置其他信号（如：心电、肌电等）的检测电极并准确记录并报告采用的电极数量、位置、作用等信息。可根据脑电信号参数（例如：潜伏期、幅值等）判定数据受污染程度，一般情况下如果某单次数据明显超出（ $\pm 200 \mu V$ ）该通道的叠加平均，则该次数据不应纳入叠加平均。根据测量信号种类不同，该剔除标准也会发生改变。应记录不同实验条件、不同受试所剔除序列的比例。对于容易产生不合格数据的受试群体，宜增加数据记录数量；
- e) 光电联合成像时的数据评估要求：可通过主观观察或者算法提取原始静息态数据中的脉搏信号：如果能够清晰观察或提取到与脉搏频率一致的波动信号，则可认为数据优良；如果观察到或提取到信号波动不明显，则可认为数据中等，如果无法提取或观察到脉搏信号即为数据较差。应取优良信号和中等信号做分析，其相应占比宜为 90%和 10%。

8 测试评价方法

8.1 测试评价流程

针对智能医疗影像辅助诊断系统开展面向医疗影像诊断业务场景的测试评价。测试评价流程见图3。

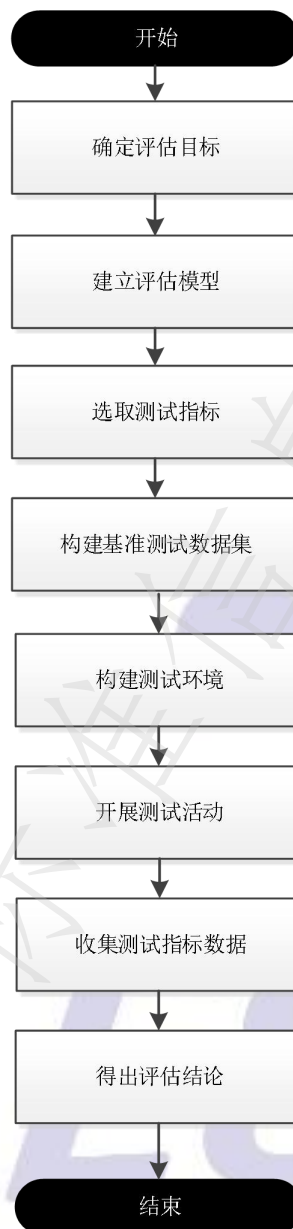


图3 智能医疗影像辅助诊断系统测试评价流程

8.2 确定评估目标

根据测试评价活动开展的不同目的，可以由医疗影像辅助诊断系统的采购方、研发方和第三方测试机构等根据实际情况协商或独立确定本次测试评估的目标。

8.3 建立评估模型

根据确定的评估目标，建立本次测试评价的评估模型。给定的评估模型为：

- a) 从测试指标中选择 1 个或多个指标；
- b) 逐一对所选定的测试指标给出本次活动的通过准则或要求；
- c) 经过测试后，选取的测试指标全部满足通过准则或要求，则为通过评估；否则为未通过评估。

8.4 选取测试指标

8.4.1 准确率

在给定的测试环境下，辅助诊断系统对给定的n个数据样本进行运算，预测为正确的样本数占总样本数的比率。

$$\text{准确率} = \frac{\text{预测正确的样本数}}{\text{选取的 } n \text{ 个数据样本}} \times 100\%$$

8.4.2 敏感度

在给定的测试环境下，辅助诊断系统对给定的n个数据样本进行运算，正例中预测为正例的样本数占总体正样本个数的比率。

$$\text{敏感度} = \frac{\text{正例中预测为正例的样本数}}{\text{数据集中正样本总数}} \times 100\%$$

8.4.3 特异度

在给定的测试环境下，辅助诊断系统对给定的n个数据样本进行运算，负例中预测为负例的样本数占总体负样本个数的比率。

$$\text{特异度} = \frac{\text{负例中预测为负例的样本数}}{\text{数据集中负样本总数}} \times 100\%$$

8.4.4 ROC 曲线下面积

在给定的测试环境下，ROC曲线下的面积，其中ROC (receiver operating characteristic curve) 表示接收者操作特征曲线。

8.4.5 PR 曲线

PR曲线是以精准率和召回率这两个为变量而做出的曲线，其中召回率为横坐标，准确率为纵坐标。

8.4.6 平均交并比

在给定的测试环境下，辅助诊断系统对给定的n个数据样本进行运算，预测的边框和真实的边框的交集和并集的比值的均值。

$$\text{平均交互比} = \frac{1}{n} \sum_{i=1}^n \frac{\text{样本 } i \text{ 的预测边框和真实边框的交集}}{\text{样本 } i \text{ 的预测边框和真实边框的并集}}$$

8.4.7 鲁棒性

在给定的测试环境下，辅助诊断系统输入给定的n个非正常数据样本，系统保持其准确率不发生较大偏移的能力。

$$\text{鲁棒性} = \frac{\text{非正常数据样本中能够保持其正常结果输出的样本个数}}{\text{基准数据集中非正常数据样本}} \times 100\%$$

8.4.8 响应时间

在给定的测试环境下,辅助诊断系统对给定的n个数据样本进行运算并获得结果所需要的平均时间。其中 T_{e_i} 表示第i个数据样本输入的时间, T_{b_i} 表示第i个数据样本结果输出的时间。

$$\text{响应时间} = \frac{1}{n} \sum_{i=1}^n (T_{e_i} - T_{b_i})$$

8.4.9 任务完成率

在给定的测试环境下,辅助诊断系统对给定的n个数据样本进行运算,辅助诊断任务正确完成任务的比率。

$$\text{任务完成率} = \frac{\text{任务完成成功的样本个数}}{\text{选取的 } n \text{ 个数据样本}} \times 100\%$$

8.4.10 推理计算的成本

给定的测试环境下,辅助诊断系统对给定的n个数据样本进行运算,系统在单位时间内完成推理计算的平均成本。其中T表示n个数据样本全部完成诊断任务所需要的时间。

$$\text{推理计算成本} = \frac{\text{辅助诊断系统的软硬件成本之和}}{n \times T}$$

8.4.11 跨平台稳定性

针对辅助诊断系统,在发生软硬件环境变化的条件下,系统保持其任务正确完成的能力。其中n表示给定的数据样本集的总数。

$$\text{跨平台稳定性} = \frac{n \text{ 个测试样本中任务完成成功的样本个数}}{\text{软硬件环境发生变化后的 } n \text{ 个数据样本}} \times 100\%$$

8.5 构建基准测试数据集

基于智能医疗影像辅助诊断系统的影像数据规范,按照实际业务的情况构建覆盖多种影像的基准测试数据集用以开展测试。对测试数据集的要求应包括:

- a) 数据模态覆盖所需的医疗影像;
- b) 数据集中包含正负样本,且标注正确;
- c) 数据集中样本分布均衡,且与真实业务场景下的数据分布一致;
- d) 数据集中包含已知干扰样本、噪声样本或野值样本等异常影像数据。

8.6 构建测试环境

根据智能医疗影像辅助诊断系统运行需要的软硬件要求,构建一致的测试环境。如无法构建出与实际使用场景相同的测试环境,则需要进一步分析由于测试环境与实际使用环境不一致所带来对测试结果的影响。

8.7 开展测试活动

在构建的测试环境下,利用已经选择的基准测试数据集进行测试。即按照被测系统的正确使用方法,将基准测试数据集中的数据全部依次输入被测系统。

8.8 收集测试指标数据

根据选择的测试指标的计算公式，收集相应的数据并进行测试指标计算。

8.9 得出评估结论

将测试指标数据与给定的指标要求进行对比，并依据建立的评估模型得出评估结论，判断此次测试评价活动的结果。



附录 A
(资料性附录)
胸部 CT 辅助诊断系统

A.1 影像数据要求

A.1.1 应用场景

本软件为胸部CT影像查看、处理和辅助诊断软件，适用于各级医院放射科、影像科、呼吸科等科室。主要用于胸部CT的DICOM影像的查看和处理，以人工智能的方式检出可疑病灶，并对病灶进行自动标记，并给出病灶位置。

A.1.2 数据格式与存储

影像文件格式为DICOM3.0。

产品可与如下系统接口：RIS (Radiography Information System, 放射科信息系统)，PACS (Picture Archive and Communication System, 医学影像计算机存档与传输系统)

A.1.3 仪器设置与成像操作

仪器设置与成像操作包括：

- a) 扫描范围：肺尖至肋膈角尖端水平；
- b) 检查设备：16层或以上多层螺旋CT；
- c) 扫描参数：螺旋扫描模式，建议螺距设定 ≤ 1 ，机架旋转时间 ≤ 1.0 s，扫描矩阵设定不低于 512×512 （具体技术参数依不同机型而定），并采用大视野（FOV=L）；没有迭代重建技术的可使用120 kVp、30 mAs~50 mAs的扫描参数，有新一代迭代重建技术的可使用100 kVp~120 kVp、低于30 mAs作为扫描参数。
- d) 重建参数：若重建层厚 ≤ 0.625 mm可以无间隔重建，若重建层厚介于 0.625 mm~ 1.250 mm之间，则重建间隔 \leq 层厚的80%；采用标准算法，或者肺算法和标准算法同时进行重建。
- e) 窗宽窗位：采用纵膈窗（窗宽350 U~380 U、窗位25 HU~40 HU）及肺窗（窗宽1500 HU~1600 HU，窗位-650 HU~-600 HU）分别进行阅片。

软件运行环境的要求：

- a) 特定软件：网络浏览器，Chrome 或兼容浏览器；
- b) 系统硬件最低配置为：英特尔至强处理器 E5-1630 v4、4G 内存、128G 硬盘、英伟达 1080 显卡。

A.1.4 影像质量

影像质量包括：

- a) 图像未进行修改、编辑，未进行有损压缩；
- b) 图像应当保持连续完整，无缺层、错层等情况；
- c) 图像重建视野应当覆盖整个胸部横截面；
- d) 图像不应出现严重的设备伪影、呼吸伪影和人为伪影。

A.2 功能结构

主要模块包含系统管理、影像管理、肺部辅助诊断。

- a) 系统登录页面（含安全管理功能）：用户认证管理、患者信息管理（具备患者信息脱敏、基础信息管理、影像脱敏与档案管理功能）；
- b) 系统具备临床功能包括：影像导入功能，影像查询和筛选功能，病灶检出及分析功能，影像查看功能，标注管理功能，多图对比功能，调窗功能。



图 A.1 胸部 CT 辅助诊断系统基本功能结构

A.3 功能要求

A.3.1 安全管理要求

安全管理要求包括：

- a) 系统研发过程严格按照现行 GDP 要求进行伦理审评和备案；
- b) 医学影像信息的采集、标注、使用、以及共享等获得医疗机构的伦理审批。以知情同意书形式获得个人信息主体的知情权与授权，或伦理机构同意免除知情通知；数据操作严格遵守现行 GCP 规范；
- c) 对待标注数据中的个人敏感信息按照各医疗机构的规定进行脱敏，对需要使用的敏感信息应采取最小必要原则；
- d) 系统严格按照 GB25000.51-2016 标准的要求保障数据的保密性、完整性、可用性。
 - 1) 应采取数据分类、重要数据备份、加密认证等措施保证数据安全；
 - 2) 系统分为管理员和一般用户两级授权机制进行不同等级用户的数据接入和使用权限控制，功过用户名、密码登录的方式，防止非授权人员进入；
- e) 系统具备日志功能，对用户认证、对数据访问进行控制、规范数据接入、使用和销毁过程进行痕迹管理，确保对数据访问行为可管、可控；对服务管理的全流程留痕，对安全、隐私、风险及事故可查询、可追溯。

A.3.2 影像数据获取要求

影像数据获取要求包括：

- a) 系统具备从本地存储或 PACS 系统导入影像数据能力；
- b) 系统具备对胸部 CT 的 DICOM 格式的医学图像进行导入的功能；
- c) 系统可根据不同医疗机构的影像质量等级划分标准，进行检测规则定制，实现在影像数据获取中自动化完成影像质量检测，并输出影像数据采集质量报告，过滤不合格“病例”数据，以确保后续环节数据可用性。

A.3.3 数据预处理要求

由于DICOM影像格式有着规范的图像组织方式，齐全的标签列表，以及一套通用的编码方法，同时为保证产品使用的数据形态尽可能贴近临床真实情况，建议在数据存储时不进行滤波、增强、重采样、尺寸裁剪、均一化等数据预处理操作。

A.3.4 影像分割要求

影像分割要求包括：

- a) 系统具对胸部 CT 的 DICOM 医学图像的分割功能，可对器官、病灶、ROI 区域进行半自动化分割，可用不同颜色区分病灶或器官；分割算法基于 level-set 分割方法进行；
- b) 系统分割的最小单元为像素；
- c) 分割处理的输入为经过预处理后的影像或者影像序列，经过处理后输出为分割后图像或者记录分割、点信息的数据结构，分割处理后应保存未分割的输入影像。

A.3.5 特征提取与特征选择要求

特征提取与特征选择要求应包括：

- a) 系统具备一种或多种特征提取功能，特征提取算法应包括统计分析、频谱分析等；
- b) 系统具备对特征进行变换的功能；
- c) 系统具有特征选择功能，抽取可最有利于区分不同模式的特征；
- d) 系统的特征将直接影响辅助诊断的结果，并是结果的展示重要信息来源，具备特征提取功能的系统应将所提取的特征和特征选择方法进行准确描述。

A.3.6 数据挖掘要求

数据挖掘要求应包括：

- a) 系统使用 faster R-CNN 为基础的算法实现对病灶的识别和检测任务；
- b) 系统需要 GPU 进行算法加速；
- c) 系统具备模型训练状态的跟踪、监控和控制。

A.3.7 结果评价要求

系统功能经有资质的医用软件检测机构进行GB25000.51-2016的检测。

系统辅助诊断性能通过国家级标准数据集和临床试验进行验证。其主要指标有准确率、敏感度、特异度、错误率、辨别力、AUC、F值、过拟合要求。

A.4 实施效果

通过业内领先图像算法对新冠肺炎CT影像的智能化定量评价，包括提示肺内疑似肺炎患者；局部性病灶、弥漫性病变、全肺受累的各类肺炎疾病严重程度分级；对病灶的形态、范围、密度等关键影像特

征定量和组学分析，精确测算疾病累计的肺炎负荷；实现对多次复查CT的全肺病变动态4D对比。从而高效、准确地辅助医生快速进行病情分析和疗效评价，为诊断、治疗提供决策依据。



图 A.2 胸部 CT 影像查看、处理和辅助诊断系统

该系统通过从新冠肺炎患者CT片中提取定量参数，可对肺炎严重程度进行自动量化评估。医生对一个病例进行常规手动量化评估需要5小时~6小时，利用AI系统只需要2分钟~3分钟，极大提升了诊断效率。通过研究上百例新冠肺炎病例，将AI系统与高年资医生的诊断结果相比较，发现定量参数与医生的评价结果相似（相关性研究中 $R > 0.87$, $p < 0.001$ ），显示高稳定性地诊断质量。这套系统的推广应用已助力全国特别是疫区重点医疗机构的大规模高质量新冠肺炎的定量诊疗工作。

附 录 B
(资料性附录)
智能医疗肺部 CT 辅助影像诊断系统

B.1 影像数据要求

B.1.1 应用场景

智能医疗CT影像辅助诊断系统可对肺部CT中的肺结节和肺炎进行识别和评估，应用场景如下：

- a) 肺结节：读取分析符合 DICOM 标准格式的图像，对图像中的肺部结节位置自动标注；并可以判断肺部结节的类型与特征，生成可操作的结构化报告，提供前后片对比、测量分割等辅助诊断功能；
- b) 肺炎：输入肺部 CT 图像可以自动勾勒肺部炎症范围，进而对肺部炎症区域进行体积和密度的测量并提示存在炎症的阳性病例，从而达到病例分诊和病灶分析的目的。

B.1.2 数据格式和存储

智能医疗CT影像辅助诊断系统基于DICOM标准进行网络数据传输，符合DICOM通讯协议。

B.1.3 仪器设置与成像操作

智能医疗CT影像辅助诊断系统在以下方面对CT设备仪器设置和成像操作提出要求，包括：

- a) 扫描范围：需包含胸部；
- b) 检测设备：CT 设备型号无限制；
- c) 扫描参数：管电压大于等于 70 KVP；
- d) 采集精度：分辨率不低于 512 像素×512 像素。

为了确保系统对不同设备的适应性，该案例在人工智能训练和算法验证上至少会包含来自6家以上不同设备制造商的影像数据。

B.1.4 影像质量

智能医疗肺部CT影像辅助诊断系统训练使用的数据，保障每例数据都进行数据质量控制，数据收集严格遵守质控要求。其中，数据类型和完整性要求应包括：不得出现缺层、错层等情况；不得进行任何修改、编辑；不得进行有损压缩等。智能医疗肺部CT影像辅助诊断系统部署于医院中时，会设置DICOM文件过滤规则，确保不满足要求的数据无法进入系统。

B.2 功能结构

CT影像辅助诊断系统基本功能结构和功能见图B.1。重视患者隐私的保护和信息的安全，提供了如下安全管理功能：

- a) 用户安全管理功能，包括多类用户和权限区分。同时可以提供患者信息脱敏与基础信息管理、影像脱敏与档案管理等相关工作软件，确保患者信息安全；
- b) CT 影像辅助诊断系统的辅助诊断功能和算法模型设计满足临床需求：辅助诊断功能包括感兴趣区标注、结节属性、疑似肺炎分诊、肺炎区域分析等功能。

算法模型设计上，通过获取影像数据、数据预处理、数据挖掘、分割、特征提取、特征选择、智能算法模型构建、模型迭代、结果评价、临床测试评价等，确保了模型算法临床使用的安全有效性。

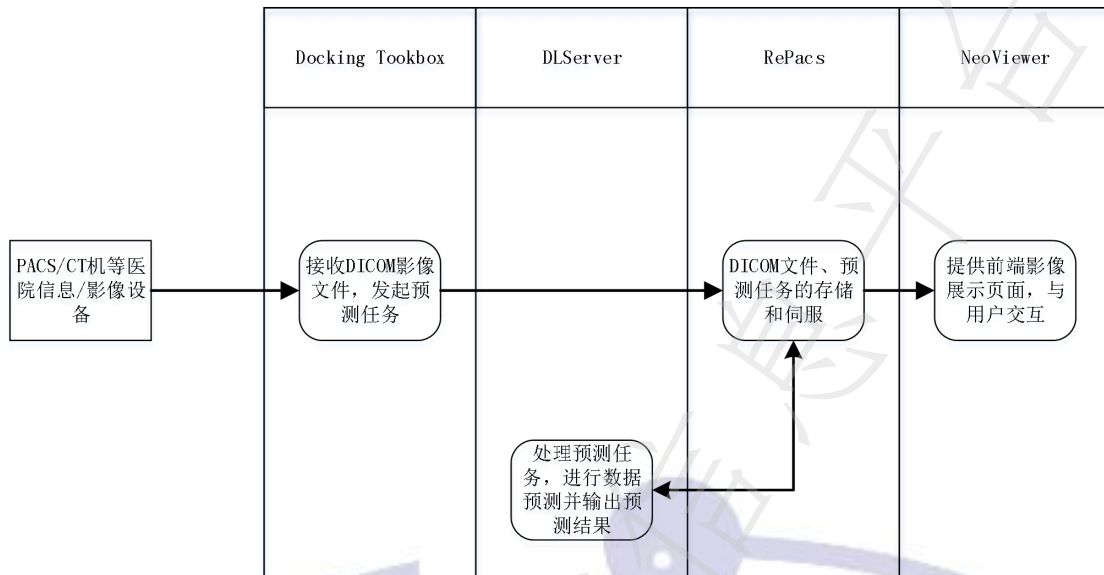


图 B.1 系统基本功能结构

B.3 功能要求

B.3.1 安全管理方式

在以下方面，确保了患者数据的安全：

- 在获得伦理批准后，开展临床试验研究；
- 医学影像信息的采集、标注、使用、以及共享都具有相应的控制流程，操作规范和记录。通过与医院签订合同，确保了个人信息主体的知情权和授权；
- 医院向开发者传递的数据，在医院会进行第一次脱敏；存入开发者的原始数据库前，会使用开发者自行研发的脱敏工具再次脱敏，确保开发者使用的影像数据已完全脱敏；
- 系统开发完成后，开发者会进行网络安全能力评价，来保障数据的保密性、完整性、可用性，采取数据分类、重要数据备份、加密认证等措施保证数据安全；
- 开发者的软件具备对接 RIS/PACS/HIS 的能力，能够与这些软件协同使用，确保用户的信息的保密和安全。

数据采集过程见图B.2，来确保患者信息的安全和完全脱敏。

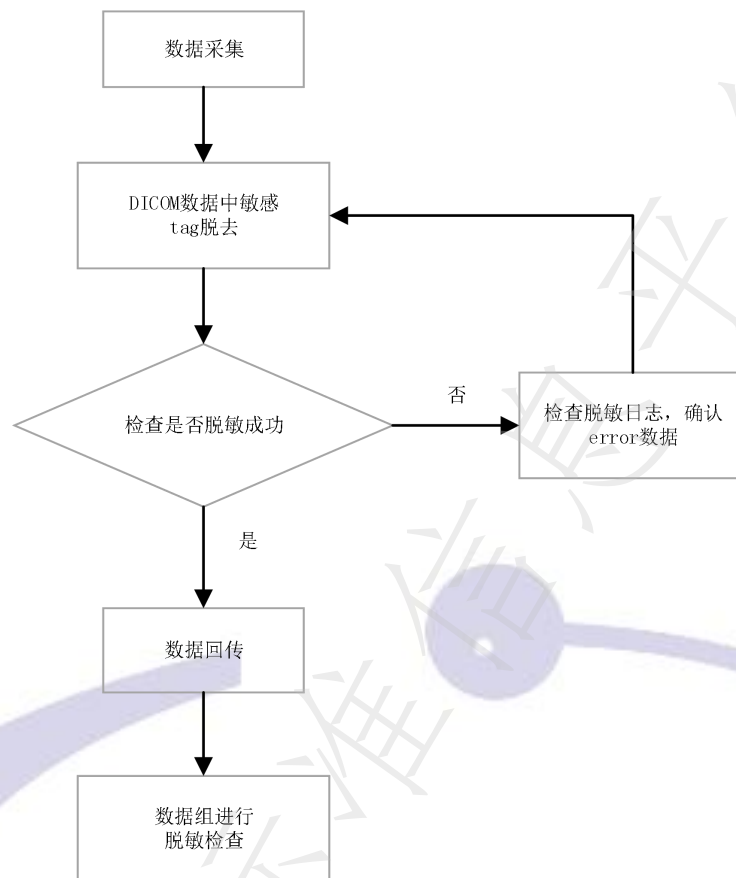


图 B.2 数据安全要求

B.3.2 影像数据获取方式

开发者定义的系统影像数据获取能力包括：

- 系统应具备从本地存储或 PACS 系统导入影像数据能力；
- 系统可直接与 CT 对接；
- 系统支持符合 DICOM 格式的数据。

影像数据获取方式，见图B.3。

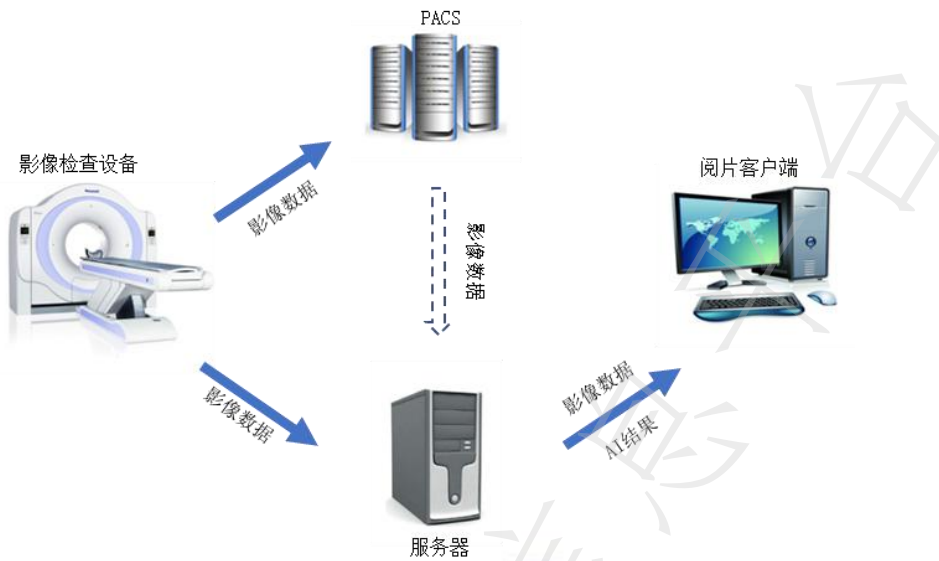


图 B.3 网络拓扑图

B.3.3 数据预处理要求

智能医疗CT影像辅助诊断系统可包含针对导入的影像数据进行预处理的功能，可根据不同诊断方向，不同影像类型，选择预处理功能。系统可具备针对输入影像数据的预处理能力，这些预处理包括边缘检测、层厚映射等。

B.3.4 智能影像诊断能力

智能影像诊断能力应包括分割、检测和分类：

- 能够自动对病灶或器官进行标记，包括肺叶、肺段、肺结节、疑似肺炎等；
- 标记能够达到像素级；
- 分割处理的输入为经过预处理后的影像或者影像序列，经过处理后输出为分割后图像或者记录分割点信息的数据结构，标记处理后保存原始的输入影像。

建立了标注资源管理制度，标注资源管理包括人员管理和基础设施管理。人员管理明确标注人员、审核人员和仲裁人员的选拔、培训、考核等要求。基础设施管理明确标注场所、标注软件等要求。人员培训考核和标注工具的确过程均形成记录，确保了影像分割、标注的质量。

B.3.5 特征提取与特征选择要求

通过深度学习模型结构的设计来进行特征提取与特征选择，针对不同预期用途的功能（如分割、检测、分类）设计最优算法结构和技术路线。

B.3.6 数据挖掘要求

模型训练，通常包括以下几个步骤：

- 数据预处理，将 DICOM 文件转换为训练数据；
- 选择合适的激活参数，常用的激活函数包括 sigmoid, ReLU, softmaxs 等，各有优缺点；
- 权重初始化，为了加快收敛，考虑使用迁徙学习；
- 学习优化算法，例如 SGD, Adagrad, Adam, AdaDelta, RMSProp 等。除提供自适应学习率之外，还针对不同模型参数使用不同比率，通常导致更平滑的收敛；

- e) 对于提供数据挖掘功能的系统支持外部算法、模型的导入，对于生成的模型，支持其储存和发布。能够直观地展示模型的分析结果，可以封装并导出分析结果，以便应用于外部评价工具；
- f) 系统可以支持使用 GPU 对特定算法加速分析；
- g) 系统可以支持对模型训练状态的跟踪、监督和控制。

B.3.7 结果评价过程

按照如下流程对结果进行评价：

- a) 根据需求中模型效果的测试指标安排测试，包括功能的程度、泛化能力、鲁棒性、以及安全性等测试。其中指标包括准确率、敏感度、特异度、错误率、辨别力、AUC、F1 值、过拟合等。
- b) 在独立与训练集的测试集上执行测试。
- c) 根据模型的输出结果和测试集的标记结果，输出测试指标值。
- d) 判断测试指标以及性能指标是否符合需求中的指标值，如符合，算法研究员进行步骤和 5)。
- e) 测试指标符合要求后，还要进行以下几项测试，并输出《模型测试报告》：
 - 1) 鲁棒性测试：用于测试模型鲁棒性，测试集一般为公开测试集或多家临床机构测试数据；
 - 2) 一致性测试：相同环境下，相同测试集测试两次，判断结果是否一致；
 - 3) 对比测试：用于不同版本模型之间的测试对比，新研发产品不适用；
 - 4) 特殊测试：用于测试老版本积累的不良事件测试集等；
 - 5) 视觉测试：数据分析师查看视觉效果，用于检查其他异常。
- f) 如未通过，进行整改，直至测试通过。

B.4 实施效果

开发者根据新的诊疗方案、一线影像数据与影像学的《诊断指南》的标准规范开发出了标准统一、诊断规范的CT医学影像人工智能辅助肺炎筛查及疫情监测系统，该系统的主要功能应包括：

- a) 快速筛查及提示功能：医生在电脑终端浏览患者 CT 影像时，AI 系统智能提示当前患者是否为可疑肺炎患者，如判断为肺炎立即预警（红色叹号），便于及时隔离，防止等待过程中的交叉感染；
- b) 精准辅助诊断功能：医生打开患者 CT 影像后，利用 AI 系统在胸部 CT 影像中对肺脏进行分割，同时可发现并分割肺野内炎症病灶，以肺叶为单位进行不同 CT 值区间体积百分比、炎症病灶体积及病灶占肺部的百分比分析，给出受累肺叶数目及肺部受累程度，辅助医生评估患者肺部受累情况，帮助医生迅速准确的判断病人病情，为诊断及治疗提供依据，提高肺炎诊疗规范与质量；
- c) 全自动智能病情及疗效评估功能：AI 系统提供全自动前后片对比功能，在医生日常工作中，通过前后 CT 片比较，能够比较客观评估病情的进展及疗效。但是该工作极其耗时和繁琐，而且人对病情细微的变化容易忽略。通过 AI 全自动对比分析功能，能够详细对比病患者前后胸部影像学变化，提供量化指标，为病情变化和疗效评估提供依据，帮助前线医生更加准确快速进行诊疗工作，打好打赢这场疫情防控阻击战；
- d) 精准的疫情状况统计及辅助上报功能：在智能诊断的基础上，系统自动统计出每天疑似病例，可以及时准确地汇集各医疗机构的疑似病例疫情数据，便于卫健委、疾控中心等政府部门对疫情进行提前配置资源和防疫措施。可给予地图可视化处理，直观掌握疫情，及时有效掌握当地区域肺炎情况；

- e) 在武汉大学同济医院的应用中，近千例病毒性肺炎阳性数据和阴性数据（其中阴性数据病人主诉往往为咳嗽或发烧，实际并未得肺炎），以病人为单位，该系统检出敏感度=98.1%，特异度=81.7%，AUC=0.981（AUC 常被用来作为 AI 模型好坏的指标，范围 0-1，越大越好）。

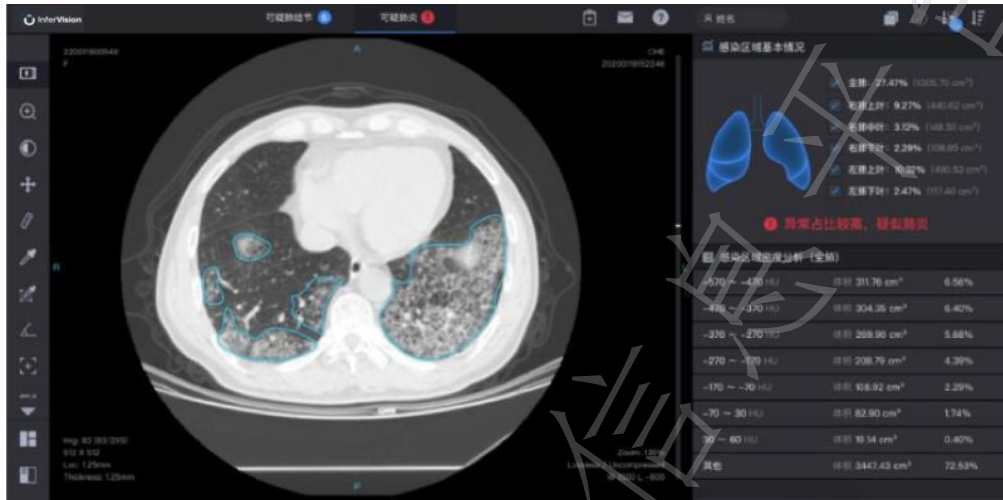


图 B.4 精准辅助诊断功能界面

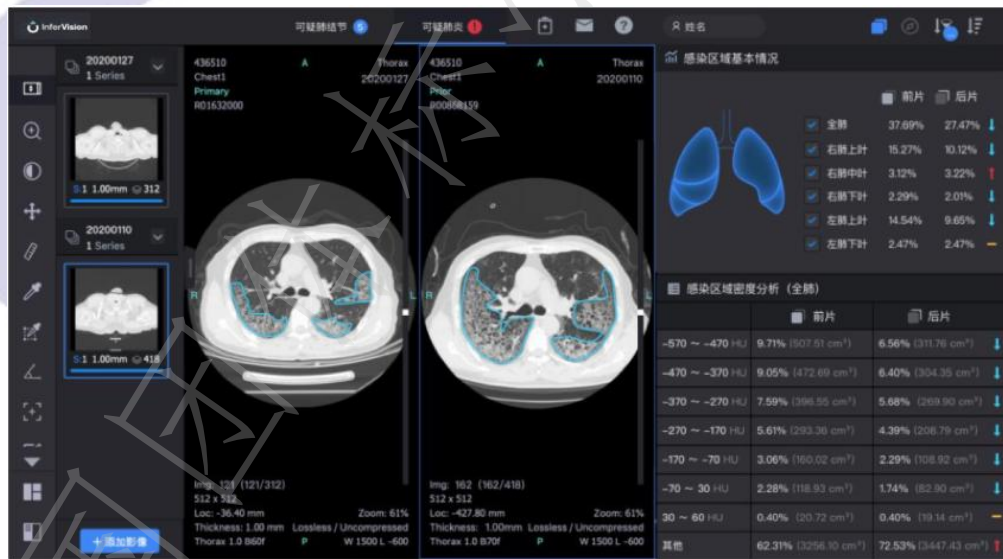


图 B.5 全自动智能病情及疗效评估功能界面