

中华人民共和国医药行业标准

YY/T XXXX—XXXX

人工智能医疗器械质量要求和评价 第1部分：
术语

Artificial intelligence medical device— Quality requirements and evaluation— Part 1:
Terminology

(征求意见稿)

(本稿完成日期：2020.06.24)

XXXX—XX—XX发布

XXXX—XX—XX实施

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
附录 A（资料性附录） 评价指标计算公式说明	23
参考文献	31

前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准的结构和编写》给出的规则起草。

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。

本文件由全国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：中国食品药品检定研究院。

本文件主要起草人：xxx等。

人工智能医疗器械质量要求和评价 第1部分：术语

1 范围

本文件规定了人工智能医疗器械质量评价使用的术语和定义；

本标准适用于人工智能医疗器械专业领域内从事科研、生产、监督检验和临床应用等方面的技术人员使用。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是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 35295: 2017	信息技术 大数据 术语
GB/T 5271.34-2006	信息技术 词汇 第34部分：人工智能 神经网络
GB/T 5271.31-2006	信息技术 词汇 第31部分：人工智能 机器学习
GB/T 5271.28-2001	信息技术 词汇 第28部分：人工智能 基本概念与专家系统
GB/T 5271.1-2000	信息技术 词汇 第1部分：基本术语
GB/T 32400-2015	信息技术 云计算
GB/T 36239-2018	特种机器人 术语
GB/T 25000.12-2017	系统与软件工程 系统与软件质量要求和评价 (SQuaRE)
GB/T 35295-2017	信息技术 大数据
GB 34960.5-2018	信息技术服务 治理 第5部分：数据治理规范
GB 18391.1-2009	信息技术 元数据注册系统 (MDR) 框架
GB/T 11457-2006	软件工程术语
GB/T 34952-2017	多媒体数据语义描述要求
GB/T 5271.1-2000	信息技术 词汇
GB/T 35273-2017	信息安全技术 个人信息安全规范
WS/T 305-2007	卫生信息数据集元数据规范
T/CESA 1037-2019	人工智能面向机器学习的系统框架和功能要求
ISO 2382:2015	Information technology—Vocabulary
ISO 16439:2014	Information and documentation — Methods and procedures for assessing the impact of libraries

3 术语和定义

下列术语和定义适用于本文件。

3.1 基础技术术语

3.1.1

人工智能医疗器械 artificial intelligence medical device

YY/T xxxx—xxxx

采用人工智能技术实现预期用途的医疗器械。

注1：如采用机器学习、神经网络、深度学习等技术实现辅助诊断等功能的医疗器械软件。

注2：如以人工智能技术从源头赋能医疗器械，形成内嵌AI算法、或基于AI芯片形成的智能硬件设备。

3.1.2

人工智能 artificial intelligence

AI

表现出与人类智能（如推理和学习）相关的各种功能的功能单元的能力。

[来源：GB/T 5271.28-2001，28.01.02]

3.1.3

模式识别 pattern recognition

通过功能单元对某一对象物理或抽象的模式以及结构和配置的辨别。

[来源：GB/T 5271.28-2001，28.01.13]

3.1.4

机器学习 machine learning

功能单元通过获取新知识或技能，或通过整理已有的知识或技能来改进性能的过程。

[来源：GB/T 5271.31-2006，31.01.02]

3.1.5

神经网络 neural network

由加权链路且权值可调整连接的基本处理元素的网络，通过把非线性函数作用到其输入值使每个单元产生一个值，并把它传送给其他单元或把它表示成输出值。

[来源：GB/T 5271.34-2006，34.01.06]

3.1.6

深度学习 deep learning

在神经网络或信念网络的情况下是对基于深层结构或网络表示的输入输出间映射进行机器学习的过程。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0320]

3.1.7

推理 inference

从已知前提导出结论的方法。

[来源：GB/T 5271.28-2001，28.03.01，有修改]

注1：在人工智能领域中，前提是事实或规则。

注2：术语“推理”既指过程也指结果。

3.1.8

特征 features

能表达模式本质的功能或结构特点的可度量属性，如大小、纹理、形状等。好的特征应能使同类模式聚类、不同类模式分离。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0386]

3.1.9

训练 training

基于机器学习算法，利用训练数据，建立或改进机器学习模型参数的过程。

3.1.10

监督学习 supervised learning

获得的知识的正确性通过来自外部知识源的反馈加以测试的学习策略。

[来源：GB/T 5271.31-2006，31.03.08]

3.1.11

无监督学习 unsupervised learning

一种学习策略，它在于观察并分析不同的实体以及确定某些子集能分组到一定的类别里，而无需在获得的知识上通过来自外部知识源的反馈，以实现任何正确性测试。

[来源：GB/T 5271.31-2006，31.03.09]

3.1.12

强化学习 reinforcement learning

由责任认定改进的学习。

[来源：GB/T 5271.31，31.03.22]

3.1.13

半监督学习 semi-supervised learning

一种学习策略，它自行利用少量的具有标记信息的样本和大量没有标记的样本进行学习的框架。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0215]

3.1.14

自监督学习：self-supervised learning

一种学习策略，通过基于数据本身设计和建立的各种标记信息来对数据本身的特征、特性进行学习，进而把学习到的数据特征网络作为主干网络迁移到对目标任务的学习中。

3.1.15

弱监督学习 weakly supervised learning

一种学习策略，通过使用有噪声的、有限的、不精确的外部信息源进行机器学习。该方法减少了对标注数据质量和数量的要求。

3.1.16

集成学习 ensemble learning

通过结合多个学习器来解决问题的一种机器学习范式。其常见形式时利用一个基学习算法从训练集产生多个基学习器，然后通过投票等机制将基学习器进行结合。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0222]

3.1.17

主动学习 active learning

YY/T xxxx—xxxx

学习过程中由学习器挑选未标记样本，并请求外界提供标记信息，其目标是使用尽可能少的查询来取得好的学习性能。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0216]

3.1.18

迁移学习 transfer learning

利用一个学习领域A上有关学习问题T(A)的知识，改进学习领域B上相关学习问题T(B)的学习算法的性能。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0219]

3.1.19

联邦学习 federated machine learning

一种多方协同建立模型的机器学习框架，各个数据源方进行数据预处理，共同建立其学习模型，并将输出结果反馈给用户。

3.1.20

交叉验证cross validation

一种利用已知数据集获取学习器最有参数，以期在未知数据集上获得最佳泛化性能。常见的有留一法和K重交叉验证法。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0236]

3.1.21

过拟合 overfitting

学习器对训练样本过度学习，导致训练样本中不具有普遍性的模式被学习器当作一般规律，降低了泛化性能；典型表现是训练集上的性能越高，测试集上的性能越低。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0238]

3.1.22

欠拟合 underfitting

学习器对训练样本学习不充分，导致训练样本中包含的重要模式没有被学习器获取，降低了泛化性能；典型表现是训练集上的性能可以继续提高，测试集上的性能同时得以提高。

[来源：计算机科学技术名词ISBN 978-7-03-059487-7，08.0239]

3.1.23

前向传播网络 forward-propagation network

在给定层内的各人工神经元之间既没有反馈路径也没有任何路径的多层网络。

[来源：GB/T 5271.34，34.02.25]

3.1.24

反向传播网络 back-propagation network

一种多层网络，它使用反向传播，以便学习期间的连接权调整。

[来源：GB/T 5271.34，34.02.30]

3.1.25

YY/T xxxx—xxxx

医学AI算法服务 AI algorithm service

满足医学诉求的AI算法在推理部署后的运行态。

算法服务接受用户的应用请求，对输入数据进行处理，返回处理结果。

[来源：T/CESA 1037-2019，3.2，有修改]

3.1.26

医学AI云服务 AI cloud service

通过云计算已定义的接口提供的一种满足医学行业独特需求的一种或多种AI能力。

[来源：GB/T 32400-2015，3.2.8，有修改]

3.1.27

医学AI边缘云服务 AI edge cloud service

通过云计算已定义的接口，部署在AI边缘计算设施上的、满足医学行业独特需求的一种或多种AI能力。

3.1.28

医学AI 系统生命周期模型 AI system lifecycle model

医学AI系统从起始到退役的整个演进过程的框架。包括：需求分析，设计与开发，验证与确认，部署，运维与监控，再评价直至退役。

注：在医学AI生存周期中，某些活动可出现在不同的过程中，个别过程可重复出现。例如为了修复系统的隐错和更新系统，需要反复实施开发过程和部署过程。

3.2 人工智能医疗器械产品分类术语

3.2.1

计算机辅助医学诊断系统 computer-aided diagnosis system

通过人的症状或迹象判断疾病或生理状态、辅助医务人员进行决策的计算系统。

3.2.2

计算机辅助医学识别系统 computer-aided detection system

具备模式识别、数据分析能力，通过识别、标记、强调或其他方式直接提醒医务人员注意医学影像或医疗器械数据的可能异常情况的计算系统。

3.2.3

计算机辅助医学分诊系统 computer-aided triage system

自动分析医学数据、给出初始解释和鉴别分类、辅助医务人员确定患者优先级和（或）就诊科室的计算系统。

3.2.4

临床决策支持系统 clinical decision support system

根据临床知识和患者数据产生辅助决策建议、由医务人员使用的计算机应用系统。

3.2.5

患者决策辅助系统 patient decision assistant system

YY/T xxxx—xxxx

向患者提供辅助决策建议、由非医务人员使用的知识管理系统，结果仅供参考。

3.2.6

医学知识库 **medical knowledgebase**

帮助临床科研人员、临床医生快速、便捷地获取疾病诊断、治疗、用药等全面、系统、动态的临床医学知识的集合，也可作为临床教学及临床诊疗的辅助参考工具。

3.2.7

医疗器械软件 **software as a medical device (SaMD)**

医疗器械软件通常包括软件组件和独立软件，软件组件是指嵌入到医疗器械中或作为医疗器械组成部分的软件，独立软件是指是指预期用于一个或多个医疗目的，无需作为医疗器械硬件组成部分即可完成预期用途的软件。

[来源，IMDRF Software as a Medical Device (SaMD): Key Definitions, 5.1, 有修改]

3.2.8

医用机器人 **medical robot**

在医疗卫生领域，用于诊断、治疗、手术、医用培训等各环节的机器人。

[来源：GB/T 36239-2018, 3.5]

3.2.9

人机交互 **human-robot interaction**

人和机器人通过用户接口交流信息和动作来执行任务。

[来源：GB/T 36239-2018, 2.5]

3.3 数据集术语

3.3.1

数据 **data**

信息的可再解释的形式化表示，以适用于通信、解释或处理。

[来源：GB/T 25000.12-2017, 4.2]

3.3.2

患者数据 **Patient data**

包含患者信息的数据。

3.3.3

个人健康数据 **personal health data**

可用于个人身体状况监测，疾病预防和健康趋势分析的数据。

注：如生物数据（基因等）、生理数据(如血压、脉搏)、环境数据(空气)、心理状态数据、社交数据以及就诊数据。

3.3.4

数据质量 **data quality**

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度

[来源：GB/T 25000.12-2017, 4.3]

3.3.5

数据集 data set

数据记录汇聚的数据形式

[来源: GB/T 35295-2017, 2.1.46]

3.3.6

训练集 training set

用于训练人工智能算法的数据集, 类标记对算法来说已知。

[来源: 计算机科学技术名词ISBN 978-7-03-059487-7, 08.0240, 有修改]

3.3.7

调优集 tuning set

用于调优人工智能算法的数据集, 类标记对算法来说未知。

3.3.8

测试集 testing/validation set

用于测试人工智能算法性能的数据集, 类标记对算法来说未知。

[来源: 计算机科学技术名词ISBN 978-7-03-059487-7, 08.0241, 有修改]

3.3.9

数据集质量 dataset quality

在指定条件下使用时, 数据集的特性满足明确的和隐含的要求的程度

[来源: GB/T 25000.12-2017, 4.3, 有修改]

3.3.10

数据集生存周期 dataset lifecycle

数据集的获取、清洗、标注、存储、整合、分析、应用、呈现、归档和销毁等各种生存形态演变的过程。

[来源: GB 34960.5-2018, 3.7, 有修改]

3.3.11

数据集偏倚 dataset bias

数据集在统计意义上偏离预期用途的程度。

3.3.12

数据元 data element

用一组属性规定其定义、标识、表示和允许值的数据单元

[来源: WS/T 305-2007, 3.1.6]

3.3.13

元数据 metadata

定义和描述其他数据的数据。

[来源: GB 18391.1-2009, 3.2.16]

3.3.14

数据质量特性 data quality characteristic

对数据质量有影响的数据质量属性的类别。

[来源GB/T 25000.12-2017, 4.4]

3.3.15

数据特征层次 data characteristic hierarchy

从不同粗细粒度表征数据特征的数据层次结构。

[来源: GB/T 35295-2017, 2.1.51]

3.3.16

参考标准 reference standard

诊断和治疗过程或基于标注过程建立的基准。

注: 参考标准可包含疾病、生理状态或生理异常以及位置和程度等信息标签。

3.3.17

金标准 gold standard

临床诊断和治疗可依据的最佳测试结果。

3.3.18

真值 ground truth

可用金标准验证的参考标准。

3.3.19

离群值 outlier

在一份数据中, 与其他观察值具有明显不同特征的那些观察值。

3.3.20

缺失数据 missing data

按照研究方案要求收集但未观测到的数据

[来源: 药物临床试验数据管理与统计分析的计划和报告指导原则2016版]

3.3.21

数据清洗 data cleaning

检测和修正数据集中错误数据项以及对数据进行平滑处理等操作的数据预处理过程。

[来源: 计算机科学技术名词ISBN 978-7-03-059487-7, 07.0392]

3.3.22

数据治理 data governance

数据资源及其应用过程中相关管控活动、绩效和风险管理的集合。

[来源: GB 34960.5-2018, 3.1]

3.3.23

YY/T xxxx—xxxx

数据挖掘 data mining

对于定量数据，通过从不同视角和维度分析、分类并总结潜在的联系和影响，以此提取模式的计算过程。

[来源：ISO 16439:2014, 3.13]

3.3.24

标签 label

附加到一组数据元素的标识符。

[来源：ISO/IEC 2382:2015, 2121626]

3.3.25

数据采集 data acquisition

数据由生成装置按照数据采集规范生成，以数字化格式存储并传输到目标系统的过程。

3.3.26

数据脱敏 data masking

对敏感信息通过去标识化或匿名化，实现敏感隐私数据的可靠保护。

3.3.27

数据标注 data annotation

对数据进行分析，建立参考标准的过程。

3.3.28

标注任务 annotation task

按照数据标注规范对指定数据集进行标注的过程。

3.3.29

标注工具 annotation tool

用于辅助标注/仲裁人员产生标注结果的一系列软件/硬件工具。

3.3.30

标注平台 annotation platform

开展标注任务的信息管理系统

3.3.31

标注规则 annotation instruction

数据需求方用于明确标注任务和标注数据的操作规范，应包含标注对象定义、所用标注工具和标注平台、标注格式、标注前的准备工作、标注后的处理工作等。

3.3.32

半自动标注 semi-automatic annotation

使用人工结合自动化工具的方式进行数据标注。

3.3.33

标注流程annotation process

产生标注结果需要遵循的步骤。

注：通常包含依次标注和独立标注两种流程，依次标注流程首先由第一标注人员进行标注，然后其标注结果交由第二标注人员进行审核，如第二标注人员认可该结果，则该结果成为最终的标注结果，否则交由仲裁人员通过仲裁方式决定最终的标注结果。独立标注流程若干名标注人员独立对原始数据进行标注，然后将各自的标注结果通过仲裁方式产生最终的标注结果。

3.3.34

仲裁arbitration

在标注人员对原始数据的标注结果不一致时用于决定最终结果的过程。

3.3.35

仲裁方式arbitration method

在标注人员对原始数据的标注结果不一致时用于决定最终结果的方式。

注：如少数服从多数方式以及仲裁人员个人决定方式。

3.3.36

人员考核personnel examination

是为了保证标注人员/仲裁人员的能力与标注要求一致的测试过程。

3.3.37

图像image

物体通过感光材料、光电传感处理设备和计算机而重显的影像。

[来源：GB/T 34952-2017，2.4]

注：如X线、CT、MRI、超声、核素显像等。

3.3.38

图形graphics

用来表示一个变量相对于其他变量变化情况的线条图，或用以代替文字说明一个概念和思想的图解和表格。

[来源：GB/T 34952-2017，2.3]

注：如生理信号、基因组图谱等。

3.3.39

文本text

以字符、符号、字、短语、段落、句子、表格或者其他字符排列形式出现的数据，旨在表达一个意义，其解释主要以读者对某种自然语言或人工语言的了解为基础。

[来源：GB/T 5271.1-2000，01.01.03]

注：如电子病历等。

3.3.40

音频audio

一种数字化动态媒体形态，用于描述声音及其顺序，给人以时序的感觉，并可进行数字人工合成。

[来源：GB/T 34952-2017，2.5]

3.3.41

视频 video

一种数字化动态媒体形态，用于描述运动图像，进行高速信息传送或显示瞬间的相互关系。

[来源：GB/T 34952-2017, 2.6]

注：如内窥镜影像、超声影像等。

3.3.42

多媒体 multimedia

综合表现文本、图形、图像、音频和视频的信息组合。

[来源：GB/T 34952-2017, 2.1]

3.4 性能指标

3.4.1 质量特性

3.4.1.1

软件质量 software quality

软件产品中能满足规定需求的性质和特性的总体。

[来源：GB/T11457-2001, 2.1293, 有修改]

3.4.1.2

软件质量保证 software quality assurance

为使某项目或产品符合已建立的技术需求提供足够的置信度，而必须采取的有计划和有系统的全部活动的模式。

[来源：GB/T11457-2001, 2.1294]

3.4.1.3

可靠性 reliability

在规定的条件下和规定的时间内，系统或部件执行所要求功能的能力。

[来源：GB/T11457-2001, 2.1334]

3.4.1.4

完整性 integrity

保护资产准确性和完备性的性质。

[来源：GB/T 25000.12-2017, 4.12]

数据创建、传输或存储未经授权不得更改的属性。

[来源：ISO/IEC 29167-19:2016, 3.3]

3.4.1.5

一致性 consistency

在文档或系统或系统部件的各部分之间，一致、标准化、无矛盾的程度。

[来源：GB/T11457-2006, 2.320]

3.4.1.6

重复性 Repeatability

由同一操作员按相同的方法、使用相同的测试或测量设施、在短时间间隔内对同一测试/测量对象进行测试/测量，所获得的独立测试/测量结果间的一致程度。

[GB/T 3358.2-2009, 3.3.5, 有修改]

3.4.1.7

再现性 reproducibility

由不同的操作员按相同的方法，使用不同的测试或测量设施，对同一测试/测量对象进行观测以获得独立测试/测量结果，所获得的独立测试/测量结果间的一致程度。

[来源：GB/T 3358.2-2009, 3.3.10, 有修改]

3.4.1.8

可达性 accessibility

组成软件的各部分便于选择使用或维护的程度。

[来源：GB/T 11457-2006, 2.20]

3.4.1.9

可用性 availability

- a) 软件（系统或部件）在投入使用时可操作或可访问的程度或能实现其制定系统功能的概率；
- b) 系统正常工作时间和总的运行时间之比；
- c) 在运行时，某一配置项实现指定功能的能力。

[来源：GB/T 11457-2006, 2.115]

根据授权实体的要求可访问和使用的特性。

[来源：GB/T 29246-2012, 2.7]

3.4.1.10

安全[性] security

保密[性]

对计算机硬件、软件进行的保护，以防止其受到意外的或蓄意的存取、使用、修改、毁坏或泄密。安全性也涉及对患者隐私、元数据、通信以及计算机安装的物理保护。。

[来源：GB/T 11457-2006, 2.1420]

3.4.1.11

鲁棒性/健壮性 robustness

在存在无效输入或急迫的环境条件下，系统或部件其功能正确的程度。

[来源：GB/T 11457-2006, 2.1397]

3.4.1.12

泛化能力 generalizability

一种评价机器学习算法的指标，描述算法对陌生样本的适应能力。

3.4.1.13

响应时间 response time

在给定的测试环境下，对给定个数数据样本进行运算并获得结果所需要的平均时间。

3.4.1.14

人工智能系统可追溯性 AI system traceability

对AI系统，在整个生存周期，利益相关方应确保对须予考虑的重要事项均存档在案，帮助全面了解本系统的产出如何导出的性质。重要事项包括：数据来源，数据源确认，所实现的过程与决策等。AI系统对其决策过程及结果进行记录的特性。

注：追溯AI系统的决策过程可用来推导AI系统决策的原因并有助于防止错误。

3.4.1.15

人工智能系统公平性 AI system fairness

AI系统做出不涉及喜好和偏袒决策的性质。

3.5 评价方式

3.5.1 质量特性

3.5.1.1

性能performance

系统或部件在给定的约束条件下实现制定功能的程度。

[来源：GB/T 11457-2006, 2.1131]

3.5.1.2

性能评价performance evaluation

为确定运行目标达到了何种有效程度而对系统或系统部件的技术评价。

[来源：GB/T 11457-2006, 2.1132]

3.5.1.3

性能测试performance testing

评价系统或部件与规定的性能需求的依从性的测试行为。

[来源：GB/T 11457-2006, 2.1135]

3.5.1.4

独立性能测试 standalone performance test

通过直接比对模型在没有医生干预的情况下产生的结果和参考标准的结果，评估AI软件的性能。

3.5.1.5

判读者性能研究 reader performance test

通过比对医务人员在独立工作和结合模型工作两种状态下判读病例数据的结果，评估AI软件的性能。

3.5.1.6

多判读者多病例研究 multi-readers multi-cases study

通过招募一组医务人员、随机组合判读人员和病例的方式开展的判读者性能研究。

3.5.1.7

YY/T xxxx—xxxx

黑盒测试 black-box testing

忽略系统或部件的内部机制只集中于响应所选择的输入和执行条件产生的输出的一种测试。

[来源：GB/T 11457-2006, 2.142、2.669]

3.5.1.8

白盒测试 glass-box testing

侧重于系统或部件内部机制的测试。类型包括分支测试、路径测试、语句测试。

[来源：GB/T 11457-2006, 2.678、2.1604]

3.5.1.9

对抗样本 adversarial examples

基于原始数据生成的、在数据特性或参数方面添加扰动的新样本。

3.5.1.10

对抗测试 adversarial test

使用对抗性样本开展的测试，或采用不同于目标人群比例的特选数据作为压力数据集进行的测试。

3.5.2 评价指标

3.5.2.1

阳性样本 positive sample

由参考标准确定为带有某一特定特征的样本。

3.5.2.2

阴性样本 negative sample

除阳性样本以外的样本。

3.5.2.3

真阳性 true positive

TP

被算法判为阳性的阳性样本。

3.5.2.4

假阳性 false positive

FP

被算法判为阳性的阴性样本。

3.5.2.5

真阴性 true negative

TN

被算法判为阴性的阴性样本。

3.5.2.6

假阴性 false negative

YY/T xxxx—xxxx

FN

被算法判为阴性的阳性样本。

3.5.2.7

目标区域 Target region

在影像评价中,根据参考标准从原始数据中划分出的若干个包含特定类别目标的最小数据子集(子集元素为像素,体素等)。

3.5.2.8

分割区域 Segmentation region

在影像评价中,人工智能医疗器械从原始数据中划分出的若干个包含特定类别目标的最小数据子集(子集元素为像素,体素等)。

3.5.2.9

病变定位 Lesion localization

算法检出病变位置正确标识出参考标准确定的病变位置。

3.5.2.10

非病变定位 Non-lesion localization

算法检出病变位置未能正确标识出参考标准确定的病变所在位置。

3.5.2.11

病变定位率 Lesion localization fraction

病变定位数量占由参考标准确定的全体病变数量的比例,也可称为召回率(见3.5.2.14)。

3.5.2.12

非病变定位率 Non-lesion localization fraction

非病变定位数量占全体病例数量的比例,非病变定位率可以大于1,也可称为平均假阳个数。

3.5.2.13

假阳性率 false positive fraction

假阳性病例数量(阴性病例中包含非病变定位)占全部阴性病例数量的比例。

3.5.2.14

灵敏度 sensitivity

召回率(查全率) recall

真阳性样本占全体阳性样本的比例;
分割区域与目标区域的交集占目标区域的比例。

3.5.2.15

特异度 specificity

真阴性病例占全体阴性病例的比例。

3.5.2.16

YY/T xxxx—xxxx

漏检率 miss rate

1减去灵敏度。

3.5.2.17

精确度（查准率） precision

阳性预测值 positive prediction value

真阳性样本占被算法判为阳性样本的比例；
分割区域与目标区域的交集占分割区域的比例；
病变定位数量占由算法确定的全体病变数量的比例。

3.5.2.18

阴性预测值 negative prediction value

真阴性样本占被算法判为阴性样本的比例。

3.5.2.19

准确率 accuracy

算法诊断正确的样本占全体样本的比例。

3.5.2.20

F1度量 F1-measure

召回率和精确度的调和平均数。

3.5.2.21

约登指数 Youden index

灵敏度与特异度之和减去1。

3.5.2.22

受试者响应曲线 receiver operating characteristics curve

ROC

通过在一组预设的阈值下计算人工智能算法在测试集上的灵敏度（Sensitivity）以及特异度（Specificity）从而产生一组（1-Specificity, Sensitivity）操作点，将操作点依次连接形成受试者操作曲线。

3.5.2.23

曲线下面积 area under curve

AUC

受试者操作曲线下的积分面积。

3.5.2.24

自由受试者响应曲线 free receiving operating characteristics curve

fROC

由算法在一组阈值设定下对于给定的测试集得到的一组病变定位率/召回率为纵轴，及非病变定位率/平均假阳个数为横轴构造的曲线。

3.5.2.25

候选自由受试者操作曲线 alternative free receiver operating characteristics curve
AFROC curve

由算法在一组阈值设定下对于给定的测试集得到的一组病变定位率/召回率为纵轴，及假阳性率为横轴构造的曲线。

3.5.2.26

精确度-召回率曲线 precision-recall curve

P-R

由算法在一组阈值设定下对于给定的测试集得到的一组召回率为纵轴，精确度为横轴构造的曲线。

3.5.2.27

平均精确度 average precision

AP

精确度-召回率曲线下的积分面积。

3.5.2.28

平均精确度均值 mean average precision

MAP

在多目标检测问题上，算法对于各类目标的平均精确度的平均值。

3.5.2.29

交并比 intersection over union

IoU

分割区域与目标区域交叠的程度，可用Dice系数（分割区域与目标区域的交集占分割区域与目标区域平均值的比例）或Jaccard系数（分割区域与目标区域的交集占分割区域与目标区域并集的比例）表示。

在自然语言处理其定义为两个集合交集大小与并集大小之间的比例。

3.5.2.30

中心点距离 central distance

分割区域中心与目标区域中心的距离，该指标反映两个集合的接近程度。

3.5.2.31

混淆矩阵 confusion matrix

含混矩阵

一种矩阵，它按一组规则记录试探性实例的正确分类和不正确分类的个数。

[来源：GB/T 5271.31-2006, 31.02.18]

注：通常矩阵的列代表人工智能的诊断结果，而矩阵的行代表参考标准的诊断结果。

3.5.2.32

Kappa系数, Kappa coefficient

用于评价人工智能诊断与参考标准诊断结果的一致性的指标。

3.5.2.33

信噪比 signal-to-noise ratio

信号平均功率水平与噪声平均功率水平的比值。

3.5.2.34

峰值信噪比 peak signal-to-noise ratio

信号最大可能功率和影响它的表示精度的破坏性噪声功率的比值。

3.5.2.35

结构相似性 structural similarity

是一种衡量两幅图像相似度的指标。

3.5.2.36

余弦相似度 Cosine Similarity

自然语言处理评价指标，通过测量两个向量的夹角的余弦值来度量它们之间的相似性。

3.5.2.37

困惑度 Perplexity

度量概率分布或概率模型的预测结果与样本的契合程度，困惑度越低则契合越准确。

3.5.2.38

字错率 (Word Error Rate)

一种语音识别评价指标，将识别出来的字需要进行修改的字数与总字数的比值。

3.5.2.39

交叉熵 cross-entropy

一种度量两个概率分布之间差异的指标。

3.5.2.40

互信息 Mutual Information

对两个随机变量间相互依赖性的量度。

3.5.2.41

服务可用性 service availability

服务客户发起服务请求后，服务可访问的时间占总服务时间的比例。

注：可用性的计算是在一系列预定义的时间段中，服务可用时间之和占预定义时间段之和的比例，可排除允许的服务不可用时间。

3.6 安全术语

3.6.1

资产 asset

对个人或组织有价值的任何东西。

[来源：GB/T 29246-2012, 2.3]

3.6.2

攻击 Attack

破坏、泄露、篡改、损伤、偷窃、未授权访问或未授权使用资产（3.5.1）的企图。

[来源：GB/T 29246-2012，2.4]

3.6.3

真实性Authenticity

一个实体正是其所声称实体的特性。

[来源：GB/T 29246-2012，2.6]

3.6.4

授权Authorization

获得访问数据和功能的权限

[来源：ISO 27789:2013，3.12]

3.6.5

保密性 confidentiality

信息不能对未授权的个人、实体或过程可用或泄露的特性。

[来源：ISO 27789:2013，2.9]

3.6.6

网络安全 Cybersecurity

保护信息和系统不受未授权活动的一种状态，未授权活动包括访问、使用、披露、中断、修改、或销毁，以使保密性、完整性和可用性的相关风险在整个生命周期内保持在可接受的水平。

[来源：ISO 81001-1，3.56]

3.6.7

利用 Exploit

明确用于通过漏洞破坏信息系统安全的方法

[来源：ISO/IEC 27039:2015，2.9]

3.6.8

抗抵赖性 Non-Repudiation

证明所声称事态或行为的发生及其发起实体的能力，以解决有关事态或行为发生与否以及事态中实体是否牵涉的争端

[来源：GB/T 29246-2012，2.27]

3.6.9

隐私 Privacy

不因不正当或非法收集和使用有关个人的数据而侵犯其私人生活或事务的自由

[来源：ISO/TS 27799:2009，3.17]

3.6.10

YY/T xxxx—xxxx

个人信息 personal information

以电子或者其他方式记录的能单独或者与其他信息结合识别特定自然人身份或者反应特定自然人活动情况的各种信息。

[来源：GB/T 35273-2017, 3.1]

3.6.11

个人一般信息 personal general information

除个人敏感信息以外的个人信息。

[来源：GB/Z 28828—2012, 3.8]

3.6.12

个人敏感信息 personal sensitive information

一旦泄露、非法提供或滥用可能危害人身和财产安全，极易导致个人名誉、身心健康受到损害或歧视性待遇等的个人信息。

注：个人敏感信息包括身份证件号码、个人生物识别信息、银行账号、通信记录和-content、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息、14岁以下（含）儿童的个人信息等。

[来源：GB/T 35273-2017, 3.2]

3.6.13

访问控制 access control

基于业务要求和安全要求，确保授权和受限的访问资产的手段

[来源：GB/T 29246-2012, 2.1]

3.6.14

脆弱性 vulnerability

可能会被威胁所利用的资产或控制措施的弱点。

[来源：GB/T 29246-2012, 2.46]

3.6.15

保密性 confidentiality

信息不能被未授权的个人、实体或者过程利用或知悉的特性。

[来源：GB/T 29246-2012, 2.9]

3.6.16

欺骗 spoofing

假冒成合法的资源或用户

[来源：GB/T 25068.3-2010, 3.21]

3.6.17

威胁 threat

可能导致对系统或组织的损害的不期望事件发生的潜在原因。

[来源：GB/T 29246-2012, 2.45]

3.6.18

YY/T xxxx—xxxx

对抗[措施] countermeasure

特定的攻击利用数据处理系统特定的脆弱性的可能性

[来源: GB/T 25069-2010, 2.1.4]

3.7 应用场景术语

3.7.1

计算机辅助 computer-aided

属于使用计算机完成部分工作的技术或过程。

[来源: ISO 2382-2015, 2121395]

3.7.2

专家系统 expert system

一种基于知识的系统,他根据由人类专家经验开发出的知识库进行推理,来解决某一特定领域或应用范围中的问题。

[来源: GB/T 5271.28-2001, 28.01.06]

3.7.3

计算机视觉 computer vision

功能单元获取处理和解释可视数据的能力。

[来源: GB/T 5271.28, 28.01.19]

3.7.4

语音识别 speed recognition

通过功能单元对人的语音所表示信息的感知与分析。

[来源: GB/T 5271.28-2001, 28.01.15]

3.7.5

自然语言处理 natural language processing

自然语言理解和生成及其衍生技术,以从文本化的人类语言中获取有意义的信息。

3.7.6

知识图谱 knowledge graph

- 1) 将海量知识及其相互联系组织在一张大图中,用于知识的管理、搜索和服务;
- 2) 特指谷歌公司开发的知识图谱。

[来源: 计算机科学技术名词 ISBN 978-7-03-059487-7, 08.0688]

3.7.7

医学图像处理 medical image processing

一类对医学图像进行图像处理的方法,包括图像重建、图像增强、图像识别、图像分割、图像配准、图像可视化等。

3.7.8

医学图像分割 medical image segmentation

YY/T xxxx—xxxx

一种医学图像处理方法，根据临床治疗或研究需求把医学图像分成若干个特定的、具有独特性质的区域，并提取出图像中包括器官、病灶等感兴趣目标的技术和过程。

3.7.9

医学图像分类medical image classification

一种医学图像处理方法，根据医学图像信息中所反映的不同特征，对不同类别的医学图像进行分类。

3.7.10

医学图像配准medical image registration

一种医学图像处理方法，将不同时间、空间、模态采集的医学图像通过算法映射到另一个坐标系的过程。

3.7.11

医学图像目标检测medical image object detection

一种医学图像处理方法，从医学图像中找出包括病灶、器官、组织等在内的感兴趣的目标，并确定其位置和类别。

3.7.12

医学图像成像加速 Medical imaging acceleration

一种医学图像成像技术，通过超分辨率重建、低剂量重建等加速算法提高各种医学成像设备的成像速度。

3.7.13

医学图像模态转换Medical imaging modality transformation

一种医学图像处理方法，从一种影像模态转换到另一种影像模态。

附 录 A
(资料性附录)
评价指标计算公式说明

A.1 辅助诊断性能

人工智能医疗器械辅助诊断功能多涉及分类问题，算法将患者数据输出两个或两个以上互斥（不相关）的类别或状态。根据输出的类别或状态的数量构成混淆矩阵，通过混淆矩阵的各个参量来评价辅助诊断性能。

A.1.1 混淆矩阵

对于分类问题，混淆矩阵的一般形式如表 A.1 所示：

表 A.1 n 分类混淆矩阵

分类	Pred_1	Pred_2	Pred_n
True_1	$N_{1,1}$	$N_{1,2}$
True_2	...	$N_{2,2}$
...
...
...
True_n	$N_{n,n}$

注：Pred_x (x=1~n) 为人工智能诊断为 x 类的类别；True_x (x=1~n) 为参考标准诊断为 x 类的类别； $N_{i,j}$ (i=1~n, j=1~n) 为参考标准的诊断结果为 i 类，被人工智能诊断为 j 类的个数；n 为分类类型个数。

二分类的混淆矩阵可简化为表 A.2 所示：

表 A.2 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	FN
	阴性	FP	TN

多分类实际可转化为二分类问题，参考标准分类为 i 类与其他类别的混淆矩阵简化形式如表 A.3 所示：

表 A.3 多分类实际可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	$TP = N_{i,i}$	$FP = \sum_{j=1, j \neq i}^n N_{i,j}$
	阴性	$FN = \sum_{j=1, j \neq i}^n N_{j,i}$	$TN = \sum_{j=1, j \neq i}^n \sum_{l=1, l \neq i}^n N_{j,l}$

A. 1. 2 灵敏度

灵敏度用 Sen 表示，表达式见式 (A1)：

$$Sen = \frac{TP}{TP + FN} \times 100\% \quad (A1)$$

式中：

Sen ——灵敏度；

TP——真阳性样本的个数；

FN——假阴性样本的个数。

A. 1. 3 特异度

特异度用 Spe 表示，表达式见式 (A2)：

$$Spe = \frac{TN}{FP + TN} \times 100\% \quad (A2)$$

式中：

Spe ——特异度；

TN——真阴性样本的个数；

FP——假阳性样本的个数。

A. 1. 4 漏检率

漏检率用 MR 表示，表达式见式 (A3)：

$$MR = 1 - Sen \quad (A3)$$

式中：

Sen ——灵敏度；

MR——漏检率。

A. 1. 5 阳性预测值

阳性预测值用 PPV 表示，表达式见式 (A4)：

$$PPV = \frac{TP}{TP + FP} \quad (A4)$$

式中：

PPV ——阳性预测值；

TP ——真阳性样本的个数；

FP ——假阳性样本的个数。

A. 1. 6 阴性预测值

阴性预测值用 NPV 表示，表达式见式 (A5)：

$$NPV = \frac{TN}{FN + TN} \quad (A5)$$

式中：

NPV ——阴性预测值；

TN ——真阴性样本的个数；

FN ——假阴性样本的个数。

A. 1. 7 准确率

准确率用 Acc 表示，表达式见式 (A6)：

$$Acc = \frac{\sum_{i=1}^n N_{i,i}}{\sum_{j=1}^n \sum_{l=1}^n N_{j,l}} \quad (A6)$$

A. 1. 8 约登指数

约登指数用 Y 表示，表达式见式 (A9)：

$$Y = Sen + Spe - 1 \quad (A9)$$

A. 1. 9 受试者响应曲线

示例如图 A.1 所示：

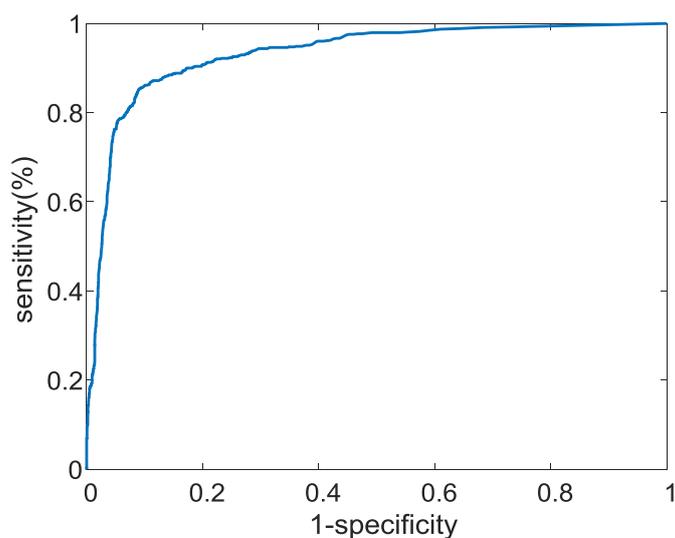


图 A.1 roc 曲线

横坐标和纵坐标值均0-1之间，曲线下面积AUC最大值为1。

A.2 辅助探测性能

人工智能医疗器械辅助探测功能多为对病例中的病变进行识别检出，如在医学影像应用中找到病变在图像中的位置。举例：在N个病例样本中，共Les个病变，其中LL个被正确识别，NLL个误诊，N个病例中有M个病例包含非病变定位。

A.2.1 召回率（病变定位率）

召回率用 Rec 表示，表达式见式（A10）：

$$Rec = \frac{LL}{Les} \times 100\% \quad (A10)$$

式中：

Rec——召回率；

LL——算法检出病变位置正确识别出参考标准确定的病变位置的数量；

Les ——为参考标准确定的全体病变数量。

A.2.2 精确度

精确度用 Pre 表示，表达式见式（A11）：

$$Pre = \frac{LL}{LL + NLL} \times 100\% \quad (A11)$$

式中：

Pre——精确度

LL——算法检出病变位置正确识别出参考标准确定的病变位置的数量；

YY/T xxxx—xxxx

NLL——算法检出病变位置未能正确识别出参考标准确定的病变位置的数量。

A. 2.3 非病变定位率

非病变定位率用 NLF 表示，表达式见式 (A12)：

$$NLF = \frac{NLL}{N} \times 100\% \quad (\text{A12})$$

式中：

NLF——非病变定位率；

NLL——算法检出病变位置未能正确识别出参考标准确定的病变位置的数量；

N——全体病例数量。

A. 2.4 假阳性率

假阳性率用 FPF 表示，表达式见式 (A13)：

$$FPF = \frac{M}{N} \times 100\% \quad (\text{A13})$$

式中：

FPF——假阳性率；

M——全体病例数量非病变定位的病例个数；

N——全体病例数量。

A. 2.5 F1度量

F₁ 度量用 F₁ 表示，表达式见式 (A14)

$$F_1 = \frac{2 \times LLF \times Pre}{LLF + Pre} \quad (\text{A14})$$

式中：

F₁——F1 度量；

LLF——病变定位率；

Pre——精确度。

A. 2.6 自由受试者响应曲线

示例如图 A.2 所示：

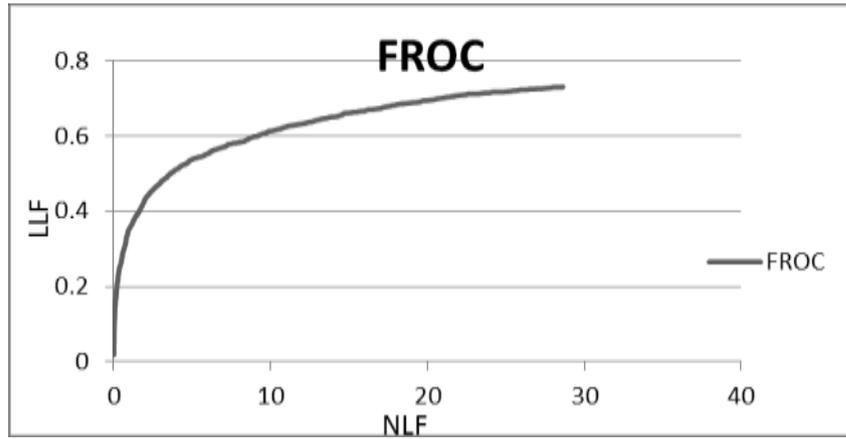


图 A.2 FROC 曲线

A. 2. 7 精确度-召回率曲线

示例如图 A.3 所示:

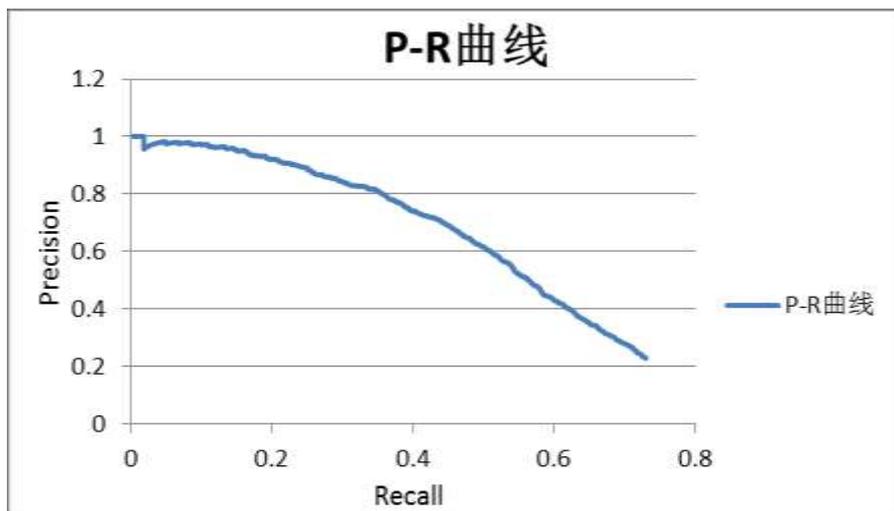


图 A.3 P-R 曲线

A. 2. 8 平均精确度

平均精确度用 AP 表示, 表达式见式 (A15):

$$AP = \int_0^1 Pre(Rec)d(Rec) \tag{A15}$$

式中:

AP——平均精确度;

Pre——精确度;

Rec——召回率。

A. 2. 9 平均精确度均值

平均精确度均值用 MAP 表示，表达式见式 (A16)：

$$MAP = \frac{\sum_{n=1}^N AP(n)}{N} \quad (\text{A16})$$

式中：

MAP——平均精确度均值；

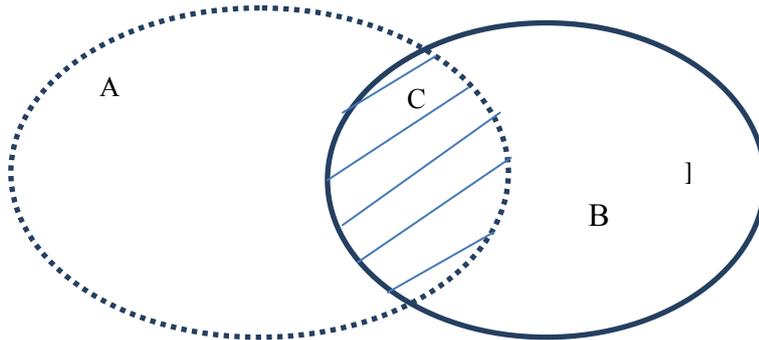
AP——平均精确度；

N——类别的数量；

n——类别n。

A.3 辅助分割性能

人工智能医疗器械辅助分割功能多为对医学影像病变进行识别检出，并进行病变边界的分割，如图 A.4 所示，A 为目标区域，B 为分割区域，C 为二者的交集 ($C = A \cap B$)。



图A.4 辅助分割示意图

A.3.1 召回率

召回率用 Rec 表示，表达式见式 (A17)：

$$Rec = \frac{A \cap B}{A} \quad (\text{A17})$$

式中：

Rec——召回率；

A——目标区域；

B——分割区域。

A.3.2 精确度

精确度用 Pre 表示，表达式见式 (A18)：

$$Pre = \frac{A \cap B}{B} \quad (\text{A18})$$

YY/T xxxx—xxxx

式中:

Pre——精确度;

A——目标区域;

B——分割区域。

A. 3.3 Dice系数

Dice 系数用 *Dice* 表示, 表达式见式 (A19):

$$Dice = \frac{2 \times A \cap B}{A + B} \quad (A19)$$

式中:

Dice——Dice 系数;

A——目标区域;

B——分割区域。

A. 3.4 Jaccard系数

Jaccard 系数用 *Jaccard* 表示, 表达式见式 (A20):

$$Jaccard = \frac{A \cap B}{A \cup B} \quad (A20)$$

式中:

Jaccard——Jaccard 系数;

A——目标区域;

B——分割区域。

参 考 文 献

- [1] 周志华. 机器学习[M]. 清华大学出版社, 2016
 - [2] 方积乾. 卫生统计学. 第7版[M]. 人民卫生出版社, 2012.
 - [3] Petrick N, Sahiner B, Armato SG, et al. Evaluation of Computer-aided Detection and Diagnosis Systems[J]. MedPhys, 2013, 40(8):87001.
 - [4] Sahiner B, Chen W, Pezeshk A, et al. Semi-parametric Estimation of the Area under the Precision-recall Curve[C]. Spie Medical Imaging, 2016.
 - [5] O. Chapelle, B. Scholkopf and A. Zien, Semi-Supervised Learning[J]. MIT press, 2006
 - [6] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning[M]. The MIT Press, 2016.
-