ICS 01.040.11.

C 50

DB13

河北省地方标准

DB 13/T XXXX—2021

|  |
| --- |
|  |

医学影像学大数据智能应用技术规范

2021 - XX - XX发布

2020 - XX - XX实施

河北省市场监督管理局   发布

目  次

**[前  言 II](#_Toc22427)**

**[引  言 III](#_Toc3249)**

[1 范围 1](#_Toc7243)

[2 规范性引用文件 1](#_Toc12017)

[3 术语和定义 1](#_Toc8674)

[4 医学影像学大数据建立流程与基本细则 3](#_Toc17940)

[4.1 医学影像学大数据概念 3](#_Toc10592)

[4.2 医学影像学大数据特点 4](#_Toc31496)

[4.3 规范建立影像大数据standardize the establishment of image big data 4](#_Toc4570)

[5 医学影像大数据的目录管理与服务要求 6](#_Toc28343)

[6 人工智能技术对医学影像学大数据挖掘和应用 13](#_Toc26134)

[6.1 快速医学影像成像方法： 13](#_Toc16458)

[6.2 医学图像质量增强的方法 14](#_Toc29226)

[6.3 智能图像分析技术 intelligent image analysis technology 16](#_Toc22216)

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。

本文件由河北省卫生健康委员会提出。

本文件起草单位：河北医科大学第二医院。

本文件主要起草人：刘怀军等。

引  言

根据医学影像学技术快速发展及智能诊断的要求，适应互联网+医疗及5G技术的需求，参考相关法规和文献，特制定本标准。

本部分描述了医学影像学大数据定义、特点、建立过程及医学影像学智能诊断规范应用。

医学影像学大数据，是指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资源,并适用于5G信息技术。

医学影像学智能诊断是指运用人工智能技术（是计算机科学的一个分支)，了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域包括医学机器人、语言识别、图像识别、自然语言处理和医学专家系统等技术要素，分析医学影像,并将影像与医疗文本记录进行交叉对比,就能够极大地降低医学诊断上的失误,帮助医生精准诊断,挽救患者生命。

人工智能在医学影像的应用主要分为两个部分：第一个部分是图像识别；第二个部分深度学习，是人工智能应用的最核心环节。这两个部分都是基于医学影像学大数据所进行的数据上的挖掘和应用。

医学影像学大数据智能应用技术规范

1. 范围

本文件规定了医学影像学大数据与智能技术的基本要求，制定了基本目录和总体结构，描述了基本内容和功能、工作流程以及各部分之间的相关性。

本文件适用于各医疗机构医学影像学大数据内容与智能技术的基本要素体系的规划和设计。

1. 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。

本文件没有规范性引用文件。

1. 术语和定义

下列术语和定义适用于本文件。

* 1. 医学影像学大数据 Medical imaging big data

医学影像学大数据是指信息网络技术与传统的公共卫生、医疗服务、医疗保障、健康、保健、医院综合管理等大业务深度融合，以及互联网+健康医疗、精准医学、人工智能等新兴领域，所产生的与医疗、卫生、健康、保健等医学影像学相关的数据。

* 1. 目录服务器 catalog server

按照目录服务器接口的要求，提供医学影像学大数据与智能技术信息资源发现和目录管理的计算机服务程序。

* 1. 元数据库 metadata repository

元数据库存储用于医学影像学信息资源共享活动建立相关目录所需要的基本元数据、扩展元数据内容。

* 1. 结果集 result set

根据目录检索请求在服务器端间的查询结果集合。

* 1. 多单元组学数据库（Aging Atlas数据库)

该数据库从常规放射学单元学组、CT影像学单元数据组、MRI单元数据组、介入性放射学单元数据组等不同层面整合了医学影像学大数据相关数据集，实现了不同条件下医学影像学各组数据的相互汇聚融合。

* 1. 医学影像学信息元数据元素 metadata element

元数据的基本单元，用以描述医学影像学大数据与智能技术信息资源的某个特性。注：元数据元素在元数据实体中是唯一的。

* 1. 医学影像学信息元数据实体 metadata entity

一组说明医学影像学大数据与智能技术信息资源相关特性的元数据元素。注：可以包含一个或一个以上元数据实体。

* 1. 医学影像学信息基本元数据 basic metadata

描述信息资源基本属性的元数据和元数据实体。

* 1. 标识符 identifier

用于标识或命名一个医学影像学数据项目并可能指出该数据信息确定性质的一个或一组字符。

* 1. 医学影像学大数据与智能技术资源标识符 big data for health care resource identifier

用于唯一标识医学影像学大数据与智能技术资源的一组字符。

* 1. 医学影像学大数据与智能技术资源分类 resource classification

根据资源的特性及其相互关系而分别排列出的不同次序、类别和体系。

* 1. 医学影像学大数据与智能技术标识符的构成 Composition of identifiers

基于卫生信息OID标识体系对医学影像学大数据与智能技术资源进行标识，依据我国《卫生信息标识体系对象标识符编号结构与基本规则》和《卫生信息标识体系对象标识符管理注册管理规程》等相关规定进行标示符的构成管理。

* 1. 医学影像学大数据与智能技术类目码 Category codes

类目码是省级代码下一级节点编码，用来区分不同的业务应用或者数据业态。

* 1. 医学影像学大数据与智能技术亚目码 Subjects code

亚目码是类目码对应节点的下一级节点编码，亚目是根据各类目领域的特点，按照业务内容的组成部分或业务流程的先后顺序进行分类。

* 1. 医学影像学大数据与智能技术细目码 Details code

细目码是亚目码后节点编码，细目编制是一个逐步完善的过程，通过医学影像学大数据与智能技术 资源目录管理系统实现大范围注册后，才可能逐步编目完成一个较为全面、科学的资源目录细目。

1. 医学影像学大数据建立流程与基本细则
   1. 医学影像学大数据概念
      1. 医学影像学大数据是由DR、CT、MR等医学影像设备所产生并存储在PACS系统（医学影像系统）内的大规模、高增速、多结构、高价值和真实准确的影像数据集合。与医院信息系统（HIS）大数据、检验信息系统（LIS）大数据和电子病历（EMR）等同属于医疗大数据的范畴。
      2. 多结构和高价值：由日益增长的医疗影像设备种类所产生的具有医学分析和指导价值的结构化和非结构化数据。
      3. 大规模和高增速：PACS系统的普及率和人口数量的庞大，是我国医学影像学大数据的大规模基础；而PACS系统和老龄人口的快速增长率是医学影像学大数据的高增速的基础，这两者共同构成了我国医学影像学大数据的形成原因。
      4. 保证医学影像学大数据真实性的相关技术：一类是直接处理技术，对患者进行影像学检查后，直接采用软件技术对影像在影像设备上进行处理，这种方式的缺点比较明显，不能对影像进行改变，只能依靠医生依据自身的经验进行病理学处理，导致了数据结果的不准确性；第二种是PACS系统来对影像进行后处理。比如PACS系统通过多维影像融合（CT/MRI/PET-CT）技术，来对图像进行分割、配准和聚类，从而尽可能保存影像数据的真实性。
      5. 多维影像融合主要包括数据预处理、图像分割、特征提取以及匹配判断这几个过程。
   2. 医学影像学大数据特点
      1. 大数据：传统数据架构无法有效处理的一种新型数据集，容量大、维度高、范围广、复杂多变
      2. “5V+2D”
      * Volume（规模）： 规模大，包括采集、储存和计算的量非常大。1个CT检查图像的存储量是200-500M，1个标准病理检查可达5G。就一所医院而言，它包含有规模巨大的临床实验数据，错踪复杂的疾病诊疗数据，带有个性化特征的居民行为健康数据以及相关管理、医保、药品、器械等数据。
      * Velocity（速度）：数据增长速度快，数据的采集、存储和计算速度也快，时效性要求高。面对大量患者的就诊、检查、治疗、随访数据快速增长，需要大数据的高速处理，及时应用。
      * Variety（多样）：种类和来源多元化。互联网时代加剧了档案数据的多元化，包含结构化表格、半结构化文本、非结构化影像，而它的具体形式包括单纯数据（检验）、动态信号（示波）、多维影像、文字绘图、视频档案，等等。
      * Value（价值）：大数据的价值密度相对较低，或者说大浪淘沙却又弥足珍贵。
      * Variability（易变性）：庞大数据不是抽样、不是标准，更易包括变异。病例脱失、过程中断、记录偏差和残缺、医生个体认识差异以及很多重复、无关、甚至相互矛盾的记录都会引发造成数据的不确定性。
      * Deadline（时效性）:提供服务的大数据需要具有一定的鲜活性和实效性。疾病的发生、发展、就诊、检查、治疗随时间变化，所以在利用数据时，只有挖掘在“保质期”内的数据，才能创造最大价值。
      * Dynamic(动态性）**：**随着时间的流逝，数据的保存价值也在动态变化。
   3. 规范建立影像大数据standardize the establishment of image big data

规范整个影像检查流程（影像数据获取、图像后处理、影像分析思维训练、结构化诊断报告、影像数据规范存储、调用）。

* + 1. 影像数据规范存储、调用 standard storage and call of image data

不同医院制度与规范的差异，某些情况下会做重复性检查，尤其是核磁共振成像、计算机断层扫描等检查费用高昂的项目。

诊断决策的时间， 可能造成治疗延误。

通过云计算解决方案，可以使数字存储的价格降低。在云端存储病人数据最安全的方法之一是分裂合并技术。受保护的健康信息被分别加密储存，进而创造一个互联网安全的影像学研究。

* + 1. 存储--云端解决方法Storage Cloud solution

云端运算具有高扩充性、灵活性、低成本且随时随地存取，使用者通过因特网就能得到云端运算的服务资源。云端平台数据库的设计，采用共同的平台和系统，遵循健康信息交换（health information exchange，HIE）协议，以及基于标准的界面能够实现医疗系统之间的互操作。需要高速的运算性能、良好的兼容性及可扩展性的计算机集群。集群主要由计算节点、存储节点、管理节点、集群辅件组成。

形式之一：列表（Lists）、集合（Sets）、有序集合（Sortedsets）、哈希表（Hashes）。键值的数据类型决定了该键值支持的操作。Redis支持诸如列表、集合或有序集合的交集、并集、查集等高级原子操作。

基于Hadoop的海量图像数据管理系统，可以实现基于Map/Reduce的Sobel边缘检测和图像并行直方图提取等图像处理操作。

* + 1. 结构化诊断报告 structured diagnostic report

结构化诊断报告对于大量医学影像的数据归档和检索具有重要意义，能够提升医疗文档中的数值记录功能。即支持自由文档报告，又支持结构化信息，还支持将文本和数据与病证表达的关键图像进行关联，形成图文报告，消除了图像与数据分离的弊端，也可实现自动诊断。

* + - 1. 统一术语的建立

本标准参考北美放射学会(RSNA)在SNOMED CT、LOINC和UMLS的基础上编制的标准化的放射学术语集RadLex。RadLex提供了能够捕述医学影像学全部信息的标准术语，为结构化报告提供了标准字典库，也为医学影像学信息资源开发利用提供了标准的检索语言。

本标准参考HL7组织发布的通过XML进行临床文档信息交换的医疗行业标准--临床文档体系(clinical document architecture，CDA)。CDA采用XML（可扩展标记语言）技术描述病历结构，其中的语义借助于HL7的参考信息模型和编码词汇表进行定义。以HL7 CDA为信息交换标准，结合XML技术，实现结构化报告的数据在不同医疗信息系统间的传输，将数据与表现形式分离，有利于数据的挖掘和分析。

* + - 1. 结构化数据的实现

基于本体(ontology)进行结构化表达在医学领域目前被认为是高效的一种方案。本体是对概念体系明确的、形式化的、可共享的规范。

在本体的参与下从医学图像对应的文本信息，如扫描协议的关键字中提取知识获得方向，再选用事先定义好的处理方法对图像进行标注建立索引，通过专业人员的判定，进行知识的确认。实现对医学影像结构化的操作转变为对概念的检索。

当用户检索时，首先从用户自然语言形式的查询中提取出概念；用得到的概念以一定的检索策略去检索概念索引库，将得到的检索结果，包括病证分布、规律以及关键图像返网给用户，关键图像、测量值、诊断结论可以以RadLex术语集，通过HL7 CDA或DICOM SR的形式输出。

1. 医学影像大数据的目录管理与服务要求
   1. 目录体系 Catalog system
      1. 交换体系 Exchange system

构建基于省、市、县区域性三级卫生健康信息平台的大数据资源目录服务平台，实现医学影像学信息的开放共享、云业务协同。省级资源目录服务平台除了实现省内资源信息的融合外，还提供城市间跨区域资源共享服务，制定市级跨域共享规则，协同调度；市级资源目录服务平台除了实现市内资源信息的融合外，还提供城市内跨区（县）域资源共享服务，制定区（县）级跨域共享规则，协同调度；县级资源目录服务平台直接提供各类数据开放服务，实现区（县）域内资源信息的融合,重点支撑上级平台协同调度功能。

* 1. 技术总体架构 Technical Master Plan
     1. 组成

医学影像学大数据资源目录体系技术总体架构包括资源信息库、资源目录中心和资源交换中心。资源信息库由资源共享数据中心、目录内容信息库、目录内容管理信息库、目录内容服务信息库、元数据库组成。

* + 1. 资源共享数据中心

资源共享数据中心存储本医疗机构或其他医疗机构用于机构或部门群间共享的信息资源，包括数据结构化数据、非结构化/半结构化文本、医疗影像数据、流数据、医院数据、系统数据等多种类型的电子化信息资源。

* + 1. 目录内容信息库

目录内容信息库存储由编目系统提取共享信息资源的基本特征编目形成的目录内容。

* + 1. 目录内容管理信息库

目录内容管理信息库存储各卫生健康信息主管部门注册并通过审核到目录服务中心的内容。

* + 1. 目录内容服务信息库

目录内容管理信息库存储用于发布的目录内容。目录服务中心使用目录内容服务信息库提供目录查询检索等服务。

* 1. 资源目录中心 Resource Directory Centre
     1. 编目管理系统

编目管理系统根据资源共享数据库的内容，提取其基本特征，按照T/CHIA 17-3实现元数据赋值，形成目录内容。应提供资源目录内容编目、资源目录内容注册等功能。

* + 1. 目录内容管理系统

目录内容管理系统实现对目录服务中心的目录内容和目录服务运行的管理，应提供目录内容审核、标识符管理、资源目录分类管理、资源关联性管理、资源共享和交换规则的设置、资源目录内容的发布管理等功能。

* + 1. 目录内容服务系统

目录内容服务系统基于目录内容服务信息库，向用户提供目录内容的资源查询检索、 资源下载、资源订阅、资源申请、资源退订、资源评估等服务。目录服务系统应提供两种方式的服务-储存资源和查找资源。

* + 1. 元数据管理系统

元数据管理系统实现对元数据的管理，应提供元数据属性管理、元数据实体管理、元数据集管理、元数据版本管理等功能。

* 1. 资源交换中心 Resource Exchange Centre
     1. 交换管理子系统

交换管理子系统对医学资源信息订阅共享交换进行运行管理。交换管理子系统提供部署维护、交换配置、运行管理、审计管理等功能。

* + 1. 交换传输子系统

交换传输子系统通过交换服务实现医学资源信息内容在不同医院或医院部分群的使用者和资源共享数据库之间的传输。交换传输子系统提供消息传输、消息代理、基础服务等功能。

* + 1. 交换桥接子系统

交换桥接子系统提供适配器，用来适应不同类型、不同网段的信息源。

* 1. 流程 flow path
     1. 资源目录的形成与发布流程 Resource Directory Formation and Release Process

目录内容服务形成与提供流程包括5个环节：

1. 准备：医学影像大数据资源提供者将可共享的医学影像学大数据汇总建立成为资源共享数据中心；
2. 编目：编目管理系统综合共享资源内容的特征信息进行数据资源的编目；
3. 注册：：目录内容信息库储存已完成注册的编目管理系统形成的的目录内容；
4. 审核：由资源目录管理者通过编目管理系统对注册的资源进行审核，审核通过的传输至目录内容管理信息库；
5. 发布：由目录管理者对审核后的信息资源目录内容进行发布。
   * 1. 资源的发现定位与交换流程 Resource discovery, location and exchange process

资源的发现定位与交换流程包括2个环节：

1. 目录查询：使用者基于浏览器等客户端查询目录服务中心目录内容，通过目录查询可以发现并定位共享信息。
2. 信息获取：使用者根据目录查询得到的定位信息，通过资源交换中心获得具体信息。使用者可以通过网络浏览、查询、下载、订阅推送等各种方式从资源交换中心获取医学影像学大数据资源。
   1. 医学影像学信息元数据描述 metadata annotations
      1. 中文名称 Chinese name

指元数据元素或元数据实体的中文名称，如“信息资源名称”。

* + 1. 定义 definition

描述医学影像学元数据元素或元数据实体的基本内容，给出信息资源某个特性的概念和说明。

* + 1. 英文名称 English name

医学影像学元数据元素或元数据实体的英文名称，一般用英文全称。

* + 1. 数据类型 data type

医学影像学元数据元素的数据类型，对元数据元素的有效值域和允许对该值域内的值进行有效操作的规定。

* + 1. 值域range

说明医学影像学元数据元素可以取数据值的范围。

* + 1. 约束/条件 constraint

说明医学影像学元数据实体或元数据元素是否必须选取的属性。包括必选、可选。

* + 1. 最大出现次数 max occurrences

说明元数据实体或元数据元素可以具有的最大实例数目。

* + 1. 备注 remarks

对元数据元素或元数据实体的详细说明与注意事项说明。

* 1. 基本元数据 Basic metadata
     1. 概述 summary

医学影像学大数据与智能技术信息资源基本元数据是描述信息资源基本属性的元数据元素和元数据实体，可用于医学影像学大数据与智能技术信息资源编目、描述和数据交换活动。

医学影像学大数据与智能技术信息资源基本元数据应包括资源标识符、标准化名称、关键字与词组、信息内容摘要、信息资源使用限制、信息资源链接地址等元数据元素和责任者、信息资源分类等元数据实体。

* + 1. 基本元数据描述 Basic metadata description
       1. 资源标识符

定义：是医学影像学大数据与智能技术信息资源的唯一不变的标识编码

英文名称：Resource ID

* + - 1. 医学影像学大数据与智能技术信息资源名称

定义：描述医学影像学大数据与智能技术信息资源内容的标题

英文名称：Resource Title

* + - 1. 医学影像学大数据与智能技术信息资源分类（元数据实体）

定义：描述医学影像学大数据与智能技术信息资源的分类信息

英文名称：Resource Category

* + - 1. 医学影像学大数据与智能技术信息资源分类方案名称

定义：资源的分类方案的名称

英文名称：Resource Category Schema Name

* + - 1. 医学影像学大数据与智能技术信息资源分类方案编码

定义：资源的分类方案的编码

英文名称：Resource Category Schema ode

* + - 1. 医学影像学大数据与智能技术信息资源分类类目名称

定义：特定资源的分类方案的类目名称

英文名称：Resource Category Name

* + - 1. 医学影像学大数据与智能技术信息资源分类类目代码

定义：特定资源的分类方案的类目代码

英文名称：Resource Category Code

* + - 1. 医学影像学大数据与智能技术信息关键字

定义：用于描述资源主题的通用词、形式化词或短语

英文名称：Keyword

* + - 1. 医学影像学大数据与智能技术信息资源摘要

定义：资源内容的简单说明

英文名称：Abstract

* + - 1. 医学影像学大数据与智能技术信息资源负责方

定义：对资源的完整性、正确性、真实性等负有责任的单位的名称和地址信息

英文名称：Point Of Contact

* + - 1. 医学影像学大数据与智能技术信息资源负责单位名称

定义：提供医学影像学大数据与智能技术信息资源的机构和部门的名称。

英文名称：Organization Name

* + - 1. 医学影像学大数据与智能技术信息资源负责单位代码

提供医学影像学大数据与智能技术信息资源的机构和部门的代码。

英文名称：Organization Name

* + - 1. 医学影像学大数据与智能技术信息资源负责单位电子邮件地址

定义：资源负责单位联系人的电子邮件地址

英文名称：Electronic Mail Address

* + - 1. 医学影像学大数据与智能技术信息资源负责方联系电话

定义：资源负责单位联系人的联系电话

英文名称：Contact Number

* + - 1. 医学影像学大数据与智能技术信息资源负责方通讯地址

定义：资源负责单位联系人的通讯地址

英文名称：Contact Address

* + - 1. 医学影像学大数据与智能技术信息资源负责方邮政编码

定义：资源负责单位所在地的邮政编码

英文名称：Zip Code

* + - 1. 医学影像学大数据与智能技术信息资源格式

定义：资源存在格式的信息

英文名称：Data Format Information

* + - 1. 医学影像学大数据与智能技术信息资源安全限制分级

定义：对资源处理的限制级别的名称

英文名称：Security Classification

* + - 1. 医学影像学大数据与智能技术信息在线资源链接地址

定义：可以获取资源的网络地址

英文名称：Online resource link address

* + - 1. 医学影像学大数据与智能技术信息资源使用收费标识

定义：资源被使用时是否向用户收取费用的标识

英文名称：Charge Sign

* 1. 技术要求 technical requirement
     1. 概述 summary

医学影像大数据与智能技术资源目录的整体功能依靠各个计算机系统来实现。根据每个系统所实现的功能不同可细分为编目管理系统、目录内容服务系统、元数据管理系统和目录功能系统。

总体要求：数据采集与应用达到快速.便捷.无损.准确.安全.可靠.高效。

* + 1. 编目管理系统 Cataloging management system

编目管理系统根据各个医学影像学大数据与智能技术资源的内容，提取其基本特征，按照相关标注实现元数据赋值，形成目录内容。

其系统应遵循以下技术要求：

* + 1. 编目对象是具体的医学影像学大数据与智能技术资源，其内容包括各个行业、部门日常长期以来形成的海量数据。具体形式可以是数据库、图片、文档等各类型的数据。
    2. 编目管理系统应该支持自动、机辅方式完成元数据元素的赋值。
    3. 唯一标识符管理功能：支持唯一标识符的分配和赋值，包括支持后段码的自动生成和管理。
    4. 标准符合性检查功能：支持政务信息资源元数据和标准一致性检查，元数据完整性检查的主要目的是保证所有必选的元素据实体和元数据元素已经赋值，标准一致性检查的主要目标是保证已填写好的元数据实体和元素据元素的取值符合编目管理系统的相关规定。
    5. 信息资源分类：按照大数据信息资源分类标准，实现对共享大数据信息资源的分类。
    6. 目录内容管理系统 Directory Content Management System

医学影像学大数据与智能技术资源目录管理系统包含外部管理系统和内部管理系统。外部管理系统就是要建立外部门户网站，对外实现医学影像学大数据与智能技术资源注册、发布、查询、调阅、打印，推送及云服务等功能；内部管理系统就是要对内实现医学影像学大数据与智能技术资源编目、目录维护、主题统计、共享监测等功能。

* + 1. 目录内容服务系统 Directory Content Service System

医学影像学大数据与智能技术资源目录服务系统是医学影像学大数据与智能技术资源目录管理系统的子系统。可细分为三个服务系统：资源共享服务系统、公共信息服务系统、辅助决策服务系统。

* + - 1. 医学影像学资源共享服务系统

医学影像学资源共享服务系统是指通过资源注册和查询，实现单一信息源对其他机构、部门的信 息资源共享，从而解决信息的完整性和一致性问题。

医学影像学资源共享服务系统应具备的基本功能包括：医学影像学共享资源注册和查询。

* + - 1. 公共信息服务系统

公共信息服务系统是指通过资源查询与推送，实现对授权人提供完整个人医学影像与智能诊断医疗大数据信息或对社会公众提供健康，医疗，卫生，保健相关信息等信息，从而解决该系统信息的可及性和开放性问题。资源共享服务系统主要实现的功能是信息推送，即向授权人提供完整个人医学影像学大数据与智能技术信息或对社会公众提供公共卫生，医疗，保健，健康医疗信息。

* + - 1. 辅助决策服务系统

辅助决策服务系统是指通过资源查询与调阅，实现多渠道健康医疗信息的采集、汇总、分析与综合应用，为影像诊断提供多样、科学的决策信息。

辅助决策服务系统应具备的基本功能包括：医学及医学影像学信息资源的查询与调阅、汇总与分析以及综合应用。

* + 1. 元数据管理系统 Metadata Management System

元数据管理系统的目的是实现对元数据的管理，应提供元数据元素管理、元数据实体管理、元数据集管理、元数据版本管理等功能。

其具体功能包括：

* + 1. 查询大数据中心底层数据表的信息，文件以树形结构进行展示；
    2. 元数据信息分类处理；
    3. 提供元数据表结构变更、查询、新增、删除等操作；
    4. 提供元数据备注信息，描述数据的新增修改，包括本数据的来源信息；
    5. 以上功能适用于具有信息数据管理权限的技术人员应用。
    6. 目录功能系统 Directory function system

目录功能系统是实现医学影像学大数据与智能技术资源目录功能的主要系统。其应具备的基本功能包括：

* + 1. 医学影像学大数据与智能技术资源注册：注册的资源应该符合医学影像学大数据与智能技术资源目录的要求，并且给每个注册的资源赋予唯一标识符；
    2. 医学影像学大数据与智能技术资源发布：将注册的信息资源加入到资源目录体系中，并且在门户网站上发布；
    3. 医学影像学大数据与智能技术资源查询：根据查询请求对目录内容信息进行查询，并返回查询结果；
    4. 医学影像学大数据与智能技术资源调阅：实现医学影像学大数据与智能技术调阅功能，方便查询者浏览信息资源；
    5. 医学影像学大数据与智能技术资源目录维护；
    6. 医学影像学大数据与智能技术资源主体统计；
    7. 医学影像学大数据与智能技术资源共享监测。
  1. 关键技术方法 Key technology approaches
     1. 数据库技术 data base technique

构建医学影像学大数据与智能技术资源目录体系的关键在于通过构建元数据库来实现元数据管理。元数据是关于数据的数据，又称中介数据、中继数据，主要是描述数据属性的信息，

用来支持如指示存储位置、历史数据、资源查找、文件记录等功能。元数据库是按照数据结构来组织、存储和管理数据的数据仓库。使用元数据目的在于：识别资源，评价资源，追踪资源及各单元组间资源在使用过程中的变化；实现简单，快速，信息资源的有效发现、查找、一体化组织和对使用资源的有效管理。

元数据库的基本结构可以分为三层：物理数据层、概念数据层、用户数据层。

根据我国医学影像学大数据与智能技术信息资源管理、应用需求现状，应综合国内外大数据资源目录体系构建元素，参考其他行业大数据资源归类方法，从资源内容、资源表示、资源管理、资源责任和资源获取5个维度构建医学影像学大数据与智能技术信息资源目录元数据库。

* + 1. 资源目录分类模型 Resource Catalog Classification Model

构建医学影像学大数据与智能技术资源目录体系的第二个关键问题在于如何实现资源的分类。因此需要构建分类模型来实现医学影像学大数据与智能技术资源的分类，建立适用于医学影像学大数据与智能技术的资源目录分类模型。

医学影像学大数据与智能技术资源分类由类目、亚目和细目3个层次组成。

* + 1. 分类编码 sorting code number

编码是标识信息资源的关键方法，依据我国《卫生信息标识体系对象标识符编号规则》和《卫生信息标识体系对象标识符管理注册管理规程》，医学影像学大数据与智能技术资源根目录应依据相关标准进行分类编码，从而实现对医学影像学大数据与智能技术资源的标识。

* + 1. 元数据采集与存储技术 Metadata acquisition and storage technology

元数据采集技术包括元数据的自动采集技术和手工采集技术。元数据采集应当支持对不同元数据内容标准的元数据进行采集，同时能够对采集的元数据进行数据完整性和逻辑一致性的检查。数据完整性主要指的是元数据内容标准中所规定的必选必填内容是否都已有值，元数据的实体和元数据元素的相互关系符合元数据内容的规定。逻辑一致性是指元数据在数据结构、数据格式和属性编码正确性方面，尤其是拓扑关系上的一致性。

另外，也需要同时查看采集日志数据，以便检查采集是否成功，随时查看采集任务的相关信息。元数据采集完成后，存储在数据库中，提供元数据统计、查询等服务。

元数据的输出必须采用成熟、主流的数据编码技术进行编码，方便元数据的输出和交换。医学影像学大数据与智能技术资源元数据的存储需要尽量使用已有的关系型数据库进行存储。

元数据是层次型数据，在存储到关系型数据库时，需要进行层次型到关系型的模型转存储的模式不会随元数据标准的变换而变化，需要在关系型数据库中建立元数据的数据字典描述元数据结构；要建立高效率的索引机制保证对元数据内容的有效检索和安全的传输，存储，打印及云服务环境。

* + 1. 目录服务与应用技术 Directory services and application technologies

目录服务技术是指通过网络查询医学影像学大数据与智能技术资源元数据以得到相关信息的技术。目录应用技术是向用户展现目录的技术。目录应用技术的核心应是元数据的展现技术。

* 1. 医学影像学大数据与智能技术资源目录服务接口

Medical Image Big Data and Intelligent Technology Resource Directory Service Interface

医学影像学大数据与智能技术目录服务包括发现和管理两种基本功能，发现功能用于对元数据进行检索，管理功能实现元数据的管理。

医学影像学大数据与智能技术资源目录体系包含五种接口，即基础接口、发现接口、管理接口、交换接口和应用接口。其中基础接口是将发现接口和管理接口中基础性的操作定义成一个公共接口，基础接口和发现接口是必选实现，管理接口、交换接口、应用接口为可选实现，有利于接口的扩展性。这五类接口实现大数据资源的发现功能和管理功能。

基础接口：提供会话管理功能和服务自描述功能，包含有目录服务初始化接口、目录服务终止接口和服务自描述接口；

发现接口：提供元数据检索功能和元数据检索结果提取功能，包含有目录检索接口以及目录检索结果提取接口，这些接口本身不提供资源，而是提供资源基本信息和如何去获得这些资源的元数据；

管理接口：提供元数据管理的功能，包含元数据管理接口；

交换接口：实现数据资源的交换和传输；

应用接口：提高目录内容管理系统的可扩展性。

医学影像学大数据与智能技术资源目录服务支持核心元数据及其扩展内容的查询，该元数据需要符合医学影像学大数据与智能技术资源目录体系对于核心元数据的要求，目录服务支持对多个元数据库的查询，元数据一般按照描述大数据资源的内容分别建立。

医学影像学大数据与智能技术资源目录服务可以是集中式的也可以是分布式的。

集中式是指元数据库和目录服务在物理上和逻辑上都部署在一个节点，分布式指的是元数据库物理分布分散、目录服务逻辑集中。

1. 人工智能技术对医学影像学大数据挖掘和应用

图像数据挖掘的主要目标是从中提取出图片的自身特征， 包括语义、质量、关联度、实体义项等。可以支持数据自动与其任务、属性相结合，对图像背后隐藏的需求进行挖掘，并通过与周边环境信息进行关联计算。

符合人工智能技术应用于医学影像学数据挖掘和辅助诊断方法。

* 1. 快速医学影像成像方法：
     1. 生成对抗网络（generative adversarial network,GAN）

生成对抗网络是一种由生成器（generator）和判别器（discriminator）组成的生成式深度学习体系结构，生成器的作用是将低质量的MR图像映射到高质量图像的流形上，判别器的作用是对映射后的图像质量进行评判。Mardani M等人提出了一种基于生成对抗网络的磁共振（magnetic resonance,MR）压缩感知（compressed sensing,CS）快速成像方法，该方法利用GAN对高质量MR图像的低维流形（manifold）进行建模，在保留感知图像细节的情况下具有很好的重组效果，并且每组图像的重组时间较短。

* + 1. 级联深度神经网络（cascaded DNN）

基于级联深度神经网络（cascaded DNN）的MR快速成像方法。

级联深度神经网络由若干个网络单位级联而成，每个网络单元包含卷积神经网络（convolutional neural network,CNN）和数据保真项两个部分，其中CNN以残差网络（residual network, ResNet）的形式构建。

CNN学习到的是降采（under-sampled）图像与满采（fully-sampled）图像之间的差异。级联深度神经网络的重建图像质量相比于传统的压缩感知方法以及基于字典学习的图像重建方法有了明显的提升，同时其重建一幅二维心脏图像的时间仅为23 ms，基本达到了准实时的效果。

* + 1. 交替方向乘子算法（alternating direction method of multipliers,ADMM）

为了将传统迭代重建方法与深度学习方法各自的优势结合起来，基于交替方向乘子算法（alternating direction method of multipliers,ADMM）的MR图像重建方法——ADMM-Net，该方法将经典的ADMM迭代重建方法利用神经网络进行重新实现。ADMM-Net对特定迭代次数的ADMM方法进行建模，在每次迭代中，利用CNN解决ADMM算法中的3个子优化问题，整个网络以端到端（end-to-end）的方式进行训练。ADMM-Net的优势在于各种参数可以通过完全数据驱动的、端到端的学习方式得到。ADMM-Net的重建结果明显优于传统方法。ADMM-Net的构建参照了经典的ADMM，其网络的重建结果具有更好的可解释性。

* + 1. 卷积框架（convolution framelets）

为了进一步证明将深度学习应用于成像方向的理论上的合理性，利用卷积框架（convolution framelets）方法从理论上加以解释。卷积框架最初用来拓展低秩Hankel矩阵（low-rank Hankel matrix）理论在逆问题中的应用。深度卷积框架神经网络（deep convolutional framelet neural network, DCFNN），在采用修正线性单元（rectified linear unit,ReLU）非线性激活函数的情况下，该网络可以实现完美重建，促进完美重建的实现。

通常，一次MR扫描可以生成多种不同对比度的图像。

现有的基于深度学习的单一对比度快速成像方法没有充分利用不同对比度图像之间的结构相似性，因此限制了其可以达到的加速比。

进一步提升重建图像的质量，实现图像重建，通过将6-echo的图像作为不同的通道输入网络中，使得在卷积过程中能够充分利用不同echo图像间的结构相似性，从而为网络的训练加入更多的约束条件，让训练过程变得更加稳定。

该方法可以实现4.2倍的MR成像加速，重建图像在均方根误差（root mean square error, RMSE）和结构相似性指数（structural similarity index,SSIM）等方面均优于单一对比度重建方法。

由于CNN卷积操作的空间局部特性，目前绝大部分基于深度学习的快速成像方法选择在图像域进行处理。

然而，一些因k空间数据不完备性造成的图像伪影却很难在图像域完美地解决。

提出了一种基于双域深度学习的MR快速成像方法，在图像域和频率域均设计了对应的深度CNN，试图从两个不同空间分别对未采集的数据进行恢复，同时图像域与频率域通过数据保真项被关联起来，从而保证重建得到的图像的真实可靠性。图像域CNN和频率域CNN在图像重建过程中的作用是不同的，相比于仅采用图像域CNN的成像方法，将二者结合起来可以获得质量更高的图像重建结果。

* 1. 医学图像质量增强的方法
     1. CT图像质量增强

在低剂量图像去噪方面，Chen H等人提出了一种基于残差自编码器（residual auto encoder）的CT图像去噪方法，该方法利用深度神经网络构建一个自编码器（auto encoder,AE），不同之处在于网络的编码器和解码器部分采用残差的方式连接。这样做的优点是可以将不同层次的图像特征结合起来，提升网络的建模能力；还可以使训练过程中误差的反向传播更加有效，提升网络的训练效果。此外，网络采用残差的方式连接，使网络实际上学习到的是噪声图像到噪声的映射，这比直接学习从噪声图像到高质量图像的映射更加容易。仿真实验结果和临床图像实验结果，相比于传统的图像去噪方法（如BM3D）,Chen H等人提出的残差自编码器方法在峰值信噪比(peak signal to noise ratio,PSNR）、SSIM（Structural Similarity）等指标上均有明显的优势，同时去噪速度也更快。

基于深度学习的图像去噪方法容易产生一定的过平滑（oversmooth）现象，这是因为网络的损失函数通常采用整幅图像的L1/L2范数，并没有对细节区域进行特别处理，而细节区域的误差在整体误差中的比例很小，所以容易在网络的训练过程中被“淹没”掉，最终导致图像细节丢失。为了解决这一问题， Wolteri nk J M等人设计了一种基于GAN的CT图像去噪方法，GAN用于学习从低剂量图像到正常剂量图像的映射，判别网络用于判别生成的去噪后的图像是否处于正常剂量图像所在的流形中，即是否和真实的正常剂量图像在视觉上相似。本质上，判别网络可以看作一个计算机自己学到的损失函数，相比于人工设计的损失函数，其能够学到更加高层次和细节化的图像特征，因此可以得到更加准确的训练结果。Wolterink J M等人提出的基于GAN的图像去噪方法能够有效地去除低剂量CT图像中的噪声，同时能够很好地保护图像的细节信息，使去噪后的图像在视觉上更加自然可信。

在某些情况下，由于物理、机械等条件的限制，只能获取到一定角度范围内的CT投影数据。传统的解析重建方法和迭代重建方法在处理这类数据不全的问题时，重建得到的图像通常包含严重的伪影和模糊。为了解决这一问题，A nirudh R等人提出了一种基于深度学习的有限角度CT图像去伪影算法（CT-Net）。其基本思想是在CT-Net的训练过程中直接学习从不完整弦图（sinogram）到CT图像的映射，损失函数结合了图像域L2范数和GAN，保证了增强后的图像具有较高的信噪比和丰富的细节信息。在应用过程中，首先利用CTNet得到增强后的CT图像，然后利用该图像补全缺失的弦图，最后采用解析或迭代重建方法利用补全后的弦图重建出最终的图像。仅采集了90°的弦图数据，利用CT-Net依然可以重建出质量较高的图像，而直接用传统的解析或迭代重建方法无法获得清晰的重建结果。

* + 1. PET图像质量增强

由于PET成像需要事先向患者体内注射放射性示踪剂（如18F-FDG），为了降低患者接受的辐射剂量，临床上对低剂量PET成像有很高的需求，然而剂量的降低会造成图像噪声的增加和对比度的下降，影响疾病的临床诊断。针对 这一问题，Xu J等人提出了一种基于残差编码解码器（residual encoder-decoder）的PET图像增强方法和传统的非局部均值（non-local means,NLM）、块匹配三维滤波（block-matching and 3D filtering, BM3D）等方法。

* + 1. MR图像质量增强

为了实现成像加速，通常会在k空间进行数据截断和填零，这会导致重建图像中存在Gibbs伪影。传统的MR图像去伪影方法通常基于k空间滤波，然而k空间滤波无法很好地区分伪影信号和有用信号，使得增强后的图像往往存在着过平滑、细节丢失等问题。增强方法，该方法基于U-Net和ResNet网络结构，可基于MT（矩阵模板库）L的MR图像增强方法能够在保护图像分辨率的情况下有效抑制Gibbs伪影。

* 1. 智能图像分析技术 intelligent image analysis technology
     1. 影像组学 Radiomics

影像组学的主要流程包括图像获取和标注、感兴趣区图像分割、影像组学特征的提取、特征值选择和降维、预测模型的训练和性能评估。

* + 图像获取和标注。影像组学大数据要求病人数据临床问题明确、格式规范、信息完整，而目前常规临床使用的影像，由于采集时的成像参数、病人体位、重建算法以及扫描仪器的不同，具有很大的差异性，因此识别和整理大量具有相似临床参数的图像数据示例（如疾病分期）也是一项挑战。同样，对于前瞻性研究而言，制定统一的数据规范也是实现影像组学研究的关键。
  + 感兴趣区图像分割。将感兴趣区域（如肿瘤等）在影像上分割出来是实现后续特征提取和信息分析的基础。通常以专家的手动分割结果作为标准，但是由于评价者自身和评价者之间的差异性以及分割工作本身的繁杂性，使手动分割不可能满足影像组学的要求。而全自动或半自动的分割方法会降低由于人的参与造成的差异性，使分割结果的可重复性更高。传统应用于医学影像的自动和半自动分割方法（如阈值分割法、边缘检测法、区域生长法、水平集法、模糊集法、活动轮廓模型法、图谱引导法）在实际中分割的精度和速度很难满足要求，近年来使用深度卷积神经网络的分割方法成为研究的热点。
  + 影像组学特征的提取。一旦确定了感兴趣区域，就可以提取高通量的影像组学特征。影像组学特征可以分为两大类：一类是使用数学公式定量计算的感兴趣区特征，包括形状特征、灰度特征和纹理特征以及小波变换、高斯变换后的特征；另一类是使用神经网络提取的图像的深度特征，但对于深度特征的具体物理意义没有直接的对照。
  + 特征值选择和降维。初步提取的图像特征一般是数以千计甚至万计的数据。为了选出可重复性好、信息量大、无冗余的特征用于最终模型的建立，一般需要对高通量的特征进行降维处理。

常见的特征降维方法如下：

* + - 方差分析。将数据列变化非常小（即列包含的信息量少）的特征直接滤除。
    - 相关性度量。数值列之间的相似性通过计算相关系数来表示，名词类列的相关系数通过计算皮尔逊卡方值来表示，相关系数大于某个阈值的两列只保留一列。
    - 组合决策树方法。对目标属性产生许多巨大的树，然后根据对每个属性的统计结果找到信息量最大的特征子集。
    - 主成分分析法。将原始的n维数据集通过正交变换转换成不相关的被称作主成分的数据集，变换后方差最大的特征即第一个主成分，其后的成分在与前述主成分正交条件限制下具有最大方差，保存前m（m＜n）个主成分就能保存最大的信息量。
    - 特征一致性度量。由手动分割结果计算得来的特征值，需要对其进行再测信度（test-retest）检验，计算一致性相关系数，将一致性相关系数小于某个值的顽健性低的特征滤除等。
  + 预测模型的训练和性能评估。依据分析的类型（即有无预测标签）可以将分析方法分为监督学习和无监督学习。聚类是最常用的无监督学习方法，分层聚类是最常用的聚类算法。监督学习常用的方法有感知机、K最近邻法、决策树、线性回归、支持向量机、随机森林、神经网络和深度学习等。影像组学分析的最终目的是使建立的模型不仅对现有的数据有很好的预测能力，而且对未知的数据也有很好的预测能力，这就要求选择适当的模型使测试误差最小。常用的模型选择方法有正则化和交叉验证。依据分析目的（即标签类型），也可以将监督学习分为分类预测和回归预测。相应的模型性能评价指标一般采用的是分类准确率和损失函数。
    1. 深度学习 deep learning

深度学习是由复杂结构或者非线性转换组成的多层神经元对数据高级抽象后进行建模分析的一种方法。它包含有多种重要的网络模型，有着不同的特点。医学图像分析也是深度学习的一个活跃的研究领域。

深度学习一般包括监督学习模式和无监督学习模式。

监督学习模式中的卷积神经网络（CNN模型）是目前医学图像分析中研究最多的机器学习算法。它是一种深层前馈型神经网络，主要由输入层、卷积层、池化层、全连接层和SoftMax层组成。卷积神经网络的优势在于它能够自动学习重要的低级特征（如线条和边缘），并且能够从低级特征中迭代地提取更复杂和更高级的特征（如形状等），其端到端的设计给模型提供了更多可以根据数据自动调节的空间，增加了模型的整体契合度。其主要原因是CNN在过滤输入图像时保留了空间关系。空间关系在放射学中至关重要，例如，骨骼边缘与肌肉、正常组织与癌组织的连接等。

为了解决卷积神经网络平移不变性和池化层信息丢失的本身缺陷，胶囊网络应运而生。它是由卷积层和可处理更多类型的胶囊组成，其最大的优点在于网络中保存了目标位置和形状的详细信息，在图像分割和目标检测方面有着非常好的发展方向。但是自身也存在着耗时长、不能同时检测同一类型两个对象的缺点。

循环神经网络（recurrent neural network,RNN）可以利用内部记忆处理任意时序的输入序列，常被用于时间性分布数据。长短期记忆（long short-term memory，LSTM）网络是RNN的一种特殊类型，它的出现解决了RNN在实际训练过程中出现的梯度消失问题。在医学图像分析领域， RNN主要用于分割任务。

无监督学习模式中的自动编码器以无监督的方式学习编码，无需训练标签，减少了模型的维度和复杂性，同时通过重建输出，可以生成与输入数据类似的新数据。在医学影像中，自编码器有两个较为常用的改进算法：降噪自编码器和卷积自编码器，不仅减少了训练时间，还发挥特征降维和特征提取的优势，在医学影像中用以解决医学图像分析中标记数据稀缺的问题。

受限玻尔兹曼机（restricted boltzmann machine,RBM）是由可见层和隐藏层组成的浅层神经网络，层间彼此连接，但层内无连接。RBM使用输入数据的反向传递生成重建，并估计原始输入的概率分布。RBM被用于降维、分类、回归、协同过滤、特征学习等，是组成深度置信网络（deep belief network,DBN）的基础部件。DBN通过训练受限玻尔兹曼机间的权重，让整个神经网络按照最大概率来生成训练数据，但是因其采用逐层训练的方式会使学习过程变慢。

GAN通过生成模型和判别模型的互相博弈，学习并产生更好的输出，它具有不需要先验假设、生成器可直接生成样本数据以及对整个图像进行评价、分析、生成三个特点。因此，它最大的优点是运行时间相对较短和生成采样数据方式相对简单。GAN及其扩展已被用来解决很多具有挑战性的医学图像分析问题，如医学图像去噪、重建、分割、检测或分类。此外，GAN合成图像的能力也被用于解决医学领域中标记数据的长期稀缺性的问题。

* + 1. 深度学习在医学影像领域的应用主要包括分类、检测、分割和配准。
  + 分类：分类主要涉及病变良恶性分类和多种疾病的鉴别。
  + 检测：DeepLesion是迄今全球规模最大的多类别、病灶级别标注的开放获取临床医疗图像数据集，含有32 735个带标记的病灶实例，包括来自全身各个部位的关键影像学发现，比如肺结节、肝肿瘤、淋巴结肿大等。
  + 分割：CT和MRI的图像分割研究涵盖了肝脏、前列腺和膝关节软骨等多种器官，但大量工作主要集中在脑部图像分割，如肿瘤分割。肿瘤分割在外科手术计划中尤其重要，可确定肿瘤的确切边界，指导手术切除。
  + 配准：图像配准用于神经外科手术或脊柱外科手术，以定位肿瘤或脊柱骨界标，便于手术切除肿瘤或植入脊柱螺钉。
  + 其他：生成对抗网络可降低低剂量CT的噪声，保留图像关键信息，减少过度平滑，同时还可以利用给定的MR图像生成对应的CT图像，使患者避免CT检查，从而减少辐射等。
  + 应建立科学的质量控制与质量保证体系，做好数据安全与数据管理，科学的开展医学影像学大数据框架下的智能诊断研究。成立专门的质量控制小组，制定工作大纲和技术路线。