

多维医疗影像数据管理系统^①



刘 丽, 蒋龙泉, 冯 瑞

(复旦大学 计算机科学技术学院, 上海 200433)

通讯作者: 冯 瑞, E-mail: fengrui@fudan.edu.cn

摘 要: 由于医疗领域数据的专业性、复杂性、多样性, 人工智能技术在医疗领域发展过程中存在样本数据缺乏的问题. 本文针对算法研究人员设计并开发了多维医疗影像数据管理系统, 系统包括数据预处理模块、数据管理模块以及数据可视化模块, 能够半自动化的生成具有分类标签的医疗影像数据, 对影像数据进行高效管理, 并且对系统中数据的分布特点以图、表的形式进行可视化展示.

关键词: 医疗影像数据; 计算机断层扫描; 数据可视化; 模态; 数据清洗

引用格式: 刘丽, 蒋龙泉, 冯瑞. 多维医疗影像数据管理系统. 计算机系统应用, 2021, 30(7): 130-135. <http://www.c-s-a.org.cn/1003-3254/7970.html>

Multidimensional Medical Image Data Management System

LIU Li, JIANG Long-Quan, FENG Rui

(School of Computer Science, Fudan University, Shanghai 200433, China)

Abstract: Since medical data are professional, complex, and diverse, sample data are lacking during the development of artificial intelligence in the medical field. In this study, we design and develop a multidimensional medical image data management system for algorithm researchers. The system includes data modules of preprocessing, management, and visualization. It can semi-automatically generate medical image data with classification labels, efficiently manage them, and display their distribution characteristics by graphs or tables.

Key words: medical image data; CT; data visualization; modality; data cleaning

人工智能技术的飞速发展, 给人类带来巨大的便利性, 人工智能技术逐渐进军医疗领域. 人工智能技术在医疗领域的应用有望带来一场医疗服务模式的革新, 从而极大的解放医生的劳动力^[1]. 以数据驱动的深度学习方法的人工智能技术常用的方法, 有效的样本数据对人工智能技术在医疗领域的发展非常重要. 因此, 开发一款能够高效管理具有标注信息的医疗影像数据的系统也非常重要.

目前, 在各大医院用于采集、存储并高效管理医疗影像数据的系统为 PACS 系统, 即图像存档与传输系统. 作为影像数据管理系统, PACS 系统具有数据采集、数据存储、数据查看、局域网内数据共享等基础

功能^[2-5]. 在欧美等发达国家对于 PACS 系统的研究比较早, 早在 20 世纪 80 年代就开始研究并且在 90 年代初期初见成效, 一些具有简单功能的 PACS 系统陆续落地. 而国内对 PACS 系统的使用起步则较晚, 20 世纪 90 年代中后期刚开始, 少数医院引入 PACS 系统, 实现简单的数字化存档^[6]. PACS 系统的使用对象是医生, 其功能也主要围绕医生设计, 目前, 医生使用该系统能够进行阅片、诊断、生成诊断报告等. 毛静设计了一款基于医学影像 PACS 的数据管理系统, 在保留其原有的采集模块、数据存储模块的基础上还增加了数据分析和处理模块, 能够对每天数据量、每月数据量、各医院数据量以及放射设备检查数据量等进行分析^[7],

① 基金项目: 上海市科委项目 (20511100800, 20511101103, 20511101203)

Foundation item: Shanghai Municipal Science and Technology Commission Project (20511100800, 20511101103, 20511101203)

收稿时间: 2020-10-21; 修改时间: 2020-11-18; 采用时间: 2020-12-01; csa 在线出版时间: 2021-06-30

能够实现对系统中数据的分布特点进行简单的分析。PACS系统的发展极大地方便了医生诊断,降低成本,但是存储的数据无法直接用于人工智能算法,导致人工智能技术与医疗数据结合的时候面临着许多问题^[8]。

为了解决人工智能技术与医疗数据结合时,有效样本数据缺乏的问题,每年各大机构都会推出一些具有标注信息的医疗影像数据集,著名的平台包括CAMELYON和MICCAI,并且举办一些医学竞赛,推动AI在临床的发展。医疗影像数据的原始数据格式包括DICOM文件、PNG文件、JPG文件或者NII文件等。其中DICOM是指医学数字成像和通信,是医学图像和相关信息的国际标准,医院用于影像检查的设备都采用DICOM标准的规范化接口^[9]。竞赛的数据虽然具有相关的标注信息,但是基于竞赛的数据存在很大的局限性,这些数据的标注只是针对每种疾病,固定的临床任务,这将导致在使用数据的时候有可能需要进行大量的预处理操作,不仅浪费时间而且如果处理不当则可能会给算法带来负面影响,降低算法的适用性。

本文旨在实现对医疗影像数据进行半自动化分类,并且对标注后的数据进行高效管理,设计并实现多维医疗影像数据管理系统。该系统的服务人群是算法研究人员,一方面可以使算法研究人员更集中于研究算法本身,另一方面有助于加速医疗辅助诊断系统落地的步伐。

1 多维医疗影像数据特点

本文重点研究面向多病种的多模态医疗影像数据,数据的复杂性在于疾病种类的多样性,多模态主要指的是数据的采集来源不同。

1.1 数据的复杂性

医疗数据最大的特点就是病种的多样性,每个病种的数据都有其自身的特点。多维医疗影像数据管理系统针对常见的4大类型疾病进行数据分析、整理并管理,包括胰腺癌数据、乳腺癌数据、肺结节数据以及肝癌数据。

1.2 数据的多模态

医疗领域的数据呈现多态性,即数据的来源丰富,通过不同的设备、不同诊断目的产生不同的数据,比如CT数据、MRI数据、病理数据等,这种在表现形式上呈不同形态的数据称医疗数据的不同模态。这些

数据都有其特自身的特点、数据格式等,通过多种模态的数据的联合诊断,能够有效地增加数据呈现的信息,从而增加医生诊断的正确率。本系统管理的数据模态主要包括CT数据、MRI数据和病理数据。

1.3 数据的多序列

对于同一模态的数据,由于观察到的组织结构的不同或者视野范围的不同则会进行不同的扫描。

CT扫描方法包括平扫、增强扫描和造影扫描。平扫即为常规扫描,一般指的是横切。增强扫描则指的是通过为患者注射一些试剂,然后在进行扫描,可以更加清楚地显示某些病变组织。造影扫描指的是医生为患者先通过注射试剂的方法为器官或组织结构形成造影,然后再进行CT扫描。

MRI扫描也会进行不同序列的扫描,比如进行DWI扫描、T1加权扫描、T2加权扫描以及T1增强扫描等,这些不同的扫描序列图像能够使不同的组织呈现更不同的效果。例如DWI对于病变非常敏感,如果有病变则DWI中病变区域会呈现亮色,而T2加权扫描则可以更好地呈现扫描部位的组织结构,可以非常清楚地看到组织边界的勾勒情况。

病理图像^[10],为了准确对肿瘤定性,则会进行不同放大倍数的扫描。肿瘤的恶性程度与组织内癌细胞的数量相关,因此为了准确的判断肿瘤的恶性程度,需要为病理切片中的癌细胞进行计数,对于病理图像,会有不同扫描倍数的切片,因此一张病理切片的大小通常都是以G为单位。

1.4 数据标注方法

本文调研了各大平台上公开的竞赛数据集,并且对分类任务的标注方法进行总结,分类数据集的标注相对来说是比较统一的,主要通过CSV的方式对原始数据进行病人级别的标注,包含病人信息与标注信息两种数据,因此本文规定,对于分类任务的数据标注详细信息如表1所示。

表1 分类数据标注属性

数据	参数	参数值
病人信息	序号	病例ID号
	性别	患者性别
	年龄	患者年龄
分类信息	标注方式	CSV文件中详细标注
	标注意义	规定0,1代表的分类级别
	图像存储格式	对图像数据存储格式的简要描述
	图像格式	PNG, DICOM

2 系统功能介绍

本节详细介绍多维医疗影像数据管理系统的总体架构以及核心模块. 该系统采用的体系结构为基于B/S的体系结构, 不管在何处都可以通过网址直接访问到本系统.

2.1 系统总体架构设计

如图1所示为系统整体架构图, 系统从上到下依次为表现层、数据处理层、基础设施层. 表现层主要方便人机交互, 对系统中页面进行可视化展示. 数据处理层主要包含系统中各个模块对于医疗影像数据的处理, 包括数据预处理模块中用到的数据筛选、去重、脱敏、模糊匹配等, 以及数据可视化模块中对数据的统计分析等. 基础设施包含数据库、磁盘阵列等硬件设施.

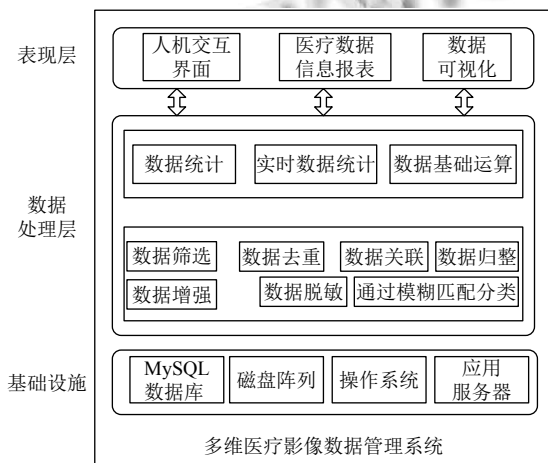


图1 多维医疗影像数据管理系统架构图

2.2 系统功能模块划分

多维医疗影像数据管理系统包含数据预处理模块、数据管理模块以及数据可视化模块3部分. 其中数据预处理模块能够实现对原始的医疗影像数据进行半自动化标注, 数据管理模块实现对不同种类的疾病进行数据管理, 数据可视化模块能够以图表的形式从不同维度对系统中拥有的数据进行数据可视化. 多维医疗影像数据管理系统的功能模块图如图2所示.

数据预处理模块包括数据清洗和数据标注模块. 数据清洗是对原始医疗影像数据进行数据清洗, 包括去除重复数据、数据脱敏操作、数据重命名等一系列操作, 最终将其处理为统一格式的数据. 数据标注模块则主要是根据医院的影像图像数据、影像报告、

病理报告之间的联系, 通过关联关系进行匹配, 最终实现从病理报告中提取出对应影像图像数据的分类标签.

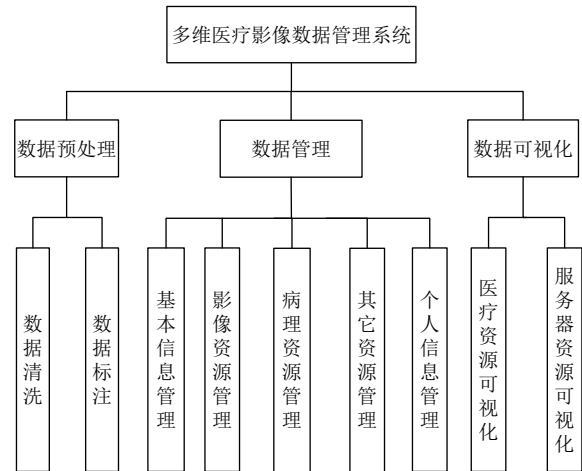


图2 多维医疗影像数据管理系统功能模块图

数据管理模块主要包含不同类型疾病的医疗影像数据的管理, 包括胰腺癌、肺结节、肝癌和乳腺癌四种类型疾病. 每种疾病包括基本信息管理、影像资源管理、病理资源管理、其它资源管理、个人信息管理. 其中每种疾病都提供一个数据获取的功能, 是为算法研究人员设计的功能, 通过该功能能够下载自己需要的医疗影像数据. 由于医疗数据模态的多样性, 本系统的设计的其它数据管理能够实现对不同数据间的差异性的兼容, 根据不同类型的数据加载与其对应的模块, 进行数据展示.

数据可视化模块主要是对系统中资源以图表的形式进行可视化分析, 包括医疗资源可视化和服务器资源可视化. 医疗资源的可视化, 是指从不同维度分析数据的分布特点, 从而有可能挖掘到更多有利于疾病诊断的关键信息. 而服务器资源可视化是对CPU、网络、内存等服务器资源进行监控, 一旦出现问题则立刻发出警告, 避免不必要的损失.

3 系统实现

3.1 关键技术

随着浏览器技术不断成熟和功能不断强大, B/S (Browser/Server) 模式具有强大的可扩展性与健壮的体系结构^[1], 本系统在开发时, 考虑到平台的兼容性和可扩展性, 采用了主流的B/S模式, 并且为了平台的可维

护性以及开发的高效性,采用前后端分离的开发模式^[12].前端开发框架为 Vue.js,其优点是数据的双向绑定,使得开发人员聚焦于业务层,后端采用 Django 框架,Python 对于医疗图像的处理有很多成熟的插件,更利于快速开发医疗数据管理系统,数据库选择了开源免费的数据库 MySQL.

3.2 数据预处理模块的实现

数据预处理模块是多维医疗影像数据管理系统的关键模块,数据进入系统的第一步都会通过数据预处理模块,该模块决定了系统数据的标注类别,数据预处理模块的涉及的关键接口如表 2 所示.

表 2 数据预处理模块关键 API 接口

接口名称	功能
mergeAllReport(path)	合并所有的影像报告,并且删除重复报告
imginfoData(path)	依次遍历影像数据,提取影像必要数据
findHospitalIDformImgReport(path1, path2)	在现有的图像数据的表格中找到对应的影像报告数据,填充表格
extraClass(path)	通过模糊匹配提取影像报告的诊断结果
fillNull(path)	填充操作
hideSec(path)	脱敏操作
banlanceData(path, type)	数据平衡操作

如图 3 所示为数据清洗操作的流程图,数据清洗是对系统中数据进行直接的操作,因此管理员才具有权限使用,管理员根据提示信息选择合适的数据,并输入对应的字段,包括影像报告、病理报告、影像数据,以及最终分类的类别,然后点击开始关联匹配,系统首先会对数据进行清洗工作,删除冗余数据以及空数据,然后从数据中提取该病例的 ID 号、性别、年龄、扫描模态 (CT、MRI)、序列 (DWI、T1、T2 等)、层厚等基础信息,分别从影像报告、病理报告中通过模糊匹配技术,获取到数据分类结果等关键字段,然后将获得的文本信息数据进行持久化操作,并且将清洗之后图像数据的路径也存入数据库中,方便后续查看影像数据.

3.3 数据管理模块的实现

如图 4 所示为平台在数据管理模块中的服务请求网络图,浏览器向服务器发起请求,请求中包含要访问疾病类型的信息,然后路由根据该信息跳转到对应的页面并且向对应的数据库服务器请求数据,将获取到的数据反馈回前端页面进而展示.

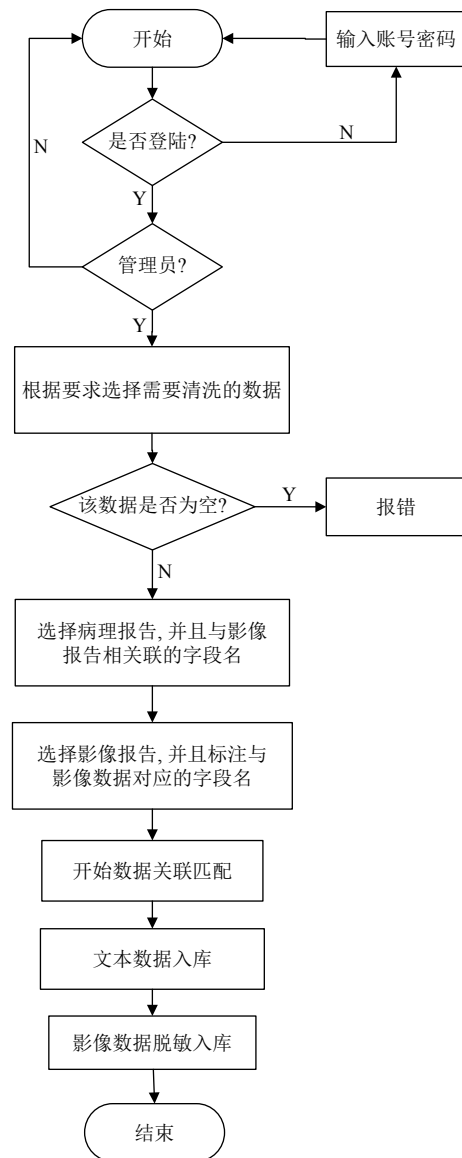


图 3 数据清洗流程图

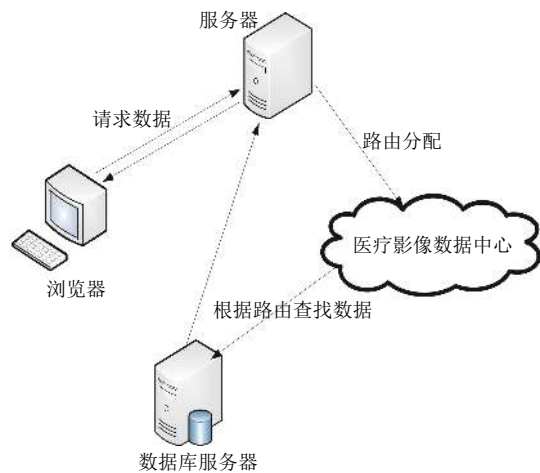


图 4 服务请求网络图

影像数据管理模块主要功能是对当前疾病所拥有的数据进行增删查改,影像数据管理比较特殊的一点是能够对 DIOCM 文件这种特殊医学文件的展示,本系统采用 dwv 框架,实现对 DICOM 图像的放大缩小浏览等功能.影像数据管理模块中数据的建模非常重要,需要构建合理的数据结构,以下是数据结构构建的关键代码.

```
class imgDataInfo(models.Model): # 影像数据信息
    # 病人基本信息
    # mainkey = models.CharField(max_length=15)
    casename=models.CharField(max_length=20)
    patientID=models.CharField(max_length=20)
    hospitalID = models.CharField(max_length=20)
    age=models.IntegerField()
    sex=models.CharField( max_length=2)
    #影像信息
    modality=models.CharField(max_length=10)
    checkMethod=models.CharField(max_length=20)
    checkDate = models.CharField(max_length=20)
    checkDevice=models.CharField(max_length=10)
    imgReport=models.CharField(max_length=500)
    reportResult = models.CharField(max_length=20)
    uploadTime=models.DateTimeField(auto_now_
add=True)
    imgPath=models.CharField(max_length=100)
    class Meta:
    db_table = "imgDataInfo" #结构化报告
    ordering = ['id']
```

3.4 数据可视化模块的实现

数据可视化模块的实现主要通过 echarts 插件完成,通过网络请求向服务端发送请求,服务端从数据库中获取到数据,经过计算后返回对应的数据,然后 echarts 根据获取到的数据绘制.医疗资源可视化包括按患者统计分析、按病种统计分析、按模态统计分析三大维度进行统计,其中每一大类中分别包括按照性别、年龄、疾病占比等不同的维度对数据进行可视化分析^[13].以下是可视化页面初始化数据的关键代码.

```
init(){
    var myChart=this.$echarts.init(document.getElementById(this.main))
    var option = {
```

```
title: { text: this.chartData.title },
tooltip: {},
legend: { data:this.chartData.lenged},
xAxis: { data:this.chartData.xdata},
yAxis: {},
series: [{
type: this.chartData.type,
data: this.chartData.ydata
}]];
// 使用刚指定的配置项和数据显示图表.
myChart.setOption(option);
}
```

4 系统的测试与应用

在实现了系统的各个功能模块后,对正常运行的系统进行关键功能的测试.如图 5 所示是对于数据预处理模块的测试,根据测试结果可以得出系统能够对指定的数据按照一定的规则进行分类;图 6,图 7,图 8 属于影像数据管理中影像数据管理模块,从打开页面到数据请求完成并渲染在视觉可接受的范围之内,测试正常;图 9 是对影像数据管理模块中的数据进行在线查看,并且其放大、缩小、滚动查阅、灰度变换等功能均满足要求;图 10 是数据统计分析模块对系统资源的统计分析的展示,测试正常.

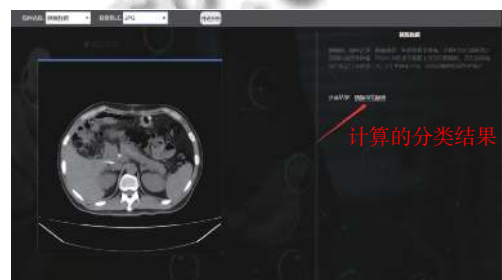


图 5 数据分类结果图



图 6 影像数据管理页面



图7 数据详细信息图

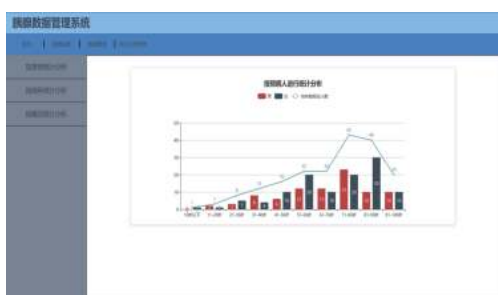


图8 按患者进行统计分析



图9 显示影像数据

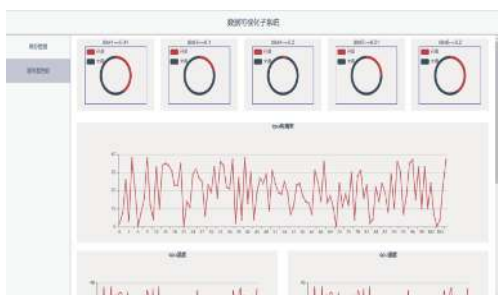


图10 数据可视化分析图

5 结语

本文在分析了当前人工智能在医疗领域存在的问题的基础上,对系统进行规划和整体设计^[14]。该系统实现了数据标注功能、数据管理和数据可视化三大功能。通过该系统能够实现对数据的高效管理,同时算法人员能够通过该系统下载具有标注信息的数据,从而为

算法研究人员提供数据支持。当然本系统还存在一些不足,比如有些功能的实现太简单,页面的设计还比较单一等。本系统在下一步工作中将会加入人工智能算法^[15],从而实现对医疗影像数据的分割标注。

参考文献

- 1 Kim M, Yun J, Cho Y, *et al.* Deep learning in medical imaging. *Neurospine*, 2019, 16(4): 657–668. [doi: 10.14245/ns.1938396.198]
- 2 吴迪. 基于云平台的眼科影像存储与管理系统的设计与实现 [硕士学位论文]. 苏州: 苏州大学, 2018.
- 3 Floyd DM, Trepp ER, Ipaki M, *et al.* Study of radiologic technologists' perceptions of Picture Archiving and Communication System (PACS) competence and educational issues in western Australia. *Journal of Digital Imaging*, 2015, 28(3): 315–322. [doi: 10.1007/s10278-014-9765-1]
- 4 杨振宇. 基于 Nodejs 的云 PACS 系统的设计与实现 [硕士学位论文]. 北京: 北京交通大学, 2017.
- 5 朱巧巧. 医院 PACS 系统的构建与发展研究. *江苏卫生事业管理*, 2020, 31(2): 215–217, 220.
- 6 黄江珊. 医学影像数据管理方法研究 [硕士学位论文]. 镇江: 江苏大学, 2019.
- 7 毛静. 基于医学影像 PACS 的数据管理系统. *电子设计工程*, 2018, 26(19): 16–19, 24. [doi: 10.3969/j.issn.1674-6236.2018.19.004]
- 8 王海星, 田雪晴, 游茂, 等. 人工智能在医疗领域应用现状、问题及建议. *卫生软科学*, 2018, 32(5): 3–5, 9. [doi: 10.3969/j.issn.1003-2800.2018.05.001]
- 9 吴志福. 基于 DICOM 标准的医疗影像管理系统的研究 [硕士学位论文]. 成都: 电子科技大学, 2012.
- 10 吴崇数, 林霖, 薛蕴菁, 等. 基于自监督学习的病理图像层次分割. *计算机应用*, 2020, 40(6): 1856–1862.
- 11 赵丹. 在线肺癌病例数据库的构建和初步应用 [硕士学位论文]. 上海: 第二军医大学, 2013.
- 12 于国清. 分析计算机网络技术在医院信息管理系统中的应用. *世界最新医学信息文摘*, 2019, 19(46): 266. [doi: 10.19613/j.cnki.1671-3141.2019.46.165]
- 13 Belinsky E, Lindsey A, Rubin Y, *et al.* Visualizing data in visual application design: US, 20060047699. 2006-03-02.
- 14 刘景娜, 宋伟, 田文英, 等. 基于 SSH 框架医院人事管理系统开发及应用. *微型电脑应用*, 2020, 36(3): 112–116. [doi: 10.3969/j.issn.1007-757X.2020.03.033]
- 15 Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks*, 2015, 61: 85–117. [doi: 10.1016/j.neunet.2014.09.003]