

中图分类号:R47

文献标识码:A

文章编号:1007-9416(2023)01-0052-03

DOI:10.19695/j.cnki.cn12-1369.2023.01.16

# 大数据在医疗领域的应用及问题探讨

贺晓松<sup>1</sup> 胡川丽<sup>2</sup>

1. 重庆工程学院大数据与人工智能研究所; 2. 陆军军医大学第二附属医院

随着经济的迅速发展和信息技术的变革,大数据正在改变全球绝大部分行业,医疗领域也不例外。本文概述了大数据概念、特征和医疗大数据的发展,总结了大数据在医疗领域的应用现状,并对大数据应用于医疗领域的挑战和展望作了深入探讨,旨在为我国医疗大数据的发展提供参考和借鉴。

随着信息化进程的不断加快,医疗健康领域逐渐进入大数据时代。大规模、多渠道、多样的数据可为医疗护理实践提供新的方法和思路,在提高医疗实践水平、医疗数字化管理、临床决策支持和养老服务等多个领域具有应用价值。然而,目前由于区域经济发展不平衡等原因导致的医疗资源分布不均,通过更加有效地利用医疗大数据,推进跨区域、跨机构、跨部门的健康医疗数据共享,充分发挥数据资源潜力、激活数据资源价值成为新的关注点<sup>[1]</sup>。另外,医疗大数据也面临着医疗信息数据安全和隐私保护不当、数据质量参差不齐的挑战,本文将深入探讨这些问题。

## 1 医疗大数据的发展和意义

### 1.1 大数据的定义和特征

大数据(Big Data)是指使用传统软件或基于互联网的平台无法管理的大量数据,它超越了传统使用量的存储、处理和分析能力<sup>[2]</sup>。国际数据中心(International Data Center, IDC)将大数据归结为“4V”特点的数据资料<sup>[3]</sup>:容量(Volume)、多样性(Variety)、速度(Velocity)、准确性(Veracity)和价值(Value)。医疗大数据保留了大数据的特点外,额外兼具有鲜明的医疗行业特征,即时效性、冗余性、不完整性、隐私性的特点<sup>[4]</sup>。

### 1.2 我国医疗大数据的发展

医疗大数据是人类与医疗及生命健康相关的活动过程产生的与健康医疗相关的数据集合。早期,医疗相关

的大部分数据都以纸质印刷形式而非电子信息化方式存储,但随着强大的数据存储,计算平台,及移动互联网的发展,医疗数据也向数字化方向快速发展。

我国医疗大数据伴随着国家政策的推动也迎来了快速发展的阶段,2015年8月国务院发布了《促进大数据发展行动纲要》,指出鼓励和规范有关企事业单位开展医疗健康大数据创新应用研究,构建综合健康服务应用<sup>[5]</sup>。2016年6月,国务院办公厅印发了《关于促进和规范健康医疗大数据应用发展的指导意见》,指出要大力推进不同医疗系统数据的融合和共享,促进医疗大数据安全规范和创新应用<sup>[6]</sup>。2016年8月国务院印发《“健康中国2030”规划纲要》,提出将大数据医疗及其产业经济发展成“十三五”时期医疗体系及医疗服务最高战略<sup>[7]</sup>。2018年国家卫健委发布《国家健康医疗大数据标准、安全和服务管理办法(试行)》,就加强标准管理、安全管理、服务管理,为医疗大数据的实施提供了政策指导。

### 1.3 数据的主要来源及特点

医疗大数据主要来源分为如下4类<sup>[8]</sup>:(1)医疗机构的患者就医数据和运营系统数据,如病种治疗和成本数据、医药、耗材、采购管理数据等,医院信息系统(HIS)、电子病历系统(EMRS)、医学影像存档与通信系统(PACS)等;(2)临床医疗研究和实验数据,主要是实验中产生的数据,也包含患者产生的数据,没有严格的边界区分;(3)制药企业研发数据,药物研发是一个复杂漫长的过程,临床试验是验证一个药物有效性和安全性的最佳方式,大量临床试验过程产生的数据;(4)可穿戴设备产生的数据,随着移动互联网技术的发展和穿戴设备核心硬件技术的成熟,各种健康设备通过云端上传方式收集用户生命体征数据,如血压、体温、心电、脉搏等。

收稿日期:2022-08-24

作者简介:贺晓松(1989—),男,湖北荆州人,硕士,讲师,研究方向:大数据、自然语言处理、推荐算法。



### 1.4 医疗大数据的意义

医疗大数据的研究和应用有如下几方面的现实意义<sup>[9]</sup>：

(1) 提高诊断准确性。医生利用医院共享的大数据系统，能够分析医院积累的大量临床数据，结合目标患者信息，能够及时、有效地为患者选择治疗方案；(2) 减少医患矛盾。通过建立移动护理信息系统，实现病人护理文档的电子化，将医生、护士、其他医务人员工作闭环执行，使患者得到有效治疗；(3) 节约医疗资源。通过云计算技术，将信息资源集成和共享，降低成本的同时提高效率；(4) 个性化医疗。通过大数据挖掘技术，对个体差异性特征在普遍特征数据分析的基础上进行优化分析，进而辅助以医疗服务子单元，满足个性化医疗服务要求。

## 2 大数据在医疗领域的应用

在医疗健康大数据越来越多样化的背景下，大数据技术也逐步从前沿技术转变为现实应用。无论是生产生活还是医疗健康行业，大数据的应用场景也越来越丰富。

### 2.1 医疗信息化建设

医疗信息化即医疗服务的数字化、网络化、信息化<sup>[10]</sup>。早期的医疗记录的常见做法是以手写笔记或打印报告的形式，随着计算机系统的出现及其潜力，医疗保健系统中所有临床检查和病历的数字化已成为当今的标准和广泛采用的做法。由Murphy、Hanken、Waters定义的电子健康记录(Electronic Health Records, EHR)是用于患者过去、现在或未来的身体/心理健康或状况有关的任何信息的计算机化医疗记录系统，方便医疗人员访问患者的医学诊断、处方、与已知过敏相关的数据，提高医疗识别和治疗效率<sup>[11]</sup>。与之类似，电子病历(EMR)、个人健康记录(PHR)、医疗实践管理软件(MPM)和许多其他医疗数据系统共同为提高医疗质量、服务效率和降低成本以及减少医疗错误提供有力保障。

### 2.2 临床医护决策辅助

在传统的医疗诊断主要依靠医生个人经验和目标患者的信息，局限性很大，由于患者的个体差异，同样诊断治理方式应用在不同患者身上的效果也可能不同。为此，可借助大数据建立一个智能诊断系统，如果将所有患者的就诊信息整合，运用数据挖掘技术分析患者信息，提高诊断准确率<sup>[4]</sup>。另外，可利用自然语言处理技术对电子病历文本进行分析，图像智能识别技术对医学影响建模，实现智能诊断。这能促进诊断更加科学化、合理化，提高诊断准确率，有效降低误诊率。

### 2.3 医疗科研领域中的技术支持

在医疗科研领域，包括治疗方案的疗效判断、新药研发、器械研制、疾病危险因素分析等各类临床科研<sup>[4]</sup>。运用大数据技术对各种数据进行筛选、分析，可以为科研工作提供强有力的数据分析支持。如：(1) 探索疾病关联关系。将离散数据进行整合与规范化，对大量、关联性的疾病数据进行分析整理，构建医疗知识图谱，可以探索疾病的关联关系；(2) 进行临床预测。通过数据挖掘技术对各类疾病的症状，监控疾病治疗效果数据进行建模，预测处于早期症状的疾病的发展趋势<sup>[12]</sup>。

## 3 大数据在医疗领域的挑战及应对

### 3.1 数据集成和共享

大型医院已经基本实施了信息化建设，但相应的数据资源仍然分散在不同的数据池中，互不相连，形成信息孤岛，导致医院之间无法共享护理数据，全社会卫生系统的相关数据无法共享和有效整合，进而影响医疗资源数据池的形成。

掣肘健康医疗数据共享的关键问题主要体现在管理机制和技术标准两个方面，因此亟需解决以下2个问题：(1) 由于我国经济发展的地区性差异，不同地区和级别的医疗机构自成体系、自建标准，不同类型医疗机构间缺乏充分的跨业务数据共享，因此应该健全协作机制，调整利益冲突并通过管理手段达成共识；(2) 医疗大数据本身多源异构性的特点，存在大量结构化、半结构化、非结构化数据，信息标准并不统一，按照标准规范建立起跨机构甚至跨区域的数据共享平台及数据交换机制<sup>[12]</sup>。

### 3.2 数据安全和隐私保护

大数据在促进数据传输和共享的同时，也带来了新的挑战，数据隐私保护就是其中之一<sup>[12]</sup>。我国对于医疗领域隐私保护的立法及政策法规的制定处于起步和发展中，尽管已有相关法律法规对保护患者隐私提出了相应的要求和规定，但是大部分规定缺乏可操作性，仅规定了对患者的保密义务，而没有规定违反该义务的后果，不利于具体司法实践操作，而且大多数条款对患者信息权利的具体内容、权利保护的方式等都没有规定。大数据时代，要搞数据共享，这在很大程度上受到严格的数据保护的制约；只有保障大数据的共享和保护，才能最大限度地发挥医疗大数据的潜在价值。

为了能够最大程度上保护公民的隐私安全，需要加强对数据的保护工作。首先在法律的内容上，应尽快制定相关政策法规；除完善相关配套法规标准外，也要在

执法和司法过程中强化保护患者隐私与个人信息,严格要求数据安全保障义务与责任,坚持知情同意原则,数据的收集、存储、使用、分析等行为应获得数据主体的知情同意,原则上应获得明示同意;其次在治理手段上,在强化隐私保护技术对个人隐私保护效果支撑作用的同时,致力提升数据加工者合法合规使用数据的自觉意识,形成注重数据伦理安全的良性氛围,倡导公民形成正确的个人隐私观。

### 3.3 数据质量保障

医疗数据在大部分应用场景中,对数据的质量都有比较高的要求。医疗数据质量风险可能导致一系列问题:(1) 即时接受诊疗的患者因为医生收到了错误的基本疾病信息受到不合理的治疗甚至承担安全风险;(2) 医疗科研有可能因为出现的错误数据而得到错误的结论,轻者会导致整个科研过程失败,耗费人力财力,重者可能会错误引导医学理论发展方向,影响整体医学发展进程。

为了提高医疗数据的准确性、真实性、适用性,须建立一套标准化数据治理模式,各级医疗机构在开展医疗数据采集、存储、整合、分析等工作的时候,以相同或相近的标准,确保内容准确、术语规范、信息数据接口一致等<sup>[13]</sup>。此外人工智能技术也正在应用于医疗数据的提取和识别过程中,通过自然语言处理、人工智能、OCR 等技术直接识别电子病历系统并完成有效数据的提取,这一过程的实现将有效减少医疗数据提取过程的人力耗费且有可能通过机器学习提高数据的准确性。这些医疗数据提取的自动化策略必然是医疗数据提取的重要未来方案,对于解决当下的医疗数据准确性安全性风险意义重大。

### 引用

- [1] 石晶金,于广军.健康医疗大数据共享关键问题及对策[J].中国卫生资源,2021,24(3): 223-227+237.
- [2] 侯黎莉.护理信息化的发展及临床应用[J].上海护理,2022,22(5):72-75.
- [3] 大数据战略重点实验室.大数据概念与发展[J].中国科技词语,2017,19(4):43-50.
- [4] 戴明锋,孟群.医疗健康大数据挖掘和分析面临的机遇与挑战[J].中国卫生信息管理杂志,2017,14(2):126-130.
- [5] 中华人民共和国国务院.促进大数据发展行动纲要[J].成组技术与生产现代化,2015(3):51-58.
- [6] 国务院办公厅.国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见[EB/OL].(2016-01-21)[2021-02-20].[http://www.gov.cn/gongbao/content/2016/content\\_5088769.htm](http://www.gov.cn/gongbao/content/2016/content_5088769.htm).
- [7] 中共中央国务院印发《“健康中国2030”规划纲要》[EB/OL].(2016-10-25)[2018-03-22].[http://www.gov.cn/zhengce/2016-10/25/content\\_5124174.htm](http://www.gov.cn/zhengce/2016-10/25/content_5124174.htm).
- [8] 杨朝晖,王心,徐香兰.医疗健康大数据分类及问题探讨[J].卫生经济研究,2019,36(3):29-31.
- [9] 李宜瑾,崔豫然.医疗大数据研究面临的机遇与发展趋势[J].信息与电脑(理论版),2019,31(19):29-31.
- [10] 汪萌.医疗信息化助力医院发展[D].南京:南京邮电大学,2017.
- [11] Miriam Reisman.EHRs:The Challenge of Making Electronic Data Usable and Interoperable[J].Pharmacy and Therapeutics,2017,42(9):572-575.
- [12] 王海星,张靓,杨志清,等.医疗大数据在临床科研中的应用探讨[J].中国医院,2020,24(7):63-64.
- [13] 叶清,刘迅,周晓梅,等.健康医疗大数据应用存在的问题及对策探讨[J].中国医院管理,2022,42(1):83-85.