

·数字经济·

基于电子健康记录的智能算法研究 综述



□李天皓 张倩 陆炜

[电子科技大学 成都 611731]

[摘要] **【目的/意义】** 大数据时代下,人工智能、区块链等信息技术得到了迅速发展。电子健康记录(Electronic Health Records, EHRs)涵盖了大量医疗的关键信息,基于EHRs的智能算法在提升智慧医疗水平方面发挥着重要作用,该领域吸引了众多国内外学者的兴趣,取得了丰硕的成果。

【设计/方法】 通过整理Web of Science和CNKI两大核心数据库的EHRs的相关研究文献,结合国家政策背景和国内外EHRs数据库的发展现状,用可视化图表分析了国内外EHRs相关研究的现状和发展趋势,并基于智能算法在EHRs的应用进行分类,展示了智能算法与EHRs相结合的研究成果。**【结论/发现】** 通过回顾性研究,揭示了EHRs与深度学习、大数据、强化学习和区块链等人工智能技术的关系,发现我国在EHRs数据库建设与利用方面的不足,并对此提出可改进的相关建议与对策。

[关键词] 电子健康记录; 电子病历; 智慧医疗; 深度学习

[中图分类号] F208

[文献标识码] A

[DOI] 10.14071/j.1008-8105(2022)-1055

Review of Research on Intelligent Algorithms Based on Electronic Health Records

LI Tian-hao ZHANG Qian LU Wei

(University of Electronic Science and Technology of China Chengdu 611731 China)

Abstract [Purpose/Significance] Under the era of big data, information technologies such as artificial intelligence and blockchain have been developed rapidly. Electronic health records (EHRs) cover a large amount of key information on medical treatment, and intelligent algorithms based on EHRs play an important role in enhancing intelligent medical treatment. This field has attracted the interest of many scholars at home and abroad and achieved fruitful results. [Design/Methodology] By collating the research literature related to EHRs in the two core databases of Web of Science and CNKI, and combining the national policy background and the development status of EHRs databases at home and abroad, we analyzed the current situation and development trend of the research related to EHRs at home and abroad with visual charts, and classified the research results based on the application of intelligent algorithms in EHRs, showing the combination of intelligent algorithms and EHRs. [Conclusions/Findings] Through a retrospective study, the relationship between EHRs and artificial intelligence technologies such as deep learning, big data, reinforcement learning and blockchain is revealed, the shortcomings in the construction and utilization of EHRs database in China are identified, and relevant suggestions and

[收稿日期] 2022-07-23

[基金项目] 国家重点研发计划(2018AAA0101003); 国家自然科学基金(71901050)。

[作者简介] 李天皓(1995-)男,电子科技大学经济与管理学院博士研究生;张倩(1998-)女,电子科技大学经济与管理学院硕士研究生。

[通信作者] 陆炜(1987-)女,电子科技大学经济与管理学院副教授、硕士生导师。E-mail: luwei@uestc.edu.cn.

countermeasures are proposed.

Key words electronic health records; electronic medical records; intelligent healthcare; deep learning

引言

电子健康记录(Electronic Health Records, EHRs),是指以计算机可处理形式存在的关于护理主体健康状态的信息存储库,通常又被称作电子健康档案或电子病历^[1]。EHRs由医疗服务机构与病人一次或多次交互产生,主要记录的信息包括病人的人口统计资料、用药记录、生命体征、临床病史、实验室检测结果以及诊断报告等信息^[2-3]。电子医疗记录(EMR)与EHRs相似,但EMR最常指的是单一的医疗事件,而EHRs包含了病人的整个医疗记录情况^[4]。目前,EHRs的定义在学术界仍缺乏统一的认识。一方面,EHRs本身的功能形态还在不断发展之中;另一方面,不同国家和组织根据自己的需求和理解给出定义^[5],但人们对EHRs应当具备的一些基本特征有着相近的认识^[6]。简而言之,学术界对EHRs认识的共同点大于分歧。

EHRs是旨在与其他医疗卫生组织和机构(例如实验室、专家、医学影像设施,药房,急救设施以及学校和工作场所诊所)共享信息,合理利用EHRs的信息对于控制公共传染病、干预卫生应急事件、预防与监测慢性疾病、改善患者护理、增强临床决策支持和提高全民健康水平均有重要的作用^[7-8]。电子健康记录系统和其他的健康数据数字化系统一样,它可以让医疗保健变得更为智能、安全、高效,在这个过程中,区块链、人工智能和大数据等智能技术发挥着巨大的推动作用^[9-10]。

本文将“电子健康记录”“电子健康档案”和“电子病历”作为关键词,结合智能算法应用进行综述。主要工作包括:对比电子健康记录的国内外研究趋势和分析EHRs的发展情况,并针对国内外结合EHRs和人工智能、大数据、区块链等智能算法的研究情况进行分析,最后提出我国发展EHRs的建议。

一、国内外EHRs研究与使用情况

对国内外EHRs的应用情况进行分析有助于梳理国内电子病历发展历程、分析国内外文献发表和主题演化情况,从而梳理EHRs主题的研究状况。

(一)国内外EHRs建设情况

我国EHRs的建立起步较晚,从2000年左右开始,逐渐有少量的大型医疗机构使用计算机代替手

写病历。在2003年非典流行时期,EHRs出现过一次快速兴起,由于当时医院建立了隔离区,无法与非隔离区直接进行物质和信息互换,很多医院使用传真机在隔离区和非隔离区之间交换电子病历文档^[11]。而国外发展EHRs的时间较早,如美国和日本,从1960年就开始将计算机技术应用于日常的病历系统中^[12-13]。目前最新的研究报告表示,美国公立医院的EHRs采用率达到了88%^[14]。2017年,我国也有94%的医院使用来自电子健康档案的电子临床EHRs数据,这些数据最常用于医院质量改进、监测患者生命体征和衡量组织绩效^[15]。

我国政府十分重视电子病历的发展情况,自2002年10月以来,出台了发展电子病历、数字化医疗的卫生信息化发展纲要文件《全国卫生信息化发展规划纲要(2003~2010年)》。在2009年形成了电子病历的基本框架和出台相应的标准《电子病历基本架构与数据标准(试行)》,随后通过试点、改革、建立分级制度等措施,逐步完善了我国电子病历的发展机制。2018年发布《电子病历系统应用水平分级评价标准(试行)》,并要求到2022年,全国二级和三级公立医院电子病历应用水平平均分级达到3级和4级。目前,随着各地医疗机构的电子病历的建立,互认共享和患者信息脱敏等问题还有待解决。2022年2月,国家卫健委发言人表示,正在研究建立全国统一的电子病历。历经初步试点、推广普及、规范建设,我国电子病历逐渐向高质量管理方向迈进。

(二)国内外EHRs研究情况

通过对国内外文献发表的情况进行分析,可以知晓主题的发文情况与趋势。本文通过伊玛目阿卜杜拉赫曼本费萨尔大学的电子资源门户访问Web of Science(WOS)核心合集(WOSCC)数据库,以EHRs为关键词索引外文文献,并对文献出版数量进行分析。WOS数据库是科学和学术出版的选择性引文索引,涵盖期刊、会议论文集、书籍和数据汇编,被全球研究学者广泛认为是最可靠的科学引文索引^[16]。而中文文献采用了CNKI数据库作为文献来源,包含了核心期刊:CSSCI,以“电子病历”“电子健康记录”和“电子健康档案”为关键词进行检索。中英文文献检索的时间区间为2000年1月~2022年3月,一共检索到26 361篇英文文章和1 141篇中文文章,并使用Pyechars进行发文

数据分析。

1. 发文量分析

根据2000~2021年的国内外发文量,绘制了国内外发文对比图,如图1所示。可以看出,自2000年以来,以EHRs为主题的文章数量逐年增加,特别是近十年来(2010年~2020年)发文量从581篇

增长到4436篇,增长近八倍,说明了EHRs正不断获得国际学者们的关注与研究。特别是,2020年以来,文章数量增长加快,达到一年600余篇的发文量。自2019年起,由于新冠肺炎的影响^[17],通过发文趋势图可以看出,EHRs发挥了重要的作用,如整合病历数据、分析致死原因^[18]等。

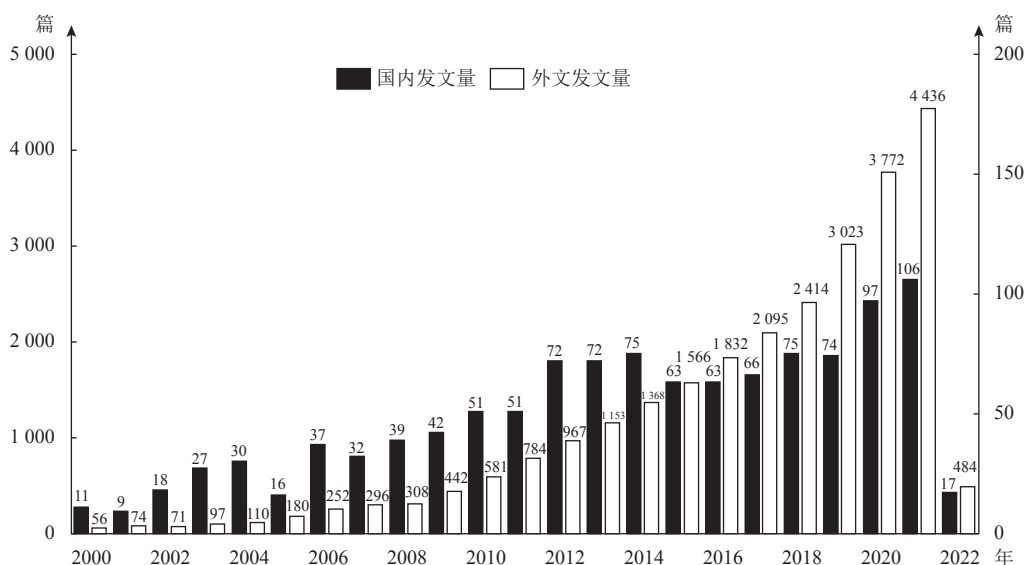


图1 国内外发文量对比图

中文发文量从2000年以来呈现了稳步上升的态势,最高是2021年,发文106篇。从数量上对比,国内发文量不足外文发文量的1/40,且在2014年后国内发文量下降。究其原因,本文进一步整理了国内机构在WOS上的发文情况,如图2所示。从图2可以看到,国内学者近些年将相关主题的文章发表在了国外的期刊上,且最高是2021年的278篇,是国内的两倍多,且2016年后,国内机构外文发文量

明显增多。一方面,说明中国学者也逐渐关注EHRs并将其纳入研究范围,他们将研究成果更多地发表在国际期刊,从而扩大研究成果的影响力,侧面反映出我国EHRs建设的有效性。另一方面,随着技术的不断发展,近年来EHRs的定义和使用也越来越规范化,但国内缺乏公开和完备的医疗电子数据库,许多学者更倾向于投稿到国外的期刊,以获得更广泛的认可。

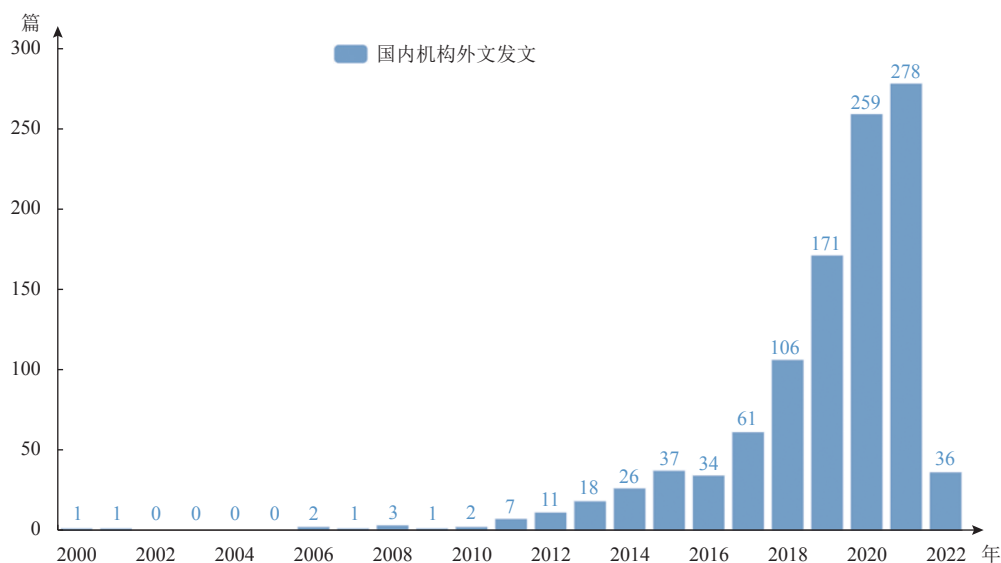


图2 国内机构外文发文情况

2. 主题词演化情况

主题词演化图能够清晰地展现研究的热点趋势与各个关键词之间的相关性。本文通过对近20年来WOS和CNKI中EHRs相关文献关键词的提取, 选择

了出现频率最高的5000个关键词, 基于自然语言处理、降维聚类等技术对数据进行处理后, 利用Biblioshiny (v1.3.2) 进行可视化, 将他们按照四个时间段制作成桑基流程图, 如图3所示。

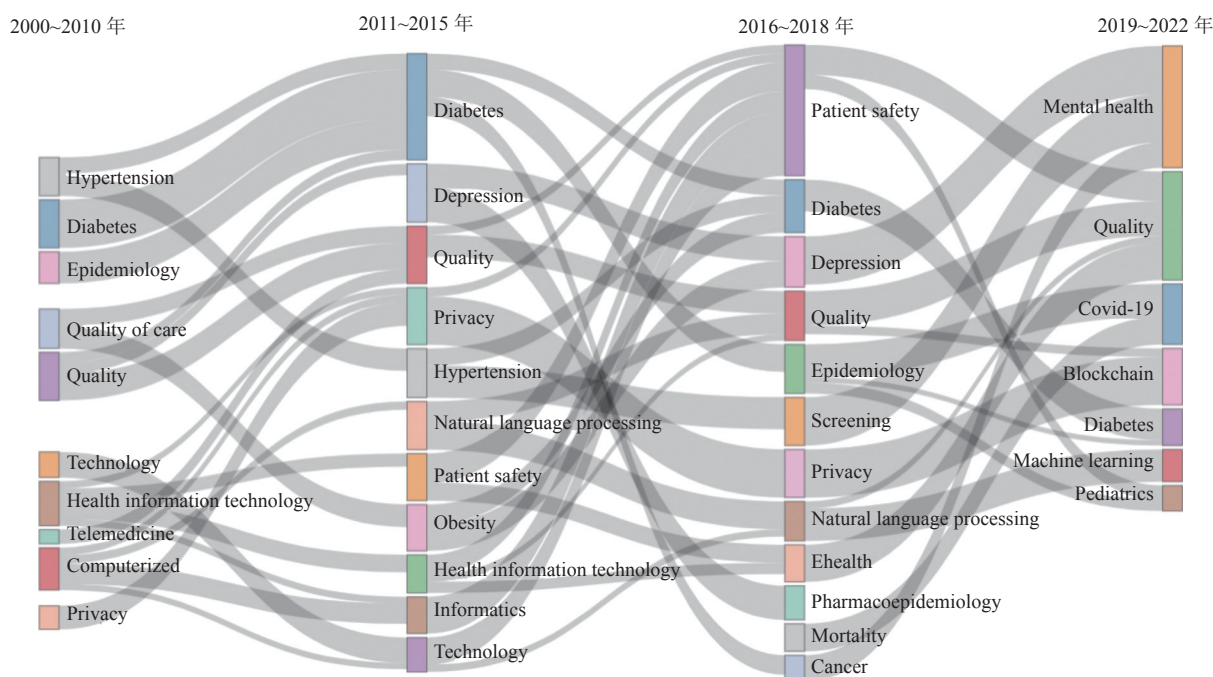


图3 外文文献主题词变化图

图3中用灰色路径表示主题关联程度, 得到了文献主题的演化图。可以从图3中看到, 第一个阶段(2000~2010)年, hypertension(高血压)和diabetes(糖尿病)是学者们利用EHRs最爱研究的主题。第二个阶段(2011~2015)年, 这一时间段出现的主题是上一时期研究过的主题的演变, 并且在内容上有联系, 例如diabetes(糖尿病), 与上一时期中的hypertension(高血压)、epidemiology(流行病学)和quality of care(护理质量)等有一定联系, 是研究延续性的体现。除了糖尿病外, depression(抑郁症)、quality(质量)和privacy(隐私)居于前列, 表明研究者们越来越关注与发挥EHRs的作用, 提升患者的生活质量和精神状态。第三阶段2016~2018年, 患者安全成为最热门的主题, 患者安全包括病人安全、用药安全、医疗保健安全等方面, 是提高医疗质量的关键。第四阶段(2019~2022)年, 随着COVID-19的爆发, EHRs被用于预测患者死亡率、预测患病人数等^[19]。与此同时, 区块链与EHRs的结合也成为了新的热点, 区块链能够很好地解决EHRs中患者隐私的问题。整体而言, 主题演化图体现了基于EHRs的研究主题在时间和内容上的变化情况, 能够更快地把

握研究热点与主题之间的变化, 为研究提供方向。

同理, 国内文献主题演化图如图4, 由于国内文献数量相对少, 本文只划分了三个时间区间进行分析。第一阶段(2000~2010年)的相关研究主要集中在医院信息系统和社区卫生服务相关研究, 此时医疗纠纷问题成为该阶段的热门主题, 医院信息系统的建设备受关注。第二阶段(2012~2015年)我国通过试点与改革逐步实现电子病历的规范化, 病案管理也成为这一时期的研究热点。第一阶段至第二阶段, 在国家信息化建设的发展进程中, 病历模板和居民健康档案都逐步实现电子化, 这一阶段的研究热点也随之发生演变。第三阶段(2016~2022年), 利用EHRs进行命名实体识别研究成为热门趋势, 大数据与人工智能技术也逐渐成为主要研究方向, 同时利用EHRs研究新型冠状病毒肺炎成为了新的研究方向。从第二阶段至第三阶段可以了解到, EHRs与区块链联系紧密, 区块链独特的加密技术和去中心化体系能够有效地保护患者隐私, 提高存储效率。

总体而言, 医疗档案与病历管理等关键词始终是研究的趋势, EHRs需要借助医疗档案管理的相关研究进行完善。随着信息化的发展, 大数据与病

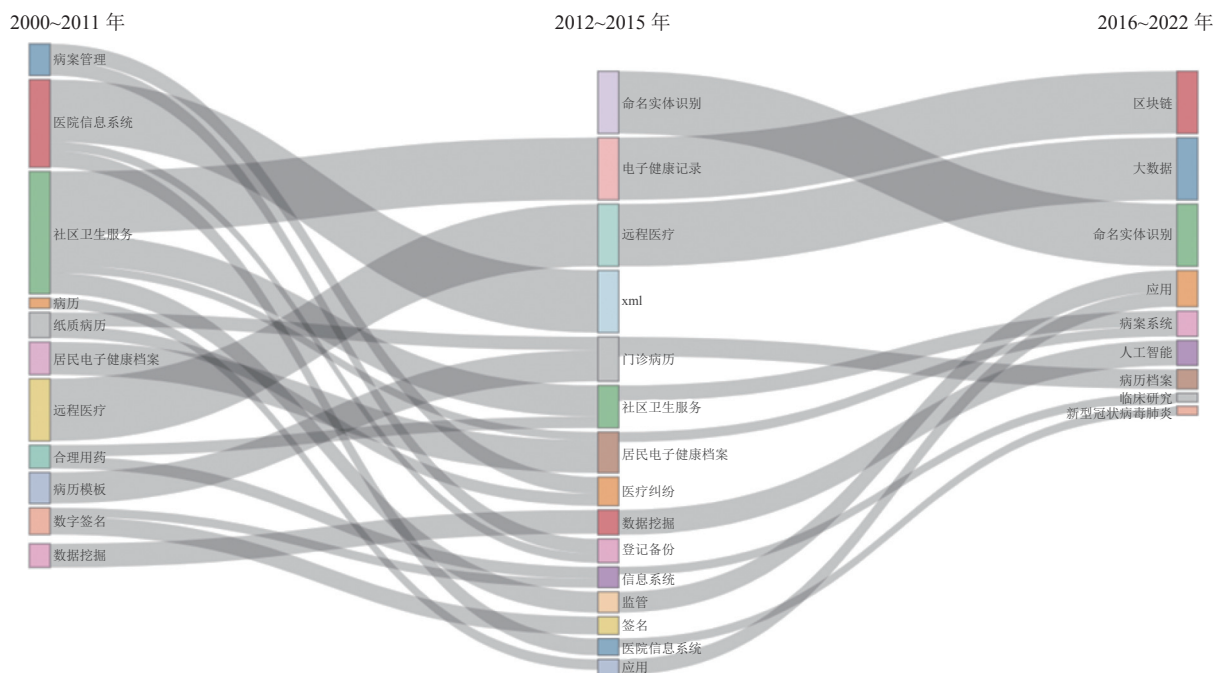


图 4 中文文献主题词变化图

历档案系统、数据库等主题词关系紧密, 人工智能与数据挖掘联系紧密, 智慧医疗也由医院信息系统延展而来。通过主题演化图, 能够了解当下EHRs的主题与之前的主题演化关系, 清楚地认识到主题变化的过程。

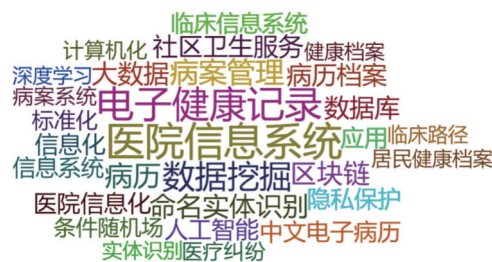
3. 关键词词云分析

词云图可以直观展示大量文本数据及其显著性。以EHRs为主题的论文中使用的关键词随着年份在不断变化, 尤其是与已经发表了很长时间的论文相比, 最近发表的论文更能展现当下的研究热

点。本文根据关键词出现的频率展现了近三年最热门的30个关键词, 得到了如图5所示的词云图, 其中单词的大小对应着出现频率的高低。可以从图5(a)中看出, 外文文献研究中最热门的关键词集中在machine learning (机器学习)、blockchain (区块链)、COVID-19与privacy (隐私)等。而机器学习则是最热门的话题, 目前也有基于EHRs的外文综述, 分别聚焦于深度学习^[21]、疾病诊断^[20]等方面进行分析, 国内尚且缺乏基于EHRs的智能算法类的综述文章。



(a) 英文词云图



(b) 中文词云图

图 5 关键词词云图分析

类似地, 从图5(b)中的中文词云图可以看出, 国内学者的热门研究点主要在医院信息系统、数据挖掘和病案管理等方面。该方面的研究多数在如何建立健全医疗电子档案^[21]、医疗电子档案的共建与分享^[22-23]和当下国内电子病历管理存在的问题^[24]等。相比于外文文献, 国内学者更看重医疗信息化

方面的研究。一方面, 信息化技术将改善医学数字信息的“孤岛”, 从而提高患者就医体验和医生工作效率, 减少医患矛盾^[25-26]。另一方面, 得益于国内政策引导, 我国从2011年开始加强了医疗信息化建设的步伐。同时, 基于大数据、区块链、深度学习和人工智能等智能技术在EHRs领域的应用也得

到了研究者的关注^[27-29]。

（三）国内外现有研究综述

随着人工智能相关的硬件与软件的发展，近年来医疗信息化记录产生的数据被学者们广泛用于研究。本文梳理了近年来现有的基于EHRs的国内外

研究情况，如表1所示。可以从表1看出，现有的国外文献聚焦于不同的具体发展方向，如自然语言处理、表征学习、疾病诊断等方面的方法综述，而国内文献聚焦于文本挖掘与实体识别、档案建设和疾病筛查等问题。

表 1 国内外现有研究综述

序号	文献	年份	主题
外文			
1	Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis ^[2]	2018年	深度学习
2	Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: a complete review ^[20]	2020年	疾病诊断
3	Applications of blockchain in ensuring the security and privacy of electronic health record systems: a survey ^[30]	2020年	区块链与隐私
4	Deep representation learning of patient data from electronic health records (EHR): a systematic review ^[31]	2021年	深度表征学习
5	Using the electronic health record to enhance care in pediatric rheumatology ^[32]	2021年	疾病护理
6	Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies ^[33]	2021年	深度时序学习
7	Data management and integration with electronic health record systems ^[34]	2021年	数据库管理
8	The impact on health outcomes of implementing electronic health records to support the care of older people in residential aged care: A scoping review ^[35]	2021年	老年人护理
9	Electronic health records: Three decades of bibliometric research productivity analysis and some insights ^[3]	2022年	文献回顾
10	Electronic health records and genomics ^[36]	2022年	基因分析
中文			
1	电子健康档案数据质量评估与治理的综述研究 ^[37]	2018年	档案建设与评估
2	基于电子病历的患者相似性分析综述 ^[38]	2018年	样本分析
3	基于EHR的医疗知识图谱研究与应用综述 ^[39]	2018年	知识图谱
4	电子健康档案研究进展综述 ^[7]	2019年	档案建设与评估
5	电子病历数据库脓毒症病例筛选方法综述 ^[40]	2020年	疾病诊断
6	中文电子病历的命名实体识别研究进展 ^[41]	2020年	实体识别
7	基于电子病历识别谵妄的研究现状与展望 ^[42]	2020年	疾病诊断
8	基于区块链技术的电子医疗数据共享研究综述 ^[43]	2021年	区块链共享
9	电子病历文本挖掘研究综述 ^[44]	2021年	文本挖掘
10	国外电子健康记录信息语义整合研究述评与展望 ^[45]	2021年	语义整合
11	机器学习在脊柱疾病智能诊治中的应用综述 ^[46]	2021年	疾病诊断

（四）国内外现有医疗电子数据库情况

数据库是EHRs的载体，国外特别是美国具有许多健全和广泛的医疗电子数据库。而我国已初步建立健康医疗数据库，但仍存在着诸多问题，如质量差的特征、缺乏统一标准、医疗机构间数据孤岛等问题。这些问题部分由于国内数字化起步较晚，多元化数据的整合较少，医疗数据呈现出数量大（因为人口基数大）的特点。

任务型数据库如Kaggle和UCI中的部分数据库，由于缺乏病人整体的治疗过程和病人基本特征等信息，按照EHRs的定义不在本文的探究范围。表2展示了部分国内外医疗电子数据库的情况。

国外医疗电子数据库有综合的急诊科ICU数据库如MIMIC^[47]和eICU，也有专科数据库如SEER肿瘤数据库和MURA骨科数据库。其中MIMIC数据库具有数据量大、存储格式规范、易读取等特点，被学者广泛用来研究。而国内的数据库存在数据库种类少、数据类型不够丰富、存储不规范不易读取等

问题。随着云计算和云存储的发展，EHRs的储存与读取问题将会得到改善，有利于学者进行后续的研究^[48]。

二、国内外基于EHRs的智能算法发展情况

目前，国内许多医院都因无法有效地利用EHRs进行数据分析来为他们的临床实践生成高质量的见解而苦恼^[49-50]。临床上产生的EHRs必须要加以利用才能发挥其作用，但EHRs储存的数据必须经过数据提取、分离、清洗等操作后才可为研究所用。针对不同的医疗场景和问题，采取合适的处理方式是解决问题的关键，本节将按照不同的智能算法方法对国内外利用EHRs的研究进行综述。

（一）时序类模型

循环神经网络（Recurrent Neural Network, RNN）算法通常用于处理时序类数据或文本类数

表 2 国外现有医疗电子健康数据库

序号	数据库名称	数据库来源	数据类型	公开情况
国外				
1	Medical InformationMart for Intensive Care (MIMIC)	美国马萨诸塞州波士顿三级学术医疗中心	时序、图像、文本	条件公开
2	i2b2	i2b2 tranSMART 基金会	文本	条件公开
3	eICU协作研究数据库	飞利浦医疗保健、麻省理工学院计算生理学实验室	数值	条件公开
4	美国SEER肿瘤数据库	美国国立癌症研究所等	数值	条件公开
5	BioLINCC数据平台	弗雷明汉心脏研究	数值	条件公开
6	MURA数据库	斯坦福吴恩达团队	图像	完全公开
7	QRESEARCH数据库	英国公共卫生部等	数值	条件公开
8	OAI公共数据集	OAI研究团队	数值、图像	条件公开
10	TCGA数据库	美国癌症基因组图谱计划	数值、图像	完全公开
11	Orphanet数据库	法国国家健康与医学研究所	数值、文本	条件公开
12	BMC临床数据仓库	波士顿医疗中心	数值	条件公开
13	美国临床试验注册平台	美国国家医学图书馆	数值、文本、图像	条件公开
14	MedlinePlus Connect	美国国立卫生研究院等	数值	完全公开
国内				
1	儿科重症监护 (PIC) 数据库	浙江大学医学院附属儿童医院	序列	条件公开
2	长庚研究数据库 (CGRD)	长庚纪念医院	数值、文本	条件公开
3	中国医院质量监控系统 (HQMS)	国家卫生健康委员会	文本、数值	未公开

注：条件公开数据库指需要通过培训或获得相应资质认证才可以免费获取的数据库

据。RNN算法通过循环神经元，使得一个序列的当前输出与之前的神经元有联系，从而向着序列的演进方向进行链式递归。由于RNN的梯度会随着时序不断积累从而出现指数级衰减，存在梯度消失问题，导致RNN的性能受到了制约，无法解决数据的长期依赖问题。LSTM作为RNN的一种变体，能够很好地处理长期时间序列数据，GRU是LSTM的一个简化版，具有与LSTM相同性能下收敛更快、参数更少的特点。GRU公式如式（1）给出：

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1}) \\ r_t &= \sigma(W_r x_t + U_r h_{t-1}) \\ \tilde{h}_t &= \tanh(W^h x_t + U^h (h_{t-1} \odot r_t)) \\ h_t &= (1 - z_t) \odot \tilde{h}_t + z_t \odot h_{t-1} \end{aligned} \quad (1)$$

图6展示了GRU的结构，图6和式(1)中的 x_t 表示输入， h_t 表示隐藏状态， $\sigma, \tanh$ 表示激活函数， r_t 与 z_t 分别表示重置门和更新门。

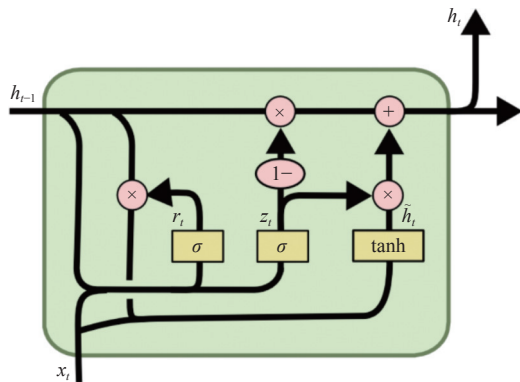


图 6 GRU网络结构图

通常RNN类算法擅长处理时序数据，被广泛应用于EHRs的研究中，如对病人的死亡风险进行预

测^[51-53]、病人再次入院间隔^[54]、或者疾病预测^[55-56]等方面。对于语言类数据，LSTM及其变体常常被用来对医学文本实体命名以提高医生查看病历的工作效率。由于在LSTM模型中，信息只能向前传播，双层循环神经网络BiLSTM很大程度上改善了LSTM对上下文信息的学习。由BiLSTM继续发展成后来的Transformer和BERT^[57]模型，都成为了自然语言处理（NLP）领域同时期的SOTA模型。生物医学文本挖掘任务的语境化语言表示模型BioBERT^[58]也有助于理解复杂的生物医学领域文本。

BERT类算法存在模型庞大、参数多、收敛缓慢等问题，在实际生产生活应用中还有一定的距离。因此，现阶段BERT研究方向多聚焦于在模型性能差异不大的情况下，尽可能地压缩模型大小^[59]。

（二）卷积类模型

卷积神经网络（Convolutional Neural Networks, CNN）在图像分类、语音识别和句子分类方面取得了优异的表现。每个卷积神经网络包含了一个卷积层和池化层，卷积层可以叠加形成深度卷积网络。卷积层通过滑动核心块，对输入数据进行卷积，从而抓住局部的数据特征，一维卷积计算公式由式（2）给出：

$$C_{1d} = \sum_{a=-\infty}^{\infty} x(a)w(t-a), \quad (2)$$

其中 x 为输入信号， w 为加权函数或卷积过滤器。经过卷积后，需要经过一个池化层以提取主要的数值特征，如最大值、平均值等。

在EHRs中，CNN常用于疾病诊断和风险预测任务中，通过提取数据表征进行学习。Cheng等使

用RNN在EHRs中进行表征分析,通过时间-事件的二维矩阵对患者信息进行表征,构建CNN模型进行表征提取和风险预测^[60]。CNN也被用来预测EHRs患者死亡率和检测不良事件^[61-62]。在预后护理方面,CNN通过对医学图像数据的利用与挖掘,可以实现对乳腺癌患者的药物反应预测^[63]。

CNN主要应用于医学图像处理。病人的影像学检查会产生许多图像数据,如胸透和CT^[64-66]等。通过对这些图像进行学习可以实现自动化的疾病诊断,如甲状腺癌的筛查^[67]等,从而有效降低医生负担,同时提高临床诊断效率。EHRs的文本数据可以辅助医学诊疗决策,CNN可以充分利用文本数据实现更好的诊疗效果。在医学文本分类任务中,如TextCNN^[68]利用预训练好的词向量通过CNN进行文本分类,并在儿科疾病诊断中得到了良好的应用^[69]。在智能疾病诊断任务中,CNN可与NLP技术相结合^[70],充分利用半结构化或非结构化的医疗文本数据,如医嘱、手术记录、护理记录、前台登记数据、既往病史等,实现疾病特征的表征学习和诊断。

尽管CNN在各种诊断任务中都达到了医生级别的准确度,但模型可解释性的缺失不利于算法的临床落地。另一方面,多学科疾病的诊断和更复杂多模态信息下的诊断,也是CNN类算法的研究方向^[71]。

(三) 生成类模型

变分自编码器(Variational Auto-Encoder, VAE)^[72]和生成对抗网络(Generative Adversarial Networks, GAN)^[73]都是生成模型(Generative model)的代表。所谓生成模型,即能自动生成样本的模型。可以将训练集中的数据点看作是某个随机分布抽样出来的样本,如果能够得到这样的一个随机模型,便可以得到这个生成模型,但这个随机分布需要通过对训练集的学习来得到或逼近。

由于GAN在实际的生成效果中比VAE更优秀,这里我们只介绍GAN的工作原理。GAN由Goodfellow于2014年提出,是一种两个神经网络互相竞争的特殊对抗过程^[73]。第一个网络为生成器G,用于生成数据,第二个网络为判别器D,用于区分生成器创造出来的假数据。GAN目标函数 V 则由式(3)给出:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{\text{data}}(x)} \log D(x) + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))], \quad (3)$$

其中, x 表示数据样本, z 表示白噪声, $p_{\text{data}}(x)$ 表示生成模型分布,而 $p_z(z)$ 表示噪声的先验分布。原始的GAN存在模式崩溃以及难以收敛等问题,

WGAN(Wasserstein GAN)将衡量生成器和判别器数据分布之间的距离公式改为 Wasserstein距离^[74],它相对KL散度与JS散度具有一定的平滑特性,理论上可以解决GAN梯度消失的问题。为了使GAN能够更好地适应卷积神经网络的架构,DCGAN(Deep Convolutional GAN)通过替换池化层、删除全连接层和使用批归一化的方式实现更好的生成效果,进一步提升了GAN的稳定性和生成结果的质量^[75]。近年的Lipschitz GAN将判别器的Lipschitz常数约束为小于等于1,避免了梯度 Uninformative 的问题,其生成样本的稳定性和质量均优于WGAN^[76]。到目前为止,GAN已经有数百种变体,如LSGAN^[77]、ACGAN^[78]等,以适应不同领域的任务。

GAN十分擅长无监督学习的任务,特别是生成逼真的医学图像,如利用GAN对胸腺图像进行扩增,用于识别胸腺癌,辅助医生进行临床诊断^[79]。同时得益于GAN的生成能力,可以对医学结构化数据进行扩增,从而减少由样本不平衡带来的训练误差^[80-81]。由于EHRs数据记录着患者的隐私问题,导致在数据共享中受到限制。GAN为EHRs数据的可替换性提供了解决方案,通过捕获多维、异构的数据特征,生成逼真的多模态EHRs数据,降低数据采集和共享的障碍,保护患者的隐私。同时,GAN也可以作为补全缺失数据的方法之一^[82],利用GAN学习已有数据的分布特征,利用生成器对缺失数据进行填补,达到比传统补全方法更好的填补效果。

目前,GAN的发展仍然面临诸多挑战,如全局收敛性的证明和对抗样本的困扰等^[83],但GAN依然是生成模型中最具有潜力的模型,未来可以利用GAN模型生成更多高清的医疗图像辅助临床医生诊断。

(四) 强化学习

强化学习(Reinforcement Learning, RL)是一种以目标为导向的智能决策技术,它以马尔可夫决策过程(Markov Decision Process, MDP)为理论基础,描述了如何根据与环境的重复交互所得到的经验在顺序决策过程中学习最佳策略。MDP通常由 $\langle s, p, a, r, \gamma \rangle$ 五元组构成,分别为状态空间 s 、转移概率 p 、动作空间 a 、奖励函数 r 和折现因子 γ 。经典的强化学习方法,如Q-learning或其变体,利用迭代计算出状态的动作价值函数 Q ,如式(4)给出:

$$Q^\pi(s, a) = \mathbb{E} \left[\sum_{t'=t}^{\infty} \gamma^{t'-t} r_{t'} \mid s_t = s, a_t = a \right], \quad (4)$$

其中, π 表示由状态 s 和对应动作 a 构成的策略

空间, Q-learning通过最大化累计奖励达到最终目标。当状态和动作空间高维且不连续时, Q-learning则无法工作, DQN(Deep Q Network)将RL和深度学习(Deep Learning, DL)相结合^[84], 利用神经网络拟合Q值, 成功解决了上述问题, 但依然存在Q估值过高的问题, DDQN则采用了两个神经网络分别估计动作的选择和动作的评估来消除Q值高估的问题^[85]。

在EHRs相关研究中, RL可以为患者提供最佳的个性化治疗方案^[86-87]。但是传统的RL方法(如Q-learning)无法解决多维状态下的决策问题。随着DL技术的不断发展和成熟, 深度强化学习(Deep Reinforcement Learning, DRL)算法与EHRs相结合的研究取得了许多进展, 如使用DQN学习最佳肝素给药策略^[88]、采用DDQN算法建议脓毒症患者的静脉注射液和升压药剂量^[89]、基于DDQN为病人提供最佳的麻醉剂量建议^[90]等, 这些方法都取得了超越人类医师的治疗效果。DL因其可解释性问题而存在一定缺陷, 研究人员将RL与博弈论结合, 在并发症的治疗决策问题中使用夏普利值解释临床特征的重要性, 为II型糖尿病患者建议个性化治疗方案^[91]。尽管单纯的RL方法已经能够成功地应用于临床医疗辅助决策, 但是智能体通过“探索”“试错”和“奖励指导”来学习, 可能导致学习到的策略威胁患者的健康, 尤其在药物的相互作用中, 决策的安全性尤其重要。为了更安全地提供辅助决策, 监督学习可以与强化学习相结合, 为患者学习一个更加“安全”的治疗方案, 以确保处方的低风险性和安全性^[92]。

随着多智能体技术不断发展, 学者们开始将多智能体强化学习应用在EHRs数据中, 以支持临床决策的研究。首先, 针对多个医生会诊治疗的场景, 学者们采用了多个智能体模拟多医生会诊场景, 实现对患者关键生理指标的最佳控制^[93]。其次, 由于患者的状态是多维的, 存在治疗周期长、治疗药物复杂等问题, 在应用强化学习解决医疗问题中, 往往面临着奖励的稀疏问题。因此, 有研究采用了多智能体分层强化学习辅助临床决策, 通过实施具有层级特点的多智能体对策略进行加速学习^[94]。

尽管RL在EHRs数据利用方面显示出可靠的前景, 但是在实际应用中还需要考虑诸多问题, 如状态和动作的处理、奖励函数的制定、智能体的探索策略设计、模型策略的评估和在多智能体强化学习中的信用分配等问题^[95]。

(五) 区块链技术

区块链技术是一种分布式数据存储技术^[96], 它将需要存储的交易信息通过hash算法编译成区块, 并通过链的形式与其他区块相连接, 具有透明、匿名、不可篡改等特点。

EHRs系统在医疗应用过程中面临着互操作性差、信息不对称和数据泄露等问题^[97]。区块链因其安全性、匿名性和数据完整性等优势, 使得在其上存储患者的医疗记录成为合理的选择。EHRs数据在上传区块链后, 任何人无法更改和删除, 确保了病人记录的准确性和唯一性^[98], 但这项技术需要电子签名、加密算法、云存储等技术的支持。电子签名能够提高数据访问的安全性, 如多权限电子签名^[99]、基于角色的身份验证^[97]等方式, 以增强签名的不可伪造性。区块链加密算法中的公钥加密^[100]、对称加密^[101]等加密技术增强了医疗数据的安全性, 但密钥的管理成为数据加密的关键, 为了防止私钥不被泄漏, 采用轻量级的密钥备份和恢复方案^[102]成了不错的选择, 或使用智能合约授权用户的密钥访问权限^[103]来管理密钥。最后, 区块链中EHRs的存储大都采用云数据库与链下数据库存储方式, 将原始数据存储存储在云端, 将数据索引存储在联盟区块链网络中, 以降低区块链存储负担和隐私泄露风险^[104]。

利用区块链技术对EHRs进行数据共享是推广EHRs研究的关键, 目前的技术普遍采用智能合约^[104]和问责机制^[105]等方式, 以提高数据共享的私密性, 如通过群签名智能合约实现匿名信息交换^[99], 从而增强EHRs数据在不同医疗机构之间的流动性, 实现医疗数据的便捷共享, 以防止数据共享过程中的隐私泄露。

基于区块链技术的EHRs系统可以方便地对患者医疗记录进行增、删、改、查和授权访问, 但是区块链技术在应用中还面临着一定的挑战, 如可扩展性和存储容量、缺乏社交技巧、缺乏普遍定义的标准。

三、政策与建议

纵观我国基于电子病历的研究现状, 不难发现, 在推动医疗智能化的过程中, 数据的整合、数据的利用和数据隐私问题仍是急需解决的关键问题, 为此, 本文提出如下建议:

1. 规范数据整合, 加大医疗信息数据库的建设力度。虽然我国的医疗信息化建设已在逐步普及,

但是由于城乡医疗资源的差异,医疗数据库的普及和受到限制^[23],数据质量不高,数据利用不足^[49]。医疗数据库的标准化建设将为医疗信息管理与利用提供助力,医疗机构应通过数据采集、样本处理及规范化存储,实现临床信息的数据整合、质量控制和数据服务的信息平台,通过医疗数据实现创新增值。

2. 打破“信息孤岛”,建立有效的EHRs数据共享与互认机制。医疗信息的共享将有助于慢性病患者^[106]和老年患者^[107]的长期治疗。目前,我国除少数大城市的大型医疗机构,大部分医疗机构之间的数据都是相互独立的^[108],各级各类医疗机构的医疗信息平台没有实现对接,数据难以共享,存在不同的医疗机构之间的化验结果互不相认等情况。为此,医疗机构应该整合医疗信息资源,统一数据标准,消除数据壁垒,落实数据的共享互认机制,改善资源之间的互通互联问题,提高数据的互操作性,杜绝“信息孤岛”困境,推进数据的整合利用^[109]。

3. 开放资源,建立标准的集成公开医疗数据集。对比发达国家的医疗信息管理现状,诸如美国、英国、丹麦等都有自己的集成的、公开的医疗数据库,供相关研究人员使用^[110]。目前我国的标准集成公开医疗数据库屈指可数,政府应该组织专业人员整合与建立针对特定疾病或人群的医疗数据集,支持行业领先企业或研究机构在医疗大数据领域的创新与应用研究,利用数据帮助研发人员解密医学规律、整合医学知识,实现跨学科的数据交互,为相关疾病研究提供资源,促进智慧医疗技术的进步。

4. 完善医疗信息安全防护体系,做好电子病历资源数据的隐私保护工作。电子病历包含患者的隐私信息,相关部门应强化数据安全意识,制定医疗数据采集、存储、传输、共享各个环节的流程规范^[111],明确行为边界和“禁区”。强化医疗信息系统的安全管理,完善数据监测和预警机制,制定医疗信息安全事件的应急措施,尽量避免可能出现的数据隐私风险^[112]。在数据存储中,需充分利用区块链在隐私保护方面的优势,与现有存储技术相配合,实现EHRs数据管理与使用过程的可溯源、可追踪、可把控。在数据共享中,可采用数据生成模型实现EHRs的可替换性,全方位保护患者的隐私。

5. 积极推动智慧医疗研究成果的转化与落地实施,实现产学研一体化。中科院健康电子研发中心

与深圳诺嘉公司共建健康大数据联合实验室^[50],将基于医疗大数据研究的新技术与新成果转化成为大众医疗服务的新应用与新产品,实现了个性化与社会化的健康管理。尽管智能算法在理论与仿真实验中被证明在提升诊疗水平方面效果显著,但是新技术、新方法的产业化尚未形成规模,尤其是在利用深度学习技术辅助临床决策中,神经网络的可解释性问题是目前急需攻克的难点,严重影响着智能策略的可信度和有效性。

四、结语

智慧医疗是我国《新一代人工智能发展规划》发展方向之一,基于智能技术的医疗,如深度学习、区块链等,已经成为当前智慧医疗的核心技术。电子病历的建设与发展为智能技术在临床医疗中的应用提供了可靠的支撑。尽管我国正在大力推动医疗信息化建设,但对于电子病历的规范化和标准化方面与其他发达国家还有一定的差距,在医疗数据的开放获取和数据共享方面存在挑战。尤其是在我国的医疗发展不平衡情况下,如农村落后于城市、贫困地区落后于发达地区,内地落后于沿海等现状,限制了医疗信息化水平的稳步提高,合理高效的推广与研究电子病历将提升现有的医疗信息水平。

参考文献

- [1] 田生湖,陈渝.我国电子健康档案研究进展与挑战[J].中华医院管理杂志,2021,37(1): 63-68.
- [2] SHICKEL B, TIGHE P J, BIHORAC A, et al. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis[J]. *IEEE Journal of Biomedical and Health Informatics*, 2018, 22(5): 1589-1604.
- [3] JABALI A K, WARIS A, KHAN D I, et al. Electronic health records: three decades of bibliometric research productivity analysis and some insights[J]. *Informatics in Medicine Unlocked*, 2022, 29: 100872.
- [4] GELLMAN M D. *Encyclopedia of behavioral medicine*[M]. New York: Springer Science Business Media, 2013.
- [5] 任冠华.电子健康档案概念解析及标准化现状分析[J].医学信息学,2015,36(1): 14-18.
- [6] GUNTER T D, TERRY N P. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions[J]. *Journal of Medical Internet Research*, 2005, 7(1): e3.
- [7] 孙玉珍,吴彧一,方毅.电子健康档案研究进展综

述[J]. 中国医疗设备, 2019, 34(5): 156-159.

[8] 糜泽花, 钱爱兵. 智慧医疗发展现状及趋势研究文献综述[J]. 中国全科医学, 2019, 22(3): 366-370.

[9] CERCHIONE R, CENTOBELLI P, RICCIO E, et al. Blockchain's coming to hospital to digitalize healthcare services: designing a distributed electronic health record ecosystem[J]. *Technovation*, 2022(2): 102480.

[10] SUN X, DOURI A, GULLIFORD M. Applying machine learning algorithms to electronic health records predicted pneumonia after respiratory tract infection[J]. *Journal of Clinical Epidemiology*, 2022, 145(5): 154-163.

[11] 张迪. 电子病历数据应用现状分析及思考[J]. 数字技术与应用, 2021, 39(1): 44-46.

[12] 穆芳洁. 国内外电子病历的发展概况及思考[J]. 中国病案, 2014(9): 40-42.

[13] 陆敏. 新医改形势下公立医院电子病历发展现状与展望[J]. 健康之路, 2014, 000(1): 30-30.

[14] SHIELDS M C, HORGAN C M, RITTER G A, et al. Use of electronic health information technology in a national sample of hospitals that provide specialty substance use care[J]. *Psychiatric Services: A Journal of The American Psychiatric Association*, 2022, 72(12): 1370-1376.

[15] 范阳东, 汪紫彤, 陈美雪. 发达国家和地区电子健康档案建设及对我国的启示[J]. 现代医院, 2021, 21(3): 470-473.

[16] BIRKLE C, PENDLEBURY D A, SCHNELL J, et al. Web of science as a data source for research on scientific and scholarly activity[J]. *Quantitative Science Studies*, 2020, 1(1): 363-376.

[17] HAGHANI M, BLIEMER M C J, GOERLANDT F, et al. The scientific literature on coronaviruses, COVID-19 and its associated safety-related research dimensions: a scientometric analysis and scoping review[J]. *Safety Science*, 2020, 129: 104806.

[18] YAN L, ZHANG H-T, GONCALVES J, et al. An interpretable mortality prediction model for COVID-19 patients[J]. *Nature Machine Intelligence*, 2020, 2(5): 283-288.

[19] TURATTO F, MAZZALAI E, PAGANO F, et al. A systematic review and bibliometric analysis of the scientific literature on the early phase of COVID-19 in Italy[J]. *Frontiers in Public Health*, 2021(6): 776.

[20] LATIF J, XIAO C, TU S, et al. Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: a complete review[J]. *IEEE Access*, 2020, 8: 150489-150513.

[21] 李桂萍. 医改下医院医疗档案管理的优化策略探讨[J]. 兰台世界, 2017(A01): 112-113.

[22] 赵统堂. 医院电子病历档案管理及其共享策略探讨[J]. 兰台内外, 2022(1): 34-36.

[23] 李阳阳. 医院电子病历档案信息共享管理现状与优

化策略[J]. 档案管理, 2022, 2: 86-87.

[24] 唐世芳, 倪芳, 郑川. 档案视角下电子病历管理的问题与对策[J]. 资源信息与工程, 2022, 37(1): 156-158.

[25] 霍添琪, 徐帆, 李红霞, 等. 基于SE-SBM模型的贵州省三级医院电子病历系统建设效率评价[J]. 中国卫生质量管理, 2022, 29(1): 16-20.

[26] 吴敏, 甄天民, 谷景亮, 等. 健康医疗大数据国内外发展及在卫生决策支持中的应用展望[J]. 卫生软科学, 2019, 33(2): 76-79.

[27] 张舒薇. 大数据时代背景下的医疗档案管理的思考[J]. 数码设计:下, 2019(4): 142-143.

[28] 柴凯杰, 丁有伟, 胡孔法. 基于区块链的中医电子病历高效查询方法研究[J]. 软件导刊, 2021, 20(12): 9-14.

[29] 曹艳林, 王将军, 陈璞, 等. 人工智能对医疗服务的机遇与挑战[J]. 中国医院, 2018, 22(6): 25-28.

[30] SHI S, HE D, LI L, et al. Applications of blockchain in ensuring the security and privacy of electronic health record systems: a survey[J]. *Computers & Security*, 2020, 97: 101966.

[31] SI Y, DU J, LI Z, et al. Deep representation learning of patient data from electronic health records (EHR): a systematic review[J]. *Journal of Biomedical Informatics*, 2021, 115: 103671.

[32] TAXTER A J, NATTER M D. Using the electronic health record to enhance care in pediatric rheumatology[J]. *Rheumatic Disease Clinics of North America*, 2022, 48(1): 245-258.

[33] XIE F, YUAN H, NING Y, et al. Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies[J]. *Journal of Biomedical Informatics*, 2022, 126: 103980.

[34] SERWER G A. Data management and integration with electronic health record systems[J]. *Cardiac Electrophysiology Clinics*, 2021, 13(3): 473-481.

[35] DENDERE R, SAMADBEIK M, JANDA M. The impact on health outcomes of implementing electronic health records to support the care of older people in residential aged care: a scoping review[J]. *International Journal of Medical Informatics*, 2021, 151: 104471.

[36] CARTER A B, ABRUZZO L V, HIRSCHHORN J W, et al. Electronic health records and genomics[J]. *The Journal of Molecular Diagnostics*, 2022, 24(1): 1-17.

[37] 熊兴江, 马敬东, 徐承中. 电子健康档案数据质量评估与治理的综述研究[J]. 中国卫生信息管理杂志, 2018, 15(6): 637-642.

[38] 贾峥, 宗瑞杰, 段会龙, 等. 基于电子病历的患者相似性分析综述[J]. 中国生物医学工程学报, 2018, 37(3): 353-366.

[39] 何霆, 吴雅婷, 王华珍, 等. 基于EHR的医疗知识图谱研究与应用综述[J]. 哈尔滨工业大学学报, 2018, 50(11): 137-144.

- [40] 周飞虎. 电子病历数据库脓毒症病例筛选方法综述[J]. 解放军医学院学报, 2020(9): 918-921.
- [41] 杨飞洪, 张宇, 覃露, 等. 中文电子病历的命名实体识别研究进展[J]. 中国数字医学, 2020, 15(2): 9-12.
- [42] 赵艳莉, 陈龄, 岳冀蓉. 基于电子病历识别谵妄的研究现状与展望[J]. 生物医学工程学杂志, 2020, 37(1): 185-188.
- [43] 邱硕, 丁润宇, 柳亚男, 等. 基于区块链技术的电子医疗数据共享研究综述[J]. 电脑知识与技术:学术版, 2021, 17(34): 7-9.
- [44] 吴宗友, 白昆龙, 杨林蕊, 等. 电子病历文本挖掘研究综述[J]. 计算机研究与发展, 2021, 58(3): 513-527.
- [45] 齐虹. 国外电子健康记录信息语义整合研究述评与展望[J]. 情报科学, 2021, 39(6): 177-184.
- [46] 刘彤彤, 杨环, 西永明, 等. 机器学习在脊柱疾病智能诊治中的应用综述[J]. 计算机科学, 2021, 48(S02): 597-607.
- [47] JOHNSON A E W, POLLARD T J, SHEN L, et al. MIMIC-III, a freely accessible critical care database[J]. *Scientific Data*, 2016, 3(1): 160035.
- [48] 陈静, 付玉敏. 基于云存储的电子健康记录隐私保护策略探讨[J]. 医院管理论坛, 2021, 38(5): 13-16.
- [49] 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.
- [50] 史钰斐. 健康医疗大数据发展应用的思考[J]. 无线互联科技, 2021, 18(19): 94-95.
- [51] XU Y, BISWAL S, DESHPANDE S R, et al. RAIM: recurrent attentive and intensive model of multimodal patient monitoring data[C]. London: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 2565-2573.
- [52] YIN K, QIAN D, CHEUNG W K, et al. Learning phenotypes and dynamic patient representations via RNN regularized collective non-negative tensor factorization[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33: 1246-1253.
- [53] LEI L, ZHOU Y, ZHAI J, et al. An effective patient representation learning for time-series prediction tasks based on EHRs[C]. Madrid: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018: 885-892.
- [54] WANG W W, LI H, CUI L, et al. Predicting clinical visits using recurrent neural networks and demographic information[C]. Nanjing: The 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD 2018), 2018: 785-789.
- [55] WU S, LIU S, SOHN S, et al. Modeling asynchronous event sequences with RNNs[J]. *Journal of Biomedical Informatics*, 2018, 83: 167-177.
- [56] JIN B, CHE C, LIU Z, et al. Predicting the risk of heart failure with EHR sequential data modeling[J]. *IEEE Access*, 2018, 6: 9256-9261.
- [57] DEVLIN J, CHANG M-W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2022-05-10]. https://www.researchgate.net/publication/328230984_BERT_Pre-training_of_Deep_Bidirectional_Transformers_for_Language_Understanding.
- [58] LEE J, YOON W, KIM S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J]. *Bioinformatics*, 2019, 36(4): 1234-1240.
- [59] 陈德光, 马金林, 马自萍, 等. 自然语言处理预训练技术综述[J]. 计算机科学与探索, 2021, 15(8): 1359-1389.
- [60] CHENG Y, WANG F, ZHANG P, et al. Risk prediction with electronic health records: a deep learning approach[C]. Miami, FL: the 2016 SIAM International Conference on Data Mining, 2016: 432-440.
- [61] WANYAN T, HONARVAR H, Azad A, et al. Deep learning with heterogeneous graph embeddings for mortality prediction from electronic health records[J]. *Data Intelligence*, 2021, 3(3): 329-339.
- [62] LI R, HU B, LIU F, et al. Detection of bleeding events in electronic health record notes using convolutional neural network models enhanced with recurrent neural network autoencoders: deep learning approach[J]. *JMIR Medical Informatics*, 2019, 7(1): e10788.
- [63] MCGUINNESS J E, LEE J, ANURAJ A, et al. Abstract PS5-21: Use of a novel convolutional neural network (CNN)-based mammographic evaluation to assess response to adjuvant endocrine therapy[J]. *Cancer Research*, 2021, 81(4): PS5-21.
- [64] LIU G, LIAO Y, WANG F, et al. Medical-Vibert: medical visual language BERT for COVID-19 CT report generation with alternate learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(9): 3786-3797.
- [65] TASHITA A, MORITA K, NII M, et al. Automated estimation of mTS score in hand joint X-ray image using machine learning[C]. Himeji: 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT), 2017: 1-5.
- [66] SHIRI I, AKHAVANALLAF A, SANAAAT A, et al. Deep residual neural network-based standard CT estimation from ultra-low dose CT imaging for COVID-19 patients[C]. Boston: 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2020: 1-3.
- [67] 林玉萍, 龙红, 李彪, 等. 基于医学影像和病历文本的甲状腺多模态语料库构建与应用[J]. 西北大学学报:自然科学版, 2021, 51(2): 198-206.
- [68] KIM Y. Convolutional neural networks for sentence classification[C]. Doha: The 2014 Conference on Empirical

Methods in Natural Language Processing (EMNLP), 2014: 1746–1751.

[69] 李小整, 王华珍, 熊英杰, 等. 卷积神经网络模型在儿科疾病预测中的应用[J]. 中国数字医学, 2018, 13(10): 11–13.

[70] YANG Z, HUANG Y, JIANG Y, et al. Clinical assistant diagnosis for electronic medical record based on convolutional neural network[J]. *Scientific Reports*, 2018, 8(1): 6329.

[71] ESTEVA A, ROBICQUET A, RAMSUNDAR B, et al. A guide to deep learning in healthcare[J]. *Nature Medicine*, 2019, 25(1): 24–29.

[72] KINGMA D P, WELING M. Auto-encoding variational bayes[EB/OL]. [2022-05-10]. <https://arxiv.org/abs/1312.6114?context=cs.LG>.

[73] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[EB/OL]. [2022-05-10]. <https://arxiv.org/abs/1406.2661?context=cs>.

[74] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[J/OL]. [2022-05-10]. https://www.researchgate.net/publication/313044478_Wasserstein_GAN.

[75] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[C]. San Juan. 4th International Conference on Learning Representations, 2016.

[76] ZHOU Z, LIANG J, SONG Y, et al. Lipschitz generative adversarial nets[EB/OL]. [2022-05-10]. <https://arxiv.org/abs/1902.05687v2>.

[77] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks[C]. Venice: IEEE International Conference on Computer Vision, 2017: 2813–2821.

[78] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier GANs[C]. NSW: Proceedings of the 34th International Conference on Machine Learning, 2017: 2642–2651.

[79] WENG S W, AMER M, MAUL T, et al. Conditional generative adversarial networks for data augmentation in Breast cancer classification[C]//International Conference on Soft Computing and Data Mining, London: Springer, Cham, 2020.

[80] ZHANG X, WANG Z, LIU D, et al. DADA: deep adversarial data augmentation for extremely low data regime classification[C]. Brighton: 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019: 2807–2811.

[81] LEE M B, KIM Y H, PARK K R. Conditional generative adversarial network- based data augmentation for enhancement of Iris recognition accuracy[J]. *IEEE Access*, 2019, 7: 122134–122152.

[82] YOON J, JORDON J, SCHAAR M V D. GAIN: missing data imputation using generative adversarial nets[C].

Sweden: The 35th International Conference on Machine Learning, 2018: 5675–5684.

[83] 梁俊杰, 韦舰晶, 蒋正锋. 生成对抗网络GAN综述[J]. 计算机科学与探索, 2020, 14(1): 1–17.

[84] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529–533.

[85] HASSELT H V, GUEZ A, SILVER D. Deep reinforcement learning with double q-learning[C]. Phoenix: The Thirtieth AAAI Conference on Artificial Intelligence, 2016: 2094–2100.

[86] MALOF J M, GAWEDA A E. Optimizing drug therapy with reinforcement learning: the case of anemia management[C]. California: The 2011 International Joint Conference on Neural Networks, 2011: 2088–2092.

[87] WANG Y, WU P, LIU Y, et al. Learning optimal individualized treatment rules from electronic health record data[C]. Chicago: 2016 IEEE International Conference on Healthcare Informatics (ICHI), 2016: 65–71.

[88] NEMATI S, GHASSEMI M M, CLIFFORD G D. Optimal medication dosing from suboptimal clinical examples: a deep reinforcement learning approach[C]. Orlando: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016: 2978–2981.

[89] KOMOROWSKI M, CELI L A, BADAWI O, et al. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care[J]. *Nature Medicine*, 2018, 24(11): 1716–1720.

[90] LOPEZ-MARTINEZ D, ESCHENFELDT P, OSTVAR S, et al. Deep reinforcement learning for optimal critical care pain management with morphine using dueling double-deep q networks[C]. Berlin: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019: 3960–3963.

[91] ZHENG H, RYZHOV I O, XIE W, et al. Personalized multimorbidity management for patients with type 2 diabetes using reinforcement learning of electronic health records[J]. *Drugs*, 2021, 81(4): 471–482.

[92] WANG L, ZHANG W, HE X, et al. Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation[C]. London: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 2447–2456.

[93] LI T, WANG Z, LU W, et al. Electronic health records based reinforcement learning for treatment optimizing[J]. *Information Systems*, 2022, 104: 101878.

[94] KAO H-C, TANG K-F, CHANG E Y. Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning[C]. New Orleans, Louisiana: The Thirty-Second AAAI Conference on Artificial Intelligence, 2018:

2305-2313.

[95] KHAJURIA R, SARWAR A. Reinforcement learning in medical diagnosis: an overview[J]. *Recent Innovations in Computing*, 2022, 832: 179-188.

[96] FANG H S A, TAN T H, TAN Y F C, ET AL. Blockchain personal health records: systematic review[J]. *Journal of Medical Internet Research*, 2021, 23(4): e25094.

[97] SHAHNAZ A, QAMAR U, KHALID A. Using blockchain for electronic health records[J]. *IEEE Access*, 2019, 7: 147782-147795.

[98] 林孟晨, 冯勇, 付晓东. 一种基于联盟区块链的电子医疗记录安全共享模型[J]. *小型微型计算机系统*, 2021, 42(10): 2161-2166.

[99] GUO R, SHI H, ZHAO Q, et al. Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems[J]. *IEEE Access*, 2018, 6: 11676-11686.

[100] AL O A, RAHMAN M S, BASU A, et al. Medibchain: a blockchain based privacy preserving platform for healthcare data[C]. Guangzhou: International conference on security, privacy and anonymity in computation, communication and storage, 2017: 534-543.

[101] ZHENG X, MUKKAMALA R R, VATRAPU R, et al. Blockchain-based personal health data sharing system using cloud storage[C]. Ostrava: 2018 IEEE 20th international conference on E-health networking, applications and services (Healthcom), 2018: 1-6.

[102] ZHAO H, ZHANG Y, PENG Y, et al. Lightweight

backup and efficient recovery scheme for health blockchain keys[C]. Bangkok: 2017 IEEE 13th international symposium on autonomous decentralized system (ISADS), 2017: 229-234.

[103] CAPECE G, LORENZI F. Blockchain and healthcare: opportunities and prospects for the EHR[J]. *Sustainability*, 2020, 12(22): 9693.

[104] LIU J, LI X, YE L, et al. BPDS: a blockchain based privacy-preserving data sharing for electronic medical records[C]. Abu Dhabi: 2018 IEEE Global Communications Conference (GLOBECOM), 2018: 1-6.

[105] XIA Q, SIFAH E B, SMAHI A, et al. BBDS: blockchain-based data sharing for electronic medical records in cloud environments[J]. *Information*, 2017, 8(2): 44.

[106] 许薇娜. 电子健康档案对慢性病目标管理的意义[J]. *黑龙江档案*, 2021(4): 34-35.

[107] 邹钰莹, 娄峥嵘. 中央层面养老服务政策内容量化评价——基于PMC指数模型的分析[J]. *电子科技大学学报(社科版)*, 2020, 22(3): 68-76.

[108] 敖虎山. 打通“数据孤岛”推动建立医疗行业“数据银行”[J]. *中国医疗保险*, 2022(3): 18-19.

[109] 舒婷, 赵韡, 刘海一. 2020年我国医院电子病历系统应用水平分析[J]. *中国卫生质量管理*, 2022, 29(1): 8-10.

[110] 舒影岚, 陈艳萍, 吉臻宇, 等. 健康医疗大数据研究进展[J]. *中国医学装备*, 2019, 16(1): 143-147.

[111] 裴莹蕾, 王娅. 构建区域医疗大数据治理体系的实践探索[J]. *卫生经济研究*, 2022, 39(5): 71-74+78.

[112] 吴凌放. 刍议医疗健康大数据对外合作中的法律风险[J]. *中国卫生法制*, 2022, 30(3): 101-103.

编辑 邓婧