中华医学会第二十六次全国医学信息学术会议

论文汇编

2020 年 10 月 22-25 日

线上·会议

目 录

专题发言

S-001 “智慧医疗：大健康 新生态”
---------- 董家鸿 1
S-002 完善新突发传染病早期症侯群智能化监测与预警
---------- 马家奇 1
S-003 互联网+医疗信息化发展与思考
---------- 徐向东 1
S-004 可视化智慧医疗创新实践
---------- 王秋萍 1
S-005 以人工智能重塑医疗：平安医疗科技的实践和探索
---------- 谢国彤 2
S-006 物联网在医院应用实践分享
---------- 全宇 2
S-007 现代智慧医院的建设与应用
---------- 李振叶 2
S-008 电子病历系统应用水平分级评价工作经验分享
---------- 刘新奎 2
S-009 医技检查智能化自动预约的实现与应用
---------- 孙凌毅 3
S-010 江苏省人民医院智慧门诊助力疫情防控
---------- 陈彦 3
S-011 精准医学大数据与全民健康
---------- 方向东 3
S-012 医疗大数据的欺骗性及其对策探讨
---------- 朱卫国 3
S-013 人工智能在脑血管病的临床应用研究
---------- 李子孝 4
S-014 面向心血管疾病的临床辅助决策支持
---------- 黄正行 4
S-015 High-throughput relation extraction algorithm development associating knowledge articles and electronic health records
---------- 俞声 4
S-016 医学数据挖掘评测论文报告与活动总结
---------- 李姣 4
S-017 图情档学科建设和学科发展
---------- 张久珍 5

S-018 从信息素养教育到泛信息素养教育
---------- 初景利 5
S-019 用户信息行为分析研究
---------- 曹锦丹 5
S-020 面向医学情报发现的智能文献计量
---------- 张嶷 5
S-021 医学信息资源建设策略
---------- 任慧玲 6
S-022 我国医学信息教育命运共同体构建模式和实施路径研究
---------- 王伟 6
S-023 我国医学信息学领域中的“群”现象研究
---------- 李后卿 6
S-024 国际医学信息学教育知识流动规律研究
---------- 胡德华 6
S-025 国际线上教学研究热点与前沿分析
---------- 崔雷 7
S-026 大数据背景下医学信息学研究生培养的思考
---------- 张云秋 7
S-027 医学信息学开放课程自主学习平台研发以及课程联盟建设
---------- 赵文龙 7
S-028 基于知识图谱的药品不良反应预警
---------- 蔡永铭 7
S-029 基于 5G 技术的医学信息素养教育面临的机遇、挑战与对策
---------- 林宏伟 8
S-030 互联网医院在线模式在新冠疫情防控中的应用实践
---------- 曾国军 8
S-031 网络社区用户健康信息推荐模型研究
---------- 张晗 8
S-032 基于人工智能技术的糖尿病及其并发症防控研究
---------- 吴辉群 8
S-033 5G 时代医院创伤调度监控指挥中心信息化建设
---------- 郭建军 9

论文报告

OR-001 疫情防控期间移动医疗会诊系统的建设与应用
---------- 杨扬,李郁鸿 10
OR-002 浅谈病案管理信息化建设的必要性
---------- 魏梦迪 15
OR-003 基于 DTFS 的 PACS 存储系统设计
----------朱彦霞,陈益洲,华南等 20

OR-004 基于 LDA 模型的新冠肺炎网络健康社区信息需求分析
----------安欣宇,陈育新,张晗 29
OR-005 鼻咽癌放疗后大出血患者智能预警系统的研究与设计
---------- 葛晓伟,程铭 29
OR-006 基于 DGCNN 和注意力机制的中文电子病历命名实体识别研究
----------胡海洋,赵从朴,姜会珍等 35
OR-007 Association Analyses Reveal Comorbidity Patterns among Older Lung Cancer Patients in China
----------冯佳,穆晓敏,王伟等 42
OR-008 基于深度残差网络的银屑病分类诊断模型研究
---------- 李鹏,罗爱静 43
OR-009 2 型糖尿病患者早期眼底图像改变的临床预测模型研究
---------- 邵家辉 56
OR-010 医院图书馆微信公众平台用户信息采纳行为影响因素及实证分析
----------李日,高歌,伦志军 57
OR-011 互联网医疗咨询平台使用意愿研究
---------- 童俞嘉,翟兴 64
OR-012 基于结构熵权法的医院图书馆数字资源评价
----------庞兴梅,田明慧,庞兴峰 75
OR-013 基于 SPO 语义网络的民族医药知识发现
----------杨雪梅,王雪,关陟昊等 82
OR-014 Protégé在构建中医药本体中的运用
----------杨郑子衿,徐倩,罗爱静 91
OR-015 医学院校“互联网+”创新创业教育实践平台构建
---------- 李学沧 101
OR-016 “人工智能”背景下基于 CDIO 模式导师制的医学信息学人才培养探索与实践
---------- 吴辉群 114
OR-017 医工融合专业人才培养质量的隐性影响因素探析——基于扎根理论的研究
----------王辅之,杨枢,张浩等 119
OR-018 基于质性分析的大学生健康信息获取行为引动研究
---------- 祝帆帆,桂晓苗 126
OR-019 精神类疾病在线健康社区用户性别差异研究——以微医网为例
---------- 王安娜,向菲 138
OR-020 疫情期间医学文献检索课在线授课效果调查与分析
----------兰雪,张晗,刘春鹤 149
OR-021 5G 智慧移动医疗研究
---------- 刘思远,胡德华 157
OR-022 新冠肺炎疫情下高校图书馆面向学生的学习科研信息服务研究
----------赵嘉玮,林鑫,郭进京等 162
OR-023 签约家庭医生对河南农村居民健康素养水平的影响
----------李梦蕾,张帆,张军亮等 171
OR-024 农村中年慢性病患者在线健康信息获取障碍研究
----------王卓鑫,吕合凯,孙伟伟等 176

OR-025 医联体内医疗信息共享的关键影响因素研究
---------- 李雨晨,刘智勇 186
OR-026 安徽医科大学信息管理与信息系统专业教学改革探讨——基于毕业生就业情况的问卷调查分析
---------- 丁晓,宋国强 199
OR-027 医学信息教育智慧云平台构建研究
---------- 徐义涛,文庭孝 204
OR-028 基于 WoS 的医学信息学领域研究现状分析
----------孙晶晶,张帆,张军亮等 212
OR-029 基于结构方程的河南省 5 家中医院门诊和住院患者满意度的比较分析
----------王晶晶,吕晖,邢晓琳等 219
OR-030 在线健康信息质量评价工具探析
----------吕合凯,王卓鑫,孙伟伟等 226
OR-031 电子烟视频传播效果影响因素研究——基于哔哩哔哩平台
---------- 谢之英,向菲 234
OR-032 河南省中医院患者满意度调查及影响因素分析
----------邢晓琳,吕晖,王晶晶等 244
OR-033 移动医疗 APP 用户隐私关注影响因素研究
----------袁珊珊,罗亚湄,周建等 248

全文刊登

PU-001 新时代下高校图书馆面对重大公共卫生事件应急响应及疫情常态化下服务策略研究
---------- 许丹 262
PU-002 健康医疗大数据多中心研究现状与发展趋势
----------刘迷迷,李丽婷,周毅 274
PU-003 基于妇幼数据中心构建信息安全保障体系的设计研究
---------- 王辉 278
PU-004 探索医院建设“互联网+护理服务”可行性分析
---------- 牛亚军 284
PU-005 腹部 CT 图像肾脏分割方法综述
---------- 承垠林,周毅 289
PU-006 癫痫脑电自动发作预测研究综述
----------马梦楠,张亮军,承垠林等 293
PU-007 基于 ARIMA-LSTM 组合模型的月门诊量预测精度研究
---------- 李琳,周毅 298
PU-008 基于独立循环神经网络改进残差架构的癫痫脑电分类研究
----------马梦楠,承垠林,韦晓燕等 304
PU-009 基于赫斯特指数的循环神经网络癫痫脑电识别研究
----------马梦楠,张亮军,承垠林等 317
PU-010 基于深度学习的癫痫患者脑电图分类与预测任务研究
----------马梦楠,张亮军,承垠林等 323

PU-011 关于防疫期间医院计算机信息系统的讨论
---------- 王东飞 330
PU-012 医学领域自然语言处理的研究热点分析——基于 CiteSpace 的知识图谱分析
----------陈婕卿,张锋,曾可等 334
PU-013 基于医联体的医疗影像云平台设计与应用
---------- 郭金鸽 343
PU-014 基于文献计量的中国临床医学院士科研产出行为分析
----------杨辰毓妍,安新颖,蔡荣 346
PU-015 近 5 年国际医学信息学计量研究
---------- 单治易,安新颖 354
PU-016 基于 Altmetrics 分析“新冠肺炎”学术成果社会影响力
----------蔡荣,安新颖,杨辰毓妍 355
PU-017 基于 FEARS 的屈螺酮炔雌醇药物不良事件信号挖掘研究
----------詹阳洋,胥学莹,陈力等 361
PU-018 面向航空医学科研团队的嵌入式学科服务
----------于丽,邓略,张莉莉等 369
PU-019 健康医疗大数据治理体系及方法思考
----------张睿,甄天民,谷景亮等 374
PU-020 我国全科医生资源配置现状及公平性研究
---------- 温楠,程艳敏 375
PU-021 基于 5G 技术的医联体信息化共享模式探研
----------喻平,李绍杰,廖廓等 380
PU-022 基于美国 FAERS 数据库的阿特珠单抗不良事件信号挖掘研究
----------白雪霏,陈力,钟薇等 385
PU-023 医联体模式下分级诊疗建设思考
---------- 王琳琳 385
PU-024 新冠肺炎疫情影响下高校医学图书馆职能建设的思考—以武汉大学图书馆医学分馆为例
---------- 王俊 386
PU-025 基于微信小程序护理员管理平台的开发与设计
----------李俊,郭秋月,贺庆等 390
PU-026 国外全科医生发展经验对我国全科医生发展的启示
---------- 王凡,程艳敏 390
PU-027 数据分析和患者细分在临床科室管理中的应用研究
---------- 牛承志,赵丹 394
PU-028 健康信息素养在新型冠状病毒肺炎疫情防控中的作用
---------- 黄尤江,郑利荣, 贺莲等 401
PU-029 在线健康社区知识交流效率研究
---------- 张亚飞,贺培凤 404
PU-030 战储卫生器材信息系统的设计研究
---------- 杨举,胡红波 409
PU-031 互联网医院实现医保在线支付的设计与实践
----------曹凯迪,张小亮,朱佳等 414

PU-032 国内外医疗数据可视化研究的知识图谱比较分析
----------刘春鹤,张晗,赵玉虹 414
PU-033 智慧教室在高等学校教学中的应用研究
---------- 杨苏彬,朴杰 423
PU-034 代写代发论文的识别、防范及学术失范的治理
----------杨春华,姚敏,刘娜等 425
PU-035 基于 5G 和人工智能技术的院前院内急救管理平台建设
---------- 张小亮 426
PU-036 基于微信小程序的移动医疗数字化多媒体/短视频医患沟通系统 “摇摇灵”的开发与实现
----------郭秋月,李俊,贺庆等 430
PU-037 面向电子病历的胃癌术后复发风险预测研究
---------- 李郁鸿,程铭 431
PU-038 中文在线健康网站质量评估
---------- 孙伟伟,赵斌 431
PU-039 护理人员护理信息系统使用行为与工作绩效关系研究
----------刘智勇,春兰,周建等 437
PU-040 基于电力载波技术的智慧医院配电网监测系统设计
---------- 贾志刚 445
PU-041 基于大数据背景下的病案统计管理分析
---------- 郭斌 445
PU-042 医疗健康中的药物物联网与大数据
---------- 徐勐 449
PU-043 基于 Spark 的电子病历多维检索系统研究与应用
---------- 李育申,葛晓伟 449
PU-044 双向聚类方法在医学文本挖掘中的实现过程及应用意义
----------杨颖,李子政,许丹等 450
PU-045 基于本体的化学药物知识表示模型构建
----------秦璐,徐倩,罗爱静 455
PU-046 从宏观结构及微观分布揭示全球免疫学科战略情报
----------杨颖,陈斯斯,徐爽等 463
PU-047 Data Mining of The Knowledge Structure in Medical Informatics by Co-citation Biclustering Analysis and Strategic Diagram
----------杨颖,徐爽,韩爽等 475
PU-048 基于 FFCA 的电子病历命名实体领域本体构建研究
---------- 王凯 475
PU-049 近 10 年来神经网络辅助诊断乳腺肿瘤技术研究的可视化分析
---------- 邱煜炎,吴福生 484
PU-050 基于专利计量的医院专利管理对策分析
----------吕鸣,邹立君,侯胜超 485
PU-051 高校规范化使用 ESI、Incites 指标助力双一流学科建设的知识情报服务实证研究
----------杨颖,韩爽,许丹等 485

PU-052 医疗健康中的大数据分析
---------- 张伟祎 498
PU-053 医院图书情报服务中区块链技术的应用
----------刘明信,崔婷婷,张延一等 499
PU-054 基于聚类分析的疾病与基因关联关系研究
----------任雪菁,安新颖,范少萍 502
PU-055 基于 FDA 不良事件报告系统的二甲双胍不良事件信号研究
---------- 张旭 502
PU-056 基于大数据的新型冠状病毒肺炎疫情防控策略的探讨
----------秦三利,齐江华,张仲男等 513
PU-057 5G+智慧医疗应用场景探索与研究
---------- 陈军 517
PU-058 数据挖掘在病案信息中应用
---------- 王春杰 517
PU-059 区域健康医疗大数据应用发展政策内容分析
----------牟燕,吴敏,何有琴 521
PU-060 中文临床自然语言理解评估基准构建
----------杨飞洪,王序文,李姣 531
PU-061 美国 9 所医学院（校）图书馆信息素养教育现状调查与分析
---------- 刘靓靓 532
PU-062 基于舆情视角下的重大突发公共卫生事件引发的信息疫情防控模式研究
——以新型冠状病毒肺炎疫情舆情防控为例
---------- 刘帅 545
PU-063 基于 CiteSpace 的我国订单定向医学生研究热点和趋势分析
----------魏艳君,牟燕,宋奎勐 550
PU-064 贵阳市第一人民医院基于电子签名的医院无纸化应用探讨
---------- 邵建华 550
PU-065 山东省病原微生物实验室及实验活动备案管理系统的研发与应用
---------- 席敏,段永璇 556
PU-066 结核病专项管理信息系统应用研究
---------- 张现辉,宋浩 561
PU-067 区块链技术在医疗卫生领域现状研究
---------- 钱旦敏 567
PU-068 基于 GM（1,1）灰色预测模型的山东省公立医院医疗费用分析及预测
----------戴涛明,宋奎勐,甄天民 577
PU-069 基于真实世界数据的奥利司他不良事件信号挖掘研究与分析
----------袁婷,王丽,张杰等 577
PU-070 基于 VGG16 模型的图像情感分类方法
----------王瑶,马梦楠,周毅 586
PU-071 国际生物信息学领域发展脉络研究及热点探析
---------- 许丹 590

PU-072 基于本体的膀胱癌临床诊疗决策支持知识库构建
----------李星雨,李锴文,林天歆等 600
PU-073 基于人工智能的肾结石评估模型研究
----------李星雨,赖聪,承垠林等 607
PU-074 基于电子病历的膀胱癌术后复发风险预测综述
----------毛钤镶,刘迷迷,周毅 613
PU-075 自然语言处理技术在临床上的应用
----------毛钤镶,刘迷迷,周毅 616
PU-076 基于 EEG 脑网络的癫痫发作状态转变研究
----------韦晓燕,马梦楠,陈子怡等 620
PU-077 基于 EEG-FMRI 的多模态信息融合技术进展
----------韦晓燕,马梦楠,陈子怡等 626
PU-078 基于注意力 U 型网络的眼底图像硬性渗出物检测
----------张亮军,承垠林,马梦楠等 631
PU-079 基于距离正则化水平集演化的视盘分割
----------张亮军,承垠林,马梦楠等 637
PU-080 新一代信息技术在 2019 新型冠状病毒肺炎防控中的研究与应用
----------刘迷迷,李星雨,周毅 640

摘要刊登

PA-001 健康大数据与精准健康管理
---------- 刘福荣 635

嘉和美康
GOODWILL

15年积淀，专注于医疗信息管理软件研发、生产、销售、集成及服务

12个分公司/办事处/合作机构遍布全国

6大产品体系，为医疗机构临床、科研、教学、运营等业务提供技术支撑

1300+
专业技术团队,全面支持用户需求

1400+
用户案例，行业经验累积丰富

更多信息，敬请登录
www.goodwillcis.com

云孚科技
全球领先的语义技术服务商

医学知识图谱

从药品说明书等医学文本中
抽取医学实体及实体间关系
实现医学知识的抽取

基于抽取知识构建医学知识图谱
实现结构化的医学知识管理能力
随时掌握流行医学实体关系变化
医学知识库增量更新，持续增长

核心产品

语义云

全栈自主研发
多项技术世界第一
便捷私有化部署

知识图谱

深度行业优化
自动挖掘知识
独家事理图谱

智能搜索

支持各类文档格式
精准理解用户意图
语义检索相关文档

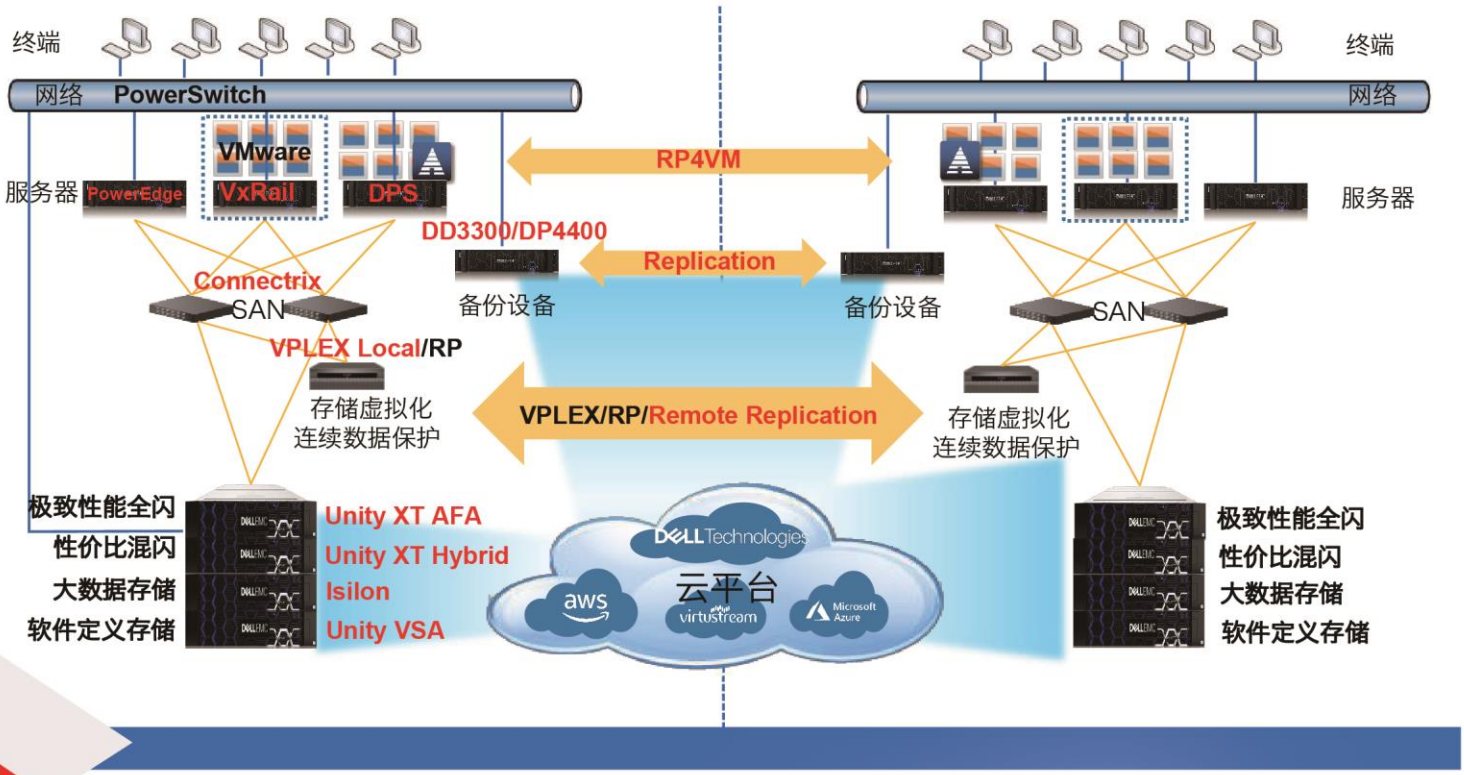
智能问答

个性化闲聊
多轮对话
融合多种问答

云孚科技（北京）有限公司
bd@yunfutech.com
北京市海淀区中关村国际孵化园D座304

长虹佳华
CHANGHONG IT

DELLEMC医疗行业IT基础架构整体解决方案



长虹佳华(股份代号3991.HK)

- 国企控股，香港上市
- 定位于新型的ICT综合服务商，为合作伙伴提供高效、专业的ICT综合服务
- 整合、优化国内外资源的专业ICT解决方案服务商与ICT产品分销商
- 专业智能终端产品生产商和服务商
- 建设完善MSP综合服务平台，提供云生态的行业综合服务
- 构建覆盖全国的开放、合作、共赢的B2B区域分销新生态

做**帮助成长 支持成功**的好伙伴

长虹佳华官方微博

长虹佳华官方微信服务号

专题发言

S-001

“智慧医疗：大健康 新生态”

董家鸿
北京清华长庚医院院长

S-002

完善新突发传染病早期症侯群智能化监测与预警

马家奇
中国疾病预防控制中心首席科学家

S-003

互联网+医疗信息化发展与思考

徐向东
国家卫生健康委统计信息中心

S-004

可视化智慧医疗创新实践

王秋萍
国家呼吸医学中心

S-005

以人工智能重塑医疗：平安医疗科技的实践和探索

谢国彤
平安科技

S-006

物联网在医院应用实践分享

全宇
中国医科大学附属盛京医院

S-007

现代智慧医院的建设与应用

李振叶
宁夏医科大学总医院

S-008

电子病历系统应用水平分级评价工作经验分享

刘新奎
郑州大学第一附属医院

S-009

医技检查智能化自动预约的实现与应用

孙凌毅
昆明医科大学第一附属医院

S-010

江苏省人民医院智慧门诊助力疫情防控

陈彦
江苏省人民医院（南京医科大学第一附属医院）

S-011

精准医学大数据与全民健康

方向东
中国科学院北京基因组研究所（国家生物信息中心）

S-012

医疗大数据的欺骗性及其对策探讨

朱卫国
中国医学科学院北京协和医院

S-013

人工智能在脑血管病的临床应用研究

李子孝
首都医科大学附属北京天坛医院

S-014

面向心血管疾病的临床辅助决策支持

黄正行
浙江大学

S-015

High-throughput relation extraction algorithm development associating knowledge articles and electronic health records

俞声
清华大学

S-016

医学数据挖掘评测论文报告与活动总结

李姣
中国医学科学院医学信息研究所

S-017

图情档学科建设和学科发展

张久珍
北京大学信息资源管理学

S-018

从信息素养教育到泛信息素养教育

初景利
中国科学院文献情报中心

S-019

用户信息行为分析研究

曹锦丹
吉林大学公共卫生学院

S-020

面向医学情报发现的智能文献计量

张嶷
悉尼科技大学澳大利亚人工智能研究所

S-021

医学信息资源建设策略

任慧玲
中国医学科学院信息所图书馆

S-022

我国医学信息教育命运共同体构建模式和实施路径研究

王伟
吉林大学公共卫生学院

S-023

我国医学信息学领域中的“群”现象研究

李后卿
中南大学湘雅医学院

S-024

国际医学信息学教育知识流动规律研究

胡德华
中南大学生命科学学院

S-025

国际线上教学研究热点与前沿分析

崔雷
中国医科大学

S-026

大数据背景下医学信息学研究生培养的思考

张云秋
吉林大学公共卫生学院

S-027

医学信息学开放课程自主学习平台研发以及课程联盟建设

赵文龙
重庆医科大学

S-028

基于知识图谱的药品不良反应预警

蔡永铭
广东药学院

S-029

基于 5G 技术的医学信息素养教育面临的机遇、挑战与对策

林宏伟
湖北医药学院

S-030

互联网医院在线模式在新冠疫情防控中的应用实践

曾国军
四川大学华西医院

S-031

网络社区用户健康信息推荐模型研究

张晗
中国医科大学

S-032

基于人工智能技术的糖尿病及其并发症防控研究

吴辉群
南通大学

S-033

5G 时代医院创伤调度监控指挥中心信息化建设

郭建军
江苏省人民医院（南京医科大学第一附属医院）

论文报告

OR-001

疫情防控期间移动医疗会诊系统的建设与应用

杨扬、李郁鸿
郑州大学第一附属医院

目的 新型冠状病毒感染的肺炎患者病情复杂、变化快、危重患者较多，患者病情牵涉多科室、多学科协作，为对患者做出完备诊断，通常需要重症医学科、呼吸科、感染科等多科室协作开展病例讨论，传统会诊模式面临过程繁琐，会诊医务人员存在感染风险等问题。如何改变传统会诊模式，优化会诊流程，降低临床医护人员跨区域工作受感染风险，提高临床医护人员工作效率

方法 通过借助移动互联网、大数据、5G 等信息化技术，基于医院临床数据集成平台，建设移动远程医疗系统（Mobile Telehealth System, MTS），改变传统会诊模式，优化会诊流程。

结果 实践表明，MTS 会诊能够实现医疗数据的有效利用，提高医院会诊管理与临床科室工作效率；面对疫情，在优化多学科之间的会诊与交流等方面能够发挥重要作用。

结论 远程医疗作为改善医疗服务的重要工具，能够有效助力于各类疫情防控，目前，日常使用的远程医疗系统缺乏移动性，借助国家“互联网+医疗健康”的政策推动，本文介绍的 MTS 会诊基于医院临床数据集成平台，支持患者病历展示、线上的案例讨论。MTS 会诊管理模式不仅能够优化多学科会诊流程，提高医院会诊管理的质量，还能加强科室之间的沟通交流，节省医务人员工作时间，提升临床科室的工作效率。

2019 年 12 月，新型冠状病毒肺炎（Corona Virus Disease 2019，COVID-19）[1]在我国武汉地区发现，随着疫情的蔓延，我国其他地区及境外也相继发现了此类患者。面对突发疫情，我院作为省内首批新型冠状病毒肺炎省级定点医院，为对患者进行集中诊治，在一院三区按照病情轻重缓急，设立“发热病房”、“隔离病房”、“重症隔离病房”共三级八个病区，由于新型冠状病毒感染的肺炎患者病情复杂、变化快、危重患者较多，患者病情牵涉多科室、多学科协作，为对患者做出完备诊断，通常需要重症医学科、呼吸科、感染科等多科室协作开展病例讨论[2-3]。针对传统会诊模式面临过程繁琐，会诊医务人员存在感染风险等问题，迫切需要借助移动互联网、大数据、5G 等信息化技术[4]，基于医院临床数据集成平台，建设移动远程医疗系统（Mobile Telehealth System, MTS），进而改变传统会诊模式，优化会诊流程，降低临床医护人员跨区域工作受感染风险，提高临床医护人员工作效率。

1 临床数据集成

数据集成是以医院管理、业务应用为出发点，提供技术支持的系统性解决方案，能够实现医院数据整合、系统集成、业务应用，为医院管理和临床应用提供服务的信息技术工具[5]。在遵守医疗数据安全规范前提下，依托医院数据集成平台，通过标准化方式整合患者病历信息，实现消息传递、数据共享与交互、系统异构集成，从而降低系统压力、减少系统风险、增加系统安全与稳定性，提高临床工作效率，为医联体、分级诊疗、患者服务、医院管理、科研辅助等应用提供技术实现方式，如图 1 数据集成平台分为四个部分：

（1）信息集成数据存储库（Integrated Data Repository, IDR）：IDR 不仅需要集成医院各临床业务数据（如门诊病人和住院病人病历数据），而且还包括医院的各类运营数据。

（2）企业服务总线（Enterprise Service Bus，ESB）：构建 ESB 机制，针对独立部署运行和异构的各类应用程序，促进应用程序之间的相互交互与通信，实现临床数据交互。

（3）业务外联系统（Business Outreach System，BOS）：实现医院内外部系统之间的信息交互，并充当接口适配器为将医院的内部和外部系统分开。

（4）所有这些组件均符合医疗信息安全标准和数据标准化。

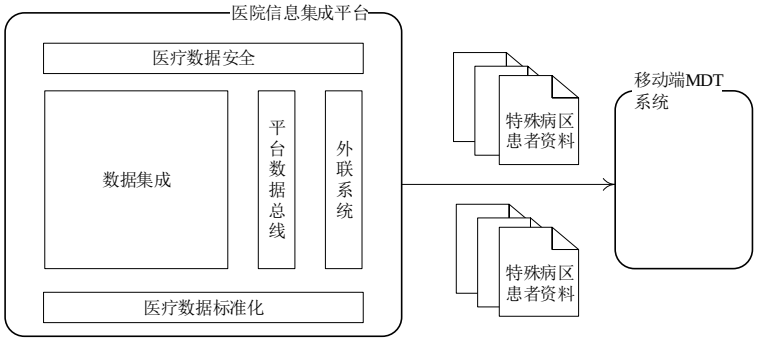


图 1 平台体系结构

基于数据集成平台，精确定义集成平台与 MTS 的数据接口，并将数据分为住院数据和门诊数据，共 22 个接口。住院数据包括住院各类病历、医嘱信息、实验室检验结果、检查报告和图像检查等。门诊数据包括患者门诊医嘱、门诊病历、实验室检验结果、检查报告和图像检查、体检信息等。MTS 将会诊所需的患者信息有效的通过集成平台获得。

临床医生在进行多学科会诊时，需要在 MTS 会诊 APP 上查阅需会诊且获得授权患者的全部病历信息，包括医嘱、PACS、LIS、各类文书等信息，可以在 MTS 会诊 APP 填写会诊意见，根据授权开立医嘱等，并将数据同步到病历系统，因此，MTS 会诊的构建需要以医院临床数据集成平台为基础。

2 MTS 建设

为保证疫情期间一线隔离病区医护人员工作效率，MTS 会诊 APP 提供简洁的界面，以消息提醒机制实时跟踪患者信息，并将患者全部信息集成到移动终端上，以便医务人员可以随时通过手机查询患者信息，进行病情展示跟踪以及病例讨论。

2.1 病情展示与跟踪

主管医生能够利用 MTS 会诊 APP 查阅及跟踪患者的病情病历，包含患者列表、病情摘要、电子病历、检验、检查报告与图像、会诊单、医嘱单等模块。

1.电子病历：系统具备调阅患者本次或历次就诊入院病历、病程记录功能；（图 2）。

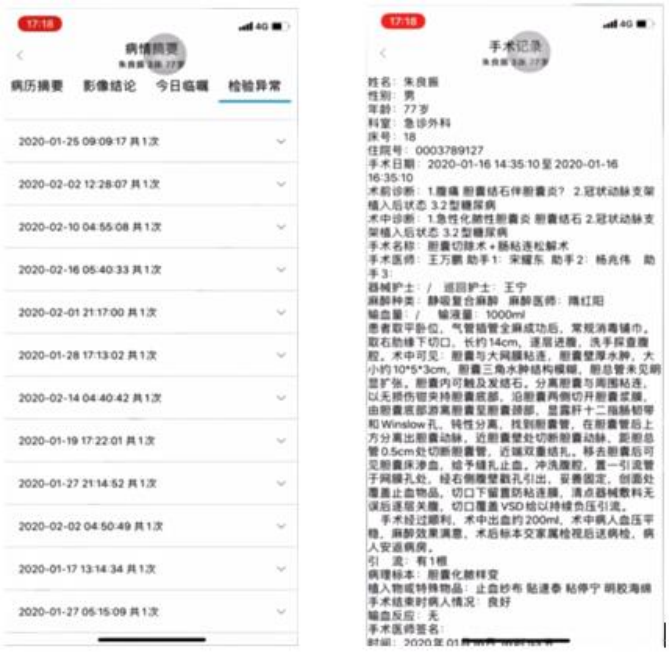


图 2 患者病情摘要与电子病历信息

2.病情摘要模块：系统具备浏览病人主诉、初步诊断、现病史、诊断依据、鉴别诊断等病历，本次就诊影像检查报告、检验结果异常值等。

3.检验模块：在查阅实验室检验结果时，实验室检验结构异常值以警告颜色提醒。

4.检查报告与图像：遵循 DICOM 标准[6]，该模块具有 DICOM 原始/压缩图像检索和 DICOM 图像处理功能；同时具有超声、内窥镜检查和病理 JPEG 图像等显示功能；通过缩略图对检查/序列进行快速导航，双击缩略图加载检查/序列功能（图 3）。

5.医疗建议模块：该模块涉及医疗状况的诊断和治疗。

6.会诊单与医嘱单：会诊医师能够通过 MTS 会诊 APP 进行会诊申请的提交、会诊申请接受、填写会诊意见等，并可根据授权为会诊患者开立医嘱。

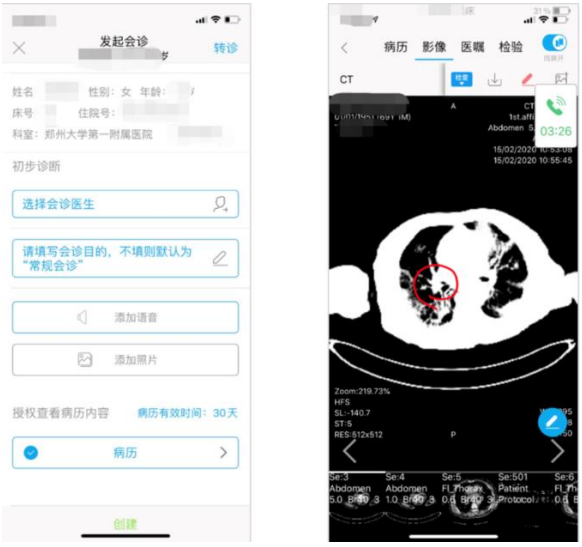


图 3 患者病情摘要与电子病历信息

2.2 病例讨论

在病例讨论模块，医务人员能够实现跨平台同屏交互，利用即时会诊呼叫/应答、语音/视频实时交互、病历与图像实时交互等功能，在不同的屏幕上进行多人移动交互式病例讨论。

以实时交互的方式，在共享患者病历数据的前提下，参加讨论的医务人员可以视频交流，病历记录、检查检验结果、图像等同步显示，并可以在同一屏幕上处理医学图像同步和图像处理标签（图 4）。

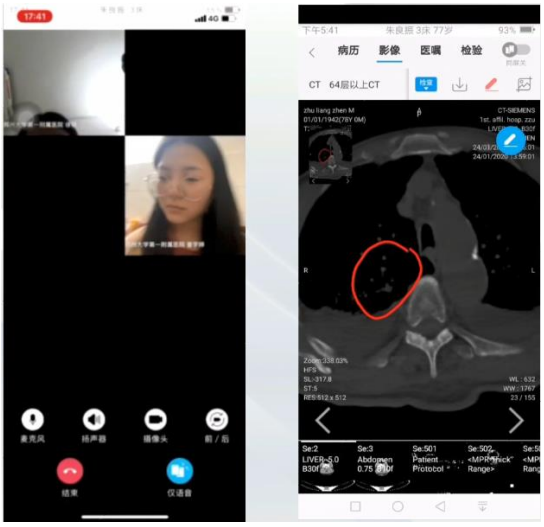


图 4 患者病情摘要与电子病历信息

在 MTS 会诊 APP 实际使用过程中，医生创建会诊请求并提交会诊请求表单，响应者收到请求，立即做出响应，并开始案例讨论。讨论结束时，参与者可以针对讨论内容输入自己的意见，也可以查阅历史意见，请求者可以选择授权共享患者数据。

3 信息网络安全措施

基于移动互联网、大数据、5G 等信息化技术，MTS 会诊 APP 需在互联网运行，并需要适时访问医院内网数据，从网络安全的角度出发，通过建立独立的网关服务器，手 MTS 会诊 APP 发送数据访问请求，网关服务器通过医院内部网从临床数据集成平台获取数据，然后发送给移动终端（图 5）。

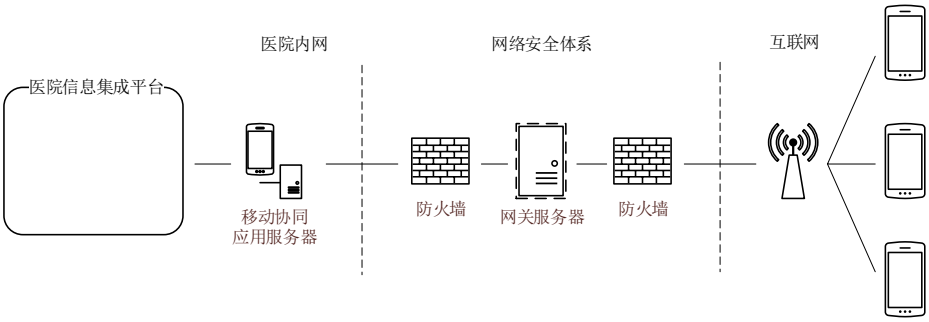


图 5 网络安全架构

在数据安全方面，从医院临床数据集成平台获取数据时，必须确保用户只能在授权范围内访问患者信息，移动端与医院内网应用服务器之间的数据访问，采用请求/响应模式，患者数据只缓存于移动端，医生退出应用程序后，缓存数据自动清除；在跨机构病例分享和会诊过程中，对患者姓名等关键身份数据可以进行脱敏处理；此外，在浏览患者病历信息时，操作医务人员水印会置顶显示，且对移动端进行截屏时，会有警告信息，以保护患者隐私（图 6）。

考虑到用户的安全性，MTS 采用单点登录，当医务人员在另一终端重复登陆时（获取授权验证码），系统会自动提示，并将原在线登陆端强制下线。移动端 MDT 系统采用手机号和医院工号共同注册机制，通过短信验证的用户才具备登录应用终端的权限，用户更换移动终端登录需再次短信验证，验证码发送注册绑定的手机号，防止盗号登录。

图 6 用户安全机制

4 应用效果

MTS 在医院新冠疫情防控过程中，被应用在八个隔离病房（如：发热门诊、发热病房、隔离病区、隔离重症病区等）与院内其它科室对患者病情进行多学科会诊并支持案例讨论，MTS 系统建设耗时 2 个月，正式上线 37 天内，一线隔离病区 65 名医生，完成新型冠状病毒肺炎相关会诊 301 次，使用 3,061 次。

MTS 的上线应用使得患者能够随时随地得到不同专科多位专家的“面对面会诊”，多位专家通过综合分析、精准研判，能够为患者提供准确、高效、更加优化的治疗方案；通过与在用医务人员沟通，MTS 会诊具有便捷、高效、简易的特点，为医院管理者、抗击疫情一线医生提供方面快捷的隔离区患者管理、多学科会诊等工作。

5 讨论

远程医疗作为改善医疗服务的重要工具，能够有效助力于各类疫情防控，目前，日常使用的远程医疗系统缺乏移动性，借助国家“互联网+医疗健康”的政策推动[7-8]，本文介绍的 MTS 会诊基于医院临床数据集成平台，支持患者病历展示、线上的案例讨论。MTS 会诊管理模式不仅能够优化多学科会诊流程，提高医院会诊管理的质量，还能加强科室之间的沟通交流，节省医务人员工作时间，提升临床科室的工作效率。

随着当前移动互联网、大数据、5G 技术的快速发展，通过整合医院医疗数据，深挖核心数据价值，形成专科专病数据库以及疑难杂症、罕见病等专家会诊知识库，并考虑利用人工智能技术，最终能够更一步的改善患者服务、提高医疗质量。但是，受制于移动终端电子签名的合法性、有效性、安全性等问题，目前，移动 MDT 会诊功能中电子签名的应用仍需要积极探索。我们的研究表明了使用移动协作技术开发 MTS 的潜力，以便利医疗保健专业人员并减少人与人之间的联系。

OR-002

浅谈病案管理信息化建设的必要性

魏梦迪
郑州大学第一附属医院

目的 随着网络信息技术的发展，信息化的病案管理也随之完善，在临床应用当中有着较高的应用范围。电子病历通常是将患者的治疗信息、个人信息等录入到计算机数据库管理系统当中，形成

针对性的病案。经过应用电子病历，可以有效提高信息利用率和共享率，提高治疗的有效性。本文旨在探究病案管理信息化建设的必要性，从而引起医疗行业对病案信息化建设的重视。

方法 通过分析病案管理信息化建设的现状和优势探究其必要性，并指出推进病案管理信息化建设的有效途径。

结果 目前病案管理信息化建设管理机制体系不完善、对信息化建设的重视不够；管理人员队伍综合素质有待进一步加强，缺乏专业知识和管理意识，工作积极性不高；各种软硬件条件存在着限制，投入力度严重不足；信息共享意识较为淡薄，缺乏共享平台支撑。信息化病案管理的优势包括：一是提升病历管理工作效率，提高信息的有效性和病案服务的科研价值；二是优化医院内部协调及管理；三是实现医疗信息共享，使患者的病历资料在不同诊疗过程中都得以共享，同时实现远程交流、远程会诊；四是提升医疗管理水平，电子病历能够通过直观的图形、表格等方式，清楚地表现出治疗的效果，有助于医疗效果评定和医院管理措施调整；五是有效促进医疗创新，可以为医护人员提供有效的参考资料，也可以使治疗方案更加具有针对性；六是节约人力成本，有效降低支出，提高医院的经营效益；七是减少安全隐患，通过借助于服务登录、登记验证以及密码钥匙安全锁等相关的技术来加强信息的安全保护，防止黑客的攻击，保护医院珍贵的资源；八是规避法律风险，可以长期较完整地保存医疗信息记录，并使所有过程有迹可循，从而为未来的纠纷处理提供有效的依据。

结论 在电子化存储模式下，病案管理自身的职能已经由单纯的病案记录、保存逐渐转变为集信息存储、共享、挖掘、监督于一体的综合型病案管理模式。这种职能的转变已经成为现代医院信息化发展的必然趋势，不仅能够减少病案管理的流程，缩短病案管理时间，极大地减轻病案管理人员的工作负担，也能够更好地提高医院整体的信息管理水平，更好地服务于患者。病案管理信息化建设是非常必要的，提高病案管理信息化建设的有效途径有：改善软硬件条件并对软件及时更新；加强人员培训，完善病案管理细则，结合工作建立相关的考核激励机制；集成化发展，使电子病案系统能及时采集、接收完整数据信息；标准化、规范化发展，增强数据互用性，促进信息共享。

一、医院病案管理的现状

（一）管理机制体系不完善，对信息化建设的重视不够。

病案是指医疗机构在开展临床工作过程中形成的，真实、完整记录病人病情变化和诊疗过程的原始记录，它客观、完整地记录了病人的病情变化和诊疗的全过程，具有价值性、技术性、专属性等特点。病案记录着很多重要资料，但是我国过去很长一段时间内都是采用纸质病案管理的

方式，要想查找相关病案存在费时费力的问题，且整理起来需要花费较大的工夫。随着信息技术的推广应用，病案管理也需要借助信息化技术来提高管理效率。但是从目前医院病案管理现状来看，很多医院主要将建设重点放在内部管理系统的信息化构建，而病案没有得到有效开发与利用，导致工作效率低下。由于医院对于病案管理工作的重视程度不高，而且在进行病案信息化建设的过程当中，往往是趋于形式，不能够真正有效深入，即使是进行了信息化建设也不能够充分发挥病案信息化建设的应有价值，尤其是相关的考核机制体系不够完善，病案信息采集与整理过程中，缺乏相应的监管措施，相关工作人员的工作积极性不高，存在着应付了事的现象。

（二）管理人员队伍综合素质有待进一步加强。

在现代信息技术快速发展的背景下，病案管理软件不断增多，推动病案管理工作的信息化发展，同时也对管理人员的专业能力提出了更高的要求。目前阶段各医院病案管理人员技能素质相对单一，病案信息化建设人员相对短缺，不仅缺乏专业知识，而且工作积极性不高。由于缺乏专业性的信息化建设管理队伍，也影响了医院病案信息化建设的发展步伐，加之病案管理人员缺乏管理意识，不能够采取行之有效的举措收集整理病案，这样往往会造成医院内部信息丢失或者是病案信息不全等现象，给医院的发展带来了很大的危害。

（三）各种软硬件条件存在着限制，基础设施不够完善。

医院在推进病案信息化建设的过程当中，一方面需要具备软件条件，能够结合医院自身实际建立病案信息化管理系统，另一方面要配备相关的硬件设施，比如说计算机、打印机、扫描仪等，只有对硬件进行有效的配备，才能够推进病案信息化建设。事实上，医院对于信息化建设的投入力度严重不足，不仅缺乏相关的硬件条件，而且各种多媒体影像设备以及录入系统的设备不够健全，直接限制了医院病案信息化建设工作的开展。

（四）信息共享意识较为淡薄，缺乏共享平台支撑。

医疗机构在医疗档案信息系统的相关软件开发应用是一个相对浩大的工程，需要相关各部门的共同努力才能构建。但目前各医疗机构对于病案管理也是以“保存”作为主要目的，“利用”的意识不高，管理方式较为封闭，不能将这些资源和信息运用到实际医疗救治中，医疗档案信息无法发挥自身作用和社会效应。另外，由于医疗档案信息共享的统一制度没有形成，系统开发滞后，缺乏平台支撑，现在各医疗机构的信息化平台均在内部局域网上封闭执行，无法实现实时共享。

二、信息化病案管理的优势

（一）提升病历管理工作效率

医疗改革需要适应社会发展的趋势，要想更好地服务于人民，需要重视数字化技术的引进，从而提高医院病案管理效率。病案管理是医院管理的重要内容，甚至直接影响医院的正常经营，因此利用现代化技术来提高医院病案管理效率具有重要的现实意义。和传统病历进行对比，电子病历不仅可以记录患者个人或者治疗信息，也可以放入 B 超、 CT 等图像信息，使其可以保存在计算机当中，成为具有个性化的电子档案。在另一方面，可以经过软件操作合理区分患者的疾病类型，便于后续检索。电子病历还可以避免由于书写而导致的错误，有助于提高信息的有效性。病案资源的开发利用成为医院科研管理的重要内容，电子病历的建设使用能够将传统的人工管理变为计算机管理，能够提高病案管理的工作效率，能够有效利用医院病案信息，从而提高了病案服务的科研价值。

（二）优化医院内部协调及管理

信息技术在病案管理中应用需要医院构建统一的病案管理系统，因此医院需要从配套的病案管理系统着手，将各部门的病案资料进行整理，从而对医院病案信息进行集中管理，并及时根据部门反馈完善病案系统。这样一来有助于医院管理与内部控制工作的实施，强调对医院管理的监管。充分利用医院现有病案档案资源，协调各部门之间的关系，从而提高病案管理工作的效率。

（三）实现医疗信息共享

在实际工作中，一些患者无法依据检查结果、表面症状进行确定针对性治疗方法，因此需要进行查找患者以往的就诊信息进行综合考量。如果患者就诊的医院有所不同，通常情况下无法从以往就诊的医院获得患者的就诊信息，因此可能会影响实际的治疗效果。而电子病历的应用，可以使患者即使在不同医院治疗过，也可以使病历资料在不同医院当中得以共享，并且也可以进行患者病情信息的远程交流，开展远程会诊，使患者的病情判断更加准确。

（四）提升医疗管理水平

纸质病历在管理上难度比较高，而且许多数据的整合对医疗人员的要求比较高。而电子病历能够经过直观的图形、表格等方式，清楚地表现出治疗的效果，有助于医疗效果评定。另外可以依据统计数据为医院调整管理措施做出参考，促进管理效果的提高。

（五）有效促进医疗创新

医疗创新对提高医疗工作质量具有重要的作用，而电子病历可以容纳大量的信息，不仅可以放入患者的治疗信息，也可以放入患者的预后等。不但可以为其医护人员提供有效的参考资料，也可以使治疗方案更加具有针对性。

（六）节约人力和成本

纸质病案管理通常要求投入大量的精力和不断扩大的储存空间，而电子病历可以保存在网络当中，无需较大的容纳空间，但是在存储信息上却有绝对的优势。不仅可以有效降低人力物力，也可以有效减少实物存储空间，因此电子病历可以有效降低支出，提高医院的经营效益。

（七）减少安全隐患

由于医院档案的存放和管理过程当中存在着诸多不便，通过借助于信息化技术推进医院档案信息化建设，能够进一步对相关的信息进行加密处理，通过借助于服务登录、登记验证以及密码钥匙安全锁等相关的技术来加强信息的安全保护，防止黑客的攻击，保护医院珍贵的资源。

（八）规避法律风险

法律规定住院病案档案保管时间为 30 年，若出现法律纠纷需要原始病历档案作为依据，因此需要做好原始数据的保护。而电子病案管理具有其天然优势，受物理环境影响较小，能够长期较完整地保存医疗信息记录，并且使所有过程有迹可循，从而为未来的纠纷处理提供有效的依据。

三、推进病案管理信息化建设的有效途径

（一）改善软硬件条件

首先应该加强重视程度，投入足够的人力物力财力保障完善信息化建设所需要的一切软硬件条件。在大数据时代背景之下必须要及时更新软件，提高对病案信息的处理能力，对各种病案信息进行有效收集、整理和分类、归纳，为病案管理信息化建设提供有效的保证。

（二）加强人员培训

应该完善病案管理的细则，结合工作建立相关的考核激励机制，把医院的病案信息化建设作为医院管理发展目标去认真考核，激发员工工作的积极性，保证病案管理信息化建设人员的综合素质。结合实际需要，要对各种人才进行引入或者加大内部培训力度，形成更加专业综合的病案管理团队，保证管理的实际效果。

(三) 集成化发展

完善的病案管理信息化平台应当包括住院系统、医嘱系统、手麻系统（手术麻醉信息管理系统）、PACS（影像归档和通信系统）、医院 HIS 系统（医院管理信息系统）、检验科 LIS 系统（实验室信息系统）、病理诊断系统、处方系统等，各系统产生大量的数据由电子病案系统进行采集并归档，形成完整的电子病案。因此平台应当开通相应接口，与 HIS 系统、住院系统等无缝对接，使电子病案系统能及时采集、接收完整数据信息。

（四）标准化、规范化发展

公共卫生领域需要全面实现信息的标准化、规范化，增强数据互用性，促进信息共享，建立有效的数据管理方式，改进分析技术和方法。一方面，电子病历发展需要和国际标准进行联系，

使自身的格式、规范可以达到既定标准，可以有效输入系统，使数据共享更加快速。另一方面，需要对签字等方面进行重视，在完善签名认证等内容之后，才能使病历得以完善、归档。特别是在签名方面，需要将手写签名转换成电子签名，并且需要盖上负责人的专用章，提高病历的专业性和规范性。

在电子化存储模式下，病案管理自身的职能已经由单纯的病案记录、保存逐渐转变为集信息存储、共享、挖掘、监督于一体的综合型病案管理模式。这种职能的转变已经成为现代医院信息化发展的必然趋势，不仅能够减少病案管理的流程，缩短病案管理时间，极大地减轻病案管理人员的工作负担，也能够更好地提高医院整体的信息管理水平，更好地服务于患者。

OR-003

基于 DTFS 的 PACS 存储系统设计

朱彦霞[1]、陈益洲[2]、华南[3]、罗刘敏[4]

1. 河南省职工医院 2. 河南省卫健委

3. 河南有线电视网络集团有限公司 4. 郑州工商学院

目的 医院信息化等级评审、智慧医院评级等一系列医疗评审制度与政策的出台与推进，对医疗相关的信息化技术提出了更高的标准与要求。随着信息化技术的发展、医疗技术水平的提高、电子病历的广泛普及，医疗信息大数据时代已悄然而至。医学影像技术的发展对疾病筛查、诊断水平的提高起到了非常重要的作用，但在医学影像存储系统中，患者的影像数据量也与日俱增，给医院信息化系统的存储能力带来了新的巨大挑战，与此同时，在系统日常业务运行过程中影像报告并发查询效率问题、系统升级过程中面临的大数据迁移问题以及存储集群性能优化、存储成本节约化等问题都成为大家普遍关注的焦点。目前基本分布式的文件存储系统技术已日趋成熟，但在医院信息系统中的应用并同时整合医院硬件资源等方面的探讨与研究较少。

方法 本文基于 DTFS（分布式文件传输）同时将医院所有硬件存储资源整体规划，采用分时段、分布式存储机制设计 PACS 分布式存储平台。

结果 可实现医院内部硬件分时段空闲资源的合理调配与高效利用，与此同时，可有效提升 PACS 数据的并行访问速度、提高临床医师阅片与诊断的效率、增强患者查询影像结果报告的体验感，极大降低大数据存储的灾备风险并可确保必要时进行系统数据迁移的速度与安全性。

结论 该系统可有效解决 PACA 数据存储优化问题、访问性能，提高数据存储的安全性。

引言

目前,医疗信息数据的存储主要采取传统的中心化 SAN 存储方式，平均每个患者医疗影像（PACS）存储数据在 500MB 以上[1]，以三级医院为例，根据国家卫生统计中心相关数据 2019 年全国三级医院平均出院人数约为 3-3.5 万人，仅单向存储 PACS 数据空间至少占约 14T，由此面临极大的存储、备份、存储扩容压力以及随着数据量增加而带来的数据库访问压力；尤其当医院进行存储硬件更换、软件系统升级时将面临具大的数据迁移压力以及数据完整性、数据一致性、数据迁移效率低下等问题。在日常的诊疗业务中，医院局域网内存在服务器及终端分时段存储、网络资源闲置或利用率不高的问题。如何在确保数据安全的前提下，提高医院硬件存储、网络资源的利用率是医院信息化管理极为关注的话题。

区块链具有去中心化、可追溯、安全可靠等优点[2-4]，网络中各节点遵守特定的共识机制实现相互间的信息及权益的获取。区块链平台可以分为公有链、联盟链和私有链[5]，公有链中所有的节点不限制组织与参与者，联盟链由若干个参与机构共同管理并遵守共同认定的联盟规则,私有链严格限制参与节点的资格，一般应用于企业或行业内部的数据管理、审计等业务中，或者比较特殊的组织内部，通过密码技术对存储信息进行保护并记录完整的存储记录。本文提出并设计了一种基于 DTFS（Distributed File Transmission System, 分布式传输文件系统）的 PACS 存储系统，充分利用现有各网络节点时段性闲置资源及私有区块链、加密算法等技术，实现分节点数据的存储、传输与备份，在 PACS 业务数据传输方面有效提高文件网络存取速度、降低数据坍塌式损毁风险。

1 国内外研究现状

近年来，国内外许多研究人员和机构运用区块链技术在医疗领域积极探索并取得了一定近展。目前，区块链技术主要应用[6-11]于电子健康病例存储与共享、医疗欺诈与理陪、药品溯源与防伪、DNA 钱包、比特币支付、药品防伪、蛋白质折叠等方面。2015 年，飞利浦与 Tierion 合作，实现将将医疗数据记录到区块链中[12]，同年，Guardtime 实现了基于 hash 计算的密码学无钥签名基础设施的区块链技术[13]；2016 年 4 月，飞利浦与区块链技术公司 Gem 共同推出了 Gem Health[14]项目以推动医疗领域间的合作，同年，PokitDok 公司与因特尔公司合作，提出“Dokchain”医疗区块链计划，拟构建医疗领域临床与财务数据的分布式网络解决方案[15]；2017 年，Patientory 发布区块链医疗应用平台[16]，为患者与医护人员之间搭建安全数据存储点，实现患者对个人医疗健康历史数据的访问；2018 年，沃尔玛获得三项有关区块链专利，其基于区块链的医疗记录系统将患

者数据存储在分布式账本中[17]，当病人遇到无意识的昏迷情况等紧急情况时，相关医疗记录可被授权访问，同年，Clinicoin 利用区块链技术提供以患者为中心的临床试验数据共享，通过令牌奖励实现生态加强；2019 年，SAP 公司与 Chronicled 合作，利用区块链技术为制药客户提供信息共享服务。

2017 年，薛腾飞等人[18]针对医疗数据共享过程中用权限审查及数据校验资源及时间消耗量大等问题，提出了一种基于区块链、共识机制等技术的数据共享模型，但该方案仍采用中心化数据存储方式、数据恢复及安全性问题尚未解决，同年，梅颖等人[19]将云存储与区块链技术相结合，提出一种分布式的个人医疗记录存储方案，该方案使用加密技术将医疗数据存储于链式结构中、利用区块链存储医疗记录公共信息与访问权限等，实现了对医疗敏感数据的安全存储；2018 年，乔蕊等人[20]基于区块链技术建立动态存储数据模型，提高了“动态数据”的可信度；2019 年,刘志豪等人[21]将云数据库与区块链相结合设计一个分布式数据存储方案,徐健等人[22]将区块链与非对称加密技术结合，有效利用非对称加密安全性高、多方协作简单等特点，实现医疗记录跨域分享的数据追踪、防篡改及身份验证的简化，张利华等人[23]针对单管理节点易遭受攻击的威胁，通过设计双区块链结构分别实现医疗记录存储与共享。在本文中，充分利用现有网络资源时段冗余的特点、通过软技术实现基于 DTFS 的 PACS 存储优化，可节省投入成本提升系统安全、运行性能。

2 系统背景技术

2.1 分布式哈希表

分布式哈希表（Distributed Hash Table，DHT)是分布式存储与传输的关键技术，广泛应用于 P2P 网络中。网络中各节点共同维护一个形如<key, value>的文件索引 Hash 表，通常将文件的 hash 值存入 key 中、将 IP 地址存入 value 中；这个 hash 表会按照一定的规则算法分割成若干份存储于网络的各节点中，每个节点仅维护一小部分 hash 表；当查找时，仅需查找相应报文路由到相应节点即可[24]。

2.2 区块链

2008 年 11 月中本聪所提出的“比特币”概念，2009 年初实践化区块链正式诞生，全球掀起区块链热潮。目前已经从以“比特币”为代表的 1.0 时发展到 3.0 时代：“以太坊”为区块链 2.0 时代的代表主要应用限于金融领域，将数字货币与智能合约技术相结合并实现了用 Pathon、JavaScript 等语言 Solidity 进行智能合约的开发[25]；区块链 3.0 时代基于共识机制、智能合约、

跨链融合等区块链技术构建区块链世界，将“去中心化”理念应用于各行各业，实现各种资产在区块链上的交易与跟踪。

区块链（Block-chain)将 P2P 协议、数字加密、时间戳、分布式数据存储等技术实现数据的去中心化链式存储，各节点间通过共识机制（主要包括工作量证明 PoW、权益证明 PoS、权威证明 PoA、DPos 股份授权证明机制、交易量证明机制 EPos 确保链上数据的一致性[26-29]。

3 平台设计

3.1 PACS 系统存储应用现状

现各医院 PACS 多采用 SAN 结构的单一存储，存储构架如图 1 所示。

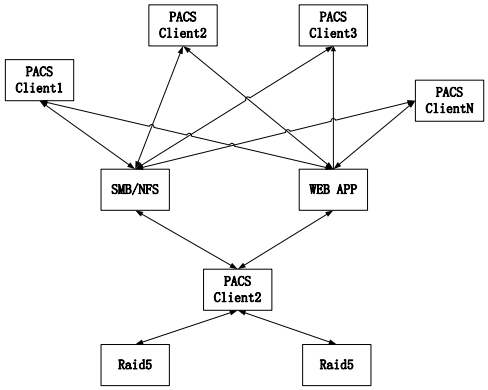


图 1 基于 SAN 的单一存储构架

以一个省级三甲医院为例，1 年至少能产生 15TB 的数据，如果每份数据文件以 20KB~10MB 颗粒度计算，将至少存储 150 万个左右小文件。由此，在此背景下，基于 SAN 存储面临的主要问题如下：

1.一般要求 SAN 实现主备镜像，主备镜像的日常数据同步需要消耗较大的存储及 I/O 压力。

2.当发生主备镜像中的某个数据存在异常错误， 主备镜像数据全部恢复耗时较大（以 15TB 数据为例，主备镜像全部恢复≥3 个小时)，同时恢复期间设备 I/O 占用率为 100%，有造成服务停止的风险。

3. 当存储的小文件达到一定量时，无论是 NTFS 还是 EXT4 等常见分区格式，对文件的检索和查询性能均会降低。

4.一般采用 SMB\NFS 等方式与第三方系统进行数据共享，如果遇到操作系统漏洞，可能导致数据被破坏或者泄露。

5.所有的数据通讯都需要通过 PACS 服务器，如果业务量大则压力无法平衡。

3.2 DTFS 整体解决方案

本于基于 DTFS 的 PACS 存储系统方案如图 2 所示，其整体设计思路如下：

1.采用 HTTP、HTTPS 和 WebDev 协议。

2.客户端支持 Windows\Linux 操作系统。

3.数据库采用分布式部署的版本控制模式，数据采用中心加分布式存储。

4. DTFS server 服务端为全量存储，但采用块文件方式存储物理文件，小文件已打包方式存储，大文件切片分割，切片信息被存储到核心数据库中。

5. DTFS node 客户端后台运行不会打扰桌面用户，可通过 server 设置分布式存储和内存使用的大小，无需全量存储数据，但需保持数据索引与服务端数据库版本同步；当使用 DTFS 的程序提取数据时，首先判断是否在本地存储，是则直接读取本地区块，否则通过 DTFS server 的 API 直接取得网内区块；DTFS node 需定期上报本地资源使用情况，保证某个节点的 I/O、CPU 和剩余空间压力不至于过大（设置临界点）。

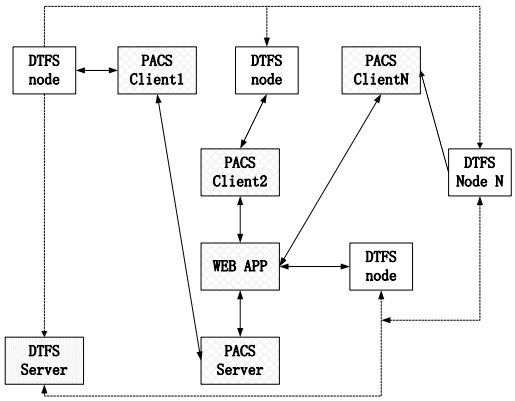


图 2 基于 DTFS 的 PACS 存储系统

3.3 DTFS 节点存储分层设计

DTFS Node 的存储方式采用三层设计架构，各层设计功能如下。

1.冷数据层

用于存储长期保留的数据。在新建 DTFS Node 时，一般采用后台多线程方式从同节点设备进行数据块同步，当 DTFS Node IO 或 CPU 占用率较高时降低写指令频率。

2.温数据层

从当前 DTFS Node 的 PC 产生的数据，优先写入本地，并将结果上报中心服务器完成同步；数据同步后其它分节点，根据主服务发布的广播指令进行数据同步，以类 CNS 方式减少或缓解客户端对服务器端请求压力。

3.热数据层

除中心服务器外，所有的 DTFS node 均有本地数据库，只存储 DTFS server 下发的指令和需要上报的指令和数据区块。

3.4 算法设计

3.4.1 DTFS node 审核

每个 DTFS node 安装后会产生一个 GUID 并提示用户选择服务器或输入服务器 IP;链接成功后并对本地资源进行上报，上报后 DTFS Server 会审核此 Node；DTFS Node 通过审核后，服务端会下发根据 DTFS Node 的硬件水平设置磁盘和内存的策略，然后 DTFS Node 转入后台运行；当一个使用 DTFS 的 APP 或者 WebAPP 通过认证机制访问 DTFS API 接口操作数据时，当设备硬件性能不足则可以设置禁用本地缓存模式直接从服务器推荐的 DTFS 节点读取数据。

3.4.2 新建文件过程

新建文件的过程也即导入 DTFS 系统的过程，具体流程如下：

1.首先根据用户 ID(UserID)获取用户角色(UserRole)，同时调用本地 DTFS API 接口对文件进行二进制流化处理并根据设定切片包大小(Buffer_length)进行数据流切片；与服务器进行握手，用<Public key, Private key>进行数据加密并获取 Hash、Did、Gid、Version 等相关信息后打包上报；上报后数据分别被存储于 DTFS Server 及当前 DTFS Node。

2.服务器获取并签入数据后，广播指令给 1/5 的在网资源较富余的 Node 并分批下发同步指令，Node 按 DTFS-Tree 结构根据资源状态进行同步，服务端可在后台查看目前 Node 节点和数据的运行状态。

3.Node 节点的策略分为 Full 模式及 Auto 模式。通常可选取服务器集群中业务原资源消耗率较低的节点标识为 Full Node，用于 DTFS 数据优先资源同步；其他终端节点标识为 Auto Node，并根据日常监控资源使用情况进行 Auto 分发同步；Auto 模式设置两种状态，即 CDN 状态和镜像状态（镜像状态包含 CDN 功能），如果所有节点达不到 DTFS 镜像模式的最低标准，则仅启动 CDN 功能，DTFS 将失去灾备恢复能力。

3.4.3 读取文件过程

文件读取过程中主要采用 NO-SQL 技术完成内存数据操作等，具体流程如下：

1. 调用请求 DTFS API 接口并传送<Gid, Did, Hash>, 接口可选择获取对应的 Version 信息及 与之相关的存储路径信息等并根据路径信息，自动从本地或路径导航指示 Node 数据块中获取相应文件数据。

2. 获取文件数据后，DTFS API 调用 Server 上存储的对应密钥对数据进行解密；解密后的文件数据存储于 Node 节点上的内存数据库中。

3. 读取文件过程结束后，DTFS API 对内存数据进行释放。

3.4.4 修改文件过程

文件修改过程支持版本控制功能，主要分为版本覆盖、版本升级两种模式，具体流程如下：

1.版本覆盖模式

Gid 不变 Version+1；Did 和 Hash 会被刷新与修改后的切片一同上传，同时服务端下发指令，将 Gid 下原有版本的数据在所有节点上声明清除已节省空间；此时所有 Node 节点获取数据时只能请求 Server 端或已经同步过的 Node 节点来获取数据。

2.版本升级模式

Gid 不变 Version+1； Did 和 Hash 产生新的与修改后的切片一同上传，同时服务端下发指令，将 Gid 下原有版本的数据在所有节点上声明 Version 升级。

3.4.5 删除文件过程

文件删除过程也采用了版本控制机制，分为彻底删除和当前版本删除两种模式，具体流程如下：

1.彻底删除模式

根据 Gid 直接删除所有历史版本，所有 Server 和 Node 节点启动就会执行，服务端只会保留删除历史。

2.版本删除模式

根据 Gid 的获取版本号并将相应版本标记删除；删除后客户端将不再获取此版本数据但其他数据节点会暂存此节点信息直至同步数据资源不足时释放。

3.5 系统安全机制

3.5.1 DTFS 标记约束

每个软件应用 DTFS 的时候，需通过平台申请账号并获取系统为软件应用分配唯一标识及会话认证 key；软件应用调用 DTFS API 的时候，会根据机制获取会话令牌用于文件操作流程。

3.5.2 文件命名约束

文件命名根据路径轨迹以冒号为分隔符，以软件应用 APP1 创建及查询 PACS 文件过程为例：创建过程，创建形如 App1:眼科:张某:2020-04-22:眼部扫描.zip 的文件；查询过程根据命名约束进行路径拆解扫描，并根据接口调用及传参情况获取相应资源。

3.5.3 数据访问原理

本地软件由 DTFS API 调用平台会串行化一个 http/https 地址用于数据本地软件的获取；DTFS API 还支持 WebDev 模式，可实现类 Windows 操作文件方式，但不支持彻底删除和选择版本读取。

3.5.4 数据同步镜像原则

镜像服务需要通过节点实时进行备份且至少保证在线的设备空间与主服务器是 1：1 的可用空间，以下对理想状态与非理念状态下存储网络状态进行了分析，分析假设：服务器端 20TB 全量数据，以全院 100 台设备计算。

1. 理想状态

理想状态下所有节点设备的可存储空间在全量数据的 5 倍左右，此基础配置下可以保证极端情况下至少有 33%的设备在线以支撑系统正常运行要求；按照每台 PC 终端可划出 1TB 的存储空间用于 DTFS 存储，根据员工昼夜轮休情况，全院 2/3 的电脑在非工作时间处于关机状态，则剩余的 1/3 的电脑已可满足服务器短的镜像同步需求。

2.非理想状态

节点不满足 DTFS 的基础需求，则 DTFS 工作采用热 CDN 模式。该模式下，对近期有访问的数据进行缓存，容量超限之后，自动删除访问率较低的数据，当核心服务器数据丢失后，不可全量恢复、仅可支持恢复所有节点当前版本数据。

3. 成本对比分析

以下将 PACS 存储分别在 SAN 模式及 DTFS 模式下的存储服务器硬件成本及运营恢复成本进行了分析对比，分析结果表明：基于 DTFS 的存储模式无论在硬件成本及运营恢复成本上均优于传统 PACS 模式。

（1）存储、服务器等硬件成本分析

传统 SAN 模式服务器采用 Raid5，有效存储 20TB，服务器、硬盘及配套硬件总成本约 6 万；采用 DTFS 模式，以 100 台 PC 客户端构建 DTFS 子节点同时考虑服务器冗余资源利用。根据市场调查目前 PC 客户端硬盘存储≥2TB，空闲存储≥1TB 可用于 DTFS，平台整体硬件成本将小于 3 万元，当客户端在线率及冗余率增高时，其 DTFS 扩容成本随之降低。

（2）运营成本分析

客户端请求服务器读取数据的时候服务器端会产生大量网络和 I/O 压力，对服务器及网络扩容需求增高带来相应运营成本不断增高；采用 DTFS 按照局域网是 1000M 计算，通过瀑布式传输可分布实现数据同步传输，常规数据的读写不直接再请求核心服务器，充分了缓解了核心服务器和核心交换机的压力，延缓了 IT 扩容的周期，相应运营成本增幅较慢。

3.5.5 数据灾难恢复

系统设计充分考虑数据灾难恢复方案：当遭遇中心服务器宕机数据彻底丢失时，首先重新部署服务器；部署完成后，各 Node 端自动链接 Server（如果服务端 IP 域名变更则需要 Node 端重新修改一遍配置），Server 端审核后，审核通过的 Node 会自动上传本地版本库的数据和版本日志，只要本地分节点的数据覆盖能达到 100%则数据可以完美恢复；若恢复过程中，Node 的数据可能会发生篡改或损坏，则 Node 节点之间需要进行批量数据块校验比对，通过服务端仲裁策略决定服务端的最终恢复相应版本数据；当冗余备份不足时，仅恢复当前节点版本数据。

4. 总结

数据储存备份压力、数据并发传输效率、数据迁移风险是 PACS 信息化平台中面临的重要问题。本文在医院 PACS 平台之上设计了一个基于 DTFS 的存储分发容灾平台，可有效延缓系统升级时间以实现资源高效利用并降低成本投入，同时通过存储及安全机制策略设计可有效提高系统的数据传输速度、提升系统安全性能并降低灾难性数据丢失风险。

OR-004

基于 LDA 模型的新冠肺炎网络健康社区信息需求分析

安欣宇[1]、陈育新[2]、张晗[1]
1. 中国医科大学医学信息学院
2. 西南大学计算机与信息科学学院

目的 2020 年初，重大传染病新冠肺炎由武汉蔓延至全国，新冠肺炎传染性强，防控难度大。为有效遏制疾病传播，群众出行受到限制，网络问答社区成为群众疾病信息交流的重要沟通渠道。用户通过网络健康问答社区提问或共享信息，以求满足自身的信息需求。本文探究疫情期间网络健康用户对于新冠肺炎的提问主题分布，为信息服务提供参考。

方法 本文搜集网络健康问答社区——快速问医生肺炎板块 4-6 月发布的 2672 条新冠肺炎提问记录，采用 LDA 模型对提问主题进行建模，利用主题一致性结合主题-词汇分布概率确定抽取 9 个主题。通过对主题进行分析，探究新冠肺炎疫情期间网络健康用户新冠肺炎相关信息需求。

结果 对用户提问的词云分析显示网络健康用户信息需求主要集中在新冠肺炎的临床症状、传播途径、检查诊断三个方面。利用 LDA 模型所提取的 9 个主题进行分析，结果显示网络健康用户对于新冠肺炎临床症状、检查诊断、用药治疗、疫区防控等方面关注较多。

结论 本文对新冠肺炎网络健康用户关注的热点问题进行主题分析，研究发现新冠肺炎网络健康用户的信息需求与国家相继颁布的诊疗方案及防控指南内容相一致，说明本研究新冠肺炎用户提问主题分析结果准确反映该阶段网络健康用户的信息需求。

OR-005

鼻咽癌放疗后大出血患者智能预警系统的研究与设计

葛晓伟、程铭
郑州大学第一附属医院

目的 针对鼻咽癌放疗后大出血患者死亡率高、并发症严重，且疗效、预后以及生存质量差等问题，考虑如何利用机器学习等方法对鼻咽癌放疗后患者进行预警，提前给与干预措施。

方法 基于建立的鼻咽癌放疗患者专科专病数据库，对选择的特征指标进行差异性分析，选择差异性较大的特征指标数据，并设计分类评估框架，选择分类预测效果最好的算法，用于建设鼻咽癌放疗后大出血患者预警系统。

结果 实现鼻咽癌患者放疗后大出血的早期识别、预警及正确干预，并能为患者推荐更为精确的个体化干预措施，具有良好的临床应用前景。

结论 探索鼻咽癌放疗后大出血的风险预测因素可明确高风险人群，及早给予干预，降低鼻咽癌放疗后大出血，仍然是一项重要的研究。本文通过运用机器学习算法预测模型，帮助医护人员提供评估治疗决策数据，对放疗后大出血进行预警，识别鼻咽癌放疗后大出血风险，帮助医护人员识别高风险患者，及早给予干预，能够有效降低其发生率，减少医疗成本。

鼻咽癌是一种发生于鼻咽腔顶部和侧壁黏膜的恶性肿瘤，其分布特点存在明显的人种与地域差异，且以黄种人和东南亚地区为主[1]。流行病学调查资料显示，欧美国家鼻咽癌年发病率小于1/10 万，而我国男性发病率高达 3.28/10 万，女性超过 1.36/10 万。由于鼻咽癌独特的放疗敏感性，以放疗为主的综合性治疗是目前重要的治疗方式，通过综合性治疗方式，鼻咽癌的五年平均生存率可达 60%左右[1]。但放疗造成的多种并发症对患者的影响不容忽视，如口腔黏膜炎、颈部纤维化、鼻腔频繁出血、分泌性中耳炎、鼻窦炎等，其中放疗后鼻咽癌患者出现大出血症状最为严重，其发生率不高，但致死率却高达 40%以上，严重影响鼻咽癌患者疗效、预后以及生存质量[2,3]。因此，如何预防鼻咽癌放疗后大出血是临床亟待解决的问题，其中早期识别、预警与干预是关键。

多数学者认为，鼻咽癌患者放疗过程中，剂量较大、正常组织和器官遭受放射性损伤、骨内动脉炎性反应及血管破溃等是引发鼻咽癌患者放疗后大出血的主要原因。高龄、肿瘤分期、合并颅底侵犯、多次放疗、肿瘤原发部位、肿瘤类型、是否合并炎症、开口门齿距＜ 1 cm、照射时接受总剂量等因素与鼻咽癌患者放疗后大出血存在一定关系[4]；此外，有学者认为，患者全身性的自身疾病，如高血压、糖尿病、动脉硬化、凝血功能障碍等也可导致并加重鼻出血[5]。

虽然，国内外学者针对鼻咽癌放疗后大出血的影响因素进行了积极的探索，但依旧存在诸多不足，例如：缺乏预警指标体系[6,7]。目前国内外研究主要是对鼻咽癌放疗后大出血的相关因素进行分析，尚未检索到鼻咽癌相关的预警指标体系。

基于以上问题，以信息化技术为手段，以预警管理理论为指导，围绕“降低鼻咽癌患者放疗后出现大出血症状的发生率及死亡率”实际需求出发，建立的鼻咽癌患者放疗专科专病数据库，选择差异性较大的特征指标数据，利用机器学习中分类评估框架，选择分类预测效果最好的算法，

设计鼻咽癌放疗后大出血患者预警系统，实现鼻咽癌患者放疗后大出血的早期识别、预警及正确干预。

2 专科专病数据库

2.1 专科专病数据库建立

由于鼻咽癌放疗后大出血患者数据存放在全院级的 HIS、LIS、PACS、电子病历、随访等不同数据库中，数据库的种类版本不同，结构化数据与非结构化病历文本数据共存，需要对鼻咽癌患者数据的全部电子病历的真实世界数据进行抽取，经过数据清洗、整合、数量验证、质量鉴定、数据池转换等预处理操作，按照统一索引实现鼻咽癌放疗后大出血患者专科专病数据的标准化、结构化信息的存储、查询及管理。

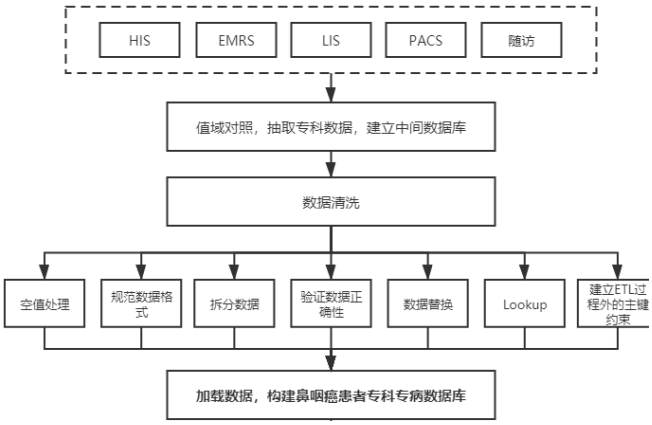


图 1 鼻咽癌放疗后大出血患者专科专病数据库结构图

2.2 病例数据及特征选取

基于专科专病数据库中某三甲医院 2016-2019 年鼻咽癌患者，抽取符合纳入实验标准的鼻咽癌放疗后大出血患者 67 例，纳入标准：①病理结果确诊为鼻咽癌并接受放疗者的患者；②一次出血总量＞300ml 或出血量＞100ml 且反复发生者；③去除鼻内镜等鼻科检查为普通鼻腔出血或者病历资料不全的患者。

以年龄、性别为基础匹配因素从专科专病数据库中，随机抽取相应例数的鼻咽癌放疗后未大出血患者作为对照组，共计 134 例患者的性别、年龄、肿瘤分期、放疗剂量、放疗次数、血凝、抗凝药、造影等八大类 33 个临床指标（详见表 1）计 9975 条数据用于本次研究。

表 1 鼻咽癌放疗后患者临床指标

指标类别	指标
个人因素	性别、年龄

主观因素	头晕、烦躁、鼻塞、出血量、鼻出血、血涕、脓涕、头疼、鼻疼、耳闷、口渴
药物因素	是否口服抗凝药
合并慢性病	高血压、糖尿病、其他慢性病
个人史	吸烟、饮酒
肿瘤因素	部位、分期、颅底侵犯、病理类型
检查指标	血红蛋白、红细胞、红细胞压积、血小板、血小板分布宽度、血管造影、凝血酶原时间
治疗因素	放疗剂量、是否再程放疗、放疗次数、是否化疗

2.3 数据特征指标分析

特征指标的选定有助于降低数据的拟合度、提高算法的精度，减少训练时间，避免对预测新数据造成干扰，采用卡方检验的方法对鼻咽癌放疗后是否出现大出血症状的患者各项指标进行差异性分析，卡方值越大，约不符合，越能用于区分，卡方值越小，越趋于符合，若卡方值相等。最终选择在鼻咽癌放疗后是否出现大出血区分上的差异均较大的前 20 项指标用于分类预测。

3 基于机器学习的分类预测

通过设计基于机器学习的分类预测评估框架来选择合适的分类预测算法，针对逻辑回归算法（LR）、分类与回归树算法（CART）、支持向量机（SVM）、贝叶斯分类器（NB）和 K 近邻算法（KNN）五种算法，采用 10 折交叉验证来分离数据并进行比较， KNN、CART 与 SVM 算法的评估结果比较好，需进一步优化，对 SVM 进行算法调参，在对放疗后大出血患者与未出血患者进行分类预测时，使用参数{'C': 0.1, 'kernel': 'linear'}时，分类预测模型的性能最好为 0.94；针对 KNN 算法，使用参数{'n_neighbors': 3}时，分类预测模型的性能最好为 0.88；针对 CART 算法，参数{'n_estimators': 160}，分类预测模型的性能最好，具有最优的准确率为 0.93。

选择召回率 R(recall)、F 度量值和精确率 P(precision)来评价鼻咽癌放疗后患者出现大出血症状与未出现症状分类预测效果。

表 2 鉴别分类预测结果

	precision	recall	f1-score
KNN	0.92	0.91	0.91
CART	0.87	0.88	0.87
SVM	0.93	0.94	0.93

根据对算法的评估结果，SVM 具有最好的准确性，选择 SVM 算法对分类。从表中结果可以看出，SVM 对正态化的数据具有较高的准确度，鼻咽癌放疗后患者大出血预测精确度达到 0.93。

4 预警系统架构设计

4.1 系统功能性需求设计

鼻咽癌患者放疗后大出血智能预警系统适用于住院患者，主要包含风险评估、风险预警、辅助决策、干预推荐、统计分析、健康教育、系统设置七个功能性模块，包含医护人员 Web 端和患者 APP 两种。

风险评估：在风险评估模块，系统与医院 HIS 系统做有接口，可自动获取鼻咽癌患者列表，医护人员登录后，可填写鼻咽癌患者疾病相关信息，对患者进行大出血风险评估，查看评估结果。根据医护人员录入评估内容，系统进行依据建立的预警模型，能够实现自动分析，快速识别风险。

风险预警：系统根据选定的机器学习分类算法，评估结果并给出预警值，判断风险等级，并根据患者风险级别显示不同颜色，发出预警信号，提醒医护人员患者存在大出血风险。

辅助决策：根据风险预警阶段不同的风险级别及患者存在的风险因素，系统自动推荐相应的干预措施，医护人员可结合患者病情、系统评估内容及结果，为患者制定个性化的智能的干预方案，从而起到辅助决策作用。

干预推荐：住院期间，医护人员为患者提供相应的干预措施，并在该模块进行记录并评价效果。

健康教育：健康教育模块包括疾病相关知识、健康饮食、大出血的预防、放疗相关知识、大出血先兆的观察方法等相关知识，供患者登录后阅读，提高患者疾病认知水平。

系统设置：根据人员角色及分工不同，设计不同的登录权限。包括医护人员登陆、患者登陆及软件工程师登陆界面。需要自行注册，设置登陆名及密码后登陆。医护人员登录可阅读评估内容，结果和干预推荐等；患者登录后可阅读健康教育相关内容。

4.2 系统实现框架与流程

根据鼻咽癌放疗后大出血智能预警系统的需求，将系统分为业务数据、数据处理、专科专病数据、算法引擎、逻辑与界面六层，各部分间耦合性低、独立性强，系统的架构设计如图 2 所示。

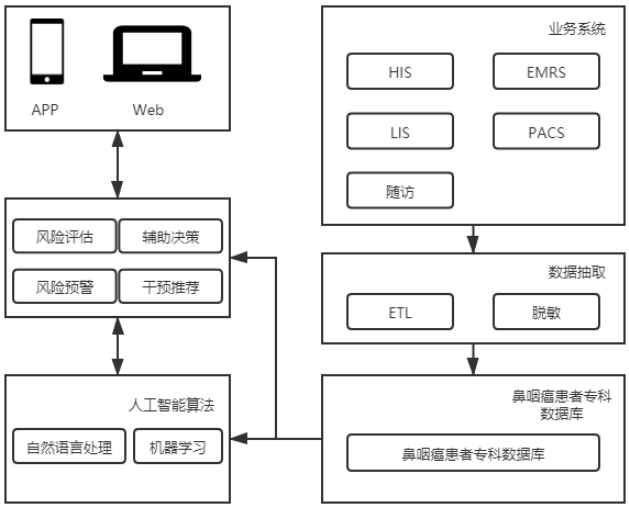


图 2 系统架构图

业务数据层包含医院的核心业务系统 HIS、EMRS、LIS、PACS 以及随访系统等。

数据处理层主要任务为获取所需的专科数据，通过将核心业务系统中所需要的数据进行抽取、转换及处理，并采用离线抽取数据与实时获取增量数据两种方法，通过设置定时器脚本增量更新，并对提取的数据进行筛选过滤。

将在数据处理层抽取的各业务系统中数据，按照住院号索引储存到鼻咽癌患者专科数据库，实现鼻咽癌患者数据的专库专存，便于系统应用。

逻辑层主要用于完成鼻咽癌患者放疗后大出血智能预警系统中风险评估、风险预警、辅助决策、干预推荐等逻辑功能。界面展示层实现用户与系统之间的交互界面显示，提供 PC 端与移动 APP 两种交互方式。

4.3 系统应用流程

医护人员登录系统后能够看到 HIS 系统同步的鼻咽癌患者，并按照要求补充填写患者相关信息。

利用鼻咽癌专科数据库中的数据以及医护人员根据实际情况录入的结构化数据或非结构化病历数据，利用自然语言处理技术、深度学习算法对非结构化病历文本数据提取特征因素，结合按照病历回顾、专家意见等形成的指标利用机器学习中效果比较好的分类预测算法进行分类，实现对患者的风险预警，并对患者风险进行智能化评估，计算风险评估值。

对于达到预警临界值的患者，判断其风险预警级别，给医护人员进行患者的风险预警提醒。

针对患者的情况，可以利用多标签的协同过滤等推荐算法，计算相同情况下患者的相似度，给出针对性干预方案，对干预方案实施后，需对方案进行评分，以便对推荐算法进行优化。

5 小结

近年来，较多医疗新技术的应用，能够有效减少肿瘤周围正常组织的受照剂量，提高放射治疗的靶区和鼻咽癌肿瘤组织的适形度，但鼻咽癌患者放疗后大出血仍是不可忽略的最凶险的并发症。探索鼻咽癌放疗后大出血的风险预测因素可明确高风险人群，及早给予干预，降低鼻咽癌放疗后大出血，仍然是一项重要的研究。本文通过运用机器学习算法预测模型，帮助医护人员提供评估治疗决策数据，对放疗后大出血进行预警，识别鼻咽癌放疗后大出血风险，帮助医护人员识别高风险患者，及早给予干预，能够有效降低其发生率，减少医疗成本。

OR-006

基于 DGCNN 和注意力机制的中文电子病历命名实体识别研究

胡海洋、赵从朴、姜会珍、李源杰、张晶、马琏、朱卫国
中国医学科学院北京协和医院

目的 运用自然语言处理技术自动化抽取和总结电子病历中的信息成为了目前十分热门的研究方向，其中医学命名实体识别是最基本和最重要的医学自然语言处理任务。由于目前基于词向量嵌入的模型性能极易受到分词效果的影响，且现有模型大多使用循环神经网络，计算速度较慢，很难达到实际应用要求。本文以临床医疗知识发现为导向，开展了相关的医疗命名实体识别及阴阳性判断研究。

方法 针对目前常用的命名实体识别方法存在的问题，本文在协和医学自然语言处理平台的基础上对不明原因发热患者的电子病历数据进行了实体标注工作，并且创新性的提出了基于多头自注意力机制的膨胀卷积神经网络模型，模型通过字嵌入和位置嵌入算法，减轻了模型对分词效果的依赖；通过使用膨胀卷积神经网络模型进行训练，进一步提升了模型识别效果，更好的理解语义信息并提高了计算速度。

结果 文中实验最终获得总体平均准确率 0.918、平均召回率 0.852、平均 F1 score 0.883 的结果，同时与其他三种常用命名实体识别模型方法进行对比，通过比较 F1 分数，本文中提出的模型相较于双向长短期记忆网络（Bi-LSTM）基线模型提高了 3.42%，相较于加入卷积神经网络输出层的双向长短期记忆网络（Bi-LSTM+CNN）基线模型提高了 2.66%，相较于加入注意力编码层的双向长短期记忆网络模型（Bi-LSTM+Attention+CNN）模型提高了 2.04%。

结论 本文针对不明原因发热的电子病历数据进行了命名实体识别与阴阳性识别研究，构建了对应的单病种自然语言提取模型，提出的基于多头自注意力机制下的膨胀卷积神经网络命名实体识别模型，相较于经典的双向长短期记忆网络模型等，能更好的提高命名实体识别效果，加快计算速度。同时本文提出的模型对临床文本中实体的阴阳性也有很好的判别。

引言

近年来随着电子病历系统的广泛使用，医院中已经积累了大量的非结构化电子病历数据，包含入院记录、病程记录、出院记录、检查记录、临床诊断、诊疗计划等内容，这些临床文本中包含着丰富的医疗信息和医学知识，对于医疗知识复用、疾病预测、药物发现、辅助诊疗等具有重要的意义[1]。电子病历中的命名实体识别技术近年来一直受到研究人员的关注，通过命名实体识别技术，能够自动识别与病人相关的医疗实体，生成结构化数据，帮助构建医学领域的知识体系。

1 研究现状

1.1 国外研究现状

由于欧美国家的医学信息化发展相对较早，英文电子病历的实体识别研究已经开展了多年，早期命名实体识别往往采用基于归纳总结的方法，即利用词典或者规则对医学实体进行映射和匹配，IBM 的 MedKAT（Medical Entities Dictionary）[2]利用国际疾病分类肿瘤学分册（ICD-O）实例化临床文本病理报告中的疾病信息，梅奥诊所中的 cTAKES（Clinical Text Analysis and Knowledge Extraction System）[3]利用医疗领域的术语词典抽取病历信息并建立映射编码，但基于规则和词典的系统高度依赖于专家归纳，系统对新实体的抽取效果差，效率较低。发展至今，目前很多机器学习和深度学习的算法已经应用到命名实体识别当中，包括隐马尔可夫模型（Hidden Markov Model,HMM）[4]，最大熵模型（Maximum Entropy,ME）[5]，条件随机场（Conditional Random Fields,CRF）[6]，以及循环神经网络（Recurrent Neural Network，RNN）和卷积神经网络（Convolutional Neural Network，CNN）等，都在命名实体识别模型中得到了广泛的运用，Bhatia 等人[7]针对医学文本命名实体识别任务提出一种新型的端到端的 Transformer 模型，采用共享编码器分别解码的方式来提高性能。Rebecka Weegar[8]等人通过集成学习方法，分别使用 SVM，Perceptron 和 CRF 和四个不同的数据集，为西班牙和瑞典的病历训练了模型。

1.2 国内研究现状

在中文电子病历实体识别领域，由于中文语言的特点和电子病历结构的不同，国内的研究技术发展历程也不同。如许源等[9]利用条件随机场与 RUTA 规则相结合的针对卒中入院记录医学实体识别方法; Cai[10]等人提出了一种基于自我注意机制的 LSTM-CRF 模型，将词性（POS）信息纳入深度学习模型中，以提高中文实体边界检测的准确性。Hui Wang 等人[11]采用了一种规则和条件随机场相结合的模型，针对肝癌病例构建了一个包含 115 份病历的语料库。Yin 等人[12]使用卷积神经网络提取汉字中的部首级特征，并使用自注意机制来捕获字符之间的依赖关系，取得了较好的效果。

针对现有模型大多使用循环神经网络，计算速度较慢的情况，本文以临床医疗知识发现为导向，在医学自然语言处理平台的基础上提出了基于多头自注意力机制的膨胀卷积神经网络模型，开展医疗命名实体识别及阴阳性判断研究。

2 基于膨胀卷积神经网络的电子病历命名实体识别模型

首先，电子病历中一个输入的文本序列为$T = (t_1, t_2, \cdots, t_n)$，$t_i$为该序列中第i个字符，我们需要生成对应的输出序列$Output = o_1, o_2, \cdots, o_n$，$o_i \in L$, L为所有标签的集合，最后通过解码算法从Output中解码出实体。

模型整体结构由图 1 所示，共由嵌入层、注意力层、卷积层、输出层 4 部分组成，嵌入层使用字嵌入和位置嵌入算法，解决了分词算法的错误，注意力层句子间的依赖性，卷积层对注意力层输出进行卷积操作，输出层由两层全联接层组成，采用切片式算法进行实体分割，后面将对每一部分进行详细介绍。

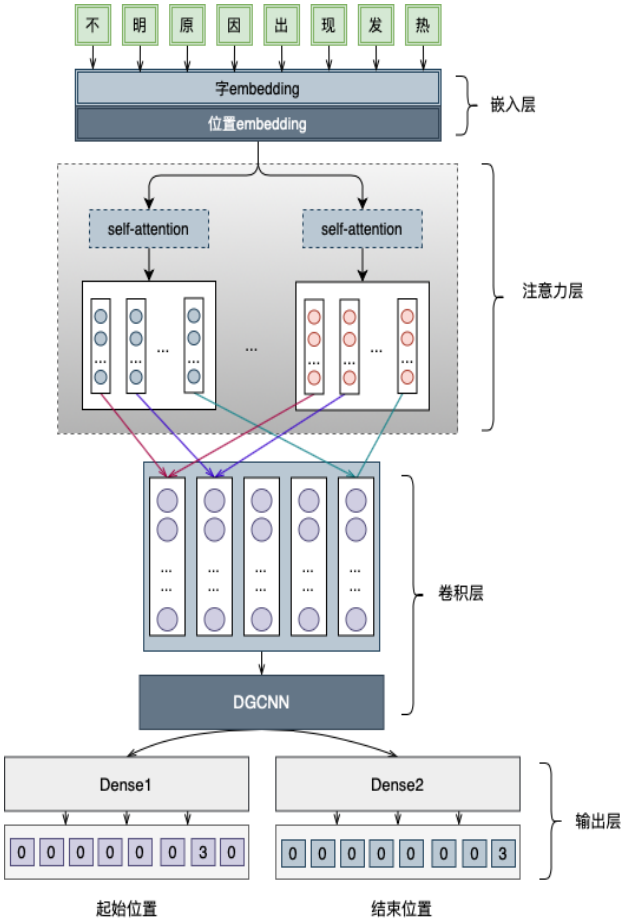


图 1 模型总结构图

2.1 嵌入层

针对常用的中文分词算法可能出现的错误，我们将临床文本的序列输入表示为字向量嵌入（word embeddings）和位置嵌入 (position embeddings)叠加组成。将序列中的每一个位置编码为一个向量，并将其融入到字向量中。字向量嵌入记录的是每个字的信息，位置嵌入记录了字的

顺序信息，通过字嵌入和位置嵌入算法，可以减轻模型对分词效果的依赖，字向量和位置向量的具体形成图如图 2 所示。

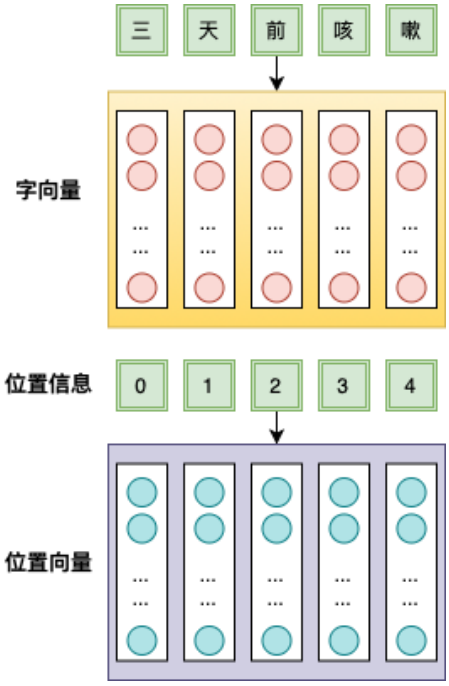


图 2 字向量、位置向量形成图

2.2 多头自注意力层

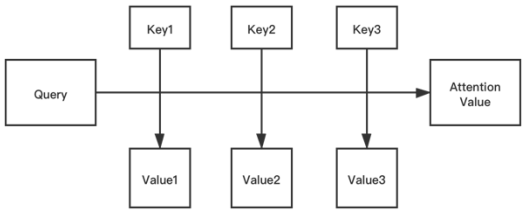


图 3 注意力机制原理

如图 3 所示，注意力机制(attention mechanism)可以被高度抽象为上述模型。注意力机制最初应用于数字图像处理领域，后来逐渐应用于自然语言处理领域的各种任务。可以将注意力函数看作一个查询 Query(Q)到一系列键 Key(K)再到值 Value(V)对的映射，通过计算 Query 和 Key 的相似度，再与对应的 Value 进行相乘，求和后得到最终的参数，具体计算公式为：

$$\mathrm{Attention}(Q,S)=\sum_{i=1}^{l_x}\mathrm{Similarity}(Q,K)*V\#(1)$$

在计算自注意力时，通常取$\mathrm{Query}=\mathrm{Key}=\mathrm{Value}$，并使用缩放点积的形式计算相似度，使内积不容易过大，缩放点积指在使用点积运算进行相似度计算的基础上，缩小$\sqrt{d_k}$倍（d_k为向量的维度），公式表示如下。

$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V\#(2)$$

多头自注意力机制是对自注意力机制的改进。首先，对 Q，K，V 使用不同的矩阵进行线性变换，重复 h 次；然后针对从每个线性变换获得的 Q，K 和 V 并行执行注意力函数生成输出；最后，对拼接后的输出做一次线性变换，多头自注意力的结果即为所得到的值。经过 h 次自注意力的计算，多头自注意力机制能够在不同的维度中学习相关的信息，强化其联系抽取的能力。本研究利用多头自注意力机制从电子病历文本的内部结构中得到字符之间的长距离依赖关系。多头注意力模型的具体公式表达如下。

$$\mathrm{MutiH(Q,K,V)Concat} = (\mathrm{h_1},\mathrm{h_2},\cdots,\mathrm{h_n})\mathrm{W^0}\#(3)$$

2.3 基于膨胀卷积神经网络的卷积层

传统卷积神经网络对序列建模受到卷积核的大小的限制，但是在自然语言处理中，文字往往是具有依赖性的，对于距离较远的依赖，如果想要抓取到这种依赖关系，必须线性的堆叠模型的层数，所以为了抓取文字的更长的依赖关系，相对于 BiLSTM 的长距离依赖关系编码，DGCNN 对局部实体的卷积编码可以达到更好的医疗实体识别效果。

膨胀卷积的工作机制为跳过与中心直接相邻的 p-1 个输入，与中心及与中心相邻的第 p 个输入连接，p 为膨胀卷积的膨胀率。在逐步扩大感受域、层数不断增加的过程中，神经网络参数呈线性增加，而感受域呈指数级增加。本文使用膨胀率依次为 1，2，5，1 的 4 层膨胀卷积神经网络，对注意力层输出的向量进行卷积，自动提取文本中特征，输出为对应的特征向量。

如图 4 所示，仅经过 3 步膨胀变化后，感受域就已扩散至输入矩阵中的全部数据。其中每层的参数数量相同，可有效减少训练时的参数，加快训练速度。

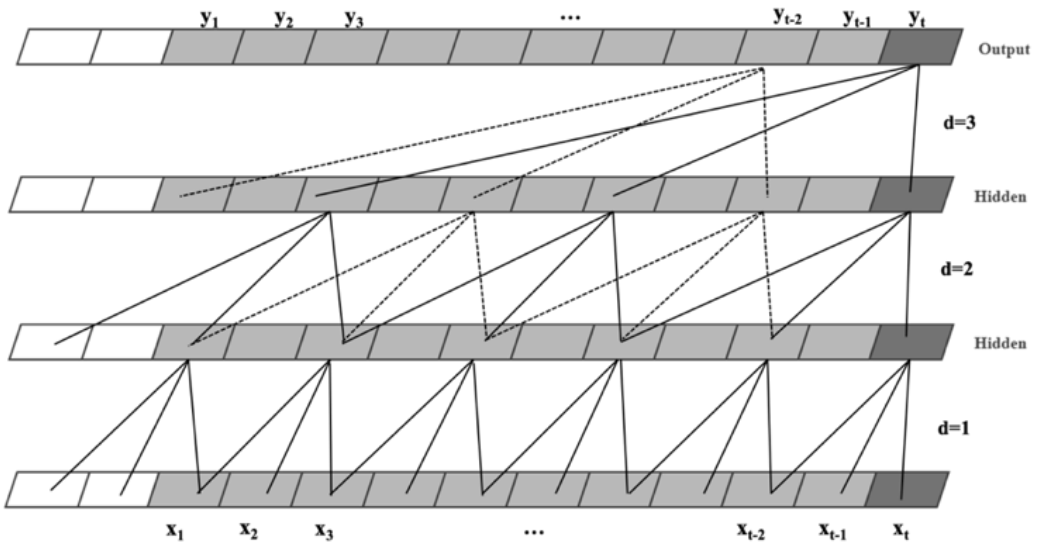


图 4 膨胀卷积神经网络原理图

2.4 输出层

2.4.1 全连接层

对于卷积层的输出$\mathrm{A} = (\mathrm{a_1},\mathrm{a_2},\cdots,\mathrm{a_n})$，分别通过两个以 sigmod 函数为激活函数的全连接层(Full Connected Layer,FCL)，得到一个实体首部预测序列$\mathrm{SP} = (\mathrm{sp_1},\mathrm{sp_2},\cdots,\mathrm{sp_n})$和一个实体尾部预测序列$\mathrm{EP} = (\mathrm{ep_1},\mathrm{ep_2},\cdots,\mathrm{ep_n})$。其中，$\mathrm{sp_i}$，$\mathrm{ep_i}$表示第$i$个位置的字符是某类实体首部、尾部的概率，

取概率值最大的那一类作为该位置的标签。使用一种切片式解码算法，对S和E进行处理，将实体属性与阴阳性从原始文本T中提取出来。切片式解码算法利用预测得到的实体首部位置和尾部位置在文本上进行切片以获得实体。

2.4.2 损失函数

模型训练使用了基于交叉熵的损失函数，其形式为：

$$\text{Loss} = \text{CE}(\text{SP}, \text{TrueS}) + \text{CE}(\text{EP}, \text{TrueE}) \#(4)$$

其中，$\text{CE}(\text{SP}, \text{TrueS}), \text{CE}(\text{EP}, \text{TrueE})$分别为首部位置预测、尾部位置预测与真实值的交叉熵，将两个交叉熵相加作为模型最终的损失函数。

3 实验设计与分析

3.1 实验数据

3.1.1 实验数据集

本次实验数据集中所用的电子病历来自于北京协和医院普通内科，共 685 份不明原因发热患者病历。病历包括入院记录，出院记录，病程记录三部分。入院记录记录了患者的主诉、现病史、既往史、体格检查等信息。病程记录是医生针对患者病情和诊疗过程所进行的连续性记录。出院记录中记录了患者出院时的诊断结论和医嘱。所有数据均进行了脱敏处理。

数据集总共包含 54300 条句向量，其中训练集 50000 条，测试集 4300 条。

3.1.2 协和医学自然语言处理平台（PUMCH-mNLP）

协和医学自然语言处理平台提供了医疗场景下病历文本分析及挖掘的核心工具，旨在帮助医院高效处理病历数据。通过该平台将非结构化的医疗文本被转化为包含重要医学信息的结构化数据后，以 json 格式输出或存储于数据库中。

3.1.3 数据实体标注

许源等[9]将入院记录中的命名实体类型分为 5 类，包括危险因素、临床症状、检验检查、治疗、药物。夏登辉等[13]构建的电子病历语料库中实体类型分为疾病、症状、检查和治疗 4 种类型。本研究借鉴上述语料库的构建经验，将电子病历中的实体分为症状、疾病、检验项目、检查项目、用药和过敏史 6 种类型，其中症状、疾病、用药、过敏史的类型区分阴阳性。

语料标注工作初期利用 PUMCH-mNLP 医学自然语言处理平台标注，然后在具有医学背景的相关人员协助修订下完成标注。

3.2 实验环境、评测方法

本文实验软件编程环境为 python3.5，使用 TensorFlow 框架构建膨胀卷积神经网络模型进行训练和测试。实验采用标准 Bakeoff 评分程序计算常用的评测指标:P(Precision,准确率)，R(Recall，召回率)，F1(召回率和准确率的调和平均值)。计算方法如下列公式所示。

$$\text{P} = \text{TP}/(\text{TP} + \text{FP}) \#(5)$$

$$\text{R} = \text{TP}/(\text{TP} + \text{FN}) \#(6)$$

$$F1 = 2 * P * R/(P + R)\#(7)$$

3.3 实验设计及参数

实验中分别使用了经典的三种模型作为对比模型，对提出的膨胀卷积神经网络模型进行多种对比实验:(1)使用基本的双向长短时记忆网络模型（Bi-LSTM）进行训练(2)使用 Bi-LSTM 算法对模型进行编码，卷积神经网络解码(Bi-LSTM+CNN)(3)使用双向长短时记忆网络模型加入自注意力机制，卷积神经网络解码(Bi-LSTM+Attention+CNN)。

其中，Muit-Attention+DGCNN 模型中字嵌入的维度为 128，序列最大程度为 100,在 CNN 模块采用 sigmod 作为激活函数，同时在 CNN 层中使用 Dropout，取值为 0.8。实验结果如表 1 所示。

模型	P	R	F1
BiLSTM	88.44	81.76	84.97
BiLSTM+CNN	91.00	81.04	85.73
BiLSTM+Attention+CNN	89.49	83.35	86.31
Attention+DGCNN	91.81	85.22	88.39

表 1 模型评价指标对比表

(单位:%)

3.4 实验结果分析

通过对比 F1 的分数，Muit-Attention+DGCNN 模型比 Bi-LSTM 基线模型提高了 3.42%，比 Bi-LSTM+CNN 模型提高了 2.66%，比 Bi-LSTM+Attention+CNN 模型提高了 2.04%。实验结果说明膨胀卷积神经网络模型在处理中文命名实体识别和阴阳性判断中有很好的优势。

4 总结与展望

4.1 工作总结

我们根据普通内科医生的实际科研需求，建立了基于多头自注意力机制下的膨胀卷积神经网络命名实体识别模型，获得了总体平均准确率 0.918、平均召回率 0.852、平均 F score 0.883 的结果。本文提供的 DGCNN 模型相较于传统的 Bi-LSTM 方法能更好的提高命名实体识别效果，加快计算速度。同时本文提出的模型对临床文本中实体的阴阳性也有很好的判别。

4.2 局限性和展望

本文仅对不明原因发热的电子病历数据进行了命名实体识别与阴阳性识别研究，构建了对应的单病种命名实体识别模型，还需要在更多的其他疾病的电子病历数据提取中去验证，当然在核磁共振报告、CT 报告、手术记录等文本提取中的应用也需要进一步研究。同时，膨胀卷积神经网络算法虽然有较好的识别效果，但是还有进一步改进的空间，后续我们将寻找更高效的算法来进行优化。

基于本文对自然语言识别模型的研究，我们将其应用到了具有自主知识产权的成果的协和医学自然语言处理平台，平台内集成了病历结构化、分词标注、实体抽取、词频统计、词云统计等多项功能模块，目前该平台已经获得了一项软件著作权。

OR-007

Association Analyses Reveal Comorbidity Patterns among Older Lung Cancer Patients in China

冯佳[1,3]、穆晓敏[1]、王伟[1]、徐鹰[2,3]
1. 吉林大学公共卫生学院
2. 美国佐治亚大学生物信息研究所
3. 吉林大学中日联谊医院

Purposes This study is aimed to identify the comorbidity patterns of older men with lung cancer in China.

Methods We analyzed the electronic medical records of lung cancer patients over age 65 in Jilin Province of China. We conducted rank-frequency analysis and social network analysis to identify the predominant comorbidity and comorbidity network. We applied the association rules to mine the comorbidity combination.

Results Our analyses discovered: (1) there are 31 additional medical conditions in older patients with lung cancer. The most frequent comorbidity is pneumonia, cerebral infraction, and hypertension. (2) Network-based analysis have revealed 7 sub-networks. (3) The association rules analysis shows the 41 interesting rules.

Conclusion This study provides evidences about comorbidity patterns of older men with lung cancer in China. Understanding the comorbidity pattern of older patients with lung cancer can assist clinicians in their diagnosis.

OR-008

基于深度残差网络的银屑病分类诊断模型研究

李鹏[1,2]、罗爱静[1]
1. 中南大学湘雅三医院 2. 湖南中医药大学

目的 银屑病是一种常见的皮肤疾病，发病的原因至今尚不明确，采用深度学习技术对银屑病进行分类诊断有助于减轻医生负担、简化诊疗流程、提高诊断质量，因此具有重要意义。

方法 文中提出一种基于深度残差网络的银屑病分类诊断模型。首先采用数据增强、银屑病图片大小调整和 tfrecord 编码等技术对网络的输入进行预处理，然后设计了一个 34 层的深度残差网络(ResNet-34)来对银屑病的特征进行提取。此外，文中采用交叉熵作为 ResNet-34 的损失函数来衡量模型的准确性，并利用 Adam 算法作为优化器来对 ResNet-34 进行训练，最终得到一个优化的 ResNet-34 模型用于银屑病诊断。

结果 基于 K 折交叉验证的实验结果表明，所提模型在召回率和 ROC 曲线方面的性能要优于其他诊断方法，可以为银屑病数据分析、疾病预防提供技术支持。

结论 文中提出一种基于深度残差网络的银屑病分类诊断模型。首先采用数据增强、银屑病图片大小调整和 tfrecord 编码等技术对网络的输入进行预处理，然后设计了一个 34 层的深度残差网络(ResNet-34)来对银屑病的特征进行提取。此外，文中采用交叉熵作为 ResNet-34 的损失函数来衡量模型的准确性，并利用 Adam 算法作为优化器来对 ResNet-34 进行训练，最终得到一个优化的 ResNet-34 模型用于银屑病诊断。基于 K 折交叉验证的实验结果表明，所提模型在召回率和 ROC 曲线方面的性能要优于其他诊断方法，可以为银屑病数据分析、疾病预防提供技术支持。

银屑病是一种常见的皮肤疾病，发病的原因至今尚不明确，采用深度学习技术对银屑病进行分类诊断有助于减轻医生负担、简化诊疗流程、提高诊断质量，因此具有重要意义。文中提出一种基于深度残差网络的银屑病分类诊断模型。首先采用数据增强、银屑病图片大小调整和 tfrecord 编码等技术对网络的输入进行预处理，然后设计了一个 34 层的深度残差网络(ResNet-34)来对银屑病的特征进行提取。此外，文中采用交叉熵作为 ResNet-34 的损失函数来衡量模型的准确性，并利用 Adam 算法作为优化器来对 ResNet-34 进行训练，最终得到一个优化的 ResNet-34 模型用于银屑病诊断。基于 K 折交叉验证的实验结果表明，所提模型在召回率和 ROC 曲线方面的性能要优于其他诊断方法，可以为银屑病数据分析、疾病预防提供技术支持。

1 引言

银屑病(Psoriasis，PS)[1]又称“牛皮癣”，是一种以表皮细胞过度增殖及免疫性炎症为特征的慢性反复发作性疾病。该病具有病程长、顽固性、常见性以及外观丑陋等特性，且病程后期可侵犯多种脏器，被世界卫生组织列入世界十大顽症之一[2]。

当前的研究大多是对名老中医治疗银屑病的用药规律进行了分析和总结[3-4]，但对于银屑病的诊断却很少探讨。事实上据统计，我国约有700万名银屑病患者，但只有少部分患者得到诊断，主要原因是由于基层医生或医院缺乏诊断能力。在2018年召开的“世界银屑病日科普活动”上，专家表示：人工智能技术(AI)结合大数据可以为患者提供更为直观的银屑病辅助诊断[5]。

银屑病是不能够完全治愈的，基本上每位患者都会复发，需要长期跟踪治疗。银屑病的诊疗难度非常高，病种主要包括寻常型银屑病、关节型银屑病、脓胞型银屑病和红皮病型银屑病等四种类型。在做诊断的时候，不仅仅要考虑到外观因素，还要综合考虑心血管、心理、胃肠道、自身免疫等多个方面，如何对疑似患者是否患上银屑病以及患上何种银屑病进行准确且快速地诊断是目前的一大难题[6]。该问题属于典型的分类问题，为此本文提出一种基于深度残差网络(Deep residual network[7], ResNet)的银屑病分类诊断模型，通过训练得到一个ResNet-34模型来对银屑病进行分类诊断，有效地提高了银屑病的识别率。

2 相关工作

深度残差网络在医学图像处理方面已经发挥了重要的作用，文献[8]将深度残差网络应用到冠状动脉CT血管造影图像斑块识别问题中，通过设计18层、50层和152层等三种深度的ResNet来重建心肺复苏术图像 (Cardiopulmonary Resuscitation, CPR)图像，以辅助放射科医师对冠状动脉斑块进行快速地诊断，极大地提高了诊断的准确性。文献[9]采用深度残差网络来实现胸片X线的诊断。文中采用模型激活热图来训练一个152层的ResNet实现X线胸片两分类：“有发现”与“无发现”的基本模型。文献[10]针对目前人工勾画缺血性脑卒中病灶精度不高、且效率低下的问题，设计了一种基于三维深度残差网络与级联U-Net的自动分割算法。文中首先利用图像的3D上下文信息并改善类不平衡现象，将脑卒中核磁共振图像(MRI)采样成图像块作为网络输入；然后利用3D深度残差网络与级联U-Net的分割模型来对图像块进行特征提取和分割处理。文献[11]针对目前头颈部肿瘤放射治疗危及器官的精确分割难度较大、手动勾画危及器官存在严重偏差等问题，提出一种基于3D深度残差全卷积网络的头颈部肿瘤放疗危及器官自动分割方法。该方法通过设计V-Net网络分割模型学习得到危及器官CT影像的深层特征和浅层特征，并提出利用器官位置先验约束采样

区域与随机采样相结合的训练样本选择策略，同时采用Dice损失函数对网络进行训练，不仅提高了网络的训练速度，而且可以保证在除脑干以外的危及器官分割中性能比其他对比方法更优。

此外，还有文献[12]针对目前皮肤癌黑色素瘤在识别过程中容易受到类内差异大和样本数据量少等问题的影响，导致识别准确率较低、且方法稳定性差的问题，提出了一种基于深度残差网络的黑色素瘤识别算法。该算法通过构建深度残差网络提取皮肤镜图像的高维特征，使用残差学习防止网络梯度退化、降低网络训练的难度，实现了黑色素瘤的有效识别。文献[13]针对ECG信号特征检测与识别在心脏病远程实时监测与智能诊断中难度很大这一问题，文中分析了影响心电图(ECG)信号采集的关键因素，提出了一种基于小波自适应阈值去噪和深度残差卷积神经网络的心律不齐信号识别算法。该算法使用小波自适应阈值技术完成ECG信号滤波，并设计了包含多个残差块(residual block)结构的20层卷积神经网络(CNN)，即深度残差卷积神经网络(DR-CNN)，对5大类心律不齐ECG信号进行了识别。

鉴于深度残差网络在图像识别方面所取得的优异表现，文中将其应用到银屑病诊断问题中，提出了一种基于ResNet-34的银屑病分类诊断模型，并通过在大规模银屑病数据集的实验验证了本文模型的有效性。

3 基于ResNet的银屑病分类诊断

3.1 ResNet的提出

在深度学习中，卷积神经网络(Convolutional Neural Network, CNN)是目前处理医学图像问题的主要手段。人们普遍认为 CNN 深度越深（参数越多），则 CNN 的非线性表达能力越强，可以进行更加复杂的特征模式的提取，理论上可以取得更好的结果。然而大量的研究表明[14]，传统的 CNN 结构随着层数加深到一定程度之后，越深的网络反而效果更差，过深的网络竟然使分类的准确率下降了，即出现网络退化问题(Degradation problem)。为了解决这一问题，何凯明等人[7]在 2015 年提出了著名的深度残差网络(ResNet)。ResNet 由一个个地残差单元相互组合而成（如图 1 所示），其特点是容易优化，并且能够通过增加相当的深度来提高准确率。其内部的残差块使用了跳跃连接(短路机制)，缓解了在深度神经网络中增加深度带来的梯度消失或梯度扩散等问题。

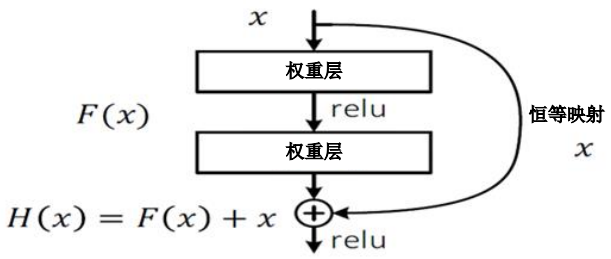


图 1 深度残差网络

3.2 ResNet原理

对于图 1 所示的深度残差网络，relu[15]表示网络的激活函数。残差单元 $F(x)$ 可以表示为：

$$F(x)=H(x)-x \tag{1}$$

在式(1)中，x 表示输入值；$H(x)$ 表示当输入为 x 时其学习到的特征。在图 1 所示的 2 层网络中，最优的输出就是输入 x，那么对于没有恒等映射的网络，就需要将其优化成 $H(x)=x$；对于有恒等映射的网络，即残差块，如果最优输出是 x，则只需要将残差单元 $F(x)$ 优化为 0 即可。后者的优化比前者简单，且更为有效，其原理如下：

设 x_l 表示第 l 层的输入值；$F(x_l,w_l)$ 表示第 l 层的残差；则有：

$$x_{l+1}=x_l+F(x_l,w_l) \tag{2}$$

经过多层残差网络的迭代后，有：

$$x_{l+2}=x_{l+1}+F(x_{l+1},w_{l+1})=x_l+F(x_l,w_l)+F(x_{l+1},w_{l+1}) \tag{3}$$

可以将式(3)重写为如下的形式，即我们可以求得从浅层 l 到深层 L 的学习特征：

$$x_L=x_l+\sum_{i=l}^{L-1}F(x_i,w_i) \tag{4}$$

根据反向传播的原理[14]，假设误差为 ε，对 x_l 求偏导，有：

$$\frac{\partial\varepsilon}{\partial x_l}=\frac{\partial\varepsilon}{\partial x_L}\frac{\partial x_L}{\partial x_l}=\frac{\partial\varepsilon}{\partial x_L}(1+\frac{\partial}{\partial x_l}\sum_{i=l}^{L-1}F(x_i,w_i)) \tag{5}$$

式(5)表示误差函数的反向传播。式子的第一个因子 $\frac{\partial\varepsilon}{\partial x_L}$ 表示损失函数到达的梯度，小括号中的 1 表明短路机制可以无损地传播梯度，而另外一项残差梯度则需要经过带有 weights 的层，梯度不是直接传递过来的。残差梯度一般不会全为-1，而且即使其比较小，有 1 的存在也不会导致梯度消失。这样一来梯度的衰减得到进一步抑制，并且加法的计算让训练的稳定性和容易性也得到了提高。所以可训练的网络的层数也大大增加了，网络的性能得到了显著提升。

3.3 基于 ResNet-34 的银屑病分类诊断流程

本文提出的基于 ResNet-34 的银屑病分类诊断流程如图 2 所示。医生或者疑似患者拍摄关于患病部位的照片上传到某一应用系统或 APP 中，系统或 APP 调用预先训练好的并已经部署完毕的 ResNet-34 对上传的图片进行分类诊断，最后输出得到是否患上银屑病以及患上何种银屑病的结论。

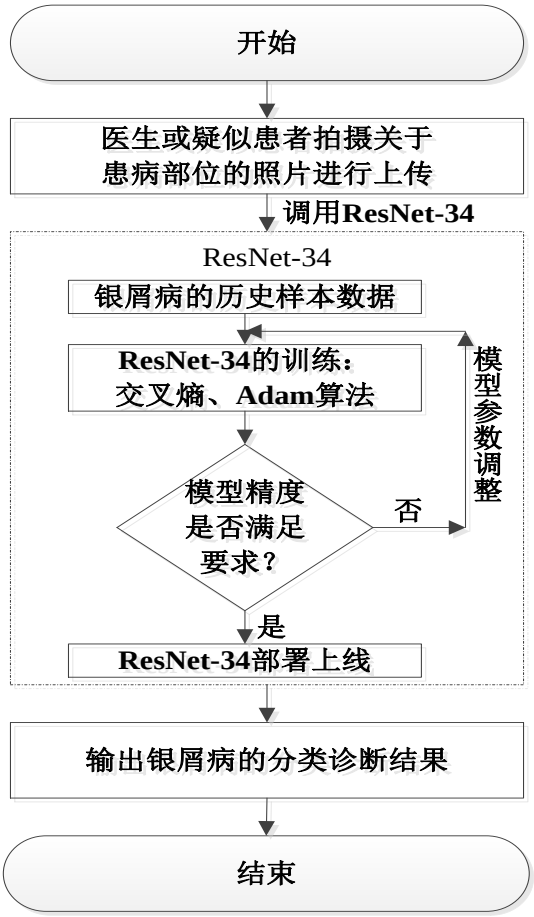


图 2 基于 ResNet-34 的银屑病分类诊断流程

4 技术细节

4.1 银屑病图片预处理

由于银屑病的病症较为复杂、发病的位置也多种多样，医生或患者在对患者的患病部位进行拍照时容易受到光照、拍摄设备以及设备像素等因素的影响，导致获得的银屑病患者图像数据经常存在大量的噪声以及图片格式不一致等问题，不便于进行进一步处理，为此本文从数据增强、图片大小调整和图片格式编码等三个方面来对样本图片进行预处理，使之满足 ResNet-34 的输入要求，从而为 ResNet-34 的训练做好准备。

4.1.1 银屑病图片数据增强

数据增强[16]也叫数据扩增，是指在不实质性的增加数据的情况下，让有限的数据产生等价于更多数据的价值。数据增强也是扩充数据样本规模的一种有效地方法。在ResNet-34的训练过程中，我们希望数据的规模越大、质量越高越好，这样训练得到的 ResNet-34才能够有着更好的泛化能力。然而实际采集数据的时候，往往很难覆盖掉全部的场景，比如：对于光照条件，在采集银屑病图像数据时，我们很难控制光线的比例，因此在训练模型的时候，就需要加入光照变化方面的

数据增强。此外，银屑病数据的获取也需要大量的成本，如果能够自动化的生成各种训练数据，就能做到更好的开源节流，且可以提高模型的鲁棒性。因此，在银屑病诊断的应用场景下做好数据增强工作是十分必要的。对于图片数据而言，目前常用的数据增强方法主要有旋转、平移、缩放和边缘填充等几种，下面给出了对银屑病图片进行数据增强的核心代码。

数据增强的核心代码：

```
#首先导入 keras 库
1. from keras.preprocessing.image import
ImageDataGenerator, 2. img_to_array, load_img
3. pic_path = r'./yinxiebing.jpg' #建立图片路径
4. augmentation_path = r'./data_augmentation' #保存路
                                              径
#定义 ImageDataGenerator，说明通过哪些操作产生新的图
片
5. data_gen = ImageDataGenerator(
                        rotation_range=30, #旋转
                        width_shift_range=0.1, #平移
                        height_shift_range=0.1, //缩
放
                        zoom_range=0.2, #边缘填充
                        fill_mode='nearest')
6. img = load_img(pic_path) #加载图片的地址
7. x = img_to_array(img) #转变成 array 格式给
                      ImageDataGenerator
8. x = x.reshape((1,) + x.shape)
9. n = 1
10. for batch in data_gen.flow(x, batch_size=1,
      save_to_dir=augmentation_path,
      save_prefix='train', save_format='jpeg'):
11.         n += 1
12.      if n > 10: #根据 ImageDataGenerator 定义的操作
               随机选择几种生成 10 张图片。
13                  break
```

4.1.2 银屑病图片大小调整

统一调整图片的大小，便于后面采用 ResNet-34 模型进行深度学习。本文以调整 data_augmentation 文件夹生成的新的银屑病图片为例，下面给出了对银屑病图片大小进行调整的核心代码。

图片大小调整的核心代码：

```
1.from PIL import Image #使用 PIL 库改变图片大小
2. import os #使用 os 库读取文件路径
3. img_path = r'./data_augmentation' #读取银屑病图片
4. resize_path = r'./resize_image' #将 resize 后的图片
                           放到 resize_image 文件夹中
5. for i in os.listdir(img_path):
6.         im = Image.open(os.path.join(img_path,i))
7.       out = im.resize((224, 224)) #resize 后的大小
为 224*224
8.         if not os.path.exists(resize_path):
9.                   os.makedirs(resize_path)
10.          out.save(os.path.join(resize_path, i))
```

4.1.3 tfrecord 编码

ResNet-34 模型的输入只能接受数值类型的数据，因此我们还需要对图片进行编码，即转换成 tfrecord 格式。tfrecord 是 TensorFlow[17]官方推荐的标准格式，能够将图片数据和标签一起存储成二进制文件，从而在 TensorFlow 中实现快速地复制、移动、读取和存储操作。训练 ResNet-34 的时候，通过建立队列系统，可以预先将 tfrecord 格式的银屑病数据加载进队列，队列会自动实现数据随机或有序地进出栈，并且队列系统和模型训练是独立进行的，这可以加速 ResNet-34 的读取和训练。下面给出了将银屑病图片转换成 tfrecord 格式的核心代码。

将图片数据转换成 tfrecord 格式的核心代码：

```
1. import os
2. from PIL import Image
3. import tensorflow as tf
4. cwd = r"./brand_picture/" #图片路径，两组标签都在
该目录下
5. file_path = r"./" # tfrecord 文件保存路径
6. bestnum = 1000 #每个 tfrecord 存放图片个数
7. num = 0  #第几个图片
8. recordfilenum = 0 #第几个 TFRecord 文件
9. classes = [] #将 labels 放入到 classes 中
10. for i in os.listdir(cwd):
11.    classes.append(i)
12.    ftrecordfilename=("traindata_63.tfrecords-
%.3d"% reco-
      rdfilenum) # tfrecords 格式文件名
13.    writer =
tf.python_io.TFRecordWriter(os.path.join(file_
            path, ftrecordfilename))
14. for index, name in enumerate(classes):
15.    class_path = os.path.join(cwd, name)
16.    for img_name in os.listdir(class_path):
17.       num = num + 1
18.       if num > bestnum: #超过 1000，写入下一个
tfrecord
19.          num = 1
20.          recordfilenum += 1
21.          ftrecordfilename =
        ("traindata_63.tfrecords-%.3d"
                              % recordfilenum)
22.          writer =
tf.python_io.TFRecordWriter(os.path.join(
                    file_path, ftrecordfilename))
23.       img_path = os.path.join(class_path,
        img_name)  #每一个图片的地址
24.       img = Image.open(img_path, 'r')
25.       img_raw = img.tobytes() #将图片转化为二进制
格式
26.       example = tf.train.Example(
27.          features=tf.train.Features(feature={
               'label':tf.train.Feature(int64_list=
                     tf.train.Int64List(value=[index])
),
               'img_raw':
```

```
tf.train.Feature(bytes_list=
                        tf.train.BytesList(value=[img_
raw])),
                }))
28      writer.write(example.SerializeToString()) #序
列化为字符串
29. write.close()
```

4.2 ResNet-34 的构建

本文构建了一个 34 层的深度残差网络(ResNet-34)来提取银屑病的特征，如图 3 所示。ResNet-34 限定输入的图片像素大小为 224×224。在 ResNet-34 中，首先采用一个大小为 7×7 的卷积核以 2 个步长对原始输入的图像进行特征提取，得到一个大小为 112×112 的特征图，紧接着采用最大池化(Max Pooling)技术来对特征进行压缩，然后经过四组大小分别为 3、4、6 和 3 的残差块进行多次反复的特征提取。其中，所有残差组中的卷积核大小都为 3×3，每组残差块中的第一个卷积核采用 2 个步长进行特征提取，其他卷积核采用 1 个步长进行特征提取。此外，同组残差块所采用的卷积核通道数均相同，分别为 64、128、256 和 512。各组输出的特征图大小依次为 56×56、28×28、14×14、7×7。最后经过平均池化和全连接层，采用 Softmax 函数进行分类映射，对输入图像是否属于银屑病进行分类诊断。

图 3 用于的银屑病分类诊断的 ResNet-34 结构

4.3 模型的训练

ResNet-34 通过改变学习目标，即由学习完整的输出变为学习残差，解决了传统卷积在信息传递时存在的信息丢失、损耗等问题，它通过将输入直接绕道传递到输出，保护了信息的完整性。在本文提出的 ResNet-34 模型中，我们采用交叉熵[18]作为损失函数来评估模型的准确性，采用 Adam 算法作为训练过程中的优化策略，采用 Softmax 函数实现银屑病的多分类诊断，下面进行详细阐述。

4.3.1 损失函数

损失函数是用来估量 ResNet-34 的预测值 $f(x)$ 与真实值 y 的不一致程度，它是一个非负实值函数。损失函数越小，ResNet-34 的鲁棒性就越好。在本文中，我们采用交叉熵作为损失函数，交叉熵能够衡量同一个随机变量中的两个不同概率分布的差异程度，在本文中表示为银屑病的真实概率分布与预测概率分布之间的差异。交叉熵的值越小，ResNet-34 的预测效果就越好。其计算公式为：

$$CrossEntropyLoss = -[y_i \log f(x_i) + (1-y_i)\log(1-f(x_i))] \quad (6)$$

4.3.2 优化器

本文采用 Adam(Adaptive Moment Estimation）[19]算法进行 ResNet-34 的训练。Adam 算法是一种在深度学习模型中将 Momentum[20]算法和 RMSProp[21]算法结合起来使用的优化算法，在训练的初始阶段，我们首先初始化梯度的累积量和平方累积量：

$$v_{dw} = 0, v_{db} = 0; s_{dw} = 0, s_{db} = 0 \quad (7)$$

然后在第 t 轮训练中，我们计算得到 Momentum 算法和 RMSProp 算法的参数更新：

$$\begin{aligned} v_{dw} &= \beta_1 v_{dw} + (1-\beta_1)dw \\ v_{db} &= \beta_1 v_{db} + (1-\beta_1)db \\ s_{dw} &= \beta_2 s_{dw} + (1-\beta_2)dw^2 \\ s_{db} &= \beta_2 s_{dw} + (1-\beta_2)db^2 \end{aligned} \quad (8)$$

其中，式(8)中 v_{dw} 和 v_{db} 分别是损失函数在前 $t-1$ 轮迭代过程中累积的梯度动量，β 是梯度累积的一个指数，一般设置值为 0.9。Momentum 算法和 RMSProp 算法主要是基于梯度的移动指数加权平均来进行训练，但是这种训练方式在迭代开始的初期会导致梯度优化的摆动幅度过大，不利于算法收敛。

所以需要对上面求得的几个值做偏差修正：

$$\begin{aligned} v_{dw}^c &= \frac{v_{dw}}{1-\beta_1^t}, \quad v_{db}^c = \frac{v_{db}}{1-\beta_1^t} \\ s_{dw}^c &= \frac{s_{dw}}{1-\beta_2^t}, \quad s_{db}^c = \frac{s_{db}}{1-\beta_2^t} \end{aligned} \quad (9)$$

通过式(9)可以求得在第t轮迭代过程中，参数梯度累积量的修正值。接下来，可以根据Momentum算法和RMSProp算法的结合来对模型的权重和偏置进行更新：

$$w = w - \alpha \frac{v_{dw}^c}{\sqrt{s_{dw}^c} + \varepsilon}$$
$$b = b - \alpha \frac{v_{db}^c}{\sqrt{s_{db}^c} + \varepsilon} \quad (10)$$

在Adam算法中，参数β_1所对应的是Momentum算法中的β值，一般取0.9；参数β_2所对应的是RMSProp算法中的β值，一般我们取0.999，而ε是一个平滑项，一般取值为10^{-8}，而学习率α则需要我们在训练的时候进行微调。综上所述，Adam算法的伪代码可以表述为：

1. 初始化$v_{dw} = 0, v_{db} = 0; s_{dw} = 0, s_{db} = 0$；
2. 在第t次迭代中，用mini-batch梯度下降法计算出dw和db；
3. 计算Momentum指数加权平均数；
4. 用RMSProp进行更新；
5. 计算Momentum和RMSProp的偏差修正；
6. 更新权重。

4.4 分类诊断

当ResNet-34训练完成之后，我们采用Softmax函数实现银屑病的分类诊断。Softmax函数主要用于多分类过程中，它将多个神经元的输出映射到(0,1)区间内看成概率来理解，从而实现多分类。ResNet-34模型的输出有5个值，代表4种常见的银屑病类型(寻常型、关节型、脓胞型和红皮病型)和正常人群(没有患病)。因此，对于每一个样本，依据Softmax函数的定义可知，它属于类别i的概率为：

$$y_i = \frac{e^{a_i}}{\sum_{k=1}^{5} e^{a_k}}, \forall i \in 1...5 \quad (11)$$

通过上式可以保证$\sum_{i=1}^{5} y_i = 1$，即属于各个类别的概率和为1。对于任意的输入样本而言，经过ResNet-34的处理后，最后只需要选择具有最大概率值所对应的类别就是该样本的诊断结果。

5 实验

5.1 数据来源

我们以2017年1月到2019年12月为一个时间周期，从湖南中医药大学第一和第二附属医院的皮肤科获取银屑病患者数据，并将其作为我们构建分类诊断模型的数据集。我们要求病人在活检前没

有接受过系统的免疫抑制治疗或停止PUVA/UV治疗至少一月以上，局部治疗至少停止2周以上，且无其他皮肤病及系统性疾病，共筛选得到数据样本40000份。

5.2 实验设置

我们在一台 8 核 16 线程的计算机上进行了实验。其中，CPU 型号为 Intel Core i9-9960X @ 3.10GHz，内 存 为 16G，操作系统为 Ubuntu 16.04 LTS 64 位。在 ResNet-34 模型中激活函数选用 relu 函数，采用深度学习框架 TensorFlow 和 Anaconda 平台实现了文中提出的基于深度残差网络的银屑病分类诊断模型。此外，为了避免模型过拟合，我们在 40000 份银屑病数据样本上采用 k 折交叉验证(在文中 k 取值为 10)来评估 ResNet-34 的预测性能，选出性能最好状态下的超参数，得到最终的模型。

5.3 评价指标和比较对象

本文将多分类问题转换为二分类问题来进行实验评价，转换的方法是采用一对其余(One-vs-Rest)的方法：将其中一个类标记为正例，然后将剩余的其它类都标记成反例。由于本文ResNet-34的输出为5类结果(4类为银屑病，1类为正常)，因此只需构建5个分类器就可以实现问题的转换。然后采用目前常用的精确率、召回率和ROC曲线等指标来评价基于ResNet-34进行银屑病分类诊断的性能。假设，银屑病患者表示正例，正常人表示反例，则可以采用如下的混淆矩阵来衡量ResNet-34的性能：

表 1 基于混淆矩阵的银屑病分类性能指标

真实值 / 预测值	正例(正常人)	反例(银屑病患者)
正例(正常人)	TP	FN
反例(银屑病患者)	FP	TN

在表 1 中，TP 表示将正例预测为正例，即真正例；FN 表示将正例预测为反例，即假反例；FP 表示将反例预测为正例，即假正例；TN 表示将反例预测为反例，即真反例。根据表 1 可知，精确率(Precision)和召回率(Recall)可以分别定义为：

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{12}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{13}$$

对于银屑病诊断任务而言，需要重点关注召回率，而不是精确率。因为银屑病在多数情况下是正例(不患病)，反例(患病)较少，两个类的样本比例差别很大。例如 100 条记录中，10 次发现

患银屑病，其中 6 次为误报，4 次为识别准确，虽然精确率降低为 94%，但是召回率却由原来的 0 上升到了 100%，虽然误报了疾病，但却没有遗漏真正患银屑病的人群。

ROC 曲线又称受试者工作特征曲线，用于刻画分类器在真正率(True positive rate，反映命中概率，纵轴)和假正率(False positive rate，反映误诊率、假阳性率、虚惊概率，横轴) 间的折中。其中 True positive rate 和 False positive rate 的定义如下所示：

$$True\ positive\ rate = \frac{TP}{TP + FN} \qquad (14)$$
$$False\ positive\ rate = \frac{FP}{FP + TN}$$

此外，我们将 ResNet-34 和目前性能较好的卷积神经网络 VGG19 在银屑病诊断上的性能做横向比较，以评价本文模型的优越性。VGG19 的实现参考文献[19]的工作。

5.4 结果分析

图 4 给出了 ResNet-34 和 VGG19 在银屑病数据集上进行分类诊断的召回率比较结果。从图 4 中可以看到，随着银屑病数据集规模的增大，两种方法的召回率都有不同程度的上升。但总的来看，本文方法的召回率总是高于基于 VGG19 的诊断方法，平均来看，本文方法的召回率要比 VGG19 高出约 9.5%。仔细分析其原因可知，这是由于：(1)相比于 VGG19 而言，ResNet-34 的深度更深，可以更好更丰富地提取银屑病的特征；(2) ResNet-34 通过引入残差的概念，在网络结构中加入了恒等映射连接，有效地解决了网络深度增加所带来的性能退化问题，因此取得了更好的性能表现。

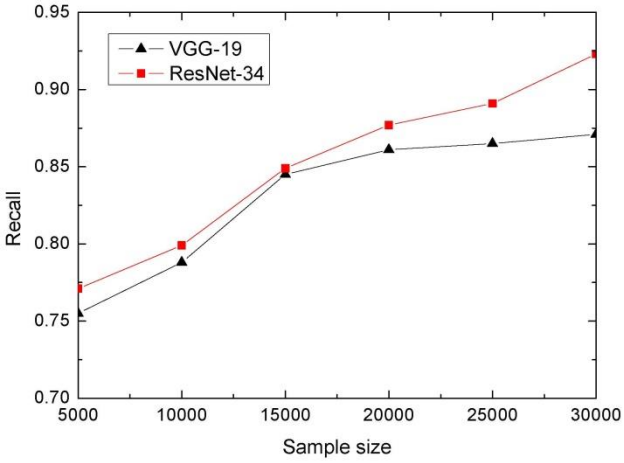


图 4 召回率比较(ResNet-34 VS VGG19)

最后，为了综合衡量 ResNet-34 和 VGG19 两种方法对于银屑病分类诊断的特异性和敏感性，下面画出它们的 ROC 曲线。对于每个测试样本，ResNet-34 和 VGG19 在进行每次分类时都会得到一个“Score”值(分值)，表示该样本多大概率上属于正例（或负例）。我们要画出 ROC 曲线，则需要一系列关于 True positive rate 和 False positive rate 的值。文中通过如下的步骤来实现：

(1)从高到低将“Score”值排序并依此作为阈值 threshold;

(2)对于每个阈值，“Score”值大于或等于这个 threshold 的测试样本被认为正例，其它为负例。从而形成一组预测数据。

(3)将观测数据的值进行连线，则可以得到 ROC 曲线。

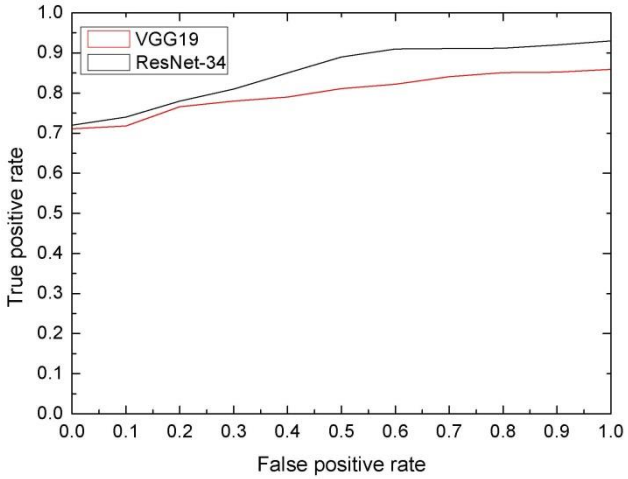


图 5 ROC-AUC 比较(ResNet-34 VS VGG19)

图 5 给出了两种方法进行银屑病分类诊断的 ROC 曲线比较结果，ROC 曲线下的面积称为 AUC（Area Under Curve）。AUC 值(面积)越大的分类器，性能越好。在图 5 中，红色线条下方的 AUC 值表示 VGG19 的分类性能，黑色线条下方的 AUC 值表示 ResNet-34 的分类性能，显然后者的 AUC 值要大于前者。这表明 ResNet-34 的性能要明显好于基于 VGG19 的方法，可以应用到真实环境下的银屑病分类诊断任务中去。

6 结束语

银屑病是一种非常难以治愈的皮肤疾病，由于发病的原因多种多样，如果准确地对银屑病进行分类诊断是目前研究的难点之一。文中提出一种基于深度残差网络的银屑病分类诊断模型，通过设计一个 34 层的残差网络来实现银屑病的精准诊断，最后的实验结果也验证了所提模型的有效性。在下一步工作中，我们将继续对银屑病的症状、证型以及用药规律等进行分析，将银屑病的诊断、治疗和用药等一体化过程构建成知识图谱，并进一步提出一种基于图卷积神经网络的银屑病用药推荐模型，从而对医生的诊疗工作提供更好的决策支持。

OR-009

2 型糖尿病患者早期眼底图像改变的临床预测模型研究

邵家辉
南通大学

目的 通过对糖尿病患者图像和正常对照组视网膜图像处理技术，通过机器视觉提取特征和先验知识提取视网膜血管形态特征，并利用 ANN 的方法对糖尿病患者与对照组进行分类，验证其有效性。

方法 使用的图片来源于 DRIVE 数据库中的正常眼底与病变图像,对视网膜图像基于小波矩的图像特征提取，基于纹理的图像特征提取，基于先验知识的视网膜形态特征提取，构建出临床预测模型。

结果 将整理后的小波矩特征矩阵作为输入后，结果病变眼底图像的正确率为 69.2%、错误率为 30.8%，正常眼底图像的正确率为 85.7%、错误率为 14.3%；病变眼底图像的正确率为 90.0%、错误率为 10.0%，正常眼底图像的正确率为 60.0%、错误率为 40.0%。整体上看，确定眼底图像是病变的正确率为 75.0%，错误率为 25.0%。具体如图 3.6 所示。纹理特征训练结果病变眼底图像的正确率为 66.7%、错误率为 33.3%，正常眼底图像的正确率为 63.6%、错误率为 36.4%；病变眼底图像的正确率为 60.0%、错误率为 40.0%，正常眼底图像的正确率为 70.0%、错误率为 30.0%。整体上看，确定眼底图像是病变的正确率为 65.0%，错误率为 35.0%。基于视网膜血管形态特征结果提示病变眼底图像的正确率为 83.3%、错误率为 16.7%，正常眼底图像的正确率为 100.0%、错误率为 0.0%；病变眼底图像的正确率为 100.0%、错误率为 0.0%，正常眼底图像的正确率为 66.7%、错误率为 33.3%。整体上看，确定眼底图像是病变的正确率为 87.5.0%，错误率为 12.5%。

结论 在糖尿病眼病的检测方面，临床医生主要依赖于眼底血管瘤、渗出等较为明显的眼底病灶进行诊断。与传统的医生定性报道视网膜血管形态不同，本实验将自动对视网膜血管的迂曲度、宽度等几何参数进行定量计算，因此，主要针对临床早期糖尿病眼病的检测，与现有的研究对象有一定的区别，并且更有临床意义。本实验比较了传统的基于计算机视觉方法提取视网膜图像特征与结合视网膜树拓扑特征进行分血管层次提取特征采用机器学习进行分类的结果。通过机器学习可以对糖尿病患者眼底图像进行辅助诊断，根据先验知识提取的视网膜形态学特征分类效果更好。在我们后面的实验中，我们还将进一步研究将机器视觉和先验知识提取的特征一起进行机器学习后的结果。

OR-010

医院图书馆微信公众平台用户信息采纳行为影响因素及实证分析

李日、高歌、伦志军
吉林大学第一医院

目的 本文从用户视角出发，构建医院图书馆微信公众平台用户信息采纳行为的影响因素模型并进行实证分析。

方法 通过结构化方程对各因素进行拟合和假设检验

结果 发现中心路径及边缘路径均可正向影响用户的信息采纳行为，并且中心路径的因素影响更大。中心路径中微信文章的内容及文章是否为原创对用户的信息采纳行为影响较大，边缘路径中文章发送的时间及标题的趣味性对用户的信息采纳行为影响较大。

结论 据此，提出优化平台管理及文章推送策略，以期为改进医院图书馆微信公众平台服务方式提供一定的参考。

本文从用户视角出发，构建医院图书馆微信公众平台用户信息采纳行为的影响因素模型并进行实证分析。通过结构化方程对各因素进行拟合和假设检验，发现中心路径及边缘路径均可正向影响用户的信息采纳行为，并且中心路径的因素影响更大。中心路径中微信文章的内容及文章是否为原创对用户的信息采纳行为影响较大，边缘路径中文章发送的时间及标题的趣味性对用户的信息采纳行为影响较大。据此，提出优化平台管理及文章推送策略，以期为改进医院图书馆微信公众平台服务方式提供一定的参考。

随着“互联网+”时代的到来，人们对信息资源的获取和利用呈现出了多元化方式。微信公众平台以其泛在化的服务方式、多形式内容推送和丰富的平台功能，成为人们获取信息的重要途径[1]。医院图书馆作为医院信息资源的集中地，理应顺应时代的发展，高效利用微信公众平台传播媒介，为用户提供全新的服务模式和阅读途径。截止 2020 年 3 月，经官方认证的三级医院图书馆的公众号达到了 34 个[2]，医院图书馆纷纷将服务扩展到微信息环境。然而，医院图书馆微信服务在蓬勃发展的同时也存在着诸多问题，主要表现在未能精准对接用户信息需求，忽略了信息传递过程中用户的主体感受。在新媒体时代，如何从用户的角度出发，优化微信公众平台信息服务方式，是目前医院图书馆需要探索和解决的问题。本研究基于精细化加工可能性模型（ELM）构建医院图书馆微信公众平台用户信息采纳行为的影响因素模型，分析影

响用户信息采纳行为的因素及各因素的作用机制，提出优化平台管理及文章推送策略，以期为改进医院图书馆微信公众平台服务方式提供一定的参考。

1 模型构建

信息采纳行为是近年来图书情报专业研究的热点。信息采纳是用户基于自身需求，主动对信息进行选择、判断，最后形成采纳与否的决策过程[3-4]，在信息采纳过程中主要受到内部因素和外部环境因素的影响[5]。ELM 是目前最具有代表性的信息采纳行为模型，近年来在医疗及电子商务行业的微信公众平台信息采纳行为的研究中得到了大量实证[6-7]。ELM 模型认为信息的采纳过程主要分为两条路径，即中心路径和边缘路径。中心路径是指信息接收者需要通过全面及深度阅读思考信息后改变其态度或行为的说服路径。边缘路径是信息接收者不需要深度思考信息内容，仅需要投入少量精力和时间就会改变其态度或行为的说服路径[8]。

本研究选择文章内容、文章是否原创、文章篇幅、文章是否有图片或表格作为模型中心路径的观测变量，文章标题趣味性、文章发送位置、文章是否有摘要介绍、文章配图选择作为边缘路径的观测变量，选取文章总阅读次数、分享次数、点赞人数作为信息采纳效果的衡量变量。提出两个假设：

H1 中心路径的观测变量对医院图书馆微信公众平台用户信息采纳行为产生影响。

H2 边缘路径的观测变量对医院图书馆微信公众平台用户信息采纳行为产生影响。

2 实证分析

2.1 数据的收集及基本情况

本研究收集了吉林大学第一医院图书馆微信公众平台发布的文章信息，数据收集时间范围为 2015 年 2 月 1 日（微信公众号建立时间）至 2020 年 6 月 5 日。共收集 219 条微信文章。收集的变量包括文章标题、文章发布的时间、文章发送的次序、文章是否为原创、文章的篇幅、文章发布时是否有摘要介绍、文章是否含有图片和表格、文章的总阅读次数、文章分享次数和点赞的人数。由两名医学图书馆专业人员对变量进行独立编码，编码规则见表 1，如果编码存在分歧，由第三名专业人员进行重新编码，最后达成一致编码意见。

表 1 变量的编码规则及变量分布情况

变量	赋	描述	频	构成

	值		数	比
文章内容				
	1	通知问候	73	34.1
	2	科室或医院成果宣传	16	7.5
	3	中文、英文数据库讲座介绍	58	27.1
	4	统计、专利和其他软件讲座介绍	25	11.7
	5	论文写作投稿或小常识等知识介绍	42	19.6
文章是否原创				
	1	转载	38	17.8
	2	原创	176	82.2
文章篇幅				
	1	大于 1000 字或小于 100 字	47	22.0
	2	500-1000 字	111	51.9
	3	100-500 字	56	26.2
文章是否有图片或表格				
	1	无图无表	50	23.4
	2	有图或表	164	76.6
标题趣味性				
	1	语言平淡，采用固定模式	111	51.9
	2	在 1 的基础上运用了特殊标点符号、数字等方法吸引读者眼球	79	36.9
	3	在 2 的基础上采用了疑问句、语气词等修辞手法	24	11.2
文章发送位置				
	1	非首位	35	16.4
	2	首位发送	179	83.6
是否有摘要介绍				
	1	无摘要	128	59.8
	2	有摘要	86	40.2
文章配图				
	1	自选图片	67	31.3
	2	采用吸引眼球的相关图片	147	68.7

发送时间

1	周五	54	25.2
2	周四	28	13.1
3	周三	37	17.3
4	周二	41	19.2
5	周一	54	25.2

2.2 信度和效度分析

本研究采用 Spss21.0 软件对收集的数据进行信度和效度分析。信度即可靠性，是指测验结果的一致程度和稳定程度[9]。通常使用 Cronbach α系数来衡量评价信度，它的取值范围是 0-1 之间，若该系数高于 0.7[10]，则认为测量方法具有较好的内部一致性，结果较可信。在本研究中，共进行了 12 个变量的内部一致性分析，Cronbach α系数为 0.806，说明本次采用的测量模型各变量间具有较好的内部一致性，可以进行下一步实证分析。

效度是指实证测验结果的有效性。效度越高表示测量结果越真实。KMO 值用来衡量研究的结构效度，一般值高于 0.5[10]，则可以认为测量方法具有较高的效度。本研究对数据进行结构效度检验，KMO 为 0.845（P<0.05），说明测量模型具有较高的结构效度。

2.3 模型的拟合

本次研究运用 AMOS22.0 软件，利用广义最小二乘法进行结构化方程模型拟合，拟合指标见表 2，除了拟合优度卡方检验χ2/df 和调整拟合优度指数 AGFI 与最优标准值稍不匹配，其他指标均符合结构方程模型的研究标准，因此可以认为本研究构建的模型具有较高的适配度，是可接受模型。

表 2 模型拟合度检验摘要表

项目	统计量	最优标准值	是否适配
χ^2	132.074	_	
df	42	_	
χ^2/df	3.145	<2	否
RMR	0.049	<0.05	是

GFI	0.905	>0.9	是
AGFI	0.853	>0.9	否
PNFI	0.663	>0.5	是
PGFI	0.578	>0.5	是
PCFI	0.691	>0.5	是
CFI	0.905	>0.9	是

2.4 假设检验

通过 AMOS22.0 软件进行结构化方程模型参数估计及含有潜在变量的路径分析。利用结构化方程的路径系数、系数的显著性和模型的 R^2 值进行假设验证。根据结构化方程模型路径分析结果，模型 R^2 值为 0.82，说明本次建立的模型可解释影响用户信息接纳行为的 82%因素。中心路径的标准化路径系数是 0.49（P<0.05），说明中心路径的研究变量可正向影响微信公众平台用户的信息采纳行为。边缘路径的标准化路径系数是 0.47（P<0.05），说明边缘路径的研究变量可正向影响微信公众平台用户的信息采纳行为。结果见表 3。在中心路径中文章的内容及文章是否原创的因子载荷较大，说明在中心路径因素中文章的内容及文章是否原创对用户的信息采纳行为影响较大。在边缘路径中，文章发送的时间及标题的趣味性因子载荷较大，在边缘路径中两者对用户的信息采纳行为影响较大。见图 1。

表 3 研究假设检验结果汇总表

研究假设	模型路径	方向	标准化路径系数	T 值	*P*	是都被验证
H1	中心路径→信息采纳行为	正	0.49	2.38	0.017	是
H2	边缘路径→信息采纳行为	正	0.47	2.30	0.021	是

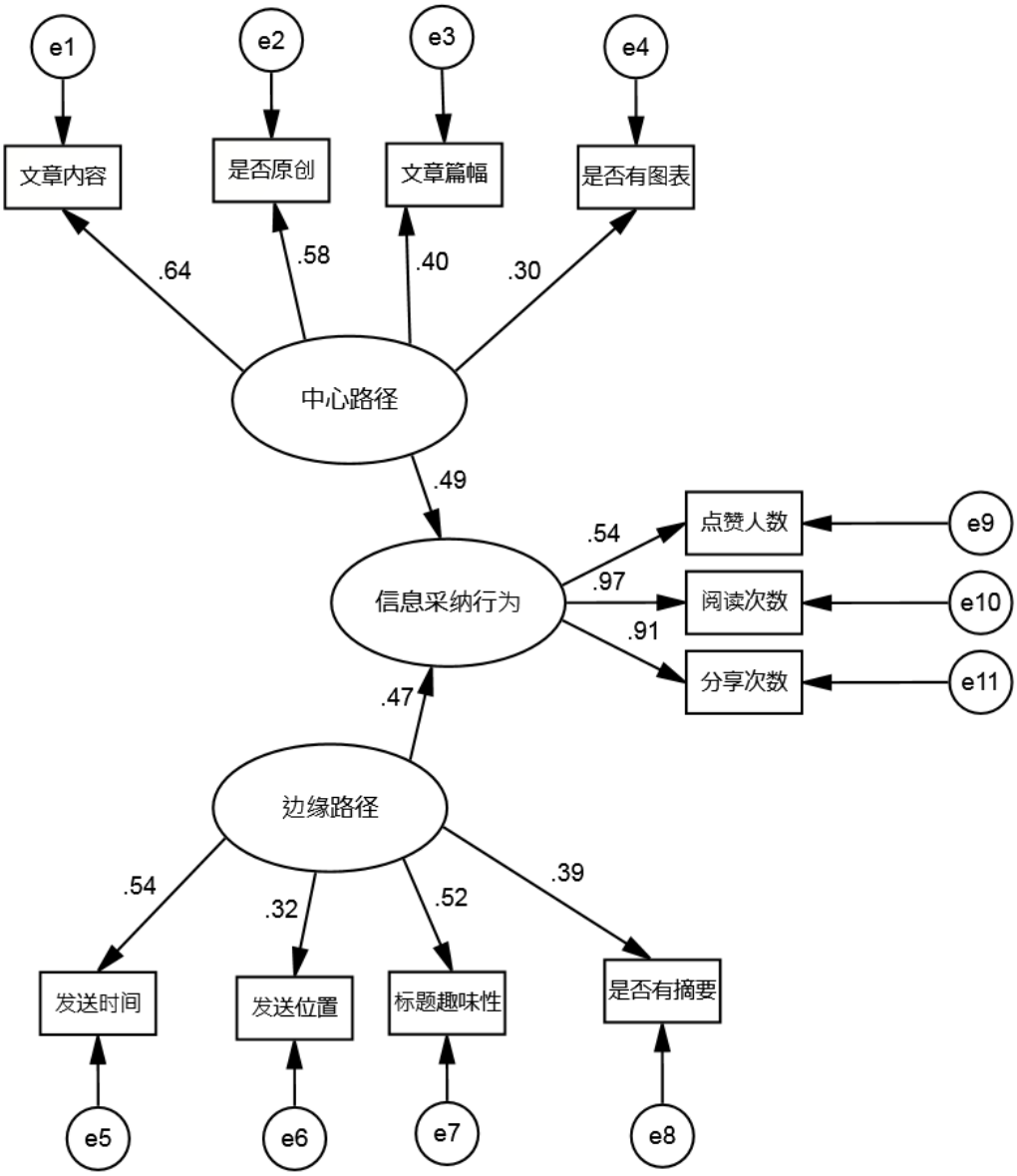


图 1 医院图书馆微信公众平台用户信息采纳行为影响因素模型

3 讨论

据十七次全国国民阅读调查结果显示，截止至 2019 年 4 月，成年国民数字化阅读比例高达 79.3%，微信、微博等全新的阅读媒介已经成为人们获取信息的重要途径[11]。微信公众平台以其传播范围广、内容形式多样及互动性强等优点成为了新媒体时代医院图书馆开展信息服务不可或缺的重要组成部分。如何高效的利用微信公众平台的阅读媒介作用，精准掌握用户的信息需求，成为了目前医院图书馆开展智慧服务的研究热点。本研究从用户角度出发，基于 ELM 模型构建医院图书馆微信公众平台用户信息采纳行为的影响因素模型并对模型进行数据实证，

分析结果发现中心路径及边缘路径均可正向影响用户的信息采纳行为，并且中心路径的因素影响更大。中心路径中微信文章的内容及文章是否为原创对用户的信息采纳行为影响较大，在边缘路径中文章发送的时间及标题的趣味性对用户的信息采纳行为影响较大。根据以上研究结果并结合医院图书馆微信公众平台研究现状，本研究提出以下改进措施和建议：

3.1 挖掘用户需求，优化信息内容

随着我国“双一流”建设和生物信息资源海量化发展，传统医学图书馆的文献传递、查收查引等服务已经无法满足医院科研用户的信息需求，知识服务、智慧服务成为了医院图书馆职能转变的方向[12]。本研究发现在中心路径中平台发布的文章内容对用户的信息采纳行为产生正向影响，因子载荷最大，平台用户对论文写作方法及投稿技巧、数据统计、专利分析等图书馆培训讲座需求较大。微信公众平台在内容管理方面，应以用户需求为导向，优化信息内容，加强专利知识、论文写作和数据分析等高品质的学术讲座培训的发布，采用线上直播和线下互动交流结合的形式，将图书馆的知识服务和智慧服务融入微信公众平台。

3.2 注重发文技巧，吸引读者眼球

在内容编撰方面，可借助热点事件或文化事件书写主题文章，开展专利知识月或论文写作月等主题活动，以主题模块形式推送文章，吸引读者兴趣。另外，应控制推送文章篇幅，做到语言简洁凝练、深入浅出。将数字、网络流行词以及排比句、疑问句等修辞手法运用到文章标题书写，吸引读者眼球。在推文形式方面，使用文字、图片、音频、视频等多种媒体形式，增强推文趣味性和可读性，提升用户的阅读体验，增加公众号文章的传播效果。

3.3 固定发文频次，增加用户粘性

用户粘性又称为用户的忠诚度，是指用户对某一服务产生好感， 形成依附性偏好， 进而重复使用的一种趋向[13]。公众平台可根据信息内容不同，固定文章的推送频率，尤其是图书馆举办的讲座培训，应固定文章推送时间，每周至少更新一次，增加用户粘性。另外有研究表明，用户喜欢在通勤过程中阅读学习型内容，平台可依据医院图书馆用户活跃时段调整推送时间，方便用户利用零散时间获取信息资源[1]。

3.4 塑造品牌效应，提升用户信任

原创能力是微信公众平台核心竞争力的重要体现，医院图书馆应该充分挖掘自身的特色，撰写如专利科普知识、论文写作投稿技巧、期刊影响因子预测等高质量的原创文章，以“微”见“效”提升公众平台的学术影响力，塑造品牌效应。此外，图书馆应结合医学专业特点，实时发布最新医学相关情报信息和热点新闻，增加用户的关注度。

OR-011

互联网医疗咨询平台使用意愿研究

童俞嘉、翟兴
北京中医药大学

目的 在这个信息时代，移动设备与互联网技术快速发展和普及，为应对人们日益增加的医疗咨询需求，线上医疗咨询服务应运而生，并开始得到更多的关注。尤其是在新型冠状病毒肺炎疫情期间，各互联网医疗咨询平台在短时间内获得了大量用户。本文结合信息系统成功模型和整合技术接受模型构建研究模型，研究互联网医疗咨询平台的使用意愿及其影响因素。

方法 以问卷调查采集数据，利用最小二乘结构方程模型，探究各种因素对公众使用互联网医疗咨询平台的影响因素。

结果 通过分析得出促进条件、社会影响、绩效期望和个人创新性对互联网医疗咨询平台的使用意愿有显著正向影响。

结论 第一，促进条件对用户使用互联网医疗咨询平台的意愿正向影响最强，这意味着用户的自身知识、技术以及周边辅助条件的支持对其采纳互联网医疗咨询平台的意愿有不小的影响；第二，社会影响作为显著正向作用于使用意愿的第二影响因素，其意义在于人们对于陌生事物的尝试动机极易受到身边人和周边环境的影响；第三，绩效期望对使用意愿也有显著正向作用，而信息质量又对绩效期望存在较强的正向作用。可见，从需方的角度来看，互联网医疗咨询平台对用户健康状态或疾病问题提供帮助的程度的确会影响用户的使用意愿。同时，从供方的角度出发，互联网医疗咨询平台需要提供更有价值的信息，提高平台的服务能力；第四，个人创新性作为个人因素对使用意愿具有显著正向作用。证实了用户对于新事物和技术的接受程度越高，其使用互联网医疗咨询平台的意愿就越强。

1 引言

当下，中国的综合国力到达了新的高度，国民各方面的物质需求充分得到满足和发展，人们开始更加关注疾病和健康相关的问题，但目前中国的医疗卫生资源尚且存在着分布不均的问题，在基层的医疗资源无法得到有效利用。

随着现代信息技术的高速发展，由互联网作为技术支持的医疗咨询平台成为解决医疗资源的供需矛盾提供新的解决方案。互联网医疗通过医疗健康信息的共享，优化医疗资源配置，以求满足人们的医疗健康需求，缓解医疗资源的供需矛盾。2019 年底，新型冠状病毒疫情爆发，互联网医疗咨询平台作为“安全医疗模式”发挥了重要作用。然而，目前其发展仍处于起步阶段，存在诸多方面还待加强。本文通过研究公众对于互联网医疗咨询平台的使用意愿，构建模型探究影响公众采纳互联网医疗咨询平台的因素并提出建议，以期改进目前的互联网医疗平台咨询模式。

2 理论和假设

2.1 相关理论模型

2.1.1 整合技术接受模型

整合技术接受模型[1]，全面地融合了多种技术接受理论领域的经典模型，最终该理论共有四个重要的潜在变量，它们分别是：绩效期望，努力期望，社会影响和促进条件。

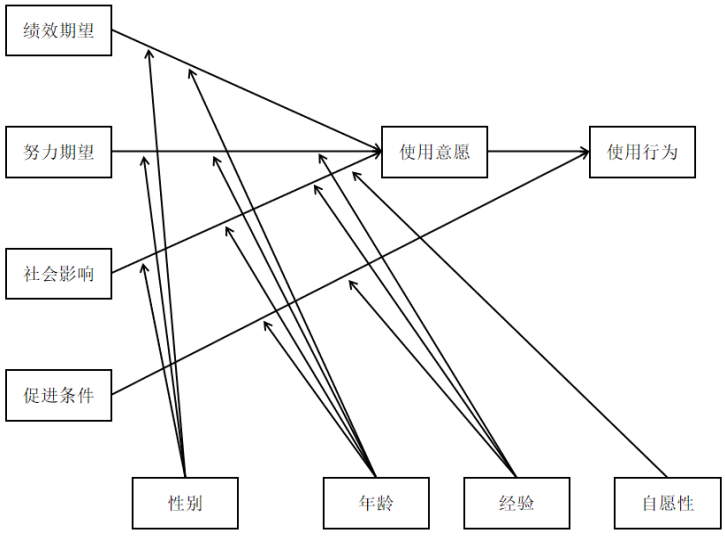


图 1 UTAUT 模型

2.1.2 信息系统成功模型

在 19 世纪末被第一次提出了信息系统成功模型，后于 20 世纪初得到改进。模型的影响因素经修改后确定为信息质量、系统质量、服务质量、使用意愿、用户满意和净收益，每个影响因素反映的内容各有不同。

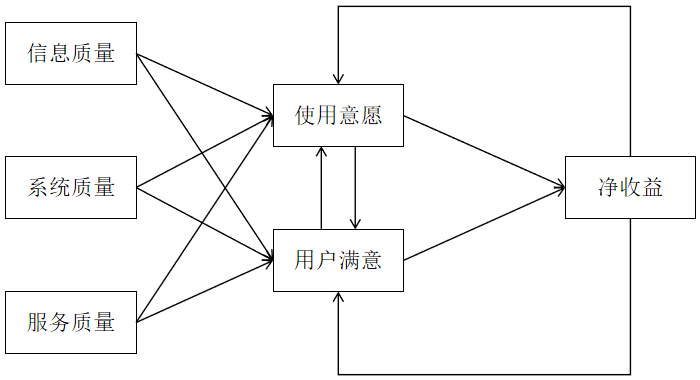


图 2　D&M 信息系统成功模型（2003）

2.2 研究模型构建

2.2.1 研究模型

整合技术接受模型和信息系统成功模型都是信息采纳相关研究的经典模型，但在具体应用的过程中各有侧重。前者侧重用户的使用感受，但对如何形成这样的使用体验缺乏分析的模式和途径；而后者侧重系统本身，在用户的使用特征的解释力不足。若将二者进行整合，以结合需求端和供给端，可使得研究过程和研究结果更加全面和立体。

因此，本次研究以公众对互联网医疗咨询的实际需求为出发点，以 UTAUT 理论模型为基础，整合信息系统成功视角，再结合个人影响因素构建研究模型，探讨互联网医疗咨询平台使用意愿的影响因素。如图 3 所示。

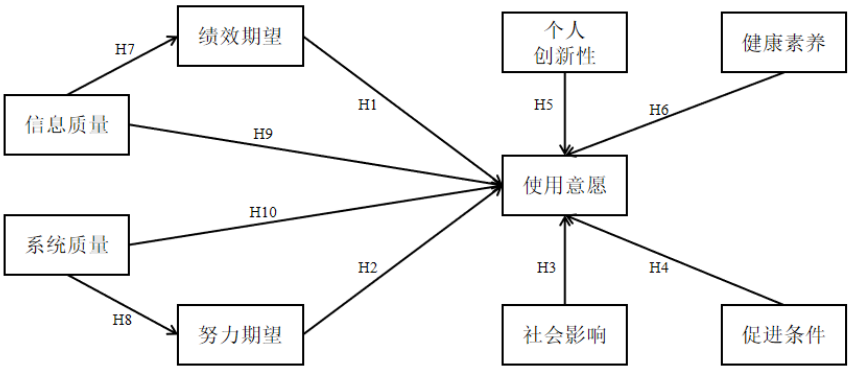


图 3　本文构建的研究模型

2.2.2 研究假设

2.2.2.1 基于整合技术接受模型（UTAUT）的假设

本文的模型主体为源于 UTAUT 模型的五个重要的内生变量，本篇文章的模型中，使用意愿（BI）是指公众使用互联网医疗咨询平台的意向的强烈程度。绩效期望（PE）是指使用互联网医疗咨询平台为公众提供便利和帮助的程度。为其提供的帮助越大，那么用户会更倾向愿意使用；努力期望（EE）的含义为用户感觉使用互联网医疗咨询平台来咨询医疗健康相关信息的难易程度；社会影响（SI）是指周围人群或者是有重要影响的人对互联网医疗咨询平台使用的支持程度；促进条件（FC）是指公众使用互联网医疗咨询平台能得到现有知识、技术和其他条件的支持程度：

假设 1（H1）：绩效期望（PE）对使用意愿（BI）有直接的正向作用

假设 2（H2）：努力期望（EE）对使用意愿（BI）有直接的正向作用

假设 3（H3）：社会影响（SI）对使用意愿（BI）有直接的正向作用

假设 4（H4）：促进条件（FC）对使用意愿（BI）有直接的正向作用

2.2.2.2 个人创新性的相关假设

个人创新性（PI）是指用户个人所具有的个性特征，反映个人率先接触和体验新事物的意愿，研究表明，当个体的个人创新性高时，也就意味着其更加倾向于接受新鲜事物并具有一定的新事物接受能力，这样的个体对于新技术更有可能具有较高水平的使用意愿。在本研究模型中将个人创新性定义为接受如互联网医疗咨询平台此类新事物的观念和内在倾向程度：

假设 5（H5）：个人创新性（PI）对使用意愿（BI）有直接的正向作用

2.2.2.3 健康素养的相关假设

健康素养（HL）的含义是用户为了达到提升或保持身心健康状态的目的，而收集和应用相关信息和服务的能力[2]：

假设 6（H6）：用户的健康素养（HL）对使用意愿（BI）有直接的正向作用

2.2.2.4 基于 D&M 信息系统成功模型的假设

在本篇文章中所应用的模型借鉴和整合了信息系统成功模型，将其中的两个潜在变量整合吸收，并作为研究技术供给端的两个方面。信息质量是对平台系统输出的信息质量开展的评估，通过准确性、时效性、相关性和完整性等来反映；系统质量是对平台系统的运行情况展开的评估：

假设 7（H7）：信息质量（IQ）对绩效期望（PE）有直接的正向作用

假设 8（H8）：系统质量（SyQ）对努力期望（EE）有直接的正向作用

假设 9（H9）：信息质量（IQ）对使用意愿（BI）有直接的正向作用

假设 10（H10）：系统质量（SyQ）对使用意愿（BI）有直接的正向作用

3 研究方法

3.1 问卷与检验

3.1.1 研究对象

本次研究的对象为疫情期间广大有卫生健康需求的公众群体，因而在设计问卷的过程中对于涉及本次研究的专业概念进行了通俗易懂的解释，在发放问卷的过程中尽可能包含更大的地域范围和更多的样本数据，并充分保证样本数据特征的均衡。

3.1.2 问卷结构

本研究的调查问卷设计分为：第一部分为受访者基本信息资料的调查；第二部分为与影响用户采纳互联网医疗咨询平台的个人因素的相关试题；第三部分为整合技术接受模型与信息系统成功模型的相关试题。此调查问卷的量表设计采用的是李克特的七级分度的量表设计方法，从“非常不同意”到“非常同意”，共有 9 个内在变量，每个内在变量有 3 - 4 个测试题项，此部分共 29 个测试题项。

3.2 问卷实施与数据统计

为了保证问卷调查回收到的样本数据能够充分有效、合理可信，本次问卷在具体实施的过程中，如图 4 所示，共分为九个步骤。通过网络渠道发放和回收问卷，共收到问卷 1368 份，经预调查问卷测试，问卷的填写时间一般在 120 秒到 500 秒以内，因而，本文在筛选有效样本时，只保留填写时间高于 100 秒且低于 800 秒，并非所有题目都填写相同数字和填写时间的答卷样本。经筛选最后保留有效问卷的数量为 906。表 1 为有效答卷样本的基本特征。

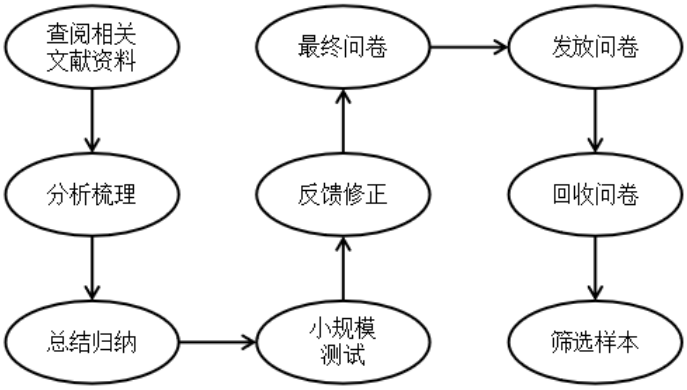


图 4　问卷实施过程

表 1　样本基本特征

指标	分类	频数	百分比
性别	男	407	44.9%
	女	499	55.1%
年龄	20 岁以下	301	33.2%
	21-35 岁	100	11.0%
	36-50 岁	437	48.2%
	50 岁以上	68	7.5%
月收入	3000 及以下	399	44.0%
	3000-5000	252	27.8%
	5000-10000	198	21.9%
	10000 以上	57	6.3%
受教育程度	高中及以下	560	61.8%
	大专	194	21.4%
	本科	142	15.7%
	硕士以上	10	1.1%

3.3 信度和效度检验

3.3.1 信度检验

本篇文章采用的整体信度检验指标分别为：内部一致性信度（Cronbach's α）、平均方差提取值（AVE）和组合信度（CR）[3]。

一般，Cronbach's α > 0.7 时认为整体量表的信度可以被接受，在 0.7 ~ 0.8 之间信度良好，在 0.8 ~ 0.9 之间信度非常好。当 AVE > 0.50，CR > 0.70 时，表明因子具有较好的可靠性。

通过 SmartPLS 2.0 的 PLS Algorithm 方法计算以上指标。如表 2 所示，各指标均高于一般标准，由此可见本次研究的问卷在结构上的可靠性和可信度水平较为理想。

3.3.2 效度检验

内容效度：本问卷所有题项是在参考同主题或领域实证研究的基础上分析总结后进行引用迁移，因而其内容效度较高。

建构效度：第一，收敛效度：即因子载荷 > 0.7，则说明问卷各题项的收敛效度较高；第二，区别效度：各潜在变量平均方差提取值的平方根值高于其与其他潜在变量之间的相关度，则区别效度较高[4]。由表 2 和表 3 可知，各个指标均高于标准，可见有模型的收敛效度和区别效度均较为理想。

表 2 测量题项及相关数值结果

变量	测量题项	测量题项内容	因子载荷	Cronbach's α	AVE	CR	来源
信息质量(IQ)	IQ1	我认为互联网医疗咨询平台能够真实准确地提供医疗健康方面的咨询信息很重要	0.898	0.925	0.816	0.946	[5]
	IQ2	我认为互联网医疗咨询平台能够提供形式多样、内容丰富的信息很重要	0.912				
	IQ3	我认为互联网医疗咨询平台提供的医疗健康咨询信息需要具备专业性和权威性	0.881				
	IQ4	我认为互联网医疗咨询平台能够及时有效地跟进时事热点并提供建议很重要	0.922				[6]
绩效期望(PE)	PE1	互联网医疗咨询平台能让我获得更多有用的医疗健康相关信息	0.914	0.907	0.843	0.942	[7]
	PE2	互联网医疗咨询平台能提高我或者周围人的健康水平	0.918				
	PE3	互联网医疗咨询平台使我更加方便地咨询医疗健康相关问题	0.923				
系统质量(SyQ)	SyQ1	互联网医疗咨询平台的界面设计应当简洁明了，条理清楚，使用灵活，容易上手	0.888	0.849	0.768	0.908	[8]

	SyQ2	互联网医疗咨询平台不应当存在垃圾广告和恶意链接，无安全问题	0.815				
	SyQ3	互联网医疗咨询平台应当集中整合资源，方便用户查询和访问	0.923				
	EE1	我很容易就能使用互联网医疗咨询软件或访问互联网医疗咨询平台网站	0.908				
努力期望(EE)	EE2	对我来说，互联网医疗咨询平台的使用方法简单易学	0.948	0.925	0.870	0.953	[9]
	EE3	我在互联网医疗咨询平台上操作起来十分轻松简单	0.942				
	SI1	若周围的家人朋友都在使用互联网医疗咨询平台，我也会尝试使用	0.867				
社会影响(SI)	SI2	对我有影响的人（家人、朋友等）的推荐对我决定使用互联网医疗咨询平台有较大影响	0.878	0.840	0.758	0.904	[7]
	SI3	媒体的推广和宣传对我使用互联网医疗咨询平台有较大影响	0.866				
	FC1	我具备使用互联网医疗咨询平台必备知识	0.835				
	FC2	我拥有使用互联网医疗咨询平台的技术条件，如智能手机等	0.843				[10]
促进条件(FC)	FC3	我在使用互联网医疗咨询平台的过程中遇到问题可以获得别人或来自平台的帮助	0.904	0.885	0.743	0.920	[11]
	FC4	我有时间和精力去了解和使用互联网医疗咨询平台	0.864				[12]
	BI1	今后我打算尝试使用互联网医疗咨询平台	0.938				
使用意愿(BI)	BI2	我很愿意向身边有需要的人推荐使用互联网医疗咨询平台	0.955	0.935	0.885	0.959	[7]
	BI3	我希望自己之后会常常使用互联网医疗咨询平台	0.929				
个人	PI1	我对新系统或新技术等有好奇心，想要	0.868	0.779	0.697	0.872	[7]

创新性(PI)		尝试					
	PI2	我愿意接受新事物，无论是在想法或行动上	0.906				
	PI3	与周围的家人朋友相比，我更早接触和使用互联网资源	0.718				
健康素养(HL)	HL1	我有意识并能够比较轻松地看懂药品说明书	0.852	0.849	0.767	0.908	[13]
	HL2	我能够理解各种媒体和宣传资料上健康问题的说明	0.919				
	HL3	当我对某种疾病感到困惑时，我知道如何获得该种疾病的相关信息	0.856				

表 3 测量模型区别效度矩阵分析表

变量	BI	EE	FC	HL	IQ	PE	PI	SI	SyQ
BI	0.941								
EE	0.658	0.933							
FC	0.818	0.758	0.862						
HL	0.470	0.514	0.527	0.876					
IQ	0.581	0.587	0.637	0.429	0.903				
PE	0.665	0.652	0.703	0.433	0.801	0.918			
PI	0.444	0.409	0.453	0.582	0.378	0.408	0.835		
SI	0.701	0.720	0.757	0.460	0.629	0.696	0.392	0.870	
SyQ	0.571	0.607	0.631	0.431	0.752	0.697	0.375	0.639	0.876

注：变量相关系数矩阵中对角线上的值为 AVE 的平方根。

3.4 模型验证

本文采用结构方程模型软件 SmartPLS 2.0 检验假设模型中对变量间因果关系的假设。根据检验结果调整各个变量以及其因果关系。

先进行模型的路径参数估计，然后使用 bootstrapping[14]对原始数据选取容量为 5000 的重抽样样本，检验其路径系数的显著性。

表 4 列出了本模型各路径的总体解释能力、路径系数和相关的 t 值。可以看出，有 PE -> BI、SI -> BI、FC -> BI、PI -> BI、IQ -> PE 和 SyQ -> EE 的因果关系得到了验证。如表 5 所示，

R2 说明了本模型的解释力。图 5 可见结果。实线路径代表差异显著，“*”表示 p<0.05，“**”表示 p<0.01，“***”表示 p<0.001；虚线路径代表差异不显著。

表 4 路径系数及其显著性

假设	路径	路径系数	t 值	支持与否
H1	PE -> BI	0.138	2.987**	是
H2	EE -> BI	-0.003	0.053	否
H3	SI -> BI	0.142	3.273**	是
H4	FC -> BI	0.601	11.547***	是
H5	PI -> BI	0.071	2.448*	是
H6	HL -> BI	0.001	0.026	否
H7	IQ -> PE	0.801	44.864***	是
H8	SyQ -> EE	0.607	19.309***	是
H9	IQ -> BI	-0.027	0.630	否
H10	SyQ -> BI	-0.000	0.001	否

注：*：p < 0.05， **：p < 0.01， ***：p < 0.001.

表 5 各内生变量的 R^2

内生变量	决定系数 R^2
绩效期望	0.641
努力期望	0.368
使用意愿	0.697

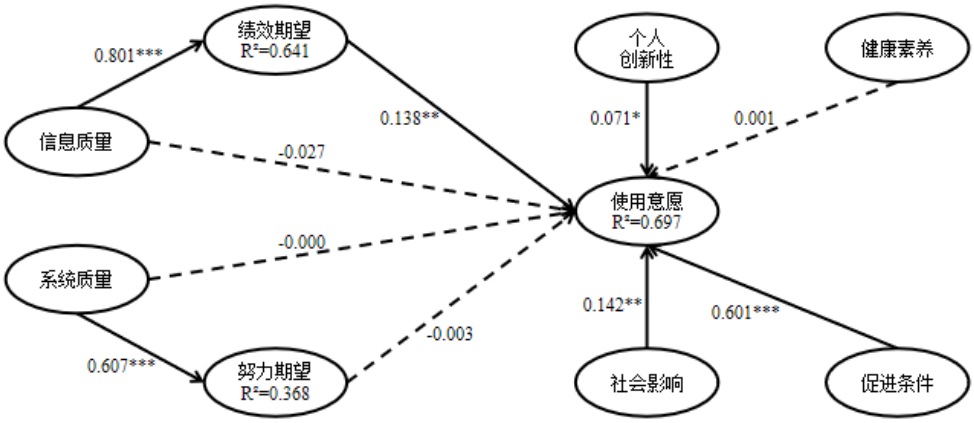


图 5 研究模型的检验结果

4 结果与建议

4.1 结果分析

促进条件、社会影响、绩效期望和个人创新性对互联网医疗咨询平台使用意愿的直接影响都被检验为显著且正向，其影响强弱程度排序为“促进条件 > 社会影响 > 绩效期望 > 个人创新性”。此外，信息质量和系统质量分别对绩效期望和努力期望有直接显著的正向影响。

第一，促进条件正向影响用户对互联网医疗咨询平台的使用意愿，在所有自变量中影响最强。这意味着用户的自身知识、技术以及周边辅助条件的支持对其采纳互联网医疗咨询平台的意愿有不小的影响；第二，社会影响作为显著正向作用于使用意愿的第二影响因素，其意义在于人们对于陌生事物的尝试动机极易受到身边人和周边环境的影响；第三，绩效期望对使用意愿也有显著正向作用，而信息质量又对绩效期望存在较强的正向作用。可见，从需方的角度来看，互联网医疗咨询平台对用户健康状态或疾病问题提供帮助的程度的确会影响用户的使用意愿。同时，从供方的角度出发，互联网医疗咨询平台需要提供更有价值的信息，提高平台的服务能力；第四，个人创新性作为个人因素对使用意愿具有显著正向作用。证实了用户对于新事物和技术的接受程度越高，其使用互联网医疗咨询平台的意愿就越强。

4.2 建议

根据本文的研究结果，笔者对如何进一步改进互联网医疗咨询平台的服务，使其为民众的医疗健康提供更多、更好的咨询服务给出如下建议：

1. 坚持正确思想领导，政策法规全面支持

公众对于互联网医疗咨询平台的重视和对其使用率的提升，需要从整个大环境上入手来进行改良，健全互联网医疗行业的法律法规：加强完善线上医疗咨询平台从业人员的吸纳和审核机制，充分调动全国医学人才资源加入线上的医疗队伍；大力促进和鼓励企业的研发和创新；完善政府和社会对互联网医疗供应者的监管机制；将线上医疗服务所产生的费用纳入医保范围。

2.改善使用环境，提升用户体验

制定科学合理的行业内规范流程以及详细精确的行业标准；做好信息发布以及安全防护的工作，保证信息的透明和安全；降低能够使用平台的设备要求，减少技术门槛；提供使用指导，辅助和引导用户使用；改进交互界面，优化使用流程；加大资源投入，以提升平台性能，加快响应速度。

3.目标人群牵头，结合全方位引导

充分发挥政府的引导作用，通过“两手抓”的策略来使宣传效果达到最大化，一方面，以个人创新性较强的群体作为主要目标人群来专攻，主要通过社群、自媒体、短视频、微博、知乎等新媒体渠道来进行线上医疗咨询的宣传；另一方面，进行广泛的全方位宣传引导，政府、社区和企业等多方共同参与，通过新旧媒体，如电视、广播、报纸、杂志、公告等方式来进行对公众的宣传引导。

4.用户需求至上，改进服务能力

以用户的需求为中心，设计高效精准的信息检索方式，从而快速准确地获知用户的诉求；提供丰富多样、合理有效的服务模式，以更加新颖和多样的信息形式和可视化方式来满足用户的不同需求；结合定期和不定期的用户调查来收集反馈用户的使用体验以及新的需求，通过整合分析进行改进；全面提升服务能力，多方位充分满足用户的需求，以切实增强用户的使用体验。

OR-012

基于结构熵权法的医院图书馆数字资源评价

庞兴梅[1]、田明慧[2]、庞兴峰[3]
1. 山东大学齐鲁医院 2. 山东大学计算机科学与技术学院
3. 山东省曹县机关事务管理局

目的 构建科学合理、符合医院图书馆实际的数字资源评价指标体系

方法 借鉴目前国际上与数字资源评价有关的标准，结合医院图书馆数据库的具体使用情况，依据评估指标的选取原则，设计了包括“利用情况、内容质量、检索系统、访问性能、使用情况、数据库商服务水平、价格”6 个一级指标和 33 个二级指标组成的医院图书馆数字资源评价体系，通过结构熵权法，确定各指标的权重；

结果 构建了各指标权重明确的医院图书馆数字资源评价体系；

结论 基于结构熵权法的医院图书馆数字资源评价体系指导性强，探讨了提高数据库利用率的方法，提出了医院图书馆资源建设的努力方向。

1 前言

随着信息技术和数字资源的蓬勃发展，各医院图书馆逐步形成了以数字资源为重心的信息资源建设格局。其中，数据库是最常见、功能强大的数字资源，有研究表明，图书馆拥有的数据库越多、类型越丰富、文献质量越高，图书馆电子信息服务能力就越强[1]，各个数据库资源各有特色，来源和内容相对独立，比如，常见的中文医学资源如同方知网的 CHKD 数据库、万方医学网、Sino Med、人卫临床知识库各有侧重，多数三甲医院图书馆都会面临遴选的困惑。同时，外文数据库是我国学习和引进国外最新技术及第一时间了解科技信息动态的信息库，在高校和专业图书馆建设中所占分量越来越重。然而，目前电子资源价格逐年攀升，特别是外文数据库价格昂贵，每年价格上涨的幅度通常要高于图书馆电子资源采购经费的增长幅度。如何在采购经费有限甚至削减的情况下，合理配置馆藏电子资源，构建强有力的资源保障体系是各医院图书馆资源建设中面临的主要问题。因此，对图书馆数字资源进行科学有效的评价成为了一项重要的课题，这对于医院图书馆调整电子资源采购策略、优化电子资源结构具有重要的指导意义。

2 数字资源评价体系的构建

2.1 数字资源评价体系的构建原则：

数字资源评价体系的构建依据目标导向性原则。医院图书馆的数字资源体系要为医院临床科研的发展提供强大的信息保障，要遵循全面系统而又重点突出的原则，一般要求中外文数据库并举，电子期刊、电子图书和二次文献数据库类型齐全，既能达到较高的文献保障率以满足大多数医疗科研人员群体的需求，同时又能契合医院重点学科发展的需求。同时，评价指标的选取按照可比性、可度量性、可获取性等原则，尽可能构建定性和定量相结合、科学合理的评价体系。

2.2 数字资源评价体系的现状

为科学合理地评价图书馆资源建设的效果，国内外已有众多学术机构和学者试图构建有效的馆藏评价标准和评价指标体系。美国研究图书馆协会为电子资源计量设计了 19 个评估指标，主

要有数字资源的可获取性、使用、成本、图书馆数字化工作、绩效评估五大指标[2]；中国高等教育文献保障系统（CALIS）对数字资源的数量和规模、内容与质量、体系与结构优化等方面进行评估[3]；马晓亭认为学术数字资源采购的核心评估标准应包括数据库服务成本、数据库可靠性、读者利用数三个方面[4]；吕旭霞从用户需求出发将电子资源的指标体系划分为用户需求的保障程度、支持程度、利用程度、满足程度和价格成本五大类[5]；刘爽从电子资源的内容、使用统计、价值与成本核算、检索系统及功能、数据提供商的服务、存档等方面构建了图书馆电子资源评价指标体系[6]。

教育部高校图工委于 2007 年颁布了《普通高等学校图书馆馆藏评价指南》（以下简称《指南》）[7]，将电子资源评价区分为基本情况评价和绩效评价两类，将电子文献基本情况评价分为“七个一级指标：文献的内容、检索系统及功能、访问性能、供应商服务、试用情况、价格因素、存档”。同时，将电子文献绩效评价概括为六个指标，包括“浏览/下载文献数量、单次使用成本、目标读者使用情况、引文分析与读者评价、图书馆员评价、影响绩效评价的几个因素”。《指南》中进一步指出“图书馆可根据每个指标的重要程度和评估目的，自行确定权重系数”。

2.3 评价体系模型设计

借鉴目前国际上与数字馆藏评价体系有关的一些主要项目，结合笔者所在的山东大学齐鲁医院图书馆（以下简称“我馆”）数字资源建设的实际经验，并在医院图书馆领域征询了意见，初步设计了医院图书馆数字资源评价体系的模型，。笔者认为数字资源的评价主要体现在内容质量、检索系统、访问性能、利用情况、成本价格、数据商提供的服务六个方面。在每个一级指标下，都有多个二级指标。如表 1。

表 1：医院图书馆数字资源评价指标

一级指标	二级指标	指标释义
内容质量	学科覆盖率	考察覆盖的临床二级学科的数量以及与医院重点学科的匹配程度
	文献类型和数量	数据库包含的期刊、图书、视频、图表和病例等类型文献及其数量
	收录年限	电子期刊、图书和二次文献的收录年限
	数据来源	考察数据是否来源于权威出版机构、专业学会等
	核心期刊收录率	考察外文文献中被 SCI、SSCI、A&HCI、EI 收录的期刊比例或中文文献中被北大核心、CSCD 收录的期刊比例
	更新周期和时滞	电子文献的更新频率，电子全文相比出版时间的滞后程度
检索系统	检索功能	是否提供可满足不同层次需求的检索功能如基本、高级、专业检索等
	查全率	能否提供满足查全需求的检索技术如分类、模糊、簇性检索等
	查准率	能否提供满足查准需求的检索技术如精确、限定、下位词检索等
	医学主题词检索	是否提供主题词检索途径、 Mesh 词是否规范

	检索界面	检索界面是否友好，是否进行资源整合提供跨库检索
	检索结果处理	检索结果是否提供处理、保存和链接功能
	个性化功能	提供个性化服务的内容和质量如定题推送等
访问性能	访问速度	在相同外界条件、相同时间测试不同数据库间的访问速度
	院外访问权限	是否提供院外访问功能，要求等同于医院 IP 地址内权限
	访问失败比例	考察访问失败的比例影响数据库正常使用的程度
	并发用户限制	并发用户数量的限制，是否影响数据库正常使用
利用情况	检索次数	一定时间段内如每月/每季度/每年的检索次数
	浏览次数	一定时间段内如每月/每季度/每年的阅览网页的次数
	下载次数	一定时间段内如每月/每季度/每年的全文下载次数
	目标读者使用率	目标读者中使用目标电子文献人数/该电子文献的目标读者人数
	单次文献使用成本	正式购买的数据库每年成本/读者全年使用次数
	用户满意度	使用二维码匿名问卷调查，采用梯度评分量化满意度，
	参考咨询馆员评价	对数据库各方面性能进行定量评价，特别是内容质量和检索系统
成本价格	价格和优惠力度	电子文献的报价和折扣幅度
	每年上涨幅度	电子文献价格的年涨幅
	镜像或永久访问权	是否提供本地镜像或停订后能继续使用已购买的数据
数据商提供的服务	访问数据统计报告	按月提供规范的使用统计报告
	宣传培训	数据商提供专业讲师培训及相关培训材料
	进科室精准培训	数据商派讲师到临床科室分专业精准培训
	解决问题的时效	对用户提出的建议和问题的反馈速度及解决问题的效果
	功能的改进	是否能根据用户需求不断改进各种功能和服务
	提供全文 URL 列表和 MARC 数据	是否为图书馆提供外文全文资源的 URL 列表和 MARC 数据

3 确定权重

3.1 方法

在评价的过程中，需要确定各个指标的权重。到目前为止，确定权重的方法有几十种之多[8-13]。这些方法根据数据来源方式的不同，可以大致分为主观赋权法和客观赋权法两类。其中，主观赋权法主要是专家根据实际经验来主观判断确定权重，例如层次分析法、德尔菲法、比较加权法等，该方法具有较强的解释力，但是客观性较差，具有一定的盲目性；客观赋权法的数据由各评价维度在对象中的现实数据集成，例如主成分分析法、离差最大化法、熵值法等，该方法切断了权重

主观性的数据来源，在大多数情况下精确性较强，但是有时所得权重与实际各指标的重要程度相悖，对于结果很难做出恰当解释，同时也不能体现决策者的建设策略。

3.2 结构熵权法

鉴于主客观赋权法各有优缺点，程启月（2010）[14]在充分分析主客观赋权法优劣势的基础上，给出了主观赋权法和客观赋权法相结合的“结构熵权法”。结构熵权法的基本思想是:通过分析系统指标及其相互关系，分解为若干个独立的层次结构，将采集专家意见的德尔菲专家调查法与定量分析相结合，对指标的重要性形成典型排序，用熵理论对典型排序结构的不确定性进行定量分析，计算出熵值和盲度分析，对可能产生潜在偏差的数据进行有效的统计处理，确定出每一层次同类指标重要程度的数值，即指标的权重，最终得出所有指标的权重。

笔者采用结构熵权法来确定医院图书馆数字资源综合评价体系中各指标的权重。

3.3 结构熵权法确定指标权重的实例

3.3.1 建立某指标下的重要性排序矩阵

根据德尔菲法专家选择的标准，即所选专家应具有较高的权威性和足够的时间耐心，人数一般限制在 10 ～ 50 人之间，选择来自山东省 10 家三甲医院图书馆的馆长或从事数字资源采访的专家，通过电子邮件的方式发放调查问卷，请专家根据自己的专业知识、经验，独立地对测评指标集的重要性进行排序打分。通常情况下认为某指标处于“第一选择”，则打分为 1；如果认为是“第二选择” ,分值则为 2； 其它依此类推。

设有 k 个专家对 A 、B 、C 、D 、 E、 F 等 n 个调查指标进行排序，由 k 张表获得指标的典型排序矩阵记为 A $(A_{ij})_{k*n}$，其中 A_{ij}表示第 i 个专家对第 j 个指标的评价。以“访问性能”为例，包括 4 个二级指标（访问速度、院外访问权限、访问失败比例、并发用户限制），请 5 位专家对“访问速度”等 4 个二级指标的重要性进行排序打分。打分结果如表 2 所示。

表 2 专家对“访问性能”重要性排序评价表

	访问速度	院外访问权限	访问失败比例	并发用户限制
专家 1	2	1	3	4
专家 2	1	2	4	3
专家 3	3	1	4	2
专家 4	2	1	4	3
专家 5	3	1	4	2

通过上述评价分数，可建立该指标下的典型排序矩阵。

3.3.2 计算结构熵值

定义定性排序转化的隶属函数为

$$\mu(I) = -\frac{\ln(m-I)}{\ln(m-1)}$$

I=A_{ij} 为专家按照“典型排序”的格式对某个指标评议后给出的重要性排序数值，I={1，2，3...j，j+1}，j 为实际最大顺序号，即参加排序的指标数量。m 为转化参数量，取 m=j+ 2。μ(I)是定义在[0 ,1] 上的变量，为 I 对应的隶属函数值，维度重要性越大，其取值越靠近 1。

此例中，4 个指标参加排序，最大顺序号取值为 4，j=4，m=4+2=6。

$$\mu(I) = -\frac{\ln(6-I)}{\ln 5}$$

分别计算出 A（A_{ij}）$_{5*4}$ 的隶属度，令b_{ij}=μ(I)，称为排序数A_{ij}的隶属度,形成隶属度矩阵 B（b_{ij}）$_{k*n}$。

$$B=(b_{ij})_{5*4}=\begin{bmatrix}0.86 & 1 & 0.68 & 0.43\\ 1 & 0.86 & 0.43 & 0.68\\ 0.68 & 1 & 0.43 & 0.86\\ 0.86 & 1 & 0.43 & 0.68\\ 0.68 & 1 & 0.43 & 0.86\end{bmatrix}$$

3.3.3 计算认识盲度和总体认识度

对于每一个因素 μ_j，视 k 个专家对指标 μ_j 的"话语权"相同，即计算 k 个专家对指标μ_j 一致看法，称为平均认识度，记作b_j， b_j=（b1j+b2j+...）/k，计算各个专家对于访问性能各二级指标的平均认识度，计算得出b_j={0.816,0.972,0.48,0.702}

定义专家对因素 μ_j 由认知产生的不确定性，称为认识盲度，记作Q_j，

$$Q_j = |\{[\max(b_{1j}, b_{2j}, \dots, b_{kj}) - b_j] + [\min(b_{1j}, b_{2j}, \dots, b_{kj}) - b_j]\}/2|$$

5 位专家对访问性能各二级指标的认识盲度记为：Q_j=[0.024,0.042,0.075,0.057]

定义 k(参加测评的全体专家数）个专家关于 μ_j 的总体认识度记作χ_j

$$x_j = b_j(1-Q_j),\quad x_j > 0$$

5 名专家对访问性能的总体认识度为 χ_j =[0.7964,0.9312,0.4440,0.6620]

3.3.4 归一化处理

为了得到访问性能 4 个二级指标的权重，需要对总体认识度进行归一化处理，得到向量α_j：

$\alpha_j = x_j / \sum_{i=1}^{m} x_j$

$$\alpha_j = [0.2810,0.3286,0.1567,0.2336]$$

通过以上典型矩阵的确定，以及“盲度分析”过程，得到的 α_j 即为 5 个专家对于“访问性能”的 4 个二级指标确定的权重，该值是专家的一致性判断，符合这些专家的群体意愿和认知，可以

看出其中第二个指标即“院外访问权限”权重值最大（0.3286），其次是“访问速度”的权重较大（0.2810）。

3.4 完成评价体系

将相同的方法应用到其它一级指标，可以计算出评价体系中所有二级指标的权重值，最终形成一个完整的医院图书馆数字资源评价体系。从我馆计算出的指标权重数值看，核心期刊收录率、检索功能、院外访问权限、下载次数、价格和优惠力度、进科室精准培训等指标的权重较大，从33个二级指标中脱颖而出，在各自一级指标中最为重要，和山东省多家医院图书馆资源建设的现状基本相符。期望该评价体系能对医院图书馆的资源建设起到实际的指导作用。

4.优化数字资源建设、藏以致用的思考和建议

4.1 广泛调查，确定需求

为了获得更多用户的真实需求，优化馆藏结构，建议采用二维码匿名调查表、座谈、采访等方式调研医院内临床专家、科研学者和医护人员等多个群体。问卷采用利克特五级量表设计，调查中对每一个问题的回答方式可分为“非常不同意”“不同意”“一般”“同意”“非常同意”五种，每项回答尽可能量化如依次赋值从“1”到“5”分，以真正得到用户在工作学习中的实时需求，提高资源采购工作的效率和质量。

4.2 医学信息素养培训教育

为使广大读者熟悉数据库，提高数据库使用技能，医院图书馆可联合数据商进行多种形式的宣传培训活动，如我馆每年度举办“数字资源使用系列培训讲座”、“中文数据库宣讲月活动”、“外文数据库宣讲月活动”等，每年平均在20场次，同时，主动走进临床科室，根据目标读者不同的学科和专业进行精准培训，图书馆提供电话、微信群、QQ群等多种途径的实时指导服务，以提高数字资源的使用率。

4.3 提供移动访问和院外访问服务

结合医院临床一线人员工作繁重、临床HIS系统与互联网必须隔离的现状，医院图书馆必须为职工提供多样化的移动服务和多媒体服务和数字信息资源的远程访问权限，VPN技术可以实现在医院外远程访问图书馆数字资源，方便职工随时随地获取图书馆的资源服务[15]。

4.4 医院图书馆根据医院的重点学科建设和重要专业项目，积极自建特色数据库和学科数据库，开展资源共享或共建，以更大范围、更高效率的全文传递和馆际互借服务来降低昂贵外文数据库的单次使用成本。

4.5 联合采购

一定地域内的多家医院图书馆可考虑组织团购，如山东省医院图书馆委员会在2019年组织全省的会员单位联合采购某一数据库，争取了较大的优惠力度，降低购买价格，从而降低使用成本。

OR-013

基于SPO语义网络的民族医药知识发现

杨雪梅、王雪、关陟昊、唐小利
中国医学科学院医学信息研究所

目的 我国民族医药发展中仍存在研究基础薄弱、缺乏科学规范的有效性和安全性依据的问题，在“一带一路”传统医药发展推广中存在诸多制约，为充分了解民族医药的疗效、作用机制等，提出基于SPO语义网络的民族医药知识发现方法。

方法 本文基于SPO语义网络和“药物-蛋白、基因-疾病”药物作用机制发现路径，通过限定实体概念类型及语义谓词获取与药物治疗、作用机制、副作用有关的SPO结构，实现对药物的治疗、作用机制、副作用的分析和研究。

结果 通过分析发现，虫草素及虫草多糖是目前研究者较为关注的有效成分，治疗对象较多，但主要涉及虫草素在肿瘤中的治疗应用；虫草的副作用主要包括对细胞的损害，其中虫草素导致的细胞调亡是目前研究者较为关注的内容。而在药物作用机制分析中，通过严格限定实体概念类型及语义谓词可以明确药物通过何种“中介”药物或基因发挥作用，从而影响疾病症状。

结论 本文提出SPO（Subject-Predicates-Object）语义网络方法，围绕药物效应、作用机制及药物副作用开展知识发现，为民族药研究利用及安全性评价提供支撑。

民族医药指中国少数民族的传统医药，是中华传统医药学重要的组成部分。各族人民在长期与疾病作斗争的实践中，积累了丰富的医疗经验和知识，创造了多种多样、独具特色的民族医药。民族医药的发展不仅可以让我国传统医药惠及更多人民，同时也是对民族传统知识文化的认知和尊重。但我国民族医药发展中仍存在研究基础薄弱、缺乏科学规范的有效性和安全性依据的问题，在“一带一路”传统医药发展推广中存在诸多制约，因此充分了解、研究民族医药的疗效及作用机制、药物副作用等，增强药物的科学规范性对民族医药的现代研究与二次开发具有重要意义。本文提出SPO（Subject-Predicates-Object）语义网络方法分析药物效应、作用机制及药物副作用，为民族药研究利用及安全性评价提供支撑，促进我国民族药国际化发展。

1 相关研究

药物知识发现目前多围绕药物疗效、用药规律、药物作用机制、药物副作用等方面开展。药物疗效是指药物改善病人的生理生化功能或病理过程，从而患病的机体恢复正常的作用结果。药物作用机制主要是研究药物如何起作用，药物对机体发挥作用是通过干扰或参与机体内的各种生理和生化过程的结果，因此药物的作用机制是多种多样的。药物副作用是指应用治疗量的药物后所出现的治疗目的以外的药理作用，副作用带来的危害越来越大，由它引起的病患占据所有医院病患的 5%，药物副作用已成为导致医院死亡的第五大原因[1]，因此药品安全及药物副作用被国内外专家学者广泛关注。

针对药物知识发现的研究方法多采用统计、机器学习、复杂网络理论等支持大规模知识计算的方式，以便从海量信息中挖掘潜在的药物隐形知识。

Patki[2]等人结合文本语义内容分析的方法，对药物副作用进行预测。分析的步骤包括对文献进行分类和标记、内容实体识别、药物副作用关系提取和潜在副作用分析等。该研究是基于普通的文本挖掘方法，一是需要大量人工参与，二是对于内容实体的识别准确性较低，不适用于大规模的药物数据挖掘和预测。李志鹏[3]把深度学习引入到药物隐含性质的研究工作中，利用卷积神经网络、Word2Vec、LDA 话题模型等手段并综合利用多个药物相关的数据库与工具去挖掘药物不良反应以及潜在的药物治疗目标。刘然[4]采用机器学习的预测算法来为各组药物进行建模，分析各个高血压的给药治疗方案及用药规律。魏巍[5]（魏巍. 药物不良反应知识发现与利用模型研究[D].）将自然语言处理、机器学习、数据挖掘、人工智能、语义分析方法应用到基于社交媒体的药物不良反应知识发现与利用过程中，实现社交媒体大数据的过滤、药物不良反应信号的抽取、药物不良反应本体的半自动构建和个性化药物不良反应自动生成。

在药物知识发现中诸多学者的方法还会涉及语义网络构建或网络分析。王雪[6]利用 SemRep 提取文献中 SPO 结构进而构建药物矛盾关系，发现并分析了学术文献中已发表的矛盾性知识并揭示其形成的特点。Pang K[7]通过 Drugbank 数据库构建药物-靶点相互作用网络来识别药物组合，利用药物综合相似性信息来预测药物相互作用。Lounkine[8]等设计了一个关联指标，来判断新的脱靶靶点之间的关联性，从而构建了一个药物靶点-药物副作用网络，基于该网络和药物化学结构来预测一个化合物是否会与脱靶靶点结合。

2 研究方法

本文利用 SemRep 抽取文献中的实体概念和语义关系，根据不同类型的 SPO 结构展开的药物效应、作用机制及药物副作用分析，基于 SPO 语义网络进行民族医药知识发现的技术路线图如图 1。

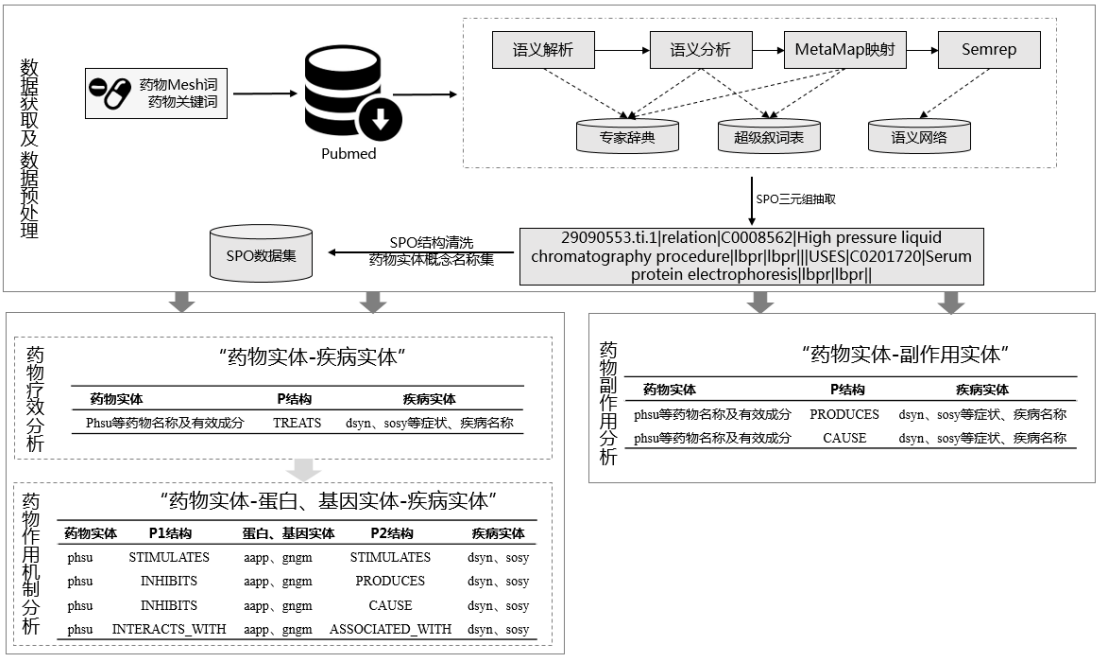


图 1 技术路线图

2.1 数据获取及数据预处理

2.1.1 数据获取

本文基于民族医药 Mesh 词及关键词，利用 Pubmed 医学文献数据库检索民族医药文献，并通过 Python 编程将文献数据转换为 SemRep 抽取 SPO 结构所需格式，提取能反映研究内容、方法、结论等重要信息的标题和摘要字段作为实体概念和关系抽取的数据源。

2.1.2 SPO 结构抽取及清洗

MetaMap[9]是一个将生物医学文本与 UMLS 超级词表中的概念进行匹配的工具，在生物医学实体识别中广泛使用。SemRep 是一个基于规则自动从文献中识别关系预测的自然语言处理系统[10]，其集成了 MetaMap 中规范化的概念实体，并运用谓词关系连接不同的实体概念[11]。SemRep 提取的关系最终按照一体化医学语言系统（Unified Medical Language System，UMLS）的规则进行输出，结果中包含多种字段信息，抽取结果如下：

----- Citation 29090553 -----

29090553.ti.1 Simultaneous determination of cordycepin and 2'-deoxyadenosine in Cordyceps genus by online SPE-HPLC.

29090553.ti.1|relation|C0008562|High pressure liquid chromatography procedure|lbpr|lbpr|||USES|C0201720|Serum protein electrophoresis|lbpr|lbpr||

本文根据 SemRep 定义的语义类型，在数据分析中选择实体概念名称、对应概念的语义类型和谓词关系开展研究。获取 SPO 结构数据后，对缺失实体概念名称或表述无意义的无效数据剔除。同时为方便开展后续研究分析，对民族医药的药物名称、药物主要成分进行分析遴选，明确研究实体概念名称集。

2.2 药物疗效分析

基于 UMLS 语义类型对抽取 SPO 结构进行筛选和分析。S 结构部分选用 Pharmacologic Substance（药理性物质 phsu）等药物名称及有效成分作为药物实体；O 结构部分选用 Disease or Syndrome（疾病与症状 dsyn）、Sign or Symptom（体征或症状 sosy）等症状、疾病名称作为疾病实体；P 结构部分，限定语义关系为 UMLS 中包含 TREATS（治疗）的谓语结构。根据以上实体概念及关系类型的限定进行 SPO 结构的筛选，找出与药物和治疗疾病相关的 SPO 结构，开展药物疗效分析。

2.3 药物作用机制分析

药物作用机制分析中根据 UMLS 语义类型对抽取 SPO 结构构建“药物实体-蛋白、基因实体-疾病实体”的关联关系。药物实体选用 Pharmacologic Substance（药理性物质 phsu）等药物名称及有效成分；蛋白、基因实体选用 Amino Acid, Peptide, or Protein（氨基酸、肽或蛋白质，aapp）；疾病实体选用 Disease or Syndrome（疾病与症状 dsyn）、Sign or Symptom（体征或症状 sosy）等症状、疾病名称。关联关系的 P 结构限制为“STIMULATES- STIMULATES”、“INHIBITS- PRODUCES”、“INHIBITS-CAUSE”、“INTERACTS_WITH- ASSOCIATED_WITH”四组语义关系对。根据以上实体概念及关系类型的限定进行 SPO 结构的筛选，根据关联关系开展药物作用机制分析。

表 1 “药物实体-蛋白、基因实体-疾病实体”的关联关系对

药物实体	P1 结构	蛋白、基因实体	P2 结构	疾病实体
phsu	STIMULATES	aapp、gngm	STIMULATES	dsyn、sosy
phsu	INHIBITS	aapp、gngm	PRODUCES	dsyn、sosy
phsu	INHIBITS	aapp、gngm	CAUSE	dsyn、sosy

phsu	INTERACTS_WITH	aapp、gngm	ASSOCIATED_WITH	dsyn、sosy

2.4 药物副作用分析

基于 UMLS 语义类型，结合 SIDER（Side Effect Resource,副作用资源）中的副作用词表对抽取 SPO 结构进行筛选和分析。S 结构部分选用 Pharmacologic Substance（药理性物质 phsu）等药物名称及有效成分作为药物实体；O 结构部分选用 Disease or Syndrome（疾病与症状 dsyn）、Sign or Symptom（体征或症状 sosy）等症状、疾病作为副作用实体；P 结构部分限定语义关系为 UMLS 中包含 PRODUCES（产生）和 CAUSE（引起）的谓语结构。根据以上实体概念及关系类型的限定进行 SPO 结构的筛选，找出与药物和药物副作用相关的 SPO 结构，分析药物的副作用。

3 实证研究

虫草是寄生于昆虫、少数真菌和植物体上的一类真菌，是广义虫草属真菌的总称，是具有营养、保健和医疗功效的宝贵生物资源。目前我国的虫草产业涉及了冬虫夏草、蛹虫草、蝉花及其相关真菌，初步形成产业，实现了较高的经济价值和社会效益。但在研究中尚存在不同学科对虫草的理解和认识存在分歧、虫草的基础研究不足、虫草副作用不明晰等问题[12]。因此以虫草民族医药开展基于 SPO 语义网络的药物知识发现研究。

3.1 虫草药物数据获取及数据预处理

通过制定检索策略(”Hirsutella sinensis”[Title/Abstract] OR ”Isaria cicadae Miquel”[Title/Abstract]) OR (Cordyceps[MeSH Terms])在 PuBMed 数据库进行检索，获取 971 条相关题录信息。利用 SemRep 语义关系提取工具从文献标题及摘要中提取 SPO，提取 5165 个 SPO 结构，经过数据清洗最终得到 5013 个 SPO 结构，涉及 54 种语义谓词、2309 个实体概念及基因。为了获取虫草的药物效应及副作用相关的 SPO，首先需确定虫草及有效活性成分的实体概念名称。不同虫草及其菌丝体的化学成分含量有差别，但是这些成分种类总体上保持基本不变，主要包括多糖类、核苷类、氨基酸类等[13]。目前研究最多、药理学功效最为显著的活性成分中有虫草多糖、虫草素、虫草酸以及虫草多肽[14]。已有研究证实这些活性成分在抗肿瘤、抗炎、抗菌、抗氧化、降血糖等多个方面具有显著的效果，被认为为临床应用提供了新的方向。而重要的虫草真菌包括枝顶孢属(Acremonium Link .ex Fr)、棒束孢属(Isaria Pers .ex Fr)、轮枝孢属(Verticillium Nees .ex Wallr.)、拟青霉属(Paecilomyces Main) [15]。从提取出的语义类型

为 Pharmacologic Substance、Fungus 的实体概念中，筛选出与虫草、虫草真菌及其有效成分有关的实体概念，进行后续的分析。虫草相关实体概念信息如表 2。

表 2 虫草、虫草真菌及其有效成分有关的实体概念名称及 CUI

Preferred name	CUI	Preferred name	CUI
Aspergillus fumigatus	C0004037	Metarhizium guizhouense	C1197375
Beauveria bassiana	C0320028	Ophiocordyceps gracilioides	C1467416
Cordyceps	C1011843	Ophiocordyceps heteropoda	C1042631
Cordyceps bifusispora	C1014110	Ophiocordyceps nutans	C1038767
Cordyceps cicadae	C1485146	Ophiocordyceps sinensis	C2313716
Cordyceps guangdongensis	C3715795	Ophiocordyceps sobolifera	C1087347
Cordyceps gunnii	C1045932	Ophiocordyceps sphecocephala	C1073371
Cordyceps kyushuensis	C1497384	Ophiocordyceps unilateralis	C1663704
Cordyceps militaris	C0319852	Ophiocordycipitaceae	C2295701
Cordyceps pruinosa	C1062524	Paecilomyces	C0030182
Cordyceps sp.	C4137597	Paecilomyces hepiali	C3012172
Cordyceps takaomontana	C1063393	Paecilomyces javanicus	C0319970
Cordycipitaceae	C2295702	Penicillium jiangxiense	C3946236
Isaria	C1029195	Adenosine	C0001443
Isaria cateniannulata	C3127832	caterpillar fungus	C0885247
Isaria farinosa	C1039585	cordycepin	C0056331
Isaria fumosorosea	C1054365	Polysaccharides	C0032594
Isaria takamizusanensis	C2631770		

3.2 虫草药物疗效分析

药物疗效分析通过限定 Subject 实体概念为任一表 2 中的实体，同时限定语义谓词为 TREATS、TREATS(SPEC)，Object 为语义类型是 Disease or Syndrome、Finding、Neoplastic Process、Pathologic Function、Sign or Symptom 的实体概念，从而获取与虫草治疗疾病相关的 SPO 语义关系。

共获取得到 18 个 SPO 语义关系，涉及 3 个与虫草相关的实体概念，18 个其他概念。如图 2 所示，其中绿色节点表示虫草的有效成分，红色节点表示疾病、症状等，节点之间的连线表示治疗关系，节点的大小表示涉及的 SPO 数量的多少。显然，虫草素及虫草多糖是研究较多的有效成分，且其治疗对象较多。其中虫草素多涉及肿瘤的治疗。

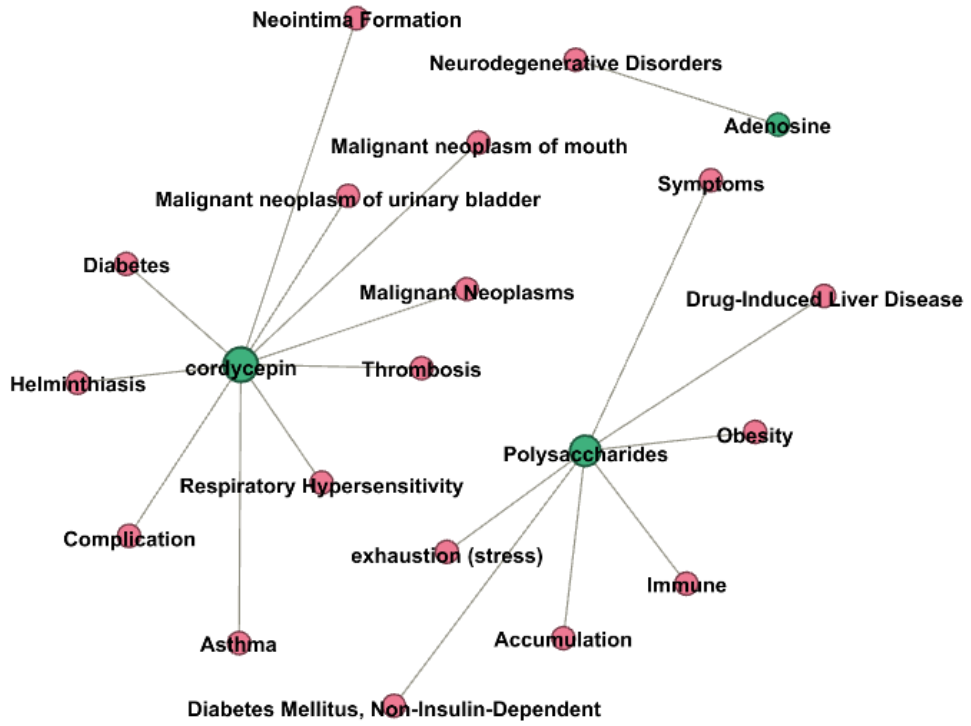


图 2 虫草及相关概念的治疗语义关系可视化

通过限定 TREATS(INFER)获取从原文中提取得到的推断的治疗关系，共 3 个 SPO。主要推断虫草素和腺苷的抗肝纤维化作用、冬虫夏草治疗慢性肾脏病。

表 2 虫草及相关实体概念的 TREATS(INFER)语义关系

PMID	Sentence	Subject concept	Predicate	Object concept
28594374	In this study, the antiliver fibrosis effects and mechanisms of ultrapure water-cultured CM (UCM), DOW-cultured CM (DCM), synthetic water-cultured CM, DOW, cordycepin, and adenosine were compared in the liver fibrosis mice induced by intraperitoneal injections of thioacetamide (TAA).	Adenosine	TREATS(INFER)	Fibrosis, Liver
25519252	BACKGROUND: Cordyceps sinensis (Cordyceps, Dong Chong Xia Cao), a herbal medicine also known as Chinese caterpillar fungus, is one of the most commonly used ingredients in traditional Chinese medicine for the treatment of people with chronic kidney disease (CKD).	caterpillar fungus	TREATS(INFER)	Chronic Kidney Diseases
31363726	The role of autophagy in the treatment of DN mice with CMP was detected by TEM and western blotting.	Polysaccharides	TREATS(INFER)	Diabetic Nephropathy

3.3 虫草药物作用机制分析

为了明确虫草及其有效成分如何作用于蛋白从而发挥影响，开展虫草药物的作用机制分析。基于表 2 的实体概念语义类型对 SPO 进行限制和筛选，获取虫草-蛋白和基因、

蛋白和基因-疾病之间关系，从而明确虫草药物作用途径。涉及虫草-蛋白和基因、蛋白和基因-疾病关系的 SPO 分别如下表 3 和表 4 所示。

表 3 虫草相关实体概念与蛋白、基因实体概念之间的语义关系

虫草相关实体概念	语义谓词	蛋白、基因实体概念
Cordyceps	LOCATION_OF	CD69 protein, human
cordycepin	STIMULATES	Interleukin-10
cordycepin	INTERACTS_WITH	Interleukin-10
cordycepin	INTERACTS_WITH	cytokine
cordycepin	STIMULATES	cytokine
cordycepin	STIMULATES	cytokine
cordycepin	STIMULATES	Interleukin-10
Cordyceps militaris	LOCATION_OF	Protein Component
Polysaccharides	STIMULATES	Ovalbumin

表 4 蛋白、基因实体概念与疾病实体概念之间的语义关系

蛋白、基因实体概念	语义谓词	疾病实体概念
Protein Component	AFFECTS	Fibrosis, Liver
Ovalbumin	CAUSES	Asthma
Interleukin-10	ASSOCIATED_WITH	Tuberculosis
cytokine	ASSOCIATED_WITH	Asthma
CD69 protein, human	AFFECTS	Pulmonary Fibrosis
CD69 protein, human	AFFECTS	Hypouricemia
CD69 protein, human	AFFECTS	Dermatitis, Atopic

基于表 1 进行严格的谓词限定，最终提取出两组“药物实体-蛋白、基因实体-疾病实体”语义关系。分别是：

①cordycepin *INTERACTS_WITH* Interleukin-10 *ASSOCIATED_WITH* Tuberculosis

②cordycepin *INTERACTS_WITH* cytokine *ASSOCIATED_WITH* Asthma

结合来源语句可以发现，虫草素对白介-10 具有调节作用，从而在结核分枝杆菌疾病通过白介-10 促进免疫调节；同时虫草素可以减少哮喘中不良的免疫反应和/或细胞因子的表达。

3.4 虫草药物副作用分析

利用 SIDER 中的副作用词表对提取出的 SPO 语义关系进行清洗和筛选，获取与虫草副作用有关的 SPO。首先通过限定 Subject 实体概念为任一表 2 中的实体，然后限定语义谓词为 PRODUCES（产生）和 CAUSE（引起）。

最终共得到 22 个与虫草副作用相关的 SPO 语义关系，共涉及 14 个实体概念，与虫草相关的共 4 个，分别为虫草素（cordycepin）、多糖类（Polysaccharides）、蛹虫草（Cordyceps militaris）及虫草（Cordyceps）。如图 3 所示，虫草的副作用主要涉及

对细胞的损害，包括细胞凋亡、细胞死亡、细胞周期停滞、胞质分裂等。其中涉及SPO数量最多的为cordycepin CAUSES Apoptosis，共涉及11个SPO，表明虫草素导致的细胞调亡是目前研究者较为关注的研究内容。

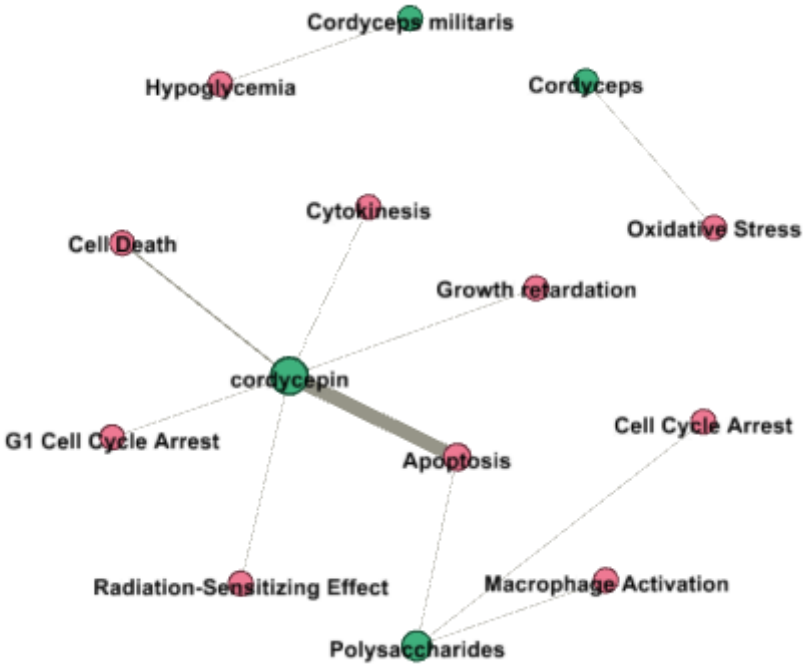


图3 虫草及相关概念的副作用语义关系可视化

4 结论与展望

本文基于SPO语义网络和“药物-蛋白或基因-疾病”药物作用机制发现路径，通过限定实体概念类型及语义谓词获取与药物治疗、作用机制、副作用有关的SPO，实现对药物的治疗、作用机制、副作用的分析和研究，并以虫草开展实证研究。通过分析发现，虫草素及虫草多糖是目前研究者较为关注的有效成分，治疗对象较多，但主要涉及虫草素在肿瘤中的治疗应用；虫草的副作用主要包括对细胞的损害，其中虫草素导致的细胞调亡是目前研究者较为关注的内容。而在药物作用机制分析中，通过严格限定实体概念类型及语义谓词可以明确药物通过何种“中介”药物或基因发挥作用，从而影响疾病症状。

后续研究中，基于SPO语义网络进一步开展知识预测，发现药物潜在治疗效果并预测药物副作用，为我国民族医药的二次研发及广泛应用提供支撑。

OR-014

Protégé 在构建中医药本体中的运用

杨郑子衿[1]、徐倩[1]、罗爱静[2]
1. 中南大学湘雅三医院
2. 中南大学湘雅二医院

目的 对已有中医药领域信息进行系统的知识表示，了解认识中医药本体的构建方法，优化对中医药信息数据来源的归纳整合，为后续知识推理及临床用药应用提供参考。

方法 本文参考中医药学语言系统(TCMLS)、《中国药典》和《中国中医药学主题词表》，利用protégé 本体编辑器构建中医药顶层框架，将方剂、中药材与中成药的概念内涵单独分析定义使中医药概念框架更为完善，重点扩展研究了方剂与人体及病因病机的知识表示，并辅以实例研究。

结果 完成以“事件”和“实体”为最顶层主要概念划分的中医药本体构建。

结论 为中医用药的应用研究提供知识推理框架，提高中医药领域知识信息的共享性和可用性。

1 引言

近年学者们在重视知识图谱与本体构建研究的同时，也在研究中医药领域的知识表示框架，例如目前最系统化的中医药术语标准化体系应为中医药学语言系统（TCMLS）、国家知识基础设施（National Knowledge Infrastructure, NKI）的中医药本体构建[1]、“中医药临床术语集（Traditional Chinese Medicine Clinical Terms, TCMCT）”等大型的术语系统。Xuezhong Zhou 以已有中医药知识体系为基础，研究开发支持中医语言知识检索、基于概念的信息检索和信息集成的中医语言系统（Unified Traditional Chinese Medical Language System, UTCMLS）本体应用程序[2]。

运用 protégé 软件构建本体是直观展示中医药知识体系的重要方式之一，本文以分析已有的中医药基本理论和知识来源为基础，通过 protégé 编辑器将中医药的知识进行系统的整理表达：定义知识语义、属性关联并扩展实例，实现中医药信息知识的共建、共享和重用。

2 中医药本体知识相关概念及联系

2.1 中医药本体概念

本体是共享概念模型的明确形式化规范说明，本体具有强语义特性，它能够以一种直观的、图形化的方式来展示领域知识，提高异构系统之间的互操作性，有助于知识共享[3]。由于中医药领

域的元数据存在完整性不足、模糊不清等问题，因此需要构建中医药物的本体，以便系统性的归纳整合数据，有利于领域知识的共享和重用。本体的知识表示过程是构建知识图谱的基础，在中医知识图谱的构建上，已有学者做出相关的理论研究和实际探索，目前医学中已有的本体数据有MeSH、OMIM、Gene Ontology、HPO 等[4]。

2.2 中医药本体构建知识来源

在本体构建方法上，目前研究多以中医药学语言系统（TCMLS）为骨架，将《中国中医药学主题词表》中的关键主题词作为术语体系和分类依据，针对中医药学知识表达和知识推理进行探索[3-5]。目前已有的中医领域本体较多，但由于中医领域的复杂性，现有研究的侧重点各有不用，主要集中于中医病证[6]、中医证候学[7]、治则治法[8]和方剂学[9]等方面的知识表达。Fang Miao[10]通过整合相关疾病，症状和处方的数据资源构建本体，提取了核心概念，概念描述，两个实体之间的关系以及中医药领域中关系的限制约束，借助 Protégé 软件从中医药领域（Traditional Chinese medicine, TCM）本体中获取语义关系。

本文知识来源主要来自《中医药学主题词表》、“中医药语言系统（TCMLS）”及《中国药典》。为了实现描述语言的转换在此引入《中医药主题词表》，由于需要在领域本体中切合专业领域的研究，因此需依据专业知识进行再次重构，使其具备更强的学术性、专业性。中医药语言系统主要分为“语义网络”和“基础词库”两部分，其中“语义网络”部分定义了 128 种语义类型，它的框架构成了中医药学语言系统的顶层架构，依据该语义体系，为了便于研究的理解，本文对中药药物的语义类型构建做了整理分类，如下图 2 所示：

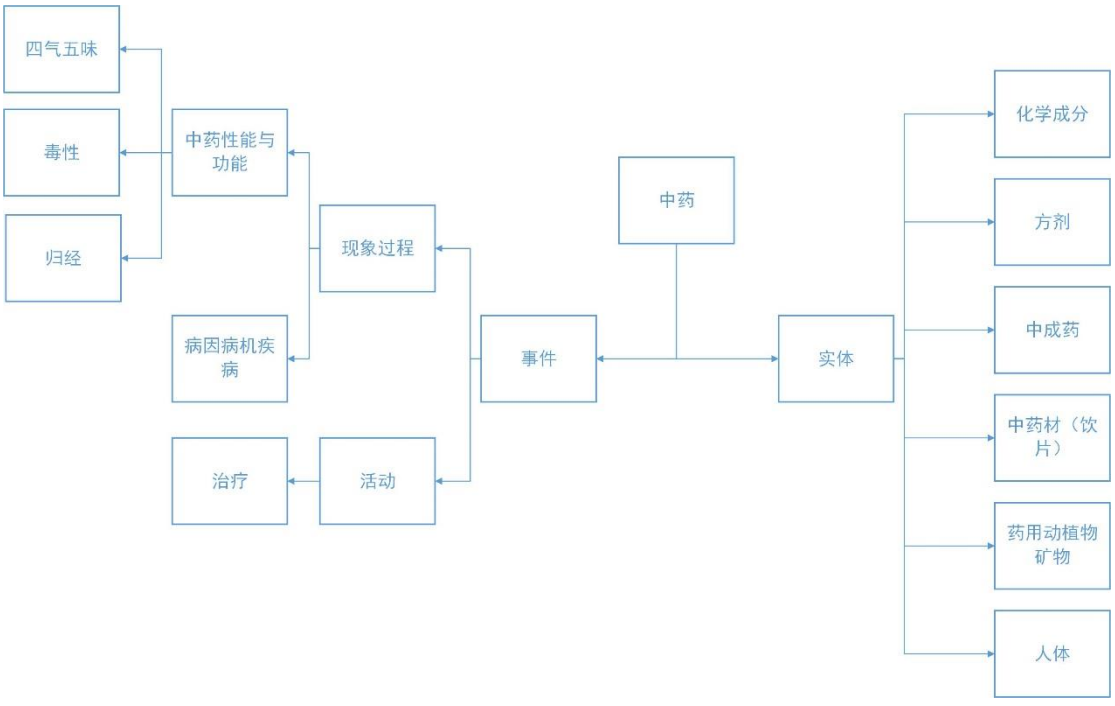


图 1 中药语义类型简化顶层本体构建

3 中医药本体的构建

3.1 本体构建的方法

本文选取由斯坦福大学提出的“七步法”来构建中医药本体，由图 2 可见七个步骤分别是:(1)确定中医药领域本体构建的专业内容范畴；(2)对现有中医药本体领域研究进行文献调研，考查是否有可以复用的本体框架；(3)依据《中医药主题词表》列出中医药领域的主要术语；(4)将定义的术语映射到类中自上而下构成等级体系；(5)明确定义中医药体系中类的属性；(6)依据中医药领域数据来源定义属性的分面；(7)本文重点是对中医药用药创建实例，因此以方剂为例建立实例。

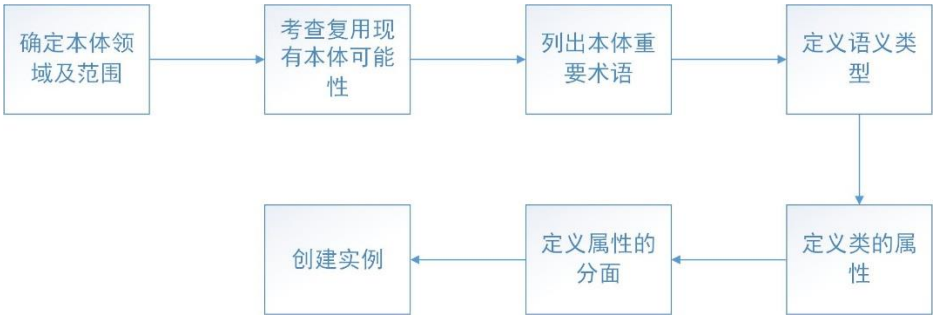


图 2 “七步法”流程图

3.2 中医药本体类的构建

选择标签 class，创建基本类结构，基于上述的 TCMLS 简化中药本体顶层构建图，将主体中药主要分成“事件”与“实体”两大类，“事件”下分二级类“活动”与“现象过程”，“实体”下分二级类“中成药”、“中药材（饮片）”、“人体”、“化学成分”、“方剂”和“药用动植物矿物”，最终建立的类目体系由图 3 所示。

图 3 protégé 类目体系

3.3 中医药本体属性定义

对已构建好的类进行对象属性和属性的分面定义，本体中类与类之间的二元关系有两种类型的属性：对象属性（Object Properties)和数据型属性（Data Properties）[11]。本文选择上而下的方式来定义层级结构，“上下位”关系用“is…a”来构建层次关系，是有指向性的单向关系层级，在其他对象属性中，“基源”和“是…的基源”，“组成”和“由…组成”两组互逆，即逆属性（Inverse properties）关系，其关系是双向性的便于后期推理，由于部分类之间的二元关系具有 Transitive(及物)属性，即上下传递性，因此本文还在关系中设置传递性选项。

本文的数据属性分类依据《中医药学主题词表》及《中国药典》，并以此进行具体的归纳总结，而后中医药用药涉及到的重要属性具体可由下表 1 所示：

表 1 中药本体类之间的属性关系

主体	关系描述	客体	关系类型
中药	主治	病因病机疾病	对象属性
	具有药性	性能	对象属性
	功效	功能功效	对象属性
	基源	药用动物植物矿物	对象属性
	影响	人体	对象属性

	进程	人体	对象属性
	特征		数据属性
病因病机疾病	病机		数据属性
	发病类型		数据属性
	发病因素		数据属性
	是…的成分	中药材（饮片）	对象属性
	is_a	中药	对象属性
	名称		数据属性
	拼音		数据属性
中成药	性状		数据属性
	鉴别		数据属性
	剂型		数据属性
	贮藏		数据属性
	用法用量		数据属性
	是…的成分	化学成分	对象属性
	is_a	中药	对象属性
	组成	中成药	对象属性
	组成	方剂	对象属性
	配伍	中药材（饮片）	对象属性
	中药名		数据属性
	拉丁名		数据属性
	拼音		数据属性
中药材（饮片）	别名		数据属性
	形状		数据属性
	用法用量		数据属性
	贮藏		数据属性
	鉴别		数据属性
	质地		数据属性
	颜色		数据属性
	浸出物		数据属性
	制剂		数据属性
	炮制		数据属性

	由…组成	中药材（饮片）	对象属性
	主治	疾病、症状、证候	对象属性
	由…组成	中药	对象属性
	方剂名		数据属性
方剂	用法		数据属性
	功用		数据属性
	方解		数据属性
	证治机理		数据属性
	配伍特点		数据属性
	是…的基源	中药	对象属性
药用动物植物矿物	名称		数据属性
	科属		数据属性

在本文的中医药本体构建中，所有术语都有属性定义的差异，即在一个父词术语下属的子类中应与其他相关的子类术语有区别，因此在考虑类的内部结构时要考虑类之间关系的精确定义。通过确定上述的属性定义，完成中医药用药的语义关系构建，中药本体构建主干部分即图 4 所示，全貌如图 5 所示：

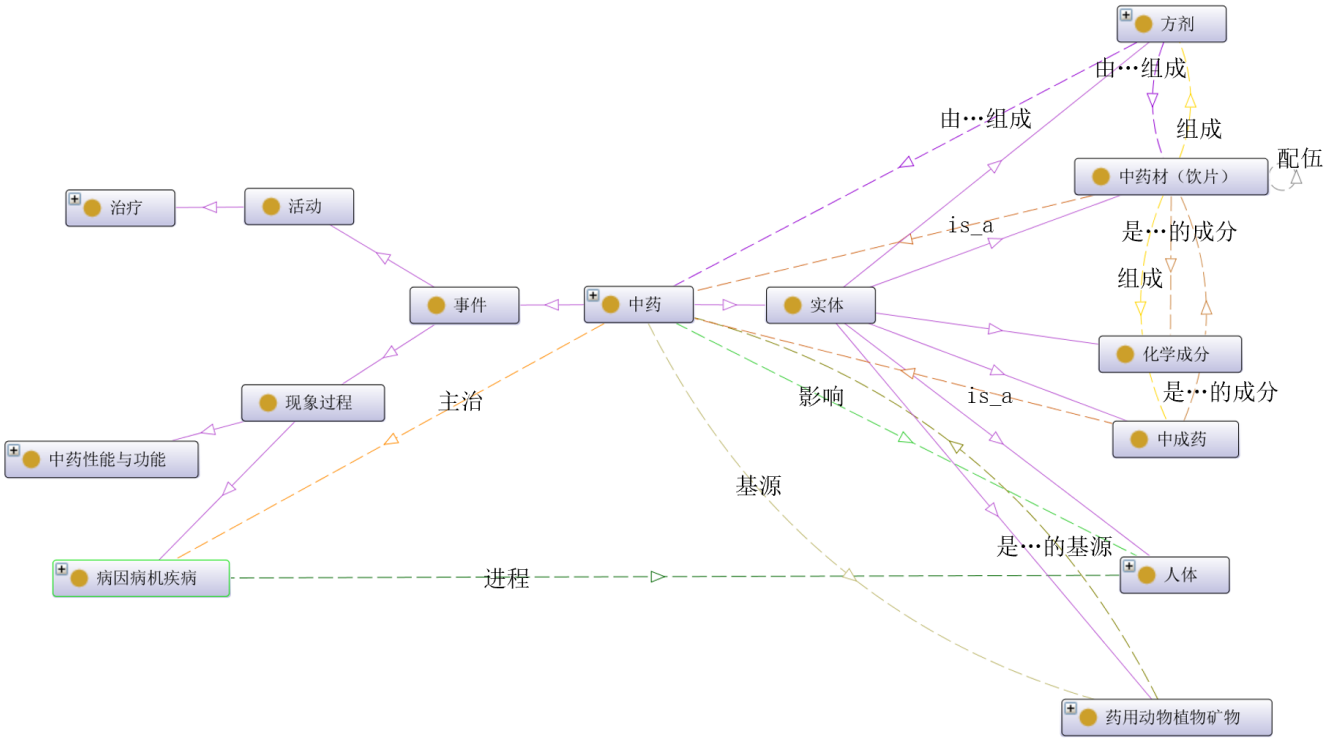


图 4 中药本体主体结构图

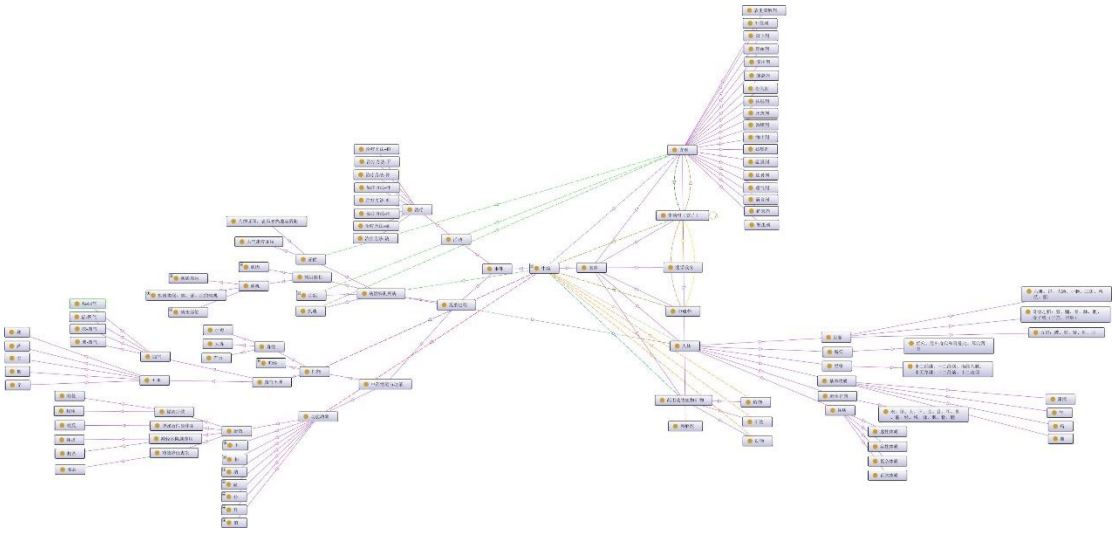


图 5 中药本体全貌结构图

3.4 创建实例

在 protégé 中，为解释药物与病因病机之间的关系，本文将方剂与症状属性相关联，如创建“症状”类的一个实例“咳嗽（cough）”，“症状”的数据属性“特征”与“病机”用作“咳嗽（cough）”的属性声明，其中“咳嗽（cough）”个体有 6 条使用记录，如下图 6 所示，其所有属性值如下：

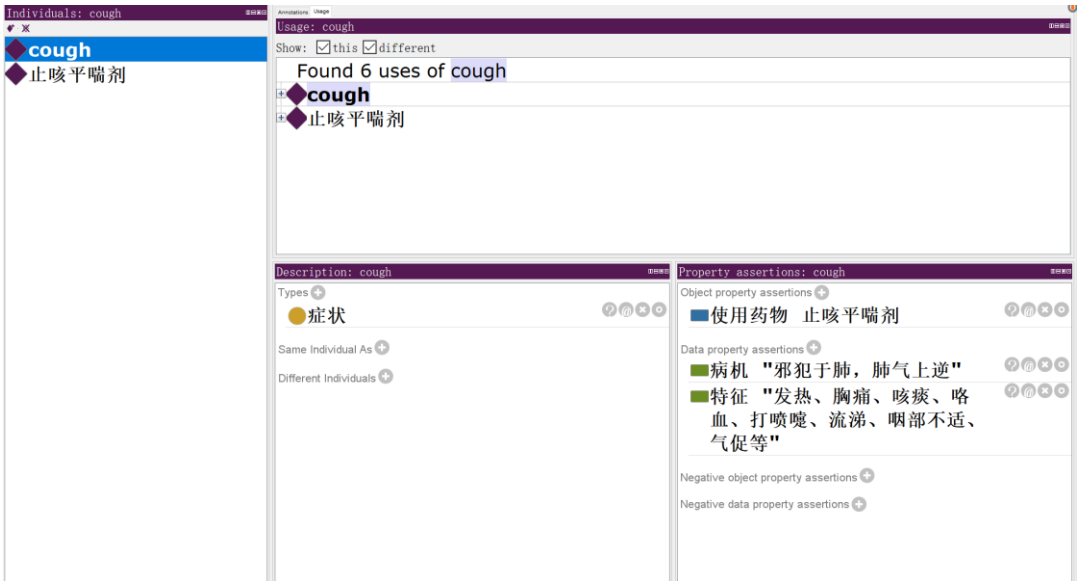


图 6 添加实例（咳嗽）

创建“方剂”类的一个实例“止咳平喘剂”，“方剂”的数据属性“功用”用作“止咳平喘剂”的属性声明，其中“止咳平喘剂”个体有 5 条使用记录，如下图 7 所示，其“功用”属性值为“有消除痰饮、止咳平喘及化痰、散结、截疟等作用”。

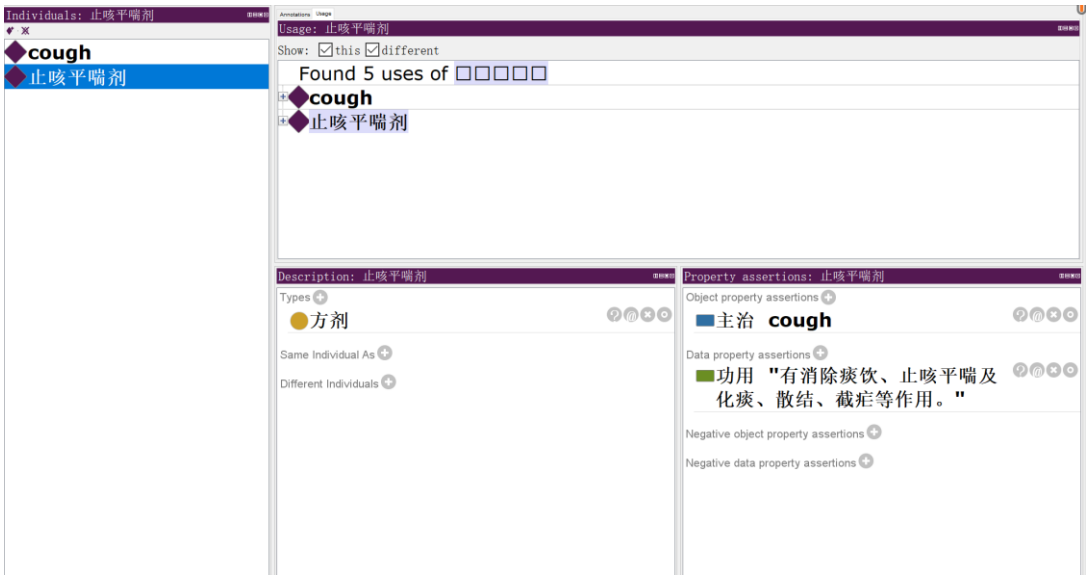


图 7 添加实例（止咳平喘剂）

中药主治症状，而中药由方剂组成，方剂与症状的语义关系为“方剂—主治—症状”，由上述语义关系可以构建二者实例关系，“咳嗽”的使用药物为“止咳平喘剂”，对应“止咳平喘剂”可以主治“咳嗽”，具体实例关系图由图 8 所示：

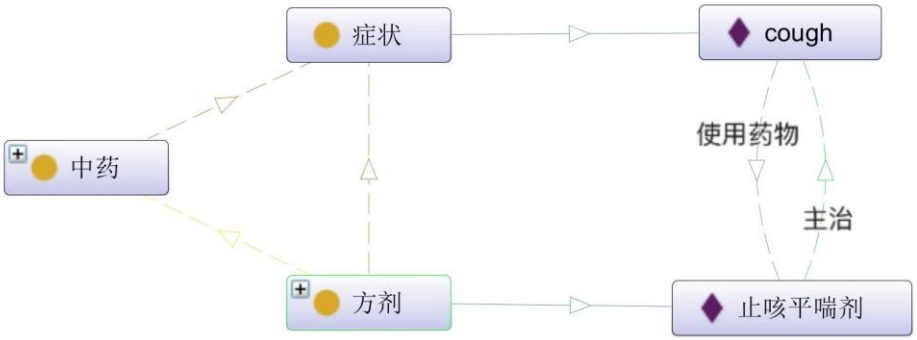


图 8 咳嗽实例语义关系图

“中药”可以影响“人体”，为具体展示中药对人体的影响关系，再构建“人体”与“方剂”的实例关系图。如创建“人体”类的一个实例“五脏”，“五脏”属于“脏腑”，“人体——脏

腑”的数据属性“名称”与“生理功能”用作“五脏”的属性声明，其中“五脏”个体有 6 条使用记录，如下图 9 所示，其所有属性值如下：

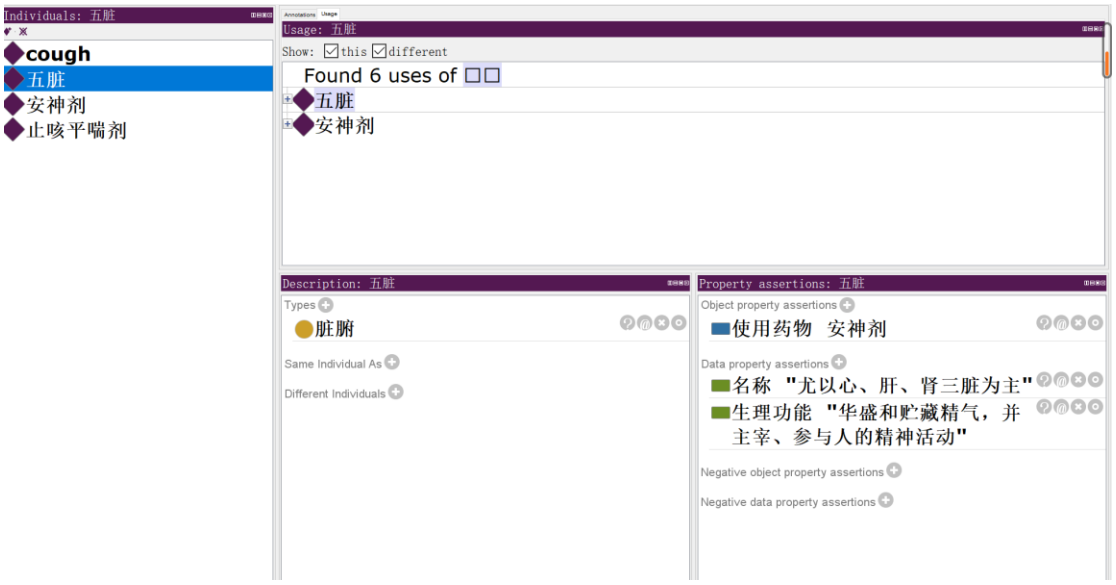


图 9 添加实例（五脏）

创建“方剂”类的一个实例“安神剂”，“方剂”的数据属性“方剂名”、“用法”、“功用”、“配伍特点”用作“安神剂”的属性声明，其中“止咳平喘剂”个体有 8 条使用记录，如下图 10 所示，其属性值具体为：

方剂名：安神剂。

功用：安神药分别有镇惊安神、养血安神、补心安神及解郁安神等作用。此外，有些安神药还有平肝潜阳、明目、收敛固涩等作用。

用法：因神志不安常以虚实夹杂证出现，故重镇安神剂和滋养安神剂往往配合应用，以求标本兼顾。

配伍特点：使用安神药需根据不同的病因及病情变化适当配伍用药。如心火亢盛者，当配清泻心火药；痰火扰心者，当配清热化痰药等。

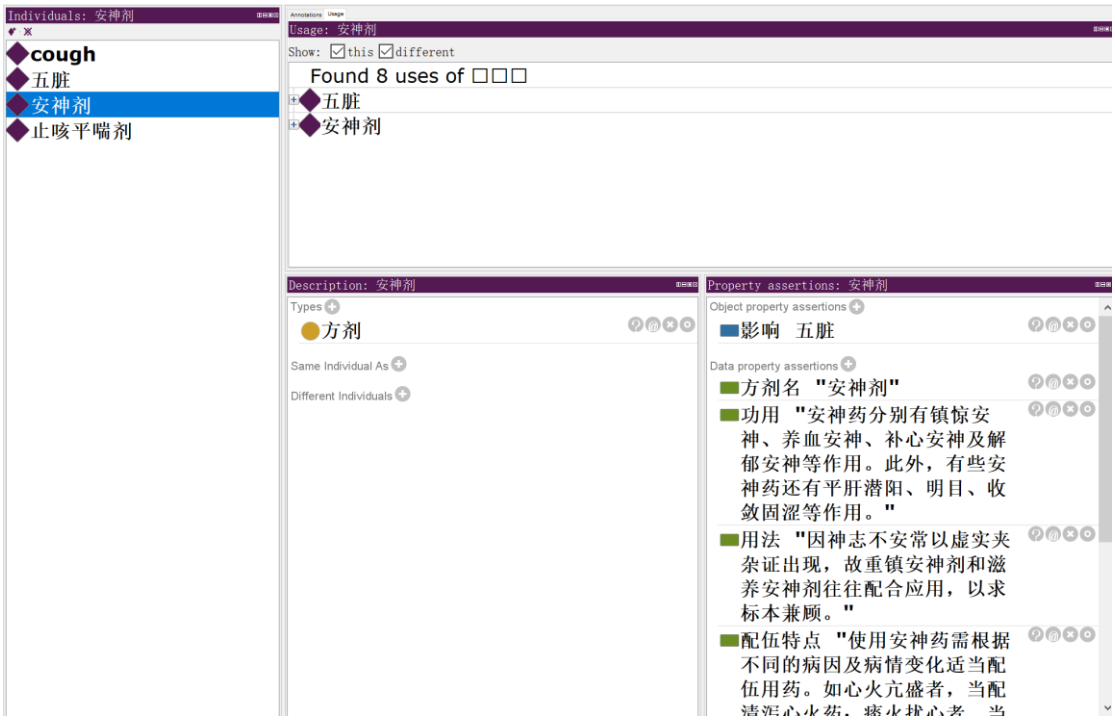


图 10 添加实例（安神剂）

中药影响人体，人体由脏腑组成，而中药由方剂组成，方剂与人体的语义关系为“方剂—影响—人体”，由上述语义关系可以构建二者实例关系，“五脏”的使用药物为“安神剂”，对应“安神剂”可以影响“脏腑——五脏”，具体实例关系图由图 11 所示：

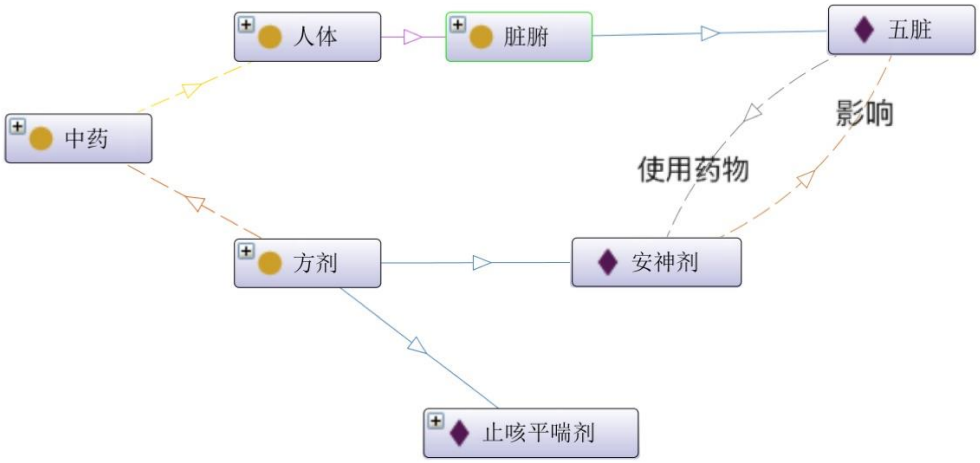


图 11 方剂与人体实例语义关系图

4 总结

本文在借鉴药典知识与 TCMLS Semantic Network 顶层本体框架的同时，针对中医药学本体构建，初步建立了中医用药本体框架，中药名称概念包含中药材、中药饮片和方剂，但各自内涵属性不同，因此本文将其概念分开描述分析，以确保其类与类之间的概念独立性和准确性[11]，使其体系更加完善，同时将中医药知识表示的复用可能性提高。不同于以往构建中医本体的单一性，将方剂和中药材合并为中药简化本体构建，本文创新点在于将方剂、中成药及中药材饮片单独分类构建，概念更为具体详尽，更有层次性；同时本文构建了方剂与症状之间的实例，解释了药物对症状的作用，具体阐述了药理方面的解释，使中药本体的理解更加深入。

在构建中医药本体的过程中也存在着一些局限性：①中医药领域的本体构建在不同研究中类的划分存在差异性，无法较好的重复利用；②中药本体的对象属性构建，在不同研究中无法完全统一，不具有统一性；③中药属性的定义具有一定的主观性，现有研究通常依靠专家评价的方式来解决该问题。具体解决方案建议将复杂概念层次化，建立较为完善统一的本体体系，以增加可利用性，还需进一步深入研究:①继续完善和发展中医药学语言系统；②参考已有的评价标准，建立评价体系指导本体构建；③利用已有的本体框架进行进一步知识推理，并设计相应的方剂应用场景，解决中医药在针对不同病因病机时的用药方案选择问题。

上述研究表明，本体构建的运用有助于中医药领域的知识表达和应用，中医药领域本体研究仍存在可探索的空间，如何将中医药信息知识全面展示的同时，提高中医药领域的本体实际应用的服务性还需探讨。

OR-015

医学院校“互联网+”创新创业教育实践平台构建

李学沧
哈尔滨医科大学大庆校区

目的 发现现阶段医学院校在创新创业教育的不足之处，通过构建“互联网+”创新创业教育实践平台，完善现有教育体系，促进医学院校创新创业教育的内涵式发展。

方法 研究对 5 所医学院校 3752 名本科生开展了问卷调查，结果表明现阶段在双创课程设置，项目指导等方面与学生的期望存在较大的差距。

结果 针对医学院校的教育资源供给与学生需求之间的矛盾，研究构建了包括五个主要功能，三层架构基于云计算和开源技术的“互联网+”创新创业教育实践平台。

结论 平台希望整合高校，政府等资源，为培养学生创新创业能力，加快项目落地，促进医学院校创新创业教育的内涵式发展做出有益的探索。

1 引言

随着创新创业教育的不断发展，国内医学院校大都开设了创新创业教育类课程，但是作为新课程由于没有统一的教学计划，大纲设置，创新创业课程的内容和师资差距较大。部分医学院校在开展创新创业教育过程中引入的通用型创业培训课程[1][2]（SYB）,由于此类课程的教学环节设计，课程内容与医学院校学科差距较大，难以与学生形成共鸣。同时由于医学院校既具有专业知识，又可以指导学生从事双创活动的教师数量较少，无法兼顾课程和学生项目指导,教学效果较差[3]。研究对国内 5 所医学院校 3752 名本科生问卷调查，发现学生对现阶段所开设的创新创业课程满意率仅为 55.8%。如何整合线上线下教育资源，利用互联网+技术，为医学院校创新创业教育赋能成为当前创新创业教育发展中的新问题。因此本研究开展了基于云计算和开源技术医学院校“互联网+”创新创业教育实践平台的构建，希望可以抛砖引玉，促进医学院校创新创业教育内涵式发展。

2 材料与方法

为了定量分析和评价现阶段医学院校学生对创新创业教育的满意程度，本研究采用问卷调查的方式对国内 5 所医学院校大二，大三的 3752 名本科生进行调查，其中女生为占比 65%，男生占比 35%。问卷调查时间历时 6 个月，采用网络问卷的形式，题目包括单选题，多选题，评分题，填空题，多选排序题等。

问卷调查的内容包括调查对象的基本情况（性别，年龄，学校，专业等），创新创业课程情况（课程种类，授课方式等），创新创业竞赛情况（竞赛种类，竞赛与专业结合度，提升创业能力等），创新创业实践情况（是否有校内外指导教师，是否有专项资金等），创新创业政策支持情况（减税政策，简化审批流程等），创新创业教育情况（丰富创业知识，培养创业精神等），教师在创新创业中的作用（师生共创，了解前沿动态，提升创业技能等）等 7 个领域，共 36 个问题，具体详见表 1。

统计采用 Excel（2010）和 RStudio（1.2.1335）统计软件进行统计分析，计数材料符合正态分布，采用人数和构成比表示；数据比较行卡方检验；检验水准 $\alpha=0.05$。

表 1 调查问卷

	题目	非常同意	比较同意	一般	比较不同意	非常不同意
创新创业竞赛	1.创业竞赛种类多样；	723(19.27%)	1044(27.83%)	1435(38.25%)	331(8.82%)	219(5.84%)
	2.参加的创业竞赛项目较容易落地；	570(15.19%)	923(24.60%)	1597(42.56%)	416(11.09%)	246(6.56%)
	3.创业竞赛项目与专业结合度较高	678(18.07%)	995(26.52%)	1497(39.90%)	354(9.43%)	228(6.08%)
	4.创业竞赛提升了创业能力；	769(20.50%)	1212(32.30%)	1364(36.35%)	229(6.10%)	178(4.74%)
	5.创业竞赛提升了创业自信心；	799(21.30%)	1248(33.26%)	1311(34.94%)	218(5.81%)	176(4.69%)
	6.创业竞赛拓展了人际关系网络；	859(22.89%)	1300(34.65%)	1256(33.48%)	178(4.74%)	159(4.24%)
	7.创业竞赛提升了团队合作能力；	971(25.88%)	1395(37.18%)	1110(29.58%)	143(3.81%)	133(3.54%)
	8.创业竞赛对于真实创业有较大帮助；	837(22.31%)	1256(33.48%)	1302(34.70%)	204(5.44%)	153(4.08%)
创新创业实践环境	9.创业实践有校内外指导教师；	859(22.89%)	1208(32.20%)	1292(34.43%)	226(6.02%)	167(4.45%)
	10.创业实践有专项创业基金支持；	767(20.44%)	1133(30.20%)	1392(37.10%)	273(7.28%)	187(4.98%)
	11.学校提供一体化的创业实践服务	714(19.03%)	1077(28.70%)	1472(39.23%)	294(7.84%)	195(5.20%)
	12.创业实践有独立的大学生创业园；	838(22.33%)	1064(28.36%)	1313(34.99%)	300(8.00%)	237(6.32%)

	13.创业实践有专的校外实践基地；	698(18.60%)	997(26.57%)	1464(39.02%)	350(9.33%)	243(6.48%)
	14.创业实践项目与专业学习结合度高；	732(19.51%)	1056(28.14%)	1458(38.86%)	298(7.94%)	208(5.54%)
创新创业政策支持	15.国家减免大学生自主创业企业税；	862(22.97%)	1183(31.53%)	1400(37.31%)	176(4.69%)	131(3.49%)
	16.地方政府简化大学生企业注册申请流程；	832(22.17%)	1171(31.21%)	1421(37.87%)	196(5.22%)	132(3.52%)
	17.学校提供创业的启动基金(无息贷款)；	818(21.80%)	1128(30.06%)	1442(38.43%)	222(5.92%)	142(3.78%)
	18.社会提供指导创业的免费培训；	815(21.72%)	1084(28.89%)	1456(38.81%)	236(6.29%)	161(4.29%)
	19.创业政策有助于提升个人创业意愿；	914(24.36%)	1333(35.53%)	1258(33.53%)	149(3.97%)	98(2.61%)
	20.创业政策对开展创业有切实的帮助；	925(24.65%)	1319(35.15%)	1259(33.56%)	144(3.84%)	105(2.80%)
创新创业教育	21.创业教育有助于丰富创业知识；	961(25.61%)	1337(35.63%)	1216(32.41%)	143(3.81%)	95(2.53%)
	22.创业教育有助于培养创新精神；	989(26.36%)	1342(35.77%)	1198(31.93%)	127(3.38%)	96(2.56%)
	23.创业教育有助于提升创业技能；	959(25.56%)	1372(36.57%)	1193(31.80%)	140(3.73%)	88(2.35%)

	24.创业教育有助于激发创业意愿；	953(25.40%)	1353(36.06%)	1202(32.04%)	148(3.94%)	96(2.56%)
	25.对学校创业教育质量总体满意；	873(23.27%)	1221(32.54%)	1328(35.39%)	206(5.49%)	124(3.30%)
创新创业课程	26.创业教育课程类型多样；	723(19.27%)	1044(27.83%)	1435(38.25%)	331(8.82%)	219(5.84%)
	27.创业课程内容与自己所学专业知识结合紧密；	608(16.20%)	907(24.17%)	1433(38.19%)	509(13.57%)	295(7.86%)
	28.创业课程内容与时代前沿趋势结合紧密；	700(18.66%)	1109(29.56%)	1410(37.58%)	317(8.45%)	216(5.76%)
	29.授课教师授课方式多样；	730(19.46%)	1118(29.80%)	1376(36.67%)	310(8.26%)	218(5.81%)
	30.授课教师具有创业经历；	707(18.84%)	916(24.41%)	1456(38.81%)	448(11.94%)	225(6.00%)
	31.授课教师具有丰富的创业教育教学经验；	796(21.22%)	1072(28.57%)	1347(35.90%)	338(9.01%)	199(5.30%)
教师在创新创业中的作用	32.有助于提升专业知识和应用能力；	1041(27.75%)	1437(38.30%)	1077(28.70%)	128(3.41%)	69(1.84%)
	33.有助于了解学科知识的前沿动态；	1076(28.68%)	1466(39.07%)	1016(27.08%)	127(3.38%)	67(1.79%)
	34.有助于提升科学研究能力；	1067(28.44%)	1441(38.41%)	1054(28.09%)	123(3.28%)	67(1.79%)
	35.有助于提升创新创业能力；	1088(29.00%)	1434(38.22%)	1045(27.85%)	122(3.25%)	63(1.68%)
	36.有助于创业	1018(27.13%)	1389(37.02%)	1141(30.41%)	132(3.52%)	72(1.92%)

	项目落地；					

3 结果

3.1 问卷结果分析

1、不同高校大学生对本校创新创业教育评价

研究将高校分为“双一流”高校和普通高校，其中“双一流”高校大学生对学校开展的创新创业教育评价较高，而普通本科院校大学生对本校创新创业教育评价非常低，提示普通院校需要在现有基础上尽可能创造条件，加大创新创业教育的投入，提升大学生的满意度。

表 2 不同高校大学生对本校创新创业教育评价比较

	“双一流”高校（n%）	普通本科院校（n%）	卡方值	P 值
非常同意	247(35.49)	626(20.48)	91.3074	6.95×10^{-19}
比较同意	230(33.05)	991(32.43)		
一般	184(26.44)	1144(37.43)		
比较不同意	28(4.02)	178(5.82)		
非常不同意	7(1.01)	117(3.83)		

2、城乡大学生个人创业意愿比较

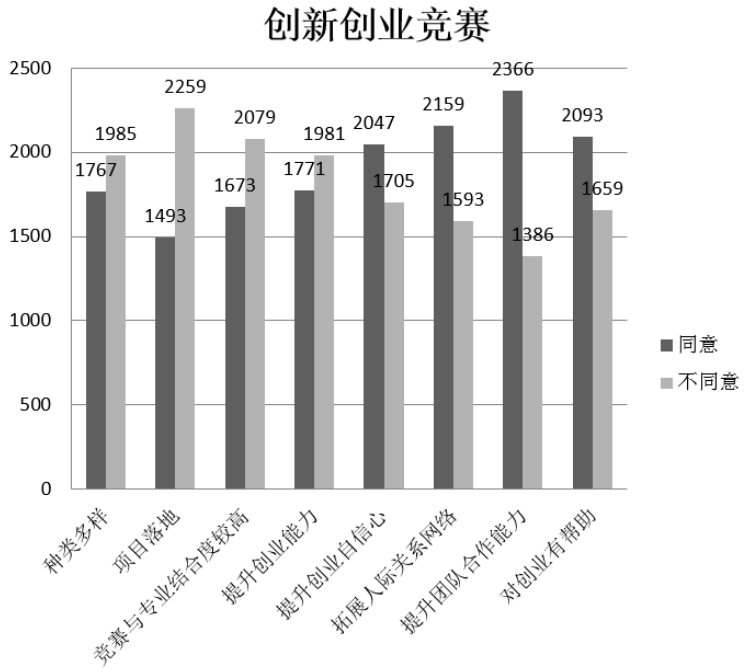
研究按照大学生生源地分为省会，直辖市或地级市，县级市或县城，乡或村，分析了城乡大学生个人创业意愿。城乡大学生个人创业意愿差异有统计学意义，城市大学生的创业意愿要更大，提示家庭和社会环境对大学生的创新创业意识培养具有一定影响。

表 3 不同地区来源大学生个人创业意愿比较

	省会，直辖市或地级市	县级市或县城	乡或村	卡方值	P 值
非常同意	78（5.14%）	43（4.09%）	32（2.7%）	29.67	0.000242
比较同意	138（9.09%）	92（8.75%）	73（6.17%）		
一般	613（40.38%）	417（39.68%）	470（39.73%）		
比较不同意	449（29.58%）	318（30.26%）	355（30.01%）		
非常不同意	240（15.81%）	181（17.22%）	253（21.39%）		

3、创新创业竞赛

当前医学院校学生对创新创业竞赛的看法是比较矛盾的，一方面大部分同学对参与竞赛的种类，项目落地的可行性，与专业结合等方面评价较差，但是一方面又高度认可参加竞赛对提升自身的创业力，自信心，拓展人际关系，团队合作能力等综合素质具有重要作用。调查表明学生赞同通过参与创新创业竞赛可以提升自身能力，但是由于现阶段双创竞赛类型的狭隘，医学相关创业项目少，参赛项目低水平重复，评委专业背景的缺乏等因素降低了学生对参与竞赛的热情，调



查问卷结果如图 1 所示：

图 1 创新创业竞赛

4、创新创业实践环境

创新创业实践环境调查反映出，现阶段学生对医学院校创新创业实践环境并不满意，主要表现在创业实践服务，实践基地，实践项目与专业学习结合度高等几个方面。由于医学院校创新创业教育开展较晚，相关的实践环境建设并没有跟上，医学院校在创新创业实践环境建设上还需要加大投入，同时由于医学专业的特点不能照搬其他学科的模式，结合医学教育特点，探索建设医学相关专业的创新创业实践环境是非常必要的。

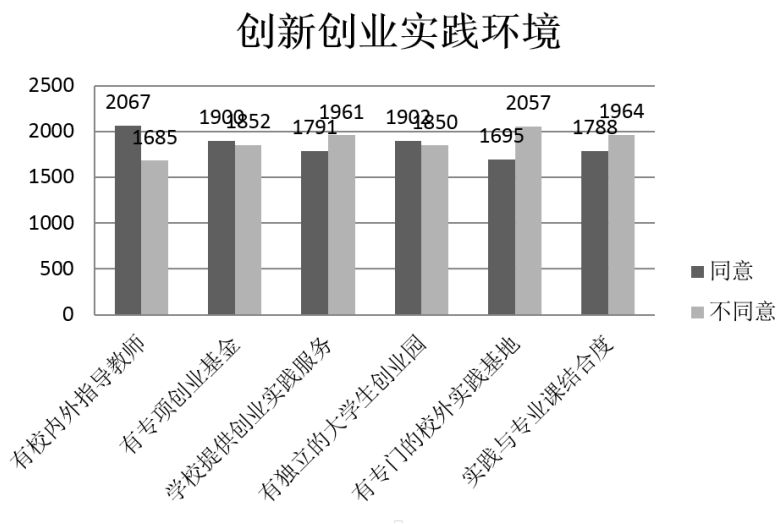
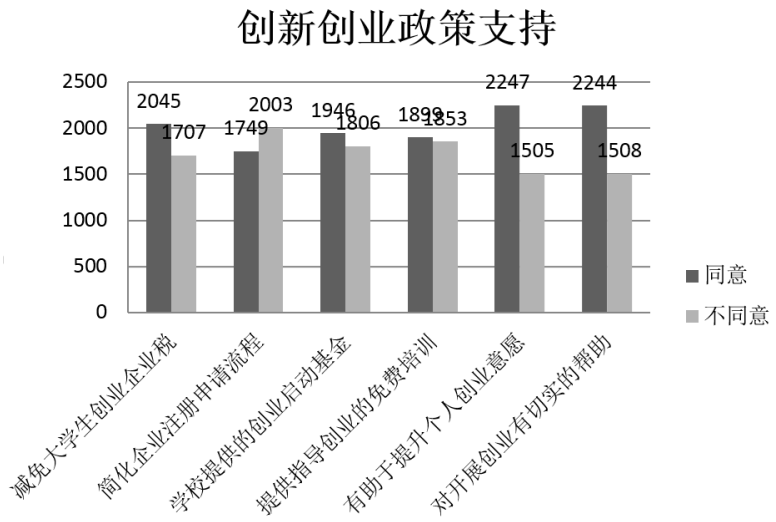


图 2 创新创业实践环境

5、创新创业政策支持

调查结果表明，医学院校学生对创新创业政策支持基本是满意的，特别是大学生开展创业有切实的帮助，提升创业意愿等方面。近几年通过学校和社会在竞赛，开始相关课程，典型报道等多个途径，对“大众创业，万众创新”相关政策的持续不断的宣传，使大学生对各级政府在大学



生创新创业相关的支持政策已经有了相当的认识。

图 3 创新创业政策支持

6、创新创业教育

通过调查，我们发现参与调查的大学生对创新创业教育在丰富创业知识，培养创新精神，提升创业技能，激发创业意愿是满意的，表明大学生对创新创业教育的作用是认可的。但是也应该看到大学生对创新创业教育质量总体满意度是偏低的，表明当前医学院校的创新创业教育处于初级阶段，有很多需要改进的地方。

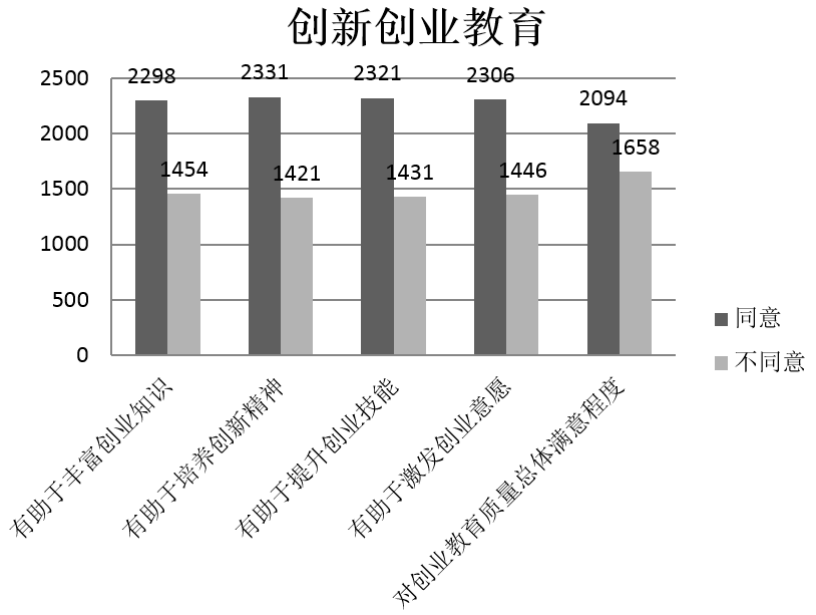


图 4 创新创业教育

7、创新创业课程

调查表明，学生对目前开展的创新创业课程普遍不满意。究其原因，现阶段医学院校创新创业由于大部分医学院校的创新创业课程开展时间较短（少于 5 年），课程积累较少，师资力量不足，课程形式单一，创新创业与专业课程内容结合的较少等原因，导致现阶段医学院校创新创业课程质量并没有获得学生的认可。因此，创新创业课程建设将是医学院校创新创业教育内涵式发展的重要内容。

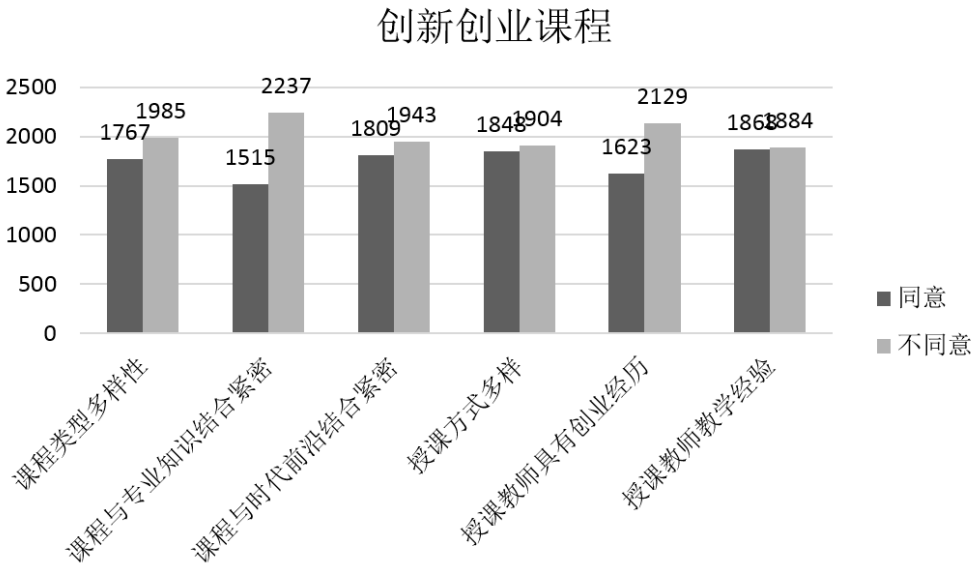


图 5 创新创业课程

8、教师在创新创业中的作用

教师在创新创业的中的作用问卷调查显示，学生对教师的作用是期望较高的，希望教师在相关活动中教师可以在提升专业知识和应用能力，提升科学研究能力，提升创新创业能力等方面发挥更为重要的作用。而对应之前的创新创业课程调查，学生对现阶段对相关教师的表现是不满意的。医学院校教师在创新创业教育过程所发挥的作用，与学生的期望存在较大差距。

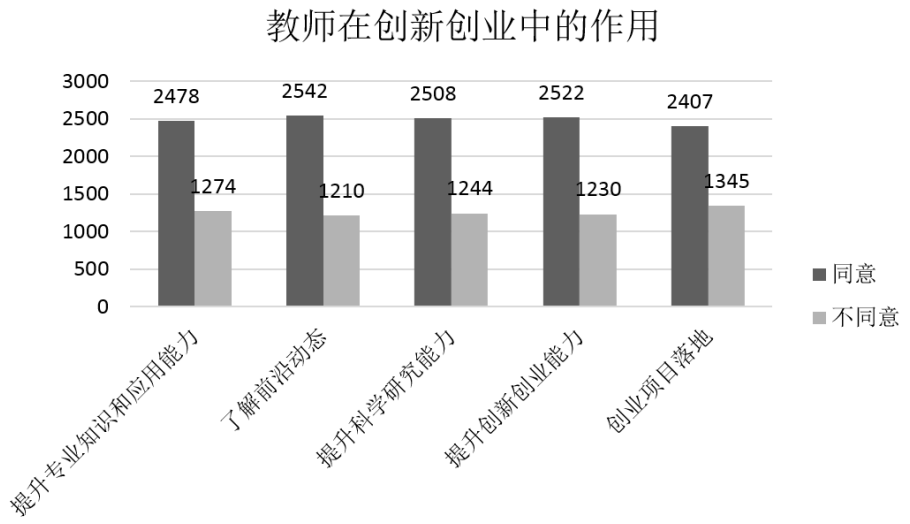


图 6 教师在创新创业中的作用

3.2 平台功能设计

“互联网+”创新创业教育实践平台需要整合学校，企业和政府多方面的资源，通过理论教学，创新实践，创业实践等多方面构建完整的创新创业生态[4][5]。在此基础上结合前期调查问卷结果，研究归纳总结了“互联网+”创新创业教育实践平台的 5 大主要功能，详细设计如下：

表 4 “互联网+”创新创业教育实践平台功能设计

1.信息发布	1.1 通知公告
	1.2 政策解读
2.在线学习	2.1 创新创业课程资源库
	2.2 视频库
	2.3 资料库
3.项目管理	3.1 项目申报

	3.2 项目审核
	3.3 项目公示
4.竞赛管理	4.1 竞赛报名
	4.2 竞赛审核
	4.3 竞赛公示
5.创业实践	5.1 项目资源库
	5.2 专家库
	5.3 创客联盟
	5.4 校内孵化
	5.5 中介服务机构

（1）信息发布

在开展创新创业活动过程中，经常有同学由于缺乏有效的信息来源渠道，错过竞赛，讲座等活动。信息发布模块为学生提供培训，竞赛，项目申报，国家和地方政策解读，方便学生及时获得相关信息，吸引更多的学生了解创新创业，参与创新创业活动。信息发布模块可以采用通过公众号，短视频，小程序等多形式，多渠道尽可能接触到学生的日常生活方方面面，实现导流功能。

（2）在线学习

互联网+时代，在线课程作为课堂内容的延伸，其作用日渐凸显[6]。创新创业课程资源库包括课件，教案，课程素材，调查问卷等教学资源形成创新创业课程资源库。视频库收录了近年来的竞赛视频，培训视频等；资料库包括调查问卷模板，项目书模板等材料帮助学生获得市场调查，商业计划书撰写等方面的资源。从而弥补医学院校在创新创业教育领域师资方面的不足。

（3）项目管理，竞赛管理

由于现阶段参与项目申报，双创竞赛的学生团队较多，而医学院校提供的评审专家较少。因此对项目，竞赛管理开展评审，及时反馈修改意见，帮助团队完善项目，提高评审效率尤为重要[7]。该模块根据现有的管理流程，实现从报名，评审到结果公示的全过程网络化，提高了评审过程的工作效率和公平性，详见图 1。

（4）创业实践

创新创业教育启发了学生，项目和竞赛提升学生能力，为今后的创业打下坚实的基础。但是走向真实创业时，学生创业团队会面临更多的问题，及时帮助学生创业团队解决问题，提供指导是当下创新创业教育过程中难点[8][9][10]。考虑到很多医学院校不具备企业孵化能力，因此设计创业实践模块整合孵化器，地方政府，高校资源来助力创业团队。其中，项目资源库收集整理了已有项目和团队信息，随时掌握过往项目的情况；专家库收集整理校内外专家信息，便于与创业团队

对接；创客联盟通过虚拟社区供学生，指导教师，专家开展交流讨论催生创意；为了帮助项目落地，校内孵化和中介服务机构对接当地孵化器，科技局等资源为创业团队提供企业注册，财务，社保，专利申请等服务，加速项目落地和成果转化的过程，提高项目转换率。

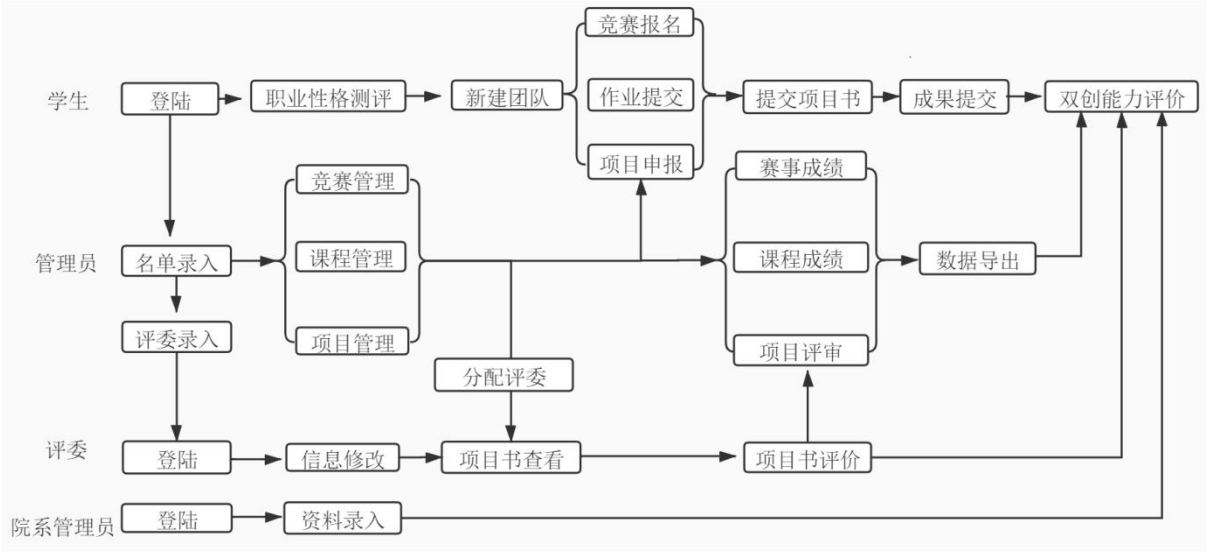


图 7 竞赛管理，项目管理业务流程图

3.3 平台架构设计

随着以云计算，开源软件为代表的现代信息技术的快速发展，使基于“互联网+”的创新创业教育实践平台实现成为可能[11]。因此研究提出了三层架构的“互联网+”创新创业教育实践平台架构设计，平台由基础设施层，数据层，业务逻辑层组成[12]，详见图 2。

（1）基础设施层

通过公有云或者私有云产品，利用容器技术（Docker）可以快速搭建平台，实现 WEB 服务器部署，防火墙，域名等功能，有效减少技术难度，降低运行维护成本。同时借助开源数据库技术实现结构化和非结构化数据的存储。目前开源统计工具 R/Python 的统计分析功能已经非常完善，借助这些分析工具，可以完成各类数据分析，并实现数据可视化。

（2）数据层

数据层包括创新创业课程资源库，视频库，资料库，项目资源库，专家库等基础数据，这些数据分为结构化数据和非结构化数据。为了实现存储和调用结构化数据和非结构化数据，平台的持久化框架采用 Spring Boot2 和 MySQL/Mogodb 开源软件解决方案。开源框架技术 Spring Boot2

支持主流数据库，提供 POM 文件简化了开发环境的配置。MySQL 和 Mogodb 联合使用可以实现结构化数据和非结构化数据存储，增、删、改、查等基本操作，为系统的功能扩展打下基础。

（3）业务逻辑层

“互联网+”创新创业教育实践平台采用前后端分离技术。其中后端采用开源通用管理系统做二次开发。在开源通用管理系统基础上，平台实现信息发布，在线课程，项目和比赛申报，审核等相关功能，同时可以还可以对大学生，院系开展各类数据收集和分析。“互联网+”创新创业教育实践平台前端实现多屏共享和数据可视化。

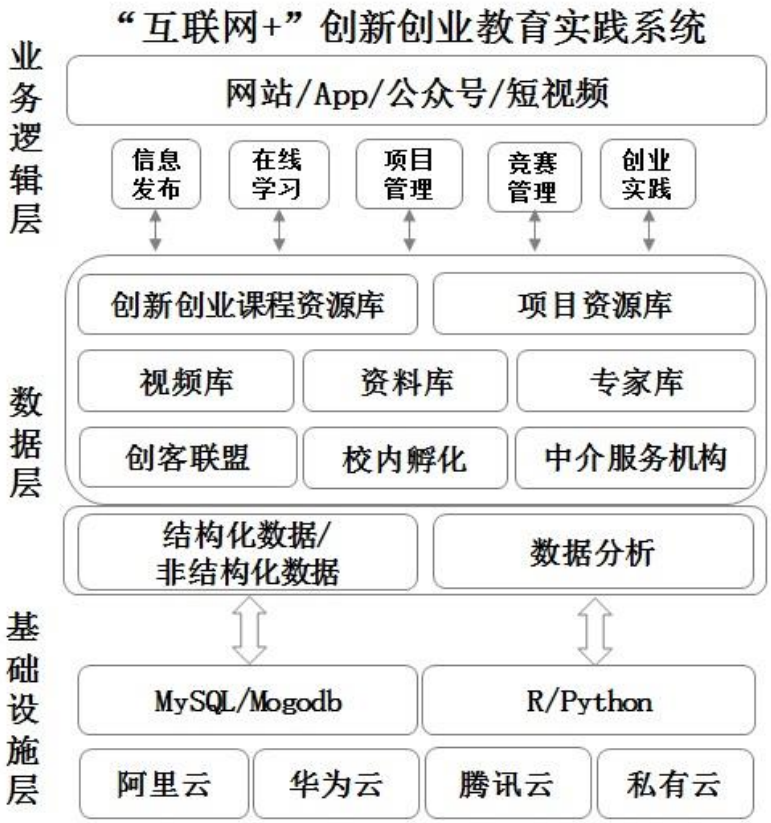


图 8 技术架构设计

4 讨论

综上，现阶段医学院校大学生了解创新创业相关政策支持，想参与创新创业竞赛又认为和所学专业联系较少，认可创新创业教育的重要性，但是对创新创业课程形式并不满意，同时期待教师在创新创业教育中发挥更大作用。问题的根源在于医学院校大学生在医学专业课程与创新创业课程融合，学科指导，参与实践活动等方面的需求强烈，与之相对应的是学校在创新创业知识学

习、实践条件，教师指导等方面供给不足。因此在现有理论课程基础上开展创新创业教育改革，利用“互联网+”建设符合医学院校学科特点的创新创业教育实践平台势在必行。

5 结论

在对国内 5 所医学院校的 3752 名本科生问卷调查基础上，针对现阶段医学院校开展创新创业教育所遇到的问题，构建了医学院校“互联网+”创新创业教育实践平台。平台包括了信息发布，在线学习，项目管理，竞赛管理，创业实践等五个主要功能，提出了基于云计算和开源技术的三层“互联网+”创新创业教育实践平台架构设计。平台在激发学生参与热情，深化课堂教学内容，缓解师资不足方面，促进创业等方面可以发挥积极作用。同时需要指出的是目前医学院校在场地，相关配套服务上还存在不足，如何进一步整合学校，企业和政府资源，加速项目落地，仍有很多工作需要做。“互联网+”创新创业教育实践平台作为当前医学院校创业生态的载体，将不断完善，促进医学院校创新创业教育的内涵式发展。

OR-016

“人工智能”背景下基于 CDIO 模式导师制的医学信息学人才培养探索与实践

吴辉群
南通大学

目的 随着医疗人工智能的不断发展，智慧医学模式在健康、疾病和诊疗、康复的理念和措施上形成全新的认识和判断，而以疾病“诊疗”为中心的现代医学教育面临着重大调整。

方法 为适应新的教育形式，教育部新医科改革明确提出发展智能医学等医学新专业。

结果 南通大学医学院医学信息学系在医学信息专业人才经过近五年本科生导师制人才培养模式的探索和实践教学中取得了一定的育人成果，在此基础上并借鉴工程教育领域构思(conceive)、设计(design)、实施（implement）和运行(operate)CDIO 教学模式优化专业培养方案

结论 ，旨在搭建“人工智能”背景下的医学信息学双创人才培养框架体系。

在以人工智能、大数据为代表的新一轮科技革命和产业变革的新背景下，各级医疗卫生机构紧跟信息化发展的步伐，正逐步进入智能医疗阶段，对诊疗数据有效分析和利用急需智能医学专业人才。在“健康中国”从治疗为主到生命健康全周期新理念下[1]，教育部围绕“三新”理念重点布局“新医科”建设，发展建设精准医学、转化医学、智能医学等医学新专业[2]。其中智能医学方向的人才培养要求多学科的交叉融合，注重学生实践操作和科研创新能力培养，能为其将来

长远发展奠定良好基础。在学科发展要求方面，借助医疗机器人、虚拟现实、增强现实、5G 网络、人工智能等技术，实现人工智能辅助的，实现医学诊断治疗全过程智能化是必然的选择[3-5]。在专业设置方面，教育部自 2018 年批准在医学学科门类医学技术类专业类别下特设智能医学工程专业，并授予工科学位。在人才培养方面，国内以天津大学、南开大学为代表的高校，2018 年开始了智能医学工程本科招生教育[6]，医科院校也纷纷开始申报该专业并探索人才培养，不同的院校结合自己的特色进行该方向的人才培养。在课程设置方面，相关的知识和能力结构无法满足需要，“临床+智能技术”整合度不高，传统单一学科背景下教学中课堂内容与科研创新严重脱节，智能技术人才尽管有对医疗数据建模的方法与工具，但缺乏对医疗数据的产生与标准知识，而传统的医学院校开设的课程中关于人工智能内容匮乏，对于应用人工智能解决临床诊疗问题的案例课程体现不够。探索以创新为导向的新医科人才培养模式，重塑新型医学信息人才的培养体系尤为重要[7]。

当前我国的医疗信息化正处于互联网医疗向智能医疗过渡的阶段[8]。南通大学从 2002 年开始的医学信息相关的人才培养也随着医疗信息化第三阶段智能医疗的开展积极转型。2015 年起，医学信息学系开展本科生导师制。本科生从大二开始，经双向选择进入教师科研团队，不仅有利于消化课堂知识，也有利于培养学生的创新能力。这项工作与大学生各类创新创业大赛挂钩，与本科毕业论文工作挂钩，尽早确立学生的研究课题。该系教师利用自身稳定的科学方向和充足的研究经费，积极指导本科生开展科学研究，训练学生的科研能力，提高学生的科研素质，培养创新型和高素质人才，并取得了可喜的成绩。本研究以南通大学医学信息学系本科生科研能力培养实践为案例，通过实行本科生导师制教学，逐步有序地将 CDIO［构思(conceive)、设计(design)、实施(implement)和运行(operate)］理念引入到本科生的实践创新能力培养当中，利用 CDIO 系统严谨、统一规划的教育理念，结合导师制模式[9,10]，有针对性地进行医学院校大学生创新训练，以探索适合自身专业特点、开拓创新精神的智能医学方向人才培养方式。

1 CDIO 教育模式的概念及特点

优秀的智能医学人才需要具备以下三种素质：第一，科学研究素质，要有强烈的科研意识、科学的科研方法和崇高的科研精神；第二，产品设计素质：程序化业务模型，对核心业务进行内容解读，做好数据化运营，处理实际数据的能力，需要始终根据市场需求进行产品设计和迭代。培养这样的人才，需要数学、统计、计算机科学、医学、人工智能等学科之间的密切合作，同时也需要和产业界或其他拥有数据的部门之间的合作[11,12]。CDIO 是工程教育与人才培养创新模式，是一种基于项目的探究式学习方式，学生参与真实产品的构思、设计、实施和运行的全过程，在项目研究过程中将知识点有机整合，在产品制作过程中把理论、实践、创新融为一体，培养学生

的工程能力[13]。近年来，因其严谨细致的规划过程和可重复性操作，其理念在医学教育领域得到应用[14-17]，在医疗信息化与相似学科领域的人才培养培养方面，也已有探索和应用[18-20]。

2 本科生导师制的含义及特点

本科生导师制是一项以教学班为基础，以学科专业活动为依托，以学生专业学习兴趣为桥梁，以导学为重点，以导向为补充，着眼于学生专业成长和教师学术发展的人才培养制度。对于充分发挥教师的主导作用，建立新型的师生关系，实施因材施教和个性化培养，推动教研一体化，构建合作与互动的育人机制，形成全程育人、全员育人、全面育人的良好氛围，建设教学研究型大学相符的优良学风，提高人才培养质量具有重要的作用和意义。项目导师是学生以赛带练、以练促赛、学用结合、实践成才的训练师、教练员和同行者实施。在指导学生以各级各类学科竞赛为主的学科专业活动，包括创新创业项目竞赛、挑战杯竞赛、学生科研项目的活动的基础上，及时对学生进行成长方向、学业规划、学习方法、创新思维、创业就业等方面的培养和指导，形成“医学人工智能项目引导的”本科生导师模式。另外，部分本科生导师还兼任学院本科生科研兴趣班导师，吸引了医学院临床医学、影像医学和口腔医学等其他专业的本科生加入到该系教师的科研团队，除此之外，本专业方向的导师还可以与其他医学学科背景导师以导师团队形式开展工作，促进了医工交叉的课题研究深入。

3 基于 CDIO 的导师制模式的实施过程

3.1 导师遴选与管理

根据 CDIO 教育模式标准中第 9 条提高教师工程实践能力和第 10 条提高教师的教学能力是该模式能否顺利实施的关键因素。导师是教学过程中除学生外的一大主体，是实施教学的主要因素，我系在开展导师制教学过程中对其有严格要求：①政治素养、专业素养及职业道德良好；②取得博士学位、或获得讲师职称三年以上、具备一定的科研基础和研究条件（至少主持或主要参加过一项省部级及以上科研项目）。遴选导师 10 名，其中正高 1 名，副高 5 名，中级 4 名，具有 CDIO 教学实战和课题申请经历的教师 9 名，每个导师每年级指导 2-3 人。在管理方面，为了保障项目质控，导师指导学生过程记录与本科生毕设设计实践管理相结合[21]，避免传统导师制模式中导师管理松紧不一、实验进度不一的情况，保证 CDIO 教学质量的可控性。为了保证项目顺利开展，系室定期组织导师培训座谈，加强沟通交流，让导师在日常交流中引导学生；系室分管毕业设计教学主任及时了解项目进展，对各小组实验进展进行监督，将 CDIO 各阶段落实到位。导师在各阶段采取不同的教学重点和方法，注重学生科研思维锻炼，培养学生独立解决问题能力。

3.2 CDIO 项目资源和实践设施保障

制定合适的项目对于 CDIO 在教育过程具有重要意义。人工智能在医疗服务领域的应用包括智能虚拟助手、医学影像辅助诊断、智能临床决策支持系统、医用机器人和精准放疗等；在公共卫生服务领域的应用包括智能健康管理、智能疾病筛查与预测；还可应用于药物研发、智能医院管理、智能医学教育和医学科研支持等领域[22]。在选择相关选题的时候要根据课程、资源和教师科研方向综合考虑。为保障教学科研实践的顺利进行，保证课题顺利进行，导师提供项目运行经费，学生也可以申请国家级、省级和校级大学生创新创业训练计划等。整合校内外相关实验室资源进行跨学科平台共享，给学生提供丰富的科研资源。我系数字医学实验教学中心面积达 500 多平方米，包括医学人工智能实验室、医院信息系统模拟实验室、远程医学实验室和医学图像处理实验室等专用实验室，拥有先进的深度学习工作站、服务器集群、网络交换机、PACS 图像显示系统、临床应用模拟系统等相关软件和硬件设备，为教学和科研工作提供了良好的条件。借助学校的智慧实验室管理平台，学生登记后可以随时进入实验室学习，极大地提高了学生学习的主动性和灵活性。

3.3 导师与学生团队教研融合模块

学生从大学二年级开始参与科研创新训练，通过 C（构思）、D（设计）、I（实施）、O（运行）的过程保障教学质量，项目以小组为单位，在整个项目实施过程中，导师在整个项目中作为引导者和监督者，负责本项目人员组织、分工、任务协调、进度监控和问题解决等，对整个课题方向及实施进程进行把控，并针对学生定期开展组会，提出自己的见解和指导。项目构思设计阶段主要在大二阶段，约 10 周时间，由导师发布各自的研究方向与项目，师生进行双选。考虑到大二开始有专业课，一般主张从学生研究兴趣着手，适当加以引导，兼顾本科知识点覆盖，例如一些编程、英语、统计课程，将本科阶段远程医学、生物医学数据分析、医用数据挖掘、医学影像信息学等课程知识点有机串联起来，增强学生对课本知识的理解及课外知识的学习扩展应用。导师引导学生了解相关课题背景、确立研究目标及内容，进行项目书撰写。为更好地引导学生入门，系部定期开展专题讲座，内容涉及文献检索、生物医学信息学等常用软件使用、常用实验技术培训等。避免了制定任务的盲目性、周期长及脱离实际等问题。实施阶段本阶段约 1 年时间，主要在大三阶段进行，也就是研究实施阶段。在该阶段要求各组根据流程完成材料准备、实验操作、数据采集分析等具体任务，项目组定期进行汇报，导师指导，系室通过本科生毕业设计（论文）信息系统进行全程管理，进行过程验收和监督。此阶段教师重在引导学生，实时纠正研究方向和目标，并定期组织进展汇报和实验讨论。系室除了提供实验资源，还可通过组织相关学术讲座，让学生在实施过程中拓宽视野、启迪思维。在学生本科阶段的最后一个学期，学生利用自己的科研成果参加竞赛、撰写论文，进行学位申请等，通过这些途径检验教学成果的有效性。

4 讨论

基于 CDIO 理念的本科生导师制教学模式，是培养具有较高科研创新能力、较强实践和学习能力、较高层次智能医学方向人才的教育模式。通过创新能力平台的搭建及 CDIO 模式导师制的实践，参与该教学模式的学生科技创新能力明显提高，参加国家级大创项目 7 人次，省级大创项目 4 人次，校级大创项目 11 人次，导师指导学生获得全国 “互联网+”大学生创新创业大赛铜奖、“挑战杯”全国大学生课外学术科技作品竞赛特等奖、“共享杯”全国大学生科技资源共享服务创新大赛优秀奖、“创青春”全国大学生创业大赛“智慧校园”主题赛优秀奖、全国高校人工智能创新大赛三等奖、全国大学生数学建模竞赛江苏省二等奖等，在 SCI 期刊及中文权威期刊发表论文 8 篇，培养的多名本科生毕业后顺利进入著名院校继续研究生学习，其中 1 人免试复旦大学攻读博士学位。CDIO 导师制培养模式让学生在本科阶段开拓眼界、养成良好的科研习惯、积累一定的科研经验，有效地克服本科生参与科学研究的盲目性和随意性，是帮助本科生打开科研大门的一次较好尝试，提高了学生的专业竞争力，也营造了学生的良好学风。

通过导师辅导及 CDIO 的严格质控，学生能积极参与到为其量身定做的项目中，主动学习和不断探索，积累的项目经验可以为今后进入到工作岗位奠定基础。然而，CDIO 导师制模式也要看到该模式的局限性。第一，CDIO 导师质量和数量有待进一步提升和扩展。在项目实施过程中，尽管对导师设立了明确的准入标准，但不可否认每个导师的带教方法、科研能力不尽相同，且高水平的具有医学人工智能产品研发经验的师资有限，这是该教育模式持续发展的难题之一[23]。第二，CDIO 工程学教育理念理念主要以项目为载体进行人才培养，在实施过程、质量把控及评价标准的建立等方面还需要进一步摸索。同时医学院校本科生专业普遍课业繁重，如何选择合适的实施时间、进行何种强度和深度的培养训练是值得思考的问题。第三，关于 CDIO 模式评价方法和标准是实施 CDIO 的关键，要客观制定标价标准，形成过程与结果评价体系有待进一步讨论完善。

如何使医学课程与人工智能专业课程衔接更紧密，构建交叉融合式医学信息学专业教学体系，让导师制更好地应用于医学信息学人才培养是当前面临的挑战。鼓励更多优秀教师参与导师制教学、对导师进行系统化标准培训、保障高水平教学质量尤为重要。充足的专项经费是 CDIO 实践教学的保障。可以根据导师的工作量和工作质量考核情况发放班导师工作津贴，根据项目导师指导的学科专业竞赛项目获奖情况并在原有学科竞赛指导奖励、挑战杯竞赛指导奖励、创新创业竞赛指导奖励、学生科研指导奖励的基础上适当给予补充项目导师工作津贴。另外，在导师的遴选管理方面可以纳入导师工作满意度的考核。内容包括学生满意度、学科专业相关的竞赛活动的参与人数和获奖情况等，其中竞赛项目包括学科竞赛、创新创业项目竞赛、挑战杯竞赛和学生科研立项，将考核结果作为教师年度考核的重要内容之一。学院每年评选一次优秀导师，并给予表彰，在晋升职务或职称时同等条件下给予优先考虑。对不能认真履行导师职责、考核不合格的将取消导师资格，并与个人年度考核挂钩，同时减少或取消导师工作津贴。总之，在“人工智能”的背景下，综合导师制和 CDIO 模式的优点，与 PBL、微课、混合教学等 [24-26] 教学模式进一步深入完善高水平医学信息学人才培养模式，是今后的努力方向。

OR-017

医工融合专业人才培养质量的隐性影响因素探析
——基于扎根理论的研究

王辅之、杨枢、张浩、张钰、孔令超、陈兴智
蚌埠医学院

目的 探析对医工融合专业人才培养质量有潜在影响的隐性因素。

方法 通过焦点小组访谈收集医工融合专业人才培养隐性影响因素的资料，应用扎根理论提取访谈资料的初始概念，基于主轴编码确定概念的副范畴和主范畴。

结果 共提取 139 条描述性子句，分类归纳后形成 29 个初始概念，最终提取 10 个副范畴和 4 个主范畴。

结论 医学院校医工融合专业人才培养质量的隐性影响因素主要体现在教学行为、学业管理、师资建设和学习环境四个方面。而要避免这四个方面对人才培养质量的影响，则需要从学生层面、教师层面和管理层面做好相关的工作。

“医工融合”是指以临床医学为载体，以人工智能、云计算、物联网、大数据等信息技术为手段，以医学领域和工学领域知识、技术、方法的交叉融合、协同创新，促进并实现医学科学、医疗技术的跨越式进步[1]。而医工融合人才的培养正是适应经济社会的发展，培养具有交叉学科思维、通晓医学和工学基本知识，能够主动发现临床实践中遇到的问题并运用工学方法加以解决的创新型人才。但是由于工科教育的特殊性，医学院校的医工融合人才培养质量已经成为一个重要的教育研究问题，正在引起人们的广泛关注[2]。

如果把医学院校的工科人才培养工作看做一个系统工程，那么从“系统论”的角度来看，高等教育人才培养质量是显性因素和隐性因素共同作用的结果[3]。显性因素包括传统意义上的课程体系建设、教学监控管理、教师治学态度等。而隐性因素则指那些对人才培养质量产生显著影响，但是又非常容易被忽视的因素[4]。包括：学生对专业的认知度、学生的学习动机，班级学风、学习环境等因素。只有均衡考虑显性因素与隐性因素在人才培养质量中的作用，才能最大程度提高医学院校医工融合人才培养质量。

目前，对于高等教育人才培养质量的显性影响因素已经开展了大量的研究，但是对于隐性影响因素的研究尚未引起足够的重视。本文基于扎根理论的技术和方法，通过对医学院校工科学生

和教师进行焦点小组访谈，收集影响人才培养质量的隐性影响因素，并构建医学院校医工融合专业人才培养质量的隐性影响因素模型，以期为今后有的放矢地提高相关专业的人才培养质量提供理论依据和决策支持。

1 研究方法

1.1 样本遴选

本研究为定性研究。调查对象是某医学院校“物联网工程”、“医学信息工程”和“信息管理与信息系统”三个医工融合专业的本科生和部分教师。采用焦点小组访谈收集医工融合专业人才培养质量隐性影响因素。焦点小组访谈分两轮进行，首先进行针对学生的小组访谈，然后进行针对教师的小组访谈。学生组访谈对象是在三个医工融合专业的二年级和三年级本科生，分别从学习成绩排在班级前 20%、20%-80%和后 20%的学生名单中各随机抽取两人，共 36 人。教师组访谈对象包括专业课教师和年级辅导员，要求访谈专业课教师有至少 5 年以上的教学工作经验，辅导员要求有至少 8 年以上的学生管理工作经验。所有受访者被告知本研究的目的和内容，在征得其同意后，将其纳入焦点小组访谈。最终，遴选 24 位本科生，6 位教师（4 位专业课教师和 2 位年级辅导员）组成访谈小组。

1.2 调查工具

本研究采用自编的非结构化访谈问卷完成数据收集。该问卷条目是基于前期已发表的相关研究论文设计的，经过一轮学生预测试和一轮专家评估后定稿。问卷条目完全采用开放式回答来收集信息，仅规定讨论的内容范围，而不预设标准答案。问卷共有 11 个开放式问题，见表 1.

表 1. 医工融合专业人才培养质量影响因素的非结构化访谈问卷

编号	问题	调研对象
1	您认为除了一些传统教学监控措施（例如：课堂教师听课，教学质量评估等）以外，针对工科教育特点，还可以在哪些方面加强教学质量的监控管理？	学生
2	您认为工科专业课任课教师的教学能力对您学习成绩存在影响吗？如果存在影响，请您谈谈有哪些影响。	学生
3	您重视与老师（学生）建立和谐的师生关系吗？您认为师生间的和谐关系是否会影响您的学业成绩？	学生、教师
4	请您谈谈对于您所读专业开设的工科类专业课程的认识和评价？	学生
5	请您谈谈合理的奖惩制度在激励学生学习积极性方面的作用？	学生

6	除了上述谈到的问题以外，您认为还有哪些因素会影响您的学习积极性？	学生
7	您认为除了绩效考核、教学督导等传统教学质量监控措施以外，还有哪些因素会影响您提高教学质量的积极性？	教师
8	您认为和谐的师生关系有利于提高医学院校工科专业的人才培养质量吗？	学生、教师、辅导员
9	请依据您的工作经验，谈谈为了提高医学院校的工科专业人才培养质量，需要加强哪些方面的学生管理工作？	辅导员
10	依据您的了解，除了校园内部环境以外，还有哪些校外环境因素可能对于工科专业人才培养质量产生影响？	辅导员
11	您认为从学校层面还可以在哪些政策、管理方面为工科学生的学习提供帮助？	教师、辅导员

1.3 资料分析

本研究基于扎根理论对访谈资料进行分析。扎根理论研究方法是由哥伦比亚大学的 Anselm Strauss 和 Barney Glaser 两位学者共同发展出来的一种研究方法。该方法是运用系统化的程序，针对某一现象来发展并归纳式地引导出理论模型的一种定性研究方法。扎根理论在社会科学研究领域被广泛应用，取得良好的分析效果[5]。本研究通过以下四个步骤完成隐性影响因素的理论建模。第一步，将访谈录音整理成文字资料，通过对文本资料的分析，提取与医工融合专业人才培养质量影响因素相关的论述，并将其拆解为独立的描述性子句；第二步，根据独立的描述性子句所反映的特征和含义对子句进行开放性编码，形成初始概念；第三步，依据开放性编码结果形成副范畴及主范畴；第四步，构建医工融合专业人才培养质量隐性影响因素的理论框架。

1.4 质量控制

为确保信息收集质量，会前准备好两个录音笔分别设置在会场两端。在会议开始后，首先向与会者充分介绍教学质量隐性影响因素的概念，并将国内外开展的相关研究现状及取得的成果和观点整理成文字材料，发放给与会人员，并安排会前的交流时间，以便受访者充分了解教学质量隐性因素的相关概念。访谈会议中，主持人一方面鼓励与会者就本课题研究内容开展头脑风暴讨论，另一方面注意控制会议节奏和讨论内容与本研究的相关性。

2 结果

2.1 开放性编码

开放性编码是基于原始资料寻找初始概念的过程[6]。本研究共归纳独立的描述性子句 139 条，分类归纳后形成初始概念 29 个。开放性概念编码提取结果及内涵示例见表 2.

表 2. 开放性概念编码及内涵实例

编号	初始概念	参考点数*	内涵实例**
1	教学环境	15	天气较热课堂里也没开空调，就这一个原因我就不想听课了（S03）
2	学生的专业认知	13	我们专业是医工融合的交叉学科，刚进校时特别想对本专业有所了解，但不知问谁（S11）
3	师生交流	10	我就比较喜欢私下找老师聊天，感觉老师也很乐意和我们聊天（S14）
4	幽默感	9	上课很幽默，被老师吸引，不知不觉就下课了（S17）
5	普通话教学	9	老师方言比较重，说话听不清，说的什么我听不懂，更难以理解（S03）
6	课堂纪律	9	喜欢老师对课堂教学纪律严格要求，感觉对我们很负责（S19）
7	自习环境	8	工科实验室太小了。而图书馆灯光偏暗，还不让自己带台灯，也不让带电脑，不方便我们自习（S19）
8	惩罚	8	还是需要加强惩罚，尤其是考试作弊，有的监考老师不太认真会让我感觉考试不公平（S24）
9	课堂行为	7	看到上课时有一些同学在玩游戏或者睡觉，我也会受影响（S22）
10	课件制作	6	课件设计得很好，没那么多文字，一些图表有助于我们理解所学的知识（S1）
11	教师形象	5	衣服很得体，一看就是很重视上课，对教学很有信心（S3）
12	语言表达	5	老师说话总在一个频道上，像老夫子，没有抑扬顿挫，让人想睡觉（S21）
13	旷课	4	看室友睡过点了，我也不想起床，所以就和他一起旷课了（S15）
14	鼓励	4	课堂上无论回答对还是错，都希望老师能给予正面的评价（S07）
15	课程情感	3	工科类的课程感觉好难，有的课程我前几节听不懂，就不想再听下去了（S05）
16	答疑解惑	3	有一位外聘的工科专业课教师，我们提出问题他就让我们自己百度（S09）
17	工科专业课衔	3	感觉我们的工科类课程之间缺少衔接，一门课上完就上完了，

	接		不知道对后续工科专业课程有什么帮助（S19）
18	教学工作激励	2	除非教学工作做得特别好，拿到教学成果奖，否则感觉教学工作做得好不好，对自己的影响不大，而拿奖对于医学院校的工科专业来说太难了。（T03）
19	校园环境	2	可能是新校区的原因吧，但是和我想象中的大学不一样，刚到学校的时候，都是小树新楼没有积淀感，很难让我产生信任感和依附感，有点失落（S11）
20	工科教师职称晋升	2	相比医科专业，在医学院校工科专业的职称晋升太难了，想起来职称上不去，就对教学工作没了劲头（T02）
21	双肩挑兼职	2	学院工科教师短缺，除了上课，我还要承担学院的行政管理工作，有时候忙不过来，感觉太累了（T03）
22	问题学生监督	2	对于不守纪律的学生必须加大管理力度，否则会带坏整个寝室（C02）
23	学业出口	2	因为是医工融合的交叉学科专业，学生都比较关心未来的就业和考研，每年毕业季，都有学生来咨询应届生的毕业和考研情况（C02）
24	教师科研压力	1	医工融合的专业特点决定了我们从事科研工作必然占用大量精力，侵占了一些本可以用于课程备课的时间（T03）
25	工科师资培养	1	我觉得对工科专业教师的教学能力在职教育很重要，但是最近一次关于教技、教法的学习好像还是三年前的一个讲座（T02）
26	外聘工科教师管理	1	学校对于外聘的工科类教师待遇过低，有时候联系外聘教师都是利用教师的私人关系（T01）
27	社会的专业认同	1	虽然用人单位非常欢迎本专业毕业生，但是社会上对交叉学科专业不了解，学生有时候说出去自己的专业，别人会不理解为什么选择这个专业（C01）
28	父母婚姻	1	我带过的有两个学生印象比较深，都是父母离异的，性格比较孤僻，有顶撞老师的现象，也是挂科困难户（C02）
29	辅导员绩效激励	1	人才培养质量的高低与辅导员绩效没有太大联系，只要学生管理工作不要出现责任事故就可以了（C01）

*参考点数：该概念在访谈中出现的次数；**S：学生；T：教师；C：辅导员

2.2 主轴编码

主轴编码是指以扎根理论为基础，明确研究问题主要范畴的动态过程。通过进一步对开放式编码结果进行分析与提炼，本研究将初始概念映射到教学设计、课堂沟通、情感建设、学生管理、

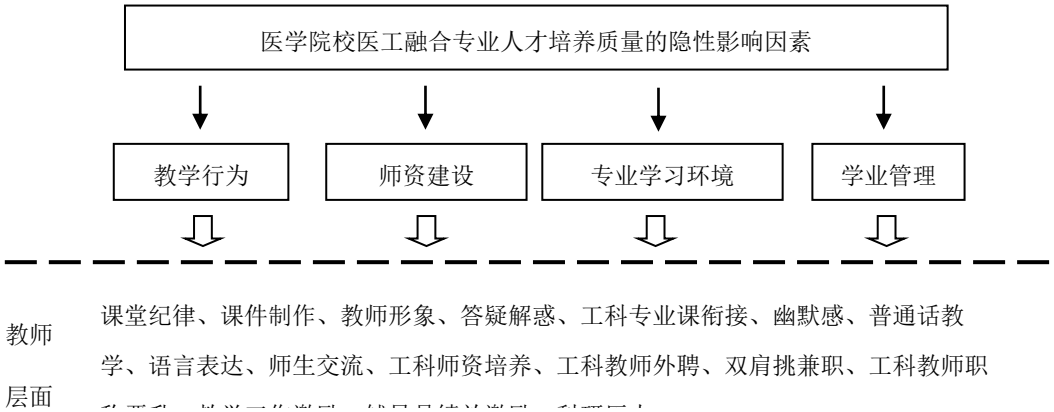
奖惩机制、伙伴影响、师资管理、教师成长、校内环境和校外环境，共计 10 个副范畴，和教学行为、学业管理、师资建设和学习环境，共计四个主范畴中。主轴编码结果见表 3.

表 3. 主轴编码结果

主范畴	副范畴	概念
教学行为	教学设计	课堂纪律、课件制作、教师形象、答疑解惑、课程衔接
	课堂沟通	幽默感、普通话教学、语言表达
学业管理	情感建设	课程情感、师生交流
	学生管理	专业认知、学业出口
	奖惩机制	问题学生监督、鼓励、惩罚
	伙伴影响	课堂行为、旷课
师资建设	师资管理	工科师资培养、工科教师外聘、双肩挑兼职、
	教师成长	教师职称晋升、教学工作激励、辅导员绩效激励、教师科研压力
学习环境	校内环境	教学环境、自习环境、校园环境
	校外环境	父母婚姻、社会的专业认同

2.3 模型构建

根据上述编码体系构建医学院校医工融合专业人才培养质量的隐性影响因素模型，见图 1。该模型为一个二维模型。第一个维度是依据主轴编码结果划分的。该维度从工作内涵的角度出发，认为教学行为、师资建设、学习环境和学业管理是影响医工融合专业人才培养质量隐性因素的四项核心工作内容，提示需要做好这四个领域中的相关工作，以帮助提高医工融合专业的人才培养质量。第二个维度是依据影响因素所属对象领域划分的。该维度从工作对象的角度出发，认为提高医工融合专业人才培养质量，需要重点从教师、学生、管理三个对象层面做好相关工作。第一个维度是主轴编码的直接结果，第二个维度是对第一个维度的扩充。



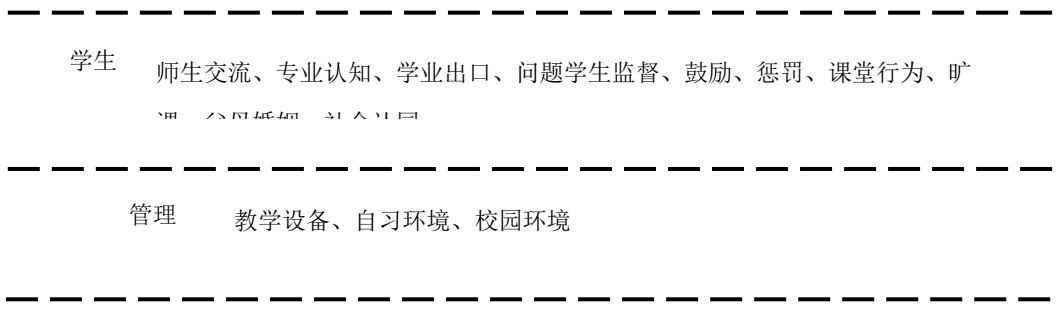


图 1. 医学院校医工融合专业人才培养质量隐性影响因素理论模型

3 讨论

医学院校开展医工融合、医理融合、医管融合专业教育是学科交叉、融合发展的大势所趋。本研究基于扎根理论对医学院校开展医工融合专业人才培养工作中可能面临的教育质量隐性影响因素问题开展了较为深入的研究。前期研究已经证明除了传统的教学软件、硬件保障体系外，还有很多因素会影响到人才培养的质量[7,8]。这些因素可能普遍存在于医学院校的工科人才培养过程中，但是没有引起人们的注意，医学院校也很少会设置专门的管理标准体系以控制这些因素对工科人才培养质量的影响。本研究结论提示我们，为了进一步提高医学院校医工融合专业教育质量，需要重点做好以下几方面工作。

第一，重视工科专业课堂教学过程中隐性因素对教学效果的影响作用。

医工融合专业不同于传统医学专业的显著特点就是授课内容涉及知识面广，更加强调知识点之间、课程之间的融会贯通[9]。这就对教师课堂教学提出了更高的要求。本研究发现更积极、有效的教学沟通，包括教学课件的优化、教师的普通话教学、授课语气、衣着形象等因素，有助于提高医学院校工科学生对所学习专业知识的接纳程度，从而有利于提高医工融合专业人才的培养质量。

第二，合理配置工科专业师资力量，加大工科师资队伍建设力度。

医学院校开展工科专业教育面临的一个重要挑战就是工科师资队伍建设[10]。在本研究中，学生和教师对该问题的相关表述也突显了医工融合专业师资队伍建设的重要性。因此解决好师资队伍建设问题尤为重要。建议重点做好以下三方面工作：首先，建立合理的校外工科教师聘用制度，以弥补医学院校开展医工融合教育可能出现的师资缺口；其次，知人善任、因材施用，合理布局师资力量。对于双肩挑教师可以适当安排其行政工作，而对于骨干教师则应该鼓励其将主要精力放在教学、科研和学科发展工作中；再次，重点做好讲师及以下职称、非博士学位教师的教学激

励工作，尤其是工科教育背景教师的教学激励工作。在评奖评优、课题申报等工作中予以扶持和倾斜，激励他们做好教学工作。

第三，细化工科专业学生管理工作内涵，提高人才培养质量。

前期研究结果表明，医学院校交叉学科学生往往缺乏对专业的清晰认识，这种对专业的不确定性会极大影响其学习积极性[11]。在本研究中也反映出类似问题。一方面是对学生的关爱，包括重视师生交流，开展专业认知教育，以及加强对学生家庭的了解；另一方面是强化学生管理中的奖惩机制，尤其是建立对于一些不适用处分处理的违规现象的约束机制，例如对于不良课堂行为、旷课等现象的处理，从而形成良好的班风、学风，促进学生的学习和健康成长。

第四，审视工科专业学习环境的顶层设计。

在学校层面，做好对于自习室、图书馆等学习环境的营造和修缮工作，尤其注重对空调、灯光、桌椅等公共配套设施的科学管理。在学院层面，重视学生社团在专业教育中的作用，利用好校内外资源，开展交叉学科的专业学术讲座和科普讲座，加强学生的专业归属感和成就感。

4 结语

医学院校开设医工融合专业教育有其独有的优势，但在教学、师资和管理等方面也存在有待完善之处。提高医学院校医工融合专业人才培养质量是一个系统工程，不仅需要考虑传统意义上的教学质量监控措施和教学软硬件设施的建设，同时还要重视隐性因素对人才培养质量的影响作用。本研究结果表明，医学院校医工融合专业人才培养质量的隐性影响因素主要体现在教学行为、学业管理、师资建设和学习环境四个方面，而要避免这四个方面对人才培养质量的影响，则需要从学生层面、教师层面和管理层面做好相关的工作。

OR-018

基于质性分析的大学生健康信息获取行为引动研究

祝帆帆、桂晓苗
湖北医药学院

目的 研究大学生群体获取健康信息的行为引动机制；

方法 借助半结构化访谈方式，获取用户调研文本，利用 NVIVO12 质性研究工具，分析行为引动的突显理念；调查问卷分析辅助研究，完善引动机制分析。

结论 大学生在在线健康信息获取行为的突显理念主要包括行为控制、行为态度、行为目标和主观态度四要素，各要素之间互相影响，最终引发用户的信息获取行为。

1，引言

信息社会中在线健康信息给大众带来了更加丰富和便利的信息获取途径，对于大众来说，获取在线健康信息需要考虑很多因素，如信息的质量、可靠性、易获取性等。目前在校的大学生群体基本都属于数字原住民，他们已经能够很好的与信息社会环境相融合。同时由于受到较高水平的教育，具备一定的自主思考能力和信息辨别能力。对于这一群体，什么因素能够影响或者刺激他们实行信息获取行为是值得研究与探讨的。

目前针对在线健康信息获取的研究主要集中在以下几点：

（1）针对某特定疾病人群的信息获取行为研究：如高血压患者，强直性脊柱炎患者的在线健康信息寻求行为，糖尿病患者等。主要集中在慢性病患者群体。

（2）针对不同属性群体的信息获取行为研究：

老年群体：大南京大学的徐孝婷学者采用问卷调查等方式研究老年人对于在线健康信息的需求研究并结合马斯洛需求层次理论提出了老年人的在线健康社区的信息需求模型；城市老年人健康信息获取渠道对于健康信息素养的影响研究；对农村常住老年人的信息获取行为研究。

农民工群体：如王秀红等人将农民工作为研究群体，基于扎根理论对农民工的的健康信息获取行为的影响因素进行分析，并构建影响因素理论结构图。

孕产妇以及妈妈群体：如山西医科大学的章娟通过调查问卷的方式研究孕产妇对于母婴保健信息获取偏好。

（3）针对信息获取行为研究方法的理论性探讨：如武汉大学吴丹教授从选择性暴露的角度梳理现有研究现状，分析信息获取行为中的信息搜寻与信息规避行为，并总结出现有的研究主要集中的情绪、决策过程等因素对于信息获取行为的影响。

通过文献梳理，近几年针对大学生的健康信息获取行为研究年发量呈现上升趋势，如图 1 所示：

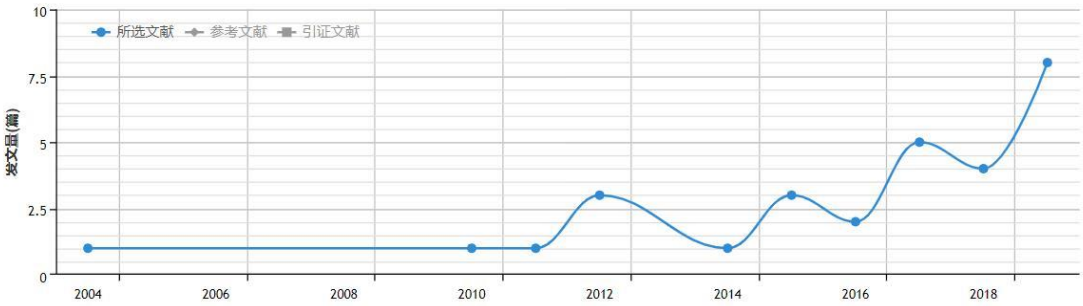


图 1　大学生健康信息获取行为研究年发文量趋势图

利用 BICOMB 进行关键词词频分析，经过数据清洗后共得到 30 个关键词，抽取词频>=3 的关键词后发现，对于大学生的在线健康信息获取的研究主要集中在以下三个方面：

表 1 大学生在线健康信息获取研究高频词统计

序号	关键词	出现频次	百分比%	累计百分比%
1	健康教育	8	14.8148	14.8148
2	新媒体	6	11.1111	25.9259
3	获取渠道	3	5.5556	31.4815

主要集中在健康教育，获取渠道分析，新媒体对于大学生群体健康信息素养的影响。

信息行为的引发往往是信息意识经外界环境的刺激被激发后引动信息用户的信息行为，信息获取是信息接受、吸收、利用和创造的源头[18]。基于此，本研究借助半结构化访谈方式，获取用户调研文本，利用 NVIVO12 质性研究工具，对大学生信息获取行为引动机制进行研究，分析能够激发大学生在线健康信息获取行为的因素。

2，研究方案设计

2.1　访谈

通过在线访谈的方式，询问三类开放式问题：目标行为的利与弊；他人或团体对行为的影响以及行为的激励因素和阻碍因素，获取调查者对于在线健康信息获取的理念基础，然后选取频次较高的几个理念作为突显理念。访谈对象为 41 位大学生，医学生有 22 位，非医学生有 19 位，共得到 41 份访谈文本。采用 Nvivo 软件对访谈文本进行质性分析。Nvivo[19]的文本编码依据是扎根理论，主要分为三阶段编码：开放式编码、主轴编码以及选择性编码。

2.2 问卷调查

本次研究还采用了问卷调查的形式收集更多的一手用户数据进行分析，问卷内容也主要是围绕上述的三个开放式问题设计，主要用来帮助进行理论饱和度检验。一共回收问卷 345 份，排除不合理的问卷（如答题时间少于 20 秒，前后答案矛盾等）后，有效问卷共有 310 份，有效回收率为 88.9%。本次调研对象涉及湖北、广西、河南、北京等 11 个地区，如图 2 所示。问卷的信度为 0.947，结构效度分析 KMO 值为 0.945。

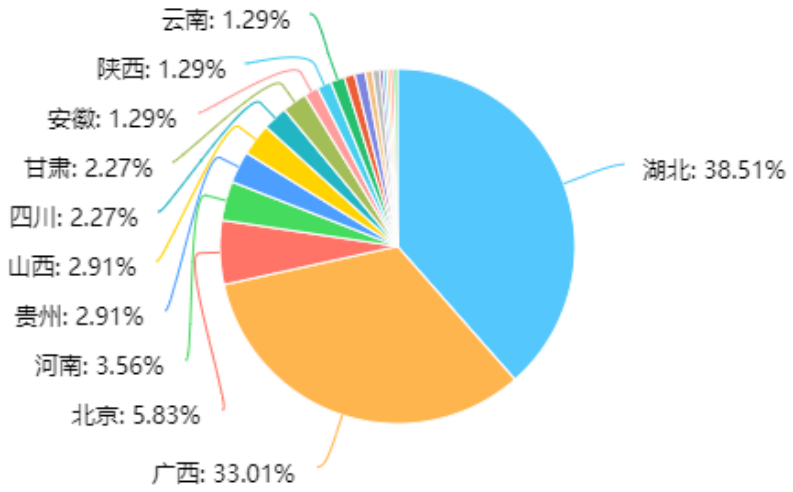


图 2　调研对象地理位置分布

3 数据分析

3.1 开放式编码

将所有访谈文本所有语句进行汇总整理，共得到 400 多句原始语句。开放式编码获取的节点都是见实编码，即自然抽取于受访者的原始话语。在进行编码时，选择 37 份访谈文本作为编码文本，剩下的 4 份文本（医学生和非医学生各 2 份）用来检验理论饱和度。编码主要由一位研究者进行，另一位研究者采用抽样编码的方式，将两位研究者的编码进行一致性检验后，调整部分有歧义的编码，并剔除频次小于 3 次的节点，初次共得到 30 个自由节点。部分自由节点如表 2 所示。

表 2　　部分高频自由节点

节点名称	参考语句（部分）

信息可靠性	（1）真实性很难得到保障；健康信息良莠不齐。 （2）良莠不齐的信息中无法判定哪种信息是具有可信度的；可能是造假的。 （3）被父母拉着看营销号，各种造谣信息看着头大。
健康状况把握	（1）根据描述结合自己的情况进行判断，从而了解自己的健康状况。 （2）便于管理和了解自身的健康状况。
信息易获取性	（1） 随时随地访问在线健康信息。 （2） 足不出户；比较方便。
信息有效性	（1） 无法正确判断自己的疾病。 （2） 容易给大众造成误导。 （3） 经常发生判断错误。
知识学习	（1）激励自己学习好专业知识；可以反复的学习。 （2）相关健康知识；很多课本上不一定有的知识。
病症联想	（1）当看到和自身症状差不多的情况的时候，会点进去看。 （2）参考其他相同症状的人。 （3）如果有一个患者跟你患了类型的病，但是病症不一样，你就会采取跟别人一样的方式来解决病情。

3.2 主轴编码

将 30 个自由节点进行二级编码，依据演绎与归纳，不断比较将近似编码链接在一起，梳理自由节点之间的逻辑关系[20]。共形成 7 个概念类属。

表 3 主轴编码薄

概念类属	自由节点
信息效用	信息可靠性；信息易获取性；信息有效性
信息心理	信息焦虑；信息迷航；从众心理； 不适可控
信息素养	用户表达能力；健康重视程度；信息技术水平；公众安全意识
系统环境	信息过载；隐私安全；医疗广告；信息全面性；平台权威性
社会信息环境	数据服务；社会认可
偏好习惯	高赞内容；高点击率；对症下药；病症联想
行为目标	自我诊断；知识学习；健康状况把握；节约成本；应急处理

3.3 选择性编码

选择性编码的主要工作是通过整合与凝练，通过概念类属整合出核心类属，通过译码范式连接主要概念类属。选择性编码见表 4 所示。

表 4 选择性编码表

核心类属	概念类属
行为态度	偏好习惯
主观规范	系统环境；社会信息环境；信息效用
行为控制	信息心理；信息素养
行为目标	行为目标

3.4 理论饱和度验证

理论性饱和验证主要是为了验证前期所提取的节点以及分析的类属关系是否已经涵盖完全。首先，对 4 份验证文本进行节点抽取和分析，看是否有新的节点产生，其次对问卷获取的数据进行分析，检验理论是否饱和。

文本分析：四份待检验文本中均无心理论节点产生。

问卷分析：

（1）信息获取主动性方面，根据五级李克特量表的分数，平均分为 3.16 分，约一半的人获取主动性一般，即有需求时才会获取，不会关注相关的公众号，在问答社区中也不会主动分享信息；23%的人比较主动，时常关注网络上的健康信息，在问答社区也会偶尔分享和交流，但是对 APP 推送的信息较为排斥；5.6%的人非常主动，会关注健康类公众号，问答社区信息交流频繁，并且还会接受各 APP 和浏览器推送的信息。

（2）获取信息类型方面：基本集中在这四个方面约 58%是养生类；61%是和自身健康状况相关；61%是获取专业相关的健康信息和专业知识；还有约 57%的被调查者表示会关注与公共卫生相关的内容。

（3）在线健康信息获取的阻碍因素：77.5%的大学生表示信息矛盾带来很大的困扰；约 72%表示信息筛选费时费力，不知道如何在获取的信息中取舍，由此带来的信息焦虑也很普遍，约 40%的人认为过多的信息会带来信息焦虑，容易自我怀疑和自我否定。

（4）在线健信息获取的正向激励：超过一半的受调查者认为信息质量对他们而言是比较重要的；超 90%的大学生认为获取的信息对自己的生活和学习是有所帮助，且愿意接受信息平台上他人分享的信息和建议，但是对于信息提供者的身份的权威性有所要求。

从验证文本和问卷分析的结果来看，访谈文本所提取的理论节点已经饱和。

4 结果分析

4.1 质性结果分析

通过文本编码归纳了核心类属有四个，行为态度、主观规范、行为控制以及行为目标。这四个核心类属都对最终行为的实行有所影响。由此绘制大学生在线健康信息获取行为引动图，如图 3 所示。

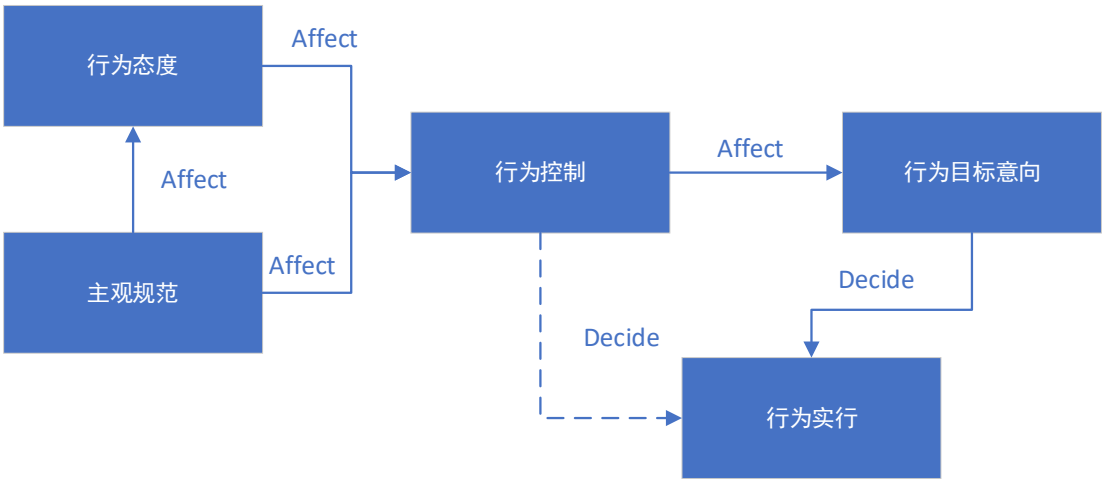


图 3 大学生在线健康信息获取行为引动图

在行为引动图中，实线部分表示行为态度、主观规范和行为态度均可以通过影响用户的行为意向来让用户决定是否实行信息行为。而虚线的行为控制指向行为实行表示当用户认为自己有足够的行为控制能力，能够决定自己在行为实行的过程中何时结束、如何进行等问题时，即使不一定有相应的行为目标也会实行信息行为。

四个核心类属之间互相影响，下面从行为态度、主观规范、行为控制和行为目标四个核心类属来分析行为引动机制。

（1）行为态度

根据期望-价值理论，行为态度是个体实行某种特定行为的重要信念以及对结果的一种预期和评价。对于大学生而言，是否通过在线获取的方式满足自己的健康信息需求首先考虑的是该行为究竟能够为自己带来什么结果以及是否可以满足自己的需求。当某一信息行为能够满足用户的信息需求与价值期望时，用户对这一行为会持有正向肯定的态度，从而促使用户产生信息行为。

用户的偏好习惯主要指用户对哪一类信息更加肯定，更容易接受。这类信息在一定程度上能够满足用户对于信息内容的期望。主要集中在对高赞和高点击率内容的倾向。这也能解释为什么用户对于信息获取平台的权威性要求比较多。当用户并不具备足够的专业知识辨别信息质量时，

用户会默认高赞和高点击率的内容在质量上可能会更高一些。此外，获取行为实行前用户是希望能够通过健康信息的获取找到对症的治疗方法并且可以从互联网上看到与自己病症相似的案例，这样的信息对于他们来说更具备说服力也更具有参考价值。

信息平台上提供的信息都具有一定的平台特性，对于那些符合用户偏好的信息平台更能够获得用户的肯定和青睐，激励用户从该平台获取信息。所以平台在前期设计、资源整合等方面需要考虑到用户的行为态度。

（2）主观规范

主观规范主要表现在外部环境对用户的信息获取行为带来的正向或负向的压力与刺激。作为信息社会原住民的大学生群体，对于外界环境更为关注，感知也更为敏感。首先体现在系统环境给用户体验带来的压力。其矛盾主要指向的是对于平台的不信任，如信息安全，自己的隐私是否会被泄露出去，平台的竞价排名问题导致医疗广告过多从而影响获取信息的真实性和可靠性存疑。因此，系统环境（包括平台的安全机制设置、平台信息是否全面和权威，信息是否只注重量而忽略质等）会在很大程度上影响用户是否要通过此平台获取信息。从本次访谈实录来看，大部分的被调查者对于现有的在线健康信息获取平台均不满意。

整个社会的信息环境主要表现在从整个国家的宏观层面是否认可在线健康信息的平台建设以及是否有比较健全的机制管控信息质量。大部分调查者表示如果从整个社会层面认可这一行为并且能够通过相关举措为用户提供高质量的健康信息将会正向激励大家。此外，和其他信息用户群体相比，大学生群体会将目光放在数据服务上。各大平台均能获取和记录用户的数据，这些访问记录能否得到良好的利用，借助大数据技术的东风为科研发展助力是很多受访者关心的问题。。而如果数据被非法利用，如窃取隐私推送医疗广告，制造假药等会阻碍用户信息获取行为。

（3）行为控制

行为控制主要是指信息用户在实行信息行为前对自己信息行为结果的预判，即自己是否有能力掌控行为的发展而不至于行为失控，偏离预期。本次研究中涉及到的行为控制类属概念主要有信息心理和信息素养。

信息心理：很多用户表示从网络上获取健康信息会使自己感到焦虑，尤其是信息之间往往会互相矛盾；此外，由于网上的信息过多，在信息筛选时受困于知识结构的限制，很难建立较好的信息筛选机制，导致越获取信息越觉得信息匮乏，在信息海洋中迷失方向，甚至会偏离预期的获取目标，导致行为失控，负向刺激用户的信息获取行为。如有受访者表示之所以不愿意获取在线健康信息是因为常常在获取过程中就不知道自己应该看哪些方面的内容，或者忘记自己最初的信息需求是什么。并且调查者纷纷表示当自我感觉自己的身体不适感处于比较轻微或者可控的情况

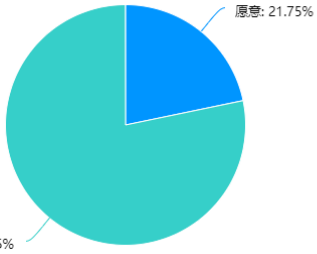
下才会考虑从互联网上获取健康信息或者进行线上问诊。在不适可控的情况下用户认为获取的信息可以帮助自己解决自身或者家人的健康问题，从而正向刺激信息获取行为。身体状况较为糟糕时用户在获取信息的过程中会逐渐产生无助感和恐慌感，这些反应都不利于信息获取行为的展开。从众心理是指当用户无法辨别信息真伪以及是否符合自己的健康状况时思维往往会极大地受大众评价的影响，特别是在公众健康危机时，用户在浏览信息的过程中更容易人云亦云，导致谣言的传播。当用户一旦成为虚假信息传播链上的一环时，其信息行为已经不再能够由自己控制，也偏离了自己最初获取信息的目的。

信息素养：信息素养主要表现在信息意识和信息能力两方面。信息意识指是否能够意识到自己需要哪方面的信息。当用户有比较清晰坚定的信息意识时，这份意识能够指引用户一直围绕自己的信息需求展开信息获取行为，而避免信息迷航信息能力即知道以什么方法获取信息以及对信息的理解和使用能力。在线健康信息的获取需要用户具备一定的信息检索和搜集能力，当用户具备较好的获取能力时可以很大程度上根据自己的需求借助不同平台和工具获取信息。而当用户没有相关的信息能力时，常常不能灵活变通获取策略和获取方式，导致信息获取行为不能完全受控于用户本身。

信息心理与信息素养使得用户能够了解什么时候应该获取信息以及当行为实施后自己是否能够按照预期计划顺利完成获取行为。

（4）行为目标

行为目标是用户获取行为的最终归宿。通过行为目标能够感知到用户的信息需求有哪些，为用户需求分析和在线健康信息平台建设提供一定的理论基础。常见的大学生行为目标主要有：自我诊断，即在不适可控的情况下从网络上获取相关信息判断自己的健康状况；知识学习：主要集中在养生类健康知识的学习。而随着新冠肺炎的爆发，大学生群体在疫情期间对于公共卫生安全类信息的获取量增多，如防疫知识、心理健康知识。对于医学生而言，医学知识的更新速度日益加快，继续深造学习的需求也越来越大，传统的课堂教学和教材已经无法满足本科生的知识需求。因此很多医学生会倾向利用互联网平台获取和本专业相关的知识，如关注医学类的公众号、查找文献、关注最新的前沿动态等等。健康状况把握主要是除了了解自身的健康状况外，往往还会关注家人或者朋友的健康状况；应急处理常见于突发身体不适又无法及时就医时会选择从健康知识社区或者问答平台上获取解决方案；节约成本主要是指用户普遍认为在线健康信息获取方便，随时随地可以搜寻，免去医院挂号和排队的时间，节约了时间成本。此外网络上的信息大部分是免费信息，在一定程度上可约资金成本。在问卷调查中可以发现，接受调查的 310 位受访者，有 223 位不愿意获取付费的健康信息，占比 78.25%，如图 4 所示。



4.2 相关性分析

（1）学科属性与信息判断的关系

通过 SPSS 的相关性分析，学科属性与是否能够借助自己的学科知识判断信息质量这两者之间呈显著相关关系。

表5 学科属性与信息判断的相关关系

相关性

		学科分类	专业知识能够帮助我判断信息质量
学科分类	Pearson 相关性	1	.196**
	显著性（双侧）		.001
	N	310	310
专业知识能够帮助我判断信息质量	Pearson 相关性	.196**	1
	显著性（双侧）	.001	
	N	310	310

**. 在 .01 水平（双侧）上显著相关。

学科分为“本科 医学生”和“本科 非医学生”，Pearson 相关性分析结果为显著相关。调查者中医学生纷纷表示可以借助自己所学的医学知识大致判断信息是否真实可靠，能够帮助自己过滤掉大部分不太可靠的信息。但是非医学专业的学生表示无法自我判断信息可靠与否。

根据此结果可以思考，大学生作为知识的有力传播者，虽然无法让非医学生掌握大量的医学类专业知识，但是如果能够在平时普及好基本医疗常识，也能够在一定程度上帮助大学生进行信息判断。当面临突发的公共卫生危机时才能更好的发挥大学生群体的作用。

（2）基本人口学信息与是否获取在线健康信息行为的关系

本次问卷所涉及的人口学基本信息包括性别（男、女），家庭住址（农村、城镇），家庭月收入（3000 以下、3000-5000、5000-10000、10000-15000、150000-20000、200000 以上），将这些信息与是否获取过在线健康信息行为之间的相关关系进行分析，无显著相关关系。在受访者中，超 90%的人有过信息获取行为。虽然对于其他群体，收入、生活环境会影响其是否会借助信息平台

获取信息，但是对于大学生而言这些并不影响信息行为，原因可能在于，一是作为接受高等教育的大学生，其信息意识比其他群体更为强烈，更能适应信息社会的信息获取和交流方式；二是其硬件环境更好，虽然有很多家庭条件贫困的大学生在原住地受限于经济条件并没有机会在网上获取信息，但是进入大学校园后，学校内较为完善的信息环境、为了社交和学习等目的大学生普遍都具备相关设备帮助其获取信息；三是信息获取能力更好，长期生活在丰富、便利的信息环境中，加之学校会进行信息素养和信息技能的培训，大学生群体的信息获取能力更好一些。因此这些基本人口学信息并没有很显著的影响群体信息获取行为。

4.3 引动机制应用探讨

目前市场上的健康信息平台主要分为四类，如表 6 所示：

表 6 健康信息平台类别表

类型	特点	缺陷	代表平台
搜索引擎类	以搜索资源为主	竞价排名，利益为导向，信任度低	百度
社交类	以资源共享，经验交流为主	缺乏审查机制，专业水平参差不齐	知乎
医疗服务类	拥有专业资源	资源集中度不高，信息少而不全	好大夫
咨询类	信息咨询	平台门槛低，医生资质审核问题	快速问药

依据引动机制，以各大类型平台现存的问题为主探讨如何实现各方效率，如大学生如何快速获得准确、有效的健康信息，平台如何吸引用户，监管部门如何保证信息的安全、可靠。

（1）信息过滤机制：有效的信息过滤机制能够帮助用户去除低质且无关的信息，给用户呈现更直接、准确的信息。首先减少广告的硬性推广，用户可选择无广告模式的服务。大学生群体普遍对于各大平台上推送的医疗类广告存在排斥怀疑的态度。这也是大学生群体不愿意获取付费信息的原因之一。平台应该首先提升用户对平台的信任度，增强用户粘性和忠诚度后才可以获得更多的用户流量，后期流量变现才能为平台带来更多更长远的利益。

其次，提供更专业的信息搜索功能，针对不同类别的用户应该提供不同的搜索方式。大学生群体对于检索结果的精准性有着较高的要求，专业的检索入口更能符合他们的需求，尤其是想获取医学专业性知识时。

（2）平台监管：网络监管和线下监管应该联合起来，线上对于信息的发布者可以以评分制的方式，定期进行评分公布，提升用户活跃度，激励用户提供更优质的信息。对于发放虚假信息或者散播谣言的用户禁止其他所有权限，仅可以浏览信息。人员准入方面：对于平台上的注册用户，其身份审查应该更为严格。定期审查医疗类从业工作者的从业资格，及时更新其工作属性，使信息更透明。线下对于投放在平台上的广告需要严格审查其真实性和合法性，比如药物其生产厂家的资质、医院的资质和环境等，做到真实可靠，并且定期公布资质审查结果。此外，国家层面也需不断完善相应的法规文件，规范信息平台的各种行为。

（3）资源整合：大学生群体对于平台信息比较关注的地方有日常养生信息、科普类信息、辟谣信息外，对于知识学习的需求比较大。平台除了可以不断丰富其他板块的信息，提供全面的资源外，还可以为医学生们创建在线交流区，学生可以和有工作经验的医生或者从业人员进行学术探讨，经验交流、问题解答等。如果能够借助知识管理的方法和理念，管理平台上的资源，进行知识挖掘，为学生提供更多学术上的资源可以提高平台效用。

5，总结

在健康中国的发展中，大学生群体对于健康知识的需求与贡献是不可忽视的非常重要的一部分。研究大学生的在线健康信息获取行为的引动机制，了解信息行为实行的前提条件是什么，才能知道如何能够更好的为该群体提供优质的信息服务，让高质量的健康信息满足他们的生活需求，更能够满足他们的学习需求。同时，青年大学生群体可以作为知识的传播者，他们所掌握的良好的健康信息可以扩散给身边的朋友和家人，也能够帮助其他群体建立良好的信息获取习惯和信息意识。

本文从信息获取行为的引动机制入手，分析刺激用户的信息意识转向信息行为的主要因素。研究的不足之处在于用户样本量有待扩充。在后续的研究中，进一步扩大样本量，除了研究本科的医学生和非医学生外，还可以将硕博士加入到研究范畴，其次在后续的研究中对于引动机制在平台上的应用可以进行更为深入的探讨，结合知识管理，构建更优化的健康信息平台。

OR-019

精神类疾病在线健康社区用户性别差异研究
——以微医网为例

王安娜[1]、向菲[2]
1. 华中科技大学同济医学院 2. 华中科技大学同济医学院医药卫生管理学院

目的 本研究从用户认知需求和表达需求两个方面探究精神类疾病不同性别用户在咨询医生的过程中择医行为和主题差异。[方法/过程]本研究通过采用主题分析方法提取问答记录中男性用户和女性用户咨询主题偏好，分析患者在社区中的认知需求差异。针对用户咨询行为差异采用 T 检验和单因素 ANOVA 检验。[结果/结论] 择医行为方面男性用户更倾向于选择预约量和问诊量多，图文问诊价格高的医生，女性用户更倾向于选择副主任医师和三甲医院医师。主题方面男性用户询问病情且期望从中得到一些建议，女性更多地集中于咨询的途径；男性关注工作、学校生活，而女性生活方面更倾向于谈论人际关系交往；男性集中于谈论有效检查诊断的手段，女性更注重药物治疗方式。研究用户的认知需求差异，可为有相似信息需求的用户之间建立沟通渠道，了解用户个人偏好从而为在线健康社区个性化推荐，增强在线咨询效果，促进医患友好交流，减少医患纠纷。

1 引言

1.1 精神类疾病患病率攀升

精神障碍是世界范围内重点关注的疾病，其特征为高发病率，高资源利用率和高残疾率[1]。世界卫生组织系列研究显示，从 1990 年到 2010 年，精神障碍占全球疾病负担中的比例增加了 37%，其中 14%至 33%的成人残疾与精神疾病有关[2]。到 2015 年，精神障碍和药物滥用占全球疾病总负担的比例上升至 18.94%。

在过去 40 年的快速发展中，中国不仅是世界上人口最多的国家，而且人们的生活方式和家庭结构也发生了巨大变化[3]。随着经济快速发展，社会竞争逐渐激烈，生活节奏不断加快，生活工作各方面的压力也随之而来，多种因素和风险持续对人们的身心健康产生负面影响[4]，由此引发的心理问题日益突出，精神疾病患病率不断攀升，精神疾病在全国各地变得更加普遍。根据流行病学调查表明，大约 1.73 亿人有精神疾病，包括焦虑、抑郁和强迫症等，其中 1.58 亿人从未接受过专业治疗，情绪障碍的终生患病率已达到 16.6%[5]。由于精神疾病治疗时间漫长，导致治疗成本过

高，精神疾病负担也相应增加[6]，现已成为全世界共同关注的问题。2013 年中国精神疾病和药物滥用的疾病负担占全国疾病总负担的 10%，占全球精神疾病和物质滥用总负担的 17%[7] 。

由于精神疾病带来沉重的疾病负担，我国政府也更加重视精神卫生工作。近年来，我国先后出台《国务院关于实施健康中国行动的意见》《全国精神卫生工作规划（2015-2020 年）》《关于加快民政精神卫生福利服务发展的意见》等文件，对精神卫生防治体系建设提出了具体要求。

1.2 在线健康社区发展迅速

如今大数据时代的到来以及信息技术不断发展，人们可以通过网络搜寻相关健康信息，可以迅速便捷的接触到大量医疗健康知识[8]。随着这些知识的普及，患者的健康意识逐渐增强，传统医疗模式逐渐发生转变，在线健康社区应运而生。相比过去传统的医疗模式，患者可以通过在在线健康社区中咨询病情，交流经验，分享健康信息，获取知识，实现自我健康管理、疾病监控、参与自身医疗过程等方面的需求。

由于我国心理健康教育起步较晚,社会大众对精神疾病的了解比较匮乏[9]。并且精神疾病污名化[10]以及大众对精神疾病的刻板印象，使得患者饱受社会偏见的消极影响，可能对自身病情难以启齿甚至延迟就医导致病情恶化。与其他疾病不同，精神疾病的治疗较少依靠物理器械，大多通过医生与患者的交流咨询途径解决患者的问题[11]。在线社区中用户通过匿名的方式咨询，可以很好的解决目前现实生活中羞于与医生探讨病情的问题，减轻对精神疾病的羞耻感。目前用户参与在线健康社区主要以咨询病情、交换信息，表达意见，寻求情感支持、与他人建立社会关系为主。不仅可以在社区中咨询医生病情相关问题，还可以同其他人分享个人的经验，讨论医疗相关信息[12]，如手术、诊断、治疗过程等。提高搜寻有关健康信息的能力，获得病情相关知识。与其他人建立社交联系，可以讨论除疾病诊断等专业知识外的其他主题[13]，例如与日常生活，药物副作用、相关症状描述等问题。还可以通过同其他用户日常交流获得情感支持[14],这些都能在一定程度上有效减轻病情[15]。患者参与社区行为当中咨询医生是得到权威医学信息最迅速、最直接的渠道，这也是在线健康社区最重要的功能，病人提问不仅是一种寻求信息的方法，同样也是病人参与医学对话的一种手段[16]。所以本文选用了在线健康社区中用户咨询内容进行研究。

1.3 性别差异影响个人需求的认知与表达

性别是个体间最根本的区别之一[17]，个人除了在生物学方面，在行为上和认知方面都会存在性别差异[18]。用户的健康信息行为，即用户寻求、获得、评估、分类和使用与健康相关的信息以执行所需的健康行为的方式[19]。不同性别的用户在参与在线健康社区过程中行为可能会因为性别不同有很大差异[20]，例如男性和女性患者在医疗卫生方面的需求通常有较大差异[21]，相比于男性患者，女性更多关注和参与与生活、家庭相关的话题。男性与女性由于在社会中的角色不同，承

担的责任和社会分工也有很大差异，这在一定程度上决定了男性和女性的健康信息行为同样存在差异。男性与女性在社区中的情感需求也不同，女性患者的情感表达力度更强[22]。

为了有效地促进和加强预期健康行为的表现，应针对用户的兴趣、素养、语言、文化背景、情感状态等个人信息进行完善，与健康相关的信息的质量要尽量符合患者和需求与期望[23]。研究在线健康社区中不同性别的用户需求差异，对今后社区系统的构建，个性化、精细化的推荐做出一定贡献，更好的为不同性别的患者提供不同的服务，完善在线健康社区系统，促进医生与患者之间的友好交流互动，增强在线精神科室咨询效果，减少医患纠纷，减轻精神疾病负担。

2 相关理论及国内外研究现状

2.1 社会性别理论及用户信息需求机制

社会性别一词用来指社会文化形成的对男女差异的理解，以及在社会文化中形成的属于男性或女性的群体特征和行为方式。性别差异的先验经历会使男人和女人的技能和态度有所不同，从而导致个人特征和行为方式有所不同[24]。由于个体在不同的政治、经济、文化等环境下所形成的社会角色不同，会导致个体在需求、行为方式、情感表达等方面不同。

信息学家科亨将用户信息需求划分为三个层次[25]：客观状态层次的信息需求，这是一种不以用户主观意志为转移的客观需求状态，属于潜在需求，是由用户所处的社会背景、工作环境、职业性质等客观条件所决定的，用户意识不到且无法表达；认识层次的信息需求，用户认识到了自己的信息需求，受用户自身知识结构、信息素养等因素影响，无法准确表达，需要对他人进行咨询；表达层次的信息需求，用户经过启示和咨询将自己的信息需求表达出来。受用户认知能力、表达能力等因素影响。用户可以准确认识且表达出个人需求。本文将用户咨询主题作为认知层次方面的需求，咨询行为作为表达层次需求来研究。

2.2 国内外研究现状

Fernandes, LD 等人分析强调了网络文化对健康和疾病过程的影响，导致医患关系，患者授权和对慢性病的个体管理发生变化，发现患者可以通过社区互动获得有关疾病和治疗的信息，分享经验和社会支持，有助于增加对治疗的依从性[26]。Xuan Liu 等人基于主题建模分析，情感分析和社交网络分析方法分析了糖尿病社区（甜蜜家园）的数据，调查慢性在线健康社区中的男性和女性参与者行为特征和需求差异[21]。Chin SiangAng 通过调查问卷的形式发现男性用户更容易建立新的友谊，而女性用户则更倾向于维持在线社区中与自身兴趣、认知水平相仿的用户的关系[27]。Pei-Lin Chen 等人通过比较老年人的基本特征、健康促进行为和生活质量的差异，分析性别差异和与生活质量相关的因素之间的关系，发现性别，教育程度，抑郁水平和健康饮食是影响老年人生活质量的关键因素[28]。

国内相关研究较少，邓胜利和管弦基于 TAM2 和 IAM 模型设计问卷，调查了问答平台用户健康信息搜寻行为的差异，发现用户选择问答平台为了得到有针对性的信息，且在提问时容易受到他人影响[29]。施亦龙与许鑫通过比较中美两个最大的在线问答社区用户对自闭症的认知差异，发现美国用户在社区中的提问更积极、详细多样，答案的选择更加完整准确[30]。司莉和舒婵以在线健康社区问答中糖尿病主题相关的 10 年数据为例，揭示在线医疗社区的患者、医生及问答记录的特征，从患者患病时长，过敏原，医生职称，医院等级，患者提问主题几个方面入手，研究医患群体问答记录特征[31]。

通过梳理文献可以发现国内外在线健康社区用户性别差异研究数量较少，研究主要集中于在线健康社区，将社区和性别相结合的研究较少，且研究在线健康社区中性别差异主要应用于用户健康行为研究，大多采用问卷调查、定量描述的方法，对于问答过程中文本的研究较少，所以本研究通过比较不同性别用户需求差异，研究医患问答特征，具有一定创新性。

3 研究方法

2.1 数据获取

本文基于 python 爬虫数据对在线医疗社区中医患问答相关特征进行分析，获取了截止 2020 年 4 月微医网精神科室医生的问答记录共 218808 条，问答数据包括患者信息（性别、年龄），医生信息（所属医院、职称、患者满意度评分、关注量、预约量、问诊量、问诊价格），咨询日期及答复日期等。

2.2 数据处理

抓取得到的原数据中存在大量无法识别字段、无意义文本、空字段等无效数据，需要对其进行预处理。依据 ICD-10 对问答记录中的疾病名称进行统一规范，由于网页无法打开或某些字段无法获取导致数据缺失，需要去掉缺省值。对医院等级、医生职称等数据进行编码。

2.3 分析方法

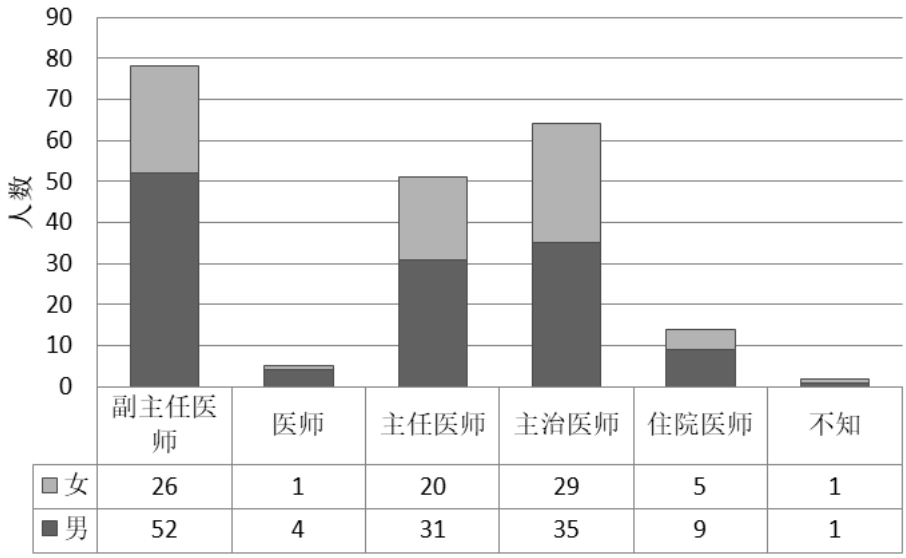
LDA 是由 Bleide 等提出的包含文档-主题-词 3 层贝叶斯文档主题生成模型，作为一种无监督机器学习方法，用来识别隐藏在文档集或语料库中的主题信息[32]。本研究通过采用主题分析方法（LDA）提取问答记录中男性用户和女性用户咨询主题偏好，判断信息质量，分析患者在社区中的认知需求差异。针对用户咨询行为差异采用 T 检验和单因素 ANOVA 检验。

3 研究结果

3.1 用户社区使用差异

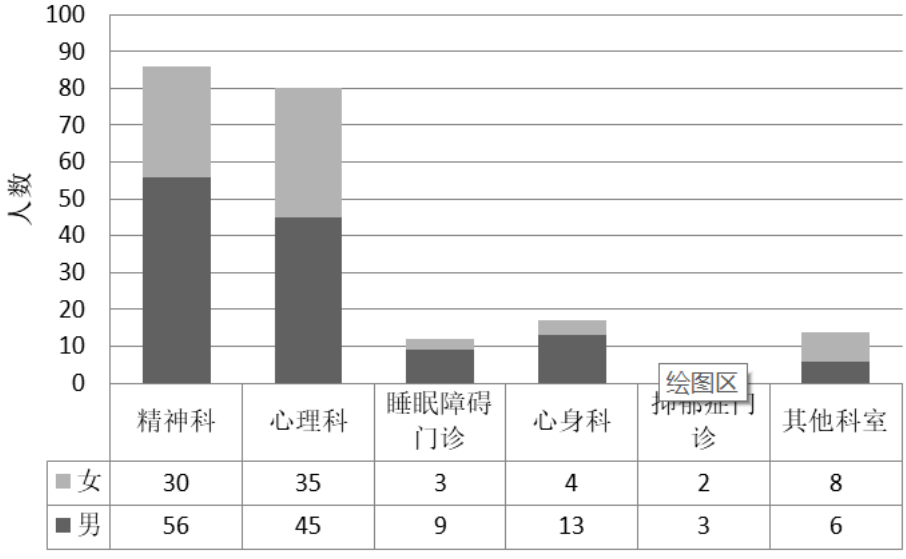
根据统计结果显示，微医网精神类疾病医生 606 名，其中有回答用户提问记录的医生共 214 名，男 132 名，女 82 名。统计有有效问答记录的医生职称、科室与性别分布结果如图 1 和图 2，

根据结果显示，拥有有效问答记录医生职称最多的为副主任医师（78），其次为主治医师（64），最少的为医师（5），其中男性副主任医师占大多数，女性主治医师居多。副主任医师共回复 53351 条问题，其次为主治医师（38416）和主任医师（38133）。大多医生来自于精神科和心理科，还有部分属于睡眠障碍等咨询门诊，极少属于其他科室。



	副主任医师	医师	主任医师	主治医师	住院医师	不知
女	26	1	20	29	5	1
男	52	4	31	35	9	1

图 1 医生性别与职称分布图



	精神科	心理科	睡眠障碍门诊	心身科	抑郁症门诊	其他科室
女	30	35	3	4	2	8
男	56	45	9	13	3	6

图 2 医生性别与科室分布图

统计向上述医生提问的用户共 12207 名，其中女性用户 7512 名，男性用户 4695 名。不同性别的用户年龄分布如图 3 所示，男性和女性用户都集中在 21 岁至 30 岁之间，超过 50%的用户年龄范围为 21 岁至 40 岁，在这个年龄层的用户精神压力较大且具备熟练使用电子产品的能力，由于精神类疾病病程较长，现实中无法立即就医的情况下，通过在线咨询自身病情便成为用户的首选。但不同性别用户各个年龄层的占比不同，虽然男性用户和女性用户集中的年龄层相同，但 21-30

岁年龄层中女性用户的占比要更大（39.4%/32%），31-40 岁中男性用户占比较大（23.7%/21.3%）。41-50 岁男性用户（12%）比女性用户（8.9%）人数多，这也值得我们关注。

根据图 4 可知，用户向医生咨询时间分布规律，有效咨询最早时间为 2014 年 11 月 26 日，该用户为女性用户，男性用户最早咨询时间为 2015 年。2014 年起，每年的咨询数量逐年攀升，2017 年至 2019 年间数量上升速度最快。

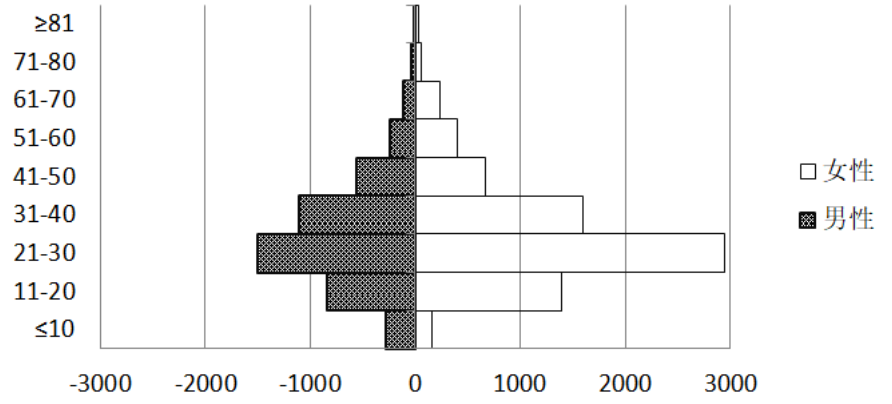


图 3 用户年龄分布图

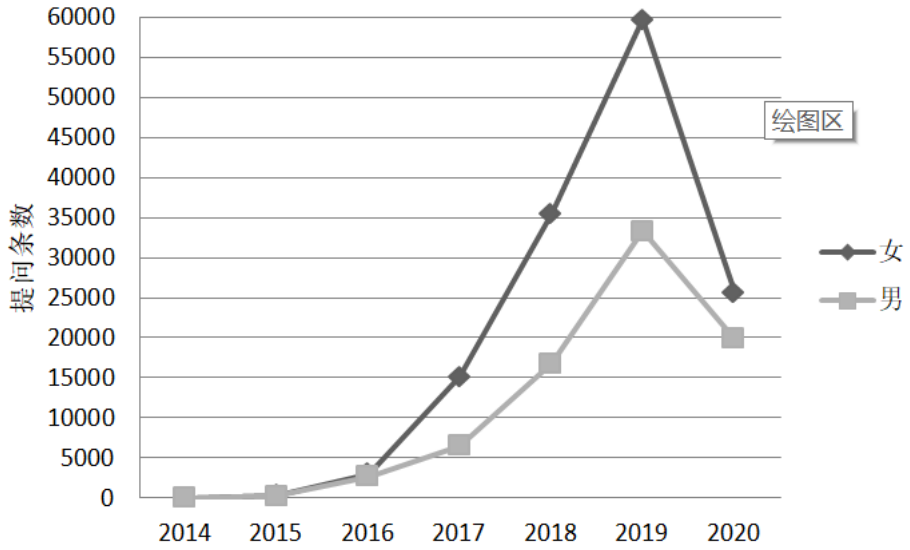


图 4 用户咨询时间分布图

3.2 用户咨询主题差异

筛选所有用户向医生咨询的内容，利用 python 结巴（jieba）分词进行分词并统计不同性别用户咨询词词频，词频排名前 40 如表 1。从表中可以看出用户排名靠前的有“孩子”、“工作”等词，说明无论时男性还是女性用户的精神状况都与家庭、社会相关。

不同性别用户咨询词虽然有部分重叠，但排名略有不同。男性用户咨询医生时大多用“问题”、“怎么”、“请问”、“是不是”等类似提问词，说明男性用户大多用提问的方式咨询病

情；而女性用户咨询词排名靠前的为“情绪”、“抑郁症”、“焦虑”、“睡眠”等词，说明女性用户在咨询医生时描述自己病情居多。女性用户在咨询过程中提到孩子、老公的次数较多，且孩子的次数排名第四，说明相对于男性来说，女性情绪更容易受到家庭影响。

表 1 不同性别用户咨询词频统计表

男性用户				女性用户			
词	词频	词	词频	词	词频	词	词频
医生	4747	正常	785	医生	9111	工作	1423
感觉	2412	药物	777	谢谢	3607	咨询	1384
医院	2018	失眠	771	医院	3529	怎么办	1380
谢谢	1916	抑郁症	766	孩子	2576	正常	1301
问题	1464	抑郁	718	情绪	2407	睡不着	1290
治疗	1347	睡觉	702	问题	2372	害怕	1281
怎么	1266	严重	683	抑郁症	2228	睡觉	1256
焦虑	1166	怎么办	665	焦虑	2174	心里	1253
孩子	1162	精神	642	需要	2116	检查	1239
请问	1039	咨询	629	抑郁	2116	药物	1221
症状	1031	应该	611	治疗	1980	心情	1194
情况	1025	工作	611	情况	1955	难受	1071
睡眠	934	时间	599	吃药	1876	服用	1068
需要	921	担心	590	睡眠	1855	喜欢	1041
是不是	910	主任	584	失眠	1773	时间	1040
吃药	879	有时候	577	有时候	1566	担心	1018
检查	833	效果	561	事情	1548	老公	989
服用	831	已经	548	严重	1511	开心	978
情绪	822	事情	539	症状	1475	影响	972
心理	818	睡不着	536	心理	1439	精神	969

在使用数据集训练 LDA 模型时，为了得到最佳的主题模型，需要确定最佳的主题数 K，Blei 等在其研究中详细阐述了使用困惑度（Perplexity）作为确定主题数的标准，通过绘制 Perplexity-topic number 曲线，寻找最佳主题数[33]。本研究计算不同性别用户咨询主题困惑度如图 5 和图 6，由图可知，男性和女性用户咨询主题困惑度在主题数为 4 的时候最低，所以将主题数 K 定为 4。

图 5 男性用户咨询主题困惑度

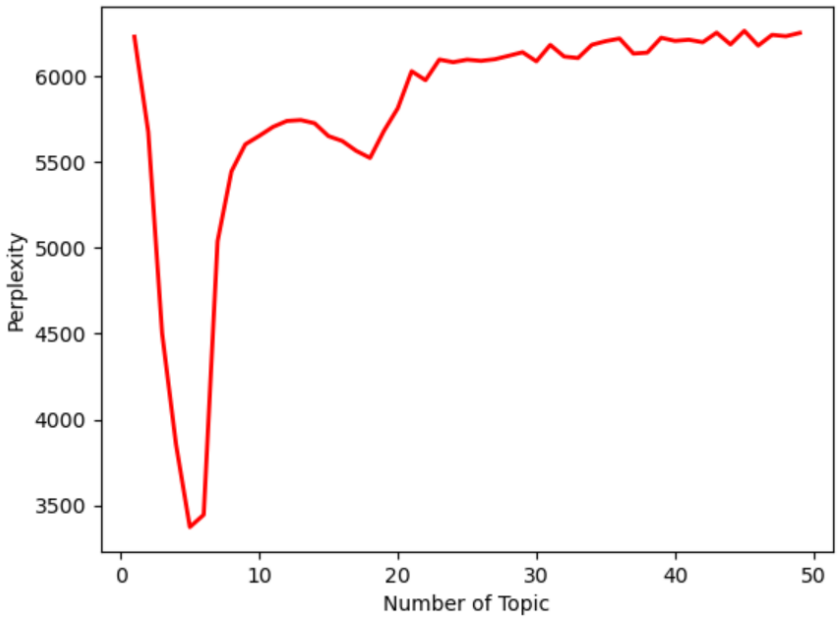
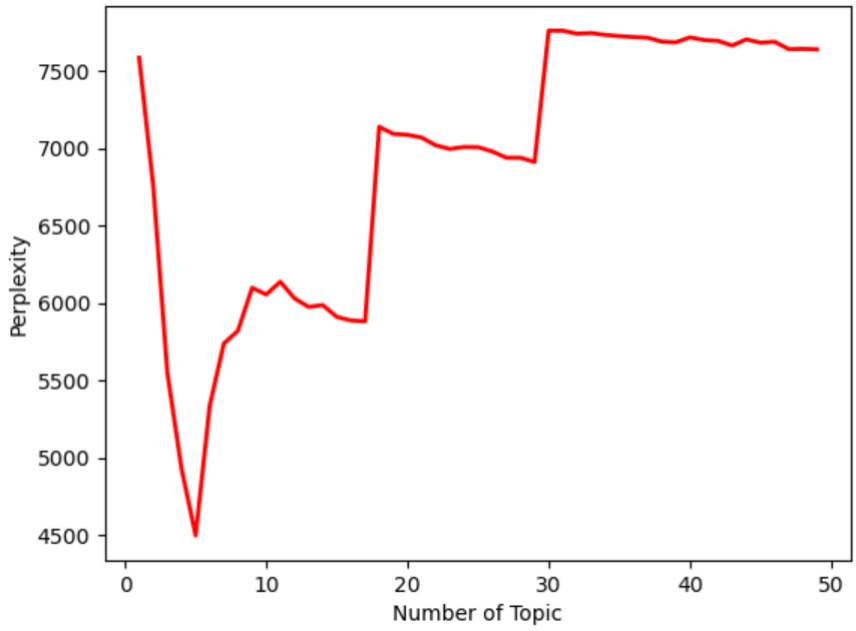


图 6 女性用户咨询主题困惑度



根据男性和女性用户主题聚类结果（表 2），将前 3 类主题分别命名为病情咨询、症状描述、日常生活，最后一类主题男性用户被命名为检查诊断，女性用户为疾病治疗。

首先症状描述方面，由于精神类疾病的症状大多相似，不同性别用户都出现情绪、睡眠等方面不适，临床表现为失眠、抑郁、焦虑等症状，所以在此方面没有较大差异。病情咨询方面，根据主题聚类结果显示，男性用户向医生咨询的时候使用“建议”“是否”等词来集中于询问病情且期望从中得到一些建议；女性则更多的使用如“预约”、“门诊”、“住院”等词语咨询，更多地集中于咨询的途径。日常生活方面，男性和女性用户谈论的内容都涉及到了家庭成员，男性

主要谈论孩子、父母，女性则谈论孩子、老公更多。除家庭以为，男性同样关注工作、学校生活，而女性生活方面更倾向于谈论人际关系交往。最后一类主题男性用户为检查诊断，由主题下词语可知男性集中于谈论有效检查诊断的手段，如心电图、CT 等。女性用户为疾病治疗，女性更注重药物治疗方式，主题下词语较多为治疗精神类疾病药品及其服用方法。

表 2 不同性别用户咨询主题

男性用户				女性用户			
病情咨询	症状描述	日常生活	检查诊断	病情咨询	症状描述	日常生活	疾病治疗
治疗	焦虑	吃药	检查	治疗	情绪	孩子	服用
孩子	症状	工作	害怕	咨询	抑郁	喜欢	停药
服用	严重	强迫	身体	药物	睡眠	学校	帕罗西汀
药物	睡眠	学校	口吃	检查	失眠	说话	盐酸舍曲林
咨询	情绪	停药	住院	预约	焦虑	害怕	半片
中药	抑郁症	朋友	心跳	在线	症状	老公	复发
是否	失眠	孩子	心电图	门诊	工作	家里	草酸艾司西酞普兰
建议	精神	父母	大便	接诊	心情	学习	看病
预约	难受	上学	血压	住院	睡不着	女儿	胶囊
诊断	头晕	母亲	CT	挂号	影响	朋友	奥氮平

3.3 用户咨询行为差异

根据 T 检验结果（表 3）可知，微医网上向精神类疾病相关医生咨询的用户中男性和女性的年龄无显著性差异，且男性和女性用户在选择医生时医生关注量、视频问诊价格及用户满意度打分均无显著性差异。由于微医网上医生的评分大多集中在 9.0 至 10.0，整体差异不大，所以用户在选择医生时满意度评分的参考价值不高，所以男性和女性用户就医选择中满意度评分无显著性差异。关注量是指在微医网平台持续关注该医生动态的人数，由于大多用户咨询医生时重点关注医生职称、所属医院等属性，持续关注某一医生的人数较少，所以男性和女性用户在选择医生时在关注量方面没有差异。在微医网上，用户向医生咨询有两种方式，一种为图文问诊，一种为视频问诊。两种方式中视频问诊的价格相对更高，可为图文问诊的两倍以上，所以采用视频问诊方式的用户较少，不同性别的用户在选择医生时对于视频问诊价格的关注同样很少，没有显著性差异。

但在医生的预约量、问诊量及图文问诊价格三个方面均值有显著性差异。根据检验结果显示，相对于女性用户来说，男性用户更倾向于选择预约量和问诊量多，图文问诊价格高的医生。

表 3 T 检验结果

	男性用户				女性用户				P 值
	个案数	均值	中位数	标准差	个案数	均值	中位数	标准差	
年龄	4674	30.51（±0.41）	29.00	14.143	7498	30.74（±0.30）	28.00	13.551	.381
关注量	4695	318.48（±15.08）	77.00	527.214	7512	304.92（±11.60）	70.00	512.919	.162
预约量	4695	557.76（±39.73）	6.00	1388.858	10329	430.54（±23.98）	0.00	1242.870	<0.001*
问诊量	4696	350.41（±13.36）	176.00	467.054	10328	192.51（±6.19）	95.00	321.178	<0.001*
图文问诊	4695	85.68（±3.42）	60.00	119.637	10329	61.87（±2.04）	50.00	105.329	<0.001*
视频问诊	4281	207.45（±5.12）	200.00	170.954	6816	210.56（±4.00）	200.00	168.749	.347
满意度	4669	9.77（±0.01）	9.80	0.235	7476	9.77（±0.01）	9.80	0.231	.600

医生职称有四种，分别为医师、主治医师、副主任医师和主任医师。医院等级为一级、二级、三级，由于三级中三级甲等占大多数，所以将三级甲等单独作为一个类别。将医生职称和医院等级按上述类别编码分析。统计用户选择的医生的职称和医院等级，发现在医师职称的选择中，男性用户主治医师、副主任医师和主任医师三种比例基本相同（30.16%、30.71%、30.16%），女性用户选择副主任医师最多（35.41%），其次为主任医师（30.23%）。医院等级选择中，男性用户选择三级甲等医院最多（77.89%），其次为二级医院（19.55%），女性同样选择三级甲等医院最多（83.32%），其次为二级医院（12.82%）。

表 4 χ^2 检验结果

	医生职称			医院等级		
结果	值	自由度	P 值	值	自由度	P 值
	37.547	3	<0.001*	110.170	3	<0.001*

本研究医生职称和医院等级任一期望频数均大于 5，可以采用卡方检验。结果如表 4 显示医生职称χ^2=37.547，P<0.001，医院等级χ^2=110.170，P<0.001，P 值均小于 0.001，说明男性用户和女性用户在选择医生时医生职称和医院等级方面的选择有显著性差异。就医生职称来说，相对于男

性，女性用户更倾向于选择副主任医师。女性和男性用户都更愿意选择来自三级甲等的医生，但女性的意愿要比男性更加强烈。

4 讨论

随着互联网的蓬勃发展，为公众提供了较为广阔且其规模日益增大的健康平台，以供用户日常参与、咨询及交流。这种参与在一定程度上可以表明患者具有足够的知识，能力和意愿参与自己的健康管理和疾病治疗过程，并有助于患者获得更好的健康管理成果，更好的治疗经验以及更低的健康管理成本[34]。

男性和女性不仅在现实生活中而且在虚拟生活中扮演的社会角色不同，从而产生不同的兴趣爱好以及物质和情感需求，双方在使用互联网和需求表达方面有不同的模式[35]。在先前的研究中发现，在使用基于文本的交流中，女性用户数量已经超过了男性，相对于男性而言，女性用户与外界建立的联系更多并更熟练地使用中介技术[35]，即女性使用社交网站用户的比例高于男性，对于使用网络作为收集信息的主要资源的态度表现更积极。这也就可以解释平台用户数量上的性别差异。

性别差异也会导致男女之间主观健康的差异[36]。通常相对于男性来说，女性的寿命更长，在整个生命过程中健康问题更多，并且慢性疾病的发生率更高，但致命的疾病发生率较低[37]。所以，男性在判断自己的健康状况时可能会更多地关注会威胁生命的条件，而女性则可能会关注慢性疾病发生的可能性[38]。精神类疾病治疗病程长，初期症状不显著，由于男性往往对自己的健康状况更为乐观[39]，女性对自身健康状况更加敏锐，所以男性对自己精神状况的关注度没有女性高。加强相关知识教育，提高男性主观健康意识，提升对自身健康状况敏锐性也是需要我们今后关注的地方。

从研究结果我们可以看到女性在咨询过程中提到较多药品名称，相对于男性用户更关注治疗方式，主题下词语较多为治疗精神类疾病药品及其服用方法，而男性用户关注的是检查方式。与医生交谈过程中，男性和女性用户日常生活方面谈论的内容都涉及到了家庭、工作、学校生活等当面，在咨询过程中，女性经常谈论自己的私人生活，例如家人和亲密朋友，男性虽然也提到了家庭成员，但更倾向于谈论工作等话题[40]，可以看出女性用户更受家庭影响，更多关注和生活、家庭相关的话题。

社会角色差异可能会导致用户在就医选择行为上的差异。对于医生的选择男性用户更倾向于选择预约量和问诊量多，图文问诊价格高的医生。女性用户更倾向于选择副主任医师。女性和男性用户都更愿意选择来自三级甲等的医生，但女性的意愿要比男性更加强烈。由于女性对家庭和家庭责任的承担过高，收入比男性少，甚至在做完全相同的工作时男女之间也可能存在工资差异

[41]，所以导致在就医选择中不同性别的患者选择不同，男性在价格方面的选择更自由。在家庭中女性更多承担照顾家人的责任，且女性更倾向于以积极态度寻求心理健康专业人员的帮助并利用与医疗卫生相关的资源[42]，所以在社区中为家人、朋友咨询的更多是女性。

中国公众对于健康关注度、健康风险意识，与受教育程度息息相关，因为缺乏相关健康知识，疏忽病情导致送医不及时或者过度在意对病情十分恐慌的病例还有很多，所以要加强健康教育，从而养成正确的防护措施和良好的个人卫生习惯，正视自身健康问题，提升对自身健康状况的敏锐性。

5 结论

本研究从用户认知需求和表达需求两个方面探究精神类疾病不同性别用户在咨询医生的过程中择医行为和主题差异。择医行为方面男性用户更倾向于选择预约量和问诊量多，图文问诊价格高的医生，女性用户更倾向于选择副主任医师和三甲医院医师。主题方面男性用户询问病情且期望从中得到一些建议，女性更多地集中于咨询的途径；男性关注工作、学校生活，而女性生活方面更倾向于谈论人际关系交往；男性集中于谈论有效检查诊断的手段，女性更注重药物治疗方式。研究用户的认知需求差异，可为有相似信息需求的用户之间建立沟通渠道，有利于医生指导不同性别的用户，有效了解男性和女性用户在医患沟通过程中更注重的因素，个人偏好等相关信息，从而为在线健康社区个性化推荐，增强在线咨询效果，完善在线健康社区系统，促进医患友好交流，减少医患纠纷做出一定贡献。

OR-020

疫情期间医学文献检索课在线授课效果调查与分析

兰雪、张晗、刘春鹤
中国医科大学

采用线上问卷的方式，从学习载体、课程形式、课程内容和自我评价 4 个方面调查了我校 230 名在线上学习《医学文献检索》课程学生的学习情况，评估线上教学效果，分析存在的主要问题及应对措施。

突如其来的新冠肺炎疫情影响了高校的正常开学和课堂授课，为保障师生的生命安全，保证疫情防控期间的教学进度和教学质量，教育部提出了“停课不停教、停课不停学”的要求。按照

教育部的统一部署，中国医科大学决定春季学期采用线上教学方式进行授课，充分利用各类在线课程平台积极开展在线教学活动，保证教学质量。为落实学校政策，适应线上教学模式，保证教学质量，我校《医学文献检索》课程积极进行调整和改革。为及时了解和总结线上教学的效果和存在的问题，保证教学效果，提升在线教学质量，课程教学团队对本学期有《医学文献检索》课程的学生进行了问卷调查，评估线上教学效果，分析存在的主要问题及应对措施。

1. 教学模式调整

本科学《医学文献检索》课程涉及 3 名授课教师，负责临床医学、基础医学、免疫检验和医学影像学 4 个专业的授课任务。线上教学并不是将课堂面对面授课模式照搬到线上，原因在于：一是线上教学无法很好地与学生及时进行沟通，了解学生对知识点的掌握情况；二是不能保证每名学生具有一个良好的、安静的学习场景；三是学生的网络和设备参差不齐；四是无法开展实验教学。面对这些问题，3 名授课教师快速展开线上教学准备，在学习线上教学平台使用的同时，进行线上教学设计。具体内容如下。

（1）线上教学平台：选择 Blackboard 和雨课堂平台进行线上教学。Blackboard 系统可上传课程介绍、授课 PPT、视频、课后作业等课程文件，方便学生下载和学习。使用雨课堂平台进行课程直播，可以在课上发送习题和弹幕互动，有助于及时掌握学生的学习情况，了解学生的课堂表现。在课程准备和测试中发现，偶尔会出现雨课堂无法登录的情况，而且雨课堂还不能支持多人语音和视频，无法进行小组汇报，因此又选取了腾讯会议作为授课平台。此外还建立了各个班级的 QQ 群和微信群，方便与学生随时进行沟通。在正式开课之前，与学生共同对以上平台进行了测试和学习。

（2）线上教学设计：将线上教学分成 3 个部分：课前、课上和课后，以减轻学生线上学习的压力和疲惫。①课前：通过雨课堂发布 PPT、视频等形式的预习材料，时间控制在 5 分钟之内能够完成。雨课堂平台可看到学生阅读预习材料的进展和时间，方便之后了解学生预习的完成情况。②课上：以 PPT 授课为主，辅以数据库实际检索的演示。为调动学生的参与度，授课过程中利用投票、选择题、弹幕等多种形式与学生进行互动，吸引学生的注意力。③课后：主要包括课后习题和补充学习材料 2 个部分，课后习题以数据库检索习题为主，学生完成后截图提交答案，时间控制在 10 分钟以内。

2.对象与方法

2.1 调查对象

以中国医科大学本科生中在本学期开展《医学文献检索》的大二、大三学生为研究对象，所学专业包括临床医学、基础医学、免疫检验和医学影像学，有效调查问卷230份。

表1 调查对象情况

项目		人数	百分比（%）
性别	男	80	34.78
	女	150	65.22
年级	大二	63	27.39
	大三	167	72.61
专业	基础医学	27	11.74
	临床药学	113	49.13
	免疫检验	29	12.61
	医学影像学	61	26.52
总计		230	

2.2 调查方法

该调查采用自制电子问卷，通过问卷星平台发放。从学习载体、课程形式、课程内容和自我评价四个方面编制调查问卷，共含有20个条目。

3.结果

3.1 学习载体使用情况

3.1.1 在线学习上网方式

使用的听课设备方面，使用电脑听课的学生最多（占 53.91%），其次为手机听课（36.65%）、平板听课（10%）。可以看到，通过移动设备（包括手机和平板）听课的学生人数占到45%左右，不可小觑。

因为文献检索课程需要学生在数据库中进行实际的检索操作，所以授课时建议学生最好利用电脑听课，从而能够在听课的同时，同步进行检索。比较采用不同听课设备听课，学生的听课效果差异，结果见图1。从图中可见，利用电脑和平板听课的学生认为听课效果较好的人数更多，而利用手机听课的学生中认为听课效果一般的人更多。平板因为可以实现分屏操作，在听课的同时，可以打开数据库网页，与老师一同进行操作。因此可以说利用平板听课与电脑听课的效果相似。

相对而言，利用手机听课存在手机界面小，无法分屏操作，而且多数数据库的移动端界面与电脑端不同。

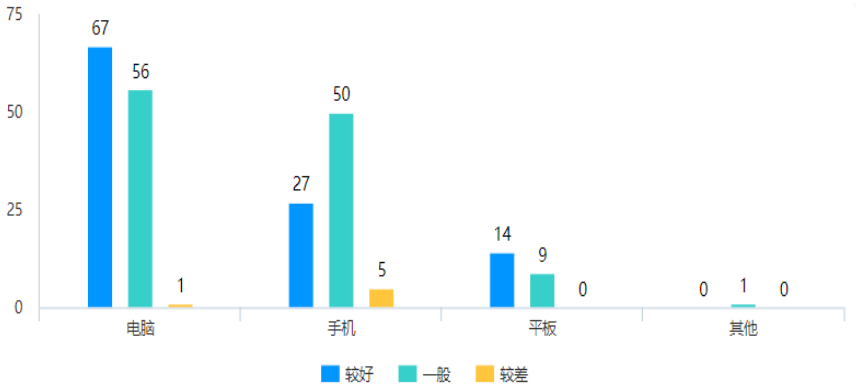


图 1 听课设备使用情况

3.1.2 网络情况

对学生线上学习的网络情况进行调查显示 38.26%的学生表示网络较好，53.91%的学生网络一般，7.83%的学生网络较差。虽然有 7.83%的学生认为网络较差，但是仅有 2.61%的学生表示网络不好影响了听课效果。大部分学生都认为目前所使用的网络能够达到一个比较好的听课效果（97.39%）。

3.1.3 教学平台

考虑到不同授课方式的特点，本门课程采取了多个授课平台，包括雨课堂、腾讯会议和腾讯课堂。学生在多个平台都进行了线上学习。调查结果显示喜欢雨课堂进行线上学习的学生人数最多，占 58.7%。喜欢腾讯会议和腾讯课堂的人数都不是很多，分别占 16.09%和 22.61%。

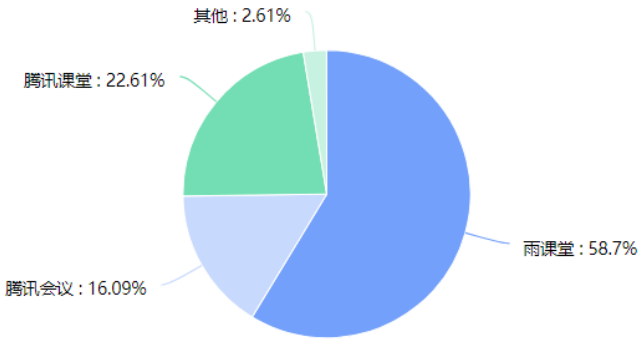


图 2 听课平台偏好情况

3.2 课程形式

3.2.1 课堂互动

为调动学生听课过程中的积极性，吸引学生注意力，教学中设计多种互动形式，包括提问、选择题、投票题和发送弹幕等。调查结果见图 3，喜欢发送弹幕进行互动的学生人数最多，占

51.74%，其次为选择题占 33.04%。对于传统面对面教学中最常见的点名提问形式则是学生最不喜欢的互动形式，喜欢人数仅占 0.87%。线上教学平台的使用丰富了传统教学的互动形式，并且更受学生欢迎。

在对互动次数多少适宜的调查中可以看到，以 45 分钟为标准，50.87%的学生认为 2 次更好，38.7%的学生认为 3 次适宜，但仍有 1.74%的学生认为可以互动 5 次以上。可见，每 45 分钟的课程与学生互动 2-3 次最好。

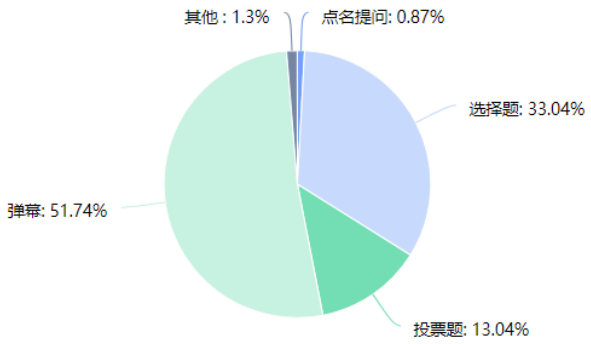


图 3 互动形式

3.2.2 授课形式

本门课程除了采取传统的 PPT 授课之外，还结合课程内容进行数据库检索演示授课，以及小组汇报的授课方式。调查结果见图 4，67.39%的学生更倾向于边演示边授课的形式。小组汇报授课形式并不是很受学生喜爱，与学生沟通之后了解到这是由于小组沟通受阻，无法很好进行交流。

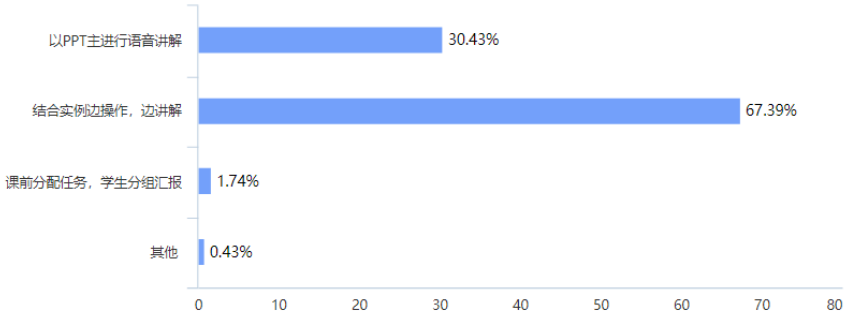


图 4 授课形式

3.2.3 补充材料

与课堂面对面授课不同，线上教学中学生受到的干扰更多，很容易被其他事物吸引。长时间的授课很难一直抓住学生的注意力，使学生保持较好的听课效果。因此本门课程采取了缩短授课时长，将一些非重点内容作为课前或课后材料发放给学生。调查显示，37.83%的学生认为课前、课中、课后的学习内容比例在 1：8：1 较好。希望增加课前、课后学习内容比例的人数相对较少。

在对课前课后学习材料方式的调查中发现，大多数学生希望补充材料是视频（60%、59.57%）。也有不少学生倾向于 PPT 形式，占 34%左右。对于不同的内容建议采取不同的形式呈现。学生指出视频更容易学习和掌握检索操作，但是 PPT 更易看知识点。

3.2.4 授课质量影响因素

对影响授课质量的可能因素进行调查发现，授课内容的呈现形式对质量的影响更大，而教师的语速、音量等并不是主要因素。排名前三的因素分别是：需结合实例进行授课、采取边演示边授课的形式、PPT 清晰。

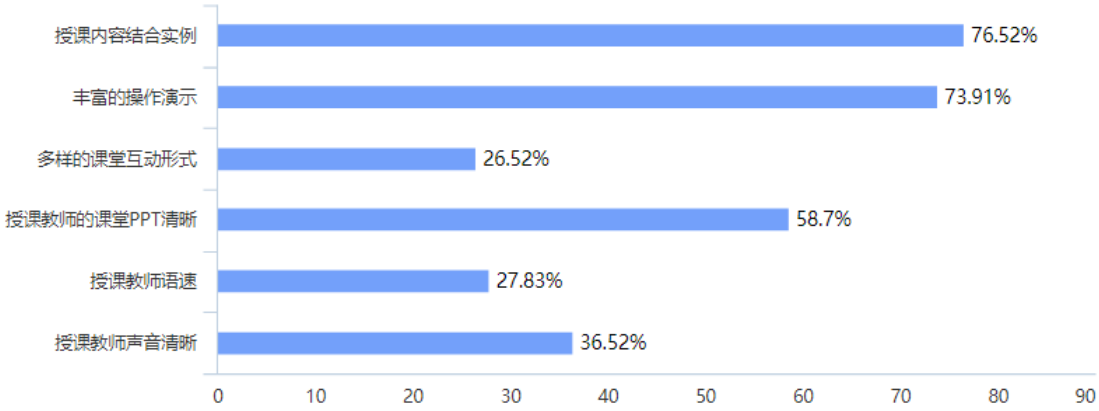


图 5 授课质量影响因素

3.3 课程内容

3.3.1 课程内容的难度、重要度和掌握度

在对课程内容的调查中发现（见图 6），学生认为相较于基础理论部分，数据库的检索更难，其中 PubMed 数据库最为困难（79.13%），其次为 Web of Science 引文检索系统（49.13%）。可以看到学生在对外文数据库的学习和掌握方面有一定难度。内容的难度影响了学生的掌握情况，仅有 18.26%的学生对 Web of Science 引文检索系统掌握良好。而 PubMed 数据库方面，虽然多数学生认为 PubMed 数据库很难，但是有 63.04%的学生表示能够很好地掌握。这可能是由于学时分布所造成的。PubMed 数据库分配了更多的学时，而 Web of Science 引文检索系统学时相对较少。

虽然学生认为中文全文数据库相对比较简单，但是有 96.52%的学生都认为中文数据库十分重要，并且掌握程度也较好。学生认为基本概念和理论知识在重要性和难度方面都较低，而且学生掌握程度也较好，可以考虑适当缩小学时。

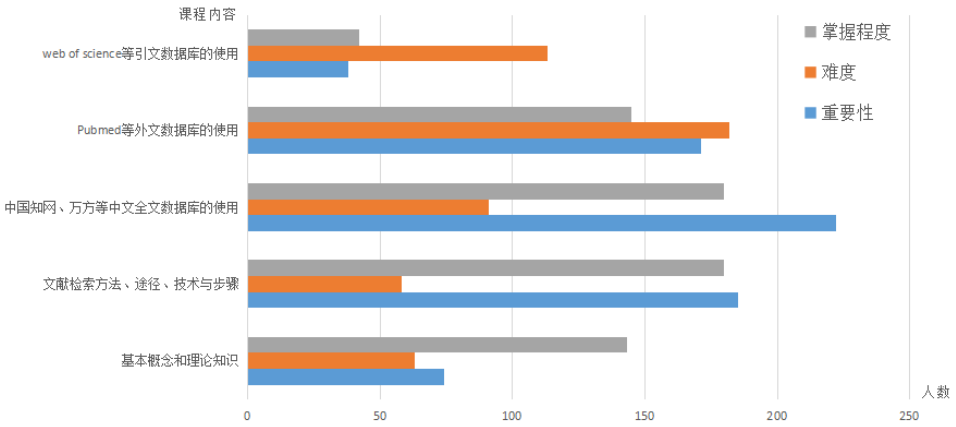


图 6 课程内容的难度、重要度和掌握度

3.3.2 课程学习的意义

通过调查发现，除了能够帮助学生掌握医学文献的检索方法和技巧（86.25%）以外，59.57%的学生认为该门课程有助于科研和学习，以及获悉科技论文引用的必要性（55.65%）。

3.4 自我评价

3.4.1 注意力评价

46.96%的学生认为自己能够在听课的大部分时间中保持较好的注意力，46.09%的学生认为自己的注意力一般，6.96%的学生表示不太能够集中注意力。46.52%的学生表示超过 30 分钟之后容易感觉疲惫和注意力不集中。仅有 13.91%表示听课 20 分钟后就感觉注意力不集中。可见，连续授课时间最好不要超过 30 分钟。

3.4.2 自学效果评价

68.26%的学生表示自学效果一般，仅有 25.22%的学生认为自己的自学效果较好。这可能也是为什么在前面调查时，大部分学生希望课前课后自学内容不要太多。

3.4.3 满意度评价

学生对课程整体的满意度达到 82.61%，其中满意占 30.87%，较满意占 51.74%。

4. 讨论与分析

医学文献检索课作为一门侧重实践操作的课程，适合线上教学，达到较好的教学效果。但经过调查发现线上教学仍然存在以下问题：

（1）学生听课所使用的网络和设备参差不齐，并且对听课效果有所影响。网络差可能会导致学生直播听课时内容不连贯、无法及时提交习题答案、与老师互动受限等问题。对此，应考虑学生因网络问题影响直播听课时，提供录课视频。虽然雨课堂平台设有直播回放功能，但是该功能也主要是在线观看，考虑为学生提供可下载的录课视频。此外，对于直播课时互动不及时等问题，

考虑课后增加与学生互动的途经和方式，如建立班级微信或QQ群，保证学生能在课后与授课教师进行沟通和交流，补交课上习题等。

在听课设备方面，大部分使用电脑和平板设备的学生认为听课效果好，而利用手机听课的效果则相对较差。并且利用手机听课的学生（36.65%）不在少数，因此需要考虑如何让使用手机听课的学生也能得到较好的听课效果和质量。手机端听课的问题在于手机屏幕小、无法分屏操作，在讲授数据库检索操作时，学生无法边听教师授课，边打开数据库网页共同进行操作。对于《医学文献检索这》样一门注重实践操作的课程来说，尤其是在数据库的教学中，学生能够跟着授课教师的步骤，共同进行检索操作是十分重要的。对此，考虑在课前向学生发放检索操作视频，让学生能够在上课时已经对数据库操作有一个基本的认识，课上如果没有跟上授课教师的操作，也可以在课后反复观看。

（2）授课时间不宜过长，课前课后布置内容不宜过多。在对学生的注意力调查中发现，大部分学生在30分钟之后开始感觉疲惫，无法集中注意力，因此连续授课时间最好控制在30分钟之内。直播授课时间的缩短，就意味着要将一部分非重点内容放在课前或课后，让学生自学。但是大部分的学生表示自学效果一般（68.26%），并且81.75%的学生希望教师授课内容和自学比例不要小于7：3。可见自学内容设置过多并不能得到好的学习效果。面向线上教学，在缩短线上授课时间的同时，不能一刀切地将剩余的大量内容作为课后学习材料发给学生，增加学生课后负担。可见，虽然要缩短授课时长，但是过多增加课前课后学习内容对于大多数学生来说是一个负担。此外，虽然大部分学生更倾向于视频形式的自学材料，但是选择PPT的学生也不占少数，有学生指出视频更容易学习和掌握检索操作，PPT更易看知识点。建议对于不同的授课内容采取不同的形式呈现。

（3）授课形式结合实例，以数据库检索演示为主。在传统课堂面对面授课时，利用数据库检索演示进行授课效果甚微。原因主要在于：①在教室授课时，许多坐在后排的学生在观看网页演示的时候经常无法看清网页，以及教师的鼠标位置；②教室授课时，受条件所限，大多数学生无法携带电脑上课，导致无法边听课边实践。而线上教学时，学生通过电脑或手机可以轻松看到教师演示时的网页界面及鼠标位置，甚至可以一边观看演示，一边同步进行检索操作。

（4）互动形式多样化，次数不宜过多。弹幕形式的互动更能体现学生的主观参与性，学生可以随时发送弹幕，不受时间、内容的限制。传统课堂上，学生即便有不理解的内容，但是多畏惧于当众提问而不敢提问。而弹幕互动则具有一定的匿名性，对于不了解的内容，学生可以直接发送弹幕提问，其他学生并不知道是谁提问的，能够一定程度上提高学生的积极性和参与度。

5.结论

通过对本学期在线上学习《医学文献检索》课程的230名学生进行问卷调查，发现《医学文献检索》的线上授课整体上取得较好的授课效果，但是仍然存在一些不足。未来需继续开展线上教学培训，提高教师的线上教学技能，重构教学内容，做好教学设计，才能实现线上教学与面对面课堂教学等同等的教学效果，并在疫情结束后顺利与线下教学有机结合。

OR-021

5G智慧移动医疗研究

刘思远、胡德华
中南大学

目的 探索5G技术在智慧移动医疗健康领域的应用场景及未来发展方向。

方法 对5G技术的特点及在不同移动医疗场景的适用性进行分析，并5G在不同移动医疗应用场景进行分类总结。

结果 5G技术的三大应用场景与移动医疗发展不谋而合，5G技术在院内数字化、移动健康及智能医疗等方面发挥优势。

结论 5G技术的发展给医疗健康领域带来了新的机遇，在大力推动5G智慧移动医疗发展应用的同时，需加强产学研结合，协同创新，充分发挥5G智慧移动医疗优势，提升医疗服务能力。

1.5G医疗健康概述

随着5G正式商用的到来以及与大数据、人工智能等前沿技术的充分融合，基于5G技术的医疗健康越来越呈现出强大的生命力和影响力，对深入推进深化医药卫生体制改革、加快健康中国建设以及推动医疗健康产业发展起到了重要的支撑作用。

5G医疗健康是指以第五代移动通信技术为依托，灵活利用有限的医疗人力和设备资源，同时借助大医院及专家先进的医疗技术及手段，在疾病诊断、监护和治疗等方面提供信息化、移动化和远程化的诊疗服务，创新智慧医疗业务应用，节省医院运营成本，促进医疗资源共享下沉，提升医疗效率和诊断水平，缓解患者看病难的问题，协助推进偏远地区的精准扶贫。

从 1G 到 4G 通信时代，通信所提供的服务尽量从各个维度满足人们的数字化消费需求，而对物体的连接缺乏总体规划和思考，而伴随着可穿戴设备、VR 、 AR 等新兴技术和应用的普及以及对封闭式场景的数字化变革（工业 4.0，智慧园区，远程医疗等）对网络提出了新的需求。同时， “互联网+”国家战略中明确指出：未来电信基础设施和信息服务要在国民经济中下沉，满足医疗、教育、生活服务、公共服务等垂直行业的信息化需求，改变传统行业，促进跨界创新。因此，5G 网络不仅带来用户体验速率的提升、时延的减少、移动性的增强等，同时还将满足各类垂直行业多样化的业务需求[1]。

当前，我国 5G 医疗健康的发展尚处于起步阶段，在顶层架构、系统设计和落地应用上还需要不断完善，但是 5G 医疗健康前期探索已取得良好的示范作用，实现了 5G 技术在医疗健康领域，包括远程会诊、远程手术、应急救援、远程监护、智慧导诊、移动医护、智慧院区管理、AI 辅助诊断等众多场景的广泛应用。

5G 的高速率、低延迟和万物互联等优势，为整个健康产业在可穿戴医疗设备、远程会诊等领域的应用发展提供了基础，为医疗健康服务行业的全方位智能化提供了可能。在急救方面，5G 帮助建立更快更稳定的急救通讯系统，保障急救工作各方面密切配合，利用 5G 传输超高清视频和智能医疗设备数据，协助在院医生提前掌握急救车上病人的病情[2]。在医院智能化方面，可以利用 5G 技术对手术视频和医疗视频等进行远程直播，结合 AR，帮助基层医生实现手术环节的异地实习。传输的高清视频与力量感知设备结合，为医生提供更真实的病况，为病人提供高质量远程会诊。查房机器人的出现帮助实现院内数字智能化，实现医生远程查房及实时监测，减轻医生的负担[3]。在疾病预防方面，5G 技术支持实时传输海量人体健康数据，协助医疗服务机构等对用户实现不间断身体数据监测[4]。可以更好地支持连续监测和感官处理装置，不断收集大量患者的实时数据，人工智能利用这些数据对大众健康进行全面且连续的记录与分析，进一步推荐更为合适的医疗保健方案。

2.5G 智慧移动医疗发展及应用场景

5G 具备高速率、低时延、大连接三大特性，分别对应三大应用场景：增强移动宽带（enhanced Mobile Broadband，eMBB）、高可靠低时延（Ultra-Reliable and Low Latency Communications，uRLLC）和低功耗大连接（massive Machine Type Communications，mMTC）三大场景[5]。这与当前移动医疗业务的发展应用不谋而合。

eMBB 场景应用主要有 5G 急救车，给急救车提供广域连续覆盖，实现患者“上车即入院”的愿景，通过 5G 网络高清视频回传现场情况，同时将病患体征以及病情等大量生命信息实时回传到后台指挥中心；还可以完成不同类型用户的可穿戴设备数据收集，实现对用户体征数据 24 小时的实

时监测[6]。uRLLC 场景主要应用于院内无线监护、远程检测及远程手术等低时延应用场景。其中无线监护通过统一采集大量病患者的生命体征信息，并在后台进行统一的监控管理，大大提升了现有的 ICU 病房的医护人员的效率。远程 B 超，远程手术等对于检测技术有较高要求，需要实时反馈，5G 技术消除现有远程检测及手术的医患之间的物理距离，实现千里之外的实时检测及手术[7]。mMTC 场景主要集中在院内，医院有上千种医疗器械设备，对于医疗设备的管理监控有迫切需求，未来通过 5G 的统一接入方式，可实现医疗器械的统一管理，并实现与其他设备及系统之间的数据联网及联动预警[8]。

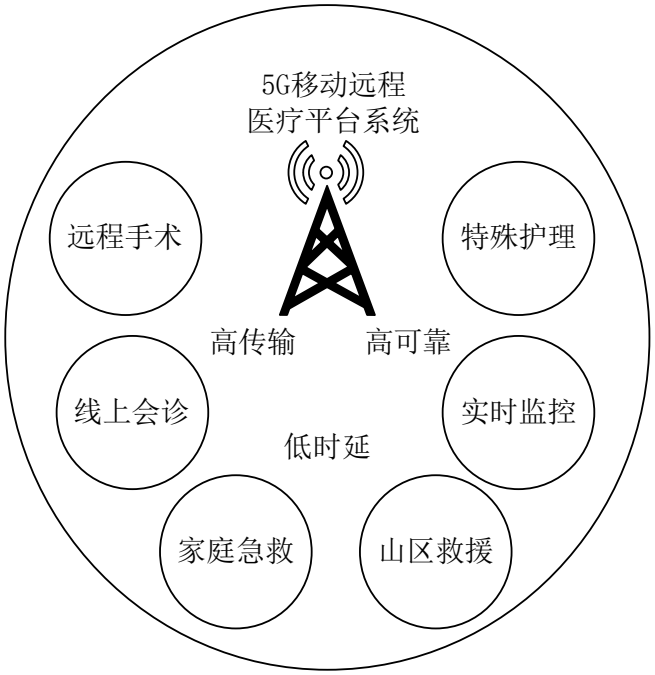


图 1 基于 5G 技术的医疗健康服务应用场景

基于 5G 技术的移动医疗发展可以解决居民看病难、医疗资源分配不均等问题。自 2009 年开始，国务院先后发布《关于深化医药卫生体制改革的意见》、《关于推进分级诊疗制度建设的指导意见》、《国务院办公厅关于推进医疗联合体建设和发展的指导意见》等系列政策文件，要求实现按照疾病的轻重缓急及治疗的难易程度进行分级，不同级别的医疗机构承担不同疾病的治疗，逐步实现从全科到专业化的医疗过程；真正实现包括电子病历、电子健康档案在内的医院间、区域间的互联互通。2018 年 4 月，国家卫健委发布《全国医院信息化建 设标准与规范（试行）》，着眼未来 5-10 年全国医院信息化应用发展要求，从软硬件建设、安全保障、新兴技术应用等方面，进一步具体规范了医院信息化建设的主要内容和要求。

国内一些大型医疗机构的智慧移动医疗服务平台现已初具规模。以华西医院、华西附二院为代表的龙头医疗机构，针对 5G 远程医疗、互联网医疗、应急救援、医疗监管、健康管理、VR 病房

探视等方面展开 5G 智慧医疗探索与应用创新研究，一方面提升医疗资源供给，实现患者和医院的信息连接，最大程度提高医疗资源效率，便利就医流程；另一方面深入挖掘医疗健康数据价值，催生新的移动及智慧医疗应用服务。

3.5G 智慧移动医疗应用场景

3.1 院内设备数字化转型，真正实现万物互联。

5G 技术融合大数据、人工智能技术及设备，可以打破业务流程之间的传统壁垒，医疗物联网、医疗云、医疗大数据应用等技术在 5G 技术的基础上，可突破医院各科室间在传统医疗模式下信息孤立的局限性，实现各部门之间有效的协调互补，提高医务人员的工作效率。

在智慧移动医疗生态下，医院将实现全方位感知患者，通过相关设备、系统和流程，实现患者信息实时测量、捕获、传递和分析。具体包括，（1）影像、病案等信息的全方位自动采集，物联化；（2）诊疗信息、监测数据等及时有效的传输互联；（3）结合人工智能技术实现智能决策支持[9]。与此同时，借助云计算、物联网、5G 技术等新兴技术构建新的 ICT 整合平台和解决方案，将实现诊疗检测护理及付费用药的全流程数字化转型。不同信息的开放和联接，不仅提升医患沟通效率和质量，更重要的是在医护之间、科室之间、医疗机构内及不同机构之间，如管理部门、保险单位、康复机构等构建联接，实现多主体业务协同。从而帮助提升医疗服务效率，提升群众就医满意度，帮助医院管理、政策制定获得更准确、全面的信息，切实提高医疗保健服务质量。

3.2 5G 移动健康设备助力健康管理，强化移动医疗。

当前我国医疗资源短缺且城乡分布失衡。仅凭医护人员对病患进行日常监测和随访，并进行健康管理档案管理维护，将严重影响日常工作。而借助于以智能穿戴设备为代表的移动健康设备，融合大数据、云存储、人工智能等技术进行健康管理的移动医疗已成为大势所趋。5G 技术的发展及深度融合，将大幅提升移动医疗的服务能力，解决医疗资源供需不平衡等诸多问题。5G 可承载大规模、高密度的物联网业务，并结合云技术实现云端智能分析。在此基础上，用户身体健康数据将通过移动健康设备进行监测记录，上传至云端大数据库分析，可及时对每一个用户的健康状况进行更为全面且准确的判断，同时医生可充分利用健康监测数据指标向患者推荐适合的疗护及术后康复方案。

5G 带来了通讯层面的全面升级，为移动医疗方案的落地提供了技术基础。但同时也意味着，与之配套的软硬件设备也要进行一系列升级，以充分发挥 5G 的技术优势。5G 时代大量的数据处理以及医疗行业对于数据准确性、稳定性、高效性等需求，对于移动医疗配套软硬件均提出了更高的要求。在移动医疗领域，要掌握及时有效的数据并提出最佳决策方案，既需要人工智能、云存储、大数据提供有效数据的挖掘和分析，又需要底端基础硬件及软件功能的支持。这是一个系统

工程，任何一环的缺失，都会对最终结果产生影响。因此，5G 时代的移动医疗，不仅需要通讯技术进行升级，相关配套产业也必须进行全面升级，才能充分融合 5G 和移动健康技术带来的优势。

3.3 5G 结合人工智能，强化基础，催生创新。

人工智能在移动健康领域的应用场景已遍地开花，包括基于用户数据的个性化诊疗方案、健康管理和疾病风险预测等，人工智能技术已逐渐成为提升医疗服务质量和效率的突出动力。而 5G 的高带宽、低时延和大连接的优势可以助力人工智能技术提供更快的响应速度、更为丰富多样的内容、更加智能个性化的应用模式以及更为直观的用户体验。5G 可以增强人工智能的处理性能，基于 5G 超大连接和网络切片的垂直行业应用，可以有效解决即时数据传输的效率问题；5G 可以降低人工智能获取数据延迟，5G 网络可以实现无线高速的低延迟联通，将有效解决运算中由于时间敏感而产生的数据问题；基于边缘云技术提升计算效率，可将业务下沉至网络边缘，以去中心化的工作方式大幅提升人工智能运算效率和可靠性。

5G 与人工智能结合赋能移动医疗，催生的医疗应用场景有个性化诊疗方案、健康管理和疾病风险预测等。在个性化诊疗方案方面，5G 与边缘云提升人工智能的运力，更加精准快速处理海量疾病诊疗数据，帮助医生基于诊疗及监测数据制定调整治疗及用药方案。在健康管理方面，5G 网络可实时传输用户健康体征数据，结合人工智能技术对数据进行智能分析，发生异常及时预警[10]。在疾病风险预测方面，基因组等数据借助 5G 快速上传至数据处理中心，借助以 AI 为基础的基因分析技术，可识别单核苷酸多态性变异与删除、拷贝变异数以及对基因的变异进行严重性评估[11]。5G 技术不仅可以大幅度提升网速，更将补齐制约人工智能发展的短板，成为驱动人工智能的新动力。每个人工智能应用都需要专属网络，根据应用需求进行实时动态调整，满足快速变化的业务需求。而通过 5G 核心网构建逻辑隔离的网络切片，具有提供网络功能和资源按需部署的能力，可满足未来医疗行业多样化的业务需求，为每个人工智能医疗应用打造个性化定制网络。

由此可见，5G 网络结合人工智能技术对于催生智慧移动医疗的创新应用具有重要的推动作用。5G 作为基础设施，如同“信息高速公路”一样，为庞大数据和信息的实时动态传递提出解决路径[12]。而人工智能，则像是高速公路的交通枢纽，基于自主学习和演化的神经网络，可与 5G 技术优势互补，赋予机器人类的智慧，同时使万物互联变成可能。二者相结合，会为整个移动健康乃至整个医疗领域的革新和行业的发展带来前所未有的提升。

4.小结

5G 智慧移动医疗作为 5G 技术与医疗健康融合的代表性发展方向，给人口老龄化与慢性病管控等问题提出了新的解决方案。同时，5G 与大数据、互联网+、人工智能、区块链等前沿技术的整合应用，对推进深化医药卫生体制改革、加快“健康中国”建设和推动医疗健康产业发展，起到了

重要支撑作用。5G 技术的飞速发展也为复合型人才培养提出了更高的要求，这无疑为医学信息学的发展提供了新的机遇和挑战。当前 5G 技术在医疗健康领域的研究及应用仍处于起步探索阶段，且缺乏统一的标准与评价方法体系，在实际应用的稳定性及安全性问题亟需得到解决。在大力推动 5G 智慧移动医疗发展应用的同时，应充分加强产学研结合，推动协同创新建设，推进医疗健康与信息化融合发展，以进一步发挥 5G 医疗健康优势，实现真正意义上的以人民健康为中心的智慧医疗发展。

OR-022

新冠肺炎疫情下高校图书馆面向学生的学习科研信息服务研究

赵嘉玮、林鑫、郭进京、蔡妙芝、任慧玲
中国医学科学院医学信息研究所-图书馆

由于新型冠状病毒疫情防控需要，高校师生面临居家防疫带来的远程教学科研挑战，高校学生的线上学习科研现状值得关注。高校图书馆作为高校的信息资源支撑机构，为师生居家研学带来了重要的信息资源及服务保障。基于此，本文对高校学生在疫情期间的线上学习科研现状及高校图书馆所提供的资源与服务进行了调研，并依据调研结果对图书馆参与高校远程教学科研过程中获得的启示进行探讨，以期为高校图书馆完善教育和信息服务职能、处理未来突发公共卫生事件提供借鉴。

引言

自2019年12月始的新型冠状病毒（COVID-19）疫情（以下简称“新冠疫情”）暴发后，国内积极采取相关防疫措施，通过封城隔离、开展社区防控工作等多方面深层次的联防联控措施，有效防止了新冠疫情的大规模扩散，目前疫情防控工作已取得了明显成效，防控形式趋于稳定。在此期间，各高校面临着新冠疫情带来的无法正常开学和在校教学科研带来的诸多挑战。教育部针对此问题提出高校在线教学组织与管理的指导意见，要求高校充分利用慕课、在线课程教育资源等网络学习资源，依托在线课程平台、网络学习空间等，开展线上教学活动，同时保证教学的进度与质量，以实现“停课不停教、停课不停学”[1]。高校学生的学习科研活动由线下转至线上，带来了前所未有的大规模网络教学科研实践。高校图书馆作为重要的文献信息资源支撑机构，服

务于高校的教学科研，在疫情期间保障了电子文献资源的正常获取使用，并为师生的居家研学提供充分的信息服务与技术支持。

基于此，本文通过问卷调查的方法了解当前高校学生的网络课程学习及科研资源获取情况，并对处在疫情中心的武汉大学图书馆及其他5所国内知名综合及医科类高校图书馆在疫情期间提供的信息资源及服务进行调研分析，同时对高校图书馆在网络课程教育及在科研中的作用及服务经验进行了探讨，以期为高校图书馆提高资源保障能力、更好地辅助网络教学科研提供借鉴。

1 新冠疫情期间高校学生的学习科研调研

为了更直观地了解高校学生在疫情期间学习网络课程的相关认识及疫情对其科研进程的影响，笔者于 2020 年 4 月利用问卷星平台对在读本科生、硕士研究生及博士研究生展开调查，共收回问卷 190 份，去除无效问卷后，有效问卷 182 份，有效率为 95.7%。在这些利用互联网进行学习科研的调查对象中，需要上网络课程（以下略为“网课”）、修课程学分的人群占总人数的 83.5%；并有 70.33%的被调查者表达了自己的科研及获取研究资源相关需求。以下将从高校学生网课学习及科研两个角度分别了解学生的居家研学现状。

1.1 高校学生网络课程学习情况

据调查了解，目前高校学生开展网课学习主要通过以下三种途径：利用网络直播互联平台及软件即时上网课、自行学习经教师录制并上传至视频平台的课程及使用开放课程平台的资源进行自主学习。基于大中小学“线上开学”后高涨的网课教学需求及新冠疫情防控工作的需要，部分原有的远程办公平台根据网络课程教学的特点，发挥技术优势，适时调整平台相应功能模块。如钉钉[2]，开放了 302 人视频会议、免费在线课堂、及支持直播互动的“群麦克”功能，提供“家校本”、“云课堂”等多样化的功能协助教师收集评改学生作业、发布在线考试题库以掌握、考核学生在线学习情况。受年轻群体青睐的新兴视频网站“bilibili 哔哩哔哩”也被高校教师用作发布教学视频的辅助平台，上传录制好的视频以供学生观看学习。除此之外，“中国大学 MOOC”等慕课平台的“国家精品在线开放课程”发挥示范引领作用，其课程负责团队提供教学服务带动高校教师团队录制高质量的在线视频课程，为广大学子供给更多的精品课程[3]。

高校教师结合教学实际、网络环境条件网络平台优势择优选取平台，组织在线教学。调查网课学习者所用软件或平台时发现存在灵活搭配选择使用的现象，其中使用比例位居前三的软件是“腾讯会议”、“钉钉”与“超星学习通”，分别占比 48.68%、46.05%与 32.89%。有 26.32%的调查对象运用“中国大学 MOOC”等 MOOC 平台学习网课，仅有 14%左右的调查对象通过“雨课堂”和“bilibili 哔哩哔哩”平台完成网课学习。在调查的网课学习者中，仅有 3.95%的调查对象认为网课软件及平台有些难用甚至难用，而大部分（59.89%）调查对象认为其比较好用。调查对象最

为推荐的网课软件是“钉钉”，是 27.63%调查对象的首选。对于学生判定网课软件或平台较好用的原因，有 59.21%的调查对象表示软件操作页面简单易上手是选择平台时首要考虑的因素，44.74%的调查对象较在意平台的回放支持或倍速观看功能，34.21%的调查对象认为互动功能相对重要，如支持“举手”向老师提问或弹幕功能。对于使用感受的影响因素，47.37%的调查对象认为软件的卡顿、音画不同步是带来负面观感的主要因素。

调查发现，关于“网课方式与传统课堂的学习效果比较”问题，接近一半的（43.42%）调查对象认为学习网课获取知识的效果和课堂学习时相差不大，而 36.84%的调查对象表示其教学效果不及传统课堂，仅有 19.74%的调查对象认为前者要优于后者。有 55.26%的调查对象认为网课的方便性及阻断疫情传播的贡献作用是在此次疫情期间最为突出的优势；有 53.95%的调查对象青睐于网课支持倍速观看的录播功能。反之，多数人认为以下几点是居家网课环境下的不足之处：“无法面对面交流，缺乏参与感”、“长时间面对屏幕，有损视力”、“客观环境（如网络延迟、设备限制等）影响听课质量”、“条件限制，无法获取教参和图书馆纸质资源”；少数人认为 “网络学习时发现有随意传播课件视频等资料，从而侵犯他人版权的现象”是其间接的不足。

1.2 高校学生科研文献信息获取和利用情况

笔者调查了疫情对高校学生科研方面的影响，有 81.63%的被调查者表示疫情下科研受到了一定程度的影响，对于“此次疫情对个人科研的影响”问题，57.5%的调查对象认为身处校外不方便查询中英文文献库；37.5%的调查对象认为受限于疫情，无法进入实验室，科研进程止步不前；32.5%的调查对象表示无法查阅参考图书馆书籍；还有 20%的调查对象表示和导师的联系交流受阻。在科研全过程无法顺利进行的情况下，高校学生利用有限的条件来尽可能地维持科研进度和状态。在校期间，多数学生选择在高校图书馆查阅馆藏书籍、利用图书馆提供的电子文献资源来扩展补充知识，疫情下，科研资料的获取渠道集中于线上文献库及平台等。

对于调查对象获取图书馆订阅的文献库的渠道，有 64.06%的调查对象使用了校内 VPN 等远程获取途径，有 28.13%的调查对象使用限时开放的公共文献平台（如知网 CNKI、SinoMed、专利之星数据库等），有 26.56%的调查对象选择了学校推荐使用的公共账号密码，还有 15.63%的调查对象表示自己拥有账号或借用他人的账号。关于高校提供的数据库或账号的可用性问题，一半以上（53.13%）的调查对象表示大部分数据库都可以正常使用并获取文摘和全文文献等；42.19%的调查对象表示存在偶尔登录不上、链接不稳定、无法正常获取文献的问题；仅有 4.69%的调查对象认为数据库或账号的可用性较差。对于图书馆提供疫情下文献获取保障措施的及时程度，40.63%的调查对象表示图书馆资源提供及时高效，37.5%的调查对象表示其比较及时，认为其一般或不太及时的占少数。

笔者调查了学生了解其所在高校的图书馆在疫情期间的服务情况，具体服务内容及提供比例见图 1。学生从图书馆微信公众号、微博及官网等处获取图书馆宣传的服务信息，从图中可见高校学生对图书馆所提供服务存在不同程度的认识，学术资源远程访问、数据库及平台使用指南及咨询服务是多为高校学生了解的图书馆服务，也是能较好反映学生远程科研需求的服务形式。

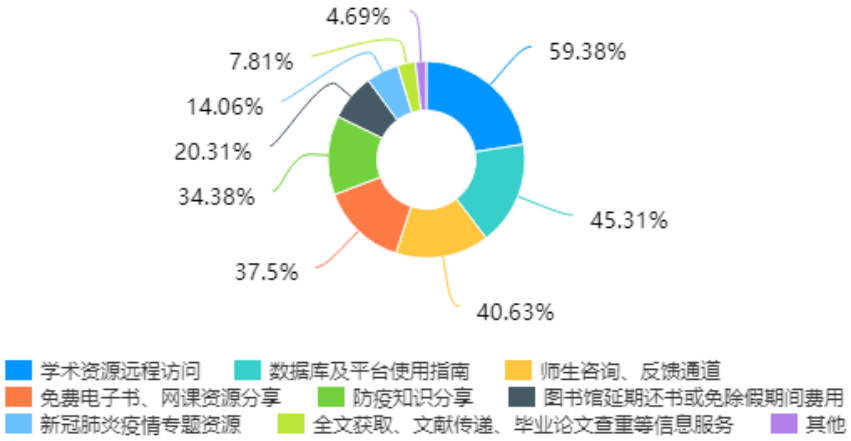


图 1 疫情期间高校图书馆面向学生提供的服务内容

关于疫情期间有科研需求的学生对图书馆服务的使用情况等，有三分之二的调查对象表示使用了远程图书馆学术资源；有 34.38%的调查对象参考过图书馆发布的课程或数据库使用指南；有 28.13%的调查对象获取了学校电子教参；有 21.88%的调查对象访问过图书馆整合的新冠肺炎专题库；有 20.31%的调查对象使用了图书馆借书延期服务；还有 14.06%的调查对象向图书馆员寻求了信息咨询帮助，如问题解答、全文获取等。其中，多数调查对象比较满意图书馆提供的各项服务，认为对自己的科研学习有所帮助，也可以进一步完善服务内容和质量。在学生使用图书馆提供的资源时是否能准确高效地获取所需文献问题上，接近九成以上的调查对象认为在科研中有时不能高效获取文献，检索方式较单一，需要掌握更专业的文献检索及获取技能，提高研究效率；仅有 9.38%的调查对象表示拥有足够的文献检索技巧，可以轻松准确地查找所需文献。至于学生期望获得的图书馆服务等，多数调查对象表示希望图书馆提供以下方面的服务：“当前能提供文献检索方面的知识，提高科研效率”“更加高效快速的全文获取、学术资源问题查找”和“更多高质量的资源推介”；结合当下网络课程发展趋势，部分学生认为图书馆“可以和考研机构、付费网课平台建立合作，提供学生所需的各类网络课程资源”；还有部分学生认为图书馆可以在微信平台上发布“有奖知识问答或竞猜活动”，增加学生与图书馆之间的互动。

分析上述高校学生在疫情期间学习科研现状的问卷调查结果，受如课本缺失、网课平台使用不便、课程学习交互不好等诸多因素影响，部分学生网课学习效果不甚理想，有学生还面临侵犯

他人知识版权的风险。疫情限制了高校学生的科研过程，在校实验暂止，大多数的学术资源需要藉由网络获取。经调查的学生中，部分学生不了解也较少使用高校图书馆面向学生科研提供的线上服务，大部分学生需要提高文献检索及获取技能以提高科研效率。

2 疫情期间高校图书馆的信息服务

图书馆作为高校的文献信息资源中心，拥有种类众多、内容丰富的纸本及数字文献资源，是以促进高校人才教育、服务科学研究为目的的学术性机构，其相应的职能为教育职能和信息服务职能[4]。由高校图书馆建立的文献信息资源保障体系，依据高校性质有所侧重地收录学科信息资源，为高校师生的教学、科研提供了有力的文献资源支持，在应对此次新冠疫情突发公共卫生事件中起到了重要作用。笔者选取武汉大学、清华大学、北京大学、上海交通大学、中国医学科学院北京协和医学院和中国医科大学 6 所具有代表性的综合类及医学类高校图书馆，对疫情期间图书馆提供的资源和服务进行了调研。

表 1 高校图书馆提供的资源及服务

图书馆名称	图书馆提供资源	疫情期间图书馆推出服务	图书馆远程资源访问方式
武汉大学图书馆	各类数据库、电子书籍及电子教参资源、网络课程、新冠相关资源及平台	1.学术资源远程访问服务；2.收集、推介疫时免费及限时试用数字资源；3.按专题推荐图书馆各类资源；4.为师生搜集推荐免费电子教参及提供教学支持；5.外借图书不计算逾期；6.线性肺炎防控专题网站、在图书馆移动 APP 上线抗疫专栏；7.向方舱医院捐赠近万册刊物助力抗疫；8.宣传开展信息素养教育网课；9.策划发起“凭君传语报平安”“加油武汉，加油小布”征集活动；10.信息咨询服务：文献传递、查收查引、读者问题咨询答疑等	WebVPN 方式，校园 VPN 客户端，CARSI Shibboleth 身份认证服务方式

清华大学图书馆	数据库、电子书资源 、新冠肺炎资源、电子教参平台	1.学术资源远程访问服务；2.收集、推介疫时免费及限时试用数据资源；3.推荐图书馆收录资源；4.实行“疫情期间借书规则”：外借图书不计逾期，无接触预约借书；5.开展在线直播讲座、研讨会；6.信息咨询服务：代检代查、科技查新、情报分析、BALIS 原文传递、接收博士后出站报告、学生毕业论文审核、读者问题咨询解答等	校外访问控制系统，CARSI Shibboleth 服务，远程漫游，SSL VPN 方式，MyLOFT 远程访问工具
北京大学图书馆	各类数据库、电子书籍及电子教参资源 、新冠肺炎相关资源	1.学术资源远程访问服务；2.中国高等教育文献保障系统 CALIS、中国高校人文科学文献中心 CASHL 正常使用；3.外借图书不计算逾期；4.送书到楼服务 5.为师生提供电子教参服务平台 6.宣传开展在线直播讲座；7.信息咨询服务：查收查引服务、知识产权信息、情报研究和分析、信息素质教育、读者问题咨询解答等	SSL VPN 方式，CARSI Shibboleth 身份认证方式，WAYless 访问
上海交通大学图书馆	各类数据库、电子书资源、交大文库在线导览、网络课程、新冠相关资源	1.学术资源远程访问服务；2.收集、推介疫时免费及限时试用数字资源 3.按照学科分类推送图书馆电子资源及工具；4.外借图书不计算逾期；5.为师生提供电子教参；6.宣传开展在线滚动培训讲座、”思源微课“”专利学堂“”等微课程；7.预约借阅纸质图书服务；8.举办线上有奖竞答、打卡活动；9.信息咨询服务：线上馆际互借、文献传递、科技查新、查收查引、情报计量分析、	VPN 方式，CARSI Shibboleth 身份认证服务方式，校园网代理，短期校外漫游

专利信息服务等

中国医学科学院图书馆	各类数据库、电子书资源、网络课程、新冠肺炎相关资源及平台	1.免费开放 SinoMed；2.发布新冠肺炎动态监测平台，新冠肺炎文献专题库，新冠防控科技文献监测快报；3 学术资源远程登陆访问；4.收集、推介疫时免费及限期试用数字资源；5.开闭馆时间，外借图书不计算逾期；6.宣传推荐在线直播讲座；6.信息咨询服务：查新查引、全文服务、信息咨询用户服务微信区、服务费缴纳等	医信通，CARSI Shibboleth 身份认证服务方式，WebVPN 方式
中国医科大学图书馆	各类数据库、电子书资源、网络课程、 新冠肺炎相关资源	1.学术资源远程登陆访问；2.收集、推介疫时免费数字资源 3.提供各类数据库指南；3.开闭馆时期间内，外借图书不计算逾期；4.宣传开展在线直播、讲座 5.举办线上有奖竞答活动 6.信息咨询服务：信息检索、查新服务、查重服务、读者问题咨询解答等	图书馆 VPN 方式，一卡通登录

从表 1 中可以看出，疫情期间高校图书馆的资源服务主要有以下特点：

（1）保障了学术资源的远程获取

以上调研的图书馆，在疫情期间均做出迅速反应，联合技术部门及数据库商，为高校师生提供了如校园 VPN，身份统一认证等多种方便快捷的途径获取校内原有的数据库、文献资源平台上的各类学术资源，利用先进的移动互联网技术手段来保障资源获取渠道的畅通。

（2）结合网络平台提供多种服务获取途径

目前，官方认证的高校图书馆活跃在微信公众号、微博等网络平台，成为师生快捷高效地获取最新图书馆动态的窗口。图书馆通过多种网络途径提供资源推介和指引，同时向师生提供便捷，精准的全方位参考咨询服务，答疑解惑，辅助教科研。

（3）整合多元资源，服务形式多样化

高校图书馆与数据库商及各类资源平台有着良好的合作基础，整合提供了丰富多元的电子资源[5]。除此之外，图书馆对新冠肺炎相关信息资源进行去伪存真，汇集权威的疫情信息，帮助师生读者做出理性参考及判断。服务表现形式多样化，尤其是医科类院校发挥其医学专业人员知识背景及资源优势，发布的新冠专题库及动态监测平台等可为新冠肺炎科研攻关提供宝贵的专业资源参考。

（4）调整服务政策及内容，辅助师生科研学习和网络教学

高校图书馆及时调整服务政策和内容，解决师生关心的图书逾期问题；制作详细的数据库或网络平台使用指南；开设各类知识分享在线讲座及研讨会；部分高校图书馆制定教参供给方案，为高校师生提供不同程度的教参支持。这些措施有效支持了师生科研学习和网络教学。

3 线上研学实践下高校图书馆服务的启示

高校图书馆提供的各项服务有效降低疫情对教学科研的不利影响，为师生适应线上研学带来平缓过渡。高校图书馆应结合高校学生学习科研需要，总结疫情下信息资源提供及服务经验，发挥本馆优势，在以下方面发挥积极作用。

3.1 开展在线信息素养教育，提高学生科研效率

在疫情期间，尽管部分高校图书馆[6, 7]积极开展的图书馆主题滚动培训、线上直播信息素养培训等作用显著，通过本文问卷调查显示出学生信息素养不足的现象依然存在。目前主要以公共选修课的形式由高校图书馆牵头开展高校信息素养教育[8]，但并非所有高校都具备充分的开课条件。对此，一些信息素养教育稍欠缺的可以和成果丰硕的高校图书馆加强沟通与合作，通过协同开展远程网络课程及讲座的方式推进高校信息素养教育资源的共建共享。高校学科馆员具有对口学科院系的专业背景优势，可以结合各院所不同学科领域的资源特点调整信息素养教学方式内容，以嵌入学科教学的信息素养主题课程、讲座、网络微课程等多种手段适应学生多样化的科研需求。图书馆通过提供形式多样的信息素养教育培养学生信息意识，提高学生文献利用能力，使其具备相应的信息素养，从而助力科研过程。

3.2 协助师生应对网课中的版权挑战

在线开放的网络课程环境下， 教师常利用如图片、视频和文献等第三方信息资源丰富课堂内容，我国《著作权法》第二十二条中包含“合理使用”第三方作品的界定：“为学校课堂教学或者科学研究，翻译或者少量复制已经发表的作品，供教学或者科研人员使用，但不得出版发行”[9]。使用环境专指课堂面授教学，而网络课程则属于此范围外，课堂教学面临着知识版权侵权的风险与挑战[10]。同时，对于教学中使用的课件作品，若学生未经许可将其任意传播，无疑是侵犯了教师的著作权。且由于“合理使用”的主体严格限定为“教学或科研人员”，若课件制作使用了

第三方成果，教师允许学生下载课件则会侵犯他人的复制权；若课件被学生随意上传至网络，则可能同时侵犯了他人的信息网络传播权[11]。

面向学生主体，高校图书馆应汇集知识产权相关知识，通过在图书馆主页设置知识产权专题栏、在微信公众号及微博等网络平台推送有关文章，从而增强学生版权意识，避免侵犯著作权人的合法权益。针对线上教学可能出现的版权问题面向教师开展咨询指导服务，以避免课程中的版权纠纷，包括通过在线讲座，研讨会等形式向教师普及相关版权知识及政策；收集、筛选和推荐各类公共领域及开放获取资源以辅助课程教学；提供“合理使用”判定，并协助教师获取所需的第三方资料的版权授权许可等[12]。

3.3 完善突发公共卫生事件应对机制，提高信息资源及服务保障能力

此次新冠疫情影响深远，高校图书馆应补足疫情期间暴露出的信息服务短板，及时收集师生对服务的反馈信息，进一步完善现有的各项资源保障服务措施。通过建立健全常态化的突发公共卫生事件应急机制，完善应急预案，提高管理意识、信息资源及服务保障能力，及以便未来出现同类事件时高校图书馆能做出快速反应，充分发挥高校图书馆的教育和信息服务保障作用[13] 。

4 结语

本文通过发放调查问卷了解新冠疫情下学生学习科研现状，并对国内6所高校图书馆的在疫情期间提供信息资源的类型、信息服务的内容形式及获取途径等进行调研比较。问卷结果展示了学生网络学习中的平台使用情况及网课学习效果；学生科研遭受一定影响，过程限缩至在线资源学习研究中。调研结果显示，高校图书馆在此期间为在校师生提供不同程度的网络资源及信息服务支持。

高校图书馆应根据网课教学中可能出现的问题及学生科研需求提供更深层次的服务，并完善突发公共卫生事件应对机制，以提高高校图书馆信息资源和服务的保障能力，为师生提供更优质的资源及服务。

OR-023

签约家庭医生对河南农村居民健康素养水平的影响

李梦蕾[1]、张帆[2,3]、张军亮[1]、孙晶晶[1]
1. 新乡医学院
2. 新乡医学院卫生信息资源研究中心
3. 新乡医学院教学质量监控与医学研究中心

目的 探讨签约家庭医生对农村居民健康素养水平的影响，为提升农村居民的健康水平提供参考。

方法 采用多阶分层整群随机抽样方法，抽取河南省 5 个地市 90 个行政村的 3943 名农村居民进行问卷调查，利用 Logistic 回归、倾向得分匹配及 c2 检验等方法进行统计学研究。

结果 进行倾向得分匹配前，有 17.18%的农村居民具备健康素养；进行倾向得分匹配后，性别、年龄、文化程度、婚姻状况、收入类型、是否患慢性病及健康状况等因素间的差异明显缩小，匹配效果较好（P ＜0.05）。

结论 农村居民的健康素养水平待提高，应加强对农村居民的健康教育；农村居民对签约服务内容的认知度不高，应加深农村居民对家庭医生签约服务的了解程度；家庭医生影响农村居民的健康素养水平，应承担起提升农村居民健康水平的重大责任。

健康素养（health literacy，HL）指个体获取、理解和处理基本健康信息或服务并做出正确健康决策的能力[1]，是衡量居民健康水平的重要指标。2018 年，中共中央、国务院在印发的《乡村振兴战略规划（2018-2022 年）》[2]中指出，要全面推进健康乡村建设，促进改善农村居民的健康状况。“健康中国 2030”规划纲要中也提出，要把农村和基层当作重点，逐步缩小城乡居民间的健康差距，从而实现全民健康[3]。家庭医生签约服务制度是深化基层医药卫生体制改革的重要举措，也是新形势下改善农村居民健康状况的有效途径[4]。河南省作为一个农业大省，截至 2018 年底，农村人口数量占我省总人口的 48.29%[5]；因此，提升农村居民的健康素养水平，对改善农村居民的健康状况具有十分重要的现实意义。本文通过调查河南省农村居民的健康素养状况并分析其与签约家庭医生之间的关系，为推进家庭医生签约服务，制定适合河南省农村居民的健康促进规划提供借鉴。

1 对象与方法

1.1 研究对象

采用多阶分层整群随机抽样方法，依据河南省东、中、西部的区域位置，随机选取 5 座城市，每座城市随机抽取 2 个代表县，每个代表县随机抽取 3 个乡镇，每个乡镇抽取 3 个行政村，采用随机抽样的方法，在每个行政村抽取 50 名居民作为研究对象来进行问卷调查。研究对象纳入标准：

1）户籍为河南省的农村居民（或在河南省居住时间超过 6 个月的农村居民）；2）年龄≥18 岁；3）能够理解问卷内容且语言交流无障碍。本研究共发放问卷 3943 份，回收有效问卷 3649 份，有效回收率为 92.54%。

1.2 研究方法

1.2.1 调查方法

以卫生部制定的《全国居民健康素养监测调查问卷》[6]为依据，自行设计《河南省农村居民健康素养调查问卷》，由经过专业培训的调查员统一进行问卷调查。调查内容包括：基本情况、健康知识、健康生活方式与行为、基本健康技能等共 37 道题目。指标定义：依据 2020 年贫困户脱贫标准[7]，人均纯收入未达到 4000 元以上即为低收入人群，反之则为高收入人群。评分标准：按照中国健康教育中心统一公布的评分要求进行打分，以百分制的原则，最终得分≥80 分即可判为具备健康素养，素养水平指具备健康素养的人在总调查人群中的占比。问卷信度效度 $P>0.85$。

1.2.2 倾向得分匹配法

倾向得分匹配（Propensity Score Matching,PSM）这一概念最早是由 Rosenbaum 和 Rubin 在 1983 年提出的，用以评估实施某项目或政策后的“处理效应”，是一种“反事实估计”分析方法[8]，多见于经济学领域。本文拟匹配一组与签约居民特征较为类似的未签约居民进行健康素养水平的比较，从而两组样本的差异可被视为签约家庭医生的作用。变量定义及赋值如表 1 所示。

表 1 变量定义及赋值

变量类别	变量名称	变量赋值
输出变量	是否具备健康素养	1=是；2=否
指示变量	是否签约家庭医生	1=是；2=否
特征变量	性别	1=男；2=女
	年龄	—
	文化程度	1=初中及以下；2=高中及专科；3=本科及以上
	婚姻状况	1=未婚；2=已婚；3=离异；4=丧偶
	收入类型	1=低收入；2=高收入
	是否患慢性病	1=是；2=否
	健康状况	1=好；2=比较好；3=一般；4=比较差；5=差

1.3 统计学方法

采用 EpiData 3.1 进行数据双录入；使用统计软件 SPSS 25.0 对数据进行统计分析。匹配容差设定为 0.02，利用 Logistic 回归模型计算倾向值，均衡性验证使用 t 检验，计量资料用 $\bar{X}\pm S$ 表示,计数资料用相对数表示,率的比较使用χ^2检验。以 $P<0.05$ 表示差异具有统计学意义。

2 结果

2.1 基本情况

被调查的 3649 名农村居民中，男性 1685 人（46.18%），女性 1964 人（53.82%）；平均年龄为 42 岁，30 岁以下者 845 人（23.16%），30-39 岁者 872 人（23.89%），40-49 岁者 613 人（16.80%），50-59 岁者 798 人（21.87%），60 岁及以上者 521 人（14.28%）；文化程度初中及以下者 1139 人（31.22%），高中及专科者 1952 人（53.49%），本科及以上者 558 人（15.29%）；低收入居民 1626 人（44.56%），高收入居民 2023 人（55.44%）；慢性病患者 1197 人（32.80%），非慢性病患者 2452 人（67.20%）；已签约居民 2178 人（59.69%），未签约居民 1471 人（40.31%）。

2.2 农村居民的健康素养水平及对签约服务内容的知晓情况

调查结果显示，有 627 名（17.18%）农村居民具备健康素养，其中 375 名（59.81%）居民签约了家庭医生。对于家庭医生的签约服务内容，1558 名（42.70%）居民知晓基本医疗服务，653 名（17.89%）居民知晓健康档案，467 名（12.80%）居民知晓家庭医生定期随访，806 名（22.09%）居民知晓健康体检，1727 名（47.33%）居民知晓健康教育讲座咨询， 2156 名（59.08%）居民知晓免费测量血压血糖，且有超过一半（52.15%）的居民曾利用过免费测量血压血糖这项服务。

2.3 农村居民签约家庭医生的影响因素及其对健康素养的影响

2.3.1 农村居民签约家庭医生的影响因素分析

在进行倾向得分匹配前，需对影响居民签约家庭医生的因素进行 Logistic 回归分析。结果显示，农村居民的性别、年龄、收入类型是影响其签约家庭医生的因素（P ＜0.05）。文化程度、婚姻状况、是否患慢性病及健康状况与是否签约家庭医生之间无显著相关性（P＞0.05）。如表 2 所示。

表 2 农村居民签约家庭医生影响因素 Logistic 回归分析

变量	B	OR(95%CI)	P
性别	-0.456	0.634（0.545,0.737）	0.000
年龄	0.079	1.082（1.021,1.147）	0.007
文化程度	-0.095	0.910（0.811,1.019）	0.103
婚姻状况	-0.085	0.919（0.832,1.015）	0.094
收入类型	-0.234	0.791（0.679,0.922）	0.003
是否患慢性病	0.030	1.031（0.874,1.217）	0.719
健康状况	-0.014	0.986（0.913,1.065）	0.723

2.3.2 匹配效果检验

在进行倾向得分匹配前，性别、年龄、文化程度、婚姻状况、收入类型、是否患慢性病及健康状况等因素在签约居民和未签约居民之间均具有统计学意义。由于这些因素的存在，致使无法对签约家庭医生对居民健康素养的影响做出正确判断，而倾向得分匹配法刚好能够消除这种组别之间的干扰因素。进行匹配后，可以发现签约居民和未签约居民在各因素方面的标准差明显缩小，说明签约居民与未签约居民在各因素的平均水平上较为接近，干扰因素得以消除，各因素在两组

之间的差异均无显著性（P ＞0.05），匹配效果比较好。如表 3 所示。

表 3　倾向得分匹配前后各变量的变化情况

变量	得分匹配	均值±标准差		t 检验	
		签约居民	未签约居民	t	P
性别	匹配前	1.51±0.500	1.58±0.494	3.815	0.000
	匹配后	1.45±0.498	1.47±0.499	0.835	0.404
年龄	匹配前	43.13±16.725	41.52±16.231	-2.888	0.004
	匹配后	42.26±16.087	43.24±17.204	-0.412	0.680
文化程度	匹配前	1.82±0.670	1.87±0.653	2.353	0.019
	匹配后	1.83±0.684	1.81±0.840	1.403	0.161
婚姻状况	匹配前	1.88±0.787	1.98±0.792	3.623	0.000
	匹配后	1.95±0.492	1.97±0.483	1.543	0.123
收入类型	匹配前	1.75±0.435	1.07±0.255	-5.808	0.000
	匹配后	1.34±0.474	1.14±0.346	-1.803	0.072
是否患慢性病	匹配前	1.36±0.480	1.28±0.449	-5.160	0.000
	匹配后	1.37±0.408	1.31±0.495	0.451	0.654
健康状况	匹配前	2.49±1.039	2.32±0.986	-5.031	0.000
	匹配后	1.93±0.258	1.93±0.253	0.272	0.786

2.3.3 农村居民健康素养水平与是否签约家庭医生的χ^2检验

倾向得分匹配后，有 208 名（22.2%）签约居民具备健康素养，而在未签约家庭医生的居民中，只有 138 名（14.8%）具备健康素养。与未签约居民相比，签约居民具备健康素养的比例更高一些。因此，是否签约家庭医生也会影响到居民的健康素养水平（P ＜0.05）。如表 4 所示。

表 4　农村居民健康素养水平与是否签约家庭医生的χ^2检验［n(%)］

	是否具备健康素养		χ^2	P
	具备	不具备		
是否签约家庭医生			17.377	0.000
签约	208（22.2）	727（77.8）		
未签约	138（14.8）	797（85.2）		

3 讨论

3.1 农村居民的健康素养水平有待提高

研究结果显示，2019 年河南省农村居民健康素养水平为 17.18%，低于 2019 年河南省居民健康素养水平（19.11%）[9],与国务院发布的《国务院关于实施健康中国行动的意见》（国发〔2019〕

13 号）[10]中提出的“到 2022 年全国居民健康素养水平不低于 22%”的目标还有一定的差距。作为基层的弱势群体，国家相关部门采取了一系列措施来提升农村居民的健康状况，但农村居民的健康素养水平仍与城市居民的相差甚远[11]。健康素养水平较低不仅会对农村居民的身体健康造成影响，而且还会加重他们的生活负担，导致出现“因病致贫”“因病返贫”的恶性循环[12]。因此，相关政府部门应以建设健康乡村为抓手，针对农村居民的特异性，积极开创易于农村居民接受的乡村健康教育形式，如以送电影下乡、戏曲广播、村委会喇叭、微信朋友圈等载体宣传健康教育，提升农村居民健康素养水平，尽早实现“2022 年全国居民健康素养水平不低于 22%”的战略目标。

3.2 农村居民对签约服务内容的知晓度不高

家庭医生签约服务以全科医生为核心，借助团队的形式与居民进行签约，并为签约居民提供连续的基本医疗服务[13]。作为健康的“守护者”、资源的“优化者”、费用的“监管者”[14]，家庭医生是提高居民健康水平的主要责任人。本次调查结果显示，17.89%的被调查居民知晓健康档案，12.80%的被调查居民知晓定期随访。由此可以看出，河南省农村居民对家庭医生签约服务内容的认知不够。由于缺乏充分的认知，许多农村居民都并不了解签约服务的具体内容，从而严重制约了基层签约服务工作的开展[15]。此外，调查结果还显示 52.15%的被调查居民知晓且利用过免费测量血压血糖这项服务，说明农村居民对家庭医生的医疗水准定位多为基础技术水平，不信任家庭医生的医疗技术。因此，相关政府部门要通过广泛宣传，让农村居民能够理解签约服务政策，从而提高他们对家庭医生的认知程度。加强基层医疗卫生人员的技术培训，号召更多的医务工作者投身基层卫生事业，整体提升家庭医生团队的技术水平，增强农村居民对家庭医生的依赖感和信任感[16]。

3.3 签约家庭医生是影响农村居民健康素养水平的重要因素

本次调查结果显示，在控制性别、年龄、文化程度、婚姻状况、收入类型、是否患慢性病及健康状况等因素后，是否签约家庭医生是影响农村居民健康素养水平的因素。22.2%签约家庭医生的农村居民具备健康素养水平，14.8%未签约家庭医生的农村居民具备健康素养水平，相较于未签约的农村居民，签约的居民具备健康素养的比例更高一些。家庭医生签约服务作为一种新型的医疗服务模式，在提高卫生资源利用、降低医疗费用、提升居民健康水平等方面有着重要意义[17]。因此，要充分注重家庭医生的建设内容，极力避免“签而不约”的现象，优化农村居民在基层卫生机构的就医体验[18]，促使农村居民与家庭医生形成良好的契约关系，以家庭医生为抓手[19]，进而提升农村居民的健康素养水平，充分发挥家庭医生在基层卫生服务机构中的重要作用 ，做好基层居民的健康守护者。

OR-024

农村中年慢性病患者在线健康信息获取障碍研究

王卓鑫[1,3]、吕合凯[2,1]、孙伟伟[2,1]、刘鹏[2,1]、潘玮[2,1]、陈兴智[2,1]、王辅之[2,1]
1. 蚌埠医学院 2. 卫生管理学院 3. 护理学院

背景 随着互联网的发展，网络已经成为居民获取健康信息的重要渠道之一。由于网络信息质量参差不齐以及农村居民文化程度较低，导致农村居民通过网络获取健康信息存在困难。

目的 评估农村社区中年慢性病患者在线健康信息获取能力，并分析其在获取在线健康信息时面临的障碍。

方法 在蚌埠市农村社区调研了70位中年居民，采用定量分析和定性分析的混合研究方法，对其在线健康信息获取行为展开深入研究。

结果 70位受访者健康信息素养自评得分普遍较高（Mean=46.2，SD=4.89），然而70%的受访者在线健康信息获取能力偏低。不同性别、教育程度以及互联网使用时间的受访者健康信息素养水平存在显著差异；主要的获取障碍包括两方面，一是健康信息素养水平不高，缺乏判断在线健康信息质量的能力；二是现有的在线健康信息质量差，并且缺乏有效的质量监控机制。

结论 农村社区中年居民对于自身健康信息素养评价较高，但实际在线健康信息获取能力与自评结果差距较大，大多数受访者缺乏筛选高质量在线健康信息的技能。

随着信息技术的快速发展，互联网已成为人们获取健康信息的重要途径。2020年4月，中国互联网络信息中心（CNNIC)发布的第45次《中国互联网发展状况统计报告》中显示，我国网民规模达9.04亿，较2018年增长5708万，互联网普及率达64.5%，较2018年底提升4.9个百分点[1]。由此可见，互联网已经逐渐渗透到居民的日常生活中。与此同时，居民对健康信息的关注日益增多，利用网络获取健康信息的需求也更加迫切。根据调查显示，68%的互联网用户会使用网络查询与健康相关的信息[2],但由于网站提提供的健康信息的质量参差不齐，导致部分用户不能够判断网络信息的准确度[3][4][5]。

近年来，我国大力发展乡村振兴战略，农村居民健康问题受到极大关注。互联网逐渐改变了农村社区居民获取健康信息的行为模式，成为居民获取健康信息的重要渠道之一[6][8]，截至2020年3月，我国农村网民规模为2.55亿，占整体网民规模的28.2%[1]，由于我国农村人口基数较大，

居民文化程度相对较低，经济水平相对落后，居民在获取在线健康信息时存在诸多问题[6][7]。因此，我们应该更加关注农村居民在线健康信息获取能力。

目前，已有大量针对在线健康信息获取行为的研究。例如，Hena Naz Dindeng 等通过评估不同类型的互联网使用人群在寻求在线健康信息方面的特征，发现不同年龄、种族以及社会经济地位的人群在使用和获取健康信息方面存在差异[9]；Matthew J Manierre 通过建立回归模型分析性别差异对健康信息获取行为的影响，发现女性相比男性更倾向于进行健康信息的搜寻[10]；Sharon Swee-Lin Tan 等通过系统地回顾现有的关于患者寻求在线健康信息及其对医患关系影响的研究，发现网络搜索健康信息可以改善医患关系[11]。虽然已有关于在线健康信息获取行为的大量研究，但大多是通过问卷调查或系统回顾的方法来研究人群的在线健康信息获取行为模式，而通过测试不同群体的实际检索操作行为并且量化获取能力的研究甚少，尤其是针对农村社区居民这一重点人群的研究较为缺乏，本文将关注点放在农村社区中年居民群体上，探析居民在线健康信息检索获取行为，评估该群体在线健康信息获取行为的能力水平。

1 研究对象与方法

1.1 研究对象

本研究调查对象面是小蚌埠镇的慢性病患者。采用两阶段抽样：第一阶段，采用单纯随机抽样在小蚌埠镇确定 10 个农村社区作为抽样社区；第二阶段，在每个农村社区采用选择性配额抽样选取 10 名受访者。对受访者开展入户调研。

样本纳入标准：① 在农村社区连续居住时间超过 6 个月；② 年龄在 30-65 岁之间；③ 确诊罹患一种或一种以上慢性病。样本排除标准：① 中途退出调查；② 身心状况不适合参与本次调研。

1.2 研究设计

（1）健康信息素养自评

本次调研采用罗丹等人汉化并修订的日常健康信息素养简单自评问卷[12]（EHIL-10），完成对受访者健康信息素养的评测。该问卷包括两部分，第一部分为调查对象的社会人口学基本情况，即性别、年龄、教育程度、职业类型、网龄等情况；第二部分为健康素养自评问卷，包括 14 个条目，经因子分析得出 4 个维度：健康信息意识（Health information consciousness，HIC）、健康信息获取（Health information seeking，HIS）、健康信息评价（Health information evaluation，HIE）和健康信息应用（Health information application，HIA）。

（2）在线健康信息检索行为测试

在调研人员不参与指导的情况下，让受访者自主利用笔记本电脑登录互联网在线查询健康信息，对受访者的测试行为全程录屏记录。该部分要求受访者利用互联网完成三个测试题目：①利用互联网，找到最佳的包含疾病健康知识的网站；②根据自身健康状况，选择并保存 3 篇健康信息文章；③选择一篇您认为关于某种慢性病最佳治疗方案的文章。

利用 DISERN 量表评估每位受访者所选择在线健康信息网站的可靠性以及所选治疗方案的信息质量。DISERN 量表由大英图书馆和牛津大学所开发[13] ，分为两个部分。第一部分共有 8 个条目，总分 40 分，用来评估网站的可靠性；第二部分共有 7 个条目，总分 35 分，用来检测治疗方案文章的质量。这两个部分的总得分可以反映出受访者获取高质量在线健康信息的行为能力水平。

（3）半结构化访谈

半结构化访谈问卷围绕“检索过程和策略”和“网站健康信息质量评价”两部分开展。由调查员与受访者进行深入交流，了解受访者选择健康信息网站的标准，甄别信息质量的方法以及在获取健康信息过程中遇到的困难。访谈过程全程录音，调研结束后，通过录音回放提取整理每位受访者对于 15 个访谈问题的回答，并提炼出农村社区中年居民获取在线健康信息面临的障碍。

1.3 统计学方法

数据录入及预处理采用 EXCEL2010,统计分析应用 SPSS16.0 完成。描述性统计分析采用均值±标准差，组内差异运用 t 检验/方差分析，p<0.05 表示结果有统计学意义。

2 结果

2.1 调查样本概况

本次调查样本的社会人口学概况见表 1。

表 1. 调查样本基本情况

基本情况	n	%	基本情况	n	%
性别			**互联网使用时间**		
男	39	55.71	1 年及以下	12	17.14
女	31	44.29	1 年-3 年	14	20.00
年龄			3 年以上	44	62.86
30-45 岁	27	38.57	**手机使用时间**		
46-60 岁	33	47.14	没用过	0	0
61-65 岁	10	14.29	1 年-3 年	6	8.57
文化程度			3 年以上	64	91.43
小学及以下	9	12.86	**慢性病情况**		
中学/中专/技校	56	80.00	罹患 1 种慢性病	42	60.00

大学	5	7.14	罹患 2 种及 2 种以上慢性病	28	40.00
研究生及以上	0	0	**首次确诊罹患慢性病时间**		
职业类型			1 年以内	7	10.00
企事业单位员工	29	41.43	1 年-3 年	24	34.29
务农	10	14.29	3 年-5 年	12	17.14
自谋职业者	19	27.14	5 年以上	27	28.57
无工作	12	17.14			

2.2 健康信息素养水平

本次调查居民健康信息素养平均得分为 46.21 分，其中，最高分为 56 分，最低分为 34 分。70 个受访者中得分不低于 40 分的人数占 91.4%（64 人），调查显示大部分居民健康素养处于中上等水平。四个子维度中健康信息应用维度（mean=7.9143，SD=0.37078）得分最低。各个维度得分统计数据见表 2。

表 2. 70 个样本健康信息素养自评描述

	维度 1(HIS)	维度 2(HIE)	维度 3(HIC)	维度 4(HIA)	HIL 总分
Mean	12.80	14.30	11.20	7.91	46.21
SD	2.23	2.98	1.04	0.37	4.89

健康信息素养自评问卷的单因素分析结果表明，除了性别（p=0.036）、教育程度（p=0.010）、互联网使用时间（p = 0.042）以外，其他社会人口学因素与健康信息素养评分结果均没有显著差异（p > 0.05）。单因素分析结果见表 3。

表 3. 健康信息素养得分的社会人口学因素差异

社会人口学因素	Mean	SD	*t*/F	*p-val*
年龄			0.249	0.781
31-45 岁	46.44	4.27		
46-60 岁	46.33	5.20		
61-65 岁	45.20	5.79		
性别			-2.135	0.036
男	45.13	5.08		
女	47.58	4.35		
教育程度			4.992	0.010
小学及以下	42.78	5.21		
中学/中专/技校	47.09	4.56		
大学	42.60	4.39		
研究生及以上	0	0		
职业类型			1.354	0.265

企事业单位员工	47.24	4.33		
务农	44.30	4.92		
自谋职业者	45.16	5.22		
无工作	47.00	5.34		
互联网使用时间	3.33	0.04	3.334	0.042
1 年及以下	43.00	5.78		
1 年-3 年	46.93	4.68		
3 年以上	46.86	4.45		
手机使用时间			0.039	0.844
没用过	0	0		
1 年-3 年	45.83	7.88		
3 年以上	46.25	4.62		
慢性病情况			0.010	0.921
罹患 1 种慢性病	46.17	4.51		
罹患 2 种以及上慢性病	46.29	5.51		
首次确诊慢性病时间			0.478	0.699
1 年以内	45.00	4.00		
1 年-3 年	46.25	6.04		
3 年-5 年	46.25	3.62		
5 年以上	46.92	4.56		

2.3 行为能力测试

（1）任务一：利用互联网，找到最佳的包含疾病健康知识的网站

70 位受访者中有 64 位受访者能通过互联网找到包含健康信息的网站，完成该项任务。其中，64.1%的受访者选择了专门的健康信息类网站，如百科名医网、中华康网、39 健康网等；23.4%的受访者将百度作为搜索健康信息的最佳网站；12.5%的受访者选择了其他类网站作为搜索健康信息的最佳网站；8.6%的受访者没有完成该项任务。通过回放录屏资料发现，绝大多数受访者选择最佳健康信息网站具有很大的随机性，很少有受访者能够根据网站访问量和权威信息等指标来选择最佳的健康网站。

（2）任务二：根据自身健康状况，选择并保存 3 篇健康信息文章

70 位受访者中 65 位受访者能够根据自身状况完成该项任务，其中，35.4%的受访者在保存文章时存在困难，每位受访者保存的 3 篇文章中基本上都包括了疾病预防和治疗类文章。在保存文

章时存在的主要困难是受访者不会操作互联网，除此之外，个别网站内下载或保存文章需要收费，只能保存网页链接以及个别网站保存文章前需进行注册，步骤比较繁琐。

（3）任务三：选择一篇您认为关于某种慢性病最佳治疗方案的文章

92.9%的受访者都顺利地完成了此项任务，利用 DISERN 量表评估文章质量的结果显示，得分均在 30 分以下，其中，90%的受访者得分低于 20 分（满分 35 分），可见大部分受访者获取的文章质量极低。通过浏览受访者所选的治疗文章可见，文章来源大多没有指明文章作者、发布机构、发布日期以及参考书目等，文章内容个人主观意见较强，且对于治疗方案的描述也较为笼统，可见大部分受访者都不具备根据文章信息来源、信息透明度等标准来甄别文章质量的技能。

在信息检索能力方面，通过回放录屏视频资料，发现全部完成该项任务的 64 位受访者中，89.1%的受访者都是直接利用搜索引擎（其中，百度搜索引擎占比 52.6%）输入慢性病治疗方案等关键词来检索文章，10.9%的受访者是直接通过医学健康资讯类网站查找疾病治疗类文章。其中，利用搜索引擎获取的文章主要来源于百科知识、经验问答等，这类文章的 DISERN 量表评分较低，涉及的治疗方案大多来源个人主观意识，也大多未能描述治疗方法的效果和风险。不仅如此，搜索引擎直接检索出的信息纷繁复杂，受访者难以做出正确判断，个别受访者甚至选择与医学健康无关的网站文章作为最佳治疗方案文章。

在获取的信息质量方面，应用 DISERN 量表（满分 75 分）对受访者的在线检索行为进行量化分析。结果表明，70 位受访者的平均得分为 37.4 分，得分在 10 分以下占 10%，得分处于 20-40 分的人数占 44.29%，得分处于 40-60 分的人数占 45.71%，可见大部分受访者的在线健康信息获取能力较弱。该量表第一部分（满分 40 分）是用来检测受访者所选网站的可靠性，平均得分为 23.8286 分，说明受访者查询的健康信息网站质量普遍偏低；量表第二部分（满分 35 分）是用来检测受访者获取的疾病治疗方案的可靠性，平均得分为 13.5714 分，说明受访者获取的文章质量普遍较低，绝大部分受访者甄别在线健康信息质量的能力严重不足，获取高质量在线健康信息的能力有限。DISERN 量表得分统计见表 4。

表 4. DISERN 量表得分统计情况描述

	Part 1	Part 2	Total
Mean	23.83	13.57	37.4
SD	9.32	5.99	13.77

2.4 半结构化访谈

56 名受访者完成了半结构化访谈任务，归纳、整理所有访谈内容，结果见表 5。

表 5. 56 名受访者的半结构化访谈结果分析

问 题	结 果	分 析
您是通过什么搜索引擎了解到健康信息网站？你之前访问过这个网站吗？	31 人使用百度，12 人使用 360 搜索，10 人使用搜狗,1 人使用 QQ 浏览器，2 人未使用搜索引擎仅用微信公号。 其中 36 人以前访问过该网站，20 人没有访问过。	使用百度作为搜索引擎的受访者占总人数的 55.4%，百度是部分居民经常使用的搜索平台。
你为什么选择这个网站而没有选择其他网站呢？	8 人认为网站比较方便；12 人表示是个人习惯；9 人肯定网站的内容；8 人肯定网站的可信度；4 人表示是随机选择的；15 人表示不清楚。	选择原因主要：搜索方便、网站来源官方、内容全面具体，与自身病情比较符合。
您如何判断该网站是提供健康信息最好的网站？	7 人通过网站知名度判断；6 人通过文章来源；12 人通过网站内容质量；20 人通过与其他网站、自身病情以及医生建议作对比；4 人表示网络信息只是参考；5 人不知道如何判断；2 人未发表观点。	主要判断标准：网站知名度高，网友评价好、网站来源官方权威可靠、网站内容丰富、广告较少、文章质量高、与自身病情相符合等。
您感觉找到这类健康信息网站有困难吗？保存感兴趣的文章有困难吗？如有困难，哪些步骤让你感到困难？	4 人表示找到网站有困难；9 人表示下载或保存文章有些困难。	大约有 16%的受访者在获取网络健康信息时有困难，主要困难在于：不会操作、不会保存链接、下载需要收费等。
您对该网站第一印象如何？即对于网站的首页布局感觉如何？	54 人表示网站还可以，不错；2 人表示可信度不高，有虚假信息；2 人没有表明态度。42 认为网站的布局合理，4 人认为网站有垃圾广告，影响阅读，3 人没有关注该方面，7 人没有表明态度。	大部分受访者都对于健康信息网站的布局安排表示认同，只有小部分受访者认为虽然布局合理，但还是有垃圾广告。
该网站在设计上哪些方面让你满意？	12 人对于网站布局很满意；14 人肯定了网站的搜索效率及内容；5 人对于网站提供专家在线答疑栏目很满意； 25 人没有给出明确回答。	大部分受访者表示，版面详尽清晰、分类明确一目了然、搜索信息全面、设置专业医师在线解答、无广告插件和链接等。
网站在设计上有哪些方面让你不满意？	18 人对于网站弹出垃圾广告很不满意；38 人表示没有不满意之处，都很满意。	不满意之处：网页经常弹出大量广告链接和咨询，影响文章阅读以及虚假宣传等。
该网站对您最有帮助的健康信息是什么？是否提供了您以前不了解的信息？	24 人认为慢性病的治疗方案最有帮助；5 人认为疾病预防知识最有帮助；11 人认为疾病康复保健有帮助；16 人未给出明确回答；54 位受访者表示学到了新知识，2 人态度不明确。	一半受访者认为健康网站带来的最大帮助是提供了疾病的治疗方案；并且从健康信息网站中能够获取新知识。
您如何判断网站信息内容的可靠性？	16 人结合自身病情症状作比对；10 人咨询医生判断；5 人通过亲身实践；6 人通过网站提供的内容是否客观专业判断；9 人凭感觉判断 ；4 人通过与其他网站的信息进行对比判	无法辨别网站信息可靠性的受访者占 10%左右，大部分受访者主要判断方法为：结合自身病情、询问医生、亲自尝试网站指导方案、内容来源专家等

	断；6 人不清楚如何判断	
您还会再来访问该网站吗？您会把该网站推荐给家人及朋友吗？	51 人表示还会再次访问，2 人表示否定，3 人未发表态度；51 人表示会推荐，2 人表示否定，3 人未发表态度。	绝大多数受访者认为网站简单易操作，提供了较为专业的信息，愿意再次访问并推荐给他人；不会再次访问的受访者表示从网络上了解的信息比较浅，不可信，因此不会推荐给其他人。

（1）在线健康信息检索策略

①搜索引擎的使用

访谈结果显示，55.4%的受访者选择百度作为查找健康信息的搜索引擎，首要原因是受访者不需要专业的检索技巧就能快速获取到多条相关信息。89.1%的受访者直接将搜索引擎作为检索健康信息的查寻入口而并没有选择专门的医学健康信息网站，说明目前权威健康信息网站的群众知晓度和选择率较低。35.7%的受访者表示此次受访是自己第一次利用网络查找有关健康的文章，说明仍有部分居民不会主动利用互联网查找健康信息知识。

②在线健康信息网站的选择

关于受访者选择该网站的原因，66.1%的受访者知道自己选择该健康信息网站的原因，主要原因有：搜索方便、个人习惯、内容全面齐全、可信度高等。但仍有 33.9%的受访者并不清楚为什么选择该网站，选择网站带有一定的随机性。关于如何判断该网站是提供健康信息的最好网站，87.5%的受访者能够通过一些方法判断网站是否可靠，但仍有 12.5%的受访者不清楚如何判断。他们认为该网站中只要涉及大量疾病治疗类健康知识就是最佳网站，而对于如何甄别站内文章的真实度，他们并不清楚。

③在线健康信息的保存

根据访谈结果显示，7.1%的受访者表示找到该类健康信息网站有困难，83.9%的受访者表示保存文章很容易，但仍有 16.1%的受访者表示保存感兴趣的文章有困难。说明大部分中年居民掌握了如何保存在线健康信息的能力，但在实际操作中仍有少部分居民能力欠缺。

（2）在线健康信息评价

①网站布局

访谈结果显示，75%的受访者认为网页的布局设计的还不错。当问及对于该网页设计的哪些方面比较满意时，44.6%的受访者表示不清楚，25%的受访者对于网页的搜索效率以及搜索所得的内容比较满意；8.9%的受访者认为网页内有专家答疑，能够比较及时的解决自己的问题。当问及对于网页设计哪些方面不满意时，67.9%的受访者表示没有不满意之处，32.1%的受访者对于网页中穿插了大量广告链接表示相当不满意。

②在线健康信息内容价值

访谈结果显示，96.4%的受访者都表示通过健康信息网站获取到了新知识，当问及该网页的哪些健康信息对于受访者最有价值时，28.6%的受访者没有给出明确回答，35.7%的受访者认为疾病的治疗方案对自己最有价值，且 91.2%的受访者表示会再次访问该网站，会将该网站推荐给身边的亲朋好友，仅有个别受访者不愿意再次访问该健康信息网站。

③在线健康信息质量评价

访谈结果显示，89.3%的受访者都能通过一些方法甄别信息的可靠性，10.6%的受访者表示网站文章五花八门，自己并不清楚如何判断其可靠性。由此可见，部分受访者不具备甄别文章可靠性的能力，在获取真实可靠的健康信息时存在困难。

（3）意见和建议

访谈结束之际，我们对访谈录音进行总结分析，归纳受访者对于获取在线健康信息的意见和建议。结果见表 6。

表 6. 受访者的建议

	建议描述
建议一	希望政府相关部门能够进行网络获取健康信息的培训，重视居民的在线健康信息知识学习，教授居民查寻获取高质量健康信息的方法
建议二	希望能够可以减少广告，规范网站相关的相关信息
建议三	尽量用通俗化语言表述专业性知识，便于老百姓理解
建议四	互联网中的信息参差不齐，希望相关网站提供更有权威性的治疗方案
建议五	希望网站可以更加完善，提供更多的免费平台

3. 讨论

本研究使用定性和定量研究相混合的方法调研了 70 位农村社区慢性病患者获取在线健康信息的现状，并综合评估了他们的健康信息素养水平、在线健康信息获取能力，并总结出他们在获取在线健康信息时所面临的障碍。通过利用问卷自评，我们发现农村慢性病患者在线信息获取水平处于中等水平，尽管他们对于自身健康信息素养评价很高，但实际的在线获取能力与自评结果差距较大。绝大多数受访者都能认识到获取健康信息的重要性，但是对于获取健康信息的主动性不强。此外，大多数受访者难以甄别网络健康信息的质量，缺乏筛选高质量在线信息的技能。虽然大多数受访者对于如何判断在线健康信息的优劣有一定的方法，但是通过行为能力的测试来看，他们并没有按照所设想的方法来获取高质量的健康信息，而仅仅是凭主观感觉来判断的，在选择的过程中带有一定的随机性，甚至一些居民选择的健康信息是带有大量广告的。

定量研究结果显示，所有受访者健康信息素养自评得分均在 30 分以上，得分不低于 40 分的人数占九成以上，说明绝大多数受访者健康素养自评水平处于中上等。60 岁以上的受访者健康素

养得分较低，且通过 SPSS 软件分析结果得出，受访者健康信息素养与其性别、教育程度以及使用互联网时间的长短有显著性差异，女性比男性健康信息素养得分要高，这与 Basch CH[14]、沈默[15]等人的研究结果相似；使用互联网超过三年以上的受访者健康素养得分要高于三年以下的受访者，这与 Camilla Melholt 等人的研究结果相似[16]。中学及以上学历的受访者健康素养得分要高于小学文化程度的，这与前期的相关研究相同[17]，但由 SPSS 结果可看出大学学历的受访者健康素养自评得分较低，这可能是由于本次抽取的样本量中大学文化程度的人数较小的缘故。对于 70 位受访者的实际在线检索行为定量分析结果显示，DISERN 量表得分处于 20-40 分的人数占 44.29%，得分处于 40-60 分的人数占 45.71%，得分处于 10 分以下的占 10%，说明大多数受访者的实际健康信息获取能力处于偏低水平。其中，利用 DISERN 量表测评文章质量的得分多在 10-20 分之间（满分 35 分），说明绝大多数受访者缺乏获取高质量信息的技能。结合健康素养自评结果与 DISERN 量表检测结果，可以发现大多数据居民对于健康信息的意识、查询、评价和应用能力的自我评价过高，实际上，农村慢性病患者在线获取健康能力的水平有限，尤其是获取高质量的疾病治疗、预防、保健和康复等健康信息的能力严重不足。

定性研究结果显示，56 位受访者获取健康信息的方式大多是通过综合门户网站（如百度）直接检索健康信息，而通过专业的健康医学网站获取健康信息的受访者占比较低；受访者大多具有较强的健康信息获取意识 ，且大多都能利用互联网查询并保存健康信息，但是对于找到高质量且可靠性高的健康信息的技能相对缺乏，甄别高质量在线健康信息的能力不足，10.6%的受访者根本不清楚如何评价一篇有关健康信息的文章，另有 16.1%的受访者是凭自己的主观感觉获取健康信息的。除此之外，大部分受访者更加关注网站内容的丰富度，而对于内容本身的真实度和可靠性关注较少，说明绝大多数农村居民对于获取高质高效的在线健康信息的意识较为缺乏。

综上所述，本研究发现农村慢性病患者通过互联网获取在线获取健康信息时面临的主观障碍主要体现在不会使用互联网，难以理解网站中有关疾病知识的专业术语，并且缺乏判断网络信息可靠性和真实性的基本方法；客观障碍主要体现在缺少权威的专业医学健康信息类网站资源，部分网站页面布局复杂，易用性差，过多的广告链接等，都使得用户的健康信息检索效率大大降低。为此，政府及相关部门应加强针对普通公众在线健康信息获取能力的教育，帮助他们获取免费的高质量网络健康信息资源；同时，加大对于在线健康信息发布和传播的监管，完善相关法律法规，营造良好的网络健康信息环境。

OR-025

医联体内医疗信息共享的关键影响因素研究

李雨晨、刘智勇
华中科技大学同济医学院

目的 医疗信息共享是实现医疗信息资源价值最大化的重要途径，但其实际落地情况不甚良好，仍存在共享不及时、不全面等问题。为弥补该研究领域的不足，明确医联体信息资源共享的关键影响因素，本研究采用定性与定量研究相结合的方式对其展开探讨。
方法 通过文献研究归纳总结影响医联体信息资源共享的有关因素；应用系统工程中的决策试验与评价实验室（DEMATEL）方法确定医联体信息资源共享的关键影响因素。
结果/结论 通过梳理文献获得 30 个医联体信息资源共享的影响因素，基于 DEMATEL 复杂系统模型确定了 11 个关键影响因素。基于此提出以下推进医联体信息共享的策略：加强顶层设计和统筹规划；强化信息共享建设相关的专项财政支持；建立统一的医疗信息资源共享平台；完善支持分级诊疗体系的信息共享机制。

1. 引言

随着我国社会发展和经济增长，公共医疗体系效率低下、医疗资源分布不均、医疗服务能力存在差距等问题开始显现[1]。2013 年 1 月，“全国卫生工作会”正式提出了“医疗联合体”的概念。此后，随着国家和地方各级卫生行政部门推进医联体建设的政策导向，医联体建设成为我国医疗改革的重要切入点。而医联体上下联动、多机构业务协同的工作机制决定了其必须以完善的信息共享体系为支撑。医联体内的医疗信息共享是打通患者就医环节、实现一体化诊疗服务和医疗信息不间断的前提，是医疗服务能够保持连续性、协调性和整体性的基础，是充分发挥医联体功能、减少医疗资源浪费的关键。国家出台大量政策鼓励医疗信息共享，但医疗信息共享实际落地情况不甚良好，仍存在共享不及时、不全面等问题。医联体内部医疗机构的信息共享情况尚不明确，缺乏完善的梳理；医疗机构的主观能动性没有被充分调动，没有明确动力与意愿去推进医疗信息共享；卫生行政部门缺乏对影响医联体内医疗信息共享的关键因素的认知，推进医疗信息共享的行动和措施缺少科学性和依据。

目前国内外关于医疗信息共享的研究主要集中在对共享机制、实现技术的研究以及系统或平台的开发与应用上，而鲜有学者从医疗机构角度对医疗信息共享的影响因素开展研究。为了弥补该研究领域的不足，使卫生行政部门和医院信息部门的工作人员深入了解医联体内信息共享的关键影响因素及其影响机制，本研究采用定性与定量研究相结合的方式对影响医联体内部医疗信息共享的关键因素展开探讨。

2. 医联体内信息共享的影响因素梳理

2.1 影响因素框架构建

信息共享具有覆盖范围广泛、类型复杂等特点，其研究可能涉及到不同部门、不同领域、不同行业人员。由于目前国内外关于医疗信息共享影响因素的研究较少，因此本研究搜集文献时将范围扩大至公共组织间信息共享的相关研究。对于公共组织信息共享的影响因素，已经有了一些较为系统的研究。既有研究提出的理论模型从不同的研究视角对不同组织和技术环境下的跨部门信息共享活动展开研究，形成了一系列较为成熟的理论模型[2-5]。从信息共享理论模型的拓展到信息文化模型的研究，再到部门间信息生态环境的研究，一方面反映了随着信息和通信技术发展给信息共享实践带来了更多机遇和可能，研究逐步走向成熟，同时也反映了这一研究领域具有复杂性、系统性等特征。本研究结合现有的公共信息资源共享影响因素框架，将医联体内信息资源共享的影响因素划分为两个类属，分别是共享支持因素和组织管理因素。其中，共享支持因素包括政策法规维度、技术能力维度、信息化建设维度、资源保障维度；组织管理因素包括组织间合作关系维度、体制机制维度、人员认知维度、交流共享过程维度。由此得到医联体内部信息共享影响因素的理论模型。

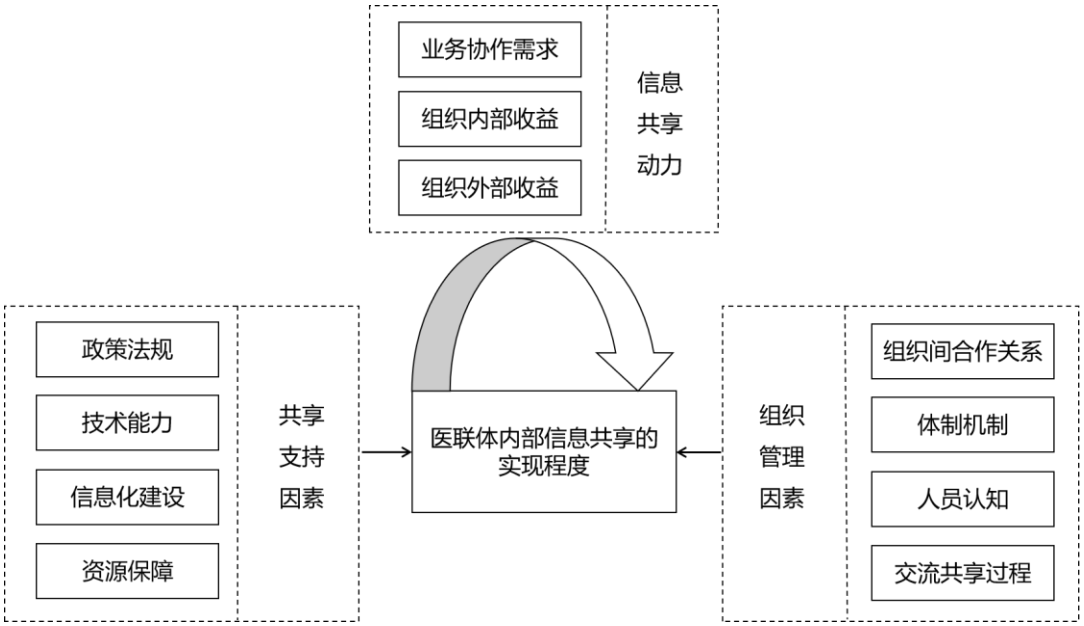


图 1 医联体内部信息共享的影响因素模型

2.2 影响因素维度分析

本研究通过应用复杂系统影响因素重要性判定的决策试验与评价实验室（DEMATEL）模型，对医联体内信息资源共享的关键影响因素进行识别。具体来说，就是基于影响因素间的相互关联程度获得直接影响矩阵，通过计算构建出各影响因素的标准化影响矩阵和综合关系矩阵，系统分析各维度下有关影响因素的综合影响指数，从而确定医联体内信息资源共享的关键影响因素。运用

DEMATEL 模型的前提在于构建出相对完整、全面的影响因素集合。为此，本研究通过检索、总结、归纳目前国内外公共组织间信息资源共享的相关研究，提取出 30 个医联体内信息共享的潜在影响因素。

2.2.1 政策法规维度

医联体信息资源共享的政策法规是指医联体内医疗机构在实施信息共享时所依据的总体规划、法规政策以及标准规范。其中，医联体信息资源共享的总体规划是指卫生行政部门或者医联体自身对于信息共享的规划和要求。医联体内的信息共享涉及多机构和多层次，在推行过程中需要来自制度体系的支持和保障。因此总体规划和制度保障体系是调动医联体成员信息共享积极性的重要力量。法规政策通过对医联体信息资源共享的目标、边界和路径进行规定，保障信息共享过程中的程序正义。标准规范是控制信息资源质量、确保共享可行性的行动指南，是信息共享实践过程的重要前提。现有研究普遍认为健全有效的法律法规体系能有效保障公共组织的信息资源共享[6]，该维度所涉及的因素主要包括政府对信息资源共享的立法、标准设计和框架制定等。本研究所整理的政策法规维度下的影响因素包括：共享总体规划(FL1)、制度保障体系(FL2) 、信息共享法规政策(FL3) 和信息管理标准(FL4) 。具体如表 1 所示。

表 1 政策法规维度下的相关因素

编号	因素	涉及此因素的部分研究或报告
FL1	共享总体规划	胡平等[7]、尤佳等[8]、那旭[9]、谢晓专等[10]、黄辉[11]、周成兵等[12]
FL2	制度保障体系	Mattessich 等[13]、Bigdeli 等[14]、Landsbergen 等[15]、杨会良等[16]、周成兵等[12]、龙健[17]
FL3	信息共享法规政策	Akbulut 等[18]、Yang 等[19]、Bigdeli 等[14]、Gil-Garcia 等[20]、Landsbergen 等[15]、Jing F 等[21]、胡平等[7]、吕欣等[22]、杨会良等[16]、张强等[23]、那旭[9]、谢晓专等[10]、韩普等[24]、何瑞文等[25]、龚立群等[26]
FL4	信息管理标准	Bigdeli 等[14]、吕欣等[22]、谢晓专等[10]、韩普等[24]

2.2.2 技术能力

技术能力维度因素主要是指能够推进医联体信息共享的信息技术方面的因素，包括对某些具体信息技术（如云计算技术、数据集成技术等）的应用能力。信息技术是信息共享的基础，医联体内不同医疗机构之间存在系统不兼容、软硬件不适配的问题，需要依靠信息技术手段才能实现不同平台、不同格式标准的医疗信息的整合。同时，由于医疗信息具有敏感性，信息共享系统的等级权限、安保能力和隐私保护能力受到绝对的重视。经过文献梳理，总结出技术能力维度下的影响因素主要包括：信息质量管理能力(FT1) 、信息系统安全管理能力(FT2) 、信息系统隐私保护能力(FT3) 、信息系统兼容性（FT4）以及信息技术应用能力(FT5)。具体如表 2 所示。

表 2 技术能力维度下的相关因素

编号	因素	涉及此因素的部分研究或报告
FT1	信息质量管理能力	Yang 等[19]、Bigdeli 等[14]、吕欣等[22]、周伟等[27]、那旭[9]、韩普等[24]
FT2	信息系统安全管理能力	Yang 等[19]、Bigdeli 等[14]、Gil-Garcia 等[20]、Jing F 等[21]、周伟等[27]、张强等[23]、谢晓专等[10]、何瑞文等[25]、龚立群等[26]
FT3	信息系统隐私保护能力	Bigdeli 等[14]、吕欣等[22]、那旭[9]、李妍等[28]、陈兰杰等[29]
FT4	信息系统兼容性	Akbulut 等[18]、Yang 等[19]、吕欣等[22]、谢晓专等[10]、黄辉[11]、韩普等[24]
FT5	信息技术应用能力（IT 能力）	Akbulut 等[18]、Yang 等[19]、Bigdeli 等[14]、Jing F 等[21]、尤佳等[8]、张强等[23]、那旭[9]、谢晓专等[10]、龚立群等[26]、邓春林[30]

2.2.3 信息化建设

信息技术是信息化的基础，信息化建设是信息共享的主要阵地。不少研究都强调过信息系统协同、数据集成、统一云平台等技术作为公共组织信息资源共享的技术支撑所发挥的重要作用。医联体成员单位间的信息资源共享主要以各医疗机构的信息化基础设施为环境，以信息共享平台为媒介，因此医联体的信息化建设情况直接影响到信息资源共享的实现。经文献梳理发现信息化建设维度下的相关因素主要包括信息化基础设施建设(FI1)及共享平台建设(FI2)。具体如表 3 所示。

表 3 信息化建设维度下的相关因素

编号	因素	涉及此因素的部分研究或报告
FI1	信息化基础设施建设	Dawes[44]、Akbulut 等[18]、Yang 等[19]、Bigdeli 等[14]、Gil-Garcia 等[20]、胡平等[7]、周伟等[27]、那旭[9]、谢晓专等[10]、龚立群等[26]
FI2	共享平台建设	周伟等[27]、杨会良等[16]、周成兵等[12]、陈兰杰等[29]

2.2.4 资源保障

资源保障维度是指用于支持医联体信息资源共享顺利实施的具体保障措施，如专项财政保障、资源汇总统筹等。除把握法律法规、技术能力、信息化建设等关键维度外，医联体信息资源共享的实施需要耗费人力、财力、物力等资源，需要良好的资源保障措施为信息共享助力。经文献梳理与提炼概括，确定资源保障维度下的相关因素主要包括：财政专项支持(FR1)、人员队伍建设(FR2)以及各方面资源汇总和统筹(FR3)。具体如表 4 所示。

表 4 资源保障维度下的相关因素

编号	因素	涉及此因素的部分研究或报告
FR1	财政专项支持	Dawes[44]、Mattessich 等[13]、Akbulut 等[18]、Bigdeli 等[14]、Gil-Garcia 等[20]、胡平等[7]、那旭[9]、谢晓专等[10]、邓春林等[30]、陈玉梅[31]
FR2	人员队伍建设	Dawes[44]、胡平等[7]、那旭[9]、谢晓专等[10]、黄辉[11]、陈玉梅[31]
FR3	各方面资源汇总和统筹	Dawes[44]、Yang 等[19]、尤佳等[8]、谢晓专等[10]

2.2.5 组织间合作关系

组织间合作关系维度是指医联体内医疗机构在参与医疗信息共享的过程中共同面临的或者需要协调的因素。在既往研究中，很多学者已经指出跨组织系统成功的最基本要素就是以信任为基础的跨组织关系。组织间信任关系有助于跨组织的信息共享的实现。具体到医联体中，组织间信任关系主要体现在组织合作的紧密程度上，也即基于不同医联体模式的合作密切性。此外，参与信息共享过程的主要利益相关者目标的一致性对信息共享系统的成功也有重要的影响作用。因此，

组织间合作关系维度下的相关因素主要包括：组织目标一致性(FC1)和组织合作紧密性(FC2)。具体如表 5 所示。

表 5 组织间合作关系维度下的相关因素

编号	因素	涉及此因素的部分研究或报告
FC1	组织目标一致性	Mattessich 等[13]、Bigdeli 等[14]、Gil-Garcia 等[20]、Jing F 等[21]、尤佳等[8]、陈玉梅[31]
FC2	组织合作紧密性	Akbulut 等[18]、Bigdeli 等[14]、Gil-Garcia 等[20]、胡平等[7]、尤佳等[8]、何瑞文等[25]、陈玉梅[31]

2.2.6 体制机制

体制机制维度包含体制和机制两个层面。通常来说，体制是指机构设置和职权划分，表现为一种长期、稳定的管理框架。组织的管理体制是指其为有效实现特定的管理目标和组织任务，依据法律法规、内部规章和标准规范制定的一整套组织和人事管理制度。医联体信息共享体制即是围绕医联体信息资源共享活动所设置部门、岗位以及相应权责分配等内容。体制维度下的相关因素具体包括：统一领导体制(FS1)、高层领导支持(FS2)、明确责任主体(FS3)等。机制与体制相比更加灵活且形式多样，是指保证制度落实的程序、规则和运作方式。目前研究中，建立明确的信息共享机制、绩效考核和激励机制、利益协调机制和监督机制对公共组织成功实施信息共享的作用已经得到广泛认可。医联体信息资源共享机制即是围绕医联体信息资源共享的具体操作程序、规则或运作方式。机制维度下主要包括医联体为实现信息共享所实施的具体措施，相关因素包括完善利益协调机制(FS4)、建立共享机制(FS5)、实行激励机制(FS6)、推行监督考核机制(FS7)。具体如表 6 所示。

表 6 体制机制维度下的相关因素

编号	因素	涉及此因素的部分研究或报告
FS1	统一领导体制	Mattessich 等[13]、Yang 等[19]、吕欣等[22]、那旭[9]、周成兵等[12]、龚立群等[26]、陈兰杰等[29]
FS2	高层领导支持	Akbulut 等[18]、Jing F 等[21]、胡平等[7]、谢晓专等[10]、黄辉[11]、 龙健[17]、何瑞文等[25]
FS3	明确责任主体	Gil-Garcia 等[20]、黄辉[11]、陈玉梅[31]
FS4	完善利益协调机制	Dawes[44]、周伟等[27]、黄辉[11]、龙健[17]、李妍等[28]、邓春林等[30]

FS5	建立共享机制	黄辉[11]、邓春林等[30]
FS6	实行激励机制	Akbulut 等[18]、黄辉[11]、周成兵等[12]、龚立群等[26]、陈兰杰等[29]、陈玉梅[31]
FS7	推行监督考核机制	Akbulut 等[18]、Yang 等[19]、谢晓专等[10]、何瑞文等[25]、龚立群等[26]、李妍等[28]、陈兰杰等[29]、陈玉梅[31]

2.2.7 人员认知

人员认知维度是指涉及医联体信息共享过程的有关领导和医务人员意识、素质和信任度等因素。就目前信息技术阶段和医联体管理体制来看，医联体内医疗机构间的信息共享的实施主要是以人为主体。不少研究者都提出过公共组织的工作人员心理和人员素质相关的影响因素。经文献梳理与总结，本研究认为人员认知维度下的相关因素主要包括人员对信息共享的认识(FH1)、人员信息共享观念(FH2)、人员对信息共享过程的信任度(FH3)、人员对信息共享机构的信任度(FH4)。具体如表 7 所示。

表 7 人员认知维度下的相关因素

编号	因素	涉及此因素的部分研究或报告
FH1	人员对信息共享的认识	杨会良等[16]、谢晓专等[10]、黄辉[11]
FH2	人员信息共享观念	周伟等[27]、杨会良等[16]、谢晓专等[10]、邓春林等[30]
FH3	人员对信息共享过程信任度	吕欣等[22]
FH4	人员对信息共享机构信任度	胡平等[7]、吕欣等[22]、龚立群等[26]

2.2.8 交流共享过程

交流共享过程维度主要是指医联体内的业务流程和信息共享流程。信息共享过程依托于业务流程的同时也会反过来影响原有的业务流程。因此，业务流程应进行相应改造、优化，使其与信息共享制度相适应，否则就会造成协作效率和管理水平的低下。在诸多信息资源共享的因素研究中，关于信息共享流程本身的因素讨论较少，但不可忽视其便利性和顺畅性在公共组织信息共享实践中发挥的作用。交流共享过程维度下的相关因素总结如下：业务流程兼容性(FP1)、信息共享流程便利性(FP2)、信息共享流程顺畅性(FP3)。具体如表 8 所示。

表 8 交流共享过程维度下的相关因素

编号	因素	涉及此因素的部分研究或报告
FP1	业务流程兼容性	Dawes[44]、Yang 等[19]、Bigdeli 等[14]、黄辉[11]、龚立群等[26]
FP2	信息共享流程便利性	吕欣等[22]
FP3	信息共享流程顺畅性	吕欣等[22]、陈玉梅[31]

3. 医联体内信息共享的关键影响因素分析

3.1 DEMATEL 模型构建与计算

本研究基于国内外相关文献回顾，从中提取出医联体信息资源共享的 30 个影响因素，基于 DEMATEL 模型对各个影响因素进行测度。首先组建专家咨询小组，咨询小组由 10 名成员组成，职业均与信息管理工作相关，包括卫健委信息中心负责人、医联体内医疗机构的信息管理人员等。向专家咨询小组成员发放问卷调查其对医联体信息资源共享的各影响因素之间的关系的评价，确定各影响因素对其他所有因素的影响程度，得出医联体内部信息资源共享的影响因素的关系体系。其中，影响度和被影响度分别指该因素造成影响程度的总和以及受其他所有因素影响程度的总和。构造关系矩阵后，经计算得出各影响因素的影响度（D）与被影响度（C），如表 9 所示。

表 9 医联体信息资源共享各影响因素的影响度和被影响度

编号	影响度（D）	rank	被影响度（C）	rank	编号	影响度（D）	rank	被影响度（C）	rank
FL1	1.28	29	0.00	0	FC2	1.27	26	0.94	24
FL2	1.61	28	0.11	3	FS1	1.18	26	0.35	24
FL3	1.71	28	0.10	1	FS2	1.23	26	0.59	24
FL4	0.14	21	0.02	1	FS3	0.34	24	0.95	24
FT1	0.13	18	0.40	23	FS4	0.35	26	0.45	24
FT2	0.13	4	0.17	24	FS5	1.27	26	1.29	24
FT3	0.11	4	0.22	24	FS6	0.43	23	0.93	24
FT4	0.29	26	0.69	24	FS7	0.45	23	0.96	24
FT5	0.25	22	0.32	22	FH1	0.14	2	1.05	24
FI1	0.57	26	0.62	23	FH2	0.08	2	1.58	30
FI2	0.78	26	0.93	24	FH3	0.07	2	0.70	30
FR1	1.23	26	0.29	17	FH4	0.07	2	0.27	24
FR2	0.27	24	0.32	17	FP1	0.31	26	0.79	24

FR3	0.30	26	0.35	17	FP2	0.27	26	0.67	24
FC1	0.74	26	0.51	24	FP3	0.29	26	0.74	24

根据医联体信息资源共享的各影响因素的影响度与被影响度结果，可以发现影响度比较高的因素包括：共享总体规划（1.28）、制度保障体系（1.61）、信息共享法规政策（1.71）、专项财政支持（1.23）、组织目标一致性（0.74）、组织合作紧密性（1.27）、统一领导体制（1.18）、高层领导支持（1.23）、建立共享机制（1.27）。这些影响因素的影响度较高，说明其在影响因素网络中更多地对别的影响因素产生影响，属于起源型因素。起源型因素主要集中在法规标准维度、体制机制维度和组织合作关系维度，因此，应当以体制改革和立法实践为抓手推进医联体信息资源共享。被影响度比较高的因素包括：共享平台建设（0.93）、组织合作紧密性（0.94）、明确责任主体（0.95）、建立共享机制（1.29）、实行激励机制（0.93）、推行监督考核机制（0.96）、人员对信息共享的认识（1.05）、人员信息共享观念（1.58）、业务流程兼容性（0.79）、信息共享流程顺畅性（0.74）。这些影响因素的被影响度较高，属于结果型因素，说明其在影响因素网络中更多地是受到其他因素的影响。结果型因素分布较为分散，主要涉及技术维度、人员认知维度、交流共享过程维度。与起源型因素相比，结果型因素对医联体信息资源共享的实践发挥更直接的影响作用。可以看出，共享机制的建立、共享平台的建设以及人员对信息共享的认知和接纳是影响医联体信息资源共享的最为重要的结果型因素。

3.2 关键影响因素识别

基于 DEMATEL 权重网络的理论模型，本研究进一步计算出了医联体信息资源各影响因素的中心度和原因度。影响因素的中心度（D+C）用以辨析该因素在整个影响因素网络中的重要性；影响因素的原因度（D-C）用以辨析该因素在整个影响因素网络中的贡献程度。医联体信息资源共享的各影响因素的中心度和原因度如表 10 所示：

表 10 医联体信息资源共享各影响因素的中心度和原因度

因素标号	中心度	原因度	因素标号	中心度	原因度
FL1	1.28	1.28	FC2	2.21	0.33
FL2	1.72	1.50	FS1	1.53	0.83
FL3	1.81	1.61	FS2	1.82	0.64
FL4	0.16	0.12	FS3	1.29	-0.62

FT1	0.53	-0.27	FS4	0.80	-0.10
FT2	0.30	-0.04	FS5	2.56	-0.02
FT3	0.33	-0.11	FS6	1.36	-0.49
FT4	0.98	-0.40	FS7	1.41	-0.51
FT5	0.57	-0.07	FH1	1.19	-0.91
FI1	1.19	-0.05	FH2	1.65	-1.50
FI2	1.71	-0.15	FH3	0.77	-0.63
FR1	1.52	0.94	FH4	0.34	-0.20
FR2	0.59	-0.06	FP1	1.10	-0.48
FR3	0.66	-0.05	FP2	0.94	-0.39
FC1	1.25	0.23	FP3	1.03	-0.45

本研究结合中心度（D+C）和原因度（D-C）对医联体信息资源共享的关键影响因素进行识别。各影响因素的中心度（D+C）和原因度（D-C）的分布如图 4-2 所示。从分布情况来看，部分影响因素既具有较高的中心度也具有较高的原因度，如 FL1（共享总体规划）、FL2（制度保障体系）、FL3（信息共享法规政策）、FR1（专项财政支持）、FS1（统一领导体制）、FS2（高层领导支持），说明这些因素在整个影响因素网络中具有相对重要的作用，相比其他影响因素具有更强的中心性和原因性。还有部分影响因素中心度很高，虽然其原因度低，但也说明许多因素通过作用于该因素间接影响医联体信息资源共享水平，其在影响因素网络中也具有重要地位，如 FI2（共享平台建设）、FC2（组织合作紧密性）、FS5（建立共享机制）、FH2（人员信息共享观念）。

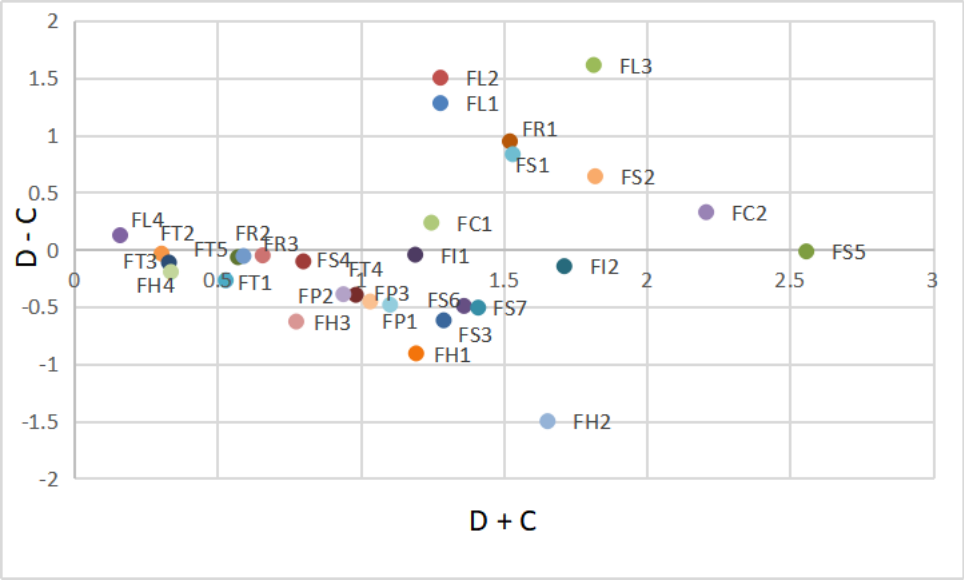


图 2 医联体信息资源共享的中心度与原因度分布

基于医联体信息资源共享的各影响因素的综合影响指数，也即影响度、被影响度、中心度和原因度等指标，本研究选取中心度和原因度排名前列的因素，通过比较和分析识别出了 11 个医联体信息资源共享的关键影响因素。按照中心度从高到低排序，这些因素分别是：FS5（建立共享机制）、FC2（组织合作紧密性）、FS2（高层领导支持）、FL3（信息共享法规政策）、FL2（制度保障体系）、FI2（共享平台建设）、FS1（统一领导体制）、FR1（专项财政支持）、FS7（推行监督考核机制）、FP1（业务流程兼容性）、FT4（信息系统兼容性）。

4. 结果与讨论

从识别出的 11 个医联体信息资源共享关键影响因素所涉及的维度来看，法规标准维度、体制机制维度入选医联体信息资源共享的关键因素达到半数，这说明关于医疗信息共享的立法、政策制定、机制建立是推进医联体信息资源共享的关键。

法规政策与标准维度下入选的两个影响因素分别为制度保障体系（FL2）和信息共享法规政策（FL3），其既具有较高的中心度，也具有较高的原因度。制度保障体系是建设医联体信息资源共享统一平台的基础，从更高层面统筹规划信息共享平台建设、明确医联体信息共享的相关责任主体，有利于信息资源共享工作的落实。信息共享法规与政策是医联体信息资源共享的法律依据，通过在个人隐私保护、共享权力范围、责任追究等方面给出法律上的明确界定，能够带有一定强制性的促进和规范医联体信息资源共享。

在体制机制维度下，统一领导体制（FS1）、高层领导支持（FS2）、建立共享机制（FS5）、推行监督考核机制（FS7）是最为关键的影响因素，其中心度和原因度均排在前列。这与目前的医

联体管理体制有关，医联体涉及多所医疗机构，在行政区划、财政投入、医保支付、人事管理等方面均存在一定的壁垒。医联体内跨机构的信息资源共享需要一个专门的部门或者核心医院对信息共享工作进行统一管理和整体协调。因此，在医联体内部需要建立统一领导体制和信息共享机制，才能打破机构之间的壁垒，进一步加强信息共享，充分发挥信息共享在提高业务协同效率方面的优势。此外，由于高层领导拥有较大的决策权，其支持对于医联体信息资源共享也具有十分关键的作用。推行监督考核机制则主要是通过将信息共享行为与绩效评估挂钩，倒逼医务人员加强对信息共享的认识和接纳。

此外，信息化建设维度、技术能力维度、资源保障维度、组织间合作关系维度和交流共享过程维度各有一个因素入选关键影响因素，分别是共享平台建设（FI2）、信息系统兼容性（FT4）、专项财政支持（FR1）、组织合作紧密性（FC2）和业务流程兼容性（FP1）。

在信息化建设维度下，共享平台建设（FI2）的中心度较高，是医联体信息资源共享成功实施的直接推手。共享平台建设是具有工具属性的影响因素，直接作用于后续的关键步骤，包括人员在平台上共享和使用数据、管理平台的数据质量等。同时，共享平台建设又是法律法规标准、体制机制改革等方面因素的综合作用结果。而提高信息系统兼容性与实现跨机构数据整合和建设共享平台都有着紧密的联系，机构间系统的数据整合是共享平台中数据可用性的重要保证。因此，信息系统兼容性（FT4）虽然原因度较低，本研究仍认为其是影响医联体信息资源共享的重要因素。

资源保障维度下的专项财政支持（FR1）兼具中心度和原因度，在复杂因素网络中处于相对重要的地位。专项财政支持是应用新的技术或方法的必要条件，无论是软硬件设施的配置还是人才队伍的建设都离不开大量资金的支持。因此，充足、稳定的经费来源对医联体信息资源共享的可持续发展尤为重要。

在组织间合作关系维度下，组织合作紧密性（FC2）具有更高的中心度，是典型的中心度极高而原因度较低的影响因素之一。组织合作紧密性受医联体模式的影响较大，同时也受到具体医联体管理体制的影响。组织合作紧密性在公共组织间信息共享的影响因素中更多地被表述为组织间信任，该因素的重要性在前人研究中较少被提及，而根据本研究分析，该因素是医联体信息资源共享的关键影响因素。

交流共享维度下，业务流程兼容性（FP1）是更关键的影响因素。医联体信息共享的目标是实现业务协同效率的提升，而业务流程的改变和适应也会作用于医联体信息资源共享，推动医疗信息共享从被动向主动转变。

5. 对策与建议

本研究以医联体信息资源共享的关键影响因素为研究问题，通过梳理文献获得了 8 个维度共 30 个医联体信息资源共享的影响因素，使用 DEMATEL 复杂系统模型获得了 11 个关键因素，分别为：建立共享机制、组织合作紧密性、高层领导支持、信息共享法规政策、制度保障体系、共享平台建设、统一领导体制、专项财政支持、推行监督考核机制、业务流程兼容性、信息系统兼容性。结合本研究发现，提出以下推进医联体信息共享的策略：

5.1 加强顶层设计和统筹规划

虽然目前医联体内医疗机构间已实现医疗信息的初步共享，但总体来说，仍存在信息资源条块分割的问题，导致信息共享价值未被充分挖掘。因此，国家和地方卫生行政部门应研究制定卫生信息共享建设总体框架，并结合各地发展现状研究制定适合本地情况的规划、政策与标准规范，推进顶层设计的落地和细化。同时，明确涉及卫生信息共享的各组织机构职责定位、权限以及资源提供和共享方式等，建立规范的多方合作机制，以确保医疗信息共享的准确性、及时性、完整性。鉴于目前卫生信息化建设已具有一定基础，在进行规划制定时应遵循连接现有信息系统而非推翻重建的理念，充分结合客观实际，制定兼具灵活性和适用性的规划。

5.2 强化信息共享建设相关的专项财政支持

有效的资金投入是开展医联体信息资源共享建设的必要保障，信息化基础设施建设、运行维护、日常管理等需要有足够的稳定、持续的专项财政支持。同时，应将资金细化至信息共享建设的每个具体环节，通过加强监管来保证信息共享建设专项资金利用的合理性。同时，充足的资金支持也是信息共享专业人才队伍建设的必要前提。信息化共享建设需要高素质的专业化人才，由专职人员负责信息共享工作，有利于简化工作流程，提高共享效率。因此，各级卫生行政部门和医联体核心医院应在财政支持下加强信息共享建设，促进医联体信息资源共享建设可持续发展。

5.3 建立统一的医疗信息资源共享平台

针对目前不同区域医联体信息共享程度存在差距、信息共享建设具备一定基础的现状以及医疗信息共享覆盖多业务领域和跨机构的特点，可采用建立第三方平台的共享模式。对于卫生信息共享资源库建设初具规模的地区，在原有共享平台基础上扩大覆盖范围；对于还未建设共享资源库的地区，投入资源支持建立区域层面统一的医疗信息资源共享平台。在技术实现方面，要求各医疗机构按照统一标准进行接口改造，依托计算机自动实现数据处理和传输，即可实现信息的一方采集、多方利用。该模式对于机构之间合作紧密性和协作机制具有较高的要求，也需要对信息共享的规则和范围进行进一步研究。

5.4 完善支持分级诊疗体系的信息共享机制

为满足覆盖全人群的连续性健康服务的需求，必须保证信息共享的全面性、及时性和准确性。如果缺乏有效的医疗信息共享机制，信息的流通就会受阻。信息共享的最终目标是提高医疗服务效率、节约医疗资源、改善居民健康管理效果，因此，应该建立完善基于医疗协同服务的信息共享机制。在信息共享建设方面，首先应考虑各方的需求，使信息共享成果能够为各方带来一定的收益，并以信息共享收益为切入点调动医联体成员单位开展信息共享的动力和积极性；其次应注重数据质量的保障和数据价值的挖掘，将分散的原始医疗数据以患者为单位进行整合，并能够及时地将信息与各相关方共享。在信息共享互操作方面，一是要实现数据规范化管理，为不同成员单位信息系统之间的数据整合提供基础；二是要完善医疗协同业务系统的功能，使其能够在患者发生转诊的同时支持互操作管理，及时更新患者诊疗信息，形成完整的健康记录；三是要保障医疗隐私信息的安全性，通过设置相应权限限制对患者信息的查询和使用。最后，应引导多方参与居民健康信息共享与利用，提高不同利益主体的参与积极性，形成良性共享机制。

OR-026

安徽医科大学信息管理与信息系统专业教学改革探讨
——基于毕业生就业情况的问卷调查分析

丁晓、宋国强
安徽医科大学

通过调查安徽医科大学信息管理与信息系统及其前身专业往届毕业生的就业情况，收集毕业生对专业教学改革的意见和建议，找出医学信息管理专业教育与社会需求之间的差距，探讨专业教学改革的新方向。

1 对象与方法

1.1 研究对象

信管专业毕业生除了第一届学生是 30 人，以后每届学生都保持在 60 人左右的规模，到 2018 年 9 月已毕业学生达 1300 余人。本次调查对象选取的是 2002 届-2016 届毕业生，主要考虑

的是这些学生已经毕业至少三年，工作相对稳定，也积累了一定的经验，提供的信息有效性相对更佳。

1.2 研究方法

调查问卷以选择题为主，结合部分主观题，内容主要涉及三个方面：毕业生的就业基本情况(职业方向和类型，工作中能力需求类型和层次) ；专业教学过程对就业能力的影响；毕业生对专业教学的评价和建议。

通过 QQ 和微信等在线平台向毕业生发放问卷，共回收问卷 203 份，其中有效问卷 196 份。安徽医科大学信管专业 2002 届-2016 届毕业生共计 870 余人，调查覆盖约 22.5%的调查对象，调查具有统计学意义，能有效说明教学中存在的问题，对于教学改革具有很好的借鉴意义。

2 调查结果及分析

2.1 毕业生就业情况基本信息

调查的 196 位毕业生中，就业地区分布以安徽为主，其次是江苏、上海、浙江。华东地区就业分布占到调查人数的 78%。就业单位分布以医院等卫生事业单位为主，其次是软件公司（就业单位分布见表 1）。毕业生当前从事的职业（岗位）以管理岗位为主，其次是行政后勤岗和技术研发岗位（就业岗位分布见表 2）。毕业生所从事工作与专业完全对口的占 24.5%，基本对口的占 33.2%，有点关联的占 27.6%，完全不对口的占 14.8%（见表 3）。毕业以来跳槽次数为 0 次的占 49.5%，跳槽 1-2 次的占 35.2%，跳槽 3 次以上的占 15.3%。调查的 196 位毕业生中现在的学历为本科的 128 人（占 63.4%），硕士及以上 74 人（占 36.6%）。

表 1 毕业生当前就业单位分布统计表

	频数	百分比
医院等卫生事业单位	101	51.5%
软件公司	38	19.4%
公务员	16	8.2%
高校、科研等其他事业单位	15	7.7%
医药企业	14	7.1%
其他	12	6.1%

表 2 毕业生当前就业岗位分布统计表

	频率	百分比
管理人员	87	44.40%
行政\后勤人员	39	19.90%

技术\研发人员	33	16.80%
专业人士（如教师、律师等）	13	6.60%
其他	12	6.10%
销售人员	10	5.10%
财务\设计人员	2	1.00%

从调查的196位毕业生数据来看，超过三分之一的毕业生选择了继续深造，毕业生毕业后工作相对稳定，绝大多数毕业生从事的工作是与专业相关的，毕业生就业服务地区以华东地区为主，就业单位以国家机关事业单位为主，企业单位为辅，就业岗位以管理岗位为主，行政和技术研发岗位为辅。从毕业生就业现状来看，与我们培养“能够在医疗卫生机构信息部门、医药软件公司等从事信息系统建设、信息管理和医疗数据挖掘的应用型高素质专门人才”的培养总目标是一致的。

表3 专业吻合度统计表

	频率	百分比	累积百分比
完全对口	48	24.5%	24.5%
基本对口	65	33.2%	57.7%
有点关联	54	27.6%	85.2%
完全不对口	29	14.8%	100.0%
总计	196	100.0%	

2.2 教学过程对就业能力的影响

在工作单位最看重的技能方面，调查反馈信息显示，实践能力、团队协作、开拓创新协调应变能力排在计算机能力和英语能力的前面。在信息专业的课程设计中最需要加强培养的能力调查中，创新思维和创新能力、人际沟通能力、和信息技术与管理能力排在前三项。就业时的主要考虑因素的统计显示信息专业毕业生考虑最多的因素是工作所在地，其次是工作前景和薪酬福利，而专业匹配度排在了后面（见图1），这从侧面反映信息专业宽口径的人才培养思路是适应社会需求的，大专业的思想可能是我们今后专业建设应该考虑的方向。

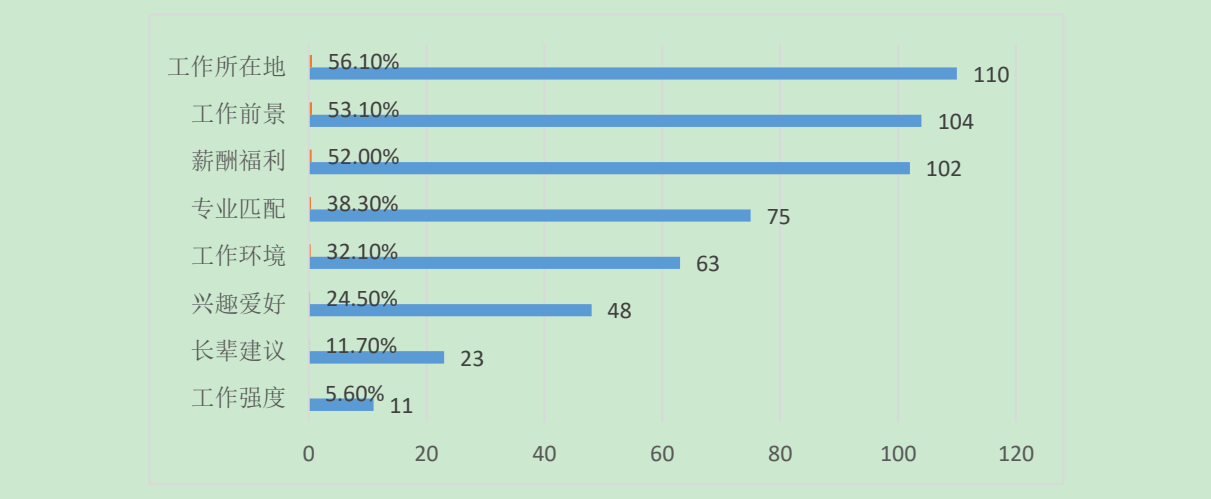


图1 择业主要考虑因素统计

在关于高校就业指导课程对于就业的影响程度的调查中，超过 85%的被调查者认为影响不大。关于专业必修课程对于就业的影响的调查中，接近 85%的调查者认为有影响或有较大影响。关于专业必修课的课程设置和时间安排是否合理的调查中，30.5%的被调查者认为合理，52.7%的被调查者认为一般，16.8%的被调查者认为不合理。关于实习的设置和时间安排合理性调查中，47.8%认为合理，44.8%认为一般，7.4%认为不合理。关于所学专业课程与工作的适用性调查中，20.2%认为完全不适用，66.5%认为基本适用，13.3%认为非常适用。在影响学习效果的主要因素调查中，学习兴趣、教学方法和教学内容及授课教师个人魅力排在前位。对工作有帮助的课程调查中，《医院信息系统》、《计算机网络》、《临床医学概论》和《管理学原理》排在前位。

从工作满意度调查来看，绝大部分被调查对象对其工作现状是满意的（见表 4）。专业满意度调查显示，绝大部分调查者对专业课程设置和专业内容设置是满意的，相对课程设置，内容设置的满意度要低一点（见表 5）。

根据调查的结果和毕业生反馈的意见，专业教学应该着重加强实践能力、沟通协作能力、创新应变能力和信息技术与管理能力的培养；专业课程设置的安排上是合理的，但是课程内容的深度还需要加强。教师的教学方法和教学内容对于教学效果影响很大，如何提升教师授课水平是专业发展应该考虑的一个重要问题。

表 4 工作满意度调查表

题目\选项	很不满意	不满意	一般	满意	很满意
目前就业总体状况	2(0.99%)	2(0.99%)	57(28.08%)	112(55.17%)	30(14.78%)
薪酬福利水平	3(1.48%)	9(4.43%)	82(40.39%)	89(43.84%)	20(9.85%)
组织文化	2(0.99%)	5(2.46%)	84(41.38%)	90(44.33%)	22(10.84%)
个人发展空间	2(0.99%)	12(5.91%)	77(37.93%)	91(44.83%)	21(10.34%)

表 5 专业课程教学满意度表

题目\选项	很不满意	不满意	一般	满意	很满意
专业课程设置	6(2.96%)	16(7.88%)	101(49.75%)	74(36.45%)	6(2.96%)
专业课程内容	6(2.96%)	22(10.84%)	99(48.77%)	70(34.48%)	6(2.96%)

3 对专业教学改革的建议

3.1 以社会需求为导向，明确培养目标，加强学生专业思想建设

医学院校信管专业是医学、计算机和管理多学科交叉的专业，培养的是复合型宽口径人才，这既是该专业的优势，同时也是其劣势所在。如何避免出现什么都学，什么都不精的局面，就需要我们明晰专业定位，明确专业培养目标。作为省属院校，该专业培养的学生主要面向华东地区，那么结合华东地区地域特点，多开展校企合作联合培养学生，既要把握市场发展动向，同时要兼顾信管学科特点，在市场需求和学科特点之间找到一个平衡点，为社会培养合格的人才和公民，而不仅仅只是劳动力。[1]

对于在校学生，除了通过开设的专业导论课加强专业思想建设，在校期间可以通过各种形式，比如请高年级优秀学生或者已经毕业的学生来交流传授经验，巩固学生的专业思想。对外也要加大宣传力度，提升学生自信心，促使学生用积极的心态来面对学习。

3.2 界定专业边界，夯实专业基础，加强信息技术能力培养，建立专业特色。

信管专业不同于计算机专业，它是以计算机为工具，来实现管理手段的技术化。因此界定专业边界，寻找专业特色是该专业发展应该考虑的重要问题。根据历届学生的岗位性质的信息反馈，我们应该在强调信息技术，奠定专业技术基础的同时，大力培养学生的信息分析能力和管理决策能力，而不的仅仅只是强调其计算机的能力。[2]毕竟和计算机专业的学生比，我们的技术能力肯定是不如别人的，而信息的分析和管理的决策能力是该专业应该也必须加强的方向。因此在课程设计上，学科特点、学生需求、社会需求这三个因素都需要考虑进去。[3]我们可以以“信息技术基础—编程语言—数据库原理—网络原理—信息系统开发”作为主线安排[4]，结合医学行业的特点，在夯实基础的同时，着力培养学生的医院信息系统开发和管理的能力，结合医疗大数据、云计算等新技术，在紧跟时代潮流的同时，构建专业特色。

3.3 加强师资队伍建设，促进学科间融合创新发展

信管专业发展受限的一个很大的瓶颈在于师资队伍的建设一直没有跟上来。师生比过低，师培的路径太少，教师的发展受限。虽然计算机学科、管理学科和信管学科背景的教师都有一定的比例，但是学科教师之间的融合交流的机会太少，大家都是在自己的学科背景下传道授业，导致学生学到的知识也是比较零散，很难成为一个体系。信管学科的优势就在于学科之间的交叉融合，而不是各种知识的叠加。[5]因此未来除了引进人才，加大学科投入，教师也要走出去，多参与各种形式的师资培养，深入企业，积累更多的实践经验，紧跟时代潮流，同时教师之间也要开展各种形式的教学研讨和科研合作，不断促进学科间的融合创新，真正使信管专业的优势得到发展。

3.4 积极开展课外活动，提高学生综合素质

信管专业是一门应用性很强的专业，学生能力的培养和锻炼除了书本的理论知识，更多的是来自于实践教学。除了教学计划里的见习和实习的机会以外，还可以开展形式多样的课外活动，来增加学生与实践接触的机会。比如可以根据学生的就业兴趣增设各种兴趣小组，鼓励学生参与创新创业、建模大赛等活动，定期开展校企之间交流活动，让学生结合市场动向，有目的的学习。创新能力、人际沟通能力和健康的心理是个人职业长效发展不可或缺的能力，因此在校期间也要多开展提高此类能力的活动，提高学生的综合素质。

4 结语

信管专业是一个年轻的应用性很强学科，专业发展的道路如何走，是一个长期的课题。未来还可以结合在校学生和用人单位的意见以及借鉴同行专业院校发展的经验，进一步修正人才培养目标，以期能获得更好的培养效果。

OR-027

医学信息教育智慧云平台构建研究

徐义涛、文庭孝
中南大学

目的 随着生物医学信息迈入大数据时代，传统的医学信息教育已无法满足对于海量生物医学数据的挖掘、处理需求，智慧教育以及智慧教育云平台等理念与技术的出现为当前的医学信息教育问题提供了解决方案。本文基于智慧教育、智慧教育平台相关概念，探讨了搭建医学信息教育智慧云平台的重要意义，通过分析智慧教育相应的功能模块，设计了医学信息教育智慧云平台的总体架构，对于架构实现的关键技术进行了介绍。旨在为我国医学信息教育的改革提供思路。

1 引言

我们已经逐步步入了 5G 时代，移动网络带宽增长和连接速度有了显著提高，在这期间，云计算、物联网、大数据技术等迅速得到普及应用，极大地推动了教育领域的创新改革，智慧教育的

概念亦在此背景下诞生。智慧教育以数字资源为核心载体，强调学生参与教学过程，同时扮演问题的提出者和解决者的角色，为学生创造更好的学习环境以及更具个性化的学习体验[1]。在此背景下的医学信息教育面临着两大挑战，日新月异的信息技术和海量增长的生物医学数据，当前该类数据已具备大数据的特征，对其进行分析与挖掘是医学信息教育中的核心内容。然而当前的医学信息教育在海量生物医学数据的分析与挖掘方面略显不足，主要体现在技术的滞后以及数据资源的封闭上。现有的网络学习平台并能从根本上改变传统的医学信息教学模式，因此利用先进的信息技术和海量生物医学信息资源搭建一个智慧教育云平台，实现智慧学习，是应对当前医学信息教育挑战的重要创新。

2 智慧教育云平台概述

2.1 智慧教育

智慧教育是整合互联网、云计算、大数据、移动通信、分布式计算等先进信息技术为一体的增强型数字教育（Enchanced-education），是对数字教育的进一步发展。它主要包括智慧教育、智慧学习、智慧科研、智慧评价、智慧管理等核心内容[2]。

智慧教育在整合现有的数字教育资源的基础上，同时将大数据挖掘和分析技术用于智慧教育产生的学习行为数据，从而明确学生学习过程中的显性变量和隐性变量。然后将分析得出的结论反馈给学习管理者、服务者和参与者，采取准确的培养方法，促进反思和教学改进，以提升教师的教学效率和学生的学习效率，在此基础上实现因材施教、个性化的教学方法[3]。智慧教育作为一种新型教育模式，在教育形态具备物联化、感知化、泛在化、智能化和个性化等特点[4]，是未来教育发展的主要趋势。

2.2 智慧教育云平台

智慧教育云平台是智慧教育理念的具体实现形式，是基于云计算技术、虚拟化技术、分布式存储等技术架构的一个智能化的，且能为不同用户提供租用或免费云服务的操作平台，通过该平台可实现智慧学习、智慧教育、智慧管理、智慧科研、智慧评价等智慧教育的主要内容。同时该类平台能够通过数据挖掘、学习行为分析等技术，向不同的学生、教师推送个性化定制内容，能够对其资源进行跨度调配，实现资源共享[5]。

在智慧教育云平台的教学模式中，教师起主导作用，学生为中心，智慧教育云平台是支撑，因材施教是关键，通过师生的良性互动，共同促进知识的掌握并培养创新能力[1]，这也是当前医学信息教育所提倡的学习模式。

3 医学信息教育智慧云平台构建的必要性

3.1 医学信息教育存在的问题

医学信息学是一门交叉学科，通过计算机及相关信息技术来处理诸如生物医学数据、信息和知识的存储、组织、检索和优化利用等一系列医学信息管理任务，以此来辅助医学领域的科研与实践，提高解决问题和制定决策的准确性、及时性和可靠性。医学信息学处理的对象是生物学的、医学的甚至是更为广义的健康数据、信息和知识[6]，目前这些数据已经具备大数据的特征。可以看出在医学信息教育中，信息技术和生物医学数据是核心内容，信息技术的学习能够有效地处理海量生物医学数据，而在实际数据处理实践中也能巩固对于信息技术的掌握，甚至给出改进方案。然而在目前的医学信息教育中存在如下问题：

（1） 课程设置滞后

作为一门交叉学科，医学信息学课程包括大量的医学基础知识、管理学、统计学知识。在最关键的计算机知识方面涉及不足，均以传统或基础课程为主，虽然有开设数据结构、高级程序语言设计、管理信息系统等课程，但缺乏大数据环境下数据挖掘与分析所需要的技术学习，例如云计算、Hadoop 分布式系统架构、MapReduce 分布式计算框架、HDFS 分布式文件系统、机器学习、可视化工具等，而且相较于信息技术的更新，医学信息教育课程的更新较为滞后[7]。

（2） 数据获取不易

生物医学数据的特征表现为海量、多态性、变化性、非结构性，医学信息学学生应该在学习的过程中接触到一线的生物医疗学数据，通过使用所学的技术对其进行挖掘、分析，提升学习效率与质量，分析结果也能为医疗机构的决策起到辅助作用。而当前生物医学数据存在数据孤岛、数据完备性缺乏、数据共享难等问题。医学信息教育中很多实践用到的数据为部分样本数据或抽样数据，学生甚至无法接触到海量的一手生物医学数据，导致了学生不能很好的掌握所学的信息技术，甚至造成导致实践与理论脱节。

（3） 信息技术重视程度不够

许多教师在授课过程中仅仅只是将信息技术作为一种辅助手段，并未体现其核心地位，这也导致了学生对于技术信息学习不够重视。同时某些教师在信息技术的教学使用中片面地强调、追求某项技术的先进性，有的时候并未选择真正适合本学科特点的信息技术手段。学生在学习过程中，也不能很好地选择适合自己研究领域的信息技术手段。

3.2 医学信息教育智慧云平台的作用

不同于其它学科领域的智慧教育云平台，医学信息教育智慧云平台加入了生物医学信息数据接口，确保学生在学习和实践中能够获得一手的生物信息数据、医疗数据等，医学信息智慧教育云平台的构建能够有效改善当前医学信息教育所存在的问题。

（1） 解决课程设置滞后问题

医学信息教育智慧云平台能够整合全网所有教学资源，提供智慧教学环境与智慧教学，通过分析学生的学习行为、学习进度等数据，可自动生成个性化课表，提高学生的学习效率。同时平台还提供先进信息技术课程的供学生检索学习或推送学习，满足学生学习热点信息技术的学习需求。

（2） 解决数据获取问题

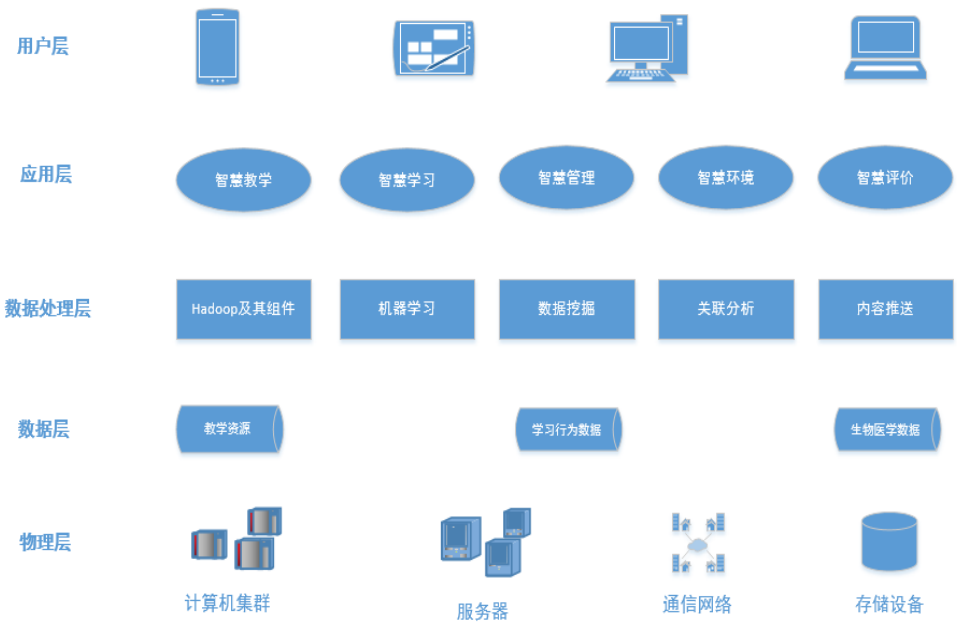
通过该平台的数据接口，教师以及学生能够获得生物医疗机构的第一手数据以供教学、交流之用，能够极大提升学生对于海量生物医学数据的处理能力以及对于信息技术的掌握。

（3） 满足多种互动需求

针对医学信息学这类实践性强的学科，医学信息教育智慧平台能够有效地加强师生互动、生生互动，提高学习的效率。

4 医学信息教育智慧云平台的总体架构

为了实现医学信息教育智慧云平台的各部分功能，本文设计了了医学信息教育智慧云平台的总体架构，平台分为 5 层，如下图所示。



（1） 物理层：物理层主要包括计算机集群、服务器、海量数据的存储设备、网络设备等硬件设施，是医学信息智慧教育云平台运行的基础。

（2） 数据层：数据层主要由三大数据库构成，第一是教学资源库，整合了全网所有与医学信息教育相关的教学资源；其次是生物医学数据库层，里面存储的海量数据可供学生学习、实践、科研使用；最后是学习行为数据层，记录着学生相关的所有学习行为数据。

（3） 数据处理层：该层采取 Hadoop 及其相关组件对于数据层的数据进行计算、分析、调度、管理。在数据分析挖掘方面部署机器学习、统计分析、关联分析、内容推送、信息可视化等模块。该部分是智慧教育云平台的核心。

（4） 应用层：该层主要有智慧教学、智慧学习、智慧管理、智慧服务、智慧环境和智慧评价等部分组成，主要为教师、学生、家长提供与应用服务，通过利用现有的智慧资源，为用户提供个性化、多样化、全方位的智慧服务。

（5） 用户层：为应用层提供各类接口，支持手机、PC 等多种类型的终端设备接入访问[5],[8]。

5 医学信息教育智慧云平台的关键技术

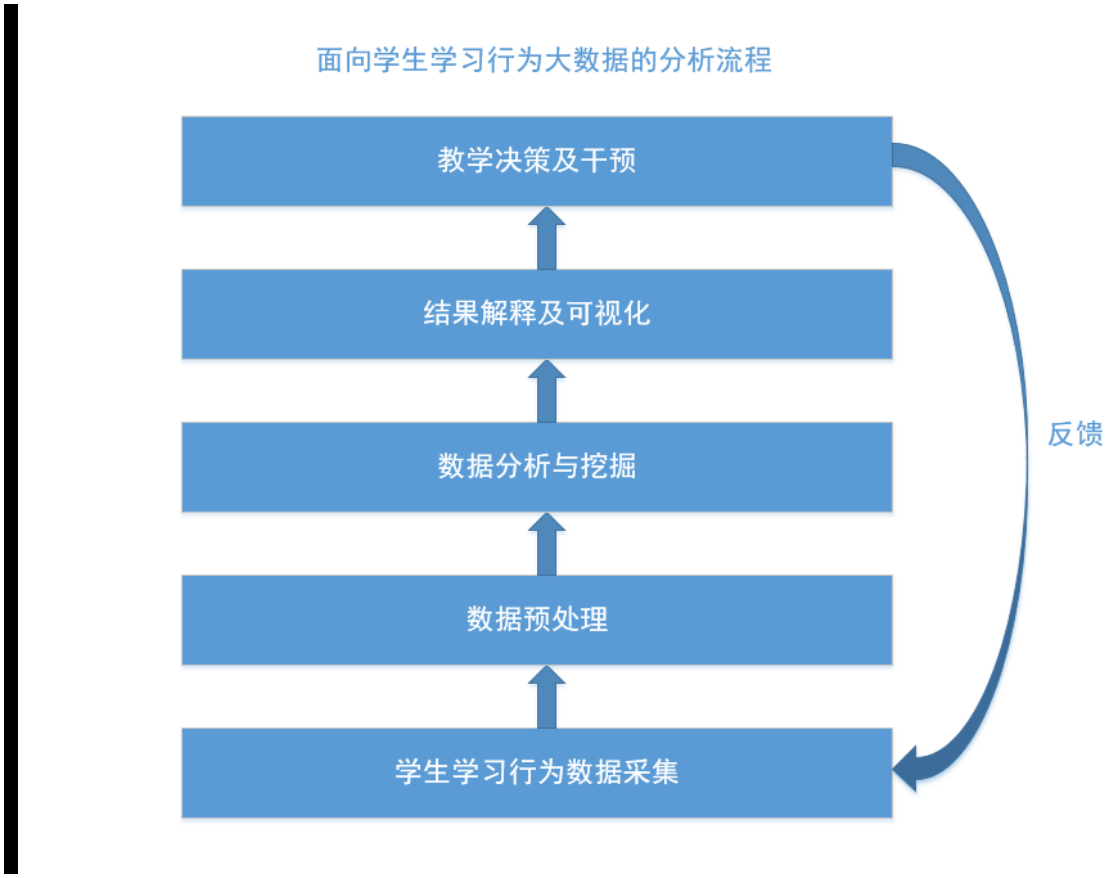
医学信息教育智慧云平台的构建需要大量的信息技术作为支撑，而在这些技术中，面向学生学习行为的分析技术以及 Hadoop 分布式计算技术则是起到了关键的作用。

5.1 面向学生学习行为的大数据分析技术

学习分析技术（Learning Analytics）起源商业零售领域，商家通过对消费者的一定时期的购买行为进行统计分析，预测客户的消费趋势。在医学信息智慧教育云平台中，是对学生日常学习和生活过程中产生的海量数据进行统计和分析，评估学生的学业进展和兴趣爱好，预测未来的发展趋势，并发现潜在问题[1]。

学习者通过医学信息教育智慧云平台进行学习的同时，会产生大量的学习行为数据，通过分析这些数据，可以得出学习者的学习习惯、学习兴趣、学习特点等隐藏的特性，从而把握学习者的学习模式，评价学习者的知识掌握，从理论上实现因材施教。教师可通过对学生学习行为数据的分析来比较不同的教学方法，分析教学的整体效果进而发现学习的薄弱环节，同时可以提出关键的指标体系，建立统计和预测模型，并帮助管理者提出更加适合学生的教育管理决策[3]。

通常而言，学生学习行为分析的流程如下所示：



（1） 采集学生学习行为数据

系统记录学生登录、退出、浏览课程、观看视频以及完成习题和测试、课堂互动、课后交流等行为数据，这一系列数据都可以反映出学生在学习过程中的具体情况。

（2） 数据集成

对收集到的数据进行预处理，包括数据转换、清洗，筛选出有潜在价值的数据。

（3）数据分析与挖掘

该部分为学生学习行为分析的核心部分，主要采用统计分析、数据挖掘、机器学习等技术，分析学生的不良学习特征、分析和比较不同院校学生的学习保持率、学习进步和完成情况，然后对学生特征进行建模并对结果进行预测分析，这是学生学习情况的结果分析和教学改进意见的主要依据，也是实现智慧教育的重要技术支撑。

（4）结果解释及可视化

根据分析和预测的结果，对结果数据的数值大小或意义进行解释，对偏高或偏低的数值给出分析结果，找出造成差异的原因，作为进行教学及管理的改进的依据。

（5）教学决策及干预

完成学生数据分析报告，使用可视化技术把解释结果形成图标或文字呈现给教学管理者，教学管理者根据分析报告采纳相应的教学策略，对学习不良的学生进行指导干预，促进学生发展[3]。

在学生学习行为数据分析的整个流程中，数据分析与挖掘是核心模块，因此选择合适的机器学习算法以及统计模型对于实现智慧教育至关重要。

5.2 Hadoop 分布式计算框架

Hadoop 是 Apache 开源组织的一个分布式计算框架，可以在大量廉价的设备组成的集群上运行应用程序，构建一个具有高可靠性和良好扩展性的并行分布式系统。Hadoop 的主要优点有：扩容能力强、成本低廉、效率高、可靠性高、免费开源以及良好的可移植性[9]。

整体来说，Hadoop 是具有高性能技术优势的，它可以以 PB 以上的数量级能力处理结构化和非结构化数据，为用户提供智能化、可拓展新的集群规模数据内容，并不断强化数据能力。所以基于 Hadoop 的医学信息教育智慧云平台在超大规模数据存储、运算与管理方面都体现出了惊人的优势，因此它可被广泛应用与教育领域中[10]。

它具备非常优秀的扩展性，可以处理分布在数以千计的低成本 X86 服务器计算节点中的大型数据，具备高容量低成本的特点。Hadoop 同时具备极强的处理混合数据类型的能力，可以管理结构化数据、服务器日志文件和 Web 点击流的数据。同时还可以管理以非结构化文本为中心的数据，如 Facebook 和 Twitter[11]。

在 Hadoop 的开发过程中，相关技术人员能够在不清楚分布式底层细节的前提下，设计分布式程序，通过集群的高性能的优势快速运算和存储大数据[12]。

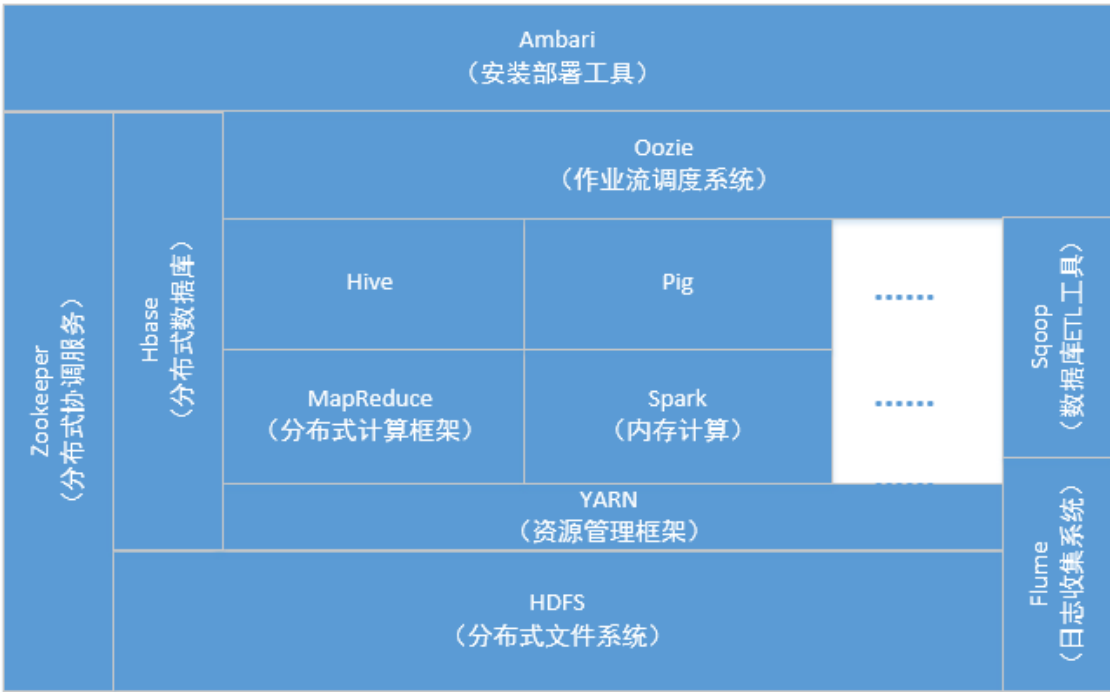
Hadoop 分布式文件系统 HDFS、MapReduce 编程模型和 HBase 分布式数据库是其三大核心技术。

HDFS 是分布式计算的存储基础，可以部署在廉价的硬件设备上，用来寻出海量数据集，因为每个数据集都进行了备份，所以它具有高容错性，并且提供了对数据读写的高吞吐率。是一个高度容错性的系统，适合部署在廉价的机器上，能够提供高吞吐量的数据访问，非常适合在大规模数据集上的应用。为了可靠的海量存储文件，各个文件以快序列的形式存储。为了保证故障容错，文件的块被复制，块的大小和复本的个数都可以配置。

MapReduce 编程就是“任务的分解与结果的汇总”，Map 将任务分解成多个子任务，Reduce 将分解后的任务处理结果汇总起来，得到最终结果。

HBase：HBase 是一个功能强大的分布式数据存储系统，基于列存储记录[13]。

Hadoop 的生态系统如下图所示：



Hadoop 生态系统

结语

综上所述，随着医学信息迈入大数据时代，医学信息教育必须有所创新，在医学信息教育中搭建一个智慧教育云平台，能够更好地跟紧信息技术发展的趋势，及时更新医学信息课程体系；同时能够更好地实现个性化、差异化的学习，让更多的医学信息专业的学生通过这各平台学习到更多的知识，并且能够做到理论联系实际，充分发挥所学知识在生物医学信息大数据分析、挖掘中的作用。伴随着信息处理技术以及教育理念的不断发展，智慧教育云平台将具备更加全面、强大的功能，智慧教育模式将推广到所有学科领域。与此同时，医疗机构通过共享医学信息数据，也是加强自身医学信息大数据平台建设的良好契机。

然而智慧教育云平台是一个相当复杂的系统，构建过程中包括一系列功能模块以及技术支撑，本文对于信息教育智慧云平台的搭建技术、智慧教育的管理模式和资源库的设计方法等没有做过多的探讨，尤其是医疗机构的医学信息数据库的共享问题，这也是将来需要进一步研究的内容。

OR-028

基于 WoS 的医学信息学领域研究现状分析

孙晶晶[1]、张帆[2,3]、张军亮[1]、李梦蕾[1]
1. 新乡医学院 2. 新乡医学院卫生信息资源研究中心
3. 新乡医学院教学质量监控与医学教育研究中心

目的 以 ISI Web of Science 数据库收录的以医学信息学为主题的期刊论文为样本，利用 CiteSpace 软件构建 2010-2019 年医学信息学研究领域的知识图谱，在对医学信息学背景知识介绍的基础上，对年度发文量，国家和机构，作者合作，文献共被引和关键词共线方面进行可视化分析，揭示其发展现状和研究热点，为医学信息学的进一步深入研究提供参考。

1 引言

我国医疗卫生领域的信息化建设是从 20 世纪 70 年代末 80 年代初开始，至今已有近四十年的历史。1993 年以来，国际信息化建设的不断发展，党和政府也对我国信息化提出了更高的要求，推动了各个行业的信息化建设。1995 年，卫生部根据党中央关于加强国民经济信息化的统一部署，适时地提出了卫生系统建设的目标和任务[1]。“九五”规划以来，我国医学信息化建设取得了明显进展，医学信息学的研究不断深入[2]。医学信息学作为一门交叉学科，总的来说就是将计算机科学的理论和方法等应用到医学领域，提高医学信息的有序化，医疗工作的高效化[3]。从整个发展历程来看，计算机科学的发展极大的推动了医学信息学研究的发展，同样医学领域的革新也能够引发科学技术的进步。为了深入开展研究工作，需要了解该领域先进国家的研究方向和内容，对学科的发展进行梳理。

因此本文基于 Web of Science 数据库中的 2010-2019 年收录的医学信息学为主题的论文进行可视化分析，来确定当前国内外的研究热点与前沿。

2 数据来源与方法

2.1 数据来源

本文数据来源 Web of Science（WoS）数据库，检索式为主题=“Medical Informatics”，文献类型=“article”or“review”，时间跨度为 2010-2019 年，精炼后共得到 1400 条记录。选择全纪录与引用的参考文献，以文本的格式下载。

2.2 研究方法

文献计量分析，主要是以文献之间各方面的属性为基础，通过数学和统计学的方式，预测研究领域的发展规律[4]。这种方法的优点有：（1）通过文献计量软件对文献进行分析；（2）按照文献各方面的属性分类，如国家和机构、作者、关键词等对文献进行分析；（3）借助 CiteSpace 等软件进行可视化分析，并将分析结果以表格和知识图谱的形式进行展示[5]。

社会网络分析，主要是对社会关系结构及属性进行研究的一种方法，主要分析不同社会单元所构成关系的结构及属性[6]。本文通过绘制机构和作者的可视化图谱来发现研究领域的整体结构和合作情况。

2.3 分析工具

本文使用 CiteSpace 软件系统，它是由美国华人学者陈超美博士开发的一款信息分析可视化工具，能够通过分时动态的知识图谱直观地展示研究领域的信息全景，识别当前研究中的关键文献、研究热点和前沿方向[7]。因其简单易用的操作方式和丰富美观的可视化效果，在国内外信息科学领域中得到广泛应用。

3 结果与分析

3.1 发文量分析

将文献根据年份进行整理，统计出医学信息学的年度发文量，折线图能够直观地看出研究领域变化趋势，对评价该领域的发展阶段和动态变化具有重要的意义。如图 1 所示，医学信息学的增长趋势近似直线（$Y=6.4X+104.8, R^2=0.5574$），表明医学信息学的发展呈波动上升趋势，研究领域不断扩展，未来会受到国内外越来越多的学者关注。

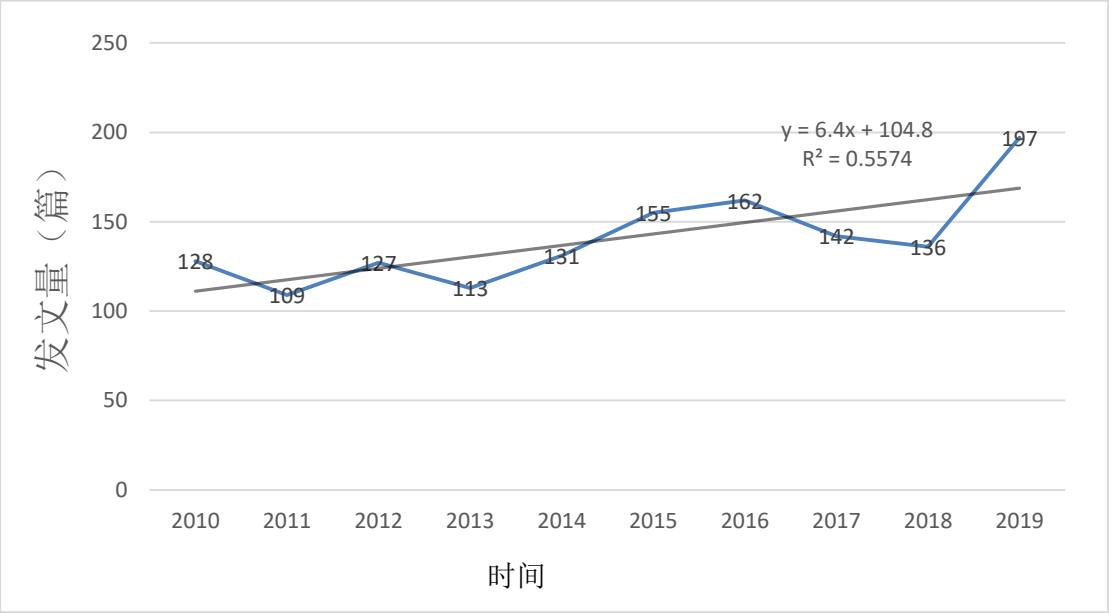


图 1 医学信息学发文量增长趋势

3.2 国家与机构分析

国家的分布情况可以反映出各个国家对该领域的关注程度，将数据导入 CiteSpace 中，在操作界面中设置相关参数：时区分割（Time Slicing）设置为 2010-2019 年，节点类型设置为 Country，其余设为默认选项。运行 CiteSpace 得到国家合作网络知识图谱，如图 2 所示有 81 个节点，452 条连线。结合表 1 我们可以看出，在医学信息学领域，美国的发文量为 635 篇排名第一，中心度为 0.32，说明美国在该领域的研究已经走在最前端，掌握丰富的研究资源，其他国家为了得到最新的信息纷纷和美国合作，美国在医学信息学领域中已经形成一个合作网络。其次是澳大

利亚和法国，发文量分别为101篇和48篇，中心度为0.14和0.15，从中可以看出，这两个国家也已经形成了自己的小合作网络，法国虽然发文量较少，但是中心度排名第二，也可以从侧面反映出法国的影响力较高。我国的发文量48篇，中心度只有0.01，说明了我们国家在医学信息学领域中仍然有很大的上升空间，需要我国学者不断研究和探索，学习美国、澳大利亚等先进国家这一领域的研究成果，提升我国的医疗卫生信息化建设水平。

通过CiteSpace软件，选择合适的参数，将节点类型设置为Institution，得到机构合作网络知识图谱，如图3所示有335个节点，节点之间的连线有934条，此次对机构研究，发现发文量多的主要是高等院校和医院，说明这两种机构是推动医学信息学研究的重要力量。发文量较多的机构有范德堡大学、哥伦比亚大学、犹他大学、哈佛大学、布列根和妇女医院和汉诺威医学院等，其中大多数来自美国，这与对国家研究的结果一致，美国在医学信息学领域中发展较快，占据绝对优势。

表1 各国医学信息学发文量

国家	发文量/篇	中心度
USA	635	0.32
GERMANY	116	0.11
AUSTRALIA	101	0.14
ENGLAND	96	0.13
CANADA	88	0.06
SPAIN	64	0.08
FRANCE	48	0.15
PEOPLES R CHINA	48	0.01
BRAZIL	43	0.04

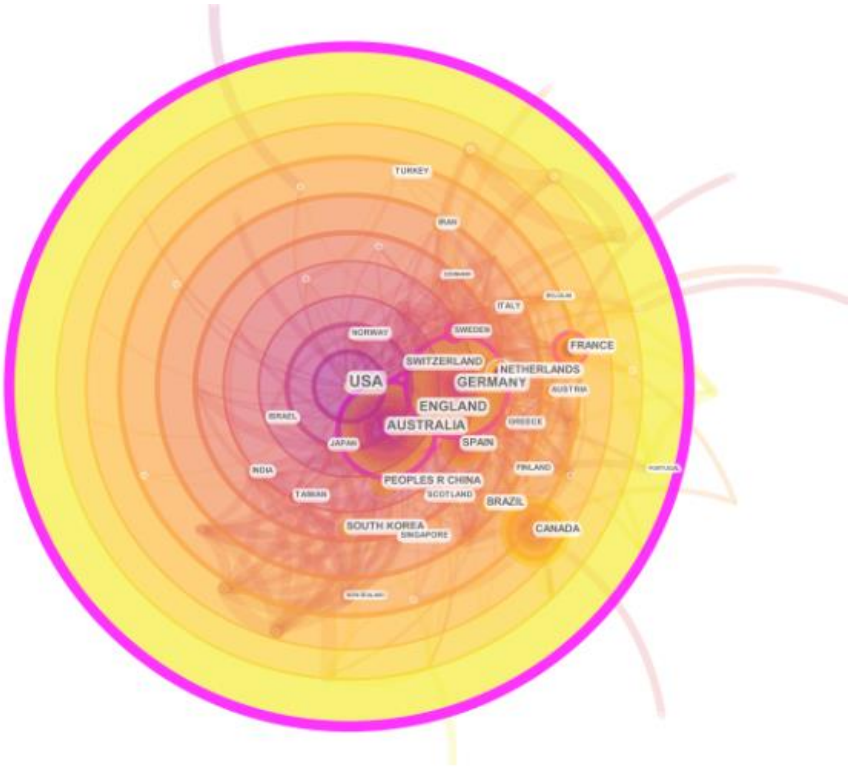


图 2 国家合作网络知识图谱

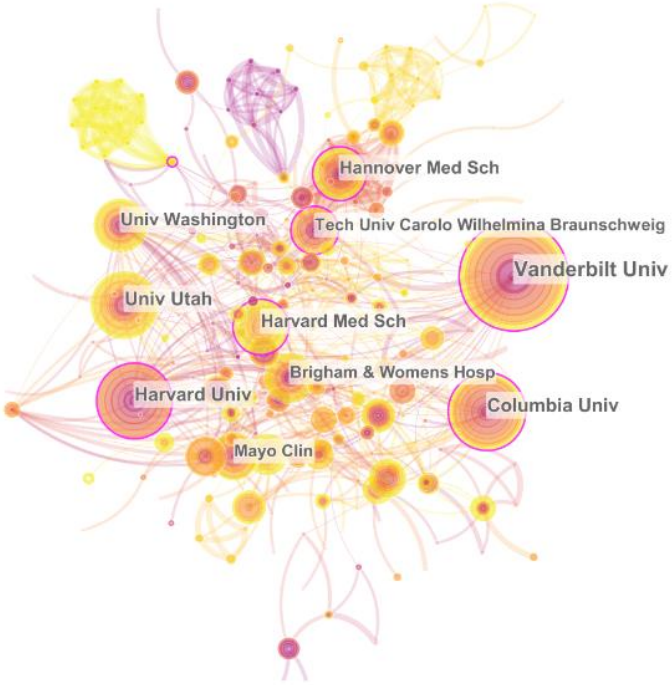


图 3 机构合作网络知识图谱

3.3 作者合作分析

通过 CiteSpace 软件，选择合适的参数，将节点类型设置为 Author，运行后得到作者合作网络知识图谱，如图 4 所示共有 404 个节点，代表不同的作者，节点之间有 664 条连线。最高产作

者是德国汉诺威医学院的 Reinhold Haux [8]共发表论文 11 篇，作者是医学信息学协会（IMIA）的主席，对医学信息学在全球视角下的目标和任务进行了研究,指出了医学信息学作为一个学科，应该致力于三个目标：更好的个体医疗和保健；更好的医疗和保健知识；更好的医疗和保健管理。其次是 R Haux[9]共发表论文 9 篇，作者主要从医学信息学的教育展开研究，指出了医学信息学课程的目标、特点以及未来的发展方向和面临的挑战。美国哥伦比亚大学医疗信息学的副教授翁春华，共发表论文 5 篇，作者主要利用大型观察性健康医疗数据网络对全球性的健康医疗问题进行研究，加强生物医学信息学领域的国际合作和教育培训工作，根据目前国内在医学研究上遇到的问题和挑战，利用数据科学和信息学的方法，促进国内的健康医疗数据研究工作。节点之间的连线表示作者之间的合作关系，从图中可以看出作者在医学信息学研究之间的交流存在一定的问题，研究力量分散，没有形成大范围的相互联系的合作网络。

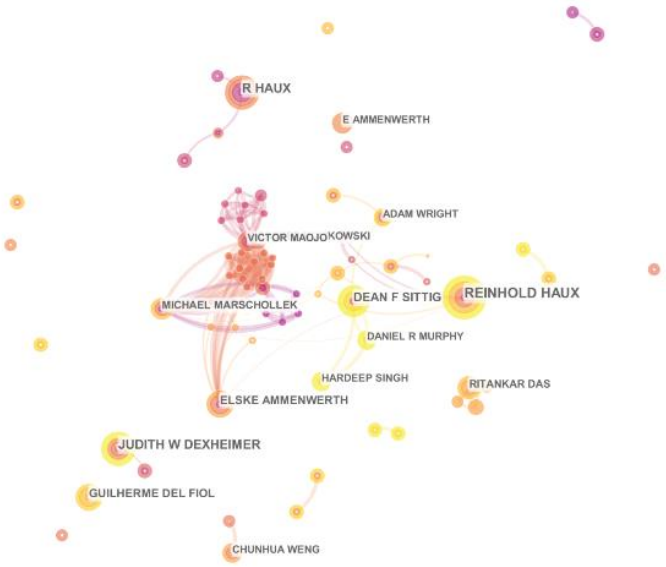


图 4 作者合作网络知识图谱

3.4 文献共被引分析

通过 CiteSpace 软件，选择合适的参数，将节点类型设置为 Cited Reference，运行后得到文献共被引网络知识图谱。如图 5 所示，共有 533 个节点，节点之间有 1887 条连线。对某一研究领域文献的被引频次进行分析，能够了解该领域内研究的高水平的论文，为以后的研究者提供有针对性的指导。按照节点在图中的大小，最突出的是由美国学者 Chaudhry B 等人[10] 于 2006 年发表的文章《系统回顾：卫生信息技术对医疗服务质量、效率和成本的影响》，其被引用频次为 36 次，文中主要介绍了卫生信息技术在提升医疗服务方面的功效，通过描述性比较研究和系统综述方法，发现决策支持系统或电子病历系统等能够有效地提高服务质量和效率。Mantas J 等人[11]于 2010 年发表的文章《国际医学信息学协会（IMIA）关于生物医学和卫生信息学教育的建议》，其被引用频次为 23 次，文中主要介绍了国际医学信息协会同意对现有的生物医学和卫生信息学建议进行修

订，这将有助于建立该领域的课程轨道，完善课程体系，并支持有关生物医学和卫生信息学教育的国际倡议。美国学者 Blumenthal D 等人[12]于 2010 年发表的文章《电子病历的“有意义使用”规定》，其被引用频次为 21 次，文中主要介绍了美国政府出台《医疗信息技术经济与临床健康法案》（HITECH ACT）来促进电子病历在医院的采用率，同时美国健康和人类服务部（DHHS）投资 259 亿美元对那些“有意义使用”电子病历的医院和医生给予财政补贴。

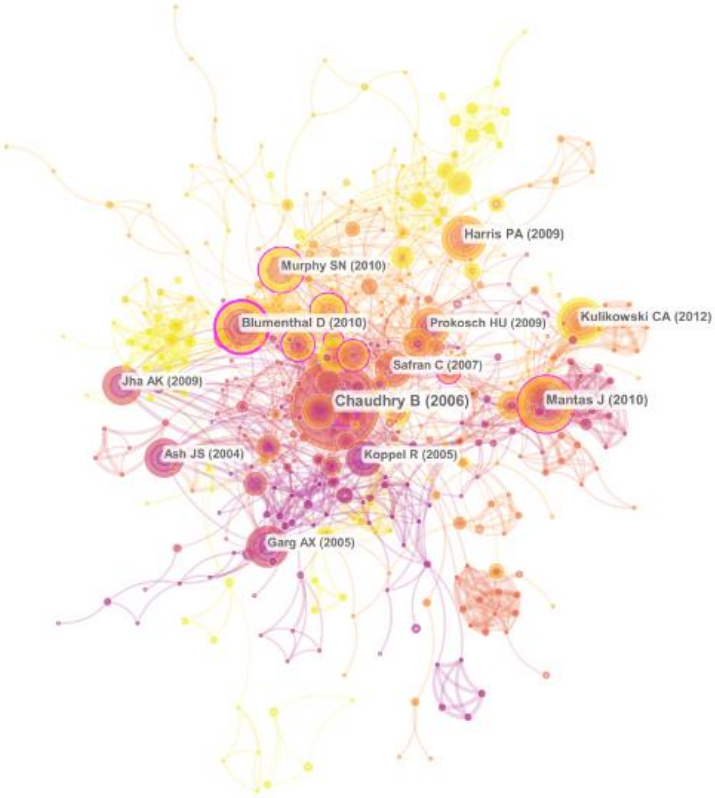


图 5 文献共被引网络知识图谱

3.5 关键词共线知识图谱

通过 CiteSpace 软件，选择合适的参数，将节点类型设置为 keyword，运行后得到关键词共线网络知识图谱。关键词是整篇文章的核心和精髓，通过对出现频次高的关键词进行分析，能够了解该领域的研究热点[13]。图谱中每个节点代表一个关键词，节点的大小表示关键词出现频次的高低，节点之间连线的粗细代表关键词之间共线的高低[14]。如图 6 所示，共有 539 个节点，节点之间有 4098 条连线。将不同写法和相同含义的关键词进行合并，可以发现出现频次高的关键词为医学信息学，电子健康档案、护理、系统、医疗保健、信息技术、质量、影响、模型、临床决策支持和数据等关键词。结合医学信息学相关知识可以将这些研究热点分为以下几类：

3.5.1 电子健康档案

又称电子健康记录，它是个人在健康相关活动中形成的数字化记录，能够存储在计算机中，通过网络得到共享的医疗健康相关的信息记录[15]。20 世纪 90 年代中后期，人们对电子健康档案的研究开始不断深入[16]。通过采集个人的健康数据，建立电子健康档案，研究者可以借助计算机对健康状况进行分析、评估和预测，推动国家医疗卫生能力和医疗保健水平的提高。另外集成电子

病历可以弥补临床传统意义上的决策支持系统的缺陷，能够提高决策支持功能，加强医疗机构的服务水平[17]。我国电子健康档案的研究起步比较晚，是伴随着卫生信息系统和信息技术的发展而开始的，对其研究还缺乏深入化和体系化的探讨，需要加大重视和参与[18]。

3.5.2 临床决策支持系统

它是一种用来改善临床决策制定的信息系统，集大数据和人工智能技术于一体，包括临床诊断决策和治疗方案决策两个功能，以历史病历和循证医学知识库中的信息为基础，通过输入病人数据和知识库进行对比，帮助临床医务工作者识别疾病，获取最佳治疗方案[19-20]。20 世纪 50 年代末，出现了最早的有档案记录的临床决策支持系统[21]，经过多年的发展，系统已经从基于规则和预先设定的流程的方法，发展成为基于融合数据为基础[22]，可以应用于各种医疗需求，具有医疗保健服务、慢性病预防和药物不良反应提醒、疾病诊疗等功能，极大地提高了医疗服务的效率。

3.5.3 数据的隐私与安全

随着现代科技的发展，数据的规模急剧增加，随着而来的也有许多问题和挑战，医疗数据包含患者的个人信息，如何更好地保护数据的安全和患者的隐私成为当前研究的热点问题。孔桂兰等人[23]为了使大量的医学用于研究的目的，建议使用对称分组密码、非对称密码和加密哈希算法，通过复合隐私保护的方法能够使临床数据仓库足够安全，来保护患者的隐私防止数据泄露。刘西蒙等人[24]提出一种新的以病人为中心的临床决策支持系统，来帮助临床医生以隐私保护的方式辅助诊断，可以有效地计算出病人的疾病风险。德国学者 Sebastian Haas 等人[25]提出了一个隐私保护信息系统的整体方案，包括数据服务和患者服务模型，用来控制向他人透露病人数据。

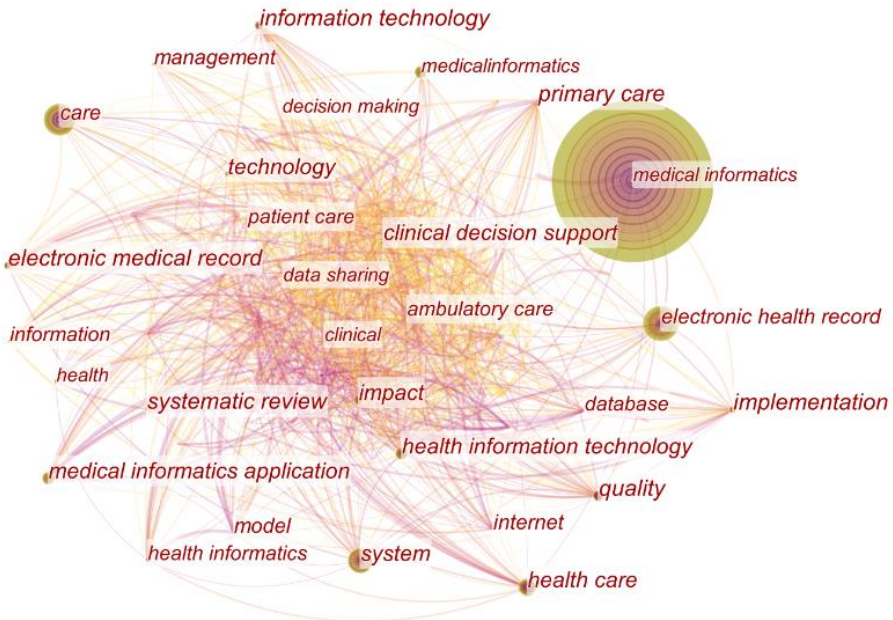


图 6 关键词共线网络知识图谱

4 结论

本文以 web of science 数据库为数据源，检索近十年主题词为 “Medical Informatics”的文献资源，通过 CiteSpace 可视化分析工具从多角度对其进行分析，得出了以下结论：第一从文献发文量来看，在所选时间段中医学信息学领域中的文献呈波动上升趋势，越来越受到学者的关注；第二从国家和机构来看，目前美国对医学信息学的研究关注度较高，形成了以自身为主的合作网络，并且拥有许多高水平的研究机构，占据了这一研究领域的绝对优势。其他国家如法国、澳大利亚利用自身的优势形成了自己的小合作网络，我国在该领域的研究还需要加强，进一步提升我国的医疗卫生水平。第三从作者合作来看，目前医学信息学研究领域合作力量分散，没有形成大范围相互联系的合作网络。第四从文献共被引来看，可以发现该领域内研究的高水平的论文，能为以后的研究者提供有针对性的指导。第五从关键词共线来看，目前医学信息学研究的热点集中在电子健康档案和临床决策支持系统等卫生信息技术的研究，大数据的发展对医学信息学的研究具有重要的作用，能够将信息技术和医疗健康深度融合，充分挖掘数据的医疗价值，推动医院的综合治疗水平。但随之而来的也有问题和挑战，医院信息系统的结构化水平低，各种数据整合和利用存在困难，数据的安全和隐私问题突出，我国对医学信息学研究还缺乏深入化和体系化，需要研究人员进一步探索。

OR-029

基于结构方程的河南省 5 家中医院门诊和住院患者满意度的比较分析

王晶晶[1]、吕晖[1,3]、邢晓琳[2]、任文杰[1]
1. 新乡医学院健康中原研究院 2. 新乡医学院公共卫生学院
3. 新乡医学院管理学院

目的 通过对门诊和住院患者进行问卷调查，找出影响患者满意度的关键因素，为提高医院卫生服务质量和医院更全面发展提供有力依据。

方法 采用现场调查，对 5 家中医院门诊和住院各 300 余名患者进行匿名调查；采用描述性分析、因子分析和结构方程来分析数据。

结果 通过因子分析和结构方程发现，影响住院患者满意度的前 2 个关键因素分别为病房环境和护士工作水平；而影响门诊患者满意度的前 2 个关键因素分别为医院环境和医生工作水平。

讨论 改善患者就诊环境，提高医务人员工作水平，尊重患者知情权和隐私权，进而不断改善患者就医体验。

随着国家医疗卫生体制改革的不断推进，“以患者为中心”的现代医疗服务理念逐渐深入人心[1]，为患者提供更优质的服务质量已成为各医院一直以来追求的目标。而患者满意度不仅是衡量医疗服务质量的重要指标，也是评价卫生资源分配有效性的途径之一[2]。因此，本研究通过选取河

南省内的五家中医院进行问卷调查，分析数据构建患者满意度模型从而找出对其有影响的关键因素。

1 研究对象与方法

1.1 研究对象与纳入标准

1.1.1 研究对象的选取

选取 5 家河南省中医院（分别为河南中医药大学第一附属医院、河南省中医院、河南中医药大学第三附属医院、河南省中医药研究院附属医院、河南省洛阳正骨医院）为调查点。采取随机抽样，2018 年 9 月在 5 家医院门诊部和住院部分别各选取 60 例患者。

1.1.2 纳入标准

门诊患者：（1）已经完成就诊；（2）意识清醒可进行完整交流；（3）愿意配合。

住院患者：（1）住院时间≥3 天；（2）意识清醒可进行完整交流（3）愿意配合。

1.2.研究方法

1.2.1 现场调查。由经过统一培训的调查员在规定的时间内到医院的门诊和住院部分别进行问卷调查，由口头询问形式进行；且当天完成和检查问卷。调查问卷是基于国家医管中心的满意度测评指标体系的基础上制定的；门诊和住院患者满意度调查问卷均分为两部分：（1）患者基本情况：均包括性别、年龄、文化程度及就诊支付方式；（2）满意度指标：住院和门诊患者满意度指标分别有 16 个和 13 个，且所有指标均采用 Likert 量表法分为 4 个等级进行评分。

1.2.2 数据统计分析。采用 epidata 双人进行录入，应用 SPSS24.0 统计软件包和 AMOS 软件进行数据分析。

2.结果

调查问卷共发放 600 份，门诊部和住院部分别回收 280 份和 266 份；其有效问卷数分别为 243 份和为 240 份；故有效回收率分别为 81.0%和 80.0%。

2.1 患者的基本情况

通过描述性分析可知，两组调查对象的年龄分布均较均匀；住院患者性别分布较均衡，但门诊患者女性所占比例较大（63.0%），两组各个文化程度水平的患者均有涉及；绝大部分（87.9%）住院患者均采用医疗保险来支付费用，而门诊则相反，仅有三分之一用医疗保险。详见表 1。

表 1 患者的基本情况

特征		住院患者		门诊患者	
		频数（n）	百分比（%）	频数（n）	百分比(%)
性别	男	116	48.3	90	37.0
	女	124	51.7	153	63.0
年龄	<30	28	11.6	59	24.2

	30-39	40	16.7	69	28.4
	40-49	44	18.3	43	17.7
	50-59	35	14.6	35	14.4
	>60	93	38.8	37	15.2
文化程度	初中及以下	87	36.3	46	18.9
	高中或中专	62	25.8	51	21.0
	本科或大专	86	35.8	133	54.7
	研究生及以上	5	2.1	13	5.3
就诊支付方式	个人自付	29	12.1	162	66.7
	新农合	58	24.2	16	6.6
	城镇医疗保险	134	55.8	53	21.8
	公费医疗	19	7.9	12	4.9

2.2 患者满意度的因子分析

2.2.1 适应性检验

结果显示，住院和门诊患者满意度指标的 KMO 值分别为 0.860 和 0.873，Bartlett 球形检验近似卡方值分别为 1798.590 和 1216.375，自由度分别为 120 和 78，P 值均小于 0.001；表示其均适合因子分析。

2.2.2 住院患者满意度指标的主成分归类

运用主成分分析法，旋转方法为凯撒正态化最大方差法，住院患者满意度指标旋转在 6 次迭代后已收敛；提取前 6 个因子作为主成分。第 1 个主因子即医生工作水平，包括医生工作态度、医生服务态度、医生沟通能力、医生诊疗水平；第 2 个主因子即护士工作水平，包括护士工作态度、护士服务态度、护士沟通能力、护士技术水平；第 3 个主因子即患者知情权，包括对所用药功能和副作用知晓度、对所用药名称知晓度；第 4 个主因子即病房环境，包括晚上病房是否安静、病房和卫生间清洁度；第 5 个主因子即辅助服务，包括费用清单清晰性、入院手续复杂性;第 6 个主因子即患者满意度，包括整体评价和推荐他人。详见表 2。

表 2 住院患者满意度指标旋转后的成分矩阵

特征	成分					
	1	2	3	4	5	6
医生工作态度	**0.872**	0.225	0.069	0.079	0.107	0.090
医生服务态度	**0.834**	0.285	0.100	0.063	0.159	0.080
医生沟通能力	**0.685**	0.429	-0.030	0.174	0.141	0.097

医生诊疗水平	**0.585**	-0.013	0.160	0.457	0.059	0.345
护士工作态度	0.294	**0.767**	0.149	0.193	0.150	0.099
护士服务态度	0.294	**0.751**	0.185	0.217	0.096	0.016
护士沟通能力	0.347	**0.660**	0.092	0.293	0.166	0.174
护士技术水平	0.446	**0.251**	0.258	0.508	0.124	0.082
对所用药功能和副作用知晓度	0.020	0.027	**0.876**	0.138	-0.009	0.109
对所用药名称知晓度	0.149	0.234	**0.836**	0.031	0.091	0.063
病房卫生间清洁度	0.188	0.268	0.138	**0.718**	0.248	-0.024
晚上病房是否安静	0.016	0.428	-0.033	**0.618**	0.119	0.311
入院手续复杂性	0.216	0.095	0.028	0.180	**0.828**	-0.021
费用清单清晰性	0.075	0.178	0.053	0.111	**0.798**	0.244
整体评价	0.144	-0.062	0.178	0.213	0.052	**0.805**
推荐他人	0.133	0.388	0.009	-0.055	0.183	**0.748**

2.2.3 门诊患者满意度指标的主成分归类

结果显示，第 1 个主因子即医生工作水平，包括医生工作态度、医生服务态度、医生沟通能力和医生诊疗水平；第 2 个主因子即医院环境，包括医院标识明确性、对医院服务设施印象和医院空间布局便利度；第 3 个主因子即辅助人员态度，包括医院挂号方便度、挂号及导诊人员服务态度、挂号及导诊人员工作态度；第 4 个主因子即患者满意度，包括整体评价和推荐他人；第 5 个主因子即患者隐私权，包括患者隐私权保护性。详见表 3。

表 3 旋转后的因子载荷矩阵

特征	主成分				
	1	2	3	4	5
医生服务态度	**0.818**	0.150	0.230	0.075	0.175
医生工作态度	**0.821**	0.129	0.197	0.065	0.077
医生沟通能力	**0.646**	0.224	0.086	0.135	0.094
医生诊疗水平	**0.530**	0.089	0.108	0.545	0.154
医院标识明确性	0.198	**0.744**	0.155	0.053	0.293
对医院服务设施印象	0.135	**0.764**	0.284	0.203	0.027
医院空间布局便利度	0.204	**0.804**	0.122	0.186	0.004
医院挂号方便度	0.124	0.215	**0.745**	0.207	-0.136
挂号及导诊人员服务态度	0.274	0.216	**0.758**	0.047	0.328
挂号及导诊人员工作态度	0.267	0.177	**0.717**	0.104	0.383
整体评价	0.035	0.319	0.271	**0.725**	-0.126

推荐他人	0.154	0.108	0.033	**0.791**	0.329
患者隐私权保护性	0.240	0.150	0.187	0.183	**0.784**

2.2.4 患者满意度结构方程模型定型

在因子分析的基础上，对数据进行验证因子分析，模型运算使用最大似然法，据此确定住院患者满意度的初始模型，接着对模型进行修正，从而确定最终模型，整体模型拟合优度均说明两个模型的拟合效果较好。详见表 4。

表 4 整体模型拟合优度评价标准

项目	绝对拟合度				简约拟合度		增值拟合度	
	GFI	AGFI	RMR	RMSEA	PGFI	NC 值	IFI	CFI
参考标准	>0.9	>0.8	<0.05	<0.08	>0.5	[1 3]	>0.9	>0.9
住院患者满意度最终模型	0.903	0.855	0.020	0.073	0.604	2.259	0.935	0.934
门诊患者满意度最终模型	0.954	0.922	0.014	0.040	0.566	1.394	0.982	0.982

2.2.5 患者满意度结构方程模型解释

运用 AMOS 软件分析，得出住院患者满意度结构方程模型（如图 1）。由模型可知病房环境、辅助服务、医生工作水平及患者知情权对患者满意度均有直接效应，其值分别为 0.251、0.213、0.180 及 0.046，而护士工作水平对患者满意度既有直接效应也有间接效应，其值分别为 0.063 和 0.153，故总效应为 0.216；病房环境与护士工作水平及医生工作水平之间均有关联，其相关系数分别为 0.843 和 0.576；而护士工作水平和医生工作水平也有关联，其相关系数为 0.723。

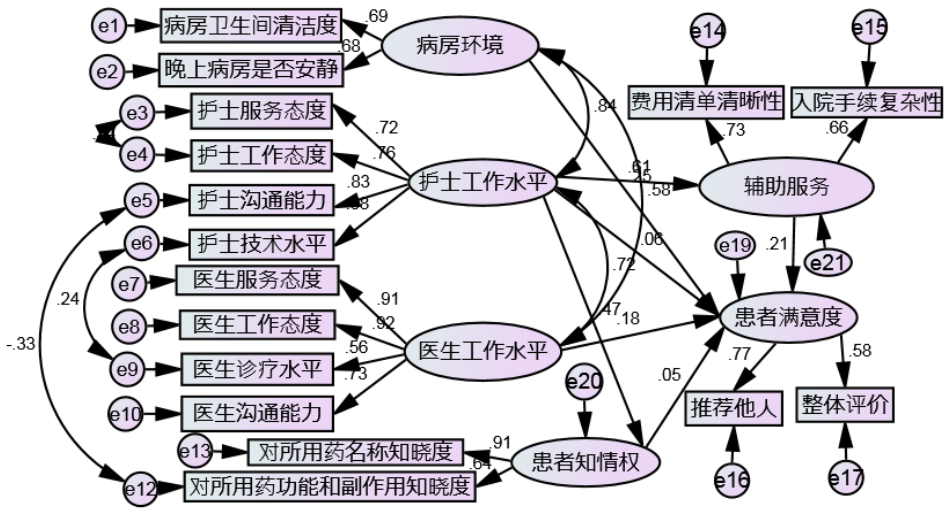


图 1 住院患者满意度结构方程模型

运用 AMOS 软件分析，得出门诊患者满意度结构方程模型（如图 2）。由模型可知医院环境和患者隐私权对患者满意度均有直接效应，其值分别为 0.434 和 0.064；辅助人员态度和医生工作水平对患者满意度既有直接效应也有间接效应，其总效应分别为 0.031 和 0.343；医院环境不仅与辅助人员态度相关联，也与医生工作水平相关联，其相关系数分别为 0.700 和 0.626；医生工作水平与辅助人员态度也相关联，其相关系数为 0.742。

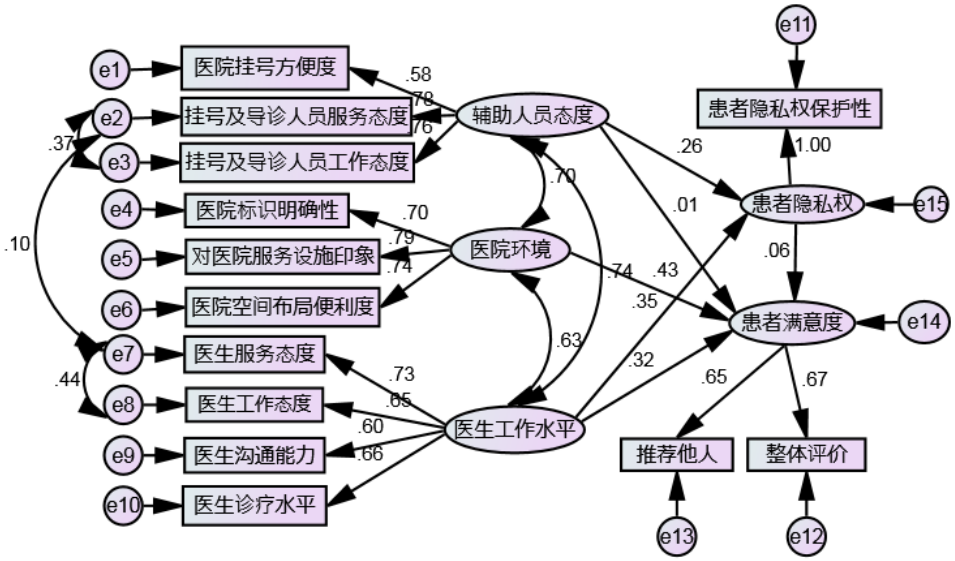


图 2 门诊患者满意度结构方程模型

3.讨论

3.1 综合分析，确定患者满意度模型

本次研究先通过因子分析，提取观测变量的公因子，即潜在变量；再通过构建结构方程去验证潜在变量之间的因果关系。从而不仅得到了影响患者满意度的因素之间的关联程度，还得到了这些因素对患者满意度的作用程度。

3.2 持续不断优化就医环境，不断提升患者就医体验

由结构模型可知，病房环境和医院环境分别对住院患者满意度和门诊患者满意度影响最大，且病房环境与医生工作水平和护士工作水平均有关联；另外，医院环境与辅助人员态度和医生工作水平均有关联。病房是患者在医院活动的主要区域，轻松舒适的病房环境不仅有助于医务人员身心健康，也可使患者更易配合医生治疗，有益病人恢复[4]。此外，医院环境既是患者与医院建立情感联接的第一媒介，其优劣状况与否也是衡量患者对医院满意与否的重要决定因素[5]。

高佳琦等[3]研究发现，就诊环境与患者满意度有显著相关关系；王镜茹等[6]研究发现，医院环境维度得分最低，提示应加强中医院硬件建设和环境改善。本研究正好也证实了此点。在衡量病房环境的观察变量中，病房卫生间清洁度的因子载荷量最大，这提示应加强对其的监管。此外，在衡量医院环境的观察变量中，对服务设施的印象因子载荷量最大；其次是医院空间布局便利度和医院标识明确性。良好的医院外部环境，不仅可以激发医生工作热情，缓解其紧张情绪，有助于其工作顺利开展，为患者提供优质高效的服务，这体现“以患者为中心”管理理念[7]，这些都提

示，应加强对医院服务设施的齐全性和便利性、医院空间布局的合理性及医院标识的明确性的建设，以提高患者满意度。

3.3 提升医生与护士工作水平，加强对患者隐私权的保护

从模型中可以看出，影响住院患者满意度的第二和第三关键因素分别是护士工作水平和辅助服务。此外，护士工作水平和医生工作水平之间相关。在衡量护士工作水平的观察变量中，护士沟通能力因子载荷量最大，其次是护士工作态度和护士服务态度，最后是护士技术水平。从中可以看出，如今患者观念的逐渐转变，越来越要求更加人性化的服务，这充分体现了“以人为本”的思想。护士通常和病人直接接触，且接触频率也很高；一旦沟通不好，会使患者对医院产生负面印象，甚至发生医疗纠纷[8]；Yellen 等的研究中也提到护理服务是决定住院期间满意度的重要因素[9]；调研中也发现患者普遍都对护士技术水平很满意。医生和护士同处在一个工作环境，对彼此之间会相互影响，尤其是医生对患者的服务态度会影响护士服务态度。

而在门诊患者满意度模型中，影响门诊患者满意度的第二和第三关键因素分别是医生工作水平和患者隐私权。而在住院患者模型中，医生工作水平是第四个关键因素，与此有所不同。此外，医生工作水平与辅助服务相关。医生的工作水平的高低是患者在就医时主要考虑的因素；在衡量医生工作水平的观察变量中，医生服务态度的因子负荷最大，这可看出患者的观念正在逐步改变，更加迫切希望在就医时受到尊重，享受到更加人性化的待遇。其次分别是医生诊疗水平和医生工作态度，最后是医生沟通能力。高佳琦等[3]发现，医生服务态度及医疗水平与患者满意度均呈正相关，这与本研究结果相一致。Junya 等[10]发现，医疗技术与患者满意度有很强正相关[9]；连慧莹等[11]也发现，患者就医时考虑最多就是医生技术水平。医生诊疗水平是其立身之本，故医院应定期对医生进行培训，以提高其诊疗水平。另外，医院应采取相应措施来提高医务人员工作态度、服务态度及医患沟通技能，以全面提升医生总体水平，增强患者满意度与获得感[12]。此外，王镜茹等[6]发现，患者隐私保护在其研究的评价门诊患者满意度指标中得分最高，而邵艳等[13]发现患者对隐私保护的满意度较低，高佳琦等[6]也发现，患者隐私与患者满意度有显著相关关系；虽然这些研究结果有相应的差异，但是这都说明患者隐私权是影响其满意度的重要因素，这都与本研究结果一致，故医院应加强对患者权的保护，如病人就诊时注意关门，做检查时注意拉帘子等。

3.4 尊重患者知情权，做好辅助人员工作

在住院和门诊患者满意度模型中，路径系数最小的分别是患者知情权和辅助人员态度。其中对药物的知晓度在衡量患者知情权重所占因子负荷最大，可见患者在治疗过程中，越来越关注与治疗自身疾病有关的信息，及越来越顾及到重视自身知情权。衡量辅助人员服务态度的观察变量中，挂号及导诊人员服务态度和工作态度的因子负荷较大；可见患者更希望在就医时得到全方位的尊重。这都体现了患者对医院有了更高需求，从以前往往仅重视其诊治水平转变到对其全方位都有了要求。

因此，为了中医院品牌的建设，应重视医院环境和病房环境的管理，使患者住之舒适，节约其就诊时间，从而使其有更好就医体验。此外，对医务人员进行全方面培训，尤其是医务人员的服务态度和沟通能力，使其得到全方位的提高，使患者得到更高效的就诊体验。

OR-030

在线健康信息质量评价工具探析

吕合凯、王卓鑫、孙伟伟、刘鹏、潘玮、陈兴智、王辅之
蚌埠医学院

目的 在线健康信息已经成为公众获取健康知识的主要途径。本文通过检索国内外有关在线健康信息质量评价的相关研究文献，介绍了五种常用的国外在线健康信息质量评价工具及其在国内外的应用现状。并分析了我国在线健康信息质量评价研究面临的问题，并探讨未来中文在线健康信息质量评价的发展趋势。

1 引言

新媒体环境下，在线健康信息传播研究的首要问题是信息质量评价。关于网络健康信息评价主要涉及两个问题，一是健康信息网站质量评价方法研究，二是网络健康信息质量评价实证研究。国内外学者和相关机构就网络信息质量的评价问题开展了大量研究。

国外较早开展了网络健康信息质量评价体系的研究。国外比较常用的健康信息质量评价工具包括：DISCERN、NetScoring、HONcode、URAC 和 OMNI 等[1]。主要的评价维度包括：信息内容（权威、归因、可信性内容、互补、数量属性、合理性、可靠性、治疗选择）；网站结构（保密、隐私、安全性、道德透明、超链接、涉及、交互性、可获取性、网站声明、财务声明、广告）；网站服务（政策、服务提供、组织）[2-6]。在评价机制方面，主要依据用户评价（DISCERN NetScoring）；第三方认证评价（HONcode、URAC）；学科信息门户评价（OMNI）[1]。其他的医疗健康网站评价体系还包括：美国医学联合会为管理下属医疗卫生机构网站而制定的工作准则[7]；健康信息科技协会（The Health Information Technology Institute, HITI）制定的“Mitretek 网络健康信息评价系统”[8]和美国卫生信息技术研究所发布的“网络医疗信息质量评价标准”[9]等。

国内对健康信息网站的评价研究较少，北京大学医学图书馆的一些学者，参考和借鉴国外医疗健康信息资源评价研究工作，结合层次分析法和专家征询法，提出了一套面向中文医疗健康信息网站的评价体系。该体系包括目标层（健康信息网站评价）、准则层（内容、易用性、其他）和方案层（权威性、相关性、准确性、及时性、全面性、友好性、交互性、无障碍性、链接的有效性和速度、链接准确性）等内容。依据层次分析法给出评价指标权重，并基于该指标体系对我国常见的 16 家健康网站进行实证评测[10]。

2 国外在线健康信息质量评价工具

2.1 HONcode

HONcode 是瑞士健康在线基金会(健康在线基金会，一个成立于 1995 年的国际性非官方非营利性组织) 制定的，适用于医学和健康网站的行为准则医疗和健康网站的准则，HONcode 解决了互联网的主要卫生保健问题之一：信息的可靠性和可信度。HONcode 为网站开发人员提供一组基本的道德参考标准，以确保读者总是知道他们正在阅读的数据的来源和目的。HONcode 是被医学健康网站的发布者所最普遍采用的参考标准。目前，全世界有 102 个国家的 7300 多个认证合格的网站使用 HONcode。HONcode 评价网络健康信息的 8 大准则分别为：权威性、目的性、保密性、可信性、合理性、网站联系方式、公开资金来源和广告政策。作为对医学健康网站进行道德规范的先驱者，HON 针对医药健康网站的管理者制定了网站应遵循的原则 HONcode，用以确保网络上健康医疗信息的可靠性与准确性；申请 HONcode 认证的医药健康网站必须要符合其中的条件，自律性加入与约束。HONcode 认证是动态的，网站在取得认证后，如被发现有违反规则的现象，将会被取消 HONcode 认证[11] ，HONcode 并不是一种法规，而是对网站管理人员导的网络道德规范，它从职业道德、医学良知等角度对网站进行引导和规范。欧美国家、特别是美国的医学健康网站获得 HONcode 认证比较多，这一方面说明欧美这些国家比较重视对网上医学健康信息进行道德规范，另一方面说明 HONcode 已被这些地区的国家广泛接受。在亚洲和其它地区的国家通过 HONcode 认证的网站比较少，这在一定程度上反映了这些国家、地区对网上医学健康道德规范的认识不足，同时也有可能是对 HONcode 的了解不多。

2.2 DISCERN

DISCERN 是由英国国家卫生服务体系(NHS) 资助，1999 年英国牛津大学医学研究所公众健康和初级卫生保健部研发，由英国国家医疗服务行政研究和发展计划资助[12]。DISCERN 起初是为评价印刷文献信息质量而开发的，但现在也被用于评价在线健康信息质量。该工具共有 16 个问题，涉及健康信息的可靠性，与治疗方案有关的信息质量，以及对健康信息的总体评价。评分采用 5 级量表(“1~5 级别”)。DISCERN 包括 15 项评估(其中 1~8 项评估信息的可信度，9 ~15 项评估信息的详尽度) 和一项总体评价，分值从 1 到 5 分，DISCERN 评分与内容评分存在着明显的联系，质量越好分值越高[13]。作为一项专门评价疾病治疗方案的工具，DISCERN 强调信息的可靠性，注重健康信息的内部特征且关注的层面更加具体，要求网站描述所有治疗的作用机制、疗效和风险，清楚地说明可能有的其他治疗选择，不治疗会发生什么，治疗选择对生活质量的影响等。

2.3 Michigan Checklist

密歇根大学公众健康网站评价工具(University of Michigan Consumer Health Web site Evaluation Checklist)，由密歇根大学于 1999 年开发。Michigan Checklist 的特点是兼具了对于健康信息发布平台和健康信息内容的评价，可用于评价健康网站本身的质量及其所发布信息的质量。Michigan Checklist 在网站受众、网站信息获取渠道及其优势比较、广告策略、网站导航

设计等方面的具体指标较多，共 11 项指标，包括权威性、时效性、信息内容、资源范围和筛选标准、受众、价值、准确性、广告、导航、速度和可访问性，设有 43 个 Yes /No 问题，根据不同权重赋分(1 ~3 分)[14]。

2.4 LIDA

英国卫生保健领域 Minervation 咨询公司于 2007 年开发。该工具与 Michigan Checklist 类似，也兼具了评价健康信息发布平台的质量，不同之处在于该工具通过增加两项附加条目，对于健康网站在信息内容制作和发布过程中的操作规范性也给出了评价标准。由 41 个问题构成，每个问题由 4 级量表评价(0~3，0 表示 Never，3 表示 Always)。问题 1~6 评价的是“可访问性”(Accessibility)，包括评价网站是否满足 World Wide WebConsortium 的标准；问题 7~24 评价的是“易用性”(Usability)，包括信息表达的清晰度、网站设计的一致性、是否设有有效的浏览和检索功能、是否有互动媒介(interactive media) 等问题；问题 25~41 评价的是“可靠性”，包括网站的更新频率、利益突出、内容编辑的方法以及内容的准确性[15]。

2.5 SAM

Suitability Assessment of Materials(SAM)，SAM 是 Doak 等于 1996 年研制，包括 6 个维度（内容、知识情况、图表、布局、学习动机、文化适应），22 个条目，具有较好的信度、效度，经国外广泛应用，已作为健康教育书面材料的评价和编写标准。SAM 的 22 个条目，每项采用“不适用”“一般适用”“非常适用”3 级评价，分别计 0、1、2 分，总分 44 分为最终得分：0~39%为不适用，40%~69%为一般适用，70%~100%为非常适用[16]。

综上所述，五种评价工具对于评价维度的设计各有侧重。五种评价工具的评测维度和定性评价标准见表 1。

表 1. 五种评价工具的对比

工具名称	评测维度	定性评价标准
HONcode	可靠性、可信度	无定性评价标准
DISCERN	可信度、详尽度、总体评价	1 分（低，存在严重的或大量不足）；3 分（中等，稍微重要或无严重不足）；5 分（高，无明显不足）
LIDA	可用性、可靠性	无定性评价标准
Michigan Checklist	内容、易用性	0 -25 分（差）；26-50 分（较差）；51-60 分（一般）；61-70 分（好）；71-80 分（优秀）
SAM	内容、知识情况、图表、布局、学习动机、文化适应	70-100 分（非常适用）；40-69 分（一般适用）；0-39 分（不适用）

3 国内在线健康信息质量评价工具

3.1 北京大学医学图书馆综合评价法

北京大学高琴在参考和借鉴国外健康信息质量评价研究的基础上，对我国医疗健康信息网站进行评价。从简单易行和方便操作考虑，参考各类标准体系，设置了内容评价指标、易用性评价指标和其他三大类指标对健康信息网站进行评价。评价过程主要是针对提出的评价指标，依据 AHP 思路，将所包含的指标进行分组，每组作为一个层次，按照高、中和低三个层次的形式排列，建立评价指标的层次结构模型，然后设计出问卷调查表，邀请信息专家给各指标赋予相对重要性的分值或权重，再用各个指标按照每个网站的具体情况赋值，最后计算得出各网站的总得分。较之完全定性化用文字表述的评价，这种方法能用数量化的方法多层次、多角度地反映网站信息的质量，其评价结果较客观、系统、全面，用户不仅可以直观了解各网站总体情况，还可以对各网站具体内容进行比较。高琴利用此方法对医疗健康信息网站进行评价得到结果为：16 个健康信息网站的总分范围在 42.704 -88.474。其中得分最高的为 39 健康网，级别为优良，接近优秀。最低的为东方健康网，级别为稍差，接近极差。高琴认为目前中文健康信息网站质量差异颇大。所有网站平均得分为 73.56，基本可以认为处于一般水平或接近良好。得分为优良或优秀的网站 5 个，占 31.25%；得分为稍差或极差的网站 2 个，占 12.5%；其余质量一般的网站占 56.25%。这些数据同样也说明了网上健康信息质量良莠不齐，用户需要谨慎利用网上医疗健康信息。如果以 60 分算及格，所评网站及格率为 87.5%，但其中没有一个达到优秀标准的，即得分在 90 以上[10]。

3.2 中文健康信息网站评价体系

唐小利等借鉴 HONcode、DISCERN 和 Eysenbach 的 TDRAC 指标分类与描述、MLA 开展的健康信息网站评价，以及近几年有关大学网站、门户资源网站和教育网站定量评价方法与指标体系，结合中文语言和网络环境的特点，通过定性和定量相结合的方法设计出了一个适合公众的健康信息网站评价体系框架。设立的一级指标共三个，包括信息内容、网站利用率与影响力、网站设计 3 个维度；设立的二级指标共 14 个，其中 12 个指标分别针对信息内容的权威性、合理性、归源性、及时性、可读性、补充性、保密性，以及网站的可访问性、互动性、透明性、功能设计、广告政策进行定性评价；定量评价采用网络信息计量学和网络链接分析方法，评价网络信息的利用率(访问量) 和影响力(链接量)；三级指标共 23 个。信息内容和网站设计维度中的原则性指标采用 Yes /No 的评分形式，程度性指标采用打分制，满分 5 分，由专家给出；网络利用率和影响力指标经计算得出；流量指标采用 Alexa 中文网站检索统计；链接指标采用 Google 高级检索进行统计。该体系与国外最常用的 2 种健康信息网站评价工具 HONcode、DISCERN 的兼容度分别为 100% 和 80%。采用该体系、Honcode、DISCERN 等同时进行评价，检验该体系与国外权威工具评价结果的一致性，最后形成一套中文健康信息网站质量评价工具，将其作为中国公众健康网健康知识筛选与评价依据，进而优化网站知识资源评价体系与采集流程，为中国公众健康网信息资源的自动建设奠定基础[17]。

4 在线健康信息质量实证研究概述

4.1 中文在线健康信息质量评价的实证研究

通过检索 CNKI，我们发现国内关于中文在线健康信息质量评价的实证研究主要起始于 2008 年，近年来呈显著上升的趋势。中文在线健康信息质量评价研究主要针对糖尿病、冠心病、肾病、注意缺陷多动障碍、儿童发热等基本的在线健康信息开展。国内的医学健康网站所通过的审核主要是我国的“信息服务管理办法”， “信息服务管理方法”作为国家的相关法规，主要是对网站的审批程序、发布范围、信息内容控制、相关处罚措施等方面进行要求和界定，各网站获审核的情况参差不齐，国际性的认证几乎没有。国内学者多采用医学和健康网站行为准则（HONcode）和网络健康信息评估工具 DISCERN 量表评估网站信息质量，诸多学者在健康网站评价的理论和实践上进行了长期、艰苦的工作，并在行为规范、资格认证、评价工具、信息过滤和评价性元数据等方面取得了一定的成绩。建立了较为完善的健康网站评价体系，刘艳丽以 Donabedian 的医疗保健质量评估模式为理论基础，构建评价网络用户健康信息质量的模型，然后根据该模型开发评价糖尿病用户健康信息网站的工具，并实施评价，最后对评价结果进行分析评价，结果发现我国糖尿病用户健康信息网站质量良莠不齐，对这些网站进行质量评价是必要的；我国用户健康信息网站建设不规范；网站的结构指标能从一定程度上反映网站的质量[18]。随着“互联网+医疗健康”的兴起，大众可以通过互联网方便快捷地获得自己所需的健康信息，但是基于公众自身的医学知识薄弱，网上信息杂乱和监管不到位等一系列问题使得一般用户难以区分在线健康信息的质量。钟乐等评估中文网站关于注意缺陷多动障碍的健康信息的质量，用搜索引擎查找关于注意缺陷多动障碍的中文健康信息，用健康信息评估工具 DISCERN 评分评估信息质量。将网站按所有人属性分组进行评估，发现非营利性组织的网站信息质量最好，教育机构和医疗机构网站信息质量最差，互联网上关于注意缺陷多动障碍的在线健康信息质量不佳。钟乐认为政府卫生部门、研究机构和公立医院等可信度高的机构应该重视网络健康信息对于患者的影响，为社会大众提供优质的健康信息[19]，目前，国内外很多临床医生、科学家、研究机构、医院或商业机构都在创办并维护健康信息网站。但国内健康信息发布与传播缺乏统一监管，其信息质量、可靠性、可读性等得不到有效保障。中文在线健康网站质量评价研究相比较于国外较少见，大多数在线健康网站质量评价研究的结果类似，在笔者通过 CNKI 查阅的相关健康信息质量评价的实例研究中，多数研究结果表明中文在线健康信息质量不佳，极少数处于中等水平，中文在线健康信息存在着信息不完整不准确、重复和无关链接过多、信息不完整、认知偏见和内容错误等一系列的问题[20]。五个典型的中文在线健康信息质量评价实证研究案例见表 2。

表 2.典型的中文在线健康信息质量评价实证研究案例

作者（年代）	题目	研究对象	质量评价工具及维度	主要结论
钟乐（2010）	中文网站中儿童注意缺陷多动障碍相关信息的质	注意缺陷多动障碍在线健康信息	可信度、详尽度（DISCERN）	互联网上关于注意缺陷多动障碍的中文健康信息质量不佳

		量评估		
朱 晨（2018）	慢性肾病相关健康类中文网站信息质量评价	肾病在线健康信息	网站行为准则（HONcode）、网站信息质量（DISCERN）	业务广告成分居多、内容归源性与隐私保护意识不强
吴 超（2018）	中文网站中关于肺癌健康信息的质量评估	肺癌在线健康信息	完整性、准确性（DISCERN）	中文网站关于肺癌健康信息的质量不佳
王元松（2019）	中国互联网网站关于糖尿病健康信息的质量评价	糖尿病在线健康信息	可靠性（DISCERN）、治疗选项（DISCERN）	重复和无关链接过多，信息不完整、认知偏见和内容错误
厉锦巧（2019）	冠心病网络健康信息的质量、可理解性和可实施性评估	冠心病在线健康信息	信息质量（DISCERN）、可理解性和可实施性（PE-MAT-P）	大多数网站的质量处于中等水平

4.2 外文在线健康信息质量评价的实证研究

通过检索 Pubmed，我们发现关于国外在线健康信息质量评价的实证研究主要起始于 20 世纪 90 年代中期，近年来呈现出逐渐上升的趋势，国外对医学或健康信息网站的评价开展较早，已经有较为成熟的评价标准和工具，政府或学术组织也已经创建并对权威的公众健康信息网站展开维护。许多医学专业人员和机构展开了大量对互联网健康信息质量的调查研究，如 1997 年，PieroImpicciatore 等调查了网上关于儿童发烧的家庭护理方面的信息。在被调查的 41 个网页中，只有 4 个网页里的内容与权威机构关于护理发热儿童的建议基本吻合[21]。此后，许多医学专业人员和机构就互联网健康信息质量陆陆续续开展了评估工作，所涉及的病种相当广泛，包括了内科、外科、妇产科、儿科、传染科和中医各科较受关注的部分疾病，癌症、男性不育症、痛经、鼻整形等若干常见病种。Eysenbach G 等人研究关于治疗癌症的中草药网络信息的质量和安全问题，他们从 10 个搜索引擎和关键词“中草药”和“癌症”中选出 200 个相关网站，其中 43 个网站符合所有入选标准。使用 DISCERN 工具对质量和安全指标进行了评估，结果显示，大多数网站在信息的准确性、信息来源的显示、信息表述的合理性或更新的规律性等方面质量较低。商业网站有许多不准确或误导性的信息，有的只强调使用中草药的积极方面，对于其副作用并不进行说明[22]。KaickerJatin 等人为确定那些介绍有关疼痛信息的网站的质量，并解释各网站之间所提供信息在质量和可读性方面的差异，对 300 个网站进行了评估。结果 29 个网站不符合纳入标准，110 个

网站重复，剩余 161 个网站符合要求。得出结论为：这些网站信息质量一般，存在不少缺陷。国外对健康信息的研究所涉及的主题相当广泛，这些研究有的对相关主题信息的整体质量进行评价，有的则只研究质量的某一个方面或几个方面的指标，如可信性、准确性、可获取性等，这些研究结果表明许多国外在线健康信息存在质量较差、可读性不高、可信度参差不齐等问题[23]。五个典型的外文在线健康信息质量评价实证研究案例见表 3。

表 3.典型的外文在线健康信息质量评价实证研究案例

作者（年代）	语种	题目	研究对象	质量评价工具及维度	主要结论
Eysenbach G（2004）	英语	关于治疗癌症的药物网络信息的质量和安全问题	治疗癌症药物在线健康信息	质量、安全性（DISCERN）	大多数网站在信息的准确性、信息来源的显示、信息表述的合理性或更新的规律性等方面质量较低
Sarah Mozafarpour（2011）	英语	评估输尿管支架在线信息的可读性、质量和普及程度	输尿管支架在线健康信息	可读性、质量和普及程度（JAMA）	许多网站的可读性远远高于标准推荐，在线信息在很多情况下是有问题的
Stephanie Robins（2016）	英语	关于男性不育症的在线健康信息:可读性、适用性和质量的评估	男性不育症在线健康信息	可信度（DISCERN）和适用性（SAM）	关于男性生育能力的高质量在线信息来源缺乏
Adam T Haymes（2016）	英语	互联网上的鼻整形健康信息的质量	鼻整形在线健康信息	可靠性（LIDA）、可读性（FRES）	鼻整形网络健康信息通常质量低，不可靠，可读性较差。
Jordan Lovett（2019）	英语	关于痛经的在线信息:可读性，可信性，质量和可用性的评估	痛经在线健康信息	可读性，可信性，质量和可用性	网上有关痛经的信息可读性普遍较差，可信度参差不齐，质量参差不齐。

5 讨论

对于在线健康信息质量评价工具的研究已经引起全球学者的广泛关注。回顾这些评价工具，我们发现虽然对于在线健康信息质量的评价维度在不同研究中具有不同的表述，但是基本都是从

信息的可靠性、可读性、可信度、完整性、适用性等方面来开展评价的。除此之外，一些评估工具，例如：LIDA、Michigan Checklist 等，还针对健康类网站的内容布局和易用性方面设计了相关评价条目。在这些评价工具中，HONcode 是一个比较特殊的工具。其特殊之处在于，除了为用户提供在线健康信息质量的评价指标外，还从在线健康信息发布者的角度，提出了在线健康信息发布所应该遵循的道德规范和准则，并提供健康网站质量认证服务。近年来，HON 又提出了 HONcode 规范，专门针对健康类 App 质量开展评价，并提供认证服务。虽然 HONcode 认证服务并非是强制实施的，但是这种在线健康信息质量认证模式为我们完善在线健康信息管理工作提供了参考。

在肯定国外在线健康信息质量评价工具研究成果和应用现状的同时，也应该看到国内对于在线健康信息质量评价研究和在线健康信息质量监控管理仍然有大量的工作要做。主要体现在以下四个方面。

第一，中文在线健康信息质量评价工具的开发没有很好地体现我国居民的信息行为特色

虽然国内已经有一些学者针对中文在线健康信息质量的评价工具开展了相关研究工作，我们在前文做了相关描述，此处不再赘述。但是相比较国外的评价工具，国内目前所开发的在线健康信息质量评价工具评价内涵相对较窄，除了针对网站的健康信息质量评价以外，尚没有开展针对移动终端网络应用平台（例如：微信公众号、微博、QQ、支付宝生活号等）的信息质量评价工具研究。而后者是最贴近百姓生活，人民群众最喜闻乐见的健康信息获取途径。因此，尽快开发针对移动终端网络应用平台的健康信息质量评价工具具有极大的理论价值和现实意义。

第二，缺少在线健康信息质量的国家认证机制

参照国外的 HONcode 在线健康信息质量认证标准，国内一些健康类网站已经申请并取得 HONcode 认证。但是我们回顾 HONcode 的认证规则可以发现，部分条目并不适用于我国健康网站的质量评价，这可能是不同国情催生的社会文化背景决定的。因此，这种认证标准是否适合中国互联网信息发布模式的现状；所开展的认证和评价工作是否能够得到客观、公正、公开的执行；认证结果是否能够对中文在线健康信息质量产生实质性的监控和提高，都是重要的研究课题。因此，这种认证模式本身是非常好的一种在线健康质量监控机制，但是具体实施时，建立何种机制，由谁来监督管理，以确保相关的认证和监控工作落实到位，切实保障我国在线健康信息的质量，需要我们进一步深入思考。

第三，评价工具的内容设计难以契合中文在线健康信息质量评价的内涵

在线健康信息质量的定义是什么，什么样的在线健康信息才是好的，从哪些方面来评价健康信息质量比较合理。这些问题一直是在线健康信息质量研究的重点。在现有的在线健康信息质量评价工具中，大概的评价维度包括：可信度、可读性、适用性、可靠性、可用性等若干方面。这里面存在两方面的问题：第一，对于这些评价维度的概念，不同评估工具的评价维度存在交叉。例如：可靠性评价条目往往和可信度评价条目存在交叉。第二，一些评价方法存在局限性，难以应用于中文在线健康信息质量的评价。例如：可读性评价中最广泛使用的 SMOG 工具，就缺少对中

文的支持。这些问题让目前的评价工具在内容设计上，难以契合中文在线健康信息质量的评价需求。

第四，缺少用户视角，难以普及推广

几乎所有的评价工具都是针对专业人士的使用所设计、开发的。评价工具中出现的“表单”、“主机托管”、“交换链接”、“利益冲突”等名词术语，对于普通公众来说难以理解，从而使这些工具难以走向社会，走向互联网的终端用户，极大限度地影响了这些工具的推广应用。可能这种设计思路没有问题，把在线健康信息质量的评价工作交给专家来完成，但是作为绝大多数的互联网终端用户来说，仍然缺乏行之有效的在线健康信息质量评价工具，让他们难以区分什么样的信息是高质量的。

综上所述，我国在线健康信息质量评价相关研究仍然存在很多不足之处，首先，中文在线健康信息质量评价工具的开发没有很好地体现我国居民的信息行为特色，目前国内所开发的健康质量评价工具适用范围狭窄，无法加以推广；其次，我们国家缺少一个在线健康信息质量的国家认证机制，多数学者自成一家，缺少规范的在线健康信息质量认证机制；再次，目前中文在线健康信息质量的评价研究多使用国外的评估工具，但这些工具不一定适用于中文的在线健康信息质量评价工作；最后，我国的在线健康信息质量评价主要是由专家学者完成，缺少用户视角。未来我国开展健康信息质量评价研究应该从我国的国情出发，从用户的角度出发，制定符合国情的评价标准，能够被普通公众使用的在线健康信息质量评价工具。

OR-031

电子烟视频传播效果影响因素研究——基于哔哩哔哩平台

谢之英、向菲
华中科技大学同济医学院

目的/意义 量化哔哩哔哩弹幕网中电子烟视频，对其播放效果影响因素进行分析，探究电子烟视频在哔哩哔哩弹幕网的传播状况和影响播放量的指标更本质的共性因子。

过程/方法 以哔哩哔哩弹幕网中的电子烟视频为研究对象，通过 Python3.8 编写网络爬虫程序和八爪鱼采集器采集视频的基本信息，通过主成分分析和多元线性回归分析探究其播放量的影响因素和找出影响这些指标更本质的共性因子。

结果/结论 研究结果表明影响这些指标更本质的共性因子分别是观众互动、up 主影响力和时间因素，三个指标都显著的正向影响视频的播放量。随后讨论了平台对于消费者态度和行为的影响，最后从用户、平台和管理者的角度对电子烟的监管提出建议。

1 研究背景

哔哩哔哩弹幕视频网站是中国年轻人高度聚集的文化社区和视频平台，该网站创建于 2009 年 6 月 26 日，被粉丝们亲切的称为“B 站”。普通游客如果想获得视频完全的观看和评论等权限必须成为正式会员，成为正式会员有两条途径：一、通过考试成为正式会员；二、通过正式会员购买邀请码成为正式会员[1]。而考试的内容主要是 ACG 文化，这也使得 B 站的用户主要以年轻人为主。B 站发布的 2019 年 Q2 财报中表示 2016 年 Z 世代的人口已经达到 3.28 亿，Z 世代指在 1995-2009 年间出生的人，按哔哩哔哩给出的数据，平台用户的平均年龄为 21 岁，新用户的平均年龄为 19.8 岁，其中未满 14 岁的人群低于 10%。在 QuestMobile2018 年 12 月 19 日发布的 Z 世代洞察报告中显示截止 2018 年 10 月，Z 世代用户突破 3.69 亿，而在 Z 世代偏爱 APP、Z 世代偏爱泛娱乐 APP 中哔哩哔哩都高居榜首。

电子烟，是一种通过雾化器雾化电子烟液向呼吸系统传输尼古丁或其他物质的产品，也称为电子烟碱传输系统[2]。来自美国、韩国、英国等国家的研究均表明未成年人使用电子烟的比例逐年增高[3-6]。深圳作为全球电子烟的主要生产销售地[7-9]，其相关调查显示，中小学生中听说过电子烟的比例为54.0%、见过电子烟的比例为33.9%、使用过电子烟的比例为3.8%[10]。2019年国家烟草专卖局、国家市场监督管理总局联合下发《关于进一步保护未成年人免受电子烟侵害的通告》文件，体现了电子烟对于未成年人的威胁，随着电子烟的流行，年轻人眼中的电子烟并不是一种帮助戒烟的产品而是一种会影响他们尝试或者开始吸烟的产品[11]，不可否认的是电子烟的受众逐渐偏向青少年并且正逐步影响着他们的吸烟行为[12]。青少年的年龄、人际关系和电子烟本身带来的新鲜感都会让青少年成为该产品的重要消费群体，而电子烟通过社交媒体的传播会更进一步对青少年产生吸引力，由此B站会对促进青少年使用电子烟方面起到一定作用。

2 研究设计

2.1 数据获取

本研究以 2020 年 4 月 5 日前 B 站电子烟相关视频为研究样本，由于 B 站对于一个关键词最多只能显示 1000 条视频，所以本研究通过“电子烟”，“蒸汽烟”，“烟油”，“烟圈”，“雾化器”，“vape”，“烟弹”7 个关键词进行检索，通过 Python3.8 编写网络爬虫程序和八爪鱼采集器得到 5314 条视频的基本信息，包括每条视频的标题、视频时长、发布时间、视频分区、视频 AV 号、标签、播放量、点赞数、硬币数、收藏数、分享数、弹幕数、评论数、是否原创、up 主 ID、up 主粉丝数、up 主视频平均播放量。

2.2 数据清洗

首先，去除了 1629 条重复和无效的视频，其次在剩余 3685 条视频中将标题或者标签中含有“电子烟”或“蒸汽烟”的视频视作电子烟相关视频，剩余的 1209 条视频则由人工筛查并得到其中 259 条视频与电子烟相关。最终排除了播放量为 0 的视频，得到 2408 条与电子烟相关的视频信息。

2.3 数据处理

首先对 2408 条视频的发布时间分布、视频分区分布以及高频视频标签进行描述性分析，其次为了探究 B 站电子烟视频播放量影响因素和影响这些指标更本质的共性因子，对视频元数据进行相关性分析、以播放量为因变量，其他变量为自变量进行主成分分析和多元线性回归分析。

3研究结果

3.1 描述性分析

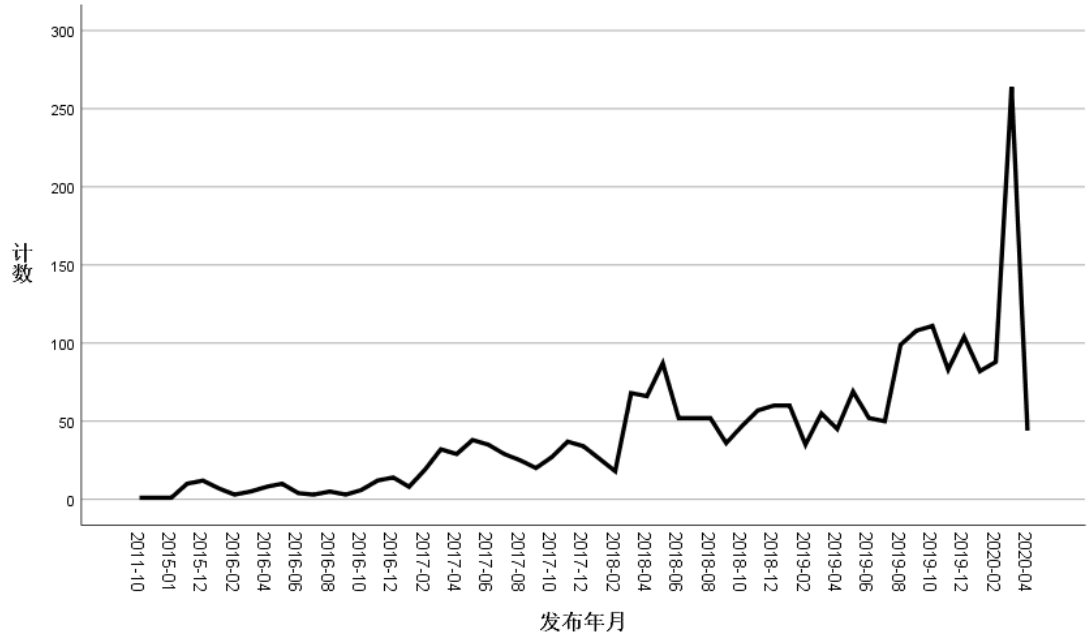


图1 电子烟视频发布时间分布图

图1为2408条电子烟视频发布时间分布图，样本中的第一个电子烟视频于2011年10月发布在B站，主要内容是多人花式吐烟圈。自2011年10月至2018年2月视频的发布量始终处于较低的水平。2018年1月16日全国烟草工作会议在北京召开，会议报告中多次提及了电子烟，自此电子烟视频的发布数量大大增加，基本超过50条/月。2019年8月至11月间，央媒点名电子烟，“健康、戒烟、无二手烟”等宣传词不符合广告法；美国疾病控制与预防中心报告了450多起与使用“电子烟”有关的严重肺病的病例；国家市场监督管理总局、国家烟草专卖局发布《关于进一步保护未成年人免受电子烟侵害的通告》，敦促电子烟生产、销售企业或个人及时关闭电子烟互联网销售网站或客户端，此时电子烟视频的发布数量也达到了几乎100条/月。而在2020年2月至3月多家电子烟品牌召开新品发布会，发布新品，因此此时的发布数量陡然升高，于2020年3月达到264条/月。由此看出电子烟视频的发布都与实时热点紧密相关，视频内容也存在多样性，包括在《关于进一步保护未成年人免受电子烟侵害的通告》发布后大量出现的科普类视频和在电子烟新品发布后的开箱视频等。

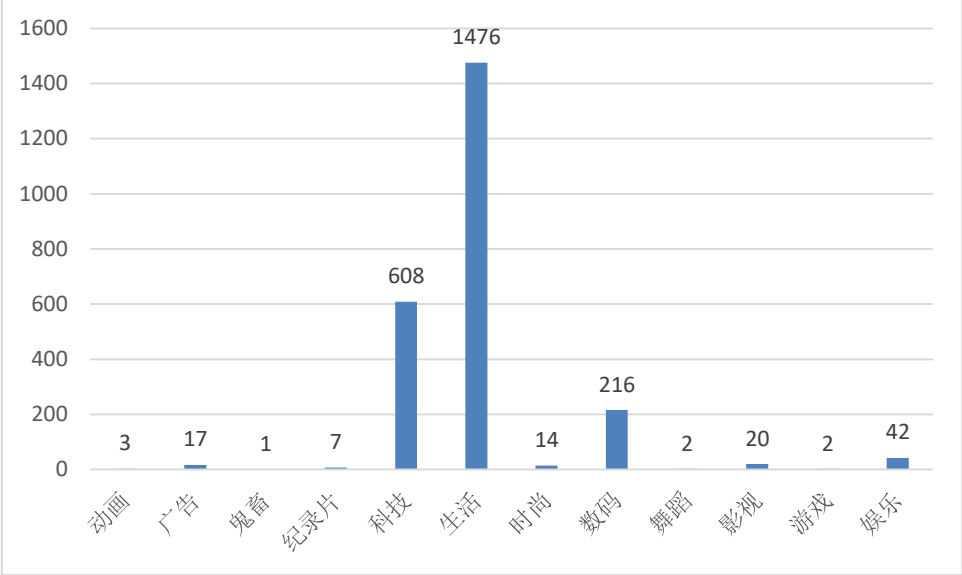


图2 电子烟视频分区分布图

图3 电子烟视频标签词云图

图2为电子烟视频分区分布图，电子烟视频的分区主要集中在生活区、科技区、数码区，这与其内容有着密切的关联，生活区视频内容主要包括电子烟开箱、花式烟圈秀、知识科普等；科技区和数码区视频内容则包括电子烟评测、电子烟知识讲解等；虽然从内容的角度来看，不同分区的视频内容存在较大重叠，但整体来看生活区视频更偏向于对于电子烟的个人理解和感受，而科技区和数码区更偏向于从电子烟机械构造和个人改造电子烟的角度来展开。

图3为以2408条视频的标签为对象的词云图，通过排除前五个与电子烟本身高度重叠的词“电子烟”、“蒸汽”、“电子”、“烟圈”、“蒸汽烟”得到，从中可以看出“戒烟”出现频率最高，在电子烟以往的宣传词中多次提到电子烟具有戒烟的效果，而这也误导了众多消费者，

从视频标签中就可以体现。其次“开箱”、“VAPE”、“测评”、“DIY”等都体现了不同的视频内容。

3.2主成分分析

表1 相关系数矩阵

		点赞	硬币	收藏	分享	弹幕数	评论数	up主粉丝数	视频时长	发布天数	平均播放量
点赞	皮尔逊相关性	1	.920**	.862**	.814**	.883**	.519**	.303**	-.058**	-.029	.477**
	Sig.（双尾）		.000	.000	.000	.000	.000	.000	.004	.150	.000
硬币	皮尔逊相关性	.920**	1	.752**	.696**	.794**	.461**	.313**	-.033	.001	.553**
	Sig.（双尾）	.000		.000	.000	.000	.000	.000	.103	.977	.000
收藏	皮尔逊相关性	.862**	.752**	1	.904**	.915**	.341**	.079**	-.029	.017	.209**
	Sig.（双尾）	.000	.000		.000	.000	.000	.000	.157	.408	.000
分享	皮尔逊相关性	.814**	.696**	.904**	1	.865**	.529**	.122**	-.030	.023	.262**
	Sig.（双尾）	.000	.000	.000		.000	.000	.000	.143	.251	.000
弹幕数	皮尔逊相关性	.883**	.794**	.915**	.865**	1	.513**	.205**	-.006	.042*	.399**
	Sig.（双尾）	.000	.000	.000	.000		.000	.000	.771	.038	.000
评论数	皮尔逊相关性	.519**	.461**	.341**	.529**	.513**	1	.373**	-.049*	.012	.488**
	Sig.（双尾）	.000	.000	.000	.000	.000		.000	.016	.556	.000
up主粉丝	皮尔逊相关性	.303**	.313**	.079**	.122**	.205**	.373**	1	-.066**	.025	.591**

数	Sig.（双尾）	.000	.000	.000	.000	.000	.000		.001	.225	.000
视频时长	皮尔逊相关性	-.058**	-.033	-.029	-.030	-.006	-.049*	-.066**	1	.195**	-.115**
	Sig.（双尾）	.004	.103	.157	.143	.771	.016	.001		.000	.000
发布天数	皮尔逊相关性	-.029	.001	.017	.023	.042*	.012	.025	.195**	1	.044*
	Sig.（双尾）	.150	.977	.408	.251	.038	.556	.225	.000		.031
平均播放量	皮尔逊相关性	.477**	.553**	.209**	.262**	.399**	.488**	.591**	-.115**	.044*	1
	Sig.（双尾）	.000	.000	.000	.000	.000	.000	.000	.000	.031	

表 1 为变量的相关矩阵，“点赞”、“收藏”、“分享”是每个视频下方的该互动按钮点击的次数，“硬币”则是该视频 up 主收到的硬币数量，正式用户每日登录便会获得一枚硬币，并且可以给 up 主的视频投币，原创视频可以投币 1 枚或 2 枚，而转载视频只可以投币 1 枚，up 主能够从中获得实际收益。“弹幕数”和“评论数”分别为该视频发布至数据收集期间累计的弹幕数量和评论数量。“up 主粉丝数”为数据收集时 up 主的粉丝数，最小为 0，最大为 up 主“柚子木字幕组”，粉丝数高达 3442434。“视频时长”和“发布天数”具体为该视频的具体秒数时长和从发布至 2020 年 4 月 5 日的具体天数。“平均播放量”为该视频 up 主所有视频累计播放量除以该 up 主累计发布的视频数。从表中可以看出 10 个变量中许多变量之间都有很强程度的相关性，大部分相关性程度很高，说明可以从其中提取主成分公共因子，因此进行主成分分析是合理的。

表2 KMO和巴特利特检验

KMO 取样适切性量数		.786
巴特利特球形度检验	近似卡方	22378.112
	自由度	45
	显著性	.000

表2为KMO和Bartlett检验的输出结果，KMO值越接近于1，表明这些变量进行因子分析的效果越好，大于0.7便可以接受，本次KMO值为0.786，说明数据可以进行因子分析。Bartlett球形检验统计量为22378.112，显著性小于0.01,因此可认为相关系数矩阵和单位阵有显著差异，所选的变量适合做因子分析。

表3 公因子方差

变量	初始	提取
点赞	1.000	.924
硬币	1.000	.811
收藏	1.000	.938
分享	1.000	.872
弹幕数	1.000	.918
评论数	1.000	.525
up主粉丝数	1.000	.742
平均播放量	1.000	.785
视频时长	1.000	.594
发布天数	1.000	.627

表3为输出的原始变量的公因子方差，提取值表示变量公因子方差的值，其中“点赞”的公因子方差为0.924，说明因子能够解释“点赞”方差的92.4%，表中可以看出所有的提取值均大于0.5，说明每个原始变量都能被主成分因子较好的解释。

表4 主成分提取结果

	初始特征值			主成分提取			因子旋转		
成分	特征值	贡献率	累积贡献率%	特征值	贡献率	累计贡献率%	总计	贡献率	累积贡献率%
1	5.021	50.207	50.207	5.021	50.207	50.207	4.421	44.211	44.211
2	1.525	15.255	65.462	1.525	15.255	65.462	2.116	21.162	65.372
3	1.190	11.902	77.364	1.190	11.902	77.364	1.199	11.992	77.364
4	.797	7.967	85.331						
5	.612	6.119	91.450						
6	.426	4.260	95.710						
7	.229	2.288	97.998						
8	.113	1.125	99.123						
9	.052	.522	99.645						
10	.035	.355	100.000						

表4为主成分提取的结果，特征值大于1是有用因子的通用标准，前三个成分的特征值λ_1=5.021，λ_2=1.525，λ_3=1.190均大于1，其累计贡献率为77.364%，也意味着前三个因子足以解释77.364%的原始变量总方差，而损失的信息量不足23%，损失的信息量较少，因此可以说明主成分提取比较理想。

表5 因子旋转载荷

因子变量	成分		
	1	2	3
收藏	.969	-.015	-.002
弹幕数	.937	.199	.035
分享	.930	.082	.006
点赞	.907	.314	-.052
硬币	.816	.381	-.018
up主粉丝数	.023	.861	-.007
平均播放量	.236	.854	-.033
评论数	.444	.572	-.005
发布天数	-.009	.100	.785
视频时长	.003	-.131	.760

表5为因子载荷旋转后的结构，旋转后主成分因子的实际意义更加清晰。第一主成分在“点赞”、“硬币”、“收藏”、 “分享”、“弹幕数”中具有较大的载荷分别为0.907、0.816、0.969、0.930、0.937，这些因素主要体现了视频质量和观众的互动程度[13]，前四个指标都表达对视频的赞赏与喜爱，点赞是对视频内容的认可，硬币则是对up主付出成本制作或搬运视频的打赏，收藏意味着该条视频可能需要反复观看，分享是对视频内容的承认也是进一步的帮助其进行传播，弹幕则主要包括信息需求、娱乐需求和社交增强需求三类[14]；第二主成分在“up主粉丝数”、“平均播放量”、“评论数”中具有较大的载荷分别为0.861、0.854、0.572，这些因素主要体现了up主的影响力；第三主成分在“发布天数”、“视频时长”中具有较大的载荷分别为0.785、0.760，这些因素主要体现了视频的时间因素。

3.3多元线性回归分析

通过主成分分析的到可以概括原始变量的三个主成分因子，将得到的三个主成分因子作为多元线性回归模型中的自变量，播放量作为因变量利用SPSS25进行多元线性回归分析，输出结果如下：

表6 模型输出结果

模型	R	R^2	调整R^2	估计标准误差	R 方变化量	F 变化量
1	.352[a]	.124	.123	.73600	.124	339.178
2	.416[b]	.173	.172	.71518	.049	143.111
3	.462[c]	.213	.212	.69754	.041	124.160

a. 预测变量：(常量)，第二主成分

b. 预测变量：(常量)，第二主成分，第一主成分

c. 预测变量：(常量)，第二主成分，第一主成分，第三主成分

表7 方差分析表

模型		平方和	自由度	均方	F	显著性
1	回归	183.730	1	183.730	339.178	.000[b]
	残差	1303.310	2406	.542		
	总计	1487.041	2407			
2	回归	256.928	2	128.464	251.161	.000[c]
	残差	1230.112	2405	.511		
	总计	1487.041	2407			
3	回归	317.340	3	105.780	217.402	.000[d]
	残差	1169.700	2404	.487		
	总计	1487.041	2407			

表8 系数输出表

模型		未标准化系数		标准化系数	t	显著性	共线性统计	
		B	标准错误	Beta			容差	VIF
1	(常量)	3.092	.015		206.159	.000		
	第二主成分	.276	.015	.352	18.417	.000	1.000	1.000
2	(常量)	3.092	.015		212.160	.000		
	第二主成分	.276	.015	.352	18.953	.000	1.000	1.000
	第一主成分	.174	.015	.222	11.963	.000	1.000	1.000
3	(常量)	3.092	.014		217.525	.000		
	第二主成分	.276	.014	.352	19.432	.000	1.000	1.000
	第一主成分	.174	.014	.222	12.265	.000	1.000	1.000

第三主成分	.158	.014	.202	11.143	.000	1.000	1.000

采用逐步回归法逐步引入自变量，模型一中预测变量只包含第二主成分，模型二中包含第一主成分和第二主成分，模型三中包含第一主成分、第二主成分和第三主成分。调整R^2值（0.123,0.172,0.212）逐渐增大，说明模型的拟合效果逐渐提高。对于模型假设，F检验显著（339.178,251.161,217.402，$p < 0.01$），说明线性回归关系成立。表8中输出的是各解释变量的回归系数、回归系数显著性检验的情况，可以看出显著水平α=0.01，表中显著性检验的概率均小于显著水平α，说明因变量与三个解释变量之间的线性关系显著，因此得到多元线性回归方程为：

$$Y = 3.092 + 0.276X_1 + 0.174X_2 + 0.158X_3$$

4讨论

本研究表明对于B站电子烟视频的传播用户的参与度和视频质量起到最大的作用（$\beta_1 = 0.276$），其次是up主的影响力（$\beta_2 = 0.174$），最后包括的视频时长和发布时间的时间因素影响最小（$\beta_3 = 0.158$），说明用户的参与度和视频质量、up主的影响力、时间因素都显著正向影响视频的播放量，这是由于高质量的视频和高度的用户参与更容易使视频登上首页被更多的人观看，up主的影响力更大则相应的拥有更大的粉丝量，自然也会带来更多的播放量，而随着视频的发布时间越来越久，观看的人数也会越来越多，视频时长较大的视频通常会导致用户分次、多次观看，因此也会带来更大的播放量。

虽然国家已经敦促各电商平台和企业下架电子烟产品、关闭电子烟商铺、撤回广告，但是社交视频平台依然广泛传播着电子烟相关视频，而在这些视频中不乏包含广告的推广电子烟产品的视频，从视频的标签就可以看出“戒烟”是除了与电子烟本身密切关联的标签外出现次数最高的，早期电子烟也存在很多打着“戒烟”口号的宣传语，因此B站视频中也存在大量错误引导的状况。B站电子烟视频的累积播放量达到了19047245次，累积收藏数达到了162666次，累计分享数达到了101810次，虽然B站推出了青少年模式，但是只有部分视频封面警示未成年人不要观看，说明对于未成年人的观看限制并不完善，这些视频可能会增加未成年人购买电子烟的机会。

从用户的角度来看，选择信息质量较高的官方账号发布的视频来了解和获取电子烟相关信息更有利，因为用户原创内容质量参差不齐，与健康安全相关的知识需要选择可靠性更强的的信息源。从平台的角度来看，对于用户原创内容的管理与审核需要更加的严谨，并且限制未成年用户对于用户原创电子烟相关视频的观看。从监管的角度来看，电子烟的营销工作在很大程度上利用了新的媒体渠道[15]，因此各社交媒体与视频网站都需要给予足够的重视，为此建议公共卫生部门加大对于社交平台和视频网站上关于用户原创电子烟视频的监管，尤其是影响力大的up主，他们的视频会有更良好的传播效果，如果是错误信息则会导致更大程度的危害，因此需要更加严格的审核其视频内容，对于反复上传错误引导类电子烟视频的up主需要作出警告和惩罚；由于用户参与度和视频质量是更为重要的影响因素，因此需要加大专业内容的制作和传播力度，提升视频质

量，于视频中多引导观众互动以达到提升传播效果的目的；制作较长时长的视频，由于B站并不是短视频平台，因此视频的时长不可过短；最后可以联合B站开展公众宣传等活动，帮助其用户更加全面的认识电子烟。

以下是本研究的局限：首先本研究只选取了 7 个最主要相关的关键词进行检索，不能排除有的视频与电子烟相关但没有被检索出来，但是本研究认为那部分视频不太能代表电子烟影响力，因为这部分视频并不依赖相关关键词检索吸引观众。其次本研究没有针对视频简介中的链接进行研究，其中可能包含推广与营销的作用。最后是本研究没有对视频的内容进行分析，这可能产生许多有价值的成果。

OR-032

河南省中医院患者满意度调查及影响因素分析

邢晓琳[1]、吕晖[2,4]、王晶晶[3]、任文杰[2]
1. 新乡医学院 2. 新乡医学院健康中原研究院
3. 新乡医学院公共卫生学院 4. 新乡医学院管理学院

目的 探讨河南省中医院患者满意度及影响因素。

方法 本文以河南省中医院为研究对象，应用较为成熟的满意度理论模型，结合中医院实际，对影响满意度的调查指标进行科学设计，对河南省 16 家中医院患者满意度的影响因素进行深入调查，运用较为专业的统计学方法对患者满意度影响因素进行分析。

结果 研究发现影响门诊患者满意度的因素主要集中在候诊环境、空间合理、环境设施、候诊秩序、流程便捷、路标清晰等，且对医患沟通最满意，对医疗收费满意度最低。影响住院患者满意度的因素主要集中在陪护探视、技术水平、路标清晰、流程便捷、服务效率、查房工作等方面，且对医疗技术的满意度最高，对医疗环境的满意度最低。

结论 医院应根据门诊患者和住院患者的需求，完善医疗服务体系，提高患者就医感；医生应“以患者为中心”，合理用药，合理检查，制定个体化的医疗方案。

患者满意度是指患者由于在健康、疾病及生命质量等方面存在一定要求，进而产生了对医疗保健服务的期望，基于这种期望对医疗服务进行主观评价[2]。在推进改善医疗体系工作中，患者满意度是衡量医疗服务、医疗质量和反映社会效益的重要指标[3]，在建设和谐社会的大环境中,患者满意度是考量监管水平及社会和谐稳定程度的关键要素[4]。它是适应医学模式转变的一种新的医疗质量评价方法，只有赢得患者的支持与信任，医疗机构才能在日益激烈的竞争中赢得优势。

1 对象及方法

1.1 调查对象

本研究在2019年10月对16家中医院的就诊患者进行抽样调查，每家医院根据就诊患者人数的多少分别抽取60-100个患者进行问卷调查，门诊患者和住院患者分别进行，采用不同的调查问卷。纳入标准:(1)调查期间在医院门诊就诊或住院的;(2)年龄18-80岁;(3)意识清楚,能正常言语交流;(4)同意配合问卷调查。最终获取门诊患者有效问卷568份，住院患者有效问卷633份，总计1201份。

1.2 调查方法

本研究查阅大量的满意度相关研究，结合中医院实际，对满意度问卷进行设计，为保证研究的全面性、客观性，分别设计了门诊患者调查问卷和住院患者调查问卷。测量尺度采用李克特 5级量表，很满意(5分)、比较满意(4分)、一般(3分)、不太满意(2分)、很不满意(1分)，分值越高，满意度越高，分值越低，表示患者对调查指标项目越不满意。无涉及即选“N”予以剔除。本研究采取了满意度调查中常用的发放问卷，现场作答，现场回收的方法，调查员统一培训。

1.3 统计方法

采用SPSS 22.0软件进行统计分析，采用统计描述、spearman相关、t检验和方差分析等统计学方法。当$P<0.05$时，表示差异有统计学意义。

2 结果

2.1 门诊患者对中医院满意度调查情况

门诊患者对中医院的满意度总得分为58.11±7.15分，处于中高水平，门诊患者对医患沟通的满意度最高，各项目均分为 4.30±0.50 分，对医疗收费的满意度最低，均分为 3.63±0.74 分，结果见表1。

表1 门诊患者满意度调查结果

分组	总分	均分
医德医风	11.69±1.58	3.90±0.53
医患沟通	8.59±1.00	4.30±0.50
流程便利	18.71±2.99	3.74±0.60
医疗环境	11.49±1.91	3.83±0.64
医疗收费	3.63±0.74	3.63±0.74
医疗技术	4.02±0.69	4.02±0.69
总分	58.11±7.15	3.87±0.48

2.2 住院患者对中医院满意度调查情况

住院患者对中医院满意度总得分为90.55±9.70，处于中高水平，其中住院患者对医疗技术满意度最高，各项目均分为4.25±0.69分，对医疗环境满意度相对较低，均分为4.05±0.56分，结果见表2。

表2 住院患者满意度调查情况

分组	总分	均分
医德医风	12.72±1.52	4.24±0.51
医患沟通	20.78±2.19	4.16±0.44
流程便利	24.41±3.13	4.07±0.52
医疗环境	20.23±2.81	4.05±0.56
医疗费用	8.16±1.29	4.08±0.64
医疗技术	4.25±0.69	4.25±0.69
总分	90.55±9.70	4.12±0.44

2.3 门诊患者满意度相关因素分析

Spearman 相关分析显示，服务态度、隐私保护、服务效率、耐心倾听、通俗易懂、挂号便捷、路标清晰、空间合理、流程便捷、交通便利、候诊环境、候诊秩序、环境设施、收费合理、技术水平均与门诊患者满意度呈现明显的正相关(P<0.05)，自费比例与门诊患者满意度呈明显的负相关(P<0.05)，见表 3。

表 3 门诊患者满意度相关因素分析

影响因素	r	P
服务态度	0.594	0.000
隐私保护	0.597	0.000
服务效率	0.452	0.000
耐心倾听	0.54	0.000
通俗易懂	0.496	0.000
挂号便捷	0.576	0.000
路标清晰	0.635	0.000
空间合理	0.737	0.000
流程便捷	0.695	0.000
交通便利	0.631	0.000
候诊环境	0.752	0.000
候诊秩序	0.697	0.000
环境设施	0.719	0.000
收费合理	0.619	0.000
自费比例	-0.104	0.032
技术水平	0.615	0.000

2.4 住院患者满意度相关因素分析

分析结果显示服务态度、隐私保护、服务效率、耐心倾听、知情服务、意见采纳、查房工作、人文关怀、住院手续、陪护探视、路标清晰、空间合理、流程便捷、交通便利、就医秩序、病区

环境、生活舒适、安保工作、环境设施、费用知情、收费合理、技术水平均与住院患者满意度呈明显正相关(P<0.05)，同样，自费比例与总体满意度呈现负相关(P<0.05)，见表 4。

表 4 住院患者满意度相关因素分析

影响因素	r	P
服务态度	0.658	0.000
隐私保护	0.604	0.000
服务效率	0.692	0.000
耐心倾听	0.366	0.000
知情服务	0.565	0.000
意见采纳	0.613	0.000
查房工作	0.688	0.000
人文关怀	0.329	0.000
住院手续	0.595	0.000
陪护探视	0.746	0.000
路标清晰	0.712	0.000
空间合理	0.656	0.000
流程便捷	0.694	0.000
交通便利	0.573	0.000
就医秩序	0.689	0.000
病区环境	0.669	0.000
生活舒适	0.631	0.000
安保工作	0.676	0.000
环境设施	0.676	0.000
费用知情	0.539	0.000
收费合理	0.583	0.000
自费比例	-0.189	0.001
技术水平	0.727	0.000

3 患者满意度调查结果分析

3.1 门诊患者满意度调查结果分析

由门诊患者满意度相关因素数据分析可以看出，影响患者满意度的因素主要集中在候诊环境、空间合理、环境设施、候诊秩序、流程便捷、路标清晰等（相关性前 6 个），医院管理部门应根据门诊患者的需求，着重在医院的环境秩序以及就医流程的便利程度两方面加以改进和完善。

近几年来医患关系一直都是社会敏感话题，暴力伤医事件的大规模发生，激起社会各界的反思，近年来政府、医院等各方面已作出了诸多努力，制定相关法律法规并出台相应政策以保证医

生与患者双方的权利[5]。门诊患者对医患沟通的满意度最高，医生更加耐心的听取患者的病情并采用通俗易懂的方式与患者交流，医患关系得到一定的缓解。

而门诊患者对医疗费用的满意度最低，与《门诊患者非技术服务满意度与影响因素研究 》研究结果一致，说明近年来，随着药品零差率的实施，药品费用得到控制，而大型医院的检查费用，仍是居民就医的主要负担所在。此外，虽说医改以来，医疗费用有所下调，但是由于医疗事故的接连发生，使医生的“防范性”心理提高，为了避免医疗事故的发生，大部分医生选择开全面检查来减少错误的发生，这使得降低医疗事故的同时，增加了患者的经济成本[1]。因此，在进一步保障医生人身安全的同时，医生应严格遵循《医疗机构从业人员行为规范》，合理用药，合理检查。有关部门应继续完善医疗费用的管制水平，公开费用明细。

3.2 住院患者满意度调查结果分析

由住院患者满意度相关因素数据分析可以看出，影响患者满意度的因素主要集中在陪护探视、技术水平、路标清晰、流程便捷、服务效率、查房工作等方面，与门诊患者有所不同的的是，住院患者对便利程度、医疗技术、医德医风、医患沟通均比较注重。患者在住院以后，往往伴随复杂的心理情绪，如极度恐慌、烦躁不安等[6]，因此对医疗服务水平就提出了更高的要求。医院应当根据住院患者的心理特点，提升各方面的服务质量，切实提高患者的就医体验。

医疗技术水平一直是患者满意与否的一个重要指标，患者满意度的高低从其特有的角度也反映了医疗技术水平的高低[7]。本次调查结果显示，住院患者对医疗技术满意度最高。原因可能是随着现代医学的快速发展，医生诊疗技术不断提升，疾病治愈率也得到提高，医生本着“以患者为中心”的理念，为患者制定个体化的医疗方案，能够得到患者的支持和信任。

住院患者对医疗环境满意度最低，在医疗环境涉及的几个因素中，患者对安保工作的满意度相对最高，对医院提供的生活、后勤保障服务最不满意。这些问题与医院管理者责任心不强、管理理念和服务水平滞后，医院文化建设没落到实处，没有真正把“以患者为中心”作为服务的根本宗旨等密切相关。医疗环境的舒适清洁程度和饮食的好坏是影响患者身心健康的重要因素,医院可以通过优化后勤管理、转变服务理念、注重细节、满足患者的切身需要来进一步提高服务质量。

OR-033

移动医疗 APP 用户隐私关注影响因素研究

袁珊珊、罗亚湄、周建、 刘智勇
华中科技大学同济医学院

目的/意义 针对移动医疗 APP 用户，通过对隐私关注影响因素进行研究，进一步了解用户的心里、性格等情况，给开发者、政府等把握用户心理动机、行为动向提供可靠的理论支撑。

方法/过程 通过文献归纳整理，提取影响移动医疗 APP 用户隐私关注的因素，主要包括个人隐私倾向、个人隐私经历、媒体报道、社会制度、医疗信息敏感度、感知有用性。提出本文的研究

假设,构建研究模型，引用成熟量表来制定调查问卷。使用社交网络方式发放问卷，统计收回 480 份问卷，手工对无效问卷进行剔除，得到有效问卷 442 份。

结果 个体隐私倾向和医疗信息敏感度正向影响移动医疗 APP 用户的隐私关注；社会制度和感知有用性负向影响移动医疗 APP 用户的隐私关注；个人的不良事件经历和媒体报道对移动医疗 APP 用户的隐私关注未证实有显著影响。

1 引言

中国互联网络信息中心（CNNIC）统计报告显示：截至 2019 年 6 月，我国拥有 8.47 亿移动互联网用户。跟随其后，在线医疗服务行业在逐渐崛起，如丁香医生、微医、用药助手、39 健康、好大夫在线等各种移动医疗 APP 的用户在不断增加。

移动医疗 APP 通过获取用户的个人信息，给用户提供个性化服务。经总结，移动医疗 APP 中的隐私信息分为三类：第一类为用户个人身份的信息（类似姓名、性别、年龄等）；第二类为用户向移动医疗 APP 提供有关个人健康的信息；第三类为用户在 APP 中所发布的含有个人感情色彩或者表现某种价值观的信息等。通过对数据的获取和进一步挖掘，使得用户的个性化服务得以慢慢实现，但同时有关个人隐私问题引起人们的关注。

2016 年 6 月 22 日发布《中国网民权益保护调查报告 2016》，该报告指出有 54%的互联网用户认为个人信息泄露很严重，并且有 21%的互联网用户认为这问题非常很严重。84%的网民认为个人信息泄露会带来的不利影响[1]。2018 年，威瑞森电信公司(Verizon)发布《2018 年的数据泄露调查报告》，报告中显示医疗保健行业(17.7%)数据泄露严重,其仅次于公共事务管理行业(26.2%)的泄露状况。2019 上半年，44.4%的网民表示去年在上网过程中遭遇过网络安全问题，例如：账号或者密码被盗、个人信息泄露等。由此可见，隐私泄露在医疗行业安全问题中较为突出，对公众造成了一定的影响。这些问题会提高用户对隐私的关注程度，用户从而会减少会拒绝提供有关个人的信息。

因此，为了移动医疗 APP 有更好的发展，能够更加了解用户的行为习惯，本研究分析影响移动医疗 APP 用户隐私关注的因素。

2 研究变量、假设与模型

本研究通过借鉴 IUIPC 量表以及相关研究提出的影响隐私关注的影响因素，确定模型所需要的指标变量，包括隐私关注、个人隐私经历、个人隐私倾向、社会制度、媒体报道、医疗信息敏感度、感知有用性。根据 IUIPC 模型，用户隐私关注可以用收集、控制和意识这三个维度进行测量。在理论研究的基础上提出研究假设，进而构建移动医疗 APP 隐私关注影响因素模型。

2.1 变量释义与研究假设

（1）隐私关注：不同学者的定义不相同，1968 年，学者 Westin[2]最早提出，隐私关注定义为在隐私情境中，主观上个人感受到的公平程度。随后，Smith[3]等学者将其定义为当自身信息被采集和

使用时，个人对此事的关注程度。Campbell 等人定义隐私关注为个人针对相应的隐私情境的主观感受[4]，在 2005 年，Rifon[5]等学者将其定义为一种与隐私泄露和隐私侵害相关联的一系列特殊的信息意识和感知。在本次研究中隐私关注指用户对个人信息被收集、储存、使用的关注程度。

最早，Smith 提出结构性隐私关注（CFIP）模型，该模型将隐私关注分为 4 个维度分别为收集、二次使用、不适当访问以及错误[6]。在互联网环境中，有学者提出新的隐私关注量表，包括信息发现、信息滥用[7]。2004 年，Malhotra[8]在 CFIP 的基础上，提出 IUIPC 模型，该模型提出隐私关注的 3 个维度：控制(control)、意识(awareness)和收集(collection)。收集是引起用户隐私关注的起点，表现出用户对收集个人信息时的焦虑[9]。控制表现出用户对于自己个人信息的控制能力，是否愿意在 APP 中提供自己的隐私。意识是用户对于个人信息被如何使用的感知程度。

基于成型的量表，杨姝[10]从四个方面（理论基础、应用领域、维度及主要贡献）对隐私关注测量量表进行述评，验证 IUIPC 量表比 CFIP 量表具有更高的稳定性，更适合于中国国情。当前，国内关于互联网隐私问题的研究多使用 IUIPC 量表。本研究主要借鉴 IUIPC 量表，使用收集、控制、意识这三个维度对隐私关注进行测量。

（2）个人隐私倾向：个人隐私倾向表示个人对隐私保持开放或保守的态度，并反映出个人对隐私操作的熟悉程度。有研究表明使用互联网时的熟悉程度和使用次数会促进用户隐私关注[14]。对隐私了解的越多，则会越多的保留自己的隐私情况。 有研究生通过实证研究得出个人隐私倾向对隐私关注有显著的影响[15,16]。因此提出假设：

H1：个体隐私倾向正向影响移动医疗 APP 用户的隐私关注

（3）个人隐私经历：个人隐私经历指的是个人曾经是否经历过隐私泄露等事件，个人隐私曾经受到过侵犯的人将更加关注隐私问题，并且处理信息也越来越谨慎[3]。2010 年，有学者[13]也验证了隐私泄露的经历对影响用户隐私关注。因此，提出研究假设：

H2：个人的隐私经历正向影响移动医疗 APP 用户的隐私关注

（4）社会制度：社会制度是指司法机关、政府、企业为保护用户隐私而采取的相关措施。从政府层面看，政府通过以法律规章为主的形式来保护消费者的网络隐私；政府的监管越弱，用户越会加强对个人隐私的关注，制定相关的政策可以降低用户对隐私的关注程度[17,18]。2017 年 6 月正式实施的《中华人民共和国网络安全法》中，给予用户强控制权，企业收集用户的个人信息时，必须得到用户的同意。从行业和企业组织层面来看，发布互联网行业自律、制定相关规章制度有助于用户了解企业如何收集和使用他们的个人信息。有调查研究表明健康 APP 较少设立单独的隐私团队或部门因此需要加强企业隐私保护组织的建设[19]。为减少用户的隐私关注，企业组织须提供安全的交易环境，禁止企业未经许可披露用户注册资料或者以其他方式用户商业目的。因此，做出如下假设：

H3：社会制度负向影响移动医疗 APP 用户的隐私关注

（5）媒体报道：媒体报道指媒体对于隐私泄露事件的报道。媒体对泄露用户隐私的不良事件的报道，满足用户的知情权。有调查显示，2011 年，CSDN 用户的信息泄漏后，大量用户立即做出更改其密码并添加防火墙等行为以保护个人隐私[20]。媒体报道隐私泄露事件越多，用户对隐私关注的程度越大，因此，做出研究假设：

H4：媒体报道正向影响移动医疗 APP 用户的隐私关注

（6）医疗信息敏感性：信息敏感性是指在特定环境下，个人对数据的关注程度[21]。信息敏感性的界限在学术界没有达成一致，它受用户的个体特征和所生活的环境等多因素影响[22]。绝大部分人不愿意向网站提供私密的信息[23]。相比于简历上的信息，人们更愿意提供联系方式，相比于医疗信息，人们愿意提供简历信息。医疗信息的敏感度与提供个人健康信息意愿呈负相关[24]。因此，提出研究假设：

H5: 医疗信息敏感正向影响移动医疗 APP 用户隐私关注

（7）感知有用性：感知有用性指用户主观上认同此项技术能够带来的收益。感知有用性影响用户对产品的使用，感知有用性的程度越强烈，那隐私关注程度就越低[6]。在 LBS 和新媒体精准广告的应用中，感知有用性会显著影响用户的隐私关注[12,13,11,24]。因此，本次研究提出假设：

H6:感知有用性负面影响移动医疗 APP 用户隐私关注

2.2 模型构建

基于上文提出的因素和研究假设，本研究建立了影响移动医疗 APP 用户隐私关注的影响因素的初始研究模型，如图 1 所示：

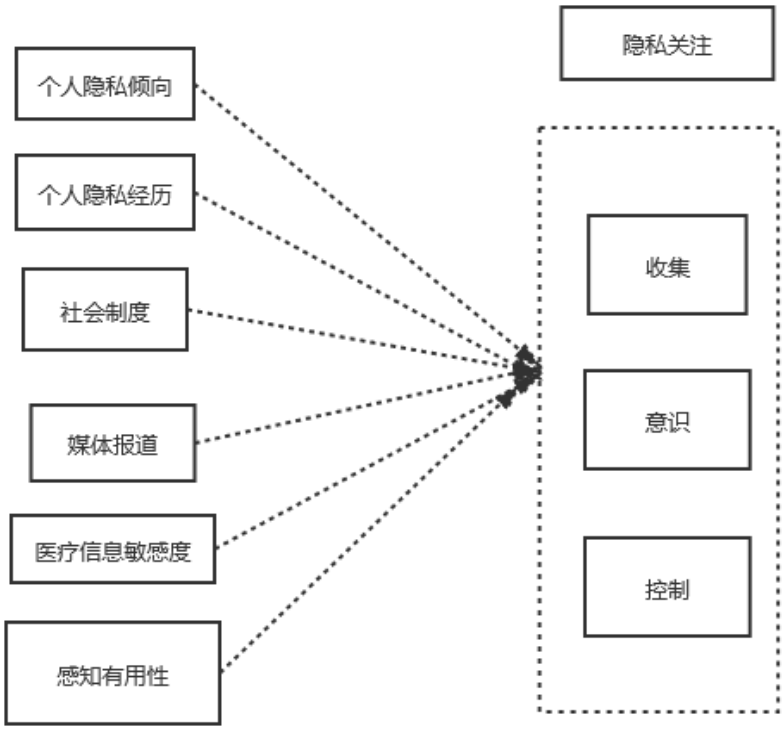


图 1 影响移动医疗 APP 用户隐私关注影响因素的初始研究模型

3 问卷设计和数据收集

3.1 问卷设计

调查问卷分为两个部分：第一部分涉及用户的性别、职业、学历、去年一年的生病频率、使用移动医疗 APP 的次数以及在疫情期间和未来大数据时代对隐私的看法；第二部分为研究变量量表。

目前，被国内外学者广泛利用的隐私关注测量量表为 CFIP 和 IUIPC 量表。根据前人的研究基础，本次研究完全借鉴 IUIPC 量表来测量隐私关注程度。该量表包括控制、意识和收集这三个维度。具体的量表信息如表 1：

表 1 移动医疗 APP 隐私关注测量量表

因素	变量	题项	来源
隐私关注	控制	1.我的个人信息是我的隐私，我应该有权控制且自主决定我的信息如何被收集、使用和共享 2.对于个人隐私保护来说，我认为个人能够控制自己的信息如何被收集或者利用是非常核心的一点	

		3.我不能或者不完全控制我的个人信息被收集或利用时，对我而言是一种侵犯	Malhotra(2004)
	意识	4.移动医疗 APP 的隐私条款中应该清晰明确的说明对个人信息的收集、处理和使用	
		5.我认为我能够知晓自己被收集的个人信息如何使用是非常重要的事	
		6.我很烦恼移动医疗 APP 随便收集我的个人信息	
	收集	7.当移动医疗 APP 向我询问或者收集我的个人信息时，我会谨慎的思考一下是否向其提供	
		8.我很烦恼移动医疗 APP 把收集到的我的个人信息提供给其他公司	
		9.我很担心移动医疗 APP 收集太多有关于我的个人信息	

本研究影响因素包括个体特征、个人隐私倾向、个人经历、媒体报道、社会制度、医疗信息敏感度、感知有用性，由于个体特征测量的标准无法统一，则不能使用量表的形式进行测量，其余影响因素的量表具体情况如表 2：

表 2 移动医疗 APP 用户隐私关注影响因素测量量表

变量	题项	来源
个人隐私倾向	1.我很担心自己的隐私信息被别人看到 2.即使有良好的移动医疗 APP 的体验经历，仍然无法减轻我把隐私信息放到网上的焦虑程度 3.我是一个非常在乎个人隐私的人，在移动医疗 APP 上做什么事都是隐私至上	M.Z. Yao, R.E .Rice, K. Wallis(2007)
个人隐私经历	1.如果我使用的移动医疗 APP 存在安全问题而导致我的个人信息受到威胁或泄露，我会感到困扰 2.如果我使用的移动医疗 APP 因追求利益而导致我的个人信息受到威胁或泄露，我会感到困扰	Coye Cheshlre, JuddAntln, &Ellzabeth.C(2010)；LaRose, R, Rifon, N, Liu, S , &lee, D(2005) Preece,J(2004)

	3.如果遇到骚扰电话等类似经历，我会感到困扰	
媒体报道	1.我会关注网络隐私泄露的负面报道 2.网络隐私泄露的负面报道，会让我在使用中感到担心 3.网络隐私的负面报道会让我减少对移动医疗 APP 的使用	白文君（2016）
社会制度	1.我认为我国的隐私保护相关法律比较完善 2.我认为我国的隐私保护相关法律很好的保护个人隐私信息 3.我认为政府在隐私方面对企业的监督管理做的很好 4.我认为目前移动医疗 APP 不会滥用它所拥有的用户个人信息	Malhotra (2004)
医疗信息敏感度	1 从个人隐私角度而言，我认为在网络环境下我真实的医疗信息是非常敏感的 2 我觉得我的医疗信息是非常重要的 3 网络环境下，我觉得我的医疗信息被泄露将会给我造成严重的损失或伤害	白文君（2016）
感知有用性	1 拥有我个人信息时，移动医疗 APP 会为我产生更大的帮助 2 使用我个人信息的移动医疗 APP 会为我提供更优质的服务 3 移动医疗 APP 对我是有用的	Xu (2008)

4.2 数据收集

由于疫情原因，本次研究采取网上发放问卷的方式来收集数据。编制成问卷星格式，通过社交网络方式发放问卷。最后共统计收回 480 份问卷。由于系统无法筛选出无效问卷，手工对无效问卷进行剔除，最后得到有效问卷 442 份。

5 数据分析与假设检验

5.1 描述性统计

本次调查共收集到 442 份有效样本，以性别、年龄、学历、职业、频率、次数以及是否愿意提供隐私信息、是否担心隐私信息被泄露等方面对被调查者的基本情况进行描述分析，具体分析结果见表 3。

表 3 被调查者基本情况描述分析

属性	类别	人数	百分比
性别	男	172	38.9
	女	270	61.1
年龄段	18 岁以下	14	3.2
	18-24 岁	319	72.2
	25-34 岁	68	15.4
	35-44 岁	20	4.5
	45-54 岁	15	3.4
	55 岁以上	6	1.4
学历	高中及以下	38	8.6
	专科	38	8.6
	本科	346	78.3
	硕士及以上	20	4.5
职业	国家机关、党政群体、企事业单位负责人	79	17.9
	专业技术人员	4	0.9
	企业相关人员	22	5
	商业、服务业人员	15	3.4
	个体户/自由职业者	282	63.8
	在校学生	8	1.8
	务农	32	7.2
频率	小于等于 1 次	140	31.7
	2-3 次	243	55
	4-6 次	42	9.5
	7-10 次	12	2.7
	大于 10 次	5	1.1
次数	1-3 次	265	60
	4-6 次	101	22.9
	7-9 次	40	9
	10 次以上	36	8.1

是否愿意提供隐私信息（疫情防控期间）	不愿意	44	10
	比较不愿意	261	59
	一般	84	19
	比较愿意	34	7.7
	愿意	19	4.3
是否担心隐私信息被泄露（未来大数据时代）	是	345	78.1
	否	97	21.9

5.2 信度与效度检验

信度采用克朗巴哈（Cronbach's Alpha）信度系数来进行验证。信度系数大于 0.8，则说明量表的信度非常好，信度系数大于 0.7，量表可以接受的。信度系数在 0.6 以上，量表需要进行修改，0.6 以下时，量表则不能接受，需要重新设计。

根据检验结果，个人隐私倾向、个人隐私经历、媒体报道、社会制度、医疗信息敏感性、感知有用性和隐私关注的 Cronbach's Alpha 值在 0.796 到 0.938 之间，都大于 0.7，问卷的整体信度好。

效度分析是检验量表的有效性，检验量表是否能把潜变量区分开。效度检验有两种方法，一个为探索性因子分析法，另一个为验证性因子分析。在做探索性因子分析之前，先使用 KMO 和 Bartlett's 球形检验测量量表能否做因子分析，KMO 测量值接近于 1，Bartlett's 检验值的显著性小于 0.001，则说明样本能进行因子分析具体分析结果，经过分析可知 KMO=0.928，Bartlett's 球形值显著（Sig.<0.001）,说明数据能够进行因子分析。

针对收集汇总的数据把 28 个观察变量划分为几个公因子，以特征值大于 1 的条件提取公共因子，抽取 7 个因子代表 28 个观察变量的大部分信息。从表 4 中可以看出通过因子分析，一共得到 7 个因素，其总体解释能力达到 71.802%，达到研究的要求。即表明这 7 个因素具有良好的代表性。各个测量项的因素负荷量均大于 0.5，且交叉载荷均小于 0.4，说明该量表具有良好的结构效度。

表 4 因子分析结果

题项	成分						
	隐私关注	社会制度	媒体报道	个人隐私倾向	个人隐私经历	感知有用性	信息敏感度
PC9	0.82	-0.168	0.142	0.093	0.108	-0.091	0.132
PC1	0.798	-0.145	0.046	0.144	0.11	-0.065	0.121
PC5	0.778	-0.069	0.051	0.149	0.144	-0.093	0.103
PC8	0.767	-0.147	0.142	0.036	0.128	-0.153	0.17
PC7	0.764	-0.103	0.072	0.152	0.059	-0.148	0.17

PC6	0.76	-0.181	0.112	0.101	0.119	-0.077	0.165
PC4	0.742	-0.146	0.118	0.145	0.017	-0.098	0.153
PC3	0.736	-0.131	0.05	0.133	0.125	-0.144	0.094
PC2	0.721	-0.169	0.137	0.08	0.072	-0.179	0.095
SS4	-0.237	0.827	-0.095	-0.125	-0.108	0.077	-0.088
SS1	-0.22	0.798	-0.099	-0.096	-0.115	0.076	-0.105
SS3	-0.2	0.772	-0.169	-0.109	-0.067	0.069	-0.084
SS2	-0.153	0.767	-0.048	-0.126	-0.136	0.113	-0.094
MC1	0.12	-0.091	0.842	0.148	0.089	-0.176	0.122
MC3	0.174	-0.185	0.834	0.09	0.076	-0.108	0.055
MC2	0.189	-0.097	0.783	0.077	0.177	-0.049	0.127
PPT1	0.188	-0.129	0.09	0.825	0.051	-0.088	0.123
PPT3	0.214	-0.126	0.084	0.795	0.121	-0.168	0.157
PPT2	0.224	-0.173	0.146	0.758	0.081	-0.083	0.051
PPE3	0.156	-0.15	0.165	-0.005	0.807	-0.136	0.149
PPE1	0.121	-0.111	0.186	0.124	0.794	-0.066	0.061
PPE2	0.228	-0.13	-0.007	0.123	0.781	-0.099	0.147
PU2	-0.205	-0.05	-0.144	-0.072	-0.076	0.81	-0.003
PU3	-0.251	0.206	-0.091	-0.203	-0.119	0.783	-0.116
PU1	-0.198	0.212	-0.102	-0.088	-0.122	0.761	-0.113
IS1	0.267	-0.122	0.128	0.144	0.088	-0.089	0.788
IS3	0.244	-0.141	0.116	0.042	0.145	-0.027	0.786
IS2	0.251	-0.084	0.072	0.152	0.132	-0.1	0.721
特征值	6.087	2.99	2.337	2.238	2.188	2.151	2.112
方差百分比	21.741	10.679	8.348	7.994	7.813	7.683	7.545
累积 %	21.741	32.42	40.767	48.761	56.574	64.258	71.802

由表 5 可知，CMIN/DF 达到要求，其中 GFI、AGFI、IFI、CFI、NFI 也都达到研究的标准，RMR（残差均方和平方根）为 0.044，小于 0.08，RMSEA（近似误差均方根）也处于理想水平，因此各个拟合指标均符合研究标准，说明模型的拟合程度好。

表 5 验证性因素模型拟合度

模型拟合指标	统计值	最优标准值	拟合情况

CMIN	475.169	——	——
DF	329	——	——
GFI	0.930	>0.8	好
AGFI	0.914	>0.8	好
IFI	0.979	>0.9	好
CFI	0.979	>0.9	好
NFI	0.934	>0.9	好
CMIN/DF	1.444	<3	好
RMR	0.044	<0.08	好
RMSEA	0.040	<0.08	好

组成信度和平均方差抽取（AVE）也是评价变量聚合效度重要指标，从表 6 可知各个测量指标标准化因素负荷均大于 0.6，平均变异萃取量（AVE）超过 0.5，组成信度（CR）都大于 0.7，表明每个变量对于收敛都具有高度相关性。

表 6 因子载荷和变量组成信度

变量	题项	因素负荷	AVE（平均变异萃取量）	CR（组成信度）
个人隐私倾向	PPT1	0.79	0.614	0.826
	PPT2	0.719		
	PPT3	0.837		
个人隐私经历	PPE1	0.718	0.587	0.809
	PPE2	0.748		
	PPE3	0.827		
媒体报道	MC1	0.845	0.65	0.847
	MC2	0.74		
	MC3	0.829		
社会制度	SS1	0.807	0.632	0.872
	SS2	0.734		
	SS3	0.766		
	SS4	0.866		
医疗信息敏感度	IS1	0.819	0.572	0.8
	IS2	0.689		
	IS3	0.755		
感知有用性	PU1	0.737	0.602	0.817
	PU2	0.667		

	PU3	0.905		
	PC1	0.818		
	PC2	0.749		
	PC3	0.752		
	PC4	0.765		
隐私关注	PC5	0.775	0.629	0.938
	PC6	0.805		
	PC7	0.785		
	PC8	0.816		
	PC9	0.866		

6 路径分析和假设检验

结构方程模型能够对抽象的概念进行估计与鉴定，分为测量模型和结构模型。首先构建结构模型，确定结构模型中所涉及的所有潜变量以及各个变量之间的关系。本研究设置了 7 个潜变量。通过之前的研究可以得出，本调查问卷通过信度和效度检验，可采用 AMOS23.0 对模型进行验证和分析，确定各个影响因素对移动医疗 APP 用户隐私关注的影响程度。根据已有研究和相关假设构建研究模型，具体见图 2：

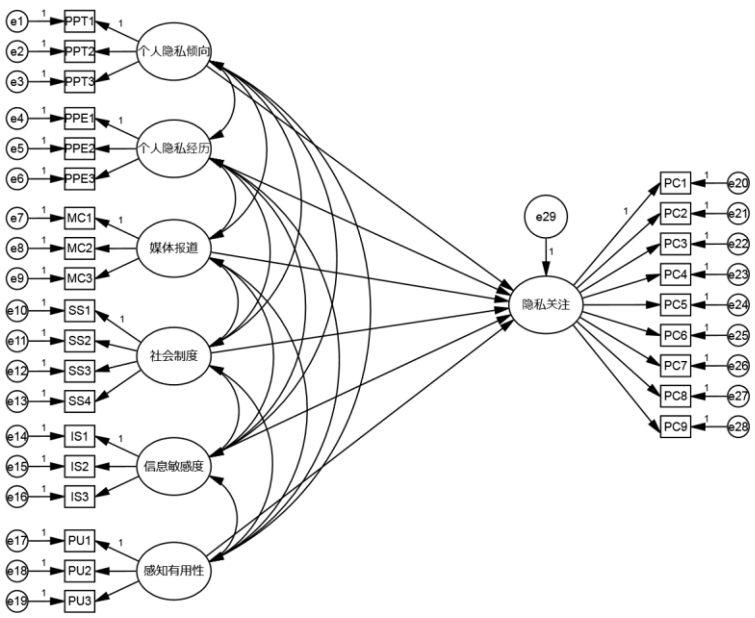


图 2 本研究的结构方程模型

经结构方程模型的检验，从表 7 可知个人隐私经历和媒体报道对隐私关注的假设不成立；其余的假设都成立。

表 7 路径系数

路径			标准化系数	非标准化系数	S.E.	C.R.	P	假设
隐私关注	<---	个人隐私倾向	0.133	0.126	0.052	2.449	0.014	成立
隐私关注	<---	个人隐私经历	0.07	0.087	0.067	1.312	0.189	不成立
隐私关注	<---	媒体报道	0.051	0.049	0.048	1.026	0.305	不成立
隐私关注	<---	社会制度	-0.181	-0.171	0.049	-3.525	***	成立
隐私关注	<---	信息敏感度	0.319	0.308	0.055	5.559	***	成立
隐私关注	<---	感知有用性	-0.199	-0.21	0.056	-3.75	***	成立

7 研究结论

本次研究探索影响移动医疗 APP 用户隐私关注的影响因素，以上研究结果表明，个体隐私倾向、社会制度、医疗信息敏感度、感知有用性对用户隐私关注有显著影响。个人不良事件经历和媒体报道对隐私关注没有显著影响。

7.1 证实观点讨论

（1）个体隐私倾向正向影响移动医疗 APP 用户的隐私关注

移动医疗 APP 用户的个人隐私倾向对隐私关注的标准化路径系数为 0.133，这表明隐私倾向偏向保守的人，隐私关注的程度会比一般人或隐私倾向偏向开放的人高，随之对自身的隐私控制权利的要求也会变高。这与之前研究结论一致，一个人对隐私的在意程度越高，则对隐私信息关注程度越多[15]。

（2）社会制度负向影响移动医疗 APP 用户的隐私关注

本研究社会制度对隐私关注的标准化路径系数为-0.181。这说明社会制度负向影响移动医疗 APP 用户的隐私关注。政府监管情况的好坏与互联网消费者隐私关注呈负相关。表明政府的监管越弱，用户越会加强对个人隐私的关注[17,25]。用户隐私的相关保护法律、政策越完善，用户的隐私的关注程度越低。因而，政府和相关的企业可以通过完善法律和规章制度对用户的隐私进行有效的保护，从而可以减少用户对隐私的关注程度。

（3）医疗信息敏感度正向影响移动医疗 APP 用户的隐私关注

本次研究移动医疗 APP 用户的医疗信息敏感度对隐私关注的标准化系数为 0.319，这表明医疗信息敏感度正向影响移动医疗 APP 用户的隐私关注。在医疗信息、简历信息、联系方式这三种信

息中进行对比时，发现人们更愿意提供联系方式，其次是简历信息，最后是医疗信息。移动医疗APP 的医疗信息敏感度越高，那用户对隐私的关注程度就越高，越不愿意提供自己的相关医疗信息。这与前人研究结论一致，医疗信息的敏感度与提供个人健康信息意愿呈负相关[24

（4）感知有用性负向影响移动医疗 APP 用户的隐私关注

本研究移动医疗 APP 用户的感知有用性对隐私关注的标准化系数为-0.199，表明移动医疗 APP 用户的有用性感知越强烈，隐私关注会越低。这与 Barkhuus &Dey（2003）的结论一致，他认为，当一些服务被认为是有用的，用户使用产品过程中的隐私关注水平会有所降低。因此，相关的企业可以通过提高技术，方便用户从 APP 中能够获得更多的便利，降低用户对隐私的关注程度。

7.2 未被证实的观点讨论

个人的不良事件经历和媒体报道对移动医疗 APP 用户的隐私关注未证实有显著影响。提高用户隐私关注是一个循序渐进的过程，当用户经历的事情越多时，用户才会更加提高注意力。移动医疗 APP 环境中，多数用户经历隐私泄露事件较少，尚不足以提高用户隐私关注。媒体报道不同受众关注度与敏感度不同，其对用户隐私关注的影响尚有待于进一步观察。

全文刊登

PU-001

新时代下高校图书馆面对重大公共卫生事件应急响应及疫情常态化下服务策略研究

许丹
中国医科大学图书馆

目的 以2020年初爆发的新型冠状病毒肺炎疫情为背景，探讨面对重大公共卫生事件高校图书馆做出的应急响应服务情况。

方法 调研42所“双一流”高校图书馆疫情响应速度和响应内容并展开相关分析。

结果 高校图书馆高度关注疫情变化并及时启动了应急响应措施，同时开展系列相关“云”信息服务。同时利用图情专业知识技能进行情报支撑以及疫情常态化下图书馆如何开展各项服务

结论 总结高校图书馆高校图书馆面对新的突发公共卫生事件快速响应支撑及疫情常态化下的服务开展提供参考借鉴，探索高校图书馆新的服务方向。

引言

2020年伊始，一场突如其来的病毒席卷中华大地，在不到两个月的时间，新型冠状病毒肺炎从武汉蔓延至全国。为有效应对疫情，党中央迅速作出部署，要求各级党委和政府及有关部门把人民群众生命安全和身体健康放在第一位，采取切实有效的防控手段和措施，坚决遏制疫情蔓延势头。全国人民积极贯彻落实习近平总书记“科学防治、精准施策”的重要指示精神，以战疫必胜的坚定信念，上下一心，团结一致，在3月底取得阶段性胜利。在此期间，广大高校图书馆也发挥了积极作用。

目前国内疫情在我国党中央的领导下得到有效控制，国外疫情状况仍不容乐观。习主席提到了“人类命运共同体”的概念，病毒不分国界，全球在抗击病毒疫情面前不分彼此，应该联合起来共同抵抗未知病毒给人类带来的困苦。在此期间，国内、外高校图书馆也在面对疫情发挥着各自的积极作用。

疫情时期，美国康奈尔大学图书馆开启了系列应对措施，在图书使用权限上开启一对多借阅，出版商也临时免费，北美大型图书馆扩大紧急状态下合理使用权限，馆员提高软件、硬件服务应变技能，跨时区、跨地域、跨国度提供服务保障。美国密西根大学在2019年末时更新了“灾难响应和回复计划”，对包括火灾、水等在内的图书馆馆藏灾难列出一份长达 102 页的详细通用说明，供读者下载查看。

我国高校图书馆始终坚持强化自身建设，也积极参与应对重大公共卫生事件。早在2003年非典时期，杨继贤就介绍了香港岭南大学图书馆在非典型肺炎事件中的危机管理和服务工作经验，提出了“非典”对图书馆工作的启示。在2009—2010年H1N1流感爆发时期，史惠媛对高校图书馆在传染病突发公共卫生事件频发的情况下危机管理模式进行了研究，指出提高危机意识、加强组

织领导、建立健全规章制度、鼓励全员参与、完善环境保障体系，是高校图书馆有效应对传染病突发公共卫生事件的重要途径[2]。在2015年H7N9禽流感疫情时期，朱虹[3]介绍了医学院校图书馆积极应对突发公共卫生事件的措施，包括建立健全应急预案和服务机制、培养主动超前的服务意识、建立专题数据库、密切相关单位的沟通联系、加强网络的建设和维护、加强馆员队伍的素质建设等。2020年新冠肺炎疫情时期，王波[4]对全国高校图书馆疫情防控期间服务创新情况进行调研，总结了高校图书馆为保护读者和馆员健康，保障高校在线教学和科研顺利开展而推出的一系列应急措施和创新型服务。

由此可见，伴随着科学技术的飞速发展，图书馆参与应对公共卫生事件的程度越来越高，发挥的作用也越来越显著。笔者对新冠肺炎疫情期间全国42所“双一流”高校图书馆的响应速度和响应内容进行了调研分析，并对新时代高校图书馆在重大公共卫生事件发生时采取的应急服务措施进行归纳总结。

1．42所“双一流”高校图书馆应急响应调研情况分析

笔者先对42所双一流高校图书馆官网平台进行调研，除2所网站受网络限制原因打不开外，其他高校图书馆网站均在首页的新闻动态、通知公告、最新资讯中发布了疫情防控信息，随后陆续推出应急服务信息资源，做到响应率占比95%。

随后对全国42所“双一流”高校图书馆微信公众号进行了查询，调研各图书馆微信公众号面对重大公共卫生事件所展现出的应急响应速度、采取的措施，以及发布信息的种类、条数、周期和特色等数据，并进行分析、探讨、总结，见表1和图1。（统计数据起始时间为2020年1月20日至2020年3月31日。）

表 1　42 所“双一流”高校图书馆在新冠肺炎疫情期间应急响应情况及资源推介

序号	高校名称	官方网站	微信推送	首发日期	通知问候	阅读推荐	服务与资源介绍	疫情期应急服务信息	信息总数	特色
1	北京大学	√	√	1.31	2	3	20	23	48	每周更新，抗“疫”服务模块
2	清华大学	√	√	1.27	1	0	30	33	64	疫期免费资源
3	北京师范大学	√	√	1.29	5	14	36	23	78	每天更新，学习战“疫”系列小图微课
4	中国农业大学	√	微信停更	–	–	–	–	–	–	–
5	北京理工大学	√	微信停更	–	–	–	–	–	–	–
6	中国人民大	受限	√	2.4	2	1	10	38	51	每周更新，同心

	学								战“疫”
7	中央民族大学	√	√	2.1	2	1	10	22	35	每周更新，共同战“疫”书香战“疫”
8	北京航空航天大学	√	√	1.28	0	0	2	22	24	每周更新
9	上海交通大学	√	√	1.27	2	4	30	24	60	每天更新，知识战“疫”
10	同济大学	√	√	2.3	3	2	13	4	22	每周更新
11	复旦大学	√	√	1.26	9	4	6	36	55	每天更新
12	华东师范大学	√	√	2.3	4	7	32	26	69	每天更新
13	南京大学	√	√	1.28	5	10	19	18	52	
14	东南大学	√	√	1.22	9	14	16	19	58	每天更新，读书战“疫”
15	浙江大学	√	√	1.28	3	16	37	12	68	每天更新
16	山东大学	√	√	1.31	10	22	26	5	63	2月每周更新，3月后每日更新
17	中国海洋大学	√	√	1.29	2	0	2	22	26	众志成城共克时艰
18	武汉大学	√	√	1.23	5	18	15	23	61	每天更新
19	华中科技大学	受限	√	1.23	2	2	14	4	22	
20	中南大学	√	√	1.27	2	8	11	9	30	
21	湖南大学	√	√	1.27	8	7	19	19	53	
22	国防科技大学	√	√	1.29	4	20	6	6	36	战“疫”不孤读
23	中国科技大学	√	√	1.28	2	8	17	7	34	-
24	吉林大学	√	√	1.29	3	16	45	19	83	宅家科研之资源推介，宅家读书
25	天津大学	√	√	1.24	10	70	59	26	165	共读战“疫”
26	南开大学	√	√	1.28	10	26	49	10	95	-足不出户尽享免费资源

27	电子科技大学	√	√	1.27	5	9	13	9	36	抗“疫”服务专栏
28	四川大学	√	√	2.1	1	4	4	8	17	-
29	华南理工大学	√	√	1.23	3	7	30	6	46	-
30	兰州大学	√	√	1.27	15	13	75	40	143	每天更新
31	西北工业大学	√	√	1.26	1	9	12	18	40	“众志成城共克时艰”疫情特供
32	哈尔滨工业大学	√	微信停更	-	-	-	-	-	-	-
33	重庆大学	√	√	1.25	3	5	1	6	15	-
34	大连理工大学	√	√	2.1	3	4	6	26	39	每周更新，战“疫”修炼指南
35	中山大学	√	√	1.26	9	22	46	39	116	免费资源共度时艰
36	西北农林科技大学	√	√	1.28	0	0	12	6	18	
37	东北大学	√	√	1.22	5	7	36	14	62	“疫”起闭关，“宅”家修炼
38	厦门大学	√	√	1.28	6	11	39	10	66	战“疫”书单
39	西安交通大学	√	√	1.26	2	2	11	15	30	“图”结一心，共克时艰
40	郑州大学	√	√	1.26	5	11	32	9	57	-
41	云南大学	√	√	1.28	4	13	30	26	73	战“疫”主题书柜
42	新疆大学	√	√	1.29	1	8	22	19	50	-
	合计				168	398	893	701	2160	

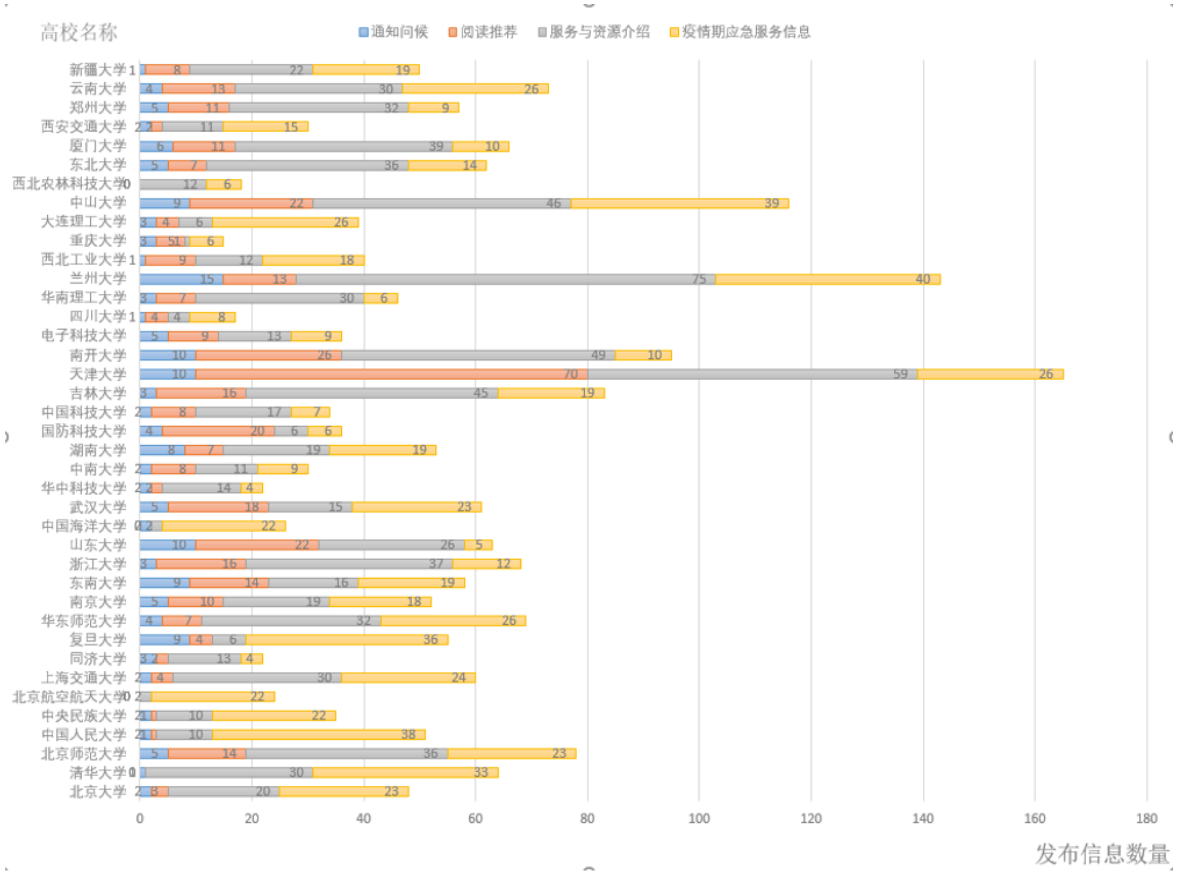


图 1　39 所高校图书馆发布信息情况一览图

调研的 42 所“双一流”高校图书馆中，除了 3 所高校图书馆微信公众号停止更新外，其他 39 所高校图书馆都在疫情期发布了应急服务措施，应急响应率占比 92.9%。

1.1 响应速度

从面对疫情的响应时间来看，在 1 月 22 日至 1 月 23 日发布消息的图书馆有 5 所，1 月 24 日至 1 月 31 日发布消息的图书馆有 28 所， 2 月 1 日至 2 月 4 日发布消息的有 6 所（图 2）。东南大学图书馆和东北大学图书馆在 1 月 22 日最先开始启动应急响应，截止到 2 月 4 日，有 39 所高校图书馆启动应急响应。可见我国高校图书馆高度关注全国疫情变化，大多及时启动了应急响应措施。39 所高校图书馆应急响应增速见图 2。

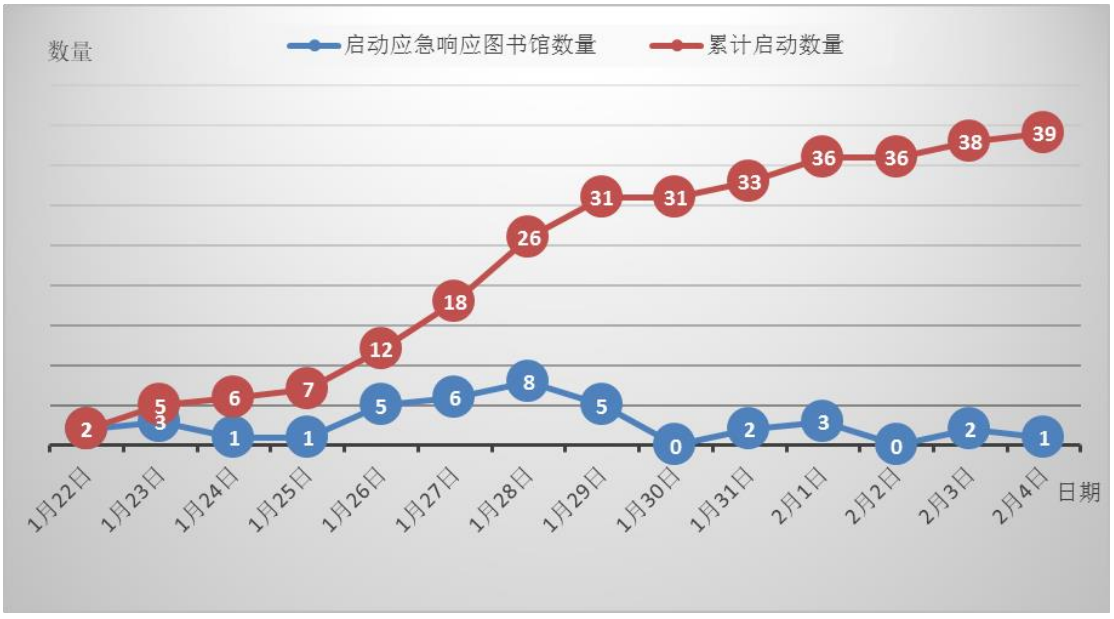


图2 疫情突发后启动应急响应的图书馆数量日变化情况

1.2 响应内容

如表 1 所示，在对各高校图书馆微信公众号在疫情期间发布的信息进行分析后，笔者主观将其归纳为通知问候、阅读推荐、服务与资源介绍、疫情期应急服务信息 4 类，这与各高校图书馆发布信息的真实意向归属可能会有差异，势必造成一定信息归属类别的误差。例如，笔者将疫情期间各图书馆发布的第一条“临时闭馆通知”归类在“疫情期应急服务信息”中，将其他通知类信息归属在“通知问候”中；将有关各项资源推介消息里含有战“疫”、抗“疫”、“疫情期”等与疫情有关字样的资源也归属在“疫情期应急服务信息”中，将其他普通类别资源推介信息则归属在“服务与资源介绍”中。此外，将带有疫情字样的相关讲座、报告、疫情相关书籍资料汇总及宣传指南、活动信息等均归入“疫情期应急服务信息”中。

总体来说，42 所“双一流”高校图书馆面对疫情，多数能够积极主动响应，及时发布应急信息，采取闭馆等措施遏制疫情蔓延。各图书馆针对疫情的首发消息多是在疫情期特殊情况下的临时闭馆通知，随后发布各种数据库资源的免费及远程使用方式方法，各大资源覆盖内容全面广范，最大程度为读者获取信息资源提供便利。在 1 月 20 日至 3 月 31 日期间，39 所高校图书馆累计发布消息 2160 条，其中通知问候类消息 168 条，占比 7.8%；阅读推荐类消息 398 条，占比 18.4%；各种信息服务及资源推荐介绍 893 条，占比 41.3%，；疫情期的应急服务信息 701 条，占比 32.5%，如图 3 所示。在此期间，发布信息超过 100 条的有 3 所学校，分别是天津大学图书馆发布 165 条信息，兰州大学图书馆发布 143 条信息，中山大学图书馆发布 116 条信息。其中发布“疫情期应急服务信息”最多的三所学校分别是兰州大学图书馆 40 条，中山大学图书馆 39 条，中国人民大学图书馆 38 条。

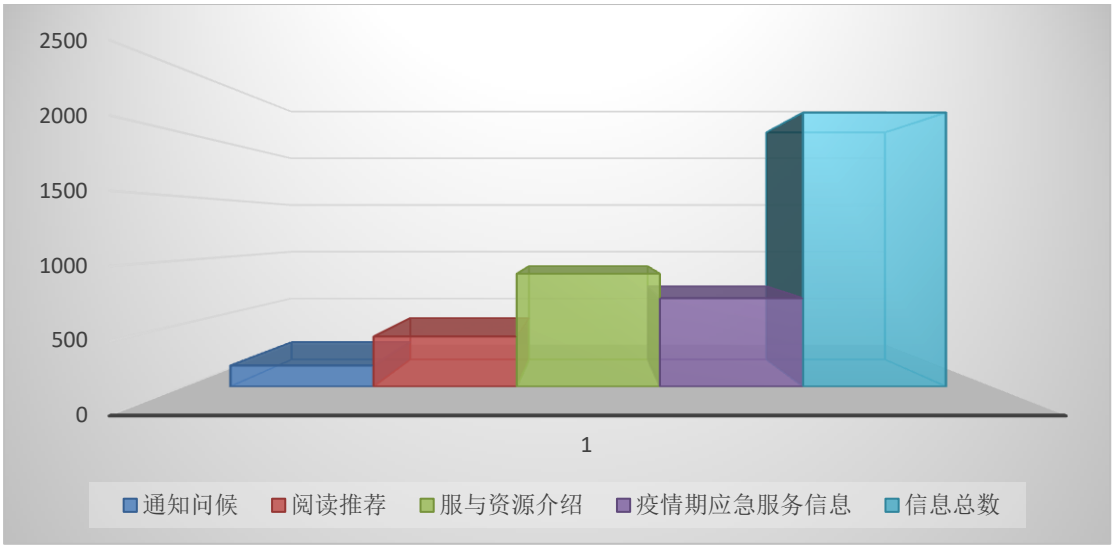


图 3 疫情期间发布消息分类情况图

在发布信息的更新频率方面，有些图书馆采取每周更新，有些图书馆是每日更新，也有些图书馆在假期前半程采取周更新，在后半程时间里采取日更新。无论是哪种方式都为各类读者提供了纷繁多样的数据库信息资源，为读者在家自学和科研提供了资源保障。

在发布信息的特点方面，主要可分为以下几个类别：一是在微信公众号中特别设置了服务专栏，例如北京大学图书馆和电子科技大学图书馆设置了“抗疫服务”专栏模块，专门提供疫情期间特别服务，如北京大学图书馆的“宅家抗疫，免费资源新体验”。二是针对疫情期的免费数据库资源推介，每所图书馆精敲细琢、别具匠心地为读者提供了系列资源信息。如中国人民大学图书馆的“同心战‘疫’”系列、中国海洋大学图书馆和西北工业大学图书馆的“众志成城、共克时艰”疫情特供资源系列、大连理工大学图书馆的战“疫”修炼指南系列、东北大学图书馆的“‘疫’起闭关，‘宅’家修炼”资源推介系列等。三是在疫情期间进行各种图书推广、介绍的读书相关系列资源推介，包括武汉大学图书馆“封城阅读计划”、东南大学图书馆“读书战‘疫’”、国防科技大学图书馆“ 战‘疫’不孤‘读’”、天津大学图书馆“共读战‘疫’”、中央民族大学图书馆“书香战‘疫’图书推介”、西北农林科技大学图书馆“打卡战‘疫’以读攻‘毒’”、厦门大学图书馆“战‘疫’书单”。此外，北京航空航天大学图书馆还推出了“北航云图书馆Bc library”为全校师生的学教学科研保驾护航，云南大学图书馆推出了战“疫”主题特色书柜。这些新颖的主题系列，是高校图书馆全体馆员智慧的结晶，体现了图书馆对战胜疫情的必胜信心和服务读者的使命初心。

2 高校图书馆应对重大公共卫生事件应急服务措施

2.1 成立疫情防控领导小组

加强领导，提供组织保障。建立由党政领导班子成员和各部门主任组成的疫情防控领导小组，明确分工，落实责任，点对点联系到每名馆员，精准排查，做到图书馆日常工作与疫情防控工作的同部署、同推进、同落实。制定疫情防控期间的工作方案和应急预案，组织对全体馆员进行视

频培训，开展桌面应急推演，及时发现问题，纠错整改，逐步完善。同时强化责任管控，严防死守，细化分工，采取有力举措全面开展防疫工作，保证图书馆各项防疫措施的及时落实，积极做好防控工作与资源服务保障工作[6-7]。例如 1 月 27 日，山西大学图书馆及时制定了《山西大学图书馆疫情防控措施》，成立了含馆领导、办公室主任、工会主席、党组成员在内的疫情防控工作领导小组，通过微信平台协调安排本单位防控工作[8]。山西财经大学图书馆成立了图书馆疫情防控工作领导小组，构建了“馆长-分管领导-部门主任”三级联防联控机制[8]。

2.2 全面闭馆降低疫情爆发风险

高校图书馆属于人员密集场所。在疫情发生后，为有效避免因人群聚集可能带来的病毒传播风险，最大限度地保障读者和一线工作人员的身体健康和生命安全，实行全面闭馆，暂停到馆读者服务[6]。身在外地学生暂不返校，本地师生不到校，统筹安排图书馆暂停开放期间的工作安排，严格闭馆防止疫情的传播扩散。例如武汉大学、华南理工大学早在 1 月 23 日便发布了暂时关闭图书馆的闭馆通知。

2.3 开展疫情期间图书馆“云服务”

高校图书馆作为高校内的服务机构，面对疫情，推出一系列线上“云服务”，包括数据库“云使用”，参考咨询“云办公”，辅助教师“云教学”等。同时，在图书馆手机客户端不时推送疫情信息，介绍数据库资源，推介优秀图书等，响应读者需求，疏导读者疫情期间焦虑情绪。比较有代表性的比如上海交通大学图书馆“知识战‘疫’”系列资源推介、北京师范大学图书馆“21 天读书战‘疫’”图书推介、浙江大学图书馆“抗‘疫’不停学，求真一小时讲座资源”资源推介系列等等。

2.4 制定抗疫后期服务预案

听从指令服从安排，在疫情得到有效控制后各高校图书馆将在确保安全的前提下逐渐恢复开放。继续加强宣传与防控，通过微信、网站、展板等多种形式和渠道，推送疫情防控相关信息，教育、引导读者和员工加强自我防护，养成健康生活习惯。开馆前，对公共区域、设施进行全面消毒，对返还图书杀毒杀菌，对进出图书馆的各类人员严格登记、测温，对馆内设施物品做到每日不定期消毒，对馆舍进行定期通风。同时加强协调保障，确保相应消毒用品、防控设备及时配置到位；开馆后，加强对入馆读者管理，核定入馆人数，错峰进入。做好进馆人员佩戴口罩和测量体温工作。对阅览自修座位进行适当调整，增加间距，避免读者在局部区域聚集；对每个区域全天进行开窗通风，做好图书馆环境卫生整治。例如安徽大学图书馆制定了《新学期开学新冠病毒感染肺炎疫情防控工作方案》[7]，天津大学、南开大学图书馆分别制定了《关于疫情过后重新开馆卫生防护的临时措施》、《复工开馆工作预案》[9]。

2.5 实行疫情日报告、零报告制度，全面掌握职工动向

严格落实学校各项防控措施，通过微信提交上报疫情日报表，引导员工严格遵守学校和所在地区防控措施要求，做好自我防护。提醒督促每位员工按时做好健康打卡信息上报；动态掌握职工寒假期间行程动向及健康情况，对在外地探亲、出差等尚未返回人员，通过网络、电话、实时

定位等形式保持联系，注重加强宣传引导和心理疏导，特别是从湖北返回的工作人员做好自主居家隔离观察。例如我馆要求每位馆员利用企业微信或者微信平台每天 12 点前填报新冠肺炎防控日报表上报个人信息，包括个人健康状态和所在位置，便于上级领导及时掌握每个人的动态。

2.6 加强疫情防控宣传教育

通过图书馆网站、QQ 办公群、微信等网络渠道，及时传达政府部门疫情防控精神文件以及疫情相关防控指南，部署疫情防控工作，开展健康教育。向教职工普及疫情防控知识和防控要求，密切关注疫情防控形势变化，及时传达学校通知和工作安排，确保全体工作人员在第一时间接收通知。2020 年 2 月 12 日，广东省高校图工委以视频会议的形式召开会议，强调在疫情防控期间要全力做好在线教学组织与平台服务保障工作，并对《广东省高等学校和中职业图书馆新型冠状病毒肺炎预防控制指引》展开讨论[10]。积极宣传抗“疫”先进典型和我国抗击疫情取得的阶段性成效，激发学生爱国情怀，引导学生全面科学正确对待疫情，增强信心，众志成城、共克时艰。

3 疫情时期图书馆的“云”信息服务

“少聚集”是疫情防控的一个重要举措，网络“云相聚”应运而生。疫情期间，各行各业“云相聚”已成为新热点。因疫情推迟开学，为了将影响降到最低，各高等院校改变传统教学模式，在网络上开展“云教学”，高校图书馆也随之启动“云服务”。

3.1 数据资源“云使用”

在疫情爆发伊始，中国知网数据库（CNKI）、万方数据库、中国生物医学文献数据库（Sinomed）等几大中国数据商即在各自网络平台和微信公众号发布了疫情期间免费使用数据库的通知，有的面对高校科研人员开放，也有的将试用时间延长到疫情结束。随后，各高校图书馆陆续在手机客户端推出介绍了各种外文文献数据库、中文文献数据库，各种学习科研数据库、电子图书数据库、多媒体数据库等资源的登陆使用方式，供读者在疫情期间免费使用。此外，图书馆通过多种方式大力宣传推广通过登录学校 VPN 系统来获取图书馆订购的全部电子资源，极大满足了疫情防控期间师生对图书馆服务的需求，为特殊时期的教学科研、人才培养提供了信息支撑。例如我馆推出的“远程数据库资源访问攻略”“指尖上的学习资源”等资讯专栏都是针对疫情期提供的数据库使用资源。同时多所图书馆全面梳理校外访问图书馆已购电子资源的常规途径，积极对接 CARSI（中国教育科研计算机网统一认证与资源共享基础设施） 等平台和远程访问站点，开辟校外访问资源新途径，同时开通区域资源文献传递系统包括 CASHL、BALIS 等，来实现文献资源获取及共享建设[11]。

3.2 参考咨询“云办公”

参考咨询服务，是图书馆作为科研技术支撑机构的重要体现。为了不影响正常对外服务，各图书馆陆续推出疫情期间“云办公”服务。例如我校图书馆参考咨询部专门开设远程网络办公服务，正常受理参考咨询、科技查新、查收查引和论文查重等申请。与非疫情期间不同，用户不必到学校提交纸质申请，只需要通过电子邮件发送委托申请，然后通过图书馆手机客户端确认委托事项即可。图书馆每天安排一名工作人员在线值班，及时受理申请，并接待线上咨询，处理有关

事宜，图书馆其他工作人员则保持弹性工作制度，在家通过登陆学校 VPN 系统使用数据库完成科技查新等任务。为减少疫情期间不必要的人员聚集，图书馆规定每周五由一名工作人员到校集中处理本周工作，出具已完成的检索证明报告和查新报告，并邮寄给客户。客户全程不需要到图书馆，减少了病毒传播风险。

3.3 辅助教师“云教学”

为贯彻教育部“停课不停学”的指示，各大高校开始尝试在线“云教学”。为了配合学校线上教学模式，迅速加强在线资源整合、服务和利用，为疫情期间线上教学和居家科研做了大量信息资源应急保障工作。在此期间，图书馆免费为各学院老师收集整理了电子教材、视频教学等参考资料，指导教师线上教学操作技能，为学校开展“线上开学”“线上教学”提供有力保障。同时，图书馆也开始在平台推出线上信息素养教育课程的教学。例如安徽医科大学图书馆[7]自 2 月 9 日以来为本科生开展网络教学，老师们认真备课并自学了各种教学平台的使用方法，不断尝试采用直播、录播、屏幕分享等方式授课，努力克服网络带来的弊端，通过视频为学生答疑解惑，督促学生按时完成学习任务。

传统教学资源以教材、教参服务为主，通过线下到馆面对面借阅来实现，已逐渐被新的服务方式所代替。特别是在特殊时期下，越来越多高校图书馆在尊重知识产权的前提下，通过整合本馆或网上开放资源，包括电子教科书、教学辅助资料，课件等资源为师生提供高质量的教学资源服务。例如上海交通大学图书馆、四川大学图书馆、电子科技大学图书馆等高校图书馆快速实施上线专门的“教学资源服务平台”，提供教学计划急需的教材和教参资料[12-13]。

4.运用图情领域科研方法，分析探测疫情情报学热点

2019 年 12 月，我国陆续出现新冠肺炎病例，到 2020 年 1 月中旬逐渐受到公众关注。按照党中央的决策部署，1 月下旬，一场抗击新冠肺炎疫情的人民战争全面打响。面对一个未知的病毒，各国科研人员都在各自岗位全力奋战，分离病毒株，研发疫苗，开发对症药物等，图情领域人员也在进行着相关研究。

4.1 情报学热点探析及可视化图谱分析

2020 年 1 月 14 日，PUBMED 更新了主题词词条，将新冠肺炎病毒冠名为“severe acute respiratory syndrome coronavirus 2[14]”,下属款目词有 9 个。经过世界卫生组织专家讨论审核，2 月 13 日，世界卫生组织将该病毒正式更名为“COVID-19[15]”，下属款目词有 10 个，分别是 2019 novel coronavirus disease、COVID19、COVID-19 pandemic、COVID-19 virus infection、coronavirus disease-19、2019 novel coronavirus infection、2019-nCoV infection、coronavirus disease 2019、2019-nCoV disease、COVID-19 virus disease，这些变更在 Mesh 词库中都已记录在册。随后在 3 月 19 日和 3 月 26 日又增加了“spike glycoprotein, COVID-19 virus[16]”和“COVID-19 diagnostic testing[17]”“COVID-19 vaccine[18]”主题词词条。

在疫情发生后，上海海事大学李杰[19]对冠状病毒运用各种可视化软件进行了一系列科学相关研究，例如运用 CiteSpace 对冠状病毒的文献分析、冠状病毒科学知识图谱分析、Scopus 中冠状

病毒研究的概况等。笔者于 2020 年 3 月 8 日在 pubmed 数据库中以上述关键词进行检索获取相关文献 966 篇，选取排序前 54 的关键词进行聚类分析获得情报学分析信息数据。聚类分为 8 个类别，分别是鼠线粒体蛋白、突变、公共卫生冠状病毒感染爆发、风险评估、疾病预防控制中心筛查、各类人群肺炎回顾性研究、SARS 相关病毒、儿童孕妇易感。图 4 展示了聚类山峰效果图，接下来将根据不同文章内容继续分类别进行后续详细分析解读。

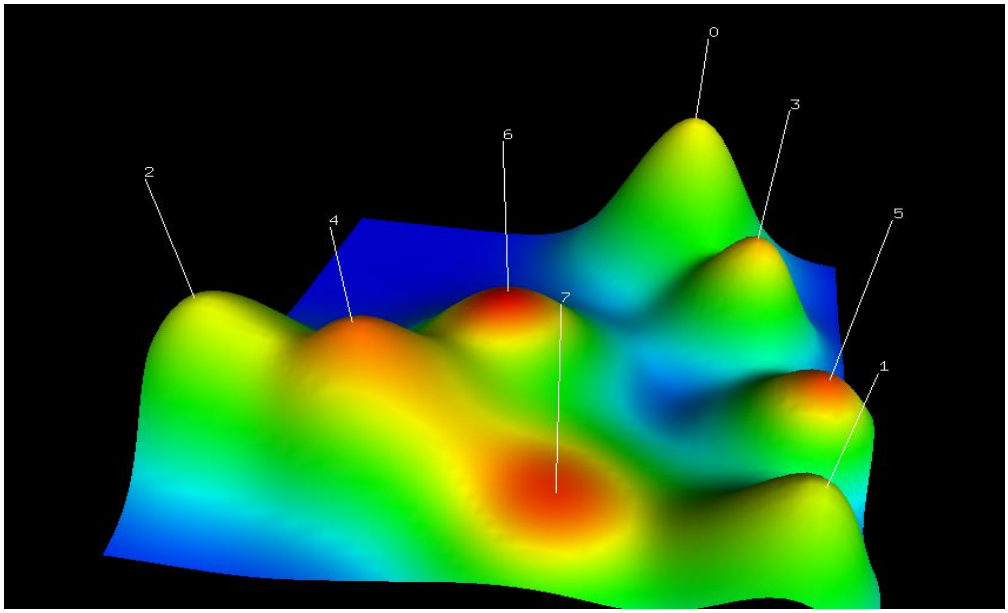


图 4 COVID-19 关键词聚类山峰图

4.2 突发监测方法预测辅助开发药物研究

突发监测是根据词频变化统计出低频却具有情报意义的突发词，探测新兴研究热点的方法，适用于某研究领域前沿趋势的探测，根据词在时间顺序上的变化来反应研究热点[18]。我馆参考咨询人员利用突发词分析与高频词分析相结合，将 SemRep 结合突发监测算法对 SARS 药物治疗发展曲线进行研究探讨，通过对 SARS 治疗药物现状及发展趋势曲线的绘制分析，为对比研究开发当前形势下的新型冠状病毒肺炎治疗新药提供参考借鉴。突发权重指数越大所代表的药物是治疗领域研究热点的可能性越大；频次越高所代表的药物是治疗领域研究重点的可能性越大，以期为文本知识发现提供新尝试[20]。

4.3 其他情报研究分析

2019 年新型冠状病毒与 2003 年 SARS 病毒具有较高的同源性,SARS 研究的历史起源、演化过程及经典文献对医务工作者及科研人员防治新型冠状病毒具有重要的借鉴意义。我馆参考咨询人员采用新兴的引文出版年光谱(reference publication year spectroscopy,简称 RPYS)分析方法,梳理 SARS 研究的历史起源和发展演化过程,为 2019 年新型冠状病毒性肺炎疫情的防治和科研工作提供情报支持。

5 其他应急响应支撑

5.1 加强疫情期信息质量的控制、筛选、甄别

在疫情爆发的特殊时期，难免一些虚假信息、不实信息会混淆干扰大众视听，蒙蔽真相，甚至引起恐慌。让谣言止于智者，图书馆馆员要根据自己的信息素养能力，甄别真实有效信息提供给大众读者，便于其了解最新动态。这便要求对应急信息的质量进行评估，选取高质量的应急信息能为应急管理提供信息支撑[21]。

5.2 提升馆员面对重大公共卫生事件所具备的专业服务能力

图书馆员需要具备信息资源挖掘整合能力及了解和掌握各种线上咨询、服务的专业知识能力，来适应和应对新时期下突发重大公共卫生事件开展服务，馆员的专业技能对于服务效果产生至关重要的作用。可以通过线上培训课程的学习提高专业技术能力来应加从容应对突发事件对图书馆开展各项服务带来的挑战。

6 疫情常态化下如何开展图书馆各项服务开展

如今全面复工复产，对于疫情的防控成为常态化，仍然不能掉以轻心，探讨面对疫情常态化下的图书馆各项功能服务工作开展情况。

6.1 面对应届毕业生的还书工作

疫情没有完全结束，毕业年级学生部分根据自愿原则部分返校，其他年级学生亦没有返校。图书馆与各个院系负责辅导员联系，为应届毕业生还书提供帮助。毕业生可以将书籍快递寄回学校图书馆，由图书馆馆员帮助其归还书籍，并做好沟通交流。

6.2 疫情常态化下图书馆网上预约借阅、到馆

为避免人员聚集，图书馆开启预约服务制度，包括到馆借书，到馆自习，到馆办理各项参考咨询服务业务，学生等外来读者进馆、出馆均需扫描二维码登记，同时红外线监测体温。

6.3 制定应急启动预案

若是发现到馆人员有体温超过37.8° 者，立即启动应急预案，拨打120将发热人员送往指定医院进行核算检测，密切接触人员隔离，直到检测结果出来，防控安全隐患解除。

7 结语

疫情是命令，防控是责任，服务是天职。面对此次疫情，各大高校图书馆积极行动，共克时艰，全力做好文献资源建设、信息技术保障工作，进行了专业、细致、热情、周到的服务，成为广大师生教学、科研、学习的坚实后盾。在疫情防控和读者服务工作中坚持“线下闭馆，线上服务”，结合实际，认真履职，精准施策，甘于奉献，不断创新服务，充分展现出高校图书馆人的风采，体现了图书馆人同心同德、众志成城，以不离不弃的服务相伴始终，共渡难关[22]。目前，新冠肺炎疫情是否能够彻底消除或者不再第二波反弹仍然不够明朗，有专家学者提出在未来的一段时间疫情常态化将成为一种可能。不管怎样，无论是疫情再次爆发或者面对常态化的服务开展，图书馆人将吸收过往经验，再接再厉，始终发挥图书馆为学校教学、科研提供文献服务保障支撑服务的功能。

PU-002

健康医疗大数据多中心研究现状与发展趋势

刘迷迷、李丽婷、周毅
中山大学中山医学院

目的 随着大数据、云计算、物联网、移动互联网和人工智能等信息技术在健康医疗领域的不断应用，健康医疗大数据呈指数增长。如何利用当前的信息技术，对这些多源、异构且分散存储的大数据进行统一管理、高效共享和挖掘利用，是当前健康医疗领域亟待解决的问题。开展多中心健康医疗大数据研究是当前主要的解决方案之一。本文主要目的是通过调研，了解和分析多中心健康医疗大数据研究的现状与发展趋势。

方法 基于多中心健康医疗大数据研究领域方面的关键词，在网络搜索引擎、文献数据库，以及医疗健康相关政府部门、事业单位和医学信息技术企业等官方网站进行调研，检索相关标准、指南、行业报告和文献报道。对检索资料进行详细分析，总结多中心健康医疗大数据研究的现状与发展趋势。

结果 对相关标准、指南、行业报告和文献报道等检索资料进行详细分析，根据数据整合利用方式方式的不同，将当前多中心健康医疗大数据研究归为三大类：基于数据汇聚、数据中心和平台的，以及基于通用数据模型的健康医疗大数据多中心研究。第一类只利用大数据技术进行数据分析挖掘的健康医疗大数据多中心研究逐渐减少；第二类利用大数据技术建立大数据库或平台，进行数据的收集、存储、分析挖掘的健康医疗大数据多中心研究更为常见；第三类基于数据标准化的、保护患者数据的健康医疗大数据多中心研究逐渐发展起来。

结论 基于健康医疗大数据的多中心研究有众多优势，但实际操作过程仍存在挑战。大数据时代背景为多中心临床研究的开展提供了先进的技术条件，结合大数据、云计算、物联网、移动互联网、人工智能和区块链等新兴信息技术开展多中心研究也有许多需要解决的问题。研究者应具备数据挖掘的科研思维，充分利用和改进大数据等新兴信息技术，努力探索解决数据获取、患者隐私保护、结果分享与发布和多中心利益冲突等问题的新思路，更加全面地开展多中心研究，创新健康医疗模式，促进医学与生命科学、卫生健康等社会事业发展。

引言

在当前信息时代，信息系统、移动设备及物联网设备的普及使用，为医疗带来应用基础和条件，同时产生了海量的全基因组测序数据、医学影像数据、半结构化和非结构化的临床文档、结构化临床记录等生命组学、队列资源、临床诊疗数据，构成健康医疗大数据。与此同时，随着社会信息化程度不断加深，以云计算、大数据、物联网、移动互联网、人工智能和区块链等新兴信息技术为核心的新一轮科学技术革命也在兴起。

整合健康医疗领域存储分散、多维度、多模态、异构的大数据，分析和挖掘其中规律，预测

未来，是健康医疗大数据蕴藏的最大价值。为此需要充分利用当前的信息技术，开展基于健康医疗大数据的多中心研究，以促进健康医疗大数据的发展与应用，给健康医疗模式带来深刻变革和推动社会发展。

1 健康医疗大数据多中心研究概述

多中心研究是由多个独立的研究中心按照同一研究方案在不同地点和单位同步进行的医学研究，包括招募研究对象、实施研究和发布研究结果。由各中心共同指定一位主要研究者总体负责，以协调各中心之间的工作。所有参与的研究中心必须严格遵守研究方案的要求，以保证数据的一致性和研究结果的可比性。

多中心研究有许多优点，主要包括：一定时间内可招募到大量研究对象；研究对象范围广，更具代表性，避免非研究因素的干扰，进而提高最终研究结果的可信度；参与的研究者众多，相互合作，共同解决遇到的挑战，以提高和保证研究设计、实施和结果发布的质量。理想的多中心研究是跨国际的研究，由多国研究者在各国多个研究中心同时开展，不同国家的研究中心，研究对象来自不同人群，能更好地代表各种人群的数据。多中心研究已成为各国医疗机构开展健康医疗研究的重要方法。

虽然多中心研究的设计原则和方法与单中心研究大同小异，但多中心研究增加了研究设计的复杂性。因此，多中心研究也存在一定的挑战，主要包括：多中心研究可能包括多种类型的观察性研究；各研究中心可能存在特殊性，因此设计研究方案时必须结合具体情况由所有研究中心共同讨论确定；多中心性难以严格地标准化干预措施和指标测量方法；为保证研究的一致性，各研究中心需同步实施研究；合并和分析各中心的数据，以及发布研究结果时，需要遵循相似性原则，例如研究对象所在地区的人口学特征等，不能简单地合并和分析数据，避免研究结果有误。

然而，大数据时代的到来给传统的多中心研究注入了新的活力。可以利用大数据技术整合、存储、分析和处理海量、多维、多模态的数据，以促进多中心研究进一步扩大样本量，以及纳入组学数据、临床诊疗数据和队列资源等范围更广泛、类型更多样的数据源，使研究结果更加可信[1]。此外，大数据技术的使用可以减少随机对照试验存在的偏倚，例如选择偏倚、数据收集偏倚、指标测量偏倚和混淆偏倚等[2]。因此，大数据技术提高了传统多中心研究的质量，基于健康医疗大数据的多中心研究越来越受关注，开展的研究项目数量逐年增加，涉及的研究中心也越来越多[3]。

2 国内外健康医疗大数据多中心研究现状

根据数据整合利用方式的不同，目前国内外开展健康医疗大数据多中心研究主要分为三类：

2.1 基于数据汇聚的健康医疗大数据多中心研究

第一类，基于数据整合的健康医疗大数据多中心研究，以疾病为中心，利用大数据技术汇聚多中心研究数据，并对其进行分析挖掘，获得可指导科研、教学和临床医疗实践的知识。Benoit Lattuca 等开展了一项多中心随机试验研究，利用大数据方法技术分析了视频对在冠状动脉造影前接受知情同意的择期住院患者的理解情况和满意度的影响[4]。Almonacid Sánchez 等提出并实施了应用大数据方法估计西班牙严重哮喘患病率的研究方案，并对其可靠性和效率进行了评估[5]。许亚

莉等开展了多中心联合筛查红细胞血型不规则抗体的研究，应用大数据分析技术，发现在患者输血前做抗筛的必要性，尤其是有输血史和（或）妊娠史的患者，为临床安全、有效、合理用血提供了科学依据[6]。贺忆培基于多中心综合医院的真实世界的大数据研究，探讨了慢性肾脏病中医证型与疾病的关联性研究[7]。

2.2 基于数据中心和平台的健康医疗大数据多中心研究

第二类，基于数据中心和平台的健康医疗大数据多中心研究，以疾病为基础建立统一的数据中心和平台，接入多家研究中心，可面向外部医疗机构开放，分享和发布研究结果。A Pession 等利用 Advanced multienter Research (AMR) AIEOP 网络，建立一个具有不同用户配置的综合数据管理系统，包括机构、数据管理器、协调员和实验室，并设置一定权限向外部用户开放，以获取和分析更多的数据和信息，最终提高医生诊断、治疗和用药的效率和准确性[8]。Samir Das 等利用蒙特利尔神经学研究所(MNI)开发的数据库和处理生态系统，来支持国内和国际的多中心数据采集，创建数据库存储库，并利用现有的软件工具包进行了大规模数据处理[9]。Suzuki Takahiro 等综合多中心的出院摘要，建立文本数据库，提高了文本挖掘的准确性，并对数据库进行分析，实现了临床应用[10]。

此外，Jihwan Park 等开展了一项多中心、大数据研究，基于收集的临床、影像学和病理学等综合数据，开发了相关的智能软件，以帮助患者和临床医生选择最佳的治疗方案[11]。Satya S Sahoo 等提出一个 Cloud wave 平台，可以满足电生理大数据的三个主要需求（可靠地存储大量的信号数据；有效地执行复杂的心电信号处理计算，以实现实时用户交互；支持基于本体驱动的 web 的可视化和查询访问协同研究），并表明云平台是利用大规模电生理数据推进多中心临床研究的新途径[12]。Grégory Operto 介绍了一个大型的分布式神经成像系统（CATI），该系统支持大规模人群的神经成像，可托管数据，有严格的数据访问权限机制，促进尽可能多的共享和开放数据，并通过基础设施服务、专家学术团队的专业知识和大规模、协调的成像设备网络的整合分析，为日益复杂的问题提供具体解决方案[13]。

近年来，随着信息技术在我国健康医疗领域的深入应用，国内学者们也开始开展第二类健康医疗大数据多中心研究。朱立峰等针对传统多中心临床研究存在的问题，推进多中心临床大数据平台建设和深度应用工作，建设了多中心临床大数据平台，并对多中心临床大数据平台中临床随访系统、核心系统和疾病分析应用系统等进行技术验证，建立了可行的技术框架[14]。李国林等构建了多中心的老年神经退行性疾病临床大数据研究平台，对该疾病的生理资料、核磁影像和神经心理学量表等临床数据进行一体化管理、数据共享和分析挖掘[15]。此外，中南大学爱尔眼科学院主导的近视眼大数据多中心研究项目和甲状腺结节精准检查及甲状腺癌大数据多中心研究项目等项目相继启动，大数据库的建立为我国各种疾病防治提供有力的数据基础，同时积极探索适合我国的、创新的疾病管理与防治模式，推动我国整个医院健康管理业务的新发展[16,17]。

2.3 基于通用数据模型的健康医疗大数据多中心研究

第一类和第二类健康医疗大数据多中心研究，虽然将可以快速积累各研究中心的数据，但并

没有对这些数据集的词汇和结构进行标准化，这使得为临床决策提供支持而真正实施多中心研究或分析比较时非常困难。而第三类，基于通用数据模型（Common Data Model, CDM）的健康医疗大数据多中心研究，健康医疗大数据多中心研究使用通用数据模型来对各研究中心的数据进行标准化，应用统一的数据结构来保障数据拥有者运行相同的分析代码。换言之，第三类健康医疗大数据多中心研究以通用数据模型为基础，各研究中心根据通用数据模型建立各自的数据库，利用分布式网络运行相同的代码或工具对数据进行分析和处理，各研究中心不需要输出可能包含研究对象隐私的数据，只需要将网络上运行的数据分析的结果整合起来，仅对外分享和发布这个整合的分析结果。

目前，国际上已经有研究协作组采用分布式研究网络开展第三类健康医疗大数据多中心研究。观测健康数据科学和信息学联盟（OHDSI）是其中之一。OHDSI 联盟开发了 Observational Medical Outcomes Partnership (OMOP) CDM，目前已出版 6.0 版，包括来自 20 多个国家的临床数据，已实现 15 亿患者的信息转化。OMOP CDM 是一套统一的数据标准，规范了观察性数据的格式和内容，支持不同来源的观察性数据通过数据的抽取、转换和加载（Extraction Transformation Loading，ETL）过程形成标准化的数据结构，在此基础上进行数据的查询和分析等应用[18]。利用这种促进多中心协作研究的通用数据模型（OMOP CDM），实现方法研究、开源的模型开发、数据分析和临床应用，以促进从临床电子病历等观察数据中挖掘可靠的医学证据和传播医学知识[19]。此外，为实现第二代测序基因数据的多中心大数据研究，Seo Jeong Shin 等对 OHDSI 联盟开发的 OMOP CDM 进行了扩展，创建了一个用于分布式研究网络的基因组数据 CDM (G-CDM)，可以解决患者隐私问题和支持基因组数据的种族多样性等，并通过比较公共数据库和临床实践产生的基因组数据的差异，验证了模型的可行性[20]。

2.4 健康医疗大数据多中心研究现状与发展趋势

总的来看，随着大数据技术在健康医疗领域的不断应用与发展，第一类只利用大数据技术进行数据分析挖掘的健康医疗大数据多中心研究逐渐减少；第二类利用大数据技术建立大数据库或平台，进行数据的收集、存储、分析挖掘的健康医疗大数据多中心研究更为常见；第三类基于数据标准化的、保护患者数据的健康医疗大数据多中心研究逐渐发展起来，可能是未来一定时期内主要的多中心研究。

此外，区块链技术的越来越受关注，从数据库建立和管理的角度来看，区块链是一种公开透明的、去中心化的、不可更改的分布式数据库，具有基于时间戳的链式区块结构、分布式节点的共识机制、基于共识算力的经济激励和灵活可编程的智能合约等代表性创新点[21]，这与目前第三类健康医疗大数据多中心研究方法技术存在相似之处，可能为多中心研究的数据库建设提供新的技术和研究思路。

3 总结与展望

虽然基于健康医疗大数据的多中心研究有众多优势，但实际操作过程仍存在挑战。各研究中心的数据来源和数据格式可能存在较大差异，使研究的通用性受到限制。目前国内外专家、学者

正努力研究使用标准化数据模型来收集和协调来源不同的健康医疗数据，但数据获取、患者隐私保护以及研究中心间的利益冲突等问题，给数据的存储、分析挖掘和结果分享带来了研究技术、研究设计逻辑、隐私伦理和法律法规等方面的挑战[22, 23]。此外，大数据技术在健康医疗领域的应用与发展需要研究者具备学科交叉的思维模式，利用学科交叉思维指导多中心研究，其应用意义不仅在于提高临床科研的质量和成效，更在于调动不同学科领域的专家根据所在学科领域的特点对数据进行多个层次的分析挖掘，开拓研究思路[24]。

综上所述，大数据时代背景为多中心临床研究的开展提供了先进的技术条件，但结合大数据、云计算、物联网、移动互联网、人工智能和区块链等新兴信息技术技术开展多中心研究仍有许多需要解决的问题。研究者应具备数据挖掘的科研思维，充分利用和改进大数据等新兴信息技术，努力探索解决数据获取、患者隐私保护、结果分享与发布和多中心利益冲突等问题的新思路，更加全面地开展多中心研究，创新健康医疗模式，促进医学与生命科学、卫生健康等社会事业发展。

PU-003

基于妇幼数据中心构建信息安全保障体系的设计研究

王辉
郑州大学第三附属医院（河南省妇幼保健院）

目的 为了保证信息系统的安全运行，必须遵循国际、国内标准和规范，通过系统的技术防护措施和非技术防护措施建立安全保障体系，为妇幼平台提供一个安全环境。

方法 安全保障体系包括基础设施安全体系、应用安全体系和安全运维体系三大部分

结果 安全保障体系包括基础设施安全体系、应用安全体系和安全运维体系三大部分。网络边界防护、系统主机防护、入侵检测与审计和完整的防病毒体系，辅之以远程访问接入及终端准入控制，构造一个切合实际的、行之有效的、相对先进的、稳定可靠的网络安全系统平台，应对日新月异的网络安全威胁。

结论 信息技术的发展极大地推动了医院信息化建设的进程，提升了医院为社会提供医疗服务的效率。但随之而来的外部网络攻击、信息数据泄露等安全问题也成为卫生行业需要高度重视并解决的问题。本文通过对妇幼数据中心构建基础安全保障体系的设计研究，提供了安全策略、安全管理、安全技术和安全运维体系设计的思路，为医院信息安全体系建设提供了一些参考。

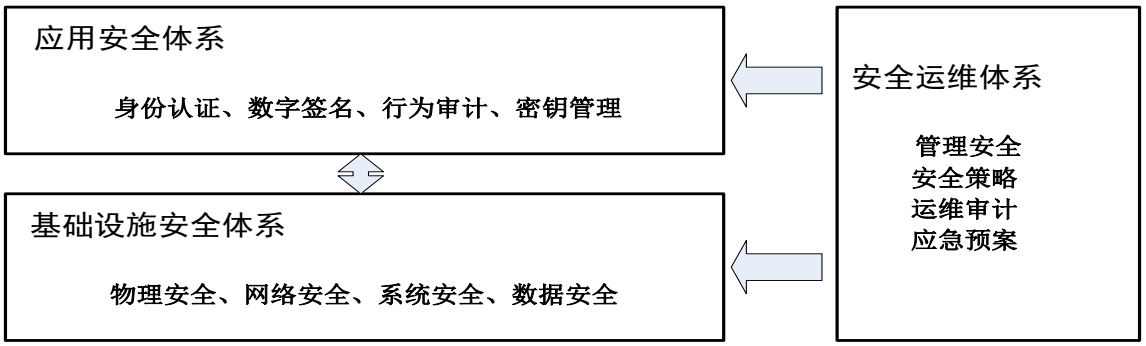
1、安全环境概述

为了保证信息系统的安全运行，必须遵循国际、国内标准和规范，通过系统的技术防护措施和非技术防护措施建立安全保障体系，为妇幼平台提供一个安全环境。

信息系统可靠安全的运行不仅关系到数据中心本身的运行，最重要的是关系所有与妇幼健康相关的保健机构的正常业务开展，因此它的网络、主机、存储备份设备、系统软件、应用软件等

部分应该具有极高的可靠性；同时为维护服务机构和服务对象的合法权益，数据中心应具备良好的安全策略、安全手段、安全环境及安全管理措施。

安全保障体系包括基础设施安全体系、应用安全体系和安全运维体系三大部分，如图所示：



2、

安全体系架构设计

三个组成部分之间是密切相关的，在实际应用过程中互相交错和引用，不一定能够严格区分，特别是安全设备产品发展很快，一件设备可能聚集着多种安全技术和应用，这使我们在技术和设备的选择上要注意选择合适的组合。

2.1 基础设施安全体系

基础设施安全构建在网络基础设施上，提供网络边界防护、系统主机防护、入侵检测与审计和完整的防病毒体系，辅之以远程访问接入及终端准入控制，构造一个切合实际的、行之有效的、相对先进的、稳定可靠的网络安全系统平台。

基础设施安全构建提供整体防护，针对关键区域，采用负载均衡系统，提供更高效率的服务，满足业务持续性的要求。避免因某个环节的不安全导致整个体系安全性能的下降。

2.2 应用安全体系

应用安全支撑平台以密码服务为核心，以身份认证为基础，提供人员身份认证、数字签名、数据加密等功能，在远程访问、网络接入，及应用服务等，强认证措施可以弥补访问控制方面的很多缺陷，数据加密可以为系统提供纵深防御。

应用安全支撑平台通过密钥管理中心、身份认证系统提供针对应用系统的安全服务，整合应用系统的认证、授权、加密操作，建立密钥与应用相隔离机制，既保证密钥使用、管理对应用系统透明，减轻应用系统开发、部署面的压力，又能进一步提供管理的安全。

2.3 安全运维体系

安全运维体系提供对整个系统安全运维支持，对安全设备进行集中的事件管理，对人员进行集中的身份管理，对整个系统提供全方面的威胁管理。对于关键设备的管理行为进行有效性认证，同时保障管理数据传输的安全。对运维操作内容进行审计。完善统一安全监控、事件管理，建立运维安全体系,保障系统的安全、可靠的运行，提供有效的安全管理。

3、基础设施安全体系建设

基础安全设施主要由物理安全、网络安全、系统安全等组成。

3.1 物理安全

系统的安全体系，首先需要考虑到物理安全。根据机房建设标准有效保障物理安全。物理安全是保护计算机网络设备、设施以及其它媒体免遭地震、水灾、火灾等环境事故以及人为操作失误或错误及各种计算机犯罪行为导致的破坏过程。保证计算机信息系统各种设备的物理安全是保障整个网络系统安全的前提。显然地震、水灾、操作失误等等都受不可预测因素的影响，所以严格意义上没有绝对的物理安全。

3.2 网络安全

信息系统的可靠、安全运行是区域内保健机构开展正常的业务活动基本保证，也是对系统正常运行的重要支撑。因此，它的网络、主机、存储备份设备、系统软件、应用软件等部分应该具有极高的可靠性；同时为保守用户秘密，维护用户的合法权益，数据中心应具备良好的安全策略、安全手段、安全环境及安全管理措施。

结合网络的特点分析，信息系统在网络层面临的主要具体威胁如下（N 代表威胁发生在网络层）：

网络层面威胁因素一览表

标号	威胁描述
N1	黑客通过 Internet 连接对妇幼保健等信息进行破坏和非授权访问
N2	黑客或内部人员从 POS 点通过网络连接对信息系统平台进行攻击或非授权访问
N3	黑客或内部人员从和妇幼平台连接的第三方网络通过网络连接对信息系统平台进行攻击或非授权访问
N4	妇幼保健数据中心中的服务器感染蠕虫或者被种植木马、后门程序而导致向外发起的非法网络连接
N5	攻击者利用分布式拒绝服务攻击等拒绝服务攻击工具，恶意地消耗网络、操作系统和应用系统资源，导致拒绝服务
N6	攻击者利用网络协议、操作系统、应用系统漏洞，越权访问文件、数据或其他资源
N7	攻击者利用网络结构设计缺陷，访问未授权访问网络
N8	攻击者和内部人员利用网络扩散病毒
N9	攻击者截获、读取、破解通信线路中的重要信息
N10	蠕虫通过 POS 连接或第三方外部网络连接扩散到信息平台
N11	蠕虫通过内部网络连接扩散到信息平台
N12	利用网络设备、防火墙的漏洞和入侵攻击导致网络基础设施瘫痪

N13	攻击者利用 DDOS 攻击使链路出现瘫痪，导致信息系统瘫痪

信息系统所涉及的信息包括：服务对象的基本健康信息、诊疗数据，以及区域内妇幼卫生资源数据等等。这些业务信息一旦遭到非法入侵、修改、增加、删除等不明侵害（形式可以包括丢失、破坏、损坏等），会对服务对象的合法权益和服务机构以及相关部门的正常业务活动造成影响和损害，可以表现为：影响正常工作的开展，导致业务能力下降，造成不良影响，引起法律纠纷等。程度表现为严重损害，即工作职能受到严重影响，业务能力显著下降，出现较严重的法律问题，较大范围的不良影响等。

网络安全是信息系统安全体系的重中之重，关系到整个信息平台的安全，一旦网络被攻破，那么大量的健康档案的相关信息将直接受到威胁。

4、妇幼数据中心基础安全保障设计

4.1 防火墙访问控制措施

防火墙在边界上针对信息系统的网络数据流入/流出提供过滤和保护，目的是阻止安全域外部连接的非授权进入内部，以及通过网络手段阻断特定的内外连接。在信息系统和外部网络的边界处应部署防火墙设备，根据会话状态信息、应用层协议、访问发起用户、被访问资源等进行访问控制规则设置，保证通过网络边界的会话得到一定控制。

4.2 网络入侵检测防御

网络入侵防御设备（IPS）在网络入侵行为进入被保护网络之前通过报警、阻断等措施为信息系统提供防护。同时，针对网络入侵进行监测，自动识别各种入侵模式，在对网络数据进行分析时与这些模式进行匹配，一旦发现某些入侵的企图，进行报警，通过对网络入侵的检测，弥补防火墙等边界安全产品的不足，及时发现网络中违反安全策略的行为和被攻击的迹象。在妇幼保健综合业务信息系统和外部网络、POS 的边界部署网络入侵防御设备对进入信息系统的网络请求进行预先阻断，避免进入内部网络。

为了简化部署，在信息系统采用高端口密度的入侵防御设备同时实现入侵防御和入侵检测的功能，即对于关键链路实现串行接入，对入侵实现实时阻断，同时采用同一台设备的其它端口对于其它网络采用端口镜像的方式接入进行实时入侵检测。

4.3 网关病毒防范

将病毒防御体系前移，建立网关病毒防御体系。对进出网络数据进行查杀，阻断来自 WEB 的文件病毒进入网络内部；阻断来自互联网的病毒或黑客对应用服务器群的病毒入侵或利用漏洞入侵；阻断来通过互联网对终端操作系统的病毒入侵或利用漏洞入侵；阻止病毒的相互传播、阻断非法/恶意 URL，彻底将病毒杜绝与网络之外。

4.4 远程访问控制

对于信息系统的信息交换，以及各保健机构远程接入的客户端或应用系统对信息平台进行远程访问，采用安全认证为信息系统提供安全服务，拒绝任何未授权用户对应用系统的访问尝试并为应用系统提供用户在该应用系统中的身份信息。

4.5 内外网安全

在妇幼保健信息系统网络中，用户接入是一个关键环节。妇幼数据中心共有 3 台服务器，其中一台为 WEB 服务器，进行对外部互联网系统进行数据交换。进行安全保障将采用基于内外防火墙的安全接入设置。采用内部防火墙及安全路由进行部署。内外网之间通过内网防火墙及安全路由器的严格设置，控制内往外连接的端口，及内网的安全设置。

内网部署内网防火墙加强对内网安全的限制，并通过安全路由强制对上网设备的安全设置检查，屏蔽对非安全设备的网络连接，使必要的内网接入公网安全性大为增强。

4.5 漏洞管理系统

根据 Gartner 调查报告，安全漏洞一般可以分为两类：配置漏洞和系统漏洞。配置漏洞是指由于默认配置、误配置等导致的安全隐患；系统漏洞是指由于系统及其依赖的子系统在产品开发中的代码问题而导致的安全隐患。这两类漏洞都需要进行及时发现和完善的管理，避免系统成为黑客、蠕虫等攻击的目标。

漏洞管理系统可以进行资产分组、分类和优先级划分，模仿黑客的行为模式，协助系统找出网络中每一台主机、网络服务的相关信息与弱点，以了解整个网络架构的变动状况，是否影响或威胁系统的安全。并且它能通过定义工作流模版规则，自动从发现的弱点产生修补问题票单，通过邮件自动发送给相关的安全管理员，并且跟踪管理员的修补过程。若管理员及时修复，它能够进行确认并自动关闭该问题票单；若该弱点没有在规定时间内修补，则会发出报警，提醒管理员需要及时修补该弱点。

4.6 主机安全软件

主机安全软件能够对包括客户端计算机和服务器在内的系统进行安全加固和威胁防范，包括防恶意代码，防入侵等等。防病毒类产品提供对计算机病毒的防治、防护侧重于防护本地计算机资源。计算机病毒防治产品是通过对内容或行为的判断建立系统保护机制，目的是预防、检测和消除计算机病毒。由于新的病毒层出不穷，尤其是在 Windows 平台上的恶意代码更是泛滥肆虐，所以，主机安全软件应当对于新病毒具有一定的主动防御机制，比如对系统目录的保护、防止蠕虫经常利用的溢出类型的攻击等等，即具有主机防入侵功能。

主机防入侵类产品对已经抵达主机的数据进行监测，从主机或服务器上采集包括操作系统日志、系统进程、文件访问和注册表访问等信息数据，并根据事先设定的策略判断数据是否异常，从而决定采取报警、控制等措施，目的是对入侵主机行为进行发现和阻止。此外，主机安全防护应当具有自动更新机制，以便于应对随时出现的新的威胁。

主机安全产品应选择网络版软件，因为网络版软件具备整体的管控能力，统一配置防毒策略、统一监控终端安全状态、统一更新病毒库、集中分析整体病毒日志、强制全网扫描病毒。终端安装的杀毒软件状况及病毒处理情况尽收眼底，实时掌握病毒情况。

4.7 系统安全加固

网络中病毒和木马的数量以几何级数增长，传统杀毒软件“特征库”的查杀方式，已很难跟上病毒增长的速度，而且随着病毒技术的发展，很多病毒利用现在操作系统底层安全性不足的缺陷，已经深入到系统内核层，这不仅增加了查杀的难度，甚至出现杀毒软件自身被病毒从底层破坏，数据资产被病毒木马任意支配的情况。

操作系统安全加固产品要求支持现在所有主流操作系统，从系统内核层实现对服务器中数据资源的强制访问控制，对操作系统的文件、注册表、服务、进程等资源实现强制访问控制，消除病毒等恶意程序的生存环境，使服务器能够免疫针对服务器操作系统的攻击,实现对已知或未知病毒程序、ROOTKIT 级后门威胁的主动防御。避免出现业务系统因新的蠕虫等感染型病毒的出现，而导致的业务网络瘫痪、业务系统服务中断等安全事故。

操作系统安全加固系统应实现服务器中安全管理员、安全审计员、系统操作员的“三权分立”机制，安全管理员通过访问控制策略，实现对信息中心等系统运维人员的“最小授权”，在策略中开放系统升级、应用程序维护等必要权限，对核心监管数据库等敏感文件的访问进行严格的授权控制，最大程度上避免因为内部人员的误操作或盗用管理员权限的恶意行为导致的安全事件。

另外操作系统安全加固系统可以建立完善安全审计制度，监控服务器中所有的操作行为，通过审计日志的查询统计功能，能够对服务器受到的攻击种类、次数做出宏观的认识，并及时发现运维人员对数据库、业务系统相关资源的非授权行为。

4.8 系统信息审计软件

安全审计类产品针对信息系统的活动信息进行审计记录及分析，目的是通过安全审计挖掘安全事件，并加以分析，得到相关信息。

在整个应用的安全体系中，安全信息来自各种系统，对安全信息的收集在整个安全管理的体系中占有重要的作用。各类系统的运行日志和管理日志是安全信息的重要来源。日志不仅可以帮助管理人员对各种安全事件和安全操作进行审计，分析系统存在的薄弱环节，发现潜在的危险，而且大量的入侵行为直接与系统日志相关，如何快速准确的收集日志信息并将这些信息加以分析保存是安全管理工作中重要的一环。

安全审计软件应考虑对系统安全策略配置状况进行审计。系统配置问题是安全问题的重要源头，比如不安全的共享、弱口令等等往往会导致严重的蠕虫病毒传播、信息泄密或入侵行为。通过采用系统安全策略审计软件对系统范围内的主机进行安全策略审计和集中汇总，使管理员能够对安全问题及时发现，并得以及时解决。

5、结 论

信息技术的发展极大地推动了医院信息化建设的进程，提升了医院为社会提供医疗服务的效率。但随之而来的外部网络攻击、信息数据泄露等安全问题也成为卫生行业需要高度重视并解决的问题。本文通过对妇幼数据中心构建基础安全保障体系的设计研究，提供了安全策略、安全管理、安全技术和安全运维体系设计的思路，为医院信息安全体系建设提供了一些参考。

PU-004

探索医院建设“互联网+护理服务”可行性分析

牛亚军
郑州大学第三附属医院（河南省妇幼保健院）

目的 本文主要介绍“互联网+护理服务”的建设方案，该护理服务模式需求强烈，国家扶持力度大，随着近年来互联网和大数据技术的发展，为该方案提供了强有力的技术支撑；同时该方案顺应了时代潮流，具有很强的创造性，利用互联网技术将医院和患者紧密的联系起来，医院的护理资源得到了最大化的利用的同时，也产生了巨大经济效益；另外，该方案的实施也将极大地提升医院的服务形象和患者的满意度；最重要的一点是，该方案的实施方便快捷，在保障安全和质量的前提下，全套服务保障项目可以快速落地和可持续运转。

方法 便捷可靠。本方案针对患者/居民、医护人员、医院管理的不同需求和特点，提供基于微信生态和 APP 的患者端应用，快速上手使用；针对护士提供专门的移动智能硬件和 APP 应用，保障护士操作需求和安全需求；针对医院运营管理，提供 PC 端信息平台和移动审核工具，满足管理的深度应用需求。

结果 自 2019 年 2 月开始试点以来，北京、天津、上海、江苏、浙江、广东等 6 地试点已经纷纷制定了各地的实施细则、服务清单、价格和收费、监管和保障措施，组织各级医疗机构开始开展上门服务。其中，广东省参与服务的医院最多，已经有约 140 家医院参与服务，各试点地区在服务开展 1 年以来提供了上万单的服务，从媒体消息来看，服务质量和人身安全有保障，表现出了良好的社会效应，起到了很好的便民惠民效果，获得了群众良好的认同和口碑。全国其他开展服务的地区：除 6 个试点地区之外，河南（济源、荥阳、洛阳、商丘等多地）、湖北武汉、湖南长沙、新疆、江西南昌等多地自主开展了互联网+护理服务尝试。和试点地区的表现一致，也取得良好的认同和社会效应。

结论 “互联网+护理服务”是时代的产物，将医疗和互联网结合发展成为新的护理模式，打破传统的护理服务模式，与时俱进，顺应了时代发展的潮流；同时医院、患者互惠互利，极大地提高了服务效率和患者的满意度；另外，对提升医院和护理人员形象产生了积极的作用，充分发挥了医院的护理能效。

1 “互联网+护理服务”背景

“互联网+护理服务”主要是指医疗机构利用在本机构注册的护士，依托互联网等信息技术，以“线上申请、线下服务”的模式为主，为出院患者或罹患疾病且行动不便的特殊人群提供的护理服务。是当前阶段国家政策鼓励、百姓需求量大的互联网+医疗健康服务之一。

在健康中国战略和分级诊疗医改策略导向下，2018 年以来，国家先后围绕“互联网+医疗健康”政策，出具了“互联网+医疗健康惠民便民服务”、“互联网医院”、“互联网+护理服务”等具体实施执行办法和通知，为医疗机构服务患者和百姓的初心，指明了发力点和方向。

2 互联网+护理服务可建设方案

2.1 服务模式分析

按照健康中国、分级诊疗等医改大政方针，遵照护理服务业改革与发展的指导意见和互联网+护理服务的指导意见，互联网+护理服务可以按照二种模式并行实施服务：

(1) 二级以上的医疗机构直接为院后患者、行动不便的群体等服务对象提供上门服务；

(2) 二级以上的医疗机构联合基层医疗机构，指导、委派基层医疗机构护士为院后患者、行动不便的群体等服务对象提供上门服务。

具体服务模式如下图所示：

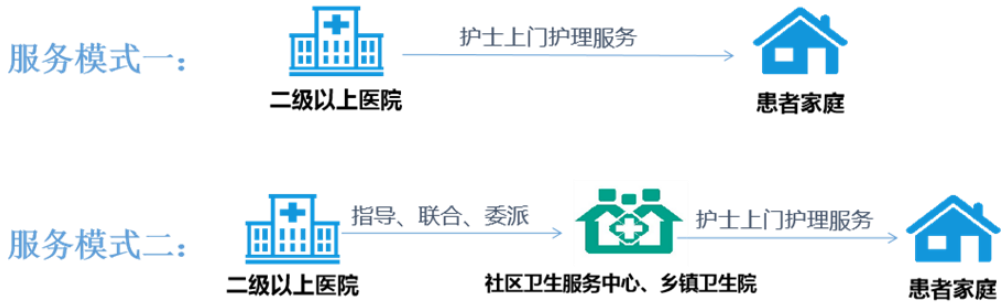


图 1 服务模式图

2.2 服务流程

互联网+护理服务采用“线上申请、线下服务”的服务模式，即在“互联网+护理服务平台系统”搭建完成后：

(1) 被服务对象或者其委托人（家属等）即可通过医院提供的患者端（一般是 APP、微信公众号等）进行下单；

(2) 医院调度人员通过运营管理平台进行服务评估，主要对下单人信息、患者信息、疾病诊断信息、疾病史进行审核和服务评估，审核通过的订单进行分派，指派本院护士、下级医联体机构（社区卫生服务中心）护士进行上门服务；

(3) 接到订单的护士了解被服务对象基本情况、进行服务沟通、领取准备耗材，进行上门服务，将医疗废弃物带回医院处置。

(4) 被服务对象或者其委托人在患者端对服务进行评价。

具体服务流程如下图所示：

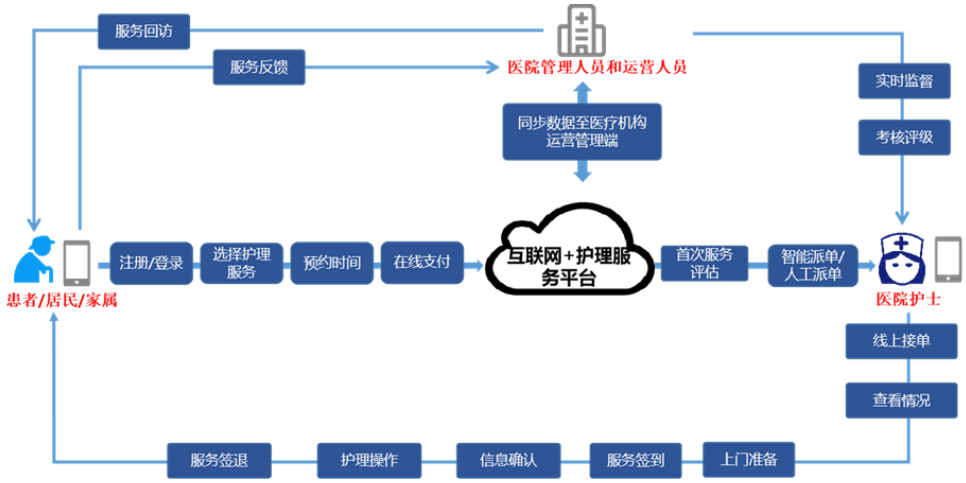


图 2 服务流程图

2.3 组织工作结构

互联网+护理服务作为一种创新型的服务模式，医院在开展服务之前需要充分的准备工作，来保障质量安全，精准对接人民群众多样化、多层次的健康需求。作为一项严肃的医疗服务，在工作开展之前，需要进行全面的人、财、物等相关准备工作。核心准备工作包括：

(1) 进行服务开展的人员准备、工具准备、保险准备等；

(2) 到物价部门进行物价审批，审批通过的价格进行公示；

(3) 参照国家政策法规建设并逐步完善院内服务制度和规范建设；

(4) 平台运营商提供软件系统，进行良好的信息化支撑；

具体组织工作结构图如下：

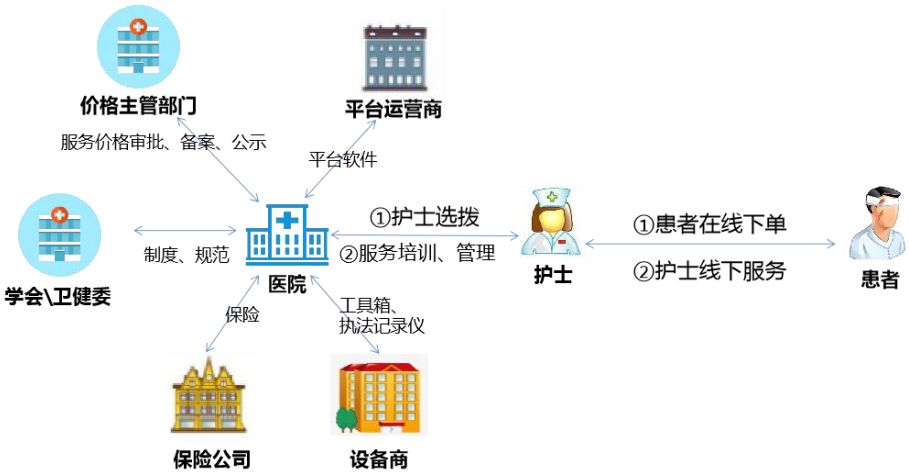


图 3 组织工作结构图

3 探索“互联网+护理服务”新模式

1. 便捷可靠。本方案针对患者/居民、医护人员、医院管理的不同需求和特点，提供基于微信生态和 APP 的患者端应用，快速上手使用；针对护士提供专门的移动智能硬件和 APP 应用，保障护士操作需求和安全需求；针对医院运营管理，提供 PC 端信息平台和移动审核工具，满足管理的深度应用需求。

2．多项措施保障服务质量与安全。通过 GPS 定位、患者审核、人脸识别技术、一键报警等方法保障护士人身安全；通过自动文书签署、护理工作记录仪等降低医疗纠纷风险，保障护士执业安全；通过上门前评估、护理前评估、护理后评估和记录，保障医疗服务质量和安全；通过硬件和网络技术保障，确保患者医疗和健康数据传输安全。

3．支持多种上门服务模式，充分满足医院当前需求和未来医共体模式下的分级协同服务。“护加家”移动信息化系统的订单和分配机制，支持（1）二级以上医疗机构直属护士团队直接上门服务模式；（2）二级以上医疗机构对医共体单位或者基层医疗机构进行指导和带教，促进基层医疗机构护理服务能力提升，由基层机构护士上门服务；（3）二级以上医疗机构对医共体单位或者基层医疗机构进行远程支持，以双方会诊联合服务的模式，患者到基层医疗机构或者基层护士上门为患者服务等多种服务模式。对多种模式的支持，能够为院内服务内容的扩展提供良好的基础，由护理服务向健康管理和指导扩展，同时，协同服务模式下，可以支持未来医共体下分级协同的业务和服务需求。

4．全套服务保障项目快速落地和可持续运转。本方案不仅提供服务开展所必须的软硬件设施，还提供配套的运营服务，以服务促项目建设效果，以项目运转反向促进软硬件产品提升，保障项目快速落地、可持续运转、高效优化。

4 基于河南省可行性分析

4.1 政策可行

2018 年 6 月，国家卫生健康委员会、国家发展和改革委员会、教育部等 11 个部门联合印发《关于促进护理服务业改革与发展的指导意见》，鼓励有条件的医疗机构对具有较高再入院率或医疗护理有较高需求的出院患者提供延续性护理服务，将护理服务延伸至社区、家庭。大力发展社区和居家护理服务。鼓励医联体内二级以上医院通过建立护理联合团队、一对一传帮带、开展社区护士培训等形式，帮扶带动基层医疗机构提高护理服务能力。鼓励二级以上医院优质护理资源加入家庭医生签约团队，为社区群众提供专业化护理服务。

2019 年 1 月，国家卫生健康委员会正式发布《关于开展“互联网+护理服务”试点工作的通知》及试点方案，确定 2 月至 12 月在北京、天津、上海、江苏、浙江、广东等 6 地试点“互联网+护理服务”。其他省份结合本地区实际情况选取试点城市或地区开展试点工作。

4.2 社会可行

社会需求层面：我国每年住院治疗患者超过 2 亿人，河南省约 2 千万人，院后患者群体庞大，患者有强烈的院后延续服务需求。

老年人、孕产妇等人群对上门护理服务需求方面。中国老年型年龄结构初步形成，并且具有总量大、增速快的特征，行动不便老龄群体有巨大的服务需求。据国家统计局统计，截至 2018 年底，我国 60 岁及以上的老年人 2.49 亿，占比 17.9%。65 岁及以上的老年人 1.66 亿，占比 11.9%。其中患有慢性病的老年人 1.5 亿，占老年人总数的 65%。失能、半失能老年人 4400 万。基数巨大并且增速也是世界第一，失能、高龄、空巢老人的增多，使得很多带病生存的老年人对上门护理

服务需求激增。而且，人数庞大、行动不便的孕产妇、婴儿群体、长期慢病患者也有互联网+护理服务的强烈需求群体。

上门护理服务，能够解决患者连续服务的需求和重点人群帮扶照料的客观需求，解决家庭、社会复诊就医不便的难题，彰显医疗机构品质品牌和社会责任感。

4.3 经济可行

对患者：通过接受上门护理服务，降低了到院就医和接受护理的交通成本、时间成本、家属陪同成本、就诊生活成本支出；同时，通过延续服务，患者享受连续康复照料服务，有助于更快的病情恢复和健康康复。

对医院：通过上门护理服务的方式，将服务延续至患者家庭，对医院的经济效益提升实现 3 个方面的促进：一是护士上门带来的直接经济效益；二是根据现在各地开展服务的经验，还可以由护士上门服务带动检测和体检等服务的开展；三是通过上门服务的开展，以及同步检测和体检服务的开展，可以带动患者来院就诊量的增加。

4.4 技术可行

互联网+护理上门服务采用“线上申请、线下服务”的服务模式，需要护士进行上门或者联合基层医疗机构为被服务对象提供服务，需要强有力的信息化支撑，满足被服务对象身份认证、病历资料采集存储、服务人员定位追踪、个人隐私和信息安全保护、服务行为全程留痕追溯、工作量统计分析等质量安全管理需求。

当前，移动互联网技术、物联网技术（IoT）、大数据技术、OCR 技术、AI 人脸识别技术的发展，已经能够支持患者便捷下单、护士实时安全保障、远程应急处理指导、医院全局视图管理等服务和管理需求，为护士和患者提供多重人身安全保障，为服务提供多重质量安全保障。

4.5 试点效应良好

自 2019 年 2 月开始试点以来，北京、天津、上海、江苏、浙江、广东等 6 地试点已经纷纷制定了各地的实施细则、服务清单、价格和收费、监管和保障措施，组织各级医疗机构开始开展上门服务。

其中，广东省参与服务的医院最多，已经有约 140 家医院参与服务，各试点地区在服务开展 1 年以来提供了上万单的服务，从媒体消息来看，服务质量和人身安全有保障，表现出了良好的社会效应，起到了很好的便民惠民效果，获得了群众良好的认同和口碑。

全国其他开展服务的地区：除 6 个试点地区之外，河南（济源、荥阳、洛阳、商丘等多地）、湖北武汉、湖南长沙、新疆、江西南昌等多地自主开展了互联网+护理服务尝试。和试点地区的表现一致，也取得良好的认同和社会效应。

5 小结

“互联网+护理服务”是时代的产物，将医疗和互联网结合发展成为新的护理模式，打破传统的护理服务模式，与时俱进，顺应了时代发展的潮流；同时医院、患者互惠互利，极大地提高

了服务效率和患者的满意度；另外，对提升医院和护理人员形象产生了积极的作用，充分发挥了医院的护理能效。

PU-005

腹部 CT 图像肾脏分割方法综述

承垠林、周毅
中山大学中山医学院

目的 医学图像分割是各类医学图像应用的前提，肾脏 CT 图像分割在临床医学图像分析和非侵入式计算机辅助诊断中有着非常重要的意义，在肾脏功能分析，患者病情诊断和治疗决策等临床领域有着重要的价值。

方法 本文对国内外研究成果进行总结，对肾脏 CT 图像分割方法进行了分类和详细的介绍，并对各种方法进行了对比分析，阐述了肾脏 CT 图像分割技术的研究方向与趋势。

结果 近年来，国内外学者在医学图像肾脏分割领域做出的研究和贡献大致可以分为两类，一类是传统图像分割方法，另外一类是基于机器学习和深度学习的图像分割方法。

结论 总体而言，肾脏 CT 图像分割方法逐步由半自动分割发展到自动分割，分割的精度也日益提升。传统的分割算法计算效率高，且很少依赖训练数据，但分割效果不理想，需要引入先验知识或人工干预来提高分割的准确率；而基于深度学习的分割算法能够实现端到端的高精度分割，但需要大量标记样本进行模型训练才能使得模型具有较高的性能和鲁棒性。

1.引言

肾脏是泌尿系统的重要部分，具有重要的代谢功能，能够对血液状态以及骨骼的发育造成影响。因此若肾脏发生相关病变则会引起发育异常，内分泌紊乱等，严重的还可能导致死亡。因此使用诊断技术实现对肾脏疾病的早发现早治疗是十分必要的。肾脏的 CT 扫描可以提供与肾脏损伤，肾脏疾病等相关的更加详细的信息。而辅助泌尿科进行诊断和治疗的智能医学成像处理系统首要任务就是进行肾脏分割，这是诸多高级处理任务如体积测量和异常检测（囊肿，肿瘤等）等的基础。

肾脏分割是医学图像分割领域的一个特定研究方向，相对于普通分割，医学图像的分割更为复杂，除需考虑常规分割中关注的区域和纹理等特征外，还要充分考虑人体解剖结构的相关信息，而对肾脏分割而言，就需要考虑肾脏类椭球形状以及其位置分布对肾脏分割的约束和指导。目前广泛应用于医学图像分割的算法繁多，就肾脏分割而言，可大体分为两部分，传统的图像分割算法以及日益成熟的基于机器学习的图像分割算法。

2.传统图像分割算法在肾脏 CT 图像上的应用

传统图像分割算法按照分割特点的不同可分为基于边界的图像分割方法，基于区域的图像分割方法，基于图论的图像分割，基于活动轮廓的图像分割以及基于机器学习的图像分割。由于基于边界的分割算法是利用边界两侧图像局部特征（灰度、颜色、纹理等）的差异性来实现图像分割任务[1]，使得其具有一定的缺陷型，该方法对噪声敏感，且是针对全局图像进行卷积操作，无法获取特定目标，由于肾脏 CT 图像中存在其他的灰度相似或重叠的组织和器官，使得该方法无法很好的应用于肾脏 CT 图像，因此本文对该方法不再过多介绍。

2.1 基于区域的图像分割方法

基于区域的图像分割方法是基于相似性准则将图像分割为不同区域。主要包括阈值法、区域生长和区域分裂合并等方法。

2.1.1 阈值分割

阈值分割[2]是一种十分经典的基于区域的并行图像分割算法，其基于图像灰度特征选定一个或多个灰度阈值，然后将图像中每一像素灰度值与阈值进行比较，并按照比较结果将像素归到对应类别中。阈值法易于实现，计算速度较快，特别时当分割目标与背景的像素值分布相差较大时，该方法能取得很好的效果。常见的阈值分割方法有最大熵法[3]、类间方差法[4]（大津法，OTSU）和最小误差法[5]等。2010 年，赵于前等[6]使用 OTSU 阈值分割和适应性区域生长算法实现了对肾脏的分割。文章使用大津法对两肾脏的最优阈值进行确定，而后使用适应性区域生长法实现对肾脏的分割。

2.1.2 区域生长和合并分裂算法

区域生长算法[7]是通过预定义的生长准则将像素或子区域逐步聚合来实现对图像的分割，该方法分割效果取决于选定的种子点和设定的生长准则。合并分裂算法[8]是通过对图像不断的分裂和合并实现对目标的分割。区域生长和合并分裂算法都属于串行区域分割算法，两种方法对均匀连续的目标有很好的分割效果，但其易受噪声影响，分割目标时若渗透到相邻目标会导致相邻目标一同被分割出来。

2001 年，Regina 等人[9] 开发了一种自适应区域生长算法，该算法能够根据肾脏待分割区域的特征自动学习其同质性标准，并通过基于随机游走[10]（Random Walk）的样本位置更新实现了区域同质性标准的更新。2006 年，Lin 等人[11]结合解剖结构，提出了一种自适应区域生长算法。文章对 30 名患者 358 张 CT 图像进行测试，Dice 系数平均值为 88％。2014 年，Belgherbi 等人[12]提出一种肾脏分割方法，文章首先基于解剖和形态学知识来去除脊椎并划定 ROI 区域，而后使用标记物控制的分水岭算法进行肾脏分割，算法在腹部 CT 图像上进行了性能评估，平均准确率 99.00%，平均特异度 99.00%，平均灵敏度 99.00%。2018 年，Wieclawek[13]提出一种 3D 标记物控制的分水岭算法用于 CT 图像中肾脏分割。该方法在 170 组腹部 CT 图像上进行了测试，Dice 系数平均值为 91.68%。

2.2 基于图论的图像分割

基于图论的分割方法将图像分割转化为图的最优划分问题，通过寻找图的最优割集实现图像的分割。常见的基于图论的分割方法有流切割[14]（Flow Cut）、Grab cut[15]、Livewire[16]和Random Walk[10]等。2012 年，Linguraru 等人[17]基于几何统计模型提出一种构建 4D 方向图的算法，并使用图割算法对 4D 腹部 CT 中肝、脾和左右肾脏进行分割，模型在 20 张 CT 图像上进行了测试，四个器官的体积重叠率超过 93.6%。陈新建等人[18]结合了主动外观模型（active appearance model，AAM）、Livewire 和图割算法，提出一种用于腹部多器官分割的新的方法，模型在一个临床 CT 数据集以及 MICCAI 2007 Grand Challenge 肝脏数据集上进行了测试，其中 TPVF（true positive volume fraction）>94.3%，FPVF（false positive volume fraction）<0.2%。Hu 等人[19]基于条件随机场模型定义了图割中使用的能量函数，并使用图划分算法来解决能量函数最小化的问题，模型使用 10 组肝病患者 CT 图像和 8 组肾脏患者 CT 图像进行了验证，其中在肾脏分割结果中模型平均灵敏度为 92.5%，平均特异度为 99.5%。2014 年，Chen 等人[20]使用一组训练图像来构建主动外观模型，并使用 Livewire 对主动外观模型进行微调以初始化肾脏形状和位置，最后使用图割方法对肾脏和肾皮质进行分割。2016 年，时永刚等人[21]基于图割和水平集算法提出一种肾脏以及肾皮质自动分割方法，文章在初始切片图割分割结果的基础上使用形态学方法演化出相邻切片的候选分割区域，直到完成全部切片的分割，最后使用水平集算法对肾皮质进行分割。

2.3 基于活动轮廓的图像分割

基于活动轮廓的图像分割方法基本思路是通过构建能量函数并求解来实现图像分割的方法。活动轮廓模型主要包括参数活动轮廓模型和几何活动轮廓模型两种。参数活动轮廓模型又称蛇模型[22] 构建了一条参数化能量曲线，通过能量函数极小值的约束对曲线进行演化，实现图像的分割。几何活动轮廓模型又称水平集方法[23]，是一种使用轮廓曲线几何特性来推动曲线演化的方法。基于活动轮廓的图像分割方法结合了高层先验知识和底层图像特征，能够很好的对图像进行分割，但该类方法对局部极小值敏感，求解相对复杂[22]。2001 年，Tsagaan 等人[24]结合肾脏形状和灰度等先验知识提出一种活动轮廓模型用于肾脏分割。模型在 33 组腹部 CT 图像上进行了测试，平均对应度约为 87%，平均距离为 1.07 体素。2011 年，Khalifa 等人[25]提出了一种基于 CT 图像对肾脏进行 3D 分割的方法，作者提出一种三维统计模型来对水平集方法进行约束，进而实现肾脏的分割。模型在 21 组图像上进行了测试，Dice 系数平均为 97.0%。2013 年，张品等人[26]提出一种基于活动轮廓和图割方法的能量最小化分割模型实现对肾脏组织的分割。模型综合了测地活动轮廓模型和 Chan-Vese 模型，并使用图割方法对离散化能量函数进行优化。模型在 30 组腹部 CT 序列图像上进行了评估，分割结果的平均 Dice 系数为 93.7%。2017 年，Skalski 等人[27]基于水平集框架提出了一种用于肾脏分割的新方法，作者在原始水平集框架的基础上引入了椭圆形状约束来限制分割曲面泄露到其他区域。该方法在 10 组肾癌患者 CT 图像上进行了测试，Dice 系数为 86.2%，平均 Hausdorff 距离为 2.37mm。

3.机器学习在肾脏 CT 图像上的应用

机器学习来源于早期的人工智能领域，机器学习可分为传统机器学习算法和深度学习算法等，传统的算法包括逻辑回归、决策树[28]、聚类[29]、支持向量机[30]、Adaboost[31]等等。深度学习算法是基于深度神经网络构建的算法。机器学习算法通过将图片中像素进行分类从而实现在图像分割领域的应用。

3.1 基于传统机器学习方法的图像分割

机器学习算法从学习方式上大致可分为监督学习、无监督学习和弱监督学习等。其中监督学习需要通过训练数据对算法模型进行训练，训练数据需要给出样本特征和对应标注结果，而无监督学习则不需要给出标注信息，模型基于样本相似度对样本进行分类。常见的监督学习算法有决策树、支持向量机、Adaboost 等，常见的无监督学习算法有 k-均值聚类、模糊聚类、自编码器、期望最大化和受限玻尔兹曼机等。2012 年，Cuingnet 等人[32]使用随机森林算法对肾脏位置进行初步定位，而后基于肾脏区域概率图使用形变模型分割 CT 图像中肾脏区域。模型在 233 张 CT 图像上进行了性能评估，Dice 系数>90%。2016 年，Jin 等人[33]基于 AAM（Active Appearance Models）和随机森林算法对肾脏进行自动分割，作者结合 3D GHT（Generalized Hough Transform）和 AAM 算法实现对肾皮质的定位，而后使用改进的随机森林算法分割出肾脏中肾皮质、肾柱、肾髓质和肾盂等部位，算法使用留一法在 37 组 CT 图像上进行了评估。

3.2 深度学习方法

基于卷积神经网络（CNN）的训练方法可以大大提高图像分割的准确性[34]，诸多优秀的图像分割算法如全卷积神经网络（FCN，Fully convolutional networks）[35]，U-Net[36]，Deeplab[37]，扩散卷积[38]以及以条件随机场为代表的后处理算法等也相继被提出。此外诸如 VGGNet[39]、ResNet[40]、DenseNet[41]等图像分类模型也通过迁移学习的方式应用到医学图像分割的特征提取中来。2016 年，Thong 等人[42]设计了一种卷积网络，其能在低分辨率和高分辨率两种模式下对 CT 图像中肾脏进行分割，作者使用 20 张图片进行了验证，左右肾脏分割结果的 dice 系数分别为 94.53%，93.07%。2017 年时永刚等人[43]基于 VGG-16 模型和迁移学习方法提出一种全卷积神经网络 K-Net，用于获取 GrowCut 的种子点标记图，而后使用 GrowCut 对肾皮质进行分割。模型分割结果的平均 Dice 系数为 91.79%。2019 年，徐宏伟等人[44]基于残差双注意力 U-Net 对囊肿肾脏进行自动分割。作者设计了一种残差双注意力模块，并将其用于 U-Net 上下采样模块中，左右肾脏分割结果的 Dice 系数分别为 96.05%，94.35%。Xia 等人[45]将 ResNet-18 CNN 集成到 SCNN（Siamese convolutional neural network）架构中，实现对已标注 CT 图像和未标注图像的相关性检索和像素区域的对应，而后使用 SIFT-Flow 模型对两图像建立像素的稠密匹配，进而实现肾脏和肾肿瘤的分割。模型在 15,500 张 CT 图像上进行了训练和测试。

4.总结

随着计算机性能和运算能力的不断提升，分割方法的不断进步，目前 CT 图像上分割肾脏的方法繁多而复杂，从简单的阈值分割，到复杂的 U-Net，分割方法逐步由半自动分割发展到自动分割，分割的精度也日益提升。总体而言，传统的分割算法对计算资源要求较低，计算效率高，且很少依

赖训练数据，易于部署和落地，但分割效果不理想，需要引入先验知识或人工干预来提高分割的准确率；而基于深度学习的分割算法能够实现端到端的高精度分割，但需要大量标记样本进行模型训练才能使得模型具有较高的性能和鲁棒性，对于医学图像通常需要使用数据扩增算法来提升样本数量。就目前而言，基于深度学习的图像分割方法成为主流趋势，因此我们需要构建优质的肾脏 CT 数据库，对医学图像扩增算法进行研究，改进和开发图像分割模型架构，探索模型压缩方法，最终实现模型的落地和应用。

PU-006

癫痫脑电自动发作预测研究综述

马梦楠、张亮军、承垠林、韦晓燕、 周毅
中山大学中山医学院

目的 近些年来，基于癫痫脑电图（electroencephalogram，EEG）的癫痫发作自动检测和诊断方面取得了重大的进展，许多创新算法已经应用于癫痫的辅助诊断和治疗中。随着研究的深入，癫痫自动检测的重点转向癫痫的发作预测。

方法 本文通过文献调查的方式，调研了近几年的癫痫脑电发作预测的文献。

结果 本文阐述了机器学习和深度学习在癫痫自动发作预测方面的应用，重点介绍了最近关于癫痫发作预测的研究结果。

结论 通过回顾性的研究，本文希望能够为癫痫诊断和发作预测技术开发的调整和扩展带来新的思路。

1、引言

癫痫发作是由大脑神经元的突然性同步放电引起，通常伴随着意志丧失和相关的非自愿异常运动。根据世界卫生组织（WHO）的统计，全球约有近 5000 万的人口患有癫痫疾病，且全球每年约有 240 万的新生病例产生[错误!未找到引用源。]。

癫痫疾病的不可控性是一大难题，虽然抗癫痫药物可以在许多癫痫患者中获得完全的癫痫发作控制，但多达三分之一的癫痫患者会患有医学难治性癫痫症（发作不可控），药物虽然可以减少癫痫发作的频率，但不能完全控制癫痫发作[错误!未找到引用源。]。由于癫痫发作的不可控性，这些患者因无法操作机动车辆等一些现实工具而在某些方面无法获得有意义的就业机会。此外还有强有力的证据表明，患有长期不可控的癫痫的病人患有永久性记忆障碍，抑郁，焦虑，自杀和其他精神疾病的 风险相比于其他人有所增加[错误!未找到引用源。]。上述因素对癫痫患者的生活质量产生负面影响，不仅影响家庭成员和看护人的生活，并且对家庭成员的医疗保健成本构成经济负担。

近几年的研究中，基于癫痫脑电图（EEG）的癫痫疾病的诊断和癫痫的发作检测等方面已取得重大进展。非线性动力学、机器学习、深度学习等技术也已应用到这一块的研究中去，且取得了

不错的研究结果，可以高度准确的帮助癫痫的诊断。Mark J Cook 团队在 13 年发表在柳叶刀上的一篇杂志，首次应用在人体身上的癫痫自动预测系统能够有效的预警癫痫的发作，为这一方面的研究打下坚定的基础[错误!未找到引用源。]。

准确的癫痫发作预测不仅提前预测患者可能发生癫痫发作的时间，还将在临床癫痫治疗上提供辅助，在设定预警的时间范围内给予一定的药物治疗等。癫痫发作预测显著改善患者的生活质量并减轻护理人员的负担，具有十分重要的现实意义。

2、传统的机器学习癫痫自动检测

机器学习涉及众多学科领域，借助学习算法能够从输人的经验数据中获取潜在工作机制的模式或特征。机器学习既可用于开发临床应用的计算机辅助诊断工具，又能探索疾病的病理生理相关机制[错误!未找到引用源。]。癫痫脑电数据的分析过程包括数据采集、预处理、特征提取选择和模型构建等环节，机器学习法均可应用于这些环节[错误!未找到引用源。]。

图 1 癫痫自动检测系统框架图

如图 1 所示，是一个基于机器学习的癫痫自动检测的系统框架图。该框架采用机器学习结合特征值提取的方法通过识别癫痫 EEG 信号中的不同状态来预测癫痫的发作。数据预处理，特征提取，训练分类器，决策等模块构成近年来研究中最常用的癫痫自动检测系统。

预处理：数据预处理通常发生在特征提取之前，旨在对收集到的数据进行噪声和伪像的去除，来获得干净的 EEG 信号进行下一步的处理。

特征提取：目前已经有很多的技术较为成熟的应用在特征提取的研究中，如表 1 所示，比如统计特征中已经使用诸如均值、方差、偏度值和峰度值的统计参数来提取 EEG 信号[错误!未找到引用源。]，研究癫痫病人不同状态期的特征值的统计学差异；非线性特征是当下研究应用最多的特征值，近似熵（ApEn）是一种用于量化时间序列波动的规律性和不可预测性的非线性动力学参数，它用一个非负数来表示一个时间序列的复杂性，反映了时间序列中信息发生的可能性。Zheng zhang 等利用不同时期癫痫病人脑电的近似熵差异，结合机器学习技术，采用支持向量机（SVM）做分类器，在癫痫脑电不同时期的分类研究中取得了不错的研究成果[错误!未找到引用源。]。相关维度（CD）通常应用于 EEG 时间序列数据以测量不同发作模式的序列复杂性，Ouyang 团队 07 年发表的文章中应用了相关维度概念[错误!未找到引用源。]。

表 1 机器学习常用特征部分统计表

方法	序号	特征

统计学特征	1	算数平均值
	2	最大值
	3	最小值
	4	标准差
	5	方差
	6	中位数
时频分析特征	1	时域分析
	2	频域分析
	3	时频分析
非线性特征	1	近似熵
	2	样本熵
	3	赫斯特指数
	4	李雅霍诺夫
	5	最大模式经验分解
	6	PCA
	7	小波分解
	8	小波包分解

分类与决策：分类与决策是癫痫自动检测系统中重要环节现在已经应用研究的分类器多如线性判别分析（LDA），SVM，随机森林（RF），贝叶斯，多层感知器（MLP），MLP 神经网络和径向基函数神经网络[错误!未找到引用源。]，反向传播神经网络（BPNN）和概率神经网络（PNN），各有优略，如表 2 所示。

表 2 机器学习分类器的优缺点

分类器	优点	缺点
线性判别分析（LDA）	1）可以使用类别的先验知识； 2）以标签，类别衡量差异性的有监督降维方式，相对于 PCA 的模糊性，其目的更明确，更能反映样本间的差异；	1）LDA 不适合对非高斯分布样本进行降维； 2）LDA 降维最多降到 k-1 维； 3）LDA 在样本分类信息依赖方差而不是均值时，降维效果不好； 4）LDA 可能过度拟合数据。
SVM	1）可以解决小样本下机器学习的问题。 2）提高泛化性能。	1）对缺失数据敏感。 2）内存消耗大，难以解释。 3）运行和调差略烦人。

	3）可以解决高维、非线性问题。超高维文本分类仍受欢迎。 4）避免神经网络结构选择和局部极小的问题。	
随机森林（RF）	1）能够处理很高维度的数据，并且不用做特征选择 2）训练速度快，容易做成并行化方法； 3）在训练过程中，能够检测到feature间的互相影响 4）对于不平衡的数据集来说，它可以平衡误差。	1）对于小数据或者低维数据（特征较少的数据），可能不能产生很好的分类 2）过拟合 3）可能有很多相似的决策树，掩盖了真实的结果
贝叶斯	1）所需估计的参数少，对于缺失数据不敏感。 2）有着坚实的数学基础，以及稳定的分类效率。	1）假设属性之间相互独立，这往往并不成立。 2）需要知道先验概率。 3）分类决策存在错误率。
神经网络	1）分类准确率高。 2）并行处理能力强。 3）分布式存储和学习能力强。 4）鲁棒性较强，不易受噪声影响。	1）需要大量参数（网络拓扑、阀值、阈值）。 2）结果难以解释。 3）训练时间过长。

Varatharajah 等人在 2017 发表的文章中使用机器学习技术研究癫痫发作预测的可能性，结合非线性动力学理论对比了支持向量机（SVM）、人工神经网络（ANN）、随机森林（RF）等分类器在癫痫脑电发作预测的优略，验证了机器学习技术在预测癫痫发作可能性上的能力[错误!未找到引用源。]。Jeff Craley 等人在 2018 年发表的文章中采用隐马尔科夫模型定义了一个癫痫传播的路径网络，模拟癫痫随时间通过大脑的空间动力学，且基于这种癫痫传播路径网络对癫痫脑电的发作间隔分类进行研究，在传统的预警之外还能为病灶区的定位提供空间信息[错误!未找到引用源。]。

3、深度学习的癫痫自动检测

深度学习（DL）一词最初在 1986 年被引入机器学习（ML），是机器学习的一个分支或子领域，后来在 2000 年时被用于人工神经网络（ANN）[错误!未找到引用源。]。人工神经网络是深度学习的基础，神经元是神经网络的组成单位。神经元在接受线性组合的输入后，经过激活函数的激活后进行非线性变换后的输出，当多个神经元聚集到一起时便是我们常说的人工神经网络。由于深度具有更强

的表示能力和泛化性能以及自我学习数据的优点，近几年的研究中越来越的人将其应用在癫痫脑电数据的研究中。

如图 2 所示，是采用卷积神经网络（CNN）[错误!未找到引用源。]进行癫痫自动预测研究的框架图。卷积神经网络是目前为止最成熟的深度学习技术，它已被广泛的应用于各种场景的研究，如医学图像识别，生理信号分析 ，阿尔茨海默病的诊断，以及健康监管等。CNN 是由卷积层、池化层和全连接层三层结构构成其基本结构。在卷积层中，事前设定好的特征即卷积核的数量与输入的 EEG 信号进行卷积进行信息的特征提取，在信号输入的方式中，Petrosian 团队使用的是一维时间序列，直接对最原始的脑电信号进行学习[错误!未找到引用源。]，而 Isabell Kiral-Kornek 团队则是对原始的脑电的信号进行预处理，将一维信号进行增维，亦取得了不错的研究成果[错误!未找到引用源。]。在 CNN 的卷积过程中，卷积神经网络选择预先设定的不同的特征进行卷积并生成特征图（输出）。由于卷积之后生成的特征图具有数据大、过拟合等缺点，卷积层之后实现的池化层一方面旨在控制网络的过度拟合，另一方面旨在控制数据的大小。现在最常使用的池化层是最大值池化，最大值池化在使特征图变小，简化网络计算复杂度的同时；进行特征压缩，最大化的提取主要特征。最终层简单而言是全连接层将所有神经元连接在一起进行预测[错误!未找到引用源。]。通过 CNN，可以直接从 EEG 数据中自动学习重要功能，不需要传统的手工标记特征。一个病人纵向数据可用性的有限原因，使得其他模型不能够有效的搭建特定病人的算法，但是该模型能够从训练 EEG 信号中自适应地学习以获取用于预测癫痫发作前的 EEG 事件。

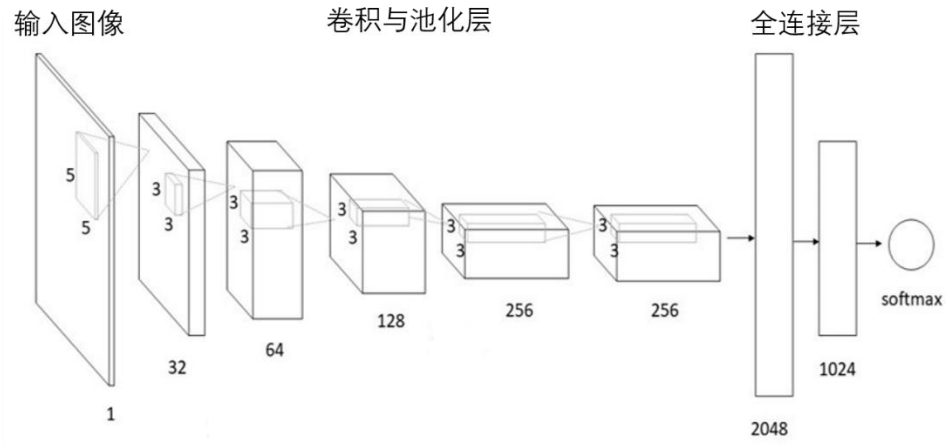


图 2 卷机神经网络结构框架图

深度学习技术的快速发展为癫痫脑电的自动检测研究带来了巨大的推动力，快速计算的特性为面向移动系统的自动癫痫检测的研究做出重大贡献。Mark J Cook 团队发表于柳叶刀上的文章[错误!未找到引用源。]进行了第一次的人体实验，不仅是在癫痫研究方面更进一步，更是证明自动癫痫预测研究的社会价值。文章采集了多名患者两年的颅内脑电数据，并利用深度学习结合采集到的数据创建一个特定与患者的预警算法，根据性能标准进行评估，在实验的 15 名患者中，直有三名患者因为一些原因没有达到预期。Mark J Cook 团队开创性的研究为后续的研究打下基础，Isabell Kiral

Kornek 等人在 Mark J Cook 团队数据的基础上作进一步的完善研究[错误!未找到引用源。]，文章采用卷积神经网络的算法部署，EEG 数据预处理上采用频谱图作为输入，医学上已经证明正常与患者脑电频谱存在统计学上的差异，并将特定于患者的算法部署到更加适合移动系统的芯片上，亦取得了非常好的研究效果。

4、癫痫自动检测的挑战与难题

癫痫的发作是一种疾病的症状，我们不能简单地理解为一种机制[错误!未找到引用源。]。因此对于现在研究中合理假设这种机制可能存在于不同患者，甚至“一刀切“的假设存在于所有患者的假设研究是现在研究中的难题。我们必须清醒的认识到算法横向发展（患者之间算法的适用性）的局限性，而特定于患者的算法部署，单纯只依靠患者本身的发作次数的数据记录（纵向数据），却不足以验证和创建特定于患者的癫痫预测算法。基于此，在研究过程中必须横向汇集数据对算法进行验证统计，正是因为这种误差的存在，在患者之间汇集数据的做法有可能并不是一种有效的策略。

在阅读和整理文献的过程中，发现癫痫预测方面的文献主要是由具有信号处理背景的电子工程师或者计算机科学家撰写的文章。这些研究很有创建性的结合医工特点，但是却很少有文章对癫痫发病症状的解释。工程师在癫痫自动检测研究上创造力十足，越来越多的算法应用到灵敏度、准确度等的提升，深度学习算法也是其一。但是现在的发展来说，深度学习就如一个“黑箱”，缺乏对其内在机理的研究。因为黑箱的原因，目前看来，人们缺乏对更复杂的预测算法的信心，成为预测算法转化社会价值的绊脚石。

5、总结

癫痫的自动预测研究能够有效的改善和提高患者的生活质量和精神面貌。本文主要是回顾了机器学习和深度学习技术在癫痫自动预测算法研究上的应用，以及面临的难题和挑战。未来发展中，深度学习将是研究的重心，但是如何转化其社会价值，改善患者生活仍将是我们研究的最终目标。

PU-007

基于 ARIMA-LSTM 组合模型的月门诊量预测精度研究

李琳、周毅
中山大学中山医学院

目的 为了探讨慢性阻塞性肺病的发病数的变动规律，提高慢性病阻塞肺病的月门诊发病数预测精度，为慢性阻塞性肺病预防与控制提供决策依据。

方法 以 2010 年 1 月-2017 年 7 月新疆地区慢性阻塞性肺病的月门诊数作为原始分析序列，提出了 ARIMA 和 LSTM（长短记忆神经网络）混合模型 ARIMA-LSTM。并比较 ARIMA 和 LSTM 的预测准确性。

结果 ARIMA-LSTM 模型能较好地拟合出新疆地区慢性阻塞性肺病的新发数量和季节性趋势，RMSPE 的值为 19.71%。

结论 ARIMA-LSTM 组合模型比单个 ARIMA、LSTM 模型预测精度高，能够提高慢性阻塞性肺病日门诊发病数预测的准确性。

1 引言

慢性阻塞性肺疾病（chronic obstructive pulmonary disease，COPD）是一种具有气流阻塞特征的慢性支气管炎和（或）肺气肿，可进一步发展为肺心病和呼吸衰竭的常见慢性疾病**错误!未找到引用源。**。全球 73.37 亿人口中慢阻肺患者高达 6 亿，其中每年有 275000 人因患慢阻肺而失去生命。在中国，2700 万人患有慢阻肺。据介绍，慢阻肺目前在全球范围内的死亡率排在单病种的第三位，与高血压、糖尿病、癌症等病种组合成了我国居民最主要的慢性病[1]。科学 、准确地分析医院慢性阻塞性肺疾病月门诊量动态变化、预测趋势 ,对于合理安排医院医生、财、物资源, 提高经济效益和社会效益都有重要意义。然而医院门诊为居民提供早期诊断 ,开展及时治疗, 具有服务面广、随意性大和可控性小等特点。门诊人次与多种因素相互关联 , 既有长期趋势、季节效应, 又有随机干扰因素 ,因而应用传统的预测方法会受一定程度的局限。

传统的时间序列分析方法有：自回归模型（AR）、滑动平均模型（MA）、自回归移动模型（ARMA）以及广义自回归条件异方差模型（GARCH）等，最具代表性的时间序列分析方法为差分自回归移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)，它是美国学者 Box 和英国统计学家 Jenkins 于 1970 年提出的基于线性数据的模型**错误!未找到引用源。错误!未找到引用源。**，在处理线性数据样本时具有独特的优势，但无法描述非线性数据的信息。在非线性系统中，也可采用人工智能的方法，如：神经网络、模糊回归及支持向量机等模型。这些预测方法虽然有各自的特点，但单项预测方法只能从某个角度提供相应的有效信息，无法反应信息源的广泛性。Bates 和 Granger 于 1969 年首次提出组合预测方法的概念，组合模型法是将不同预测方法所得的预测结果组合起来形成一个新的预测结果的方法，可以有效提高预测精度。

长期短期记忆网络（Long Short Term Memory，LSTM）是一种特殊类型的深度学习模型，该模型能够保持状态并识别时间序列上的模式。在每次处理过程中，递归架构能使状态在更新权重时保持或者传递下去。因此，LSTM 模型的单元架构在短期持久化的基础上实现了长期持久化。该模型对涉及自相关性的序列预测问题很有用。因此，本文采用 ARIMA-LSTM 组合模型对 COPD 的月门诊量进行预测，以 ARIMA 模型揭示月门诊量变动的线性规律，以 LSTM 模型揭示月门诊量变动的非线性规律，并对月门诊量时间序列进行预测及比较预测精度，以验证组合预测模型的精确性和有效性。

2 方法

2.1 ARIMA 模型

ARIMA 模型，记为 $ARIMA\ (p,d,q)\times(P,D,Q)s$。它是由 ARIMA（p,d,q）模型的基础上

增加了季节性趋势。其中，p 是自动回归的阶数，d 是几阶差分，q 是移动平均阶数； P 是自动回归阶数，D 是差分阶数，Q 是移动平均阶数，S 是周期步长。因为短期相关性和季节性是乘积关系,则该模型的完整结构如下：

$$\nabla^d \nabla_S^D x_t = \frac{\Theta(B)\Theta_S(B)}{\phi(B)\phi_S(B)}\varepsilon_t$$

式中

$$\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$
$$\Phi(B) = 1 - \phi B - \cdots - \phi_p B^p$$
$$\Theta_S(B) = 1 - \theta_1 B^S - \cdots - \theta_Q B^{QS}$$
$$\Phi_S(B) = 1 - \phi_1 B^S - \cdots - \phi_P B^{PS}$$

其中， x_t为t时刻非平稳时间序列，ε_t为零均值白噪声序列，B 是延迟算子，Δ_t表示 d 阶差分，$\Phi(B)$表示自回归系数，$\Theta(B)$表示移动平均系数，$\Theta_S(B)$为季节移动回归系数，$\Phi_S(B)$为季节自回归系数。

2.2 LSTM

LSTM 网络是一种特殊的循环神经网络(Recurrent Neural Network, RNN)类型，是为了解决模型梯度弥散的问题而提出的**错误!未找到引用源。**。LSTM 弥补了 RNN 的梯度消失和梯度爆炸、长期记忆能力不足等问题，使得循环神经网络能够更加有效地利用长距离的时间序列信息**错误!未找到引用源。**。

其引入了记忆门单元和门限限制，实现了对长距离信息的有效利用，并解决了梯度消失的问题。t 时刻，给定输入 x，LSTM 的隐层的输出表示 h_t 的具体计算过程如下：

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$
$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$
$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \odot \tanh(c_t)$$

其中，W 表示连接两层的权重矩阵，b 表示偏移列向量。LSTM 将信息存放在循环网络正常信息流之外的门控单元中。这些单元可以通过开关判断存储的信息内容以及写入或读取信息的时机。图 2 所示是 LSTM 网络单元的的结构。f 遗忘门，表示对于当前时刻的输入 x，它决定了从上一时刻传来的信息要丢弃的部分。i 表示输入门，它决定在 t 时刻应该更新哪些值，$\tilde{c}$ 是一个候选值的向量，将 i 和$\tilde{c}$ 组合起来得到 C 对神经元状态进行更新。o 是输出层，决定神经元状态需要输出的部分。h 是整个网络的输出。

2.3 ARIMA-LSTM 组合模型

本文采用 ARIMA-LSTM 组合模型对 COPD 的月门诊量进行预测，把月门诊量时间序列看成是线性自相关结构和非线性结构两部分构成。首先，用 ARIMA 模型预测 COPD 的月门诊量数据，ARIMA 模型的预测结果包含数据的线性特征，非线性特征存在于 ARIMA 模型的预测误差中**错误!未找到引用源。**。其次，用 LSTM 模型对 ARIMA 模型的预测误差进行预测，使非线性特征包含在 LSTM 的预测结果中。最后，用 ARIMA 的预测结果与 LSTM 的预测结果相加得到组合预测模型的预测值。

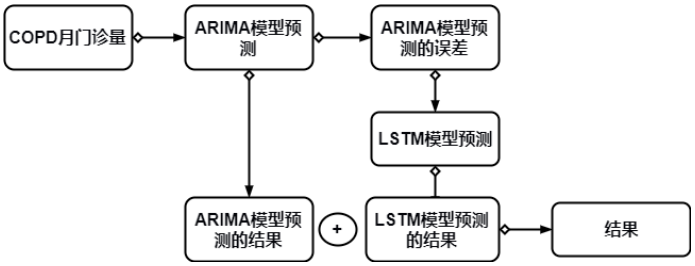


图 1 ARIMA-LSTM 组合预测模型流程图

具体预测步骤如下:

步骤 1: 用 ARIMA 模型对线性自相关结构 l_t 进行预测，预测结果为 $\hat{l}_t$ ，原序列与模型预测结果的残差 e_t，即：

$$e_t = y_t - \hat{l}_t$$

步骤 2: 根据步骤 1 得到的残差序列 e_t，利用步骤 1 的模型的阶数确定输入残差阶数，利用 LSTM 模型对残差进行预测，预测结果为 $\hat{e}_t$

步骤 3: 利用两种模型的预测结果组合成为最终预测的结果 $\hat{y}_t$，即 $\hat{y}_t = \hat{l}_t + \hat{e}_t$ 至此，对时间序列的线性和非线性部分分别建立 ARIMA 和 LSTM 模型，并将两者预测结果有机结合起来，能够充分发挥 ARIMA 与 LSTM 的各自优势，取长补短，优势互补。

3 结果

本文慢阻肺发病数据来自新疆乌鲁木齐某三甲医院 2010 年 1 月-2017 年 7 月门诊就诊人次真实数据。数据情况如图 1，从图中可以看出从 2010 年 1 月-2016 年 12 月门诊就诊人次呈递增趋势。在本次预测模型中，2010 年 1 月-2016 年 12 月的数据作为训练集，预测 2017 年 1 月-7 月的月门诊量。并采用均方根百分比误差（RMSPE）可用于评价模型的可靠性。

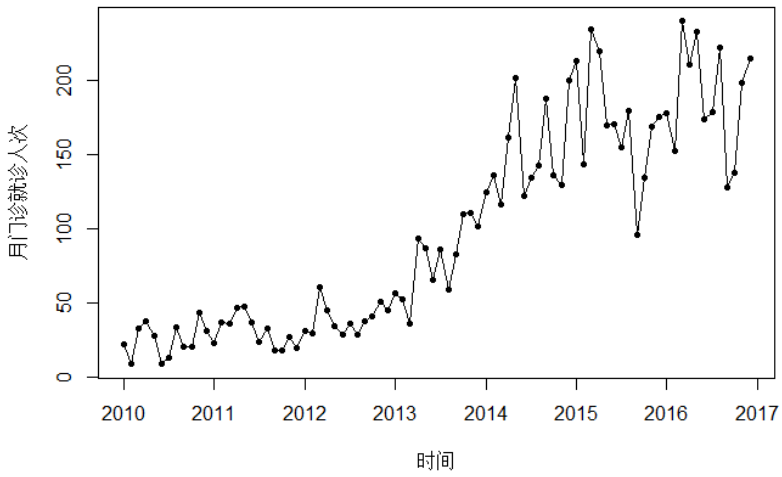


图 2 新疆地区慢性阻塞性肺病每月门诊就诊人次

3.1ARIMA 模型预测结果

在建立 ARIMA 模型之前，需要对时间序列模型进行平稳性检验和纯随机性检验。本研究采用时序图和自相关图结合假设检验（单位根检验 p=0.296）的方法判断该时间序列是非平稳时间序列。使用 1 阶 12 步差分，之后再次对序列进行单位根检验，p 值为 0.013，获得平稳性时间序列。采用 Box.test 统计信息计算 Q 统计量和 LB（Ljung-Box）统计量，判断该 COPD 时间序列不是白噪声序列。因此对该时间序列进行 ARIMA（p,d,q）×（P，D，Q）模型建模。根据时间序列的自相关图和偏自相关图，确定该时间序列最好的模型时 ARIMA（2,1,2）×（0,1,1），采用 LB（Ljung-Box）检验统计量进行模型残差检验，其 1-12 阶延迟下的 LB 统计量的值均大于 0.05，则该模型的残差序列为白噪声序列，这样的模型信息是提取是充分的。利用该模型对 2017 年 1 至 7 月的门诊数进行预测。结果见表 1。

3.2 ARIMA-LSTM 预测结果

用 LSTM 模型对 ARIMA 模型的残差进行建模预测，使用最优参数下训练得到的模型。最后根据公式 $\hat{y}_t = \hat{l}_t + \hat{e}_t$ 得到组合模型的预测结果，见表 1。

表 1 三种预测模型的预测结果比较

时间	实际值	ARIMA 模型预测结果	LSTM 模型预测结果	ARIMA-LSTM 组合模型预测结果
2017 年 1 月	185	188	171	188
2017 年 2 月	260	175	205	172
2017 年 3 月	294	257	244	263

2017 年 4 月	313	237	229	236
2017 年 5 月	296	230	217	234
2017 年 6 月	212	196	239	206
2017 年 7 月	222	202	214	212

从表 2 中 RMSPE 指标可以知道，非线性的 LSTM 模型比线性的 ARIMA 模型的预测效果好，ARIMA-LSTM 组合预测比单个的 ARIMA 和 LSTM 模型预测精度高，说明 ARIMA 和 LSTM 模型分别对线性和非性性问题有其相对的优势，但它们对于复杂的、不稳定的时间序列都不是最优的模型。ARIMA-LSTM 组合模型很好地结合了 ARIMA 和 LSTM 模型各自的优势，有效改善了预测精度，提高了预测的准确性。

表 2 三种预测模型的预测精度比较

模型	RMSPE
ARIMA 模型	20.23%
LSTM 模型	20.01%
ARIMA-LSTM 组合模型	19.71%

4 结论

慢阻肺是一种严重的慢性呼吸系统疾病，患病人数多，病死率高。由于慢阻肺呈缓慢进行性发展，严重影响患者的劳动能力和生活质量。目前慢阻肺在全球已成为第四位致死原因，引起了世界各国的重视。在我国，慢阻肺同样也是一种常见病，严重影响了广大人民的身体健康。2002 年，慢阻肺是全球第 5 位引起死亡的疾病，预计到 2030 年，将上升至全球死亡原因的第 3 位，带来的经济负担将位于世界疾病负担的第 5 位。在我国，慢阻肺为主要疾 病组成的慢性呼吸系统疾病多年来在农村中居第 1 位致死性疾病，在城市中居第 4 位。流行病学资料显示，我国 40 岁以上人群中慢阻肺的患病率达 8. 2%。

时间序列分析能够展示研究对象在一定时期内的发展变化趋势与规律，因而可以从时间序列中找出变量变化的特征、趋势以及发展规律，从而对变量的未来变化进行有效地预测。慢阻肺的发病情况有着一定的季节性和周期性，而 ARIMA 乘积季节模型主要拟合具有趋势性和季节性规律波动的序列，并结合了自回归和移动平均法的长处，具有不受数据类型束缚和适用性强的特点，是短期预测较好的模型。但是 ARIMA 模型时要求时间序列是平稳的，且不是白噪声序列。对于非平稳的时间序列 ,通常采用对序列进行差分,而达到平稳。LSTM 时间网络则不受控于时间序列的平稳性，可以直接对序列进行训练且可以得到较好的结果。

本文使用 ARIMA-LSTM 组合模型来预测 COPD 月门诊量变动，该模型很好地结合了传统预测和神经网络预测的各自优势，以 ARIMA 模型预测 COPD 月门诊量变动的线性趋势，LSTM 模型预测月门诊量

变动的非线性规律。实例表明，该组合模型能提高预测的精确度，准确把握农月门诊量变动趋势，比单个模型的预测结果更合理、更可靠**错误!未找到引用源。**。

PU-008

基于独立循环神经网络改进残差架构的癫痫脑电分类研究

马梦楠、承垠林、韦晓燕、陈子怡、周毅
中山大学中山医学院

目的 癫痫是神经系统的疾病之一，在全球范围内人口基数众多。传统的确诊手段绝大部分依赖于专职神经科医生对脑电图的阅读，耗时、效率低下且具有主观性。近年来，使用深度学习来实现癫痫患者的脑电的自动诊断吸引了越来越多的关注，但是深度神经网络在癫痫发作检测方面的潜力尚未得到充分的开发。

方法 这篇文章中，我们在传统的残差网络的架构中，使用一维卷积神经网络来进行替换，并组合独立卷积循环神经网络形成新的循环残差网络，实现癫痫脑电的自动诊断。首先，使用一维卷积神经网络的残差网络架构学习脑电的重要特征，然后使用循环神经网络学习序列间的关系，最后输出分类结果。

结果 我们的方法在伯恩大学的小样本数据集与真实世界的私人数据集上都取得不错的效果，得到100%的准确率。

结论 我们提出的模型在理想条件与真实世界中都能快速准确的识别出脑电的种类，能够为临床癫痫脑电的检测提供自动检测能力，希望能够癫痫脑电的发作预测提供积极的意义。

1 引言

癫痫是一种慢性脑功能障碍综合征，全世界有近 6500 万人受到癫痫的困扰，约占世界人口的1%[1]。引起癫痫的病因多种多样，病程长期反复，不仅严重危害患者健康，还会给患者带来巨大的精神压力[2]。脑电图的发展提供了一种无创和低成本的有效技术，可以在临床中试验探测大脑皮层的脑活动和相关疾病[3, 4]。大量研究表明，癫痫脑电与正常脑电显著不同。癫痫患者的脑活动通常包括发作间歇期和发作期，当脑活动从一种状态向另一种状态演变时脑电信号会发生较为明显的变化，因而，脑电图是临床诊断癫痫的重要依据。

专职神经科医生的脑电图阅读是目前确定癫痫病症的主要手段。但人工观察和检测长程脑电信号是一项复杂而耗时的工作，繁重的工作量容易引起医护人员疲劳而导致人工检测结果的不准确。而且往往要通过专业经验来做出判断，因此脑电图的自动检测与分类就越发显得迫切与重要。

近几年的研究中，基于癫痫脑电图（EEG）的癫痫疾病的诊断和癫痫的发作检测等方面已取得重大进展。非线性动力学、机器学习、深度学习等技术也已应用到这一块的研究中去，且取得了不错的研究结果。基于脑电时间序列的循环神经网络[5]与基于二维脑电的卷积神经网络[6]是近年来

的研究热点，前者关注于时间序列的前后关系，后者则更关注于脑电的特征提取。不同于传统的小剂量计算研究，深度神经网络更加关注于的大数据研究，但是依旧存在一些现阶段不能解决的问题。

本文结合循环神经网络与卷积神经网络的优点，提出了一种新的神经网络的架构—循环卷积神经网络。这种新的架构采用一维卷积神经网络提取原始脑电的特征，然后使用独立循环神经网络学习时间序列之间的关系，最终由分类函数实现对癫痫脑电的自动诊断。我们使用波恩大学数据和私人数据集进行训练与测试，为了避免癫痫诊断算法中的个体特异性，我们保证实验所使用的数据集来自于不同的个体，并与相关的前沿研究进行比较，我们所提出的方法普遍取得了较好的效果。二分类任务中甚至在准确率上得到 100%的结果。这些结果说明我们所提出的新的架构能够有效实现癫痫脑电的自动检测。

2 相关工作

2.1 卷积神经网络预测模型

卷积神经网络（CNN）是一类包含卷积运算且具有深度结构的前馈神经网络，具有局部感知和参数共享特性，是图像处理领域最热的研究点只之一[7]。一个完整的 CNN 包含卷积层、降采样层和全连接层[8]，如图 2 所示。卷积层通过卷积运算使得输入的原始脑电信号的某些特征得以增强，并降低噪声的影响；降采样层也就是我们常说的池化操作，可以减少数据处理量的同时保留有用的信息；全连接层得到神经网络提取到的图片特征，图 2 中全连接层经由“softmax”输出最终的结果。

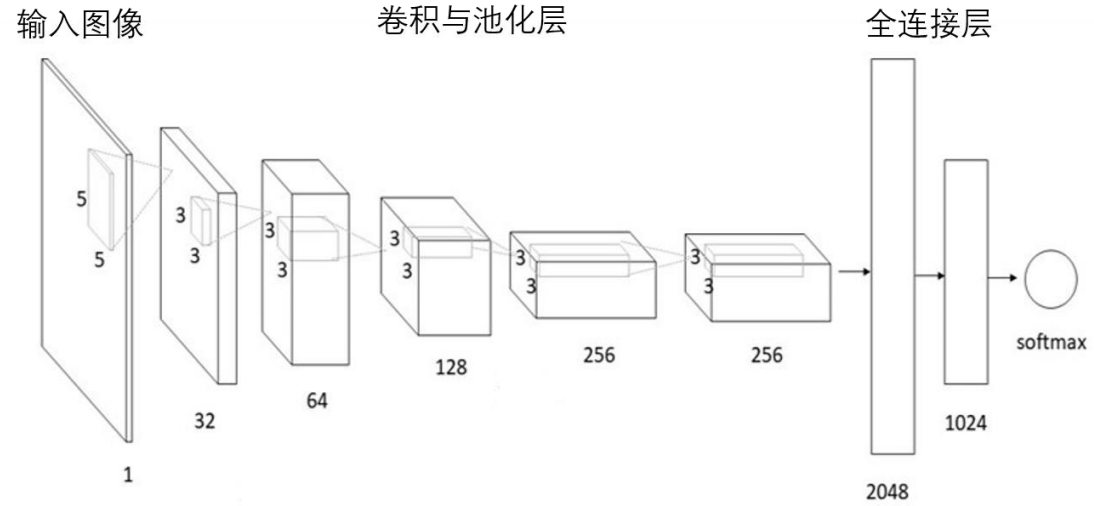


图 2 卷积神经网络结构示意图

图 3 是常见的卷积神经网络癫痫发作预测系统的框架图。系统的关键在于脑电原始数据的重

建，即输入图像的重建，现存的方法中依据卷积运算的维度重建为一维、二维以及三维的脑电图。近几年使用 CNN 的癫痫发作预测研究，分别使用了不同维度的卷积特征搭建模型[9-13]。卷积神经网络在癫痫发作预测往往使用重构图像处理，伴随着脑电数据维度改变，可能会造成重要信息的流失。池化层的存在也会导致许多非常有价值的信息的丢失，同时也会忽略掉整体与部分之间的关联因此，基于卷积的模型研究还有待进步。

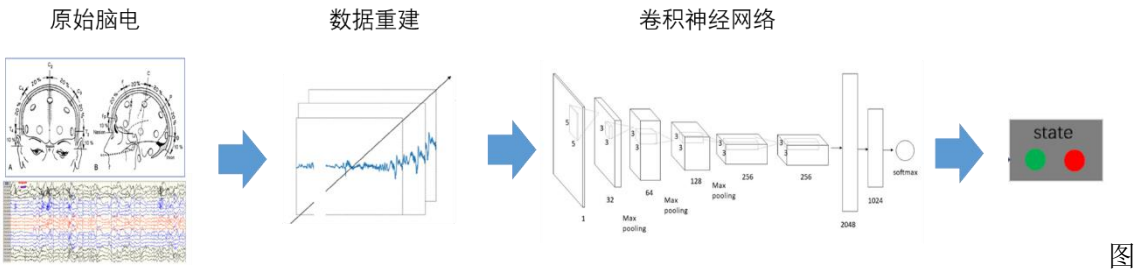


图 3 使用卷积神经网络的癫痫发作预测系统框架图

2.2 循环神经网络预测模型

循环神经网络（RNN）在自然语言处理等时间序列分析领域已经取得巨大成功及广泛应用[14]。循环神经网络包括输入层，隐藏层与输出层，图 4 为一个完整的循环神经网络结构。其中，x 表示输入的数据；s 表示样本在时刻 t 处的记忆，即隐藏层；o 表示输出样本，U 和 V 表示输入和输出的样本权重。不同于传统的前向反馈神经网络，RNN 引入了定向循环，如图 4 的右侧部分，t 时刻的输出除了与当前时刻有关外，还与前一时刻相关[15-17]。基于 RNN 能够处理输入之间前后有关联的问题，特别是 EEG 这种时间序列的分析优势及其明显。

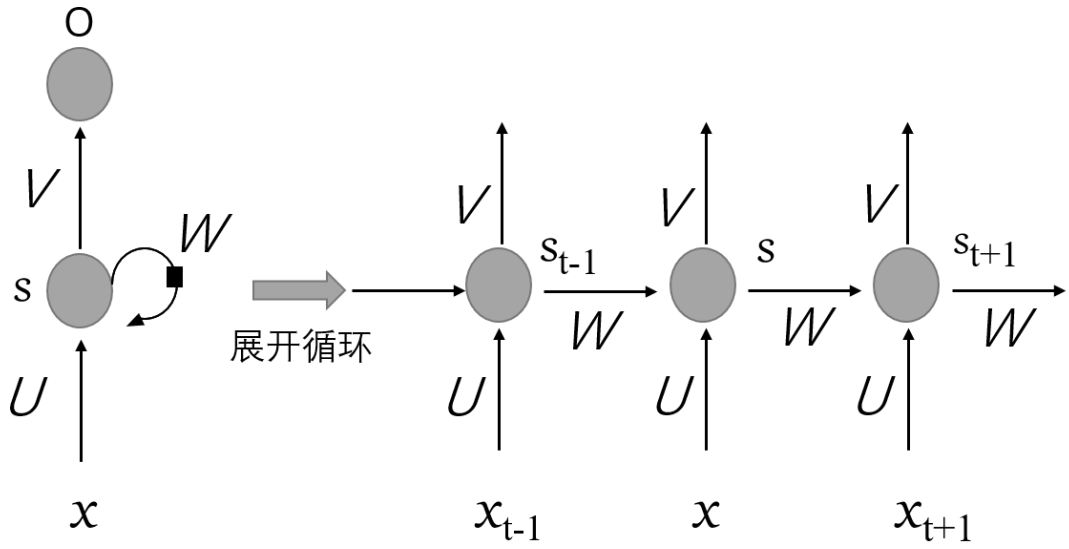


图 4 循环神经网络结构示意图

不同于卷积神经网络的 EEG 分析，原始脑电数据无需重建，RNN 可以直接处理 EEG 序列，这样保证了 EEG 最大程度的信息保留。图 5 所示为一个循环神经网络的癫痫发作预测系统框图。CNN 优势明显，但是仍然存在缺点，RNN 情况类似。因为每个循环层网络的参数共享，RNN 随着层的深度增加，会出现梯度爆炸或者梯度消失的问题。为了解决这些问题的存在，一些变体 RNN，比如 GRU、LSTM、biLSTM 等应用在癫痫发作预测研究上[5, 18-22]。

2.3 循环卷积神经网络预测模型

前面我们列举了 CNN 和 RNN 在癫痫发作预测方面的研究，CNN 具有更好的空间信息捕捉能力，而 RNN 则更擅长分析时间序列的关系。我们知道临床采集到的 EEG，不单单只是时间的累计结果，不同导联之间的相互作用以及脑网络节点之间的空间关系同样重要。考虑到 CNN 和 RNN 的优势不同，一种新的架构思想应用在 EEG 的分析中[6]。图 6 所示为一个融合了 CNN 和 RNN 的循环卷积神经网络癫痫发作预测系统，基本思想为在固定大小的时间滑动窗的移动下，CNN 网络学习序列间的空间特征并提取出来，按照时间顺序输入到循环神经网络中，经由 RNN 学习序列间的时间关系，最终给出分类结果。

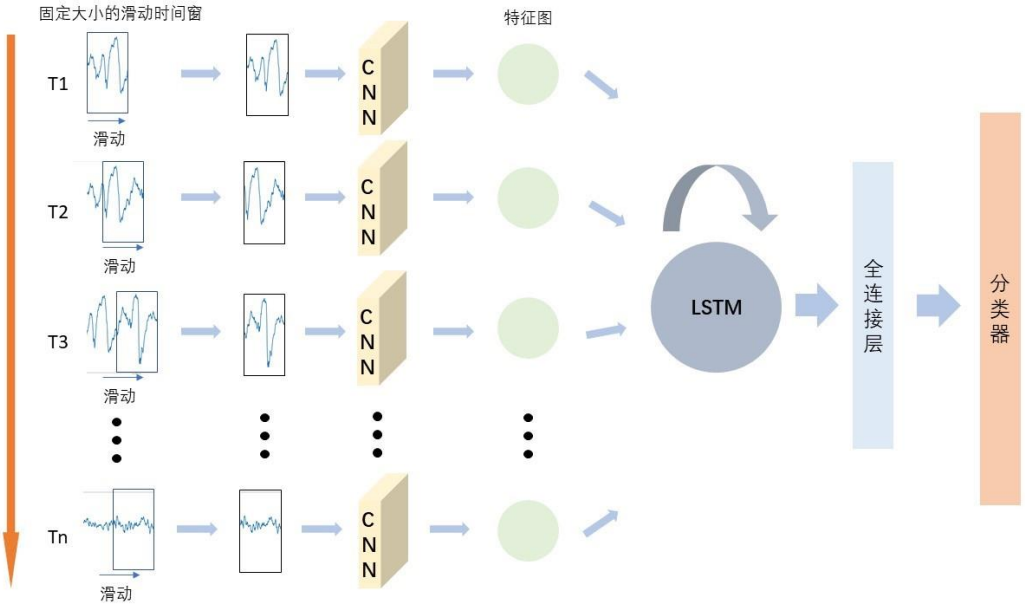


图 6 使用循环卷积网络的癫痫发作预测框架图

基于这种新的架构的研究开展的还不充足，David 等人[23]提出了一种基于卷积和递归神经网络端到端的简单架构，从 28 例患者的 119 例癫痫发作中创建的运动捕捉库中提取时空表示。随后，将测试表示形式与来自于测试数据集分离的五个异常癫痫发作的库之间的余弦相似度距离用于识别具有不符合已知行为的异常模式的癫痫发作测试。基于临床影像资料提取的运动特征用于癫痫的发作检测为癫痫疾病的诊断提供了更多的可能，虽然这种架构没有使用脑电数据，但是为多模态的研究，为基于脑电和影像的癫痫发作预测提供可能。Meysam 等[15]基于临床检测的脑电提出了一种卷积和长短期记忆循环神经网络的癫痫发作检测模型，这种模型可以将重构成二维图片的多

通道脑电数据卷积特征输入到 LSTM 中学习序列之间成的关系，最终输出分类结果。Xiaoyan Wei 等[24]提出了类似的研究，不同的是图像的重建维度，在文章中，该团队将脑电图时间序列转换为二维脑电图，进行多通道融合成三维结构。提出了一种可行的方法——长期递归卷积网络（LRCN），实现了癫痫发作的端到端自动预测模型。卷积网络块用于从数据中自动提取深度特征。结合 LSTM 神经网络块来从流数据中区分不同的图像序列和识别前置节段。用独立数据对模型进行了检验，提供了比以往研究中手工设计的方法和单个深层神经网络更高的癫痫发作预测灵敏度和 0.04FP/h 的低错误预测率。这种在外部结构的重构方式使我们研究的一个重点，但是一些团队考虑在神经网络的内部实现两种网络的结合，李飞飞团队[25]提出了一种具有强记忆力的 E3D-LSTM 网络，用 3D 卷积代替 2D 卷积作为 LSTM 网络的基础计算操作，并加入自注意力机制，使网络能同时兼顾长时和短时信息依赖以及局部时空特征抽取。这为视频预测、动作分类等相关问题提供了新思路，是一项非常具有启发性的工作。

3 数据

3.1 公开数据集

本次实验使用的公开数据集是来自伯恩大学脑电数据集[26]。该数据集已经被预处理成为理想条件下的脑电，被广泛的使用。该数据库包含五个分组，每组 100 个时间序列，每个时间序列长度为 23.6 秒。A，B 组来自健康志愿者静息状态的脑电信号，C、D 和 E 组都来自于患者脑电，区别在于 E 组记录与患者癫痫发作时间，而另外两组在整个记录期间无癫痫发作。

3.2 私人数据集

本次实验的私人数据集来自于新疆医科大学第一附属医院神经内科脑电图室 2013 年至 2016 年间 15 例癫痫患者(男 5 例，女 10 例，年龄 6-51 岁)。头皮电极按照国际 10-20 系统放置。采样频率为 500 Hz。可用的脑电图记录的总持续时间大约为 540h，发作开始和偏移时间间隔由临床专家在目视检查后手动注释，共计 168 次发作。按照实验设计要求与分期定义，将脑电分类为发作间期，发作前期与发作后期。EEG 数据集的信息表示在表 1 中。

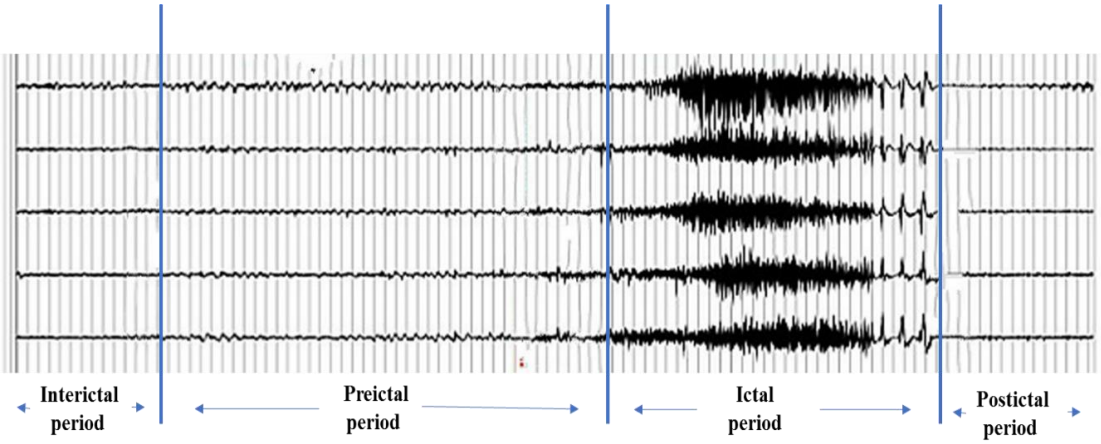
Table 1 Characteristics of the collected data

ID	Sex	Age	Type	Time	Number of Seizures	IT
1	F	36	SPS	48	10	654 s
2	F	22	SPS, CPS	48	12	274 s
3	F	36	CPS	48	14	1386 s

4	F	40	SPS	24	6	302 s
5	M	6	SPS	48	21	453 s
6	F	16	SPS, CPS	24	7	329 s
7	F	16	SPS, CPS	24	8	254 s
8	F	28	CPS	24	5	400 s
9	F	31	SPS	24	9	423 s
10	M	51	SPS	72	30	1064 s
11	M	20	SPS, CPS	48	19	4072 s
12	M	46	SPS	24	6	208 s
13	F	15	CPS	48	8	137 s
14	女	28	SPS	24	5	824 s
15	男	39	SPS, CPS	24	4	895 s

3.3 分期定义

癫痫发作过程分为四个状态，包括发作状态、发作前状态、发作后状态和发作间状态[27-29]，如图 1 所示。临床专家已经对资料中癫痫发作的起始点和终止点进行了标记。为了预测未来的癫痫发作，癫痫发作预测系统的关键是将发作前期与发作间期分开。



发作前期(Preictal)是被定义为在 ictal（或癫痫发作）周期之前的数据段，该段可以从 iEEG 信号中清楚地识别。然而，不同研究对 preictal 的长度有不同的定义，从 10 分钟到 60 分钟不等，

没有多少或任何理由[1, 30]。此外，在这项研究中，每个患者数据包含至少两次癫痫发作，但两次发作之间的间隔可以是长或短。通过不同学者的工作，现在最好的发作前期定义发作前 30 分钟。对于距离前一次发作不到 30 分钟的癫痫发作，我们认为它们只是一次发作，并且使用先导性发作的发作作为联合发作的发作。

发作间期定义为信号中既不是发作期也不是发作前期的部分。在我们的实验中，通过文献调查，定义发作后至少 1 h 和离发作至少 40min 的数据作为发作间期。本文将每次发作前 35 min 中 30 min 的数据作为发作前期数据，为了保持数据样本集中正负样本的平衡,

4 方法

4.1 系统设计框架

本研究将残差网络架构与循环神经网络相结合，并使用 1d 卷积神经网络替代残差架构中的二维卷积操作，建立癫痫发作预测的时空深层学习模型。利用 1DCNN 从图像集中自动提取信号特征，对不同阶段进行分类。此外，使用 indrnn 神经网络来区分不同类别，然后提供癫痫发作预测。输出结果显示分类函数是否认为当前脑电图片段处于前置阶段。在此过程中，indrnn 网络从卷积步骤学习多帧图像的抽象特征，然后识别癫痫脑电。本文的癫痫脑电自动检测模型的构建过程如图 3 所示。该模型运行在 Python 3 的高性能计算机上。

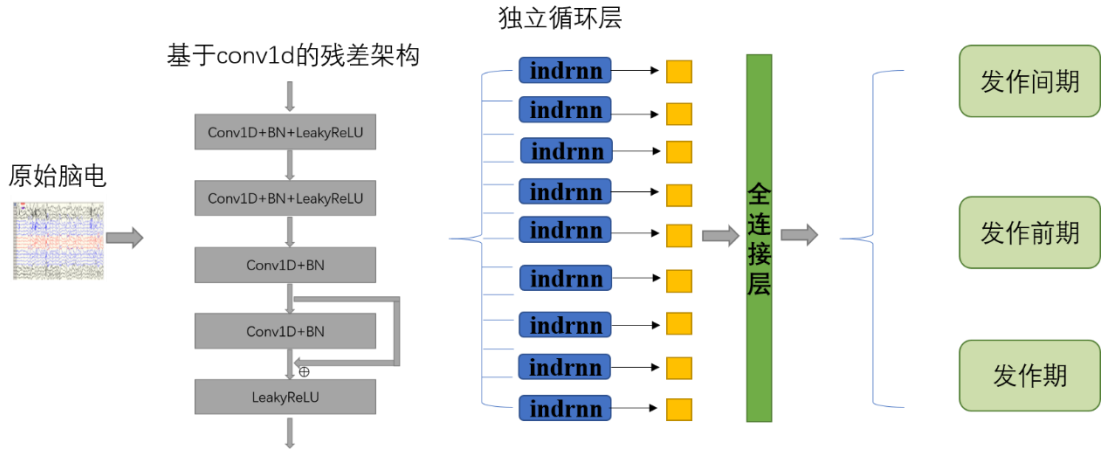


图 5 本实验的癫痫脑电识别框架图

在本研究中，将 indrnn 网络与前面提出的卷积神经网络进行结合，创建一个时空的深度学习模型用于癫痫发作预测。传统的卷积神经网络将输入的数据经过卷积和池化处理再输入到全连接层后输出结果。本模型的输入层先连接着完整的卷积神经网络，脑电信号作为图像输入到卷积神经网络中，输出一个经过变换的固定长度的特征向量，多次连续输出得到一个特征序列，随后输入到 indrnn 中，indrnn 将输入映射到隐含层中，并更新隐含层，最后输出 t 时间步的预测分布结果，使用 softmax 确定类别。在这个过程中，独立循环人工神经网络（indrnn）中学习多帧图像的抽象特征的变化规律，输出的结果经由分类函数确定当前 EEG 片段是否处于发作前期，进而做出发作预警。

4.2 独立残差网络架构

Layer	Hidden Layer	Related parameters (filters, kernels, stride)		
BLOCK1	Conv1D+BN+LeakyReLU	64	8	1
	Conv1D+BN+LeakyReLU	64	5	2
	Conv1D+BN	64	3	1
	Conv1D+BN	64	1	1
	Add			
	LeakyReLU			
BLOCK2	Conv1D+BN+LeakyReLU	128	8	1
	Conv1D+BN+LeakyReLU	128	5	2
	Conv1D+BN	128	3	1
	Conv1D+BN	128	1	1
	Add			
	LeakyReLU			
BLOCK3	Conv1D+BN+LeakyReLU	64	8	1
	Conv1D+BN+LeakyReLU	64	5	2
	Conv1D+BN	64	3	1
	Add			
	LeakyReLU			
	GlobalAveragePooling1D		2	None
	indRNN+BN	128		
	indRNN+BN	128		
	Fully connected	256		
	Softmax	n_class		

4.2.1 残差网络

CNN 能够提取 low/mid/high-level 的特征，网络的层数越多，意味着能够提取到不同 level 的特征越丰富。并且，越深的网络提取的特征越抽象，越具有语义信息。如果深层网络的后面那些层是恒等映射，那么模型就退化为一个浅层网络。那现在要解决的就是学习恒等映射函数了[31]。残差网络经过这些年的发展，对于深层次的网络学习具有很大的优势。本实验使用的残差架构是一

个三残差块的卷积网络，如图所示为第一个残差块的结构图，第二个残差块的拼接与第一个相同，第三个如图右所示，略有不同。

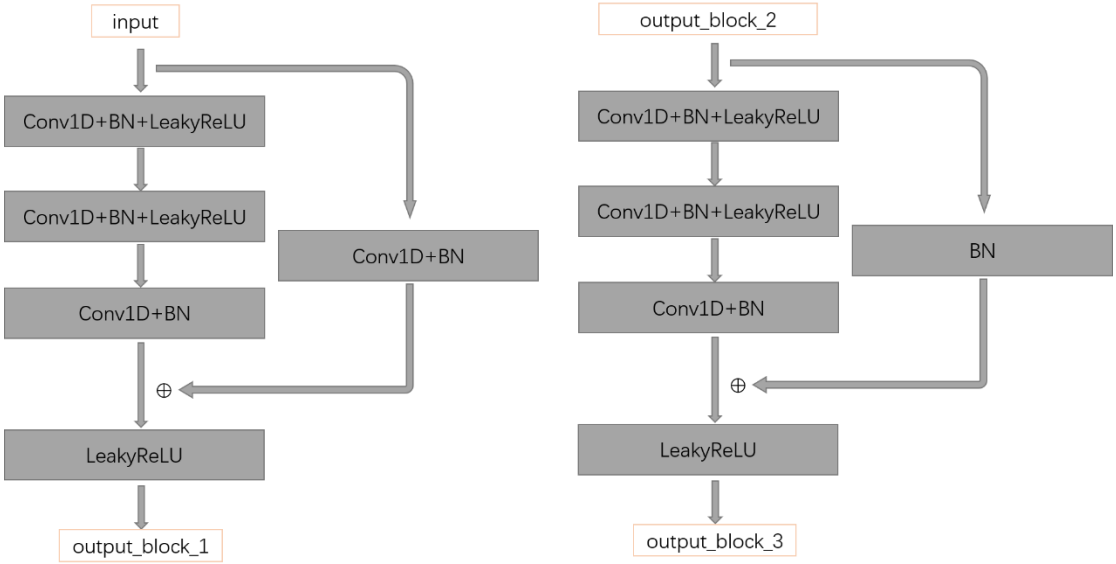
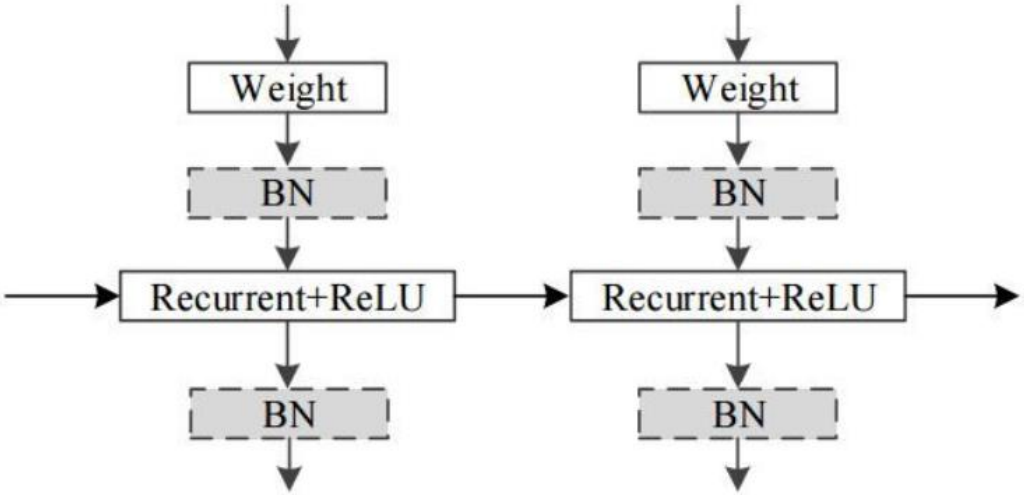


图 6 残差网络图

4.2.2 独立循环神经网络

循环神经网络（RNN）已在动作识别 、场景标注 、语言处理 等序列学习问题中获得广泛应用，并且成果显著。与卷积神经网络（ CNN ）等前馈网络相比，RNN 具有循环连接，其中最后的隐藏状态是到下一状态的输入。由于循环权重矩阵不断相乘，RNN 的训练面临着梯度消失和梯度爆炸的问题。长短期记忆（ LSTM ）和门控循环单元（ GRU ）等若干 RNN 模型可用来解决这些梯度问题。然而，在这些变体中使用双曲正切和 Sigmoid 函数作为激活函数会导致网络层的梯度衰减。因此，构建和训练基于 RNN 的深度 LSTM 或 GRU 其实存在困难。

IndRNN 中的每个神经元独立地处理一种类型的时空模型[32]。传统上，RNN 被视为时间上的、共享参数的多层感知器。与传统的 RNN 不同的是，本文提出的 IndRNN 神经网络为循环神经网络提供了一个新视角，即随着时间的推移（即通过 u）独立地聚集空间模式（即通过 w ）。不同神经元之间的相关性可以通过两层或多层的堆叠来加以利用。在这种情况下，下一层的每个神经元处理上一层所有神经元的输出。如图所示。



4.2.3 网络配置

我们的识别网络是通过使用“SGD”作为优化函数进行全局优化，“categorical_crossentropy”作为损失函数来进行训练的。随机梯度下降（Stochastic Gradient Descent, SGD）由于学习速率快并且可以在线更新，常被用于训练各种机器学习和深度学习模型。残差网络中使用 1D 卷积网络进行替换，目的是可以直接处理最原始脑电数据，而无需进行脑电增加维度的预处理。卷积单元部分的参数选择为卷积核长度为（8,5,3），卷积核数量为（64,128,128），激活函数选择为“LeakyReLU”，卷积部分主要是为了捕捉脑电数据短期时间相关性；循环单元部分网络主要是为了区分脑电类别，随着网络深度的增加循环网络存在的梯度消失与梯度爆炸问题愈加严重，此次实验选择独立循环神经网络，神经元数量设置为 128，最后使用“dense”输出分类结果，激活函数选择为“softmax”。批处理大小设置为 64，迭代次数设置为 512。通过使用公开数据集与私人数据集的对比，充分考虑了实验中的各种噪声水平。

5 实验与结果

这一节我们主要是在理想条件的公开数据集与来自于临床实验的真实数据集测试我们提出的癫痫检测性能，评估指标使用灵敏度（Sens），特异性（Spec）和分类准确性（Acc），详细的会在下面段落进行详细说明。

5.1 实验设计

理想条件下的公开数据集与临床实验数据集的区别在伪影与噪声，这些都会对检测的结果产生影响。我们首先在公开数据集上进行测试，然后再真实数据集上测试已进行对比。

5.1.1 模型的训练与测试

癫痫脑电分类研究是对于 EEG 信号片段的正确识别，识别出已经定义的癫痫不同时期。图 展示了我们实验模型的训练与测试。本实验所使用的的公开数据集共包含了三种不同的脑电，分别是健康患者组，癫痫患者癫痫未发作组以及癫痫患者癫痫发作组。因此我们的实验按照对照原则设计出了二分类和三分类任务来检测算法的性能。二分类任务实验设计是健康患者组与癫痫患者组的脑电识别；三分类任务实验设计是健康患者组与癫痫未发作组以及癫痫发作组。除了公开数据集，我们的实验考虑到了真实世界中的噪声与干扰等问题，选择了特定类型的癫痫患者的发作间期、发展前期与发作期脑电进行测试。真实数据具有流数据的特性，因此在我们的实验中，我们将其形式化为窗口化的流数据。对于每一个 10 秒的脑电片段进行识别，输出脑电类别。因为本次实验不涉及预测的任务，因此我们的脑电分类识别只取了时间窗口，而抛弃了事件分类任务。

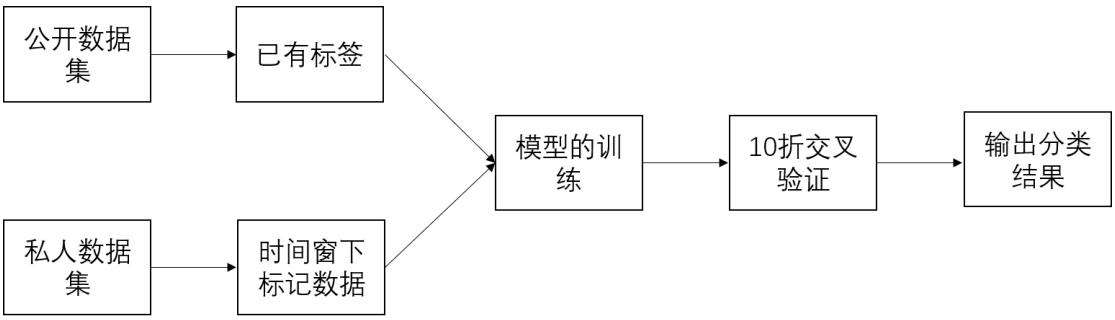


图 实验模型的训练与测试

在这项研究中使用了十倍交叉验证方法。 首先，脑电信号被随机分为十等份。 EEG 信号的十份中的九份用于训练本文中的神经网络，而其余十分之一的 EEG 信号用于测试系统的性能。 通过移动测试和训练数据集，此策略重复十次。 本文中报道的准确性，敏感性和特异性值是从十次评估中获得的平均值。

5.1.2 评估指标

评价癫痫检测方法的性能，有三个参数最重要也最常用，分别是敏感性（Sen），特异性（Spe）和总识别率（Acc）。Sen 表示识别系统对于发作前期脑电数据的敏感程度，即衡量分类器对发作前期的捕捉能力,Spe 表示识别系统正确识别发作前期脑电数据为发作前期的特异性能力，即衡量分类器对非发作前期的识别能力，Acc 表示识别系统的识别能力，具体公式表示如下：

$$\mathrm{Se} = \frac{TP}{TP + FN} \tag{3-1}$$

$$\mathrm{Sp} = \frac{TN}{TN + FP} \tag{3-2}$$

$$\mathrm{Acc} = \frac{TP + TN}{TP + FN + TN + FP} \quad (3\text{-}3)$$

其中，TP 表示被正确划分为发作前期的个数，FN 表示脑电的发作期被划分为发作前期的个数，TN 表示被正确划分为发作期的个数，FP 表示被错误的划分为发作期的脑电的个数。

5.2 理想条件下检测癫痫

我们首先检查提出的深度学习方法，以清除没有伪像和噪声的 EEG 信号。 在进行脑电图预处理（即数据分割和整形）之后，脑电图将被馈入我们的深度神经网络模型，其最终目标是有效的脑电图特征学习和分类。

5.2.1 二分类结果

癫痫脑电的二分类任务中，第一类是区分健康患者的正常脑电图与癫痫患者的癫痫脑电图，在我们的公开数据集中属于数据集 A、B 与数据集 C、D、E 之间的分类。但是有一些学者认为癫痫患者的非发作期的发作间期等脑电在临床上亦应该被视为正常脑电，在我们的实验中，我们主要是识别非癫痫发作活动脑电（数据集 A、B、C、D）与癫痫发作活动脑电（数据集 E）。考虑到每个数据集有 100 个信号，因此我们的分类任务中存在类不平衡的问题。基于这些考虑，在我们的实验中，我们一方面考虑数据扩增，另一方面对于非癫痫活动脑电进行随机选择。此外在我们的评估指标中，也不再单纯的信赖特异性，灵敏度与准确度同样是我们的关注点。

表中数据表明，基于 1DCNN 和 indrnn 的癫痫脑电识别模型能够有效的进行病症识别。对于发作病症与未发作的识别实验结果，Acc 与 Spe 都取得了 100%，在基线算法中处于最优表现。

表 2 二分类结果

Method	Spe	Sen	Acc
LSTM	94.62	89.67	93.33
1DCNN	96.17	94.37	95.36
INDRNN	93.57	91.57	93.52
RESNET(1DCNN)	98.69	96.78	97.47
RCNN	100	97.50	100

5.2.2 三分类结果

非癫痫活动脑电中又可以细分为正常脑电与发作间期脑电，因此我们的实验又做了三分类的实验。结果如表所示，我们的算法在基线算法中处于最优表现。

表 3 三分类结果

Method	Spe	Sen	Acc
LSTM	89.58	90.42	91.26
1DCNN	94.87	89.43	93.82
INDRNN	92.68	90.67	91.53
RESNET(1DCNN)	94.28	92.79	95.47
RCNN	97.68	96.52	98.33

5.3 真实情况下检测癫痫

真实世界数据存在着噪声和干扰，例如肌肉活动、眼动干扰和环境噪声等。真实世界下的模型性能才能为临床带来真正的意义。考虑这些，我们的实验收集了临床环境下的 15 例患者的脑电数据，按照已有的癫痫分期定义分为发作间期，发作前期和发作后期。对照理想条件下的进行了二分类（发作活动与非发作活动）和三分类（三个时期的识别）实验。

5.3.1 二分类结果

癫痫发作即为病症表现，真实世界数据存在干扰，相应的结果也有降低，但仍然是在基线方法中处于最优。

表 4 二分类结果

Method	Spe	Sen	Acc
LSTM	84.79	83.24	85.64
1DCNN	89.58	84,89	88.73
INDRNN	85.58	85.63	83.41
RESNET(1DCNN)	89.79	88.76	90.57
RCNN	91.42	86.58	90.74

5.3.2 三分类结果

上文中已介绍癫痫的分期定义，为发作间期，发作前期与发作期。三分类实验中，我们的方法在基线方法中依旧处于最优。

表 5 三分类结果

Method	Spe	Sen	Acc
LSTM	85.54	82.38	84.47
1DCNN	86.39	85.35	87.39
INDRNN	83.56	86.73	84.65
RESNET(1DCNN)	89.93	87.48	91.83
RCNN	90.61	85.42	92.11

6 结论

我们的研究描述了一种可以直接处理原始脑电的癫痫自动检测新方法。我们的研究证明了深度学习在癫痫检测与发作预测方面的潜力，以及卷积神经网络与循环神经网络的结合的可能性。希望本研究能促进癫痫发作预测系统的进一步发展。

PU-009

基于赫斯特指数的循环神经网络癫痫脑电识别研究

马梦楠、张亮军、承琅林、韦晓燕、周毅
中山大学中山医学院

目的 癫痫的发作预测对于提高癫痫患者的生活质量以及生存率有着至关重要的作用，正确的捕捉发作前期是进行预警的关键。

方法 本研究结合非线性动力学与人工神经网络，提出一种基于赫斯特指数的发作前期捕方法。实验中使用 16 名患者的脑电数据，结合非线性动力学赫斯特指数与循环神经网络进行前期捕捉。

结果 实验结果表明，该方法能够有效的区分发作前期与发作期的脑电。

结论 本文旨在对后续的研究提供借鉴意义。

1 引言

癫痫发作是由大脑神经元的突然性同步放电引起，通常伴随着意志丧失和相关的非自愿异常运动。根据世界卫生组织（WHO）的统计，全球约有近 5000 万的人口患有癫痫疾病，且全球每年约有 240 万的新生病例产生[1]。

癫痫疾病的不可控性是一大难题，虽然抗癫痫药物可以在许多癫痫患者中获得完全的癫痫发作控制，但多达三分之一的癫痫患者会患有医学难治性癫痫症（发作不可控），药物虽然可以减少癫痫发作的频率，但不能完全控制癫痫发作[2]。由于癫痫发作的不可控性，这些患者因无法操作机动车辆等一些现实工具而在某些方面无法获得有意义的就业机会。此外还有强有力的证据表明，患有长期不可控的癫痫的病人患有永久性记忆障碍，抑郁，焦虑，自杀和其他精神疾病的风

险相比于其他人有所增加[3, 4]。上述因素对癫痫患者的生活质量产生负面影响，不仅影响家庭成员和看护人的生活，并且对家庭成员的医疗保健成本构成经济负担。

近几年的研究中，基于癫痫脑电图（EEG）的癫痫疾病的诊断和癫痫的发作检测等方面已取得重大进展。非线性动力学、机器学习、深度学习等技术也已应用到这一块的研究中去，且取得了不错的研究结果，可以高度准确的帮助癫痫的诊断。Mark J Cook 团队在 13 年发表在柳叶刀上的一篇杂志，首次应用在人体身上的癫痫自动预测系统能够有效的预警癫痫的发作，为这一方面的研究打下坚定的基础[5]。准确的癫痫发作预测不仅提前预测患者可能发生癫痫发作的时间，还将在临床癫痫治疗上提供辅助，在设定预警的时间范围内给予一定的药物治疗等。癫痫发作预测显著改善患者的生活质量并减轻护理人员的负担，具有十分重要的现实意义。

本研究使用赫斯特指数并在数据处理中辅助其他非线行动力学指标，结合循环神经网络（RNN）研究癫痫发作前期的捕捉。框架流程如图 1 所示。经过降噪等预处理后的多导脑电数据，计算每个导联的赫斯特指数，结合其他非线性特征指标，基于 Pearson 相关系数进行导联的优化与筛选，降低无关导联的影响，最后使用双层的 LSTM 分类器进行分类，输出发作前期捕捉结果。

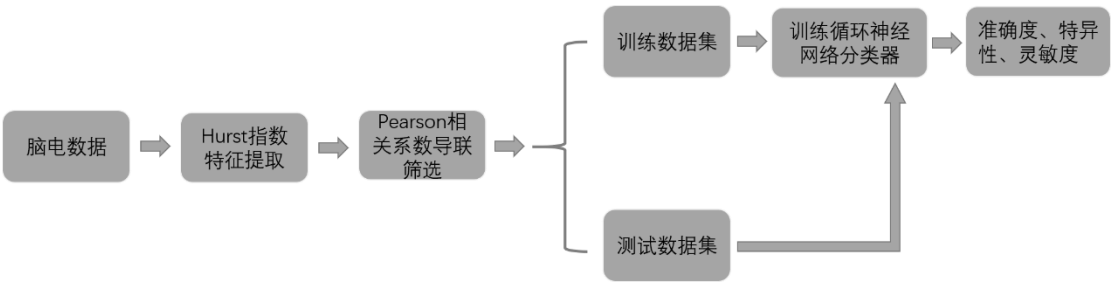


图 1 研究总体框架

2 对象与方法

2.1 实验数据

数据收集自 2013 - 2016 年新疆某医院神经内科脑电图室 16 例癫痫患者(男 5 例，女 10 例，年龄 6-51 岁)。头皮电极按照国际 10-20 系统放置。采用双极导联法同步记录 22 导联脑电图信号，采样频率为 500hz。每位患者脑电的连续记录时间为 24-72 h，可获得的脑电图总记录时间约为 540 h，发作时间和偏移时间间隔由临床专家目视检查后手工标注;共有 164 次发作缉获，平均每人 11 次。住院时间为 11675s(约 194 分钟)。

2.2 分期定义

癫痫发作的过程通常有四个时期，发作间期（正常期），发作前期，发作期和发作后期[6]。其中。通过识别发作前期能够提前预测癫痫发作以便提供临床治疗，是发作预测至关重要的一个环节。对于发作前期的定义，不同的研究有不同的时长定义，没有统一的标准，现在一般采用发作前 30min 定义为发作前期。

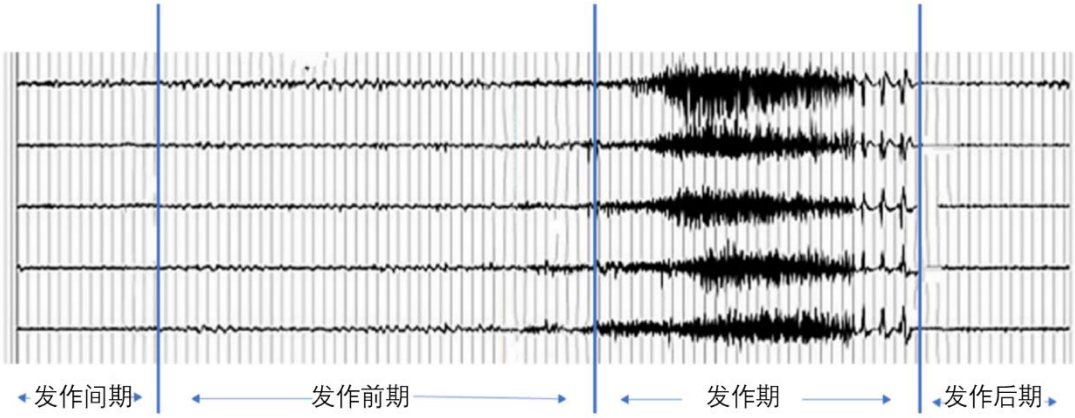


图 1 癫痫脑电状态的不同分期

2.3 赫斯特指数

赫斯特指数(Hurst Exponent)是用来衡量时间序列是否有长期记忆的一个指标[7]。赫斯特指数的值定义为 0 和 1 之间，值接近 0.5，说明被观察的序列是随机的和不具有相关性的；值越接近 1，表明序列越具有时间相关性和趋势性。癫痫脑电作为一个多维的时间序列，通过计算赫斯特指数高于 0.5 也从侧面证明了癫痫脑电的时间相关性，为癫痫发作的预测提供理论依据。

墨尔本大学的研究团队使用 Hurst 指数对 120 名癫痫患者脑电信号进行了长期观察，发现患者的 Hurst 指数在发作前期有一个陡峭的下降点，接着在发作期有若干个下降点，整体趋势是一个不断下降的过程，然后在发作后期则有一个上升点，之后保持不变[8]。如图 2 所示，再次癫痫发作的预测提供理论依据。

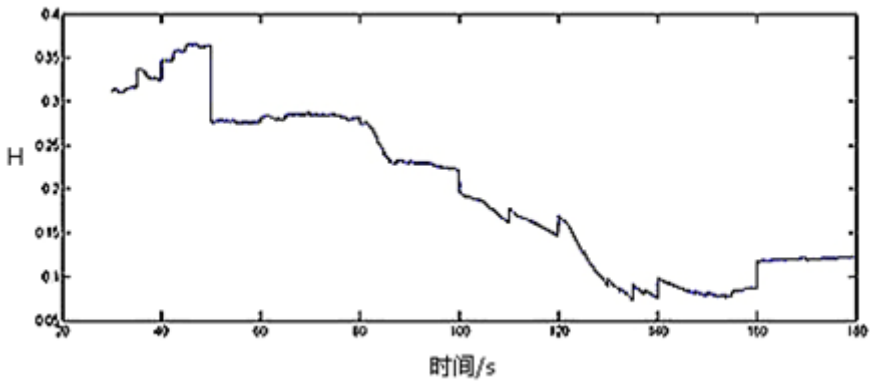


图 2 癫痫患者脑电赫斯特指数变化图

2.4 Pearson 相关系数

Pearson 相关系数（Pearson correlation coefficient，PCCs），又名积差相关系数，它是 20 世纪由英国统计学家 Pearson 提出的线性相关计算方法[9]。根据癫痫放电的原理，对于部分性癫痫研究来说，其致痫灶附近的电极变化较为明显，但同时也有研究表明致痫灶远离区域也有明显的动力学变化，大脑作为一个混沌系统，考虑全部的电极信息是非常重要的，但同时无关的导联信息

会影响信号的准确度，会附带冗余的信息使得模型性能降低，增加计算时间。因此考虑并在此基础上进行优化是进行癫痫预测分类的准确性保障。因此本研究在此基础上，纳入全局导联信息，考虑导联筛选。其具体算法如下：

假设双变量X和Y，该双变量间的 Pearson 相关系数可有：

$$\rho_{X,Y}=\frac{cov(X,Y)}{\sigma_X\sigma_Y}=\frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X\sigma_Y}=\frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \quad (2\text{-}1)$$

或者是：

$$\rho_{X,Y}=\frac{\sum XY-\frac{\sum X\sum Y}{N}}{\sqrt{(\sum X^2-\frac{(\sum X)^2}{N})(\sum Y^2-\frac{(\sum Y)^2}{N})}} \quad (2\text{-}2)$$

其中N为变量的数量，E为数学期望，cov为协方差。但是 Pearson 系数计算有一定的应用条件，即随机变量，且变量呈现正态分布，散点图呈现线性趋势，变量相互独立。但是事实上很多时候现实中的数据很难满足这个要求，一般常用其 Spearman 秩相关系数来计算。

2.5 循环神经网络

循环神经网络（RNN）在自然语言处理等时间序列分析领域已经取得巨大成功及广泛应用[10]。循环神经网络包括输入层，隐藏层与输出层，图 3 为一个完整的循环神经网络结构。其中，x 表示输入的数据；s 表示样本在时刻 t 处的记忆，即隐藏层；o 表示输出样本，U 和 V 表示输入和输出的样本权重。不同于传统的前向反馈神经网络，RNN 引入了定向循环，如图 4 的右侧部分，t 时刻的输出除了与当前时刻有关外，还与前一时刻相关。基于 RNN 能够处理输入之间前后有关联的问题，特别是 EEG 这种时间序列的分析优势及其明显。

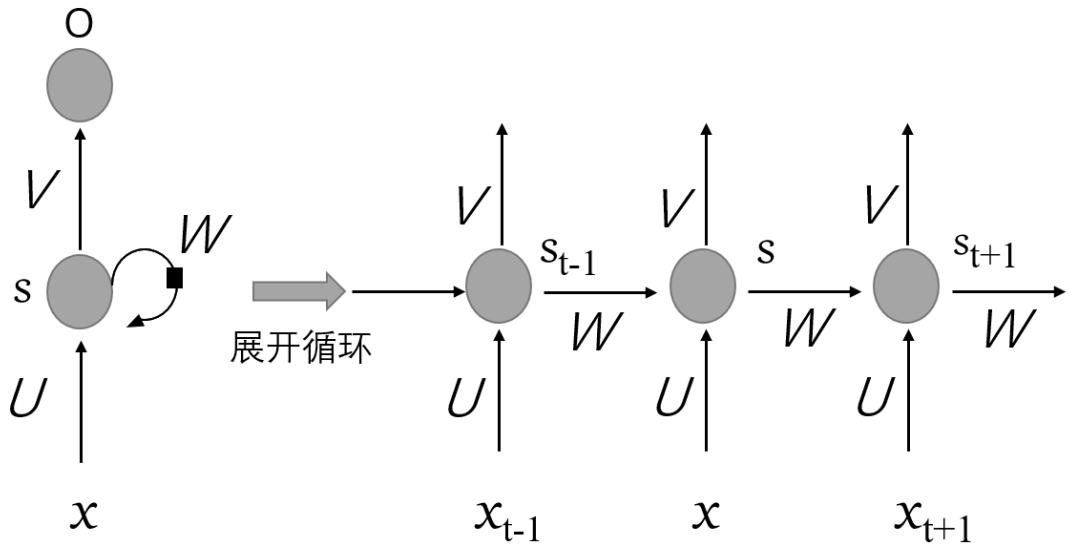


图 3 循环神经网络结构示意图

LSTM 又被称为长短记忆循环神经网络，是 RNN 的一种特殊形式，具有学习长时程依赖的能力。LSTM 独特的门控机制，能够有效的解决 RNN 中普遍存在的梯度爆炸和梯度消失问题[11]，考虑到

EEG 信号也是一种时间序列，本文首先将经过预处理和导联优化之后数据的赫斯特指数作为特征输入到 LSTM 中学习序列之间的关系，如图 4，训练模型的结构主要是两层的神经网络，第一层主要是由 LSTM 结构组成的 RNN 层，输入特征维数为（5000，1），步长选择为 30，隐藏层的节点为 128，第二层是全连接层结构组成的分类层，分类函数选择使用“softmax”。采用 Adam 算法进行参数优化，学习率为 0. 001，mini _batch 方法进行网络训练，batch 大小为 100。网络模型由 keras 框架实现。

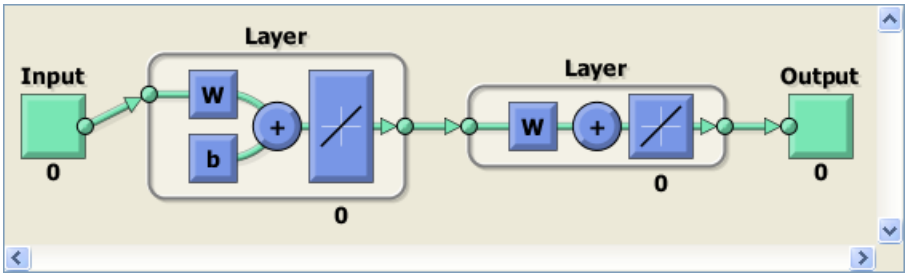


图 4 训练模型结构图

3 结果分析

3.1 评价指标

评价癫痫预测方法的性能，有三个参数最重要也最常用，分别是敏感性（Sen），特异性（Spe）和总识别率（Acc）。Sen 表示识别系统对于发作前期脑电数据的敏感程度，即衡量分类器对发作前期的捕捉能力,Spe 表示识别系统正确识别发作前期脑电数据为发作前期的特异性能力，即衡量分类器对非发作前期的识别能力，Acc 表示识别系统的识别能力，具体公式表示如下：

$$\mathrm{Se} = \frac{TP}{TP + FN} \tag{3-1}$$

$$\mathrm{Sp} = \frac{TN}{TN + FP} \tag{3-2}$$

$$\mathrm{Acc} = \frac{TP + TN}{TP + FN + TN + FP} \tag{3-3}$$

其中，TP 表示被正确划分为发作前期的个数，FN 表示脑电的发作期被划分为发作前期的个数，TN 表示被正确划分为发作期的个数，FP 表示被错误的划分为发作期的脑电的个数。

3.2 导联优化选择结果

表 1 各导联相关性的分析

指标	相关性（大于 0.7）
Hurst 指数	F3, T4, C3, C4, Afz, F4, T6, T3, T5, P3, F7, P4, Fz, Cz, M2,

	Fp1, O2
近似熵	T5, T4, T6, P3,P4, C4, Afz, C3, F4, T3, Fz, F8, F3, Pz, F7, O1, M2, M1
样本熵	F3, Afz, T5, T6, T4, P3, T3, F8, F7, Fz, Fp1, C4, Fp2, M1
最大 Lyapunov 指数	T3, F3, AFz, F3, Fz, Cz, Pz, T3, P3, P4, F2, C3, F3, T5, C2, C4, Cz, AFz, F3

预实验中发现，单纯只是赫斯特指数进行相关性的分析造成的实验误差过大，因此本研究结合了多个非线性指标对导联进行优化选择。由表 1 可知，其 Hurst 指数，近似熵，样本熵以及最大 Lyapunov 指数刻画的导联相关性具有相似性，即使使用不同的指标进行相关性分析，其导联相关性不会因此改变。此外，23 号患者在左额（Fp1，F3），左中央（C3），左颞下颌切迹（M1），左前颞（F7），左中颞（T3），左后颞（T5）导联上的相关性较强。对比病人病历发现，属于颞叶癫痫病人。其在这些导联附近棘波尖波等异常波连续出现。因此确定经过导联筛选后其整理后的导联范围为 10 个，即 Fp1，F3，C3，M1，F7，T3，T5，T6，P3，C4。

3.2 癫痫脑电发作前期的捕捉

本研究基线方法选择为支持向量机(SVM)，如表 2 所示，相比于传统的机器学习方法，非线性动力学指标与机器学习的结合有效的提升了结果，而神经网络的使用对于结果的提升也有明显的效果。

表 2 结果对比

算法	灵敏度 (Sen)	特异性 (Spe)	准确度 (Acc)
SVM	96.2	96.8	95.4
LSTM	96.7	97.9	97.2
Hurst+SVM	96.7	97.2	96.3
Hurst +LSTM	97.3	99.4	98.6

如表 2 所示，输入数据为癫痫患者发作前期和发作期的脑电数据片段，得到的发作前期识别灵敏度为 97.3%，特异性为 99.4%，准确度为 98.6%，均为基线方法的最优。表 2 中设置有不同的对比结果，算法仅使用 SVM 和 LSTM 的实验对比中可以发现深度学习的优势；而结合非线性动力学特征则有效的提升了结果。

4 结论与讨论

癫痫的预测研究能够有效的改善和提高患者的生活质量和精神面貌。本文主要以癫痫脑电发作前期的捕捉为重点，结合非线性动力学，将经过导联优化选择后的数据，经过赫斯特指数提取之后作为输入输送到循环神经网络中，输出发作前期的识别率。对比与基线方法，本研究取得

的结果有了一个很大的提升，对于后续的研究提供借鉴意义。未来发展中，深度学习将是研究的重心，但是如何转化其社会价值，改善患者生活仍将是我们研究的最终目标。

PU-010

基于深度学习的癫痫患者脑电图分类与预测任务研究

马梦楠、张亮军、承垠林、韦晓燕、周毅
中山大学中山医学院

目的 近些年来，基于癫痫脑电图(electroencephalogram，EEG)的癫痫发作自动检测和诊断方面取得了重大的进展，许多创新算法已经应用于癫痫的辅助诊断和治疗中。随着研究的深入，癫痫自动检测的重点转向癫痫的发作预测。如果可以在癫痫发作之前预测癫痫的发作，那么在临床上将产生深远的意义，治疗给药或者通过其他的治疗途径进行缓解。

方法 本文针对癫痫发作预测研究模型类别进行研究，介绍了卷积神经网络和循环神经网络在癫痫发作检测与预测的应用。此外一种新的融合了卷积与循环网络的融合架构应用在癫痫预测的研究。

结果 通过回顾性的研究，本文希望能够为癫痫的诊断和发作预测技术开发的调整和扩展带来新的思路。本文阐述了机器学习和深度学习在癫痫自动发作预测方面的应用，重点介绍了最近关于癫痫发作预测的研究结果。

结论 通过回顾性的研究，本文希望能够为癫痫诊断和发作预测技术开发的调整和扩展带来新的思路。

1、引言

癫痫发作是由大脑神经元的突然性同步放电引起，通常伴随着意志丧失和相关的非自愿异常运动。根据世界卫生组织（WHO）的统计，全球约有近 5000 万的人口患有癫痫疾病，且全球每年约有 240 万的新生病例产生[1]。

癫痫疾病的不可控性是一大难题，虽然抗癫痫药物可以在许多癫痫患者中获得完全的癫痫发作控制，但多达三分之一的癫痫患者会患有医学难治性癫痫症（发作不可控），药物虽然可以减少癫痫发作的频率，但不能完全控制癫痫发作[2]。由于癫痫发作的不可控性，这些患者因无法操作机动车辆等一些现实工具而在某些方面无法获得有意义的就业机会。此外还有强有力的证据表明，患有长期不可控的癫痫的病人患有永久性记忆障碍，抑郁，焦虑，自杀和其他精神疾病的 风险相比于其他人有所增加[3]。上述因素对癫痫患者的生活质量产生负面影响，不仅影响家庭成员和看护人的生活，并且对家庭成员的医疗保健成本构成经济负担。

图 1 所示为一个完整的癫痫发作自动预测系统框图，它包括了数据预处理、特征提取和算法模型等。近些年来的研究重点是基于算法模型的优化来提升癫痫发作预测的精准度，且随着深度学习的发展，卷积神经网络和循环神经网络在预测方面的应用也越来越广泛。本文依据现有算法

模型的类别进行介绍，综述这些神经网络在癫痫发作识别与预测方面取得的研究进展。

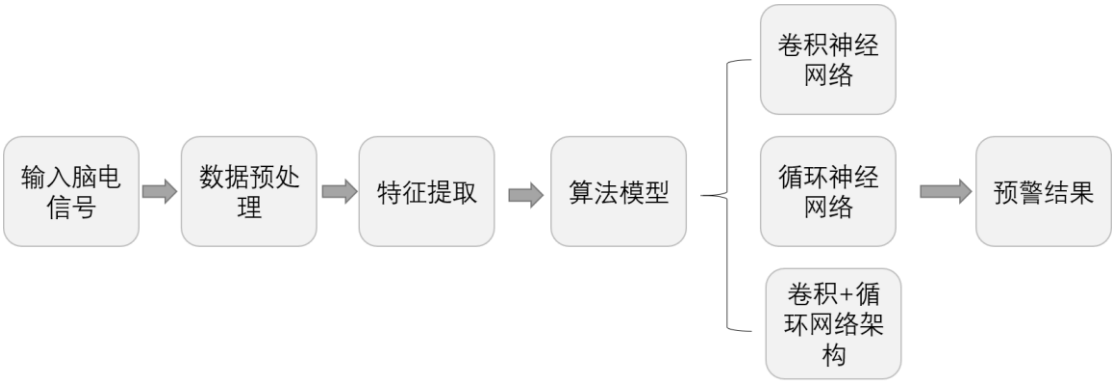


图 1 癫痫自动预测系统框架

准确的癫痫发作预测不仅提前预测患者可能发生癫痫发作的时间，还将在临床癫痫治疗上提供辅助，在设定预警的时间范围内给予一定的药物治疗等。癫痫发作预测显著改善患者的生活质量并减轻护理人员的负担，具有十分重要的现实意义。

2、基于卷积神经网络的癫痫发作预测模型

卷积神经网络（CNN）是一类包含卷积运算且具有深度结构的前馈神经网络，具有局部感知和参数共享特性，是图像处理领域最热的研究点只之一[4]。一个完整的 CNN 包含卷积层、降采样层和全连接层，如图 2 所示。卷积层通过卷积运算使得输入的原始脑电信号的某些特征得以增强，并降低噪声的影响；降采样层也就是我们常说的池化操作，可以减少数据处理量的同时保留有用的信息；全连接层得到神经网络提取到的图片特征，图 2 中全连接层经由“softmax”输出最终的结果。

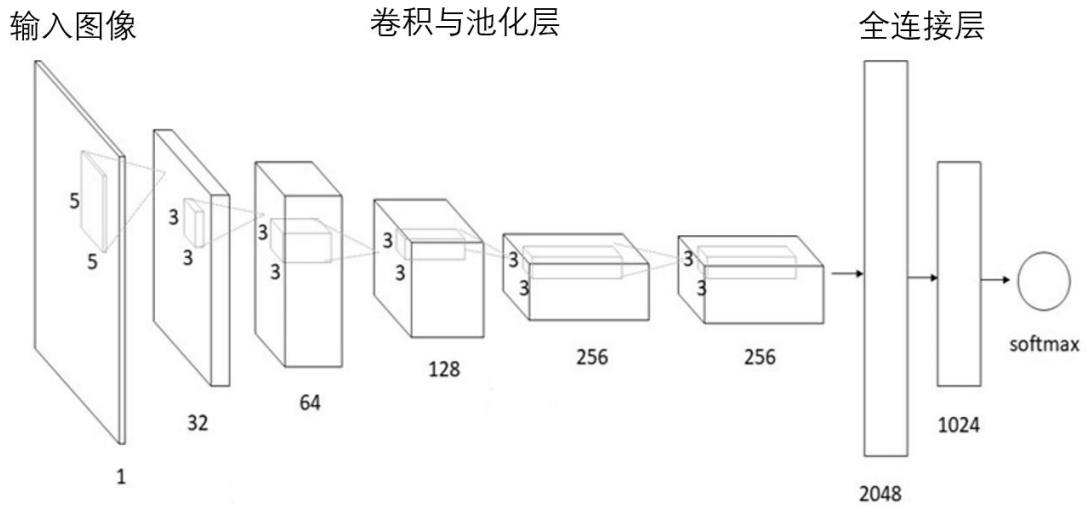


图 2 卷积神经网络结构示意图

图 3 是常见的卷积神经网络癫痫发作预测系统的框架图。系统的关键在于脑电原始数据的重建，即输入图像的重建，现存的方法中依据卷积运算的维度重建为一维、二维以及三维的脑电图。

表 1 列举了最新的近几年使用 CNN 的癫痫发作预测研究，分别使用了不同维度的卷积特征搭建模型。卷积神经网络在癫痫发作预测往往使用重构图像处理，伴随着脑电数据维度改变，可能会造成重要信息的流失。池化层的存在也会导致许多非常有价值的信息的丢失，同时也会忽略掉整体与部分之间的关联因此，基于卷积的模型研究还有待进步。

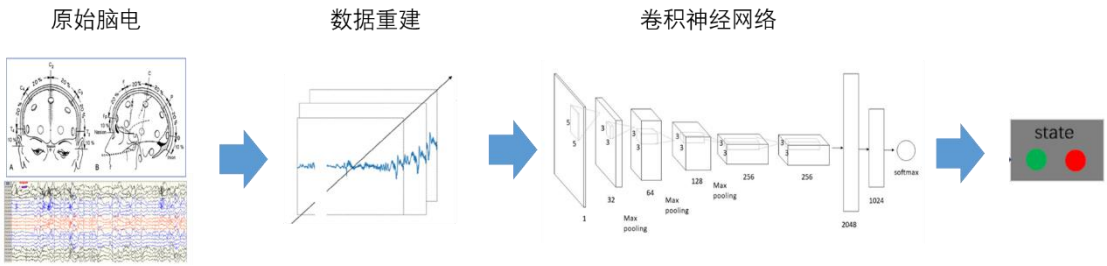


图 3 使用卷积神经网络的癫痫发作预测系统框架图

表 1 基于 CNN 的癫痫发作预测模型

作者	研究内容	重建维度	数据集	模型评价
Yuan Zhang et.al[5]	·通用空间模式（CSP） ·卷积神经网络（CNN）	2 维时频图	·波士顿儿童医院数据集	·灵敏度 92.2% ·误警率 0.12 / h
K.G. van et.al[6]	·卷积神经网络（CNN）	2 维脑电图	·私人数据集	·AUC = 0.924
Nhan et.al[7]	·傅里叶变换（FFT） ·卷积神经网络（CNN）	2 维时频图	·弗赖堡 EEG 数据集 ·波士顿儿童医院数据集 ·美国癫痫协会癫痫发作预	·灵敏度 81.4%、81.2%、75% ·误警率 0.06 / h、0.16/ h、0.21 / h

			测挑战数据集	
Ahmed M. et.al[8]	·自编码机 ·卷积神经网络（CNN）	1 维脑电图	·波士顿儿童医院数据集	·100%分类准确度
Xiaoyan Wei et.al[9]	·卷积神经网络（CNN）	多导 3 维脑电图	·私人数据集	·准确率超 90% ·灵敏度 88.90% ·特异性 93.78%
Ahmet et.al[10]	·卷积神经网络（CNN）	频谱等重构的 3 维重构图像	·波士顿儿童医院数据集	·灵敏度 85.7% ·误警率 0.096 / h

3 基于循环神经网络的癫痫发作预测模型

循环神经网络（RNN）在自然语言处理等时间序列分析领域已经取得巨大成功及广泛应用[11]。循环神经网络包括输入层，隐藏层与输出层，图 4 为一个完整的循环神经网络结构。其中，x 表示输入的数据；s 表示样本在时刻 t 处的记忆，即隐藏层；o 表示输出样本，U 和 V 表示输入和输出的样本权重。不同于传统的前向反馈神经网络，RNN 引入了定向循环，如图 4 的右侧部分，t 时刻的输出除了与当前时刻有关外，还与前一时刻相关。基于 RNN 能够处理输入之间前后有关联的问题，特别是 EEG 这种时间序列的分析优势及其明显。

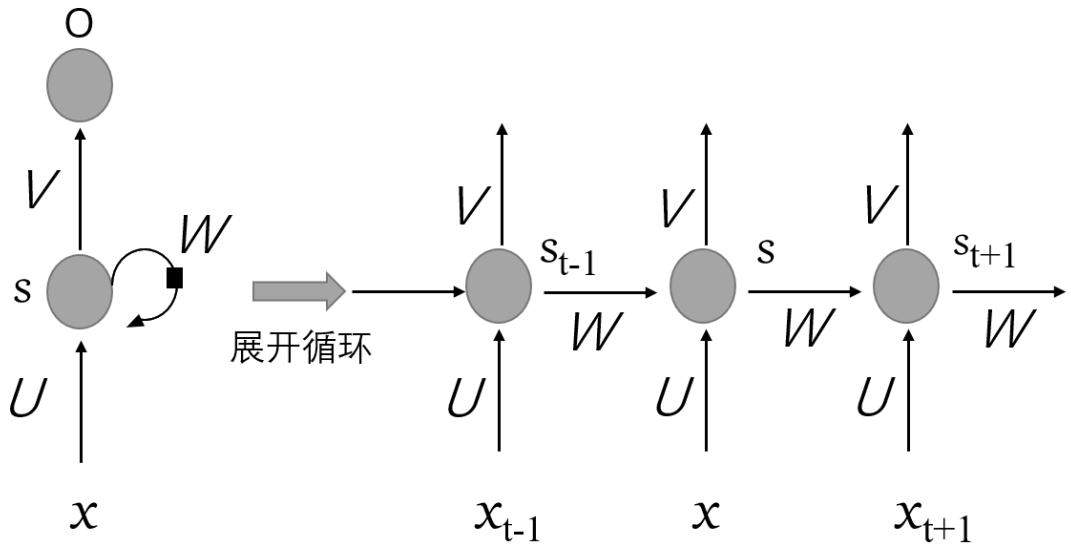


图 4 循环神经网络结构示意图

不同于卷积神经网络的 EEG 分析，原始脑电数据无需重建，RNN 可以直接处理 EEG 序列，这样保证了 EEG 最大程度的信息保留。图 5 所示为一个循环神经网络的癫痫发作预测系统框图。CNN 优势明显，但是仍然存在缺点，RNN 情况类似。因为每个循环层网络的参数共享，RNN 随着层的深度增加，会出现梯度爆炸或者梯度消失的问题。为了解决这些问题的存在，一些变体 RNN，比如 GRU、LSTM、biLSTM 等应用在癫痫发作预测研究上。表 2 列举了近几年使用 RNN 癫痫预测的一些研究工作。

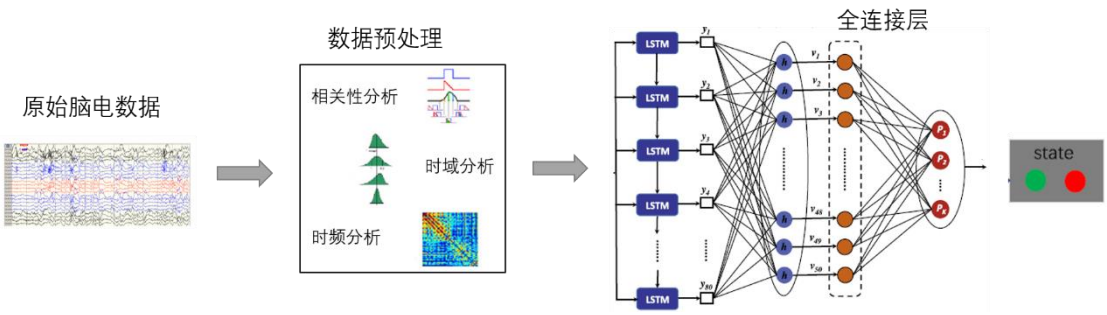


图 5 使用循环神经网络的癫痫发作预测系统框架图

表 2 基于 RNN 的癫痫发作预测

作者	研究内容	数据集	模型评价
Kostas et.al[12]	·特征提取 ·长短期记忆循环神经网络（LSTM）	·波士顿儿童医院数据集	·灵敏度 99.28% ·特异性 99.28 ·误警率 0.107F P/h（15 min）
Ahmad	·门控循环单元	·美国国家神	·灵敏度 0.887

et.al[13]	（GRU） ·长短期记忆循环神经网络（LSTM）	经疾病门口网站数据集	·特异性 0.996
Sachin et.al[14]	·门控循环单元（GRU）	·弗赖堡 EEG 数据集	·准确度 98%（5s）
Xinghua Yao et.al[15]	·递归神经网络（RNN） ·独立神经网络（ind-rnn）	·波士顿儿童医院数据集	·分类性能随片段长度的变化而变化
Arthur et.al[16]	·递归神经网络（RNN）	·私人数据集	·证明在几分钟内存在“发作前期”是完全可行的
Zachary et.al[17]	·长短期记忆神经网络（LSTM）	·私人数据集	·准确度 0.8643 ·F1 分数 0.3035

4 基于循环卷积网络的癫痫发作预测模型

前面我们列举了 CNN 和 RNN 在癫痫发作预测方面的研究，CNN 具有更好的空间信息捕捉能力，而 RNN 则更擅长分析时间序列的关系。我们知道临床采集到的 EEG，不单单只是时间的累计结果，不同导联之间的相互作用以及脑网络节点之间的空间关系同样重要。考虑到 CNN 和 RNN 的优势不同，一种新的架构思想应用在 EEG 的分析中[18]。图 6 所示为一个融合了 CNN 和 RNN 的循环卷积神经网络癫痫发作预测系统，基本思想为在固定大小的时间滑动窗的移动下，CNN 网络学习序列间的空间特征并提取出来，按照时间顺序输入到循环神经网络中，经由 RNN 学习序列间的时间关系，最终给出分类结果。

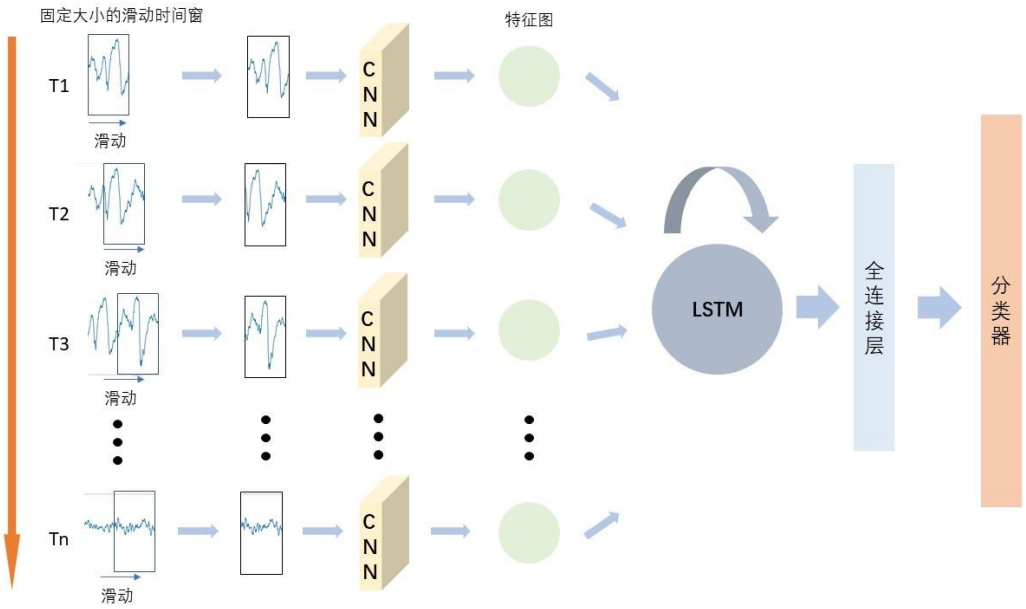


图 6 使用循环卷积网络的癫痫发作预测框架图

基于这种新的架构的研究开展的还不充足，David 等人[19]提出了一种基于卷积和递归神经网络端到端的简单架构，从 28 例患者的 119 例癫痫发作中创建的运动捕捉库中提取时空表示。随后，将测试表示形式与来自于测试数据集分离的五个异常癫痫发作的库之间的余弦相似度距离用于识别具有不符合已知行为的异常模式的癫痫发作测试。基于临床影像资料提取的运动特征用于癫痫的发作检测为癫痫疾病的诊断提供了更多的可能，虽然这种架构没有使用脑电数据，但是为多模态的研究，为基于脑电和影像的癫痫发作预测提供可能。Meysam 等[20]基于临床检测的脑电提出了一种卷积和长短期记忆循环神经网络的癫痫发作检测模型，这种模型可以将重构成二维图片的多通道脑电数据卷积特征输入到 LSTM 中学习序列之间成的关系，最终输出分类结果。Xiaoyan Wei 等[21]提出了类似的研究，不同的是图像的重建维度，在文章中，该团队将脑电图时间序列转换为二维脑电图，进行多通道融合成三维结构。提出了一种可行的方法——长期递归卷积网络（LRCN），实现了癫痫发作的端到端自动预测模型。卷积网络块用于从数据中自动提取深度特征。结合 LSTM 神经网络块来从流数据中区分不同的图像序列和识别前置节段。用独立数据对模型进行了检验，提供了比以往研究中手工设计的方法和单个深层神经网络更高的癫痫发作预测灵敏度和 0.04FP/h 的低错误预测率。这种在外部结构的重构方式使我们研究的一个重点，但是一些团队考虑在神经网络的内部实现两种网络的结合，李飞飞团队[22]提出了一种具有强记忆力的 E3D-LSTM 网络，用 3D 卷积代替 2D 卷积作为 LSTM 网络的基础计算操作，并加入自注意力机制，使网络能同时兼顾长时和短时信息依赖以及局部时空特征抽取。这为视频预测、动作分类等相关问题提供了新思路，是一项非常具有启发性的工作。

5 癫痫发作预测的挑战与难题

卷积神经网络在癫痫脑电的检测与发作预测研究的应用已经越来越深入，但是同样存在一些不可避免的困难与挑战。从癫痫脑电本身出发，大部分的检测预测模型过程中总会不可避免的有

脑电重构的步骤，重构的脑电大多数为二维或者三维的图像。虽然基于重构脑电的卷积神经网络的研究取得了不错的研究成果，但是对于脑电重构过程中的信息流失的不可判断是现在面临的主要问题；从卷积神经网络本身出发，因其平移不变性和池化的存在，会使得网络在训练的过程中，出现对数据量要求高和大以及过拟合的问题的出现。

循环神经网络的记忆特性令其在时间序列的数据处理具有得天独厚的优势。在普通的全连接网络或 CNN 中，每层神经元的信号只能向上一层传播，样本的处理在各个时刻独立，但是在 RNN 中，神经元的输出可以在下一个时间段直接作用到自身。这就导致了一个严重的问题，梯度爆炸或者梯度消失以及无法解决长时程的依赖问题。

循环卷积网络架构是一种综合了卷积与循环网络优势的新架构，但是依旧缺乏这一方面的研究。现阶段的研究主题思想是外部两种网络的嵌套，实现对脑电数据时空的抓取。但是依旧有另外一种思路，内部改变，用卷积操作替代循环网络中的线性操作。对于脑电数据的时空抓取为准确的癫痫发作提供可能，为癫痫灶的定位亦提供借鉴意义。

6 总结

癫痫发作预测一直是患者最关心的问题，同时也极具挑战性，主要是癫痫预测没有具体客观依据，诱因不明确，前驱症状不易被他人观察，对癫痫发作的预测带来不确定因素。本文介绍了癫痫检测与发作预测研究基于深度学习技术的应用、难题与挑战。按照算法技术分类，重点介绍了卷积神经网络、循环神经网络与循环卷积神经网络的优缺点以及未来发展方向。癫痫的自动预测研究能够有效的改善和提高患者的生活质量和精神面貌，希望本文能够为癫痫发作预测研究提供借鉴意义。

PU-011

关于防疫期间医院计算机信息系统的讨论

王东飞
郑州大学第一附属医院

目的 讨论新冠肺炎疫情期间医院信息工作与非疫情期间的侧重，回顾医学信息标准化发展历程与发展方向

方法 对近来所发生的信息安全事件进行了简要的发散性分析，简单描述的标准化与大数据化实际应用中可能面临的困难和已经产生的问题

结果 医学信息的特殊性和其关键的作用

结论 指出医学信息的特殊性和其关键的作用，强调了其必须在法律的保护和监督下才能稳步良性发展

1.医院防疫期间信息岗位的调整与挑战

医院是一个经常面临挑战的地方，其信息岗位也同样如此，因为疫情的需要，所有的病人要先分为发热和不发热，要在步入门诊楼的第一步就进行体温遥感信息分析，同时在相应业务系统，分化流程。对于隔离病区，我和同事们在隔离期间是分组上班的，以防传染后在整个处室蔓延，尽可能的控制在几人小组中间以防万一。同时在年前就对各楼层进行过例行网络电井巡查，在收到命令后第一时间保证隔离病区的各项信息网络工作正常展开，在隔离病区过年开始接收病人封闭前就已经架设好了并区内隔离区，办公区，休息区等之间的网络和软硬件。幼儿在年前送回老家，再次回来已经是四月份解封。还有的同事因为工作单位是医院，居住的小区和家人不希望他回家，也暂居同事家中，如此数月，众志成城，圆满完成了防疫信息保障任务，这对我们所有人都是一次挑战，不仅仅是技术层面，包括了家庭，心理，生活的各个方面。而应对疫情的技术支持，从来就不是单独的，因为各个医疗信息系统组成的信息结构是复杂的，相互支持的，下面就此进行简要讨论。

1.1 医学信息输出终端的信息为大数据提供基础

医疗信息终端大致可分为功能型，已郑州大学第一附属医院为例，有医院分诊叫 号系统，医疗放射胶片打印系统，检查结果自助打印系统，门诊自助挂号系统及 自助挂号机，还有微信和网络的自助挂号系统等， 以上这些系统是面向患者的 。面向医院内部医护人员使用的系统终端则有，护理系统，门诊电子病历系统和 病房电子病历系统，重症及手术麻醉系统，红会系统，合理用药系统等几十个在 用或停用的系统终端。其信息也是以患者身份信息为基准在数据库中进行关联的［1］。医疗终端的载体基本包含了当下所有的移动设备，PC 机，笔记本电脑，平板电 脑，手机客户端。其基本覆盖了所有的患者信息，从图片信息到文字影响信息。 其输出的终端信息包含纸质的病例文件，胶片打印的影像文件，病理蜡块，影像图像电子文档，甚至微信支付宝的电子回执也归属于医疗信息。信息输出的多样化为医疗信息大数据提供了丰富的资源和操作基础，也给大数据提出了切实的难题，那就是如何将这些信息转化为可以高效利用的大数据标准化信息。

1.2 医学信息标准化所产生的问题

信息化在方便的同时，其也在以极快的速度产生着更多更大的和更多格式的数据，例如影像检查，每个患者的影像检查后的数据大约在 300MB 到 500MB 左右，而胶片仅打印典型化的三十到六十图，而影像报告也是总结性的下定论，其数据量相当于一个 30 分钟左右的电影，而这个数据是需要存储在医学信息系统中等待复查调阅的。如果是一百个这样的病人，其数据量就是以 TB 为单位的了。而医院的检查数量要远大于这个数量级，这就要求在尽心大数据应用之前，进行大数据瘦身，数据量越大会造成数据分析速度越慢，特别是同一个人的冗余数据越多，对其本身病史的大数据分析就影像越大。标准化的另一个问题就是信息泄露，信息泄露听起来像是个安全问题，可实际上在技术层面上标准信息的加密难度造成了信息安全等级对防火墙的依赖要大于通信信息安全，因为你不能对大量的医疗信息进行加密处理，这样加密解密本身就会对使用效率造成困扰，而通讯信息不管是建立临时通讯密道，或是编译反编译信息都要比医学信息处理的量小很

多，故而医学信息的攻击事件时有发生。，例如 2017 年发生 的一系列医疗信息被盗取牟利事件 5 月，黑客倒卖医院数据落网引发广州医药圈 震荡；8 月，英国 NHS 在 8 月份遭到匿名黑客组织的攻击，被窃取的信息包含 32,501 条行政数据以及 120 万条病历记录。9 月，某部委医疗服务信息系统遭“黑 客”入侵，超过 7 亿条公民信息遭泄露，8000 余万条公民信息被贩卖；10 月，网 络攻击事件使香港宫颈癌疫苗断货；10 月，伦敦一知名整形医院遭网络攻击，黑 客称名单中有皇室成员。这只是当年发生的一些例子，其来源或者漏洞可大致分为，数据存储的安全问题；配置错误的安全设置；缺乏实际的安全解决方案来保 护数据；使用未经修补和不再受支持的软件，并具有可利用的漏洞。

2 医疗信息大数据分析使用中的硬件限制

谈数据必谈存储，数据存储主 要分布于个人的终端设备和各医疗机构自己的机房数据库，这就面临着容量问题 ，延迟问题，安全问题，和运营成本。像之前提到的一个人的医疗信息不仅限于文本文字，其 中还包括其所做的所有检查检验所带来的核磁，超声等影像资料，还有一些视频 陈述描述甚至手术影像资料等等，这是一生的数据，所以海量数据存储系统也一 定要有相应等级的扩展能力。与此同时，存储系统的扩展一定要简便，可以通过 增加模块或磁盘柜来增加容量，甚至不需要停机。延迟问题来源于大数据，比如 郑州大学第一附属医院其早间门诊高峰时刻，调取患者核磁超声等影像资料会和 护士将药品发送到药房取药的摆药单时间发生重叠，这是由于医护人员工作习惯 造成的认为延迟因为晨会交接班，查房，下医嘱，审核医嘱，生成摆药单，配药 取药，这是一套顺序的不变的流程，所以会造成在某个时间段必然延迟增加。而 能解决的延迟是指比如在不特定的时间点调取某个医疗信息所需要消耗的时间， 其规律为调取的数据越大延迟越大，调取的时间点越久远，延迟越大，调取数据 需要跨平台跨数据库，延迟越大，所以医疗机构一般会有自建机房已就近存储就 近维护，而其投入的成本又远小于专业数据存储服务机构，造成了专业性能方面 的差距。数据拿到了，但是对海量的数据进行分析，不仅仅需要专业的医疗知识去明确目标，还需要高效的算法和硬件设备去遍历庞大的数据库。

医疗信息的大数据，大”，意味着代价不菲。成本控制是关键的问题。 想控制成本，就意味着我们要让每一台设备都实现更高的“效率”，同时还要减 少那些不必要的部件或者信息。目前，像重复数据删除等技术已经进入到主存储 市场，而且现在还可以处理更多的数据类型，这都可以为大数据存储应用带来更 多的价值，提升存储效率。在数据量不断增长的环境中，通过减少后端存储的消 耗，哪怕只是降低几个百分点，其效果就非常明显。此外，自动精简配置、按照医疗信息的使用频次，年代久远排序分配资源也可以提升存储的效率。很多大数 据存储系统都包括归档组件，尤其对那些需要分析历史数据或需要长期保存数据 的机构来说，归档设备必不可少。

2.1 医学信息标准化的实践应用

标准化同时指的不仅仅是信息格式信息内容的标准，其涵盖了硬件端口接入的标准，安全协议的标准，使用的标准，商业化的标准，往期信息存储的标准甚至信息报废销毁的标准等等。这需要整个相关产业行业的共同参与和支持因此，医院在其现有硬件设备的条件限制下所选择的

的存储管理都提供纯软件的形式，可以直接安装在用户已有的、通用的或者现成 的硬件设备上。从单位容量存储成本的角度看，价格低廉速度慢的 TB 存储老旧设 备，依然可以用来建立类似病案档案馆存储一些二三十年前的基本不被调用或者 信息不全面没有学术研究参考价值的历史数据，与个医疗机构建立各自机房维护医疗信息相对的，就是医疗信息的集中存储，外包专业网络安全机构维护，其中的好处就是深度发掘医学信息的价值，不是指私人隐私的买卖，这在任何时候都是违法的。医学信息的标准化价值在于可以针对某个病种，家族遗传等建立完整的模型，分析地区疾病原因，精炼数据，为医学行业的发展提供助力，同时标准化也意味着得到国家力量的帮助。比如，自 2003 年起，英国的 NHS（National Health Service）采取行动建立一个集中的电子病历登记系统。该系统受到英国政府门户网站的保护，该门户由微软建立。这个程序被称为电子文件开发和实施方案（ErDIP）。医疗信息泄露或者买卖，其根本上是获取其中的商业利益，个人黑客也好，受雇于公司的雇佣黑客也罢，其最终目的是享有信息转化而来的商业利益，毕竟仅从技术层面上来说，医疗信息数据从来都不是最难获取的，正是因为其获利与所承担的风险不匹配，即低风险高收益，所以才有了屡禁不绝的医疗信息泄露事件。所以从法律层面上加强保护使违法成本远大于风险收益也是必须要考虑的。

2.2 医学信息标准化必然需要国家法律的参与

因为医学信息的敏感性和必须性，其走的每一步都需要有法可以，受到法律的保护和监督，以美国为例，自 1974 以来，美国通过了多项联邦法律，规定患者、医生和其他被覆盖实体对医疗数据的隐私权和保护。其中最全面的法律是 1996 年通过的健康保险便携性和责任法案，（HIPAA）[2]，该法案在 2013 又附加了综合隐私标准作为修订案。HIPAA 提供了医疗隐私的联邦最低标准，制定了使用和披露受保护健康信息（PHI）的标准，并对违法行为提供民事和刑事处罚。该法案涵盖所有医疗参与实体必须遵守的法律规定的任务，医保、医保结算，从公立医疗到私立医疗，由医院到社区诊所。和电子医疗信息传输的运营商等参与者也相应的受到 HIPAA 规则和条例的约束。之后在 2008，国会又通过了《2008 号遗传信息非歧视法案》，其目的是禁止基因性别筛查或者因携带可致病基因而拒绝其医疗保险投保或就业的个人进行基因歧视，从基因层面保护医疗信息。法律还规定，所有接触持有的个人遗传信息的机构或个人必须保护其持有的信息并禁止任何非法律允许的情况下披露该遗传信息。

2013，在遗传信息非歧视法案被通过后， HIPAA 法案也相应的修改了其与之对应的法规，将遗传信息包括在受保护健康信息（PHI）的定义中。该规则还通过扩大业务伙伴的定义来扩展 HIPAA，即由信息服务供应商或医疗参与人员扩大到包括发送或访问 PHI 的任何实体，如健康信息咨询服务的公司等等。各手机终端制造商，网络运营商在提供轻便只能快捷的服务的同时也被要求对所接触的信息具有保护的义务与功能。比如手机上的远程身份验证锁定功能就是针对这样的需要做出的改进，使终端设备遗失的情况下能首先保证终端数据的及时封闭销毁。

医学信息的数据标准化和大数据应用已经在稳步进行，从智能手环上的心率信息，睡眠信息等运动信息的记步对比功能等等，到家庭健康信息随测，手机远程会诊到医院开展的分病种信

息大数据。我国的法律法规也仅涉足了医疗卫生规范，食品药品安全监督等方面，最近一次颁布新规也是2010年的电子病历书写规范，未对已经乱象分生的医疗信息数据方面做出有效的规范或者有意义的尝试，仅靠各个医疗机构对其信息的采集保护是力有未逮且不名正言顺的，毕竟医疗机构的职能不是信息安全与监管保护。

PU-012

医学领域自然语言处理的研究热点分析 ——基于CiteSpace的知识图谱分析

陈婕卿、张锋、曾可、姜会珍
北京协和医院

目的 探讨自然语言处理技术在医学领域的研究思路和发展趋势，剖析研究热点。

方法 以Web of Science数据库为基础，检索2000年至2020年期间的相关文献，采用CiteSpace软件对文献进行科学计量分析并绘制知识图谱。

结果 共纳入文献1029篇，研究数量逐年递增，案例研究、电子病历应用、早期预测、跨语言索引和发现不良事件等是研究的热点方向。

结论 根据自然语言处理技术在医学领域的研究热点分析，对同类研究提供了参考和建议。

1 引言

自然语言处理（Natural Language Processing, NLP）是通过建立形式化的计算模型来分析、理解和处理自然语言的学科。NLP 是人工智能（Artificial Intelligence, AI）研究中的一个子领域，也是一门交叉学科，研究者们希望通过 NLP 研制出能够理解并生成人类语言的计算机系统，或者建立形式化的适用于计算机处理的语言模型和研制分析处理语言的各种算法[1]。NLP 的应用领域非常广泛，包括：机器翻译、人机对话、信息检索、信息提取、自动文摘和文本分类等等。

近年来，随着 NLP 技术的不断发展和创新，其在医学领域所发挥的作用也越来越显著。包括：利用 NLP 技术完成电子病历文本信息的语义识别、进行临床医学文献报告的精准语义识别，也可以推广到对电子病历质量控制、病案检索与发现、临床风险预警模型构建等临床应用的需求中，结合语义推理和特征学习方法，进行典型临床应用场景下的知识服务关键技术的研究。通过 NLP 技术实现高质量的医学语义识别，对于临床信息、文献和报告等文本资料的利用和价值提升都具有重要意义。

为梳理在医学领域中，NLP 技术的发展趋势、应用现状和当今的研究热点，本文收集了自2010年至2020年于Web of Science（WOS）中收录的涉及NLP技术在医学领域应用的文献，采用文献计量学的研究方法，从多个角度分析 NLP 技术在医学领域的研究思路和趋势，剖析研究动态和热点，为今后的同类研究提供理论依据。

2 资料与方法

2.1 研究资料

本研究的数据来源于 WOS，选择 Web of Science™ Core Collection 数据库，检索公式为：主题（“Natural Language Processing”）AND 主题（medical），检索语言为“English”，检索起始时间为 2010 年 1 月 1 日至 2020 年 5 月 23 日，共得到 1669 篇文献。进一步精炼检索结果，筛选文献类型为“Article”，最终得到有效文献共计 1029 篇。检索与下载的日期为 2020 年 5 月 23 日。将纳入的文献用纯文本格式导出，包含题名、作者、关键词、摘要、研究机构、期刊、资助基金和文献来源等。利用 SCI 数据库的检索分析功能得出文献发文量的年代分布、国家/地区分布、期刊分布、研究机构分布和基金资助机构分布等。另外，采用 CiteSpace 软件绘制知识图谱，分析该领域的研究热点，包括关键词的共现网络分析和文献的共被引分析。

2.2 研究方法

本研究采用基于 Java 平台的 CiteSpace 软件（版本号：5.6.R2）绘制科学知识图谱，它是一个面向科研评价领域的常用工具，是专门针对科研文献数据设计的可视化分析软件[2]。

在 CiteSpace 软件中，时间参数设置为 2000 年至 2020 年，时间区间间隔的参数 Years Per Slice 定义为 1，在阈值选项中设置 Top N per slice 为 50，运行软件绘制知识图谱。在 CiteSpace 的知识图谱中，分析对象用节点显示，该对象的引用频次越高，节点越大。共被引关系以连接线表示，两者首次共被引的时间由连线颜色表示，联系强度用连线的粗细及长短表示[3]。

在关键词的共现网络分析中，同时选取 Term 和 Keyword 两类文献节点进行文献主题的共词（co-words）分析。其中，Term 分析的功能是对文献中名词性术语的提取，主要从文献的标题、摘要、关键词和索引词位置提取；Keyword 主要是提取文献作者的原始关键词[2]。另外，对高频关键词节点中的高被引文献进行共被引分析，总结了近二十年来 NLP 技术在医学领域的研究热点和发展趋势。

3 结果与讨论

3.1 文献分布

（1）文献发文量的年代分布

根据 2000 年至 2019 年期间等文献统计数据[注]，NLP 技术在医学领域的研究发文量呈现总体上升趋势，如图 1 所示。从 2000 年的 9 篇，到 2019 年的 202 篇，发文量增长了 22 倍。2016 年至 2019 年的增长幅度较过往年份更为显著，说明近五年内 NLP 技术在医学领域的研究越来越受到学术界的关注。

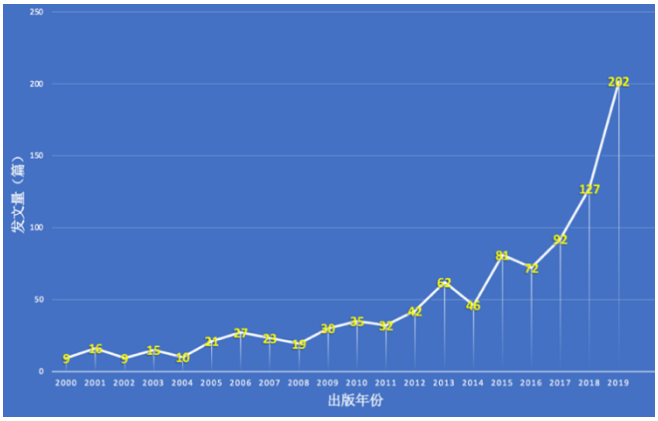


图 1 2000 年至 2019 年相关研究发文量线图

注：由于 2020 年的文献数据不足一年，故未纳入此部分的数据进行分析。

（2）文献发文量的国家/地区分布

2000 年至 2020 年期间，共有 42 个国家/地区发表过关于 NLP 技术在医学领域的研究论文。如图 2 所示，发文量最多的国家是美国（630 篇），占总发文量的 61%；其次是中国（83 篇），占总发文量的 8%，虽然我国在 NLP 研究领域的起步较晚，但是随着近年来我国不断加强国际间的学术交流和合作，发文量逐渐跃升至第 2 位；另外，法国、英格兰、德国、加拿大、澳大利亚、西班牙、意大利和日本分列发文量的第 3～10 位。

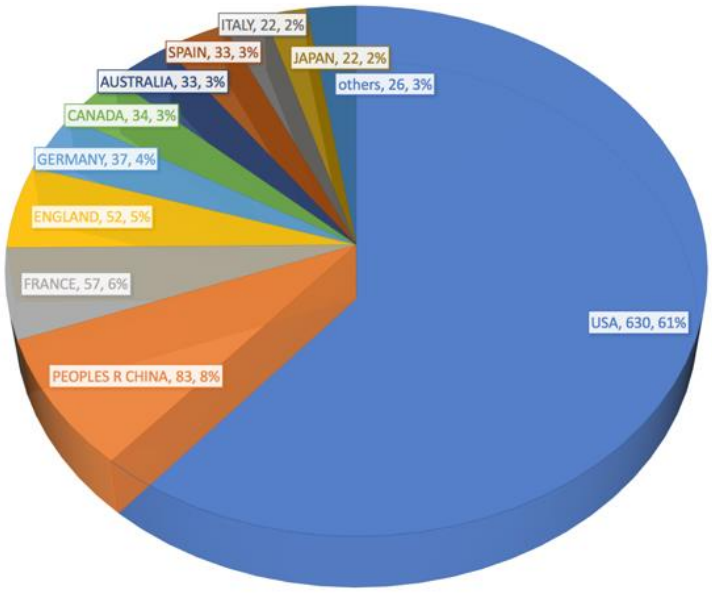


图 2 2000 年至 2020 年相关研究在不同国家/地区发文量饼图

（3）文献期刊分布

表 1 2000 年至 2020 年相关研究文献来源期刊分布（前 10 位）

序号	文献来源期刊	文献记录数（篇）
1	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS	123

	ASSOCIATION	
2	JOURNAL OF BIOMEDICAL INFORMATICS	105
3	INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS	60
4	BMC MEDICAL INFORMATICS AND DECISION MAKING	40
5	JMIR MEDICAL INFORMATICS	28
6	BMC BIOINFORMATICS	26
7	METHODS OF INFORMATION IN MEDICINE	20
8	JOURNAL OF MEDICAL INTERNET RESEARCH	17
9	PLOS ONE	17
10	ARTIFICIAL INTELLIGENCE IN MEDICINE	15

2000 年至 2020 年期间，收录关于医学领域 NLP 的研究论文的期刊共计 364 个，文献来源期刊分布见表 1。其中《JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION》收录的文献数量最多，高达 123 篇。

（4）研究机构分布

表 2 2000 年至 2020 年相关研究机构分布（前 10 位）

序号	研究机构	文献记录数（篇）
1	UNIV UTAH	63
2	MAYO CLIN	57
3	BRIGHAM WOMENS HOSP	55
4	VANDERBILT UNIV	53
5	HARVARD UNIV	50
6	MASSACHUSETTS GEN HOSP	46
7	HARVARD MED SCH	45
8	COLUMBIA UNIV	41
9	MIT	31
10	STANFORD UNIV	31

UNIV UTAH 发文（63 篇）位列第一，其次是 MAYO CLIN（57 篇），第三位的是 BRIGHAM WOMENS HOSP（55 篇）。发文量前十的研究机构分布见表 2，其中有 3 所来自医院或者医疗机构（包括：MAYO CLIN, BRIGHAM WOMENS HOS, MASSACHUSETTS GEN HOSP），其余的 7 所均来自高校，可见高校是医学领域 NLP 技术的主要研究机构。

（5）基金资助机构分布

表 3 2000 年至 2020 年相关研究的基金资助机构分布（前 10 位）

序号	基金资助机构	文献记录数（篇）
1	UNITED STATES DEPARTMENT OF HEALTH HUMAN SERVICES	320
2	NATIONAL INSTITUTES OF HEALTH NIH USA	303
3	NIH NATIONAL LIBRARY OF MEDICINE NLM	128
4	NATIONAL NATURAL SCIENCE FOUNDATION OF CHINA	35
5	NIH NATIONAL CANCER INSTITUTE NCI	29
6	NIH NATIONAL INSTITUTE OF GENERAL MEDICAL SCIENCES NIGMS	22
7	NIH NATIONAL HEART LUNG BLOOD INSTITUTE NHLBI	20
8	CANADIAN INSTITUTES OF HEALTH RESEARCH CIHR	19
9	NIH NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES NCATS	16
10	NIH NATIONAL CENTER FOR RESEARCH RESOURCES NCRR	16

在 1029 篇文献中，获得基金资助的文献共 818 篇。其中，支持最多的机构是 UNITED STATES DEPARTMENT OF HEALTH HUMAN SERVICES，高达 320 篇，文献的基金资助机构分布见表 3。另外，在中国学者发表的论文中，国家自然科学基金资助的项目发文量最大，共计 35 篇。

（6）研究高被引文献分布

表 4 2000 年至 2020 年相关研究高被引文献分布（前 10 位）

序号	文献标题	第一作者	文献来源期刊	出版年份	被引次数
1	Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications	Savova, Guergana K.	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	2010	625
2	A simple algorithm for identifying negated findings and diseases in discharge summaries	Chapman, WW	JOURNAL OF BIOMEDICAL INFORMATICS	2001	404
3	2010 i2b2/VA challenge on concepts, assertions, and	Uzuner, Oezlem	JOURNAL OF THE AMERICAN MEDICAL	2011	315

	relations in clinical text		INFORMATICS ASSOCIATION		
4	Discontinuation of Statins in Routine Care Settings A Cohort Study	Zhang, Huabing	ANNALS OF INTERNAL MEDICINE	2013	310
5	Measures of semantic similarity and relatedness in the biomedical domain	Pedersen, Ted	JOURNAL OF BIOMEDICAL INFORMATICS	2007	249
6	Automated encoding of clinical documents based on natural language processing	Friedman, C	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	2004	231
7	The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text	Rindflesch, TC	JOURNAL OF BIOMEDICAL INFORMATICS	2003	227
8	The SIDER database of drugs and side effects	Kuhn, Michael	NUCLEIC ACIDS RESEARCH	2016	222
9	MedEx: a medication information extraction system for clinical narratives	Xu, Hua	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	2010	221
10	Robust Replication of Genotype-Phenotype Associations across Multiple Diseases in an Electronic Medical Record	Ritchie, Marylyn D.	AMERICAN JOURNAL OF HUMAN GENETICS	2010	200

对所有文献的被引次数进行频次分析，在 2000 年至 2020 年期间，高频被引文献来自于美国的 Savova, Guergana K.研究团队，研究成果源自梅奥医学中心的 cTAKES 系统，主要进行临床的文本分析和知识提取研究。关于 cTAKES 系统，杨晨柳[4]指出梅奥医学中心的在系统架构上采用开源框架，开发模块化组件、构建临床语料库和注重创新等设计理念，也为临床文本自然语言处理提供了新方向。其他高被引文献分布的前十位信息详见表 4。

3.2 研究热点分析

（1）关键词的共现网络分析

对高频的关键词进行分析得到本研究的关键词共现网络分析图谱，如图 3 所示。在共现网络分析过程中，我们进行关键词的聚类。采用软件中的 LLR 模式（对数极大似然率）进行聚类命名，聚类的每一个区域对应一个标签，#号后面为聚类标签的名称。从图中可以看到，一共生成 13 个聚类，顺序是从 0～12，数字越小，代表聚类中所包含的关键词越多。每个聚类是多个紧密相联系的关键词组成。聚类的关键词标签也体现了在这一领域的研究热点方向。

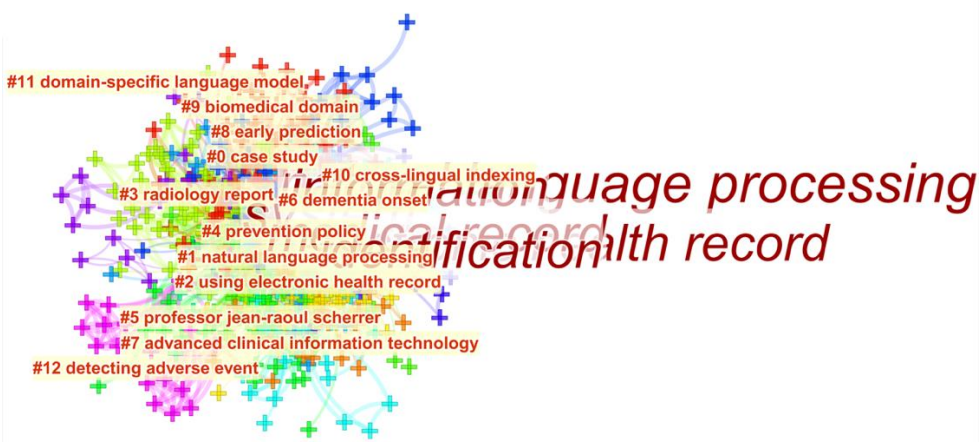


图 3 2000 年至 2020 年相关研究关键词共现网络分析图谱

根据 CiteSpace 软件的 Timeline view 功能分析，可以得到近二十年的高频关键词演化，2000 年至 2020 年关键词共现网络 Timeline 视图，如图 4 所示。图中左侧部分为时间轴，时间轴上实现部分为关键词的出现节点，节点之间的连线粗细代表共现的强度。例如：#1 natural language processing 聚类中的关键词高频出现在 2004 年至 2020 年的研究文献中。通过关键词的共现网络分析，可以看到案例研究（case study）、电子病历应用（using electronic health record）、放射报告（radiology report）、预防政策（prevention policy）、先进临床信息技术（advanced clinical information technology）、早期预测（early prediction）、跨语言索引（cross-lingual indexing）、特定领域语言模型（domain-specific language model）和发现不良事件（detecting adverse event）等是研究的热点方向。另外需要说明，#1 natural language processing 和#9 biomedical domain 与检索词一致；#5 professor jean-raoul scherrer 为人名；#6 dementia onset 可能为突现词。

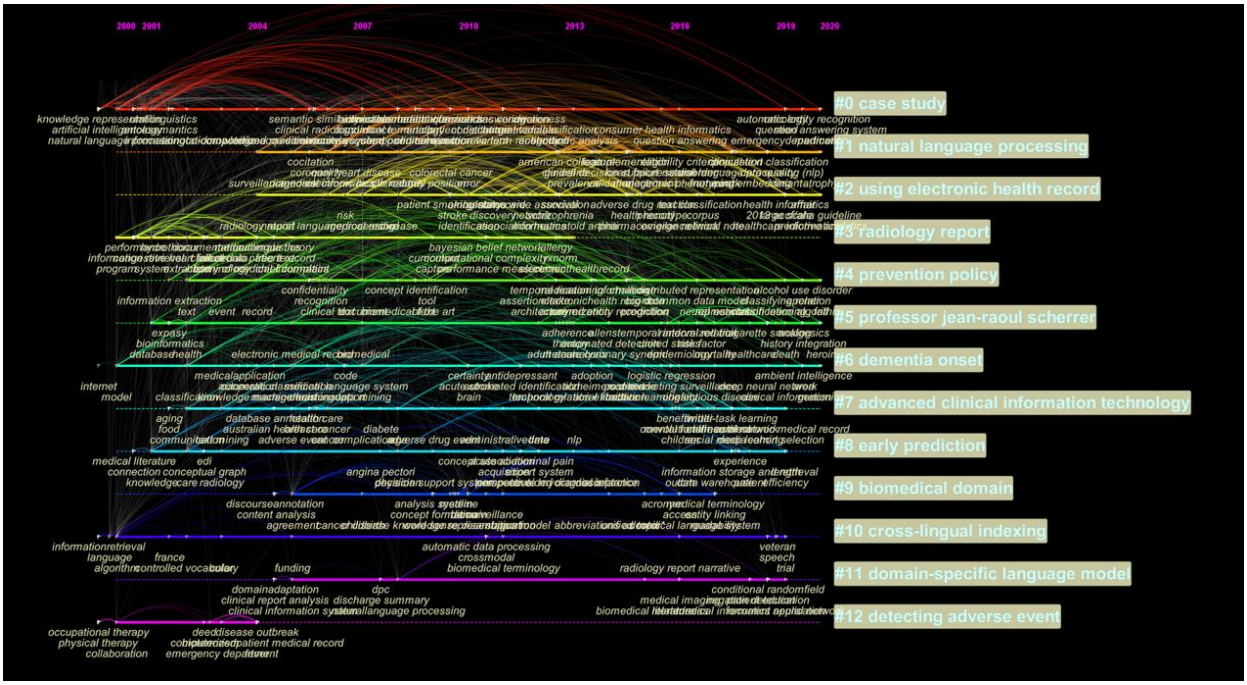


图 4　2000 年至 2020 年相关研究关键词共现网络 Timeline 视图

回顾近三年 NLP 技术在医学领域的研究，也同样可以看到大量与上述关键词有关的研究工作。易应萍[5]等选取了 300 篇临床的入院记录进行案例研究，通过基础模型构建和 NLP 技术等方法，有效地提升医学文本实体识别的准确性。胡佳慧[6]等提出中文电子病历等标注思路，进一步构建可视化交互平台，基于电子病历文本中字与词的特征，综合利用 NLP 和机器学习方法开展临床命名实体识别的实证研究，并证明进行电子病历标注方法和所构建的标注平台适用于临床文本处理，能够促进医学临床文本资源的知识关联化。任彦与国外研究者 E.Pons[7]等进行合作，利用 NLP 技术快速提取临床指征和 CT 检查结果，得到放射报告中的可用信息。张小亮[8]构建基于 NLP 的临床合理用药知识图谱，并用优化后的机器学习模型预测未标注的药品说明书。宋杰[9]通过使用 NLP 技术和人工标注法分别对 1599 例次数据同时进行处理，验证了 NLP 对于分析护理不良事件非结构文本的可行性。

（2）文献的共被引分析

共被引分析（Co-Citation Analysis）的概念由美国的情报学家 Henry Small 于 1973 年首次提出，是指两篇文献共同出现在第三篇施引文献的参考文献目录中，那么这两篇文献形成了共被引的关系[10]。

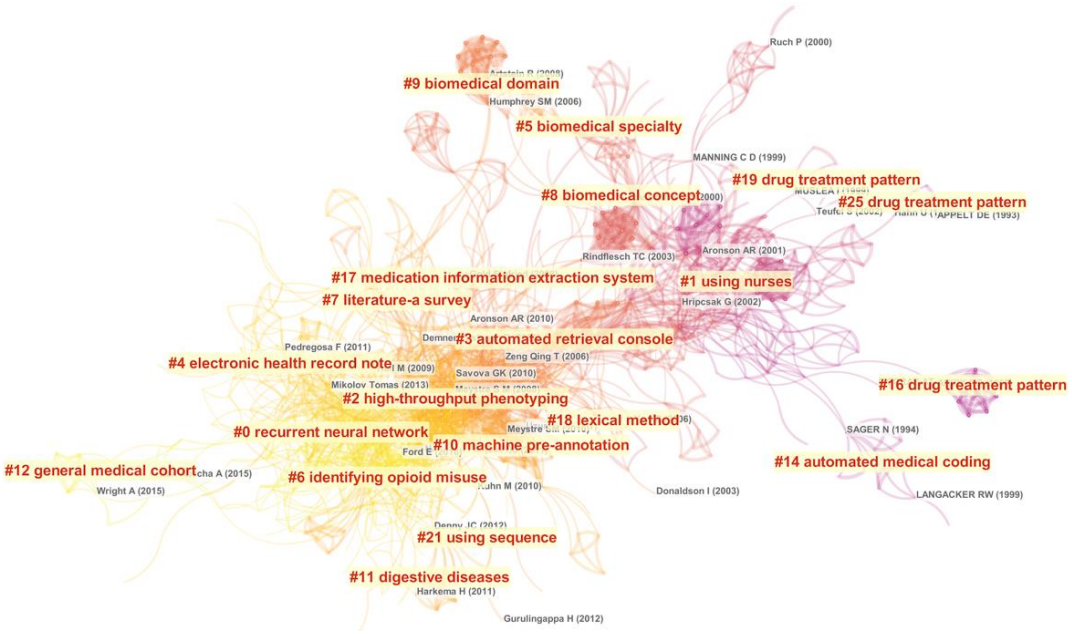


图 5 2000 年至 2020 年相关研究文献共被引网络的聚类图谱（局部）

在共被引分析图谱中，节点的大小反应文献被引用的次数，文献的共被引反应了单个文献的引用次数，节点之间的连线粗细代表共被引的强度。同样通过 CiteSpace 软件进行聚类分析，得到文献共被引网络的聚类图谱，如图 5 所示。

从图中可以看到，一共生成了 26 个聚类，顺序是从 0～25。研究热点方向主要集中在生物医学领域，体现 NLP 技术在生物医学领域应用的特异性。另外，也在研究算法上不断进行创新，包括：循环神经网络（recurrent neural network）、高通量表型（high-throughput phenotyping）等。

4 结论

在本研究中仍有不足之处，一方面由于不可能穷尽 NLP 在医学领域的全部检索词，所以可能存在数据漏检问题；另一方面因为 CiteSpace 软件在分析过程中很难对突现词进行识别和筛查，所以有可能会对研究前沿的预测产生影响。

综上所述，通过近二十年的医学领域 NLP 研究趋势和发展方向的文献计量分析和可视化分析，总结了该领域的研究热点，可为同类研究提供参考。未来，研究者们不仅可以利用 NLP 技术在医学案例研究、电子病历应用等方面，结合医学知识体系，提升为临床医生提供临床知识服务的质量和效果，而且基于生物医学领域的 NLP 技术算法创新，也将成为该领域的重要发展方向。

PU-013

基于医联体的医疗影像云平台设计与应用

郭金鸽
郑州人民医院

目的 作为中原地区的三甲医院在看好病、看准病的同时，郑州人民医院与卫宁健康战略合作，从医院管理到临床流程精准落实。

方法 通过构建医联体影像云平台，实现医联体内影像数据共享

结果 提升医联体内医疗机构服务能力和信息化建设水平，让患者更加便捷的享受高质量的影像诊断服务

结论 致力于找到适合医院发展并为患者提供便利的线上线下一体化服务。

1、国家政策驱动：

2018 年 7 月 12 日，为更加深入“互联网+医疗健康”的发展，国家卫生健康委员会、国家中医药管理局联合发布《关于深入开展“互联网+医疗健康”便民惠民活动的通知》

国家卫生健康委办公厅于 2018 年 12 月 3 日和 2019 年 3 月 5 日分别印发了《电子病历系统应用水平分级评价标准（试行）》和《医院智慧服务分级评估标准体系（试行）》

2020 年 2 月 3 日，为贯彻落实党中央、国务院关于新型冠状病毒感染的肺炎疫情防控工作的总体部署，国家卫生健康委办公厅下发了《国家卫生健康委办公厅关于加强信息化支撑新型冠状病毒感染的肺炎疫情防控工作的通知》

2、医院发展需求：

郑州人民医院与卫宁健康科技集团股份有限公司战略合作，探索以郑州人民医院总部为核心，覆盖医联体内 18 家医疗机构的信息化发展。本着以患者为中心，做“有温度”的医院，把优质医疗服务资源下沉到医联体内，实现资源共享，提升区域内整体服务效能。

3、疫情环境影响：

新冠肺炎疫情发生以来，为做好疫情防控管理，降低居民在疫情期间病毒交叉感染和传播风险，居家隔离成了全民防疫的“特效药”。我院作为疫情防控定点医院，为缓解诊疗压力，给患者提供优质服务，发展互联网医院、互联网诊疗尤为重要。

1、云平台服务的建设

基于信息安全负责医疗信息集中管理，对信息分类管理。尤其是在疫情期间患者关注医院 app 或者公众号，不出门就可以预约检查，检查完成后直接在手机等移动终端看影像报告，并通过在线咨询选择对应的挂号医生进行报告解读。

1.1 可共享信息，患者影像信息，疑难病例等信息上传云平台集中管理；

1.2 不可共享信息，患者医疗流程产生的信息，如挂号，缴费等由基层自行管理，各医疗机构对云平台提供接口供应用 App 调用；

1.3 云端的影像数据全部经过脱敏加密。保障互联网传播过程中病人隐私、数据完整性。

1.4 医疗信息单点应用，根据用户主要分为 2 类

1.4.1 患者主要功能：

（1）个人医疗信息管理：所有个人医疗信息历次查询；个人信息根据时间范围授权给诊断医生。

（2）就诊流程实时处理：预约；挂号；缴费；就诊时间预估与提醒，减少等待和往返；检查时间和结果预估与提醒；云取结果、二维码分享；院内导航，定位当前位置，导航就诊科室。

1.4.2 医生主要功能：

（1）远程会诊、AI 辅助诊断；

（2）权限范围内不同场景诊疗信息诊断查阅共享，不接触患者，通过视频、语音等方式进行会诊。

1.5 医联体间

（1）基层患者或医生通过二维码、云平台预约上级医院检查，精确到专家、时间、检查间，无需再挂号排队，直接云取结果。

（2）AI 应用扩展到下级医疗机构。辅助基层进行骨折、肺结节、新冠肺炎、脑卒中等检测。

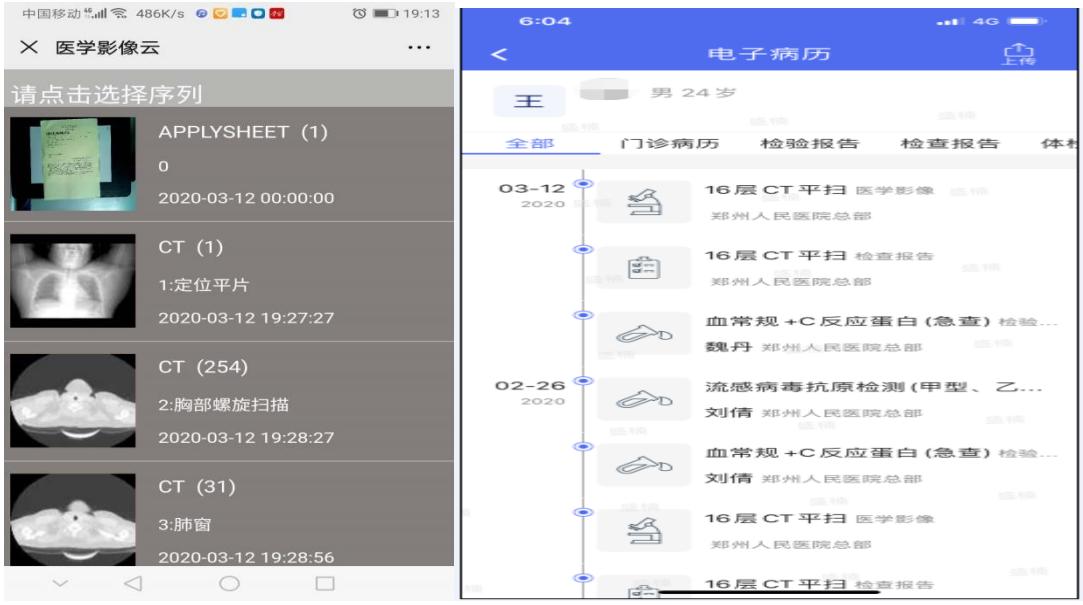
2、应用效果

通过云平台打破了就诊患者到现场排队挂号就诊，缴费，检查、等待取报告，复诊等，尤其在疫情期间缩短了时间、地域限制，大大助力了患者和医生在线上诊疗，为实现线上线下一体化提供了支持。

（1）患者到不同医院都需要重新建立个人信息档案患者缺少个人卫生信息管理和历史就诊记录；

（2）打破了各个医院之间信息不共享；患者拿一个医院的影像报告到另一家医院由于就诊信息不全，检查结果不互认，需要重新检查。耗时，费力，费钱；

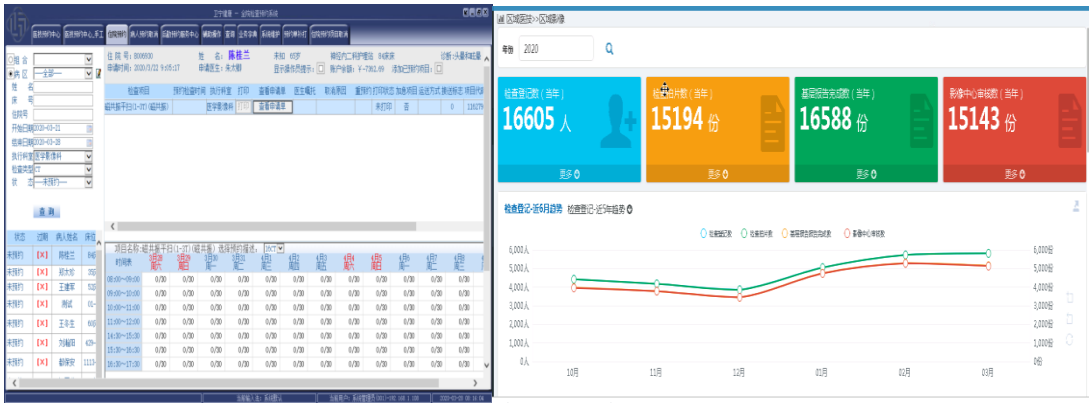
（3）患者不用在担心影像丢失或者刻录光盘或者拷贝图像困难，根据影像可以查看当时的申请单



（4）医生在不同位点通过移动移动终端（手机、pad，pc 机）远程辅助诊断；

（5）医联体间的关联多了，医联体间可以看到优势医院的预约资源，直接给患者预约检查，增加

体验不好。



3、展望

以医联体为单位，整合影像数据和诊疗信息将其延伸移动终端，助力互联网医院建设。依托医疗信息云平台连接医联体资源，共享影像数据，为医联体建设提供切实的服务。

通过优化医联体内的就诊流程，达到省时、省力、省钱。实现影像数据灵活使用，让患者更加便捷享受高质量的影像诊断服务，助力互联网医院线上线下一体化。

影像云平台有助于建立全民电子健康档案，实现流程创新，完善监管体系，推进区域分诊、支持医疗科研水平提升。

PU-014

基于文献计量的中国临床医学院士科研产出行为分析

杨辰毓妍、安新颖、蔡荣
中国医学科学院医学信息研究所

目的 随着科技发展，中国医学事业为世界医学发展作出巨大贡献，研究科研人才产出行为对于人才成长模式研究以及相关政策的制定都有着重要意义。此前关于人才科研行为的研究多从奖项和教职晋升角度出发，一些中文论文针对院士的研究又过于宏观，故而本研究聚焦中国临床学科院士，通过对其科研产出行为的梳理，以及院士评选对研究者科研产出行为影响的分析，为中国临床学科人才研究以及科研管理提供一定参考。

方法 本研究选取 32 名现阶段仍然在医院进行临床工作的院士，通过对其简历进行梳理，聚焦其 SCI 论文发表行为，使用多种统计学分析方法探索其科研产出行为规律，使用线性回归方法探索 SCI 论文发表相关指标对于院士评选是否产生较大影响，也使用 Wilcoxon signed-rank test 分析了院士评选对于研究者产出行为又造成何种影响。

结果 中国临床学界精英人才在地域上呈现集中趋势，他们发表被 SCI 收录的英文论文时间较晚，多数开始于 2004 年。多数人在生理年龄 50 岁左右科研论文产出数量最多，此后逐渐下滑。该数据集中论文相关的计量指标不能用来预测 32 位院士的评选，即不能验证 SCI 论文发表行为对院士

评选有非常明确的影响。这些被试者在被评选为院士之后，除了一区文章占比基本上保持不变，其他发表 SCI 论文相关的各个指标均有显著提升。

结论 中国临床医学院士群体开始发表 SCI 论文时间较晚，但其后总体产量非常高，在本研究选取的数据集上，SCI 论文发表并没有对院士评选产生显著影响。但成功当选院士后研究者的产出指标总体上涨，国际合作机会增加，作为主要作者参与论文发表数量也有所上升，但其论文总体质量提升并不显著。这种提升有可能是研究者自身自然发展结果，但也不排除是马太效应即资源倾斜的体现。

1. 引言

随着科技发展，中国医学事业为世界医学发展作出巨大贡献。目前中国对于科研人员有许多评选项目和荣誉称号，其中中国科学院和中国工程院院士是中国科学界个人最高的荣誉称号，一定程度上可以代表该学科在中国的尖端水平。

此前关于人才科研行为的研究有从奖项出发的，比如 R. Bjørk 研究了不同学科诺贝尔奖得主取得获奖成果的产出年龄以及获奖成果引用分析[1]，Jones and Weinberg 对诺贝尔奖得主完成获奖作品的年龄也进行了研究[2]。也有从教职晋升角度出发， James W. H. Sonne 研究了美国西部物理理疗方向学者获得终身教职的科研产出因素[3]。一些中文论文针对院士进行了研究，比如刘俊婉对中国科学院院士进行科研生产力和影响力的分析，也对院士获得称号前后的科研产出进行了对比[4]。李文丽在对中国所有生物医学领域的两院院士进行了科研产出和论文合作情况的一般性描述分析[5]。但是这些研究过于宏观，选取的院士来自多个学科。

虽然纯粹定量评价的结果被不当使用，学界对于单纯文献计量评价的使用提出许多质疑，但是之前多年的实践中，尤其是医学领域，例如 SCI 发文数、发表文献期刊的影响因子等还是研究者评价的重要标准之一，并且这些定量的评价标准还是具有较高的客观性，能够一定程度上公正的衡量研究者在科研活动上的水平。

考虑到中国院士的选举要从具有较深资历和名誉的科学家中提名投票，那么在当选这一最高荣誉称号后，院士势必会拥有更加丰富的科研资源。邱均平在研究中指出职称评定的先后顺序会对潜力相同科学家造成影响，较早获得荣誉的科学家会因为累积效应取得更高的科研成果[6]。尽管目前许多研究者开展了针对科学家的科学成就在社会年龄上分布的分析，比如早在 1979 年，赵红洲就对多国科学家进行分析，发现杰出科学家多在 25-45 岁之间作出突出成就[7]， Akenes 对挪威大学科研人员的学术产出和年龄之间的关系进行了研究，得出研究人员的研究水平随着年龄增长，但是达到一定峰值后开始下滑[8]。考虑到院士们获得称号的年龄大部分超出了学界公认的科学家作

出突出成就的年龄段，那么院士在一个较大的年龄获得称号后是否会迅速累积较大的科研资源，院士称号的获得对他们科研产出是否造成较大影响也是需要进行探索的。

2.数据选取和预处理

研究对象的基本信息来自于官方网站上提供的个人简历，考虑到院士实际年龄和中国学界发表 sci 论文成为主流趋势的时间，选取中国科学院生命科学和医学部和中国工程院医药卫生学部 2011、2013、2015 和 2017 四批院士，从中剔除研究方向为基础医学等方向的对象，选取 32 名在现阶段依然在医院工作的院士。通过人名-署名机构检索的方式在 Web of Science 数据库中进行检索。检索时间为 2019 年 1 月 19 日，检索范围均为 1900 至今，设置检索核心合集为 Science Citation Index Expanded (SCI-EXPANDED)和 Social Sciences Citation Index (SSCI)。本研究采用正规计数法计算研究者发文数量，一篇论文的每一位作者都算发表一篇论文。影响因子采集自 Thomson Reuters Journal Citation Reports (JCR)，当年发表文章的影响因子采集自上一年度该期刊影响因子。

3.数据分析

本研究选取的 32 名院士中 13 名为中国科学院院士，19 名为中国工程院院士。他们从事外科、心血管病、妇产学、肿瘤学等方向临床工作和研究。这 32 名院士现今所属机构分别位于北京（16），上海（9），湖北省（2），江苏省（1），山东省（1），辽宁省（1），重庆(1),河北省（1）。

表 1 出生年份分布情况

出生年份	1940-1944	1945-1949	1950-1954	1955-1959	1960-1964	1965-1968
数量	2	1	7	10	11	1

我们依次采集了 32 位院士被 sci 收录的论文相关数据，发现最早一位发表 SCI 论文的院士是在 2004 年发表第一篇 SCI 论文，当时他 46 岁，其余大多数院士都是在 2006 年开始进行 SCI 论文的发表。我们计算了每篇论文发表时其作者的生理年龄，结果如图 1 所示。

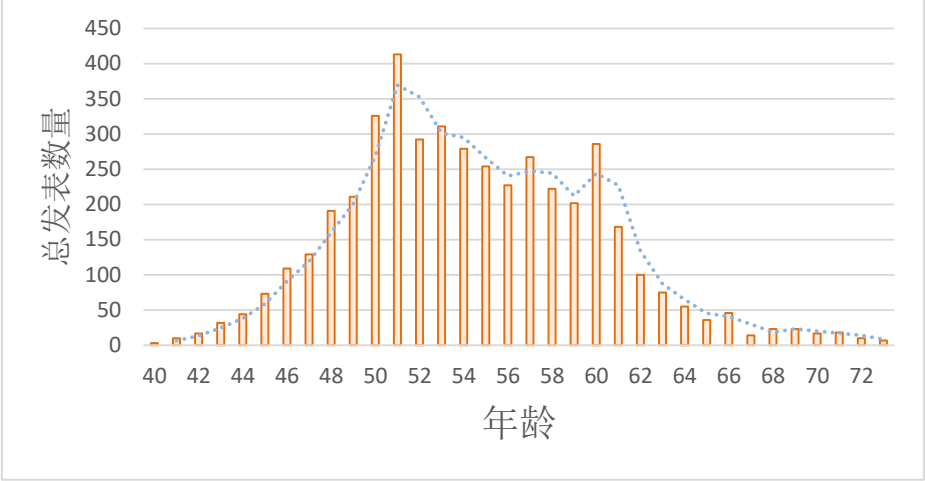


图 1　发表 SCI 论文数量和年龄关系

表 2.　32 名研究对象被评选为院士前发表 SCI 论文相关情况

学者生产力因素	平均值	最大值	最小值	中位数	标准差
年龄	57.59	71	49	57	5.49
总发表数	88.16	329	1	56	82.54
署名一作/通讯发表数	35.50	95	0	25	29.21
总被引频次	502.16	1553.30	42.77	300.44	423.72
平均被引频次	7.60	42.77	3.75	6.16	6.67
总影响因子	249.31	774.93	14.26	148.31	212.15
平均影响因子	3.53	14.26	1.87	3.06	2.17
一区发表数量	43.97	187	1	23	46.54
一区发表数量占比	0.50	1	0.16	0.49	0.19

通过 Spearman 相关分析来探究对象成功当选院士的年龄以及上述发表文章相关指标间是否存在相关关系（表 3）。总体来看，当选院士的年龄和发表 SCI 论文相关指标均没有明显相关关系，当选年龄和这些指标整体之间为负相关。总影响因子和发表论文总数呈明显正相关关系（p= .950**）的同时，影响因子均值与发表论文总数呈较为明显的负相关关系（p= -.486**）。署名为第一作者或通讯作者发表文章数量和总影响因子（p= .860**）以及一区文章数量（p= .826**）间呈明显正相关关系。

表 3.　学者生产力因素 Spearman 相关关系分析结果

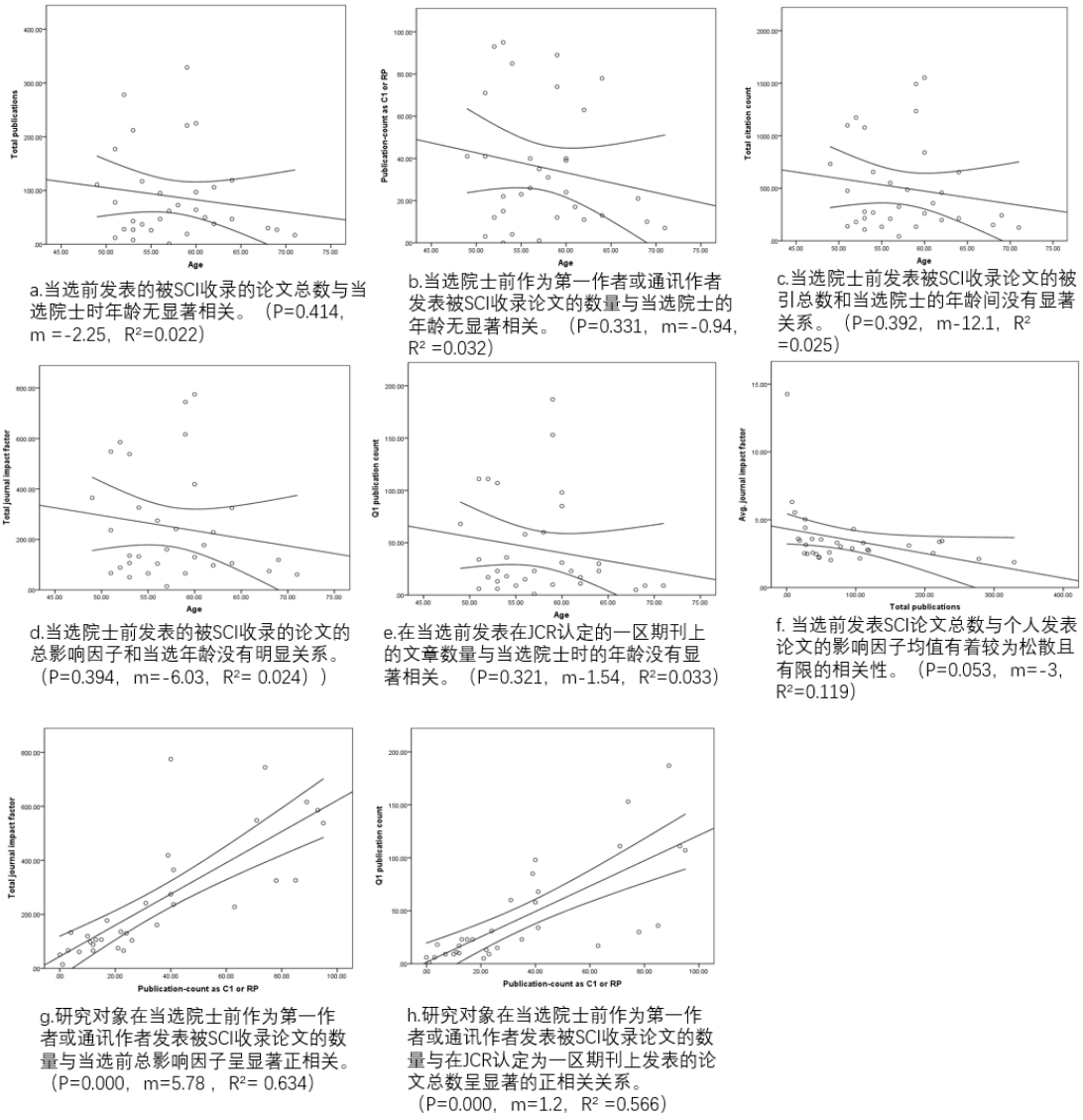
	年龄	总发表数	署名一作/通讯	总被引频次	平均被引频次	总影响因子	平均影响因子	一区发表数量	一区发表数量

			发表数						占比
年龄	1.000	-.063	-.154	-.129	-.087	-.129	-.088	-.208	-.327
总发表数	-.063	1.000	.923**	.950**	-.494**	.950**	-.486**	.929**	-.070
署名一作/通讯发表数	-.154	.923**	1.000	.860**	-.537**	.860**	-.531**	.826**	-.154
总被引频次	-.129	.950**	.860**	1.000	-.253	1.000**	-.245	.934**	.042
平均被引频次	-.087	-.494**	-.537**	-.253	1.000	-.253	1.000**	-.350*	.456**
总影响因子	-.129	.950**	.860**	1.000**	-.253	1.000	-.245	.934**	.042
平均影响因子	-.088	-.486**	-.531**	-.245	1.000**	-.245	1.000	-.342	.460**
一区发表数量	-.208	.929**	.826**	.934**	-.350*	.934**	-.342	1.000	.223
一区发表数量占比	-.327	-.070	-.154	.042	.456**	.042	.460**	.223	1.000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

通过线性回归来进一步确定这些 SCI 发文相关指标和院士当选间是否存在一定关联关系，置信区间为 95%。总体而言，这些和论文发表相关的指标与院士当选年龄没有显著或密切的相关性。在这一组线性分析中，研究对象当选院士的年龄与当选前发表的被 SCI 收录的论文总数没有相关性（P=0.414， R^2=0.022），与在一区期刊上发表的论文数量也无关（P=0.331， R^2=0.032）。当选院士的年龄与发表文章总被引次数、总影响因子和发表在一区期刊文章数量的相关系数分别为 0.392、0.394、0.321，均无相关性存在。除了当选年龄外，其他在 Spearman 相关分析中较为突出的指标也做了相关分析。当选前发表的被 SCI 收录论文总数与影响因子均值间有着较为松散且有限的相关性。（P=0.053，R^2=0.119），但作为第一作者或通讯作者发表被 SCI 收录论文的数量与当选前总影响因子呈显著正相关关系（P=0.000， R^2= 0.634），与 JCR 认定为一区期刊上发表的论文总数也呈显著的正相关关系。（P=0.000， R^2=0.566）



a.当选前发表的被SCI收录的论文总数与当选院士时年龄无显著相关。（P=0.414，m =-2.25，R^2=0.022）

b.当选院士前作为第一作者或通讯作者发表被SCI收录论文的数量与当选院士的年龄无显著相关。（P=0.331，m=-0.94，R^2 =0.032）

c.当选院士前发表被SCI收录论文的被引总数和当选院士的年龄间没有显著关系。（P=0.392，m-12.1，R^2 =0.025）

d.当选院士前发表的被SCI收录的论文的总影响因子和当选年龄没有明显关系。（P=0.394，m=-6.03，R^2= 0.024））

e.在当选前发表在JCR认定的一区期刊上的文章数量与当选院士时的年龄没有显著相关。（P=0.321，m-1.54，R^2=0.033）

f. 当选前发表SCI论文总数与个人发表论文的影响因子均值有着较为松散且有限的相关性。（P=0.053，m=-3，R^2=0.119）

g.研究对象在当选院士前作为第一作者或通讯作者发表被SCI收录论文的数量与当选前总影响因子呈显著正相关。（P=0.000，m=5.78，R^2= 0.634）

h.研究对象在当选院士前作为第一作者或通讯作者发表被SCI收录论文的数量与在JCR认定为一区期刊上发表的论文总数呈显著的正相关关系。（P=0.000，m=1.2，R^2 =0.566）

图 2　各 SCI 发文指标之间相关性分析

最后，院士这一中国科学界最高荣誉称号的评选对于研究者研究产出行为有无影响呢？本研究选取院士当选前后 3 年数据进行分析。国际合作是一位研究者合作范围扩大、影响力增高的重要表现形式，故而采集每位院士评选前后 3 年发表论文的国外合作者署名数作为衡量研究者合作范围的指标。如表 4 和图 3 所示，除了发表一区文章占比外，其余指标都有着明显增长。就中位数而言，除了发表一区文章数量占比较当选前下降，以及篇均影响因子保持不变，其他文献相关指标在当选后都有一定程度的增长。标准差的变化说明数据离散程度增加。使用 Wilcoxon signed-rank test 对指标在当选前后变化进行检验，国际合作发文署名数量（p=0.05）前后有一定程度的变化，一区文章数量占比（p=0.446）无显著变化。发表被 SCI 收录文章总数（p=0.020），署名为第一作者或通讯作者发表文章数量（p=0.002），总影响因子（p=0.009）和

篇均影响因子（p=0.014）在当选前后均有显著变化，从 Fig.10 中可以明显观察到当选后这些指标均值和中位数的增长。

表 4. 院士当选前后发文行为统计

	平均值		中位数		标准差		P 值
	当选前	当选后	当选前	当选后	当选前	当选后	
总发表数	44.34	53.72	30.50	38.50	34.49	44.82	0.020
署名一作/通讯发表数	9.16	16.41	5.50	11.50	8.37	14.98	0.002
国际合作发文次数	24.00	40.38	7.00	10.00	65.55	86.95	0.050
总影响因子	84.06	109.35	52.44	77.91	96.60	127.85	0.009
平均影响因子	1.86	2.16	1.25	1.25	1.42	1.98	0.014
一区发表数量	24.38	29.28	13.50	22.50	22.26	25.49	0.011
一区发表数量占比	0.55	0.55	0.55	0.52	0.20	0.20	0.446

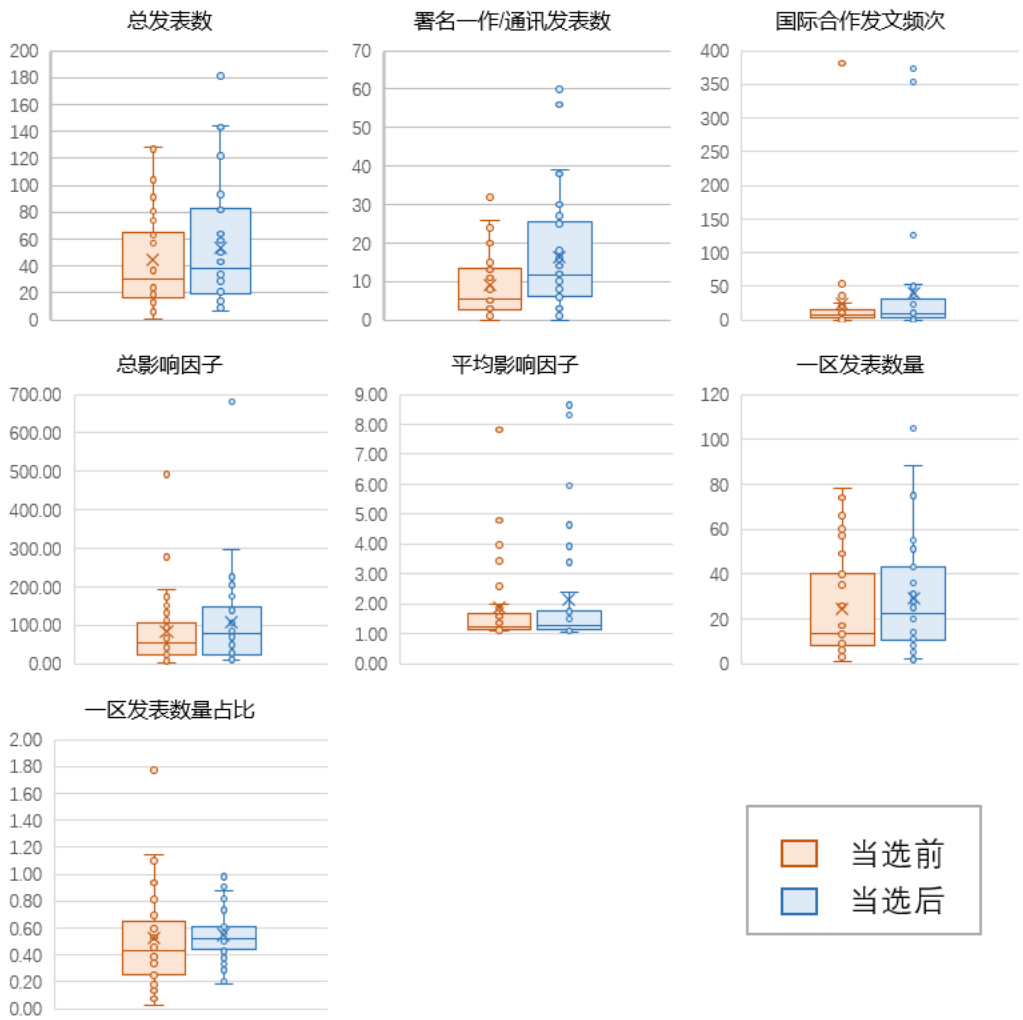


图 2　32 名对象在被评选为院士前 3 年和后 3 年间发表被 SCI 收录论文的相关指标

4. 讨论与结果

上述分析表明，中国临床学界精英人才在地域上呈现集中趋势，他们年龄分布较为集中，多数出生在 1955-1965 年之间，他们发表被 SCI 收录的英文论文时间较晚，多数开始于 2004 年。虽然在生理年龄 50 岁左右科研论文产出数量最多，此后逐渐下滑。

该数据集中论文相关的计量指标不能用来预测 32 位院士的评选，即不能验证 SCI 论文发表行为对院士评选有非常明确的影响，但是这些指标之间的关系可以为我们揭示一些情况。在被评选为院士之前，这些学者在学界已经具有很高的声誉。通过对他们发表 SCI 论文情况的梳理，可以了解到在被评选为院士之前发表 SCI 论文相关各类情况。学者发文总量越高，篇均影响因子有着较为明显的下降趋势，但作为论文主要贡献者发表文献越多的时候，总影响因子和一区文章的数

量都有着明显的上升。这一定程度上说明，当学者作为论文主要参与者时，论文的质量是有着较为明显的提升的，起码在发表期刊的影响力上是如此表现的。

中国近年对于临床工作者职称和奖励的判定上较为重视 SCI 论文，选取对象开始发表 SCI 论文来参与国际学术交流的时间较晚，但此后发文量飞速上升，这种行为的动机故而是因为学术交流需要，也符合中国学者参与国际学术研讨的潮流，但是否也与中国科研人员评价管理的制度有关呢？本研究选取的文献计量相关的指标，经过统计均与院士评选不存在明显关系。院士评选更像是整体评估后做一个定性的同行评议。

Wilcoxon signed-rank test 揭示了这些被试者在被评选为院士之后，除了一区文章占比基本上保持不变，其他发表 SCI 论文相关的各个指标均有显著提升。尤其是国际合作署名的次数，这说明院士的当选不仅为学者带来了国内的各种科研资源，其国际名誉也有着显著提高。但团队的壮大和资源的增长势必会对科研工作以及团队管理带来较大的挑战，院士在当选后三年间发表文献的数量和质量都有着非常大的提升，但是否会和权威的力量以及资源上不合理的偏斜有关呢？

该数据集得出的上述结论对于其他较早批次的院士不可以做简单的数据扩展和推理，但可以一定程度上反应较为年轻的中国临床医学界精英学者在文献发表上的情况。此项研究没有区分不同研究者所处机构和地域的经济发展等情况，也难以获得他们各年度具体获得的科研经费。这项研究的价值在于展示了中国临床医学界一线精英学者参与国际论文交流的产出行为，探索 SCI 论文产出和评选可能存在的关系，以及评选对于论文产出造成的影响。因此，本文涉及的描述性数据可以供科研管理者、对中国临床学界有兴趣的研究者以及院士自身进行参考。

PU-015

近 5 年国际医学信息学计量研究

单治易、安新颖
中国医学科学院

目的 通过对 WOS 数据库中医学信息学领域的文献进行多角度分析，以识别该领域的研究前沿和研究热点，以期望对医学信息学领域提供些许理论帮助。

方法 采用文献计量学的方法对 2015 至 2019 年国际医学信息学主题领域进行分析，分析内容包括年度发文量、研究领域、发文国家、重要发文机构研究热点。使用 CITESPACE 进行可视化，对高频关键词进行聚类分析。

结果 基于 2015—2019 年 WOS 数据库中医学信息主题领域 18103 篇文献，发现文献量持续增长，美国在该领域具有很强的主导地位。研究总结医学信息主题领域十大研究热点和排名前 25 的具有高强度变化的关键词。

结论 随着互联网+模式的快速兴起，云计算等相关技术的高速发展，国际上对医疗数据的挖掘、处理、分析需求不断增加，帮助实现以患者为中心的医疗服务，帮助医生提高诊断的精确度。医疗数据中含有患者的隐私信息，值得重点保护。我国医院必须不断提高信息化、数字化水平，才能跟上国际精准医学的新阶段。同时，移动医疗在慢性病、吸烟等方面的应用越来越重要，这得益于互联网技术的高速发展。相信在 5G 网络环境下，远程医疗技术会发展得更成熟。医学信息学作为医疗卫生行业和电子通信行业的桥梁，将起到更大的作用。

PU-016

基于 Altmetrics 分析“新冠肺炎”学术成果社会影响力

蔡荣、安新颖、杨辰毓妍
北京协和医学院/中国医学科学院医学信息研究所

目的 “新冠肺炎”的全球大流行，引发全球公共卫生危机，科研人员迅速反应，相关研究不断展开，研究成果也陆续产出。面对疫情社会公众急切需要相关健康信息来指导自己的防护，对研究成果也有一定的关注度。研究成果产生的社会影响力会帮助社会公众克服过度恐慌情绪，对病毒和疾病有更科学的认识，更加坚定战胜疫情的决心。通过分析社交媒体、新闻媒体和政策文件对研究成果的关注热点，得出不同数据来源对相关研究方向的异同点，以及关注点随时间变化趋势为主流媒体加强应急科普和舆论引导提供了方向。

方法 “新冠肺炎”的全球大流行，引发全球公共卫生危机，科研人员迅速反应，相关研究不断展开，研究成果也陆续产出。面对疫情社会公众急切需要相关健康信息来指导自己的防护，对研究成果也有一定的关注度。研究成果产生的社会影响力会帮助社会公众克服过度恐慌情绪，对病毒和疾病有更科学的认识，更加坚定战胜疫情的决心。通过分析社交媒体、新闻媒体和政策文件对研究成果的关注热点，得出不同数据来源对相关研究方向的异同点，以及关注点随时间变化趋势为主流媒体加强应急科普和舆论引导提供了方向。

结果 分析被引频次和 Altmetric Score 的相关性发现呈中度相关。通过分析社交媒体、新闻媒体和政策文件对研究成果的关注热点，发现不同数据来源关注主题数量和方向的异同点，社交媒体关注主题多样，政策文件关注的主题数量较少。疾病的爆发和流行是社交媒体、新闻媒体和政策媒体共同关注的内容。公众心理和精神问题受社交媒体和新闻媒体关注较多，政府文件这方面的提及却较少。病毒的致病机理与特性受新闻媒体的较多关注，社交媒体对此关注较少，可能由于公众缺乏专业知识、对内容理解不够。从分时段公众关注热点来看，公众的关注热点随时间推移向抑郁与心理健康主题转移。

结论 论文产生学术影响力和社会影响力存在一定差异。社交媒体用户更加关注自身安全和个人防护。主流媒体在进行科普时，可通过公众易理解的形式加入病毒基础知识这方面的内容，使公众能够增长知识，对病毒有更加清晰的认识。“新冠肺炎”的全球大流行给公众的心理健康造成一定的压力，政府及主流媒体应加强疾病知识科普，缓解公众的紧张情绪，加强对“新冠肺炎”患者及隔离人员的心理健康状况关注，医务人员以及青少年、孕妇、精神病患者等特殊人群的心理问题也应给予足够的重视，充分体现人文关怀。

引言

“新冠肺炎”疫情的发生威胁着全世界人民的生命健康，给全球公共卫生带来巨大压力和挑战。2020 年 2 月 11 日，世界卫生组织将“新型冠状病毒肺炎”命名为 COVID-19，国际病毒分类委员会将新型冠状病毒命名为 SARS-CoV-2。目前，国内抗击疫情的形式持续向好，但国外疫情仍然不容乐观。科研人员不断地“攻关”“克难”，取得一系列研究成果。社会公众在持续关注国内外“新冠肺炎”确诊人数之余，也关注着最新的科研成果以及研究进展。为及时了解舆论方向和对舆情进行监测，有学者对微博数据的分析，认为“新冠肺炎”这一话题参与度很高，政府、科研机构和科学家具有很高的信任度，要提高在应急事件中的科普能力[1]。

影响力又可以分为在学术环境中产生的影响力，和社交网络中产生的影响力。重大研究成果的取得是抗击疫情的一剂“强心针”，研究成果产生的社会影响力会帮助社会公众克服过度恐慌情绪，对病毒和疾病有更科学的认识，更加坚定战胜疫情的决心。虽然近年来我国居民的健康素养水平稳固上升，但传染病防治素养与科学健康观素养、健康信息素养以及慢性病防治素养等相比水平仍然较低[2]，公众对健康信息的理解程度有限[3]，因此，对公众进行突发传染病的应急科普是十分必要的。探明“新冠肺炎”相关研究成果产生的影响力，比较不同数据来源关注热点的异同，发现社会公众对“新冠肺炎”相关研究成果的关注热点，以及社会公众关注点的转移情况，为疾病科普提供方向支持，准确推送社会公众最感兴趣的研究进展和疾病知识。

替代计量指标基于网络社会媒体平台评价各种类型学术成果产生的社会影响力。替代计量指标是一个集成指标，数据来源多样、内容丰富，随着网络、社交媒体的发展，不断有新的指标产生。替代计量指标具有及时性、细粒度、社会化和波动性等特点，及时性[4]体现为学术成果通过网络社会媒体平台进行传播是即刻就能进行的，一些在线传播在论文发表后的几天内开始，几天之后就能够进行相关数据的统计，替代计量指标很大程度上克服了传统文献计量指标的时滞问题。细粒度体现在替代计量学是论文级别的计量。社会化体现在替代计量指标可以衡量学术成果产生的对学术界之外社会的影响[5]。波动性体现在传统文献引用中一篇论文的被引量只会增加或持平，替代计量学受删除博文、微博以及评论等因素影响，可能出现波动的情况。

1 数据来源及方法

1.1 数据来源

研究成果的形式多样，包括学术论文、专利、PPT、数据集、软件等，专利具有法定的保护期限，在较短时间很难获取，PPT、数据集和软件形式很难用简单的检索大批量获取。以学术论文为研究对象，替代计量数据以 Altmetric.com 为数据源。

1.2 数据获取

本文通过 PubMed 使用增补概念作为检索词[6]，检索式为(“COVID-19”[Supplementary Concept]) OR“severe acute respiratory syndrome coronavirus 2” [Supplementary Concept]，检索到论文 15 308 篇，利用论文的 PMID 在 Web of Science 中进行检索，获得与“新冠肺炎”主题有关的 SCI 论文，选择文献类型为“Article”的论文纳入，“Review”、“Editorial Material”、“Letter”等文献类型不纳入，纳入 3 647 篇论文作为数据集，检索时间为 2020 年 8 月 3 日。Altmetric.com 数据通过论文的 DOI 号使用 R 语言程序进行获取，最终得到拥有 Altmetric Score 的论文 3 019 篇，Altmetrics 数据获得的时间为 2020 年 8 月 3 日。

1.3 研究主题可视化

VOSviewer 是知识图谱软件，可以进行共现网络分析与可视化，使用 VOSviewer 进行关键词共现聚类，实现关注主题可视化。

2 社会影响力分析

2.1 总体概况

所有论文 Altmetric Score 平均值为 281.17，中位数为 12.85，Altmetric Score 大于 100 的论文占比 21.8%，Altmetric Score 小于 200 的论文占比 15%，Altmetric Score 大于 1 000 的论文占比 5.6%，Altmetric Score 大于 5 000 的论文占比 1.2%，有 0.4%的论文 AS 大于 10 000。约有 38.03%的论文在 Altmetric.com 纳入所有研究前 5%，可以看出“新冠肺炎”相关盐焗与其他研究相比获得关注度更高。Altmetric Score 分布具有低分值段论文密度高，高分值段论文密度低的特点。对所有论文被引频次和 Altmetric Score 做相关分析，被引频次和 Altmetric Score 均不符合正态分布，因此采用 spearman 秩相关分析，相关系数为 0.506，相关性在 0.01 级别上显著，相关系数在 0.4-0.6 之间，被引频次和 Altmetric Score 呈中等程度相关。

2.2 分数据来源分析

本文将数据来源划分为社交媒体、新闻媒体及政策文件三大类型。社交媒体包括 Twitter、Facebook、Blogs 和 YouTube。新闻媒体包括主流权威的英文和非英文媒体，Altmetric.com 的范围是包括 BBC、New York Times、光明网、人民网的 2 900 多家新闻媒体。政策文件包括政府指南、报告和白皮书等。论文被各数据来源的提及情况如表 1 所示。

表 1 论文被各数据来源提及情况

数据来源	提及论文数	提及次数平均值	提及次数最大值
Twitter	2 970	301.24	32 856
Facebook	932	1.13	60
Blogs	964	1.42	74

YouTube	179	0.11	18
News	1 061	12.06	637
政策文件	317	0.17	8

2.2.1 社交媒体

有 6 篇论文被 Twitter 提及超过 2 万次，分别是 32 480 次、32 856 次、23 836 次，20 925 次、27 431 次和 21 170 次。与 Twitter 相比，Facebook 提及论文数和提及次数少一些，论文最高被 Facebook 提及次数为 60 次，这篇论文也是 Twitter 提及最高的论文，其发表在《NATURE MEDICINE》，研究表明外科口罩可以防止有症状的个体传播人冠状病毒和流感病毒[7]，公众对这篇论文的极高关注度，体现出公众对自身安全的防护以降低自身被感染风险。Twitter 和 Facebook 大约有三分之二的高关注论文是相同的。博客相比 Twitter 和 Facebook 篇幅更长，在对论文进行提及时付出时间成本也要更多，提及论文数也较少，最高被提及次数为 74 次。论文也以视频的形式被提及，被提及次数是以发布视频的用户数，并非发布视频的数量也以视频的形式被提及，视频形式相比传统文字图片模式更加直观，有 179 篇论文被 YouTube 提及，最多被提及 18 次。社交媒体使用用户大多为普通网民，根据不同来源高关注论文关键词主题聚类的结果看（如图 1），在关注点上社交媒体各数据来源具有极大的相似性，主要是疾病的爆发与传播、疾病诊断、临床表现及相关指南制定、药物研发和公众心理和精神问题。

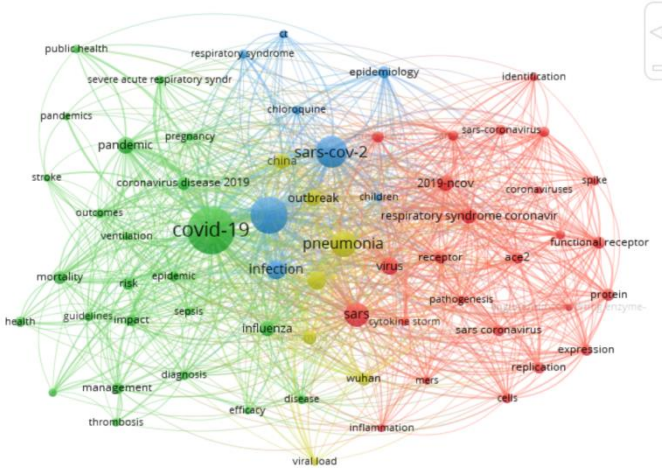


图 1 Twitter 高关注论文关键词主题聚类

2.2.2 新闻媒体

论文被新闻媒体提及的比例也达到了 35.14%，有 5 篇论文被新闻媒体提及的次数超过 500 次。新闻媒体更注重信息的时效性与准确性。根据关键词聚类形成的主题（如图 2）可以看出疾病的流行与爆发、病毒的致病机理与特性、疾病临床特征、公众心理和精神问题。最受新闻媒体关注的论文，也是研究论文集合中唯一被新闻媒体提及超过 600 次的论文，是一篇关于疾病潜伏期的研

究，估计 COVID-19 潜伏期的长度，并描述其公共卫生影响，研究得出 COVID-19 的中位潜伏期大约 5 天，类似于 SARS，结果支持对可能接触 SARS-CoV-2 的人进行隔离或积极监测的建议，在极端情况下延长监测时间也合理[8]。

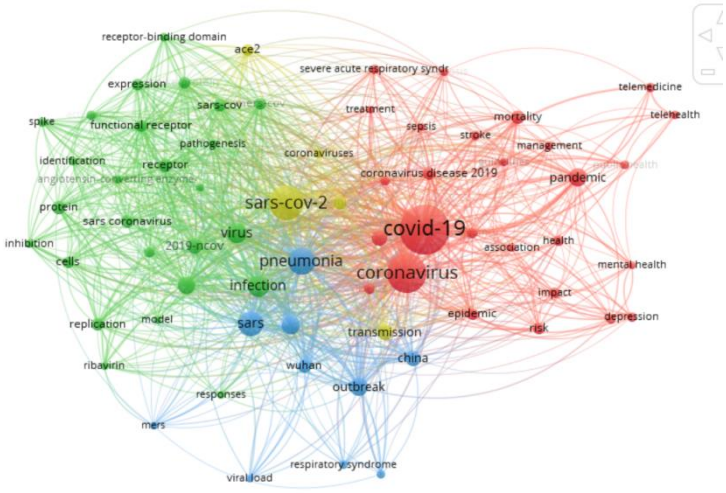


图 2 新闻媒体关注论文关键词主题聚类

2.2.3 政策文件

政策文件是政府机构和相关组织行动的“风向标”，政策文件的出台对疫情防控起到至关总要的作用，一篇论文被政策文件引用也体现其具有较高的价值。论文被世界卫生组织（World Health Organization，WHO）、英国国家健康与临床卓越研究所(National Institute for Health and Care Excellence,NICE)、美国国家经济研究局(The National Bureau of Economic Research，NBER)、联合国粮食及农业组织（Food and Agriculture Organization of the United Nations，FAO）等组织提及。对被政策文件提及论文进行主题聚类（如图 3）以及较高提及次数论文的分析发现，政策文件更加关注的方向是疾病流行与传播、对疾病严重程度的预估[9]，由此可以看出政策文件关点更为宏观，关注疫情严重程度对经济、农业及公共卫生的影响。

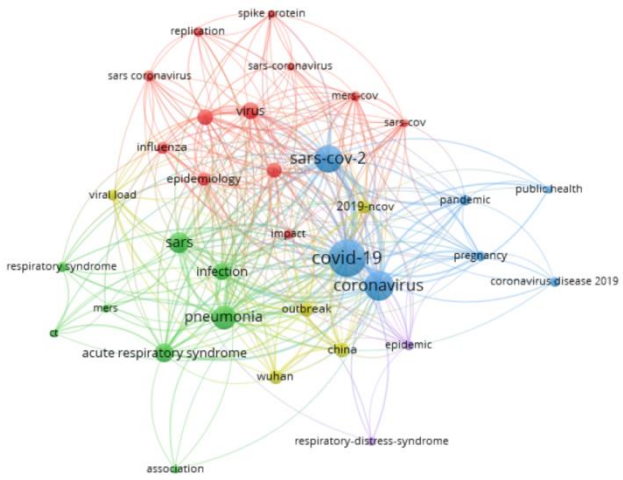


图 3 政策文件关注论文关键词主题聚类

2.3 分时段分析

Altmetric.com 提供论文的 Altmetric Score 历史信息，可以了解论文在过去某一段时间的得分情况。将时间分为四个段分别为一周之内、一个月至一周间、三个月至一个月间，以及三个月之前。各个时间段 Altmetric Score 较高的三分之一的论文为高关注论文。在四个时间段均获得高关注的论文有 209 篇，随着新成果的不断产出以及疫情形势的变化，既关注最开始的“开创性”的研究，也关注最新的成果，社会公众在获取相关信息时有时效性的要求。对各个时间段进行关键词主题聚类（如图 4、图 5）发现，在疫情发生初期，疾病的爆发与传播、病毒检测、疾病的流行病学、疾病死亡率以及疾病防控是较为受到关注的主题。随着国内外疫情形势的不断变化，社会公众关注主题也发生改变，除了一直受到关注的主题，也开始关注药物研发、医务人员、普通大众以及特殊人群（青少年、孕妇、精神病患者）的精神心理问题[10,11]等几个方面。

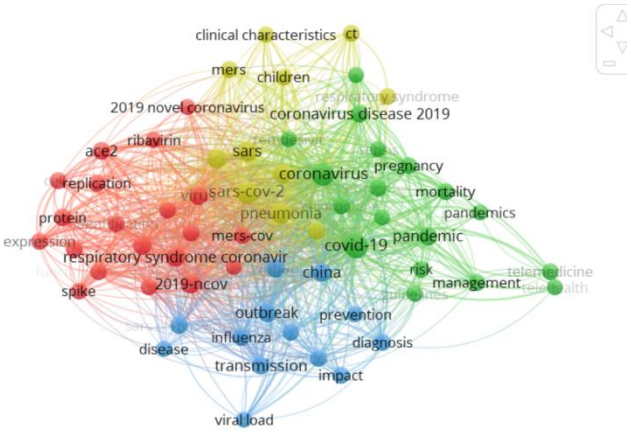


图 4 Altmetric Score 获取三个月之前高关注论文关键词主题聚类

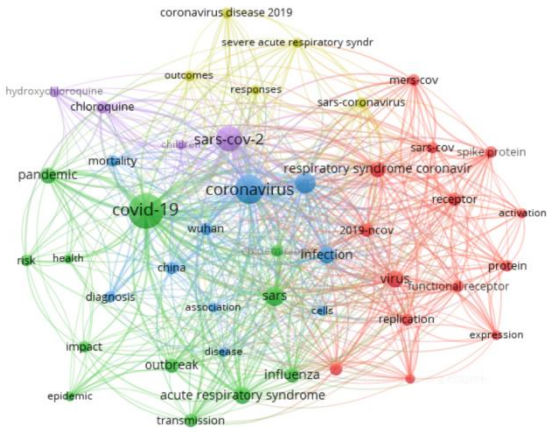


图 5 Altmetric Score 获取一周之内高关注论文关键词主题聚类

3 结语

被引频次和 Altmetric Score 呈中等程度相关，被引频次越高不代表 Altmetric Score 越高，论文产生学术影响力和社会影响力存在一定差异。社交媒体是论文覆盖率最高的数据源，其次为新闻媒体和政策文件。社交媒体、新闻媒体和政策媒体由于背后主体的不同，在关注点上也有异

同，社交媒体关注的主题多样，政策文件关注的主题数量较少。疾病的爆发和流行是社交媒体、新闻媒体和政策媒体共同关注的内容。社交媒体关注点更多从个人角度出发，与自身的防护与安全相关，新闻媒体和政策文件更多从宏观角度出发，新闻媒体发挥健康科普职能，政策文件关注疫情发生对经济社会造成的影响。

病毒的致病机理与特性受新闻媒体的较多关注，社交媒体对此关注较少，可能由于公众缺乏专业知识、对内容理解不够，在进行科普时，可以通过公众容易理解的形式加入这方面的内容，使得公众能够增加这方面知识，对病毒有更加清晰的认识。公众心理和精神问题受社交媒体和新闻媒体关注较多，政府文件这方面的提及却较少，从分时段公众关注热点来看，公众的关注热点随时间向抑郁与心理健康发生转移，“新冠肺炎”的全球大流行给公众的心理健康造成一定的压力，政府及主流媒体应加强疾病知识科普，缓解公众的紧张情绪，加强对患者及隔离人员的心理状况关注，医务人员的心理问题也应给予足够的重视，体现人文关怀。

PU-017

基于 FEARS 的屈螺酮炔雌醇药物不良事件信号挖掘研究

詹阳洋[1]、胥学莹[1]、陈力[2]、段思羽[3]
1. 四川大学华西第二医院成华妇女儿童医院
2. 四川大学华西第二医院 3. 成都市龙泉驿区第一人民医院

目的 挖掘屈螺酮炔雌醇不良事件信号并进行分析，了解其不良事件发生特点，提示临床用药可能存在的安全性风险，为临床及患者安全使用提供参考。

方法 采用报告比值比（reporting odds ratio,ROR）法和比例报告比值（proportional reporting ratio，PRR）法，对美国 FDA 不良事件报告系统（FDA Adverse Event Reporting System, FAERS）数据库中 2015 年至 2019 年共 5 年共 20 个季度屈螺酮炔雌醇导致的不良事件报告进行挖掘与分析。

结果 获得以屈螺酮炔雌醇为首要怀疑药品的不良事件涉及 2412 个患者，合计报告 7243 例。使用 ROR 和 PRR 法共得到 279 个信号，不良事件累及 25 个系统。

结论 屈螺酮炔雌醇的不良事件信号主要集中在各类神经系统、生殖系统及乳腺、精神系统血管与淋巴管系统。其中血栓相关不良事件占信号强度排名前 50 不良事件的 58.29%。挖掘出 104 个新的不良事件，集中在各类神经系统 69 个。临床应密切关注用药所带来的各类神经系统以及血栓的不良事件。

复方口服避孕药是目前全球范围广泛使用的高效避孕方法之一，是含有低剂量雌激素和孕激素（与女性体内天然的雌激素和孕激素相似）的复合甾体激素制剂，正确使用，复方口服避孕药的避孕有效率可达99%以上[1]。2017年中国有统计的人工流产数量为962万，约占全世界5500万人工

流产总数的17.4%[2]。人工流产会带来诸多近、远期并发症，如人工流产综合反应、子宫穿孔、出血、继发不孕、月经失调、子宫内膜异位症等，危害女性生殖健康[3]。

屈螺酮炔雌醇片是一种新型低剂量的单相口服避孕药，是炔雌醇和屈螺酮的复方制剂，2001年5月得到FDA批准，成为被广泛应用的激素类避孕药之一[4]。除高效避孕外，雌-孕激素避孕药还可用于治疗痤疮、多毛症、月经过多、痛经、盆腔痛和经前期综合征。雌-孕激素避孕药还能降低卵巢癌和子宫内膜癌的风险[5]。

由于短效复方口服避孕药在全世界普遍推广使用，大量人群使用的情况下对其不良反应的监测也是十分必要的，现国内外对于短效复方口服避孕药安全性研究大多为血栓栓塞的风险[6-10]，较少提及其他的不良事件。因此利于数据挖掘寻找短效复方口服避孕药潜在的药品不良事件信号也是十分有意义的。

目前国内外常用的药品不良反应信号检测方法主要是非均衡性测量法，又称比例失衡法（measureof disproportionality）[11]。其大致原理是在一个包含所有报告的数据库中，当某特定药物事件组合（drugevent combination，DEC）明显高于整个数据库的背景频率，并且达到了一定标准，就认为产生了一个信号[12]。

1.资料与方法

1.1数据来源

美国不良事件报告系统(FDA Adverse Events Reporting System, FAERS)是美国收集上市后药品不良反应的自发呈报系统，是药品安全性研究的重要数据来源。本研究数据基于 FAERS 数据库中 2015 年至 2019 年共 5 年 20 个季度的不良良事件报告。研究数据以 FDA 官网批准以及维基百科上所有关于屈螺酮炔雌醇的商品名及通用名为关键词，从 FAERS 数据库下载对应季度的 ASCII 数据包，导入 MY SQL（5.7 版）数据库，经过 demo 表去重，核对 drugname 项符合“屈螺酮炔雌醇”，共检索到 11342 例药物不良事件(Adverse Events，ADE)报告，报告数据包括患者基本信息、药品信息、不良事件名称、结果、日期和报告国家等内容。**1.2 挖掘方法**

本研究同时应用比值比法（ROR）和比例报告比值法（PRR），该方法基于四格表（表 1），通过比较目标药物的不良事件报告例数和其他所要药物的报告例数（背景数据）来判断是否“失衡”，挖掘潜在的 ADE 信号。

1.3 数据标准化

本文采用国际医学科学组织理事会《国际医学用语词典》（Medical Dictionary for Regulatory Activities，MedDRA）中药物不良反应术语集的首选系统器官分类（System Organ Class，SOC）和首选术语（Preferred Term，PT）。本研究下载 MedDRA 23.0 版，用于首选语(PT)、高位语(High Level Term, HLT)、高位组语(High Level Group Term, HLGT)、系统器官分类(SOC)编码和语言汉化[13]。

1.4 数据的筛选与计算

对标准化后的数据中药品不良事件报告不完全、信息不确定报告及其他包含异常信息报告进行排除。筛选出报告数≥3 的 PT，按照比例失衡法四格表（表 1），计算出相应 PT 值及其对应的 ROR 值、PRR 值和相应的 95%CI 下限（表 2）并统计信号数，再进一步剔除 ROR 与 PRR 比值的 95%CI≤1 的报告[10]。

表 1 比例失衡法四格表

Tab 1 Four grid table of proportional imbalance method

药品种类	目标不良反应报告数	其他不良反应报告数	合计
目标药物	a	b	a+b
其他药物	c	d	c+d
合计	a+c	b+d	a+b+c+d

表 2 ROR 法与 PRR 法公式及阈值

Tab 2 Formulas and thresholds of ROR and PRR methods

方法	公式	阈值
ROR	$ROR = \frac{(a/c)}{(b/d)} = \frac{ad}{bc}$ $SE（\ln ROR） = \sqrt{（\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}）}$ $95\%Cl = e^{ln(ROR)\pm 1.96\sqrt{（\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}）}}$	a≥3； ROR 的 95%CI 下限>1，则提示生成一个信号
PRR	$PRR = \frac{a/（a+b）}{c/（c+d）}$ $SE（\ln PRR） = \sqrt{（\frac{1}{a}-\frac{1}{a+b}+\frac{1}{c}-\frac{1}{c+d}）}$ $95\%Cl = e^{ln(PRR)\pm 1.96\sqrt{（\frac{1}{a}-\frac{1}{a+b}+\frac{1}{c}-\frac{1}{c+d}）}}$	a≥3； PRR 的 95%CI 下限>1，则提示生成一个信号

2 结果

2.1 药品不良事件的基本信息

2015 年至 2019 年，由屈螺酮炔雌醇导致的不良事件报告数共计 11342 例，筛选后共涉及 2412 个患者，合计 7243 个不良事件。报告中有 6 例男性患者，占总人数比例为 0.25%。在已知年龄的患者中，平均年龄为 31 岁。2015 年至 2019 年分别报告 958、516、319、332、285 个，呈下降趋势。报告国家前五位分别是美国、日本、巴西、俄罗斯、中国，美国报告数占 53.23%，中国占 3.19%。报告者大多数为消费者，占比为 54.44%，医务人员上报占比总和为 33.08%。报告基本信息见表 3。

表 3 2015-2019 年屈螺酮炔雌醇相关不良事件报告基本信息

Tab 3 Basic information of adverse event reports related to drospirenone and ethinyl estradiol from 2015 to 2019

信息	分类	患者个数/N	构成比/%
总计	/	2412	100
性别	男性	6	0.25
	女性	2383	98.80
	不清楚	23	0.95
年龄	＜18 岁	117	4.85
	18-35 岁	1104	45.77
	≥36 岁	574	23.80
	不清楚	617	25.58
年份（2015-2019）	2015	958	36.64
	2016	516	16.92
	2017	319	20.35
	2018	332	14.52
	2019	285	11.56
报告国家（前五位）	美国	1284	53.23
	日本	311	12.89
	巴西	88	3.65
	俄罗斯	88	3.65
	中国	77	3.19
报告者	消费者	1313	54.44
	其他卫生专业人员	352	14.59
	医师	311	12.89
	律师	267	11.07
	药师	135	5.60
	不清楚	34	1.41

2.2 药品不良事件信号挖掘结果

数据挖掘技术共挖掘出 1479 种 PT，11342 例报告，使用 ROR 和 PRR 法进行计算，合并同名并删除非 ADE 的 PT，再以 a≥3，95%CI（ROR）下限>1 为筛选条件，共得到 279 个信号，合计 7243 例药品不良事件报告，，屈螺酮炔雌醇相关药品不良事件报告信号强度居前 50 位的信号详见表 4。

表 4 屈螺酮炔雌醇信号强度前 50 位排序

Tab 4 The top 50 rankings of drospirenone and ethinyl estradiol signal intensity

PT	报告数	ROR	95%Cl(ROR)	PRR	95%Cl(PRR)
横窦血栓*	23	434.706	276.176	433.826	275.823
撤退性出血	30	316.959	215.220	316.123	214.841
激素避孕过程中的异位妊娠	4	314.477	109.096	314.367	109.093
疾病恐惧*	48	289.558	213.586	288.336	212.923
卵巢破裂*	4	279.535	97.796	279.437	97.793
盆腔静脉血栓形成	44	267.893	195.265	266.858	194.720
异常撤退性出血	34	242.728	169.765	242.003	169.419
腋静脉血栓形成	10	238.969	123.744	238.760	123.699
反常栓塞*	5	219.422	86.896	219.326	86.890
口服避孕药服药期间怀孕	58	185.741	141.748	184.797	141.205
卵巢手术*	3	176.878	54.155	176.831	54.156
大脑动脉血栓形成*	12	160.697	89.107	160.528	89.065
视网膜动脉血栓形成	5	152.180	61.171	152.113	61.167
浅表性血栓性静脉炎*	69	139.593	109.241	138.750	108.732
隐斜视*	3	138.051	42.743	138.015	42.744
大脑静脉窦血栓形成*	24	137.052	90.524	136.764	90.408
上矢状窦血栓*	14	134.209	78.020	134.045	77.973
动脉栓塞*	15	110.665	65.726	110.520	65.683
输卵管卵巢炎*	3	104.816	32.768	104.789	32.769
慢性肾盂肾炎*	4	96.762	35.420	96.728	35.419
肺梗死*	25	90.706	60.698	90.509	60.617
非急性紫质病症*	7	88.668	41.540	88.613	41.533
锁骨下静脉血栓形成	14	85.563	50.068	85.459	50.038
腔静脉血栓形成	19	78.378	49.516	78.248	49.470
乳腺纤维腺瘤	4	72.571	26.729	72.546	26.728
海绵窦血栓形成	3	59.580	18.875	59.564	18.876
高危妊娠	4	58.964	21.792	58.944	21.792
肠系膜静脉血栓形成	8	56.767	28.090	56.728	28.084
失写*	6	53.411	23.722	53.383	23.719
黄褐斑	4	53.151	19.673	53.132	19.673
颈静脉血栓形成	14	51.638	30.356	51.576	30.339

胆道运动功能障碍	3	51.455	16.340	51.442	16.341
大脑动脉闭塞*	10	50.208	26.787	50.165	26.778
肺动脉血栓形成	7	49.667	23.444	49.637	23.440
产品检测错误	4	48.693	18.044	48.676	18.044
脑静脉血栓形成*	16	47.000	28.610	46.935	28.590
深静脉血栓形成	471	46.878	42.703	44.973	41.123
妊娠期出血	15	45.915	27.503	45.856	27.485
子宫不规则出血	104	43.133	35.483	42.747	35.227
血栓切除术*	3	41.015	13.065	41.004	13.066
呼吸暂停发作*	3	40.720	12.972	40.709	12.973
脑血栓形成*	19	40.515	25.709	40.449	25.686
经期前综合症	8	39.842	19.775	39.814	19.771
肝腺瘤	3	39.581	12.614	39.571	12.614
马尾综合征*	4	38.904	14.453	38.891	14.453
损伤	531	38.068	34.864	36.332	33.409
肺栓塞*	532	38.031	34.833	36.294	33.377
脑缺氧*	8	37.845	18.791	37.819	18.787
子宫腺肌病	3	37.734	12.032	37.724	12.032
卵巢囊肿出血	4	37.179	13.818	37.167	13.818

注：“*”代表未在屈螺酮炔雌醇说明书中出现的 PT

2.3 不良反应信号累及系统器官

根据 MedDRA 的系统器官分类（SOC）对有信号的 PT 进行 SOC 分类排序，有信号的 SOC 为 25 个，不良事件信号累及系统与新的 ADR 信号详见表 5。

表 5 屈螺酮炔雌醇导致药品不良事件信号累及系统

Tab 5 Drospirenone and ethinyl estradiol cause adverse drug event signals to affect the system

累及器官、系统	信号数(个)	报告数(例)	报告数构成比（%）
各类神经系统疾病	75	1488	20.54
精神病类	25	1098	15.16
血管与淋巴管类疾病	25	939	12.96
全身性疾病及给药部位各种反应	9	905	12.49
各类损伤、中毒及手术并发症	9	752	10.38
呼吸系统、胸及纵隔疾病	6	575	7.94

生殖系统及乳腺疾病	32	427	5.90
各种肌肉骨骼及结缔组织疾病	5	181	2.50
胃肠系统疾病	8	181	2.50
心脏器官疾病	14	149	2.06
妊娠期、产褥期及围产期状况	6	91	1.26
肝胆系统疾病	7	79	1.09
社会环境	5	59	0.81
各类检查	8	44	0.61
皮肤及皮下组织类疾病	6	44	0.61
良性、恶性及性质不明的肿瘤（包括囊状和息肉状）	8	41	0.57
产品问题	2	35	0.48
眼器官疾病	7	35	0.48
感染及侵染类疾病	7	30	0.41
各种先天性家族性遗传性疾病	4	28	0.39
各种手术及医疗操作	3	25	0.35
代谢及营养类疾病	4	16	0.22
血液及淋巴系统疾病	2	9	0.12
耳及迷路类疾病	1	8	0.11
内分泌系统疾病	1	4	0.06
总计	279	7243	100.00

2.4 各类神经系统

屈螺酮炔雌醇致药品不良事件报告数中，各类神经系统不良反应上报数最多，共 1488 例，排名前 20 的 PT 见表 6。

表 6 各类神经系统疾病报告数前 20 排名

Tab 6 Top 20 rankings of reports of various neurological diseases

有信号的 PT	报告数（例）	报告数构成比
脑血管意外*	243	16.33%
头痛	174	11.69%
感觉减退*	75	5.04%
轻偏瘫*	47	3.16%
言语障碍*	44	2.96%

大脑梗死*	43	2.89%
偏头痛	43	2.89%
失语*	42	2.82%
惊厥发作*	39	2.62%
认知障碍*	32	2.15%
晕厥*	31	2.08%
短暂性脑缺血发作*	31	2.08%
平衡疾病*	28	1.88%
注意障碍*	28	1.88%
构音不良*	28	1.88%
缺血性卒中*	28	1.88%
偏瘫*	27	1.81%
脑损伤*	25	1.68%
大脑静脉窦血栓形成	24	1.61%
失忆症*	23	1.55%
总计	1055	70.90%

注：“*”代表未在屈螺酮炔雌醇说明书中出现的 PT

3 讨论

3.1 不良反应发生的人群特点

本研究共纳入屈螺酮炔雌醇导致的不良事件报告数共计7243例，2412个患者。报告中有6例男性患者，屈螺酮炔雌醇FDA说明书批准适应症仅用于女性，用于男性系用药错误，是否是上报时误填无法考证。在已知年龄的患者中，平均年龄为31岁，中位数为30岁，符合女性使用短效口服避孕药进行避孕的高峰年龄段相符合。报告国家来源中，美国占比最高，占53.23%，中国上报比例仅占3.19%，可能与中国上报不良事件的途径主要通过国家不良反应监测中心有关。报告者大多数为消费者，占比为54.44%，医务人员上报占比总和为33.08%。中国目前的上报主体为药品生产企业、药品经营企业、医疗机构，患者报告途径的建立是提高数据可利用性的一个方向[14]。

3.2 不良事件信号分析

3.2.1 不良事件所涉及的系统

经ROR 法和PRR 法计算，PRR和ROR数值越大，信号越强[15] ，信号越强表明药物与不良事件之间的相关性越大[16]。

得到信号较强的不良反应集中在神经系统、生殖系统及心血管系统，与说明书上描述的不良反应大致重合，但说明书中相对于神经系统方面的不良反应描述偏少，临床医生以及患者在使用时需更加关注用药所带来的各类神经系统的不良事件。

横窦血栓、撤退性出血、激素避孕过程中的异位妊娠虽然报告例数不多，但信号很强，提示临床医生如果患者用药出现以上情况，优先考虑是由屈螺酮炔雌醇导致的不良反应。在信号强度排名前50的PT中涉及血栓发的不良事件共1322例，占前50排名的58.29%。提示血栓不良事件发生率在各类不良事件发生中居首位，与说明书提示相符。究其原因，血栓栓塞的风险可能与孕激素的类型、雌激素的量或雌激素的药理学活性有关[17]。临床医生以及患者在使用时还需要关注血栓相关事件，如果有血栓风险患者应尽量避免使用雌孕激素类药物进行避孕。

3.2.2 新的不良事件信号

在本研究中，未在屈螺酮炔雌醇说明书中出现的不良反应共有104种，其中各类神经系统占69种，占比66.35%，在信号强度排名前50的不良反应中，其中新的不良反应26种，占52%。并且在各类神经系统疾病报告数排名前20的不良反应中，仅有头痛和偏头痛在说明书有记载。

新的不良反应集中在各类神经系统，且多数为较严重的不良反应，提示临床医生在应用屈螺酮炔雌醇期间警惕这些不良反应带来的脑部损伤，一旦发生此类不良反应立即采取应对措施。

3.3 研究局限性

本研究利用美国FAERS数据库为自发呈报系统，数据库容量大，但是仍具有局限性。该数据无法对患者的原患疾病、合并用药情况，及发生ADE的时间逻辑关系进行分析，且大部分由消费者进行上报，可能存在不准确的情况。屈螺酮炔雌醇FDA说明书黑框警告：吸烟与口服避孕药并用会增加发生严重心血管事件的风险。这种风险随着年龄的增长而增加，尤其是在35岁以上的女性中，并且随着吸烟数量的增加而增加，但该数据无法提供患者是否吸烟。本研究数据主要反应了美国、日本的研究结果，也有少量我国上报的数据，中国药品不良反应监测中心的数据暂时无法获得，无法与之对比分析，但可为中国人群提供用药参考。

4 结论

本次研究基于自发呈报系统利用 ROR 和 PRR 对屈螺酮炔雌醇药品不良事件信号进行数据挖掘，所得到的结果与屈螺酮炔雌醇说明书较为一致。屈螺酮炔雌醇的不良事件信号主要集中在各类神经系统、生殖系统及乳腺、精神系统血管与淋巴管系统。其中血栓相关不良事件占信号强度排名前 50 不良事件的 58.29%。挖掘出 104 个新的不良事件，其中集中在各类神经系统 69 个。临床应密切关注用药所带来的各类神经系统以及血栓的不良事件，尤其是其说明书中提及较少的各类神经系统的不良事件。

PU-018

面向航空医学科研团队的嵌入式学科服务

于丽、邓略、张莉莉、钟方虎
空军特色医学中心

目的 探讨了嵌入航空医学科研团队学科化服务的新模式

方法 全程嵌入科研一线，针对航空医学科研人员的在课题申报、课题研究及课题结题阶段不同信息需求提供学科化、知识化、个性化的信息服务。在课题申报阶段嵌入式学科服务的主要内容是嵌入情报调研和科技查新服务。在课题研究阶段嵌入式学科服务的主要内容是嵌入科研定题跟踪服务和科研数据管理服务。在课题结题阶段嵌入式学科服务的主要内容是嵌入引文/论文的收录检索服务和专利信息咨询服务。

结果 嵌入式学科服务能够准确地满足科研团队在不同阶段 、不同层次的需求 ，进而为科研人员提供及时、准确 、科学的信息，更好地助力科研创新。

结论 通过构建高效的科研信息服务环境和提供最优质服务，助跑科研项目，促进科研活动的繁荣和发展,我们相信嵌入科研团队的学科化服务将在未来的科研课题研究及科技创新中发挥越来越重要的作用。

引言

学科馆员嵌入科研一线提供学科化服务是图书馆适应新形势和新需求而诞生的一种新型服务模式[1]。空军特色医学中心图书馆作为医学专业图书馆，是中心研究部开展课题研究和科研创新的重要支撑机构，主要担负着为中心科研人员提供文献资源保障和航空医学文献信息学科化服务。近年来，随着信息环境、用户信息行为和信息需求的不断改变，嵌入式学科服务已经逐渐成为图书馆学科服务的必然趋势[2]。新形势下，航空医学图书情报人员如何更大地发挥医学专业图书馆的资源、技术等优势，主动全程嵌入科研一线，针对航空医学科研人员的不同信息需求提供学科化、知识化、个性化的信息服务，更好地助力科研创新，这是我们为之探索及实践的新课题。

1 嵌入式学科服务的内涵

嵌入式学科服务是一种新型学科服务理念和模式，其核心在于融入用户的信息环境，以用户的需求为中心，充分发挥馆员在信息获取、甄别、分析、重组方面的优势[3]，根据学科、课题来获取、组织并提供信息资源，将信息服务延伸到用户的学习、教学和科研过程中[4]。

嵌入式学科服务是将学科服务的资源、技术、工具、 人员嵌入到用户的需求中，与用户紧紧绑定，以支撑科研，创新服务。图书馆员应以全新的学科信息专家的角色，融入用户的团队，嵌入用户的学习、科研过程，实现图书馆员与用户深度合作，有效地在用户的科研过程中发挥不可或缺的作用[5]。对嵌入式学科服务而言，首先，它面向一个过程或阶段；其次，它更具针对性;再者，它注重实际服务效果，强调对用户需求的满足程度。此外，嵌入式服务体现一种更加主动的、个性化的服务态度。通过服务内容嵌入、 服务手段嵌入、 服务技术嵌入等方式，采用各种途径与技术手段将图书馆的各种资源与服务主动嵌入到用户信息环境中，让用户随时随地都能获得图书馆提供的信息服务需求。从而提升图书馆在教学和科研工作中的地位和影响力[6]。

2 嵌入科研团队学科服务特点

本文研究的“嵌入科研团队学科服务”，是指作为科研学术机构的图书馆以用户的课题需求为中心，为适应科研环境的泛在化、数字化、虚拟化、个性化而推出的一种“融入科研、嵌入过

程、资源到位、服务到人”的服务战略。学科馆员以研究课题合作者的身份，或作为科研团队的有机组成部分，将学科服务的重心直接嵌入科研一线、渗透到课题申报、研究、结题、成果发表以及社会影响评价等科研的各个环节来达到支持科研服务的目的[7]

2.1 嵌入科研团队学科服务应具有前瞻性和创新性

为了确保所承担的课题研究具有前瞻性和创新性，项目负责人在申报科研项目或制定研究计划时，最为关注的是学科本身的发展变化动态及相关领域国内外研究开发现状研究热点，诸如:有哪些人、哪些机构、哪些国家在从事该领域相关研究？已经取得的研究成果是什么？函待解决的重大问题是什么?未来发展趋势如何等。迫切希望能够提供全程化全方位的信息保障,满足其专业性系统性的信息需求。因此，为科研人员提供及时、有效的趋势前瞻性预测服务是科研团对队学科化服务的重要课题之一，也是能否将学科服务嵌入到项目立 项、研究、结题等一系列科研活动中的重要影响因素之一[8-10]。

2.2 嵌入科研团队学科服务应具有准确性和科学性

科研团队需求具有分散性 、层次性和阶段性的特点[11]。科研团队通常作为科研工作最小的组织单位，可以说是整个科研的组成细胞，作为科研单元的项目组拥有自己的科研领域或研究方向，也有自己的科研团队，在科研活动的各个阶段有着不同的信息需求和要求 ，需要根据科研人员不断变化的研究需求提供灵活多变个性化无缝服务。这就需要学科服务能够准确地满足科研团队在不同阶段 、不同层次的需求 ，进而为他们提供及时、准确 、科学的信息。

2.3 嵌入科研团队学科服务应具有针对性和系统性

科学技术飞速发展推动科研成果不断涌现。在科研过程中，科研人员需要及时发现和持续跟踪同行科研进展，通过了解和借鉴他人的研究成果、技术与方法，指导和修正自己的研究侧重点和突破方向，加速推进自己的科研进程，提高科研效率，他们需要以最便捷的方式随时关注本领域最新研究动态。因此 ，嵌入科研团队的学科服务 应该根据科研团队对科技情报的持续性需求 ，有针对性地对科研团队专注的某一领域进行信息收集 、信息组织、信息分析和信息管理，随时跟进其需求的变化，灵活而有效地不断调整先期的服务策略和内容，真正实现情报产品与科研团队需求的一对一对接 ，充分体现服务科研团队的针对性和系统性。

3 嵌入航空医学科研团队学科化服务实践

根据科研项目的生命周期，科研项目的全过程通常可以划分为项目申报阶段、项目研究阶段、项目结题与成果转化三个阶段。不同阶段对信息情报资源和各学科专业知识的需求各具特点.针对科研项目研究的不同阶段，学科馆员必须随时掌握科研人员的信息需求，准确把握服务方向，开展有针对性的学科服务。

3.1 嵌入航空医学科研课题立项阶段的学科化服务

无论科研人员申报国家级、军队级、省部级还是厅级等各级别的科研项目，都需要按照科研项目申报要求，详细阐述项目的研究背景、国内外研究现状、发展趋势、研究思路、研究内容、研究方法、创新点等，认真撰写科研项目申报书。该阶段对情报信息的需求是最为迫切的，在此

阶段嵌入式学科服务的主要内容是嵌入情报调研和科技查新服务。学科馆员首先需要与科研人员进行深度信息交流与沟通，了解科研人员的研究领域与研究方向，明确所拟定项目的大环境与现实意义，揭示项目研究的热点、未来发展方向和趋势，寻找研究突破口。然后根据科研团队的科研需求，选定数据库与检索词、制定检索策略，综合利用图书馆各种数字资源和检索工具，全面搜集有价值、重要的中外文文献，并对其进行筛选、甄别与提炼，及时为科研人员推送与研究领域相关的专题调研报告、前沿动态分析等，供研究人员对研究项目做出可行性分析。最后，通过对比检索文献与科研项目查新点，同科研人员一起评估拟申报科研项目的新颖性与创新性，完成科技查新报告。同时，在形成申报书的过程中，为申报书的编写提供进一步的数据和材料，为申报项目的新理论、新方法、新技术、新价值等提供数据支撑，提高项目获准立项的成功率[12-14]。以笔者跟踪中心研究部军事医学重大研究课题“长航时飞行航空医学保障关键技术研究”为例，该课题在撰写立项申请书之前，笔者曾检索了大量相关文献并对其进行了查新，论证了该课题的研究价值与意义，为该课题的成功立项、撰写了相关的科技查新报告和信息分析研究报告。

3.2 嵌入航空医学科研课题研究阶段的学科化服务

课题选定后进入课题研究阶段，研究阶段主要是对课题研究的整体布局、基本结构和研究思路进行规划和实施，是研究过程中的创造性劳动阶段[15]。在研究阶段中遇到的信息难点、知识方面等具体问题、非常需要获取大量信息、知识支持。另一方面，课题研究作为一个动态发展和不断变化的事物,研究阶段仍需不断了解国内外相应知识的最新进展情况，以持续补充新知识、完善课题保证课题的创新性。这个阶段是科研项目研究最核心的阶段，事关整个项目完成质量的高低，在此阶段嵌入式学科服务的主要内容是嵌入科研定题跟踪服务和科研数据管理服务。

科研定题跟踪服务是针对科研需要,持续的、系统的向科研人员提供所需文献资料，做好定题服务,一是要与课题参与人员进行沟通,准确了解课题的主要内容、研究过程及文献信息需求,为课题前期调研、开展立项、中期成果到结题验收的全过程,提供信息资料的收集、查询、检索、分析、整理等系统服务;二是在紧密跟踪课题研究阶段过程中，密切关注课题的阶段性研究趋向，帮助课题组把关，帮助检查方案设计、思路设计、目标设计、内容设计、方法设计及搜集课题内容重点性及科学性，根据课题的计划及进展情况,为课题组提供具有前瞻性和广泛性的文献综述、研究报告等信息,以便于科研人员了解学术前沿,摸清研究动态,开启科研新思路[16]。三是在科研项目接近尾声的时候,预测其市场竞争力，宣传科研成果,促进科研成果转化尽快将科研成果转化成生产力等。

科研数据管理是需要对科研所需要的期刊、会议论文、及科研报告、试验方法、结果等进行信息存储、组织及利用[17]。嵌入式学科馆员需要熟悉科研数据的主要类型与格式，利用元数据管理、数据提交、检索、权限管理等方面的优势，对科研项目研究过程中所产生的各种实验数据、图片、研究报告、论文进行管理，提供科研数据管理服务。如：空军特色医学中心图书馆在嵌入“十二五”全军重大课题，开展学科化服务期间，根据科研人员所承担的项目与研究课题，对航空医学重要的数据库、网站、专家博客等网络信息源进行定期扫描，经过分析加工整理成专题信

息，并通过海报、微信、邮件等方式主动及时推送科研项目研究领域的最新资讯、学术动态和同行研究进展，行之有效地提供了定题跟踪服务，同时，随时随地帮助科研人员对研究过程中产生的珍贵数据进行存储监管和存档，通过数据规范化的组织与管理，从而方便科研人员的检索及利用相关数据，提高科学研究效率[18-19]。

3.3 嵌入航空医学科研课题结题与成果转化阶段的学科化服务

课题研究成果的鉴定是课题研究的最后一个环节。也是整个课题研究活动中的关键工作。这个阶段主要内容是：书面报告的形成，图件整理数字化，汇报评审，推广应用等。科研人员需要明确科研成果的被引用情况与被收录状况，掌握科研成果的社会影响力;了解研究领域的专利申报与授予情况，市场应用情况，争取申报国内外专利，并实现科研成果的转化; 在此阶段嵌入式学科服务的主要内容是嵌入引文/论文的收录检索服务和专利信息咨询服务。在成果鉴定前,学科馆员要协助课题组负责人把课题的全部材料收集齐全, 严格审查、把关,确保归档的文件材料的质量。科研成果发表后，学科馆员利用对 Web of Science，EI，CSSCI，ESI等各种收录数据库的熟练掌握，以及中国社科索引、人大复印报刊等文摘数据库，出具科研成果被收录与被引用的检索报告，提供引文/论文收录检索服务;在项目成果的转化或专利申请时，根据科研人员的成果，协商确定专利申报类型，检索国内外专利网站与专利数据库，对比已有专利与研究成果的差异，协助科研人员撰写专利申请文件，帮助科研人员进行专利申报，提供专利信息咨询服务。为科研成果的鉴定、评估、验收、转化、奖励等提供客观依据，保证相关环节的权威性和科学性。去年，笔者持续跟踪5年的一项重大课题结题，其科研成果很快在“医疗卫生装备”杂志上刊登，科研成果发表后，我们为其提供了文章被引及收录情况的检索报告，使科研人员掌握自己研究成果的影响力，通过对科研项目进行科学评价，了解了科研团队或所处的国际化水平，在检索了相关专利数据库的基础上提出了专利申请并已顺利获得批准。

通过嵌入式学科服务实践探索，架起了图书馆和科研人员之间沟通的桥梁，建立了通畅的需求与保障渠道，促进了图书馆资源与服务的利用[20]。同时紧密围绕中心研究部核心工作，显著提升了图书馆在中心的影响力。

结 语

嵌入式学科服务以“融入一线、嵌入过程”为图书馆新的服务理念，其发展方兴未艾。作为应对信息环境变化学科服务推出的新策略和新方向，其赋予了学科服务新的生机和活力，为图书馆转型发展找寻到新的着力点和生长空间[21]。目前国内很多同仁都在积极探索图书馆嵌入科研团队的工作方式和方法，努力探寻出一条适合科研课题组需求的图书馆信息服务的模式。实践证明，嵌入科研团队的学科服务，会帮助科研人员更方便、高效获取所需要的信息、节省大量的时间和精力、提高科研工作效率、早出创新成果；可以使学科馆员充分施展自己的潜能和价值，通过构建高效的科研信息服务环境和提供最优质服务，助跑科研项目，促进科研活动的繁荣和发展,我们相信嵌入科研团队的学科化服务将在未来的科研课题研究及科技创新中发挥越来越重要的作用。

PU-019

健康医疗大数据治理体系及方法思考

张睿、甄天民、谷景亮、段永璇、宋燕、韩志琰、宋奎勋
山东省医药卫生科技信息研究所

目的 由于在数据采集、处理等过程中缺乏相关标准约束，导致健康医疗大数据大多存在质量差、价值低的问题，进而限制了医疗机构监管、公共卫生监测开展实施、卫生政策制定研究以及医学研究的开展。本文对当前健康医疗大数据存在问题进行综合分析，在此基础上提出以数据生命周期为总线的数据治理路径，并结合具体问题提出解决方法。

方法 考虑到不同数据源维度间的相关性，以及数据生命周期，构建了健康医疗大数据治理体系，分为数据采集、数据处理、元数据、分析四个环节。这四个环节环环相扣，每一个阶段的数据质量都会对下一阶段的数据产生影响，对健康医疗大数据治理通过数据分析作为总的检验，进一步指导规范各个阶段的数据标准，通过多轮完整性、准确性、一致性检验的数据迭代，为数据分析提供坚实的基础。

结果 目前，健康医疗大数据的主要来源是各级医院信息系统电子病历、疾病预防控制中心公共卫生服务数据、政府医药卫生行政数据等，由这些数据逐渐汇聚成区域的大数据平台，在这个过程中无论信息平台建设以及管理上，都有相关规范，然而这些规范的实施还有待提升，这些基础数据仍存在缺失、错误等质量问题。数据治理也需从底层数据收集阶段做起。针对数据完整性、准确性、一致性的治理提出相应思路和具体措施。

结论 在健康医疗大数据建设初期，由于数据集在类型、结构、组织以及数据粒度上具有异质性和不均匀性，造成原始数据质量不高甚至会影响数据分析最终效果。以规范数据采集、数据处理、元数据存储各个阶段的数据标准数为基础，在采集数据以及数据融合阶段对数据集进行高效的组织，对数据结构以及类型等进行规范，并且对不同数据源的数据集成和融合技术技术进行改进。对数据的准确性、完整性、一致性进行校验和处理，从而形成规范数据进行存储，不仅可以提高数据分析的效率和准确率，还可以降低健康医疗大数据的软硬件运行成本，提高数据搜索的速度。另外对数据更新周期进行管理，各个医疗体系产生数据的速度和规模不仅对存储体系提出了紧迫的要求，医药健康大数据的价值更取决于数据的更新速度，同样数据治理的速度也要跟进，伴随数据更新速度进行多轮数据治理迭代，快速实时的数据治理是提高数据质量的保障，同时完善数据分析效果。

PU-020

我国全科医生资源配置现状及公平性研究

温楠、程艳敏
山东省医药卫生科技信息研究所

目的 分析 2015-2017 年我国各省市全科医生资源配置情况及公平性。

方法 计算我国 2015-2017 年全科医生按人口和经济配置的基尼系数和泰尔指数，并对泰尔指数进行分解。

结果 2015 到 2017 年，我国全科医生数由 188649 人增加到 252717 人，增长了 33.96%，年均增长率 15.74%。每万人口全科医生数从 1.37 人增长到 1.82 人，年均增长率 15.26%。基尼系数从 0.2469 下降到 0.2423，人口配置泰尔指数从 0.1063 下降到 0.0987，经济配置泰尔指数由 0.0659 下降到 0.0467。在人口维度下，组内差距对不公平性贡献率达到 60%以上；而在经济维度下，组内差距贡献率达到了 99%以上。

结论 我国全科医生队伍不断壮大，但总量仍不足；我国全科医生分布不平衡，经济社会发展程度对全科医生的分配影响较大；我国全科医生的培养机制有待优化。

一、健康医疗大数据现状

2016 年 6 月，国务院办公厅发布《关于促进和规范健康医疗大数据应用发展的指导意见》，将健康医疗大数据应用发展纳入国家大数据战略布局。同年底国家卫生计生委启动健康医疗大数据中心与产业园建设国家试点工程[1]，全面推动大数据应用与健康医疗行业的深度融合。

根据国务院办公厅印发的《关于促进和规范健康医疗大数据应用发展的指导意见》，健康医疗大数据是指基于传统的公共卫生、医疗服务、医疗保障、药品管理、计划生育、综合管理等业务，延伸至互联网、物联网、人工智能等领域，运用信息网络技术产生的与健康医疗相关的数据[2]。从广义上讲，健康医疗大数据泛指存储在医疗卫生领域内各种管理信息系统中的数据，包括医保政务、医疗机构、公共卫生、药品销售、保险等不同系统内的信息。从狭义上讲，健康医疗大数据是指医生为患者治疗过程中产生的各种记录与数据，包括患者基本信息、检验信息、医学影像信息、临床信息、电子病历和费用信息等[3]。

健康医疗大数据不仅具有大数据的 5V 特性，即数据体量巨大（volume）、增长与处理速度快（velocity）、数据结构多样（variety）、价值密度低应用价值高（value）与真实性（veracity）[4]，而且由于医疗数据的多态性（文字描述、检验数据、电磁信号、影像、音视频、动画等不同形态），使得健康医疗大数据还有多态性、不完整性、时间性和冗余性。随着卫生健康服务体系信息化建设的不断完善，电子病历、医保记录、疾病监测等基于日常业务产生的数据，不断累积，数据激增，给健康医疗大数据的发展带来了巨大挑战，由于在数据采集、处理等过程中缺乏相关标准约束，导致健康医疗大数据大多存在质量差、价值低的问题，进而限制了医疗机

构监管、公共卫生监测开展实施、卫生政策制定研究以及医学研究的开展。本文对当前健康医疗大数据存在问题进行综合分析，在此基础上提出以数据生命周期为总线的数据治理路径，并结合具体问题提出解决方法。

二、健康医疗大数据质量问题

医疗大数据的质量治理体系是一项复杂的系统工程，包括数据的产生、采集、转换、存储、传递 、分析等生命周期过程[5]，现在阶段主要有以下问题：

完整性不够：在数据采集过程中，由于各个系统信息孤岛问题严重，区域卫生信息平台覆盖不足，而相关大数据发展规划还未涉及的地区，那么该地区的医疗机构的数据无法收集，又或者一些偏远医疗机构信息化程度有限，缺乏获取自然环境、社会环境以及经济属性信息的途径，导致数据的完整性不够。

准确性不够： 在数据产生过程中，医务人员填写医疗信息系统中，手动填写会导致数据错误遗漏以及不准确等问题。医用自动化设备如编写体征采集设备，缺乏统一度量和验证，导致数据采集偏差。进一步在数据采集过程中，由于医院 HIS 系统的不同品牌，其底层数据管理模块规范不一，那么在数据整理数据格式转换过程中会导致错误。

一致性不够：即多源数据间的相关性。如何把从不同医院采集来的数据，在就诊情况、处方单情况、付费情况以及时间等各方面进行相关性对应，做到数据格式以及数据映射的一致性。

三、构建健康医疗大数据治理体系

考虑到不同数据源维度间的相关性，以及数据生命周期，构建了健康医疗大数据治理体系，分为数据采集、数据处理、元数据、分析四个环节。这四个环节环环相扣，每一个阶段的数据质量都会对下一阶段的数据产生影响，对健康医疗大数据治理通过数据分析效果优劣作为总的检验，进一步指导规范各个阶段的数据标准，通过多轮完整性、准确性、一致性检验的数据迭代，为数据分析提供坚实的基础。

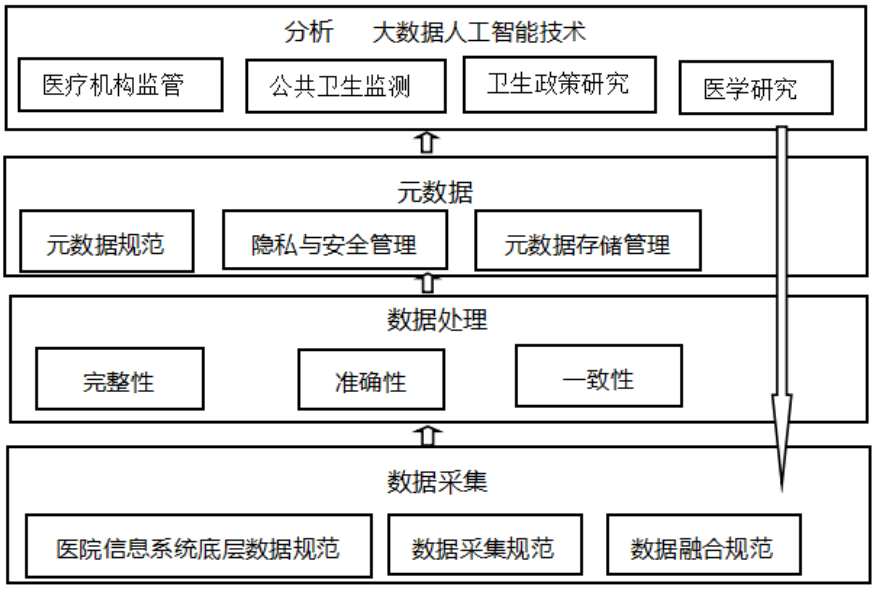


图 1 健康医疗大数据治理体系

在数据采集阶段，标准化是基础。从数据产生的源头来看，医院信息系统的医疗数据要填写规范，尽量减少错误值或缺失值的产生。并且不同的医院信息系统的底层数据模块标准要统一，以便在采集和融合入库阶段减少工作量和错误的产生，同样也要制定数据采集和不同数据源融合的规范。

在数据处理阶段，首先对各个源数据，即各个医院信息系统提取的数据做完整性、准确性和一致性的数据检验。再对数据进行筛选、重组、结构化、预处理，将残缺值和错误值去除，其余数据做到一致性。然后再对逐渐融合后的大数据逐一进行多次这样的检验和处理。

元数据管理阶段，由于不同的医药大数据平台支持不同类型的存储工具，所以要对规范元数据生成、存储以及数据脱敏、数据使用权限等方面做相应规范。

对健康医疗大数据进行分析和应用对医疗机构监督、公共卫生监测以至患者、临床医生以及卫生政策制定者产生了深远意义。从患者角度进行挖掘和分析，可以帮助其获得更好的健康及疾病管理。从疾病数据角度分析，可以对疾病进行流行病学的相关研究，帮助医生诊断以及新药的研制。从医药卫生政策的角度分析，可以客观有效的评估现有医疗体系，从而为卫生政策制定提供重要依据。

四、从数据集看数据治理具体措施

目前，健康医疗大数据的主要来源是各级医院信息系统电子病历、疾病预防控制中心公共卫生服务数据、政府医药卫生行政数据等，由这些数据逐渐汇聚成区域的大数据平台，在这个过程中无论信息平台建设以及管理上，都有相关规范，然而这些规范的实施还有待提升，这些基础数据仍存在缺失、错误等质量问题。数据治理也需从底层数据收集阶段做起。各医院的信息系统虽然有不同品牌，但是其基本数据情况涵盖门诊和住院的摘要、诊断、用药以及费用结算等内容[6][7]，概略总结为表 1，现结合数据完整性、准确性、一致性易出现的问题并提出相应治理策略。

门诊摘要	医疗服务-门诊摘要信息OUTPATIENT
	医疗服务-门诊摘要信息_诊断OUTPATIENT_DIAG
	医疗服务-门诊摘要信息_用药OUTPATIENT_DRUG
	医疗服务-门诊摘要信息_费用结算表OUTPATIENT_SETT
	医疗服务-门诊摘要信息_费用OUTPATIENT_FEE
住院摘要	医疗服务-住院摘要信息INPATIENT
	医疗服务-住院摘要信息_会诊INPATIENT_CONSULT
	医疗服务-住院摘要信息_用药INPATIENT_DRUG
	医疗服务-住院摘要信息_费用结算表INPATIENT_SETT
	医疗服务-住院摘要信息_费用INPATIENT_FEE
	医疗服务-住院摘要信息_入院诊断INPATIENT_INDIAG
	医疗服务-住院摘要信息_出院诊断INPATIENT_OUTDIAG
	医疗服务-住院摘要信息_症状INPATIENT_SYMP

表 1 主要数据关系表

1、数据不完整

数据不完整会使数据分析阶段丢失大量有用信息，使数据模型中所表现出的不确定性更加显著，数据中蕴含的规律更难把握，过多缺失值会使数据分析模型导致不可靠的结果。

以门诊数据为例，数据完整性是指病人就医的过程中的情况在各个表中都有记录，即病人就诊后即在门诊摘要表中有一条记录以 patient_id 进行唯一标注，然后与诊断表中相关记录也对应同一个 patient_id，同理在用药表以及费用结算表该病人的记录都对应同一个 patient_id。如果在某个表中病人记录丢失，也就是病人就诊在门诊摘要表中有记录，但是在诊断表、用药表或结算表中没有相关记录，这样的数据不完整性会在数据分析过程中产生错误，以各家医院每年门诊/住院就诊人数、各家医院每年门诊/住院总就诊费用为例进行分析。例如，分析该医院在某个时间段门诊人均费用，用结算表中该时间段产生的门诊总费用除以该时间段所有的就诊人次，如果门诊摘要中所有就诊病人记录条数或门诊费用结算表中病人记录条数有缺失，都会导致最后的门诊人均费用结果有偏差。如果多家医院都有这样的数据不完整的情况，那么在计算一个区域内所有医疗机构中某段时间内的门诊人均费用偏差就会更大。

针对不同表单记录条数有缺失的情况，数据治理的措施是比对各个表中记录条数，将缺失记录进行标注，根据不同的统计指标进行有条件的剔除。例如，计算某家医院某时间的门诊人均次，如果门诊摘要表中的病人记录条数与门诊结算表中的病人记录条数不一，就用门诊结算表中病人记录条数作为分母进行计算，可以使数据偏差减小。

同样除了表中所列门诊或住院摘要表，有关诊断以及手术等相关信息，填写会有缺省值，对以上这些数据不完整的情况，需要对数据质量进行初步检验，以及各个表中的一些关键字段填写的缺省率，可以督促规范医院信息系统的填写，进一步在数据收集和整合过程中减少错误，也可以在数据分析过程中根据统计需求去除缺省的记录，以减少数据偏差。

2、数据准确性

数据准确性最主要源头是各医院手工填写的准确性，数据的不准确性很大的原因是是人工填写导致或者是在不同数据源进行融合过程中产生的。数据的准确性分析就是检验数据是否有录入错误以及不合常理的数据，数据异常值会给数据分析结果带来不良影响。通过分析数据异常值产生的原因，也可以发现原因从而进行改进。

各家医院信息系统中诊断表药物详情以及手术详情有关的表中，不仅需要规范数据字典，针对不同的药物类型、治疗类型、医保类型等，如果数据字典规范不一致或者填写不规范，对数据解读和数据分析产生困难。例如有关药物表中药物类型、剂量、剂型等字段数据不准确，该条有关用药情况的数据的属性描述就产生混乱。药物类型字段的填写，决定该项药物是否为中药、西药或者是中成药就产生错误，导致在有关各种药物开具比例以及相关费用统计分析上产生数据偏差。

另外关于费用结算表也不乏出现数据错误现象，导致就不同病种、人均、自付费比例、就诊及住院费用等统计分析项目进行衡量的时候导致数据偏差。有关费用项目出现错误值，即样本中的个别值，其费用数值明显偏离其余的观测值，分析错误值就是找到这些离群点。

可以先对数据做简单统计量分析，初步判断那些数据量是不合理的，最常用的方法是找到最大值和最小值来判断某条数据是否超出了合理范围。例如，某病人某次门诊就诊产生费用如果小于 0，则该项数据存在异常。进一步使用箱线图分析[8]，可以找到异常值的上限和下限。具体做法是：定义 QL 为下四分位数，表示全部观察值中有四分之一数据取值比它小；QU 称为上四分位数，表示全部观察值有四分之一数据取值比它大；IQR 称为四分位数间距，IQR=QU-QL,其间数据包含了全部观察值的一半。异常值通常被定义为小于 QL-1.5IQR 或者大 QU+1.5IQR。箱形图可以直观反映数据分布的本来面貌，上下四分位数不易扰动，在判断异常值比较客观，具有优越性。利用箱线图可以有效找到数据异常值，对于门诊处方结算费用找到上位异常值和下位异常值，在数据分析过程中进行相应处理。或者是根据出现的异常值，从数据记录、数据收集、数据融合以至数据存储过程中逐一分析其错误产生原因，制定相应的方法进行数据治理。

3、一致性

由于以上叙述存在的数据不完整性和数据不准确性，进一步导致了数据的矛盾性和不相容性。对不一致的数据进行分析，可能会导致与事实相违背的挖掘结果。而且在数据集成过程中，由于来自不同的数据源，对不同格式的数据处理过程中也会造成不一致性。例如门诊费用表和门诊费用结算表针对同一张处方单各自根据自己表单中项目进行计费，结果不一。分析错误产生原因，从下而上，可能是填写错误，也有可能是在数据融合过程中数据格式不统一以及相关数据冗余等原因造成的。

五、总结

在健康医疗大数据建设初期，由于数据集在类型、结构、组织以及数据粒度上具有异质性和不均匀性，造成原始数据质量不高甚至会影响数据分析最终效果。以规范数据采集、数据处理、元数据存储各个阶段的数据标准数为基础，在采集数据以及数据融合阶段对数据集进行高效的组织，对数据结构以及类型等进行规范，并且对不同数据源的数据集成和融合技术技术进行改进。对数据的准确性、完整性、一致性进行校验和处理，从而形成规范数据进行存储，不仅可以提高数据分析的效率和准确率，还可以降低健康医疗大数据的软硬件运行成本，提高数据搜索的速度。另外对数据更新周期进行管理，各个医疗体系产生数据的速度和规模不仅对存储体系提出了紧迫的要求，医药健康大数据的价值更取决于数据的更新速度，同样数据治理的速度也要跟进，伴随数据更新速度进行多轮数据治理迭代，快速实时的数据治理是提高数据质量的保障，同时完善数据分析效果。

PU-021

基于 5G 技术的医联体信息化共享模式探研

喻平、李绍杰、廖廓、成幼玲、周艳红
湘潭市中心医院

目的 利用 5G 高速、安全互联技术，推行共享模式，为医联体信息化规划、建设寻找出新的方向。
方法 针对医联体信息化建设特点，结合基于 5G 专网的云计算 2.0 的桌面技术，规划设计了支撑医联体运行的信息化网络系统及应用场景，并对实际应用效果进行了总结

结果 利用 5G+云桌面技术，构建医联体整体信息化系统架构的模式，使得医联体内各成员单位的信息流更畅通，网络安全更有保障，信息系统的稳定性更高、用户的体验更好。
结论 为互联网＋医疗服务信息化建设提供了有益的参考和借鉴。

1 背景

近年来，医联体被列入新医改的重点举措后，全国各地各级医疗机构纷纷响应。2017 年成为医联体体制框架搭建的高峰期，2018 年，医联体建设进入加速推进期，2019 年随着分级诊疗的持续推进，各地不断夯实医联体建设基础，并形成了各具特色的医联体运行模式[1]。2017 年 4 月湘潭市中心医院医疗健康集团正式挂牌成立以来，市中心医院联合社区卫生服务中心、诊所，以及湘潭县、湘乡、韶山医疗机构等 30 多家单位，成立了医疗服务联合体。为加速医疗对接，研究和建立支持多异地、易维护、安全稳定的信息化架构和应用系统势在必行，但基层医疗机构信息化水平较低[2]、基础薄弱等原因，实际工作难度大。而医联体改革，重点是“联”，上海医联体改革通过输出技术、派出专家、开展规范化培训、搭建远程网络协作平台， 开设区级医院示范门诊，以信息化的互联互通带动区级医院与三级医院在就诊流程、诊疗规范、 医疗服务质量上的均等化。因此，如果能直接使用三甲医院信息化优势资源，实现共享、共用，不失为一种好的方案选择。随着 5G、云计算技术飞速发展，医联体信息化整体规划与建设有了新的选择。

2 现状

医联体信息化通常做法是统筹规划、分级建设。如苏北人民医院医疗集团就是总院组织专家多次调研、规划、指导两家分院区信息化建设，针对基层医院信息化基础相对薄弱等问题，按照总院总体部署，改造托管医院信息化基础设施，升级系统软件功能，建设远程会诊、远程影像、远程 B 超、远程心电、远程查房等应用系统，打通医联体内上下级医院信息系统，畅通双向转诊渠道，实现双向转诊流程线上线下一体化运行[4]。然而在实际操作中，存在诸多难点。

2.1 基层信息系统运维难度大

三甲医院信息化体系相对完善，IT 技术人员有较高的水平，而基层单位基本没有专业 IT 技术人员。另一方面，近年来勒索、蠕虫病毒等从终端机入侵至服务器端现象严重，对于终端机的安

全防范问题变得尤为突出。目前基层的网络信息安全薄弱，与医联体间数据流转量增大，加大了网络带宽需求和系统安全风险。

2.2 终端维护半径长、维修效率低

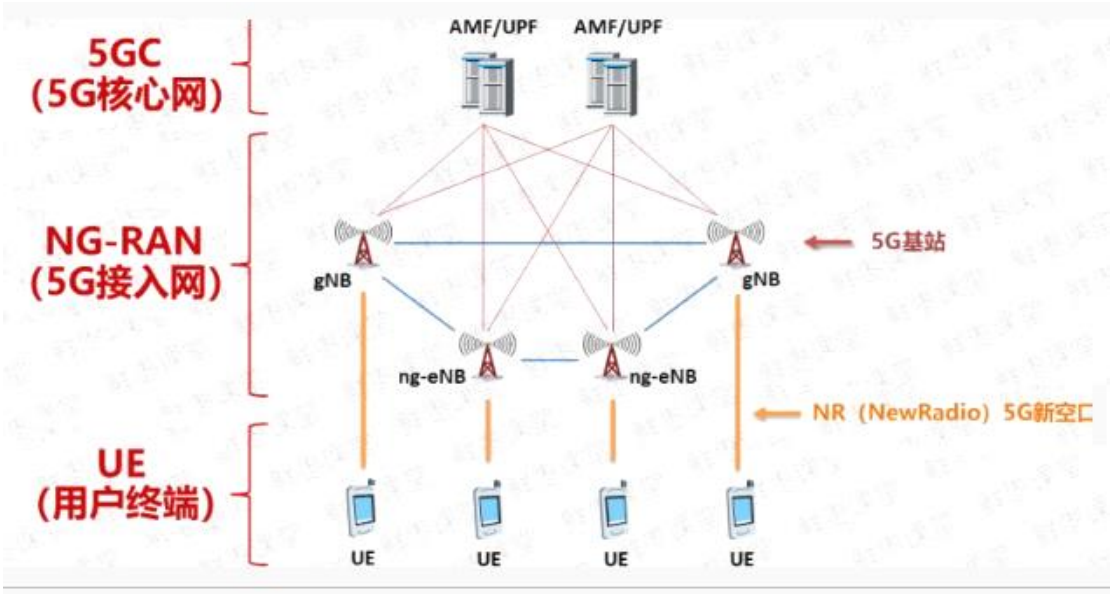
三甲医院信息技术人员有限，医联体联合单位越多，工作范围越广，故障处理时间越长。同时普通 PC 机等硬件网络故障处理或是系统升级扩容更是流程长和过程繁琐，对于普通 PC 机的安全防范问题也尤为突出。

3 基于 5G 云桌面技术

目前，基于云计算架构的桌面系统在企业级用户中得到了广泛认可和普遍应用，而 5G 作为全新的高速率、高稳定性、低时延的连接架构，为智慧医疗的建设提供了重要的技术支撑。将 5G 独特的切片技术与云桌面结合，对医院现有的网络改造后进行基础网络切片管理，形成若干专用网络，为数据传输提供安全、高效的数据通道。

3.1 5G“高带宽、低时延、高容量”促进医疗效率提升

5G 的 SDMA 是利用 Massive MIMO 的波束成型技术来实现的，由于可以控制电磁波在特定方向传播，因此在不同空间方向的用户可以同时使用全部频谱资源进行通讯，提升了基站的容量。此外，由于信号能量集中在特定的方向，还能减少对其它接收机的干扰。医院可采用 5G 技术，实现医院内部的自动缴费机、自动查房等终端接入。



图一 5G 基础结构图

5G 技术与承载网相结合，实现云计算在 5G 技术下的承载。医院采用 5G 作为数据传输，大大提升了现有网络的无线传输能力。采用现有网络有线承载和无线网络 5G 承载相结合的改造方案，为医院后续的远程医疗、机器人引导、物联网、语音交费等业务的信息化支撑打下了良好的网络基础。

3.2 5G 技术与云桌面技术的结合

云桌面技术将应用和数据进行云化处理，使得应用和数据不会孤立存储在某台服务器或存储器中，而是在一个可无限扩充的云化资源池中。医生走诊及医护移护等医疗场景中云桌面应用随着医生和医护人员移动，大大提高了医护人员的工作效率。

5G 技术使云桌面单独存在于一张网络中，在用户使用该网络时，则进行身份校验，检验不通过，则不可使用该网络，有效保障了移动中的医护人员使用网络的安全性。

医院后续则在 5G 技术和云桌面硬件环境下，将搭建未来人工智能和物联网的多项业务应用，对推进医疗信息化、提升医疗效率具有很好的价值。

3.3 5G 网络切片与无代理杀毒协同

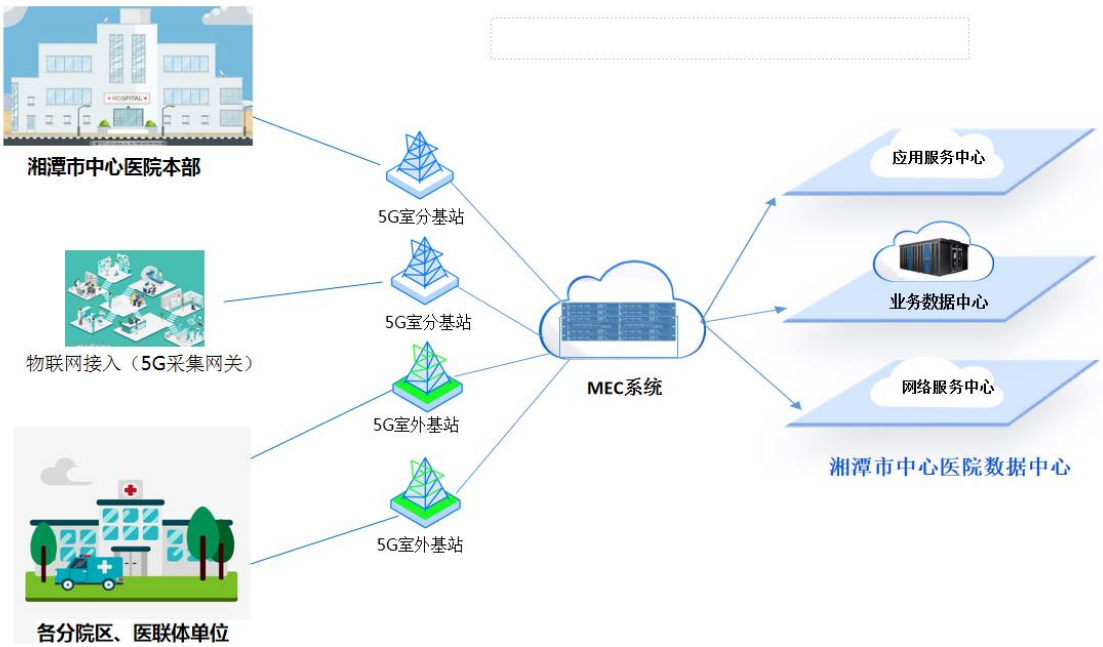
网络切片是一种按需组网的方式，可以让医院在统一的基础设施上切出多个虚拟的端到端网络，每个网络切片从无线接入网到承载网再到核心网在逻辑上隔离，适配各种类型的业务应用。医院使用 5G 网络，将 WIFI、承载网、核心网和无线网进行切片，每个网络用户使用一张切片，并在编排时对每张网络逻辑隔离，保证了专网的安全性和可靠性。而在使用专网时，用户终端在后台编排系统中采用预设白名单或申请切片需求部署机制，保证了用户数据在传输和存储过程中的安全。

而云桌面部署在 5G 网络中的无代理杀毒，有效解决了传统杀毒软件安装使用中繁琐问题，提高了虚拟基础设施利用率和终端的安全性，大幅提高了系统维护效率，降低了维护成本。同时也提升了用户的体验度。

4 医联体内信息化共享模式

4.1 医联体网络规划

在医院现有的系统网络架构上，直接利用运营商现有 5G 室分基站连通院内网络，5G 室外基站连通分院区及其它各处医联体单位。而另一端通过 MEC 系统接入医院主干网络，利用 MEC 直接在网络边缘存取数据，具有更低时延；同时数据在网络边缘卸载，具有更高的安全性。连接示意图如图二所示：



图二　医联体 5G 网络接入示意图

4.2 应用场景

将湘潭中心医院建设的各类信息系统，装入云桌面系统（如下图三所示），让各分院区、授权医联体单位共享使用。临床用软件包括医护工作站、电子病历、病人 360 全息视图、放射报告、B 超报告、不良事件上报等等近 20 个应用系统；通用办公用了类应用软件覆盖医院 OA 系统、人事系统、物资领用系统、固资系统、预算系统、廉洁风险防控等。

图三 云桌面系统共享常用软件

采用云桌面共享使用信息系统的模式提升了医联体成员单位互联网+医疗服务能力。医护人员可以在不同院区专网、也可以用 5G 网络通过安全度高的 SSL 加密隧道接入个人专属桌面系统环境，随时随地方便操作自己熟悉的软件系统，查看患者信息、录入诊疗结果，提高了工作效率。

图四 云桌面系统连接示意图

而对于湘潭中心医院下级紧密型医联体单位如五里堆社区卫生服务中心、姜畲镇卫生院，则通过 5G 直接使用云桌面系统，这样既可以减少设备维护工作量，也可以方便院内各级、各类应用系统直接连通使用。如中心医院的专家到社区坐诊，遇到疑难病人可以方便为病人开设三级医院的检查、住院证等；遇到慢性病人则可以方便查询病人住院的就诊记录、检查结果等。

三级医院医生或社区医生还可以通过随身带着的 PAD、手机记录病人诊疗情况、开设检查检验，及时查阅报告等，大大提高了用户的体验度和工作效率。

4.3 应用优势

1、利用运营商资源，大大缩短了网络建设周期，降低了信息系统运维成本；

2、利用 MEC 技术，在 5G 网络边缘存取数据，具有更低时延，满足医疗图片等大数据流低延时传输和网络安全管理的需求；

3、成熟的云桌面技术，能够根据终端需求自动调整桌面的内存大小、CPU 资源，相对传统 PC 方式更为灵活，大幅降低了系统维护人工成本。同时，相比传统 PC 方式下的操作系统补丁、病毒库更新、外接设备控制等方面，云桌面系统更安全。

4、通过云桌面可以便捷的将临床需要用到的应用系统推送到医联体各机构点或随身携带的移动设备使用。

结语

5G+云桌面技术的应用，给医联体整体信息化系统架构建设带来新的模式，使得医联体内各成员单位的信息流更畅通，网络安全更有保障，信息系统的稳定性更高、用户的体验更好。随着国家 5G 建设步伐的加快，这一模式将会进一步得到推广应用，为医联体模式的高效运行奠定坚实了信息化基础。

PU-022

基于美国 FAERS 数据库的阿特珠单抗不良事件信号挖掘研究

白雪霏[1,3]、陈力[3]、钟薇[2]、黄佳[2]
1. 四川大学华西药学院 2. 成都市新都区人民医院
3. 四川大学华西第二医院

目的 采用美国 FDA 不良事件报告系统（FAERS）挖掘阿特珠单抗的不良事件信号，为其安全风险控制和临床合理用药提供参考。

方法 采用比例失衡法对美国 FAERS 数据库中 2017 年第一季度至 2020 年第一季度共 13 个季度的阿特珠单抗相关不良事件报告进行数据挖掘。

结果 获得阿特珠单抗报告数 11140 个，不良事件信号 440 个，涉及 23 个系统器官，发现了如卵巢静脉血栓症、子宫积水、超进展肿瘤、回肠穿孔等多个新的药物不良事件。

结论 阿特珠单抗发生不良事件的信号主要集中在全身性反应和呼吸、胃肠及血液等系统，建议临床用药时应注意监测患者的不良事件，及时采取干预措施。

PU-023

医联体模式下分级诊疗建设思考

王琳琳
郑州大学第三附属医院（河南省妇幼保健院）

目的 随着医改的深入，为推动形成分级诊疗格局，合理利用资源，破解无序就医乱象，以医联体为载体，建立分级诊疗，逐渐成为媒体和公众热议的焦点。本文简要介绍了医联体模式，分级诊疗内涵和建设原则，最后对医联体模式下分级诊疗制度的建设进行了剖析。

方法 首先有利于提升基层医疗服务能力。通过组建医联体，充分发挥区域内核心医院的技术辐射和带动作用，促进优质医疗资源下沉。其次有利于形成医疗机构之间的分工协作机制。通过组建医联体，对区域内现有医疗资源进行整合和机构调整，引导不同级别、不同类别医疗机构建立目标明确、权责清晰的分工协作关系。三是有利于实现双向转诊模式。鼓励慢性病医疗机构加入医联体，畅通医联体内不同医疗机构之间的转诊机制，有利于为患者提供疾病诊疗-康复-长期护理连续性服务。

结果 医联体可通过创新远程医疗服务模式，将一个区域内的三级医院、二级医院、乡镇卫生院与社区医院紧密连接起来，实现资源互通，有效推进分级诊疗。通过建设发展医联体，在各级医疗机构之间建立统筹协调和分工合作机制，着力提升基层医疗机构的诊疗水平、运行效率，合理分流患者，可切实促进“基层首诊、双向转诊、急慢分治、上下联动”的分级诊疗建设推进。

通过重点专科对口扶持、绿色通道、远程会诊、业务指导、慢病团队管理及绩效考核评估等远程医联体服务内容，可以使上端医院、下端医院、医生、患者四方联动，在解决基层群众难以享受医疗资源问题的同时，实现对基层医院的帮扶，提高基层医院整体医疗水平，促进医疗资源均衡化。

结论 尽管医联体具备多种优势，也被炒得如火如荼，但相当一部分的医联体依旧是“叫好不叫座”，现在医联体的建设缺乏整体规划，发展极不平衡。而且，许多医联体建设的最终目的并非是为提高基层医疗卫生服务能力，只是为了扩大病人来源。在医联体中，大医院希望多吸收病患提高收入，小医院希望大医院的专家医生来院会诊、手术，以提升自身水平和知名度，进而提高医院的收入，而区域内的服务对象数量基本是稳定的，因此此消彼长不可避免。这为将要开展的分级医疗服务，埋下了严重隐患。如何破解这种现象，关键还是要看政府，不仅仅是行政命令，还要有财政导向和其他力量的配合，既要考量整体水平的提高，也要兼顾弱势群体的生存和发展。

PU-024

新冠肺炎疫情影响下高校医学图书馆职能建设的思考——以武汉大学图书馆医学分馆为例

王俊
武汉大学图书馆

目的 探讨新冠肺炎疫情影响下高校医学图书馆职能建设，促进医学图书馆全面建设。

方法 以武汉大学图书馆医学分馆为例，从医学信息教育、医学信息研究、医学师生管理与服务三个层面进行探讨分析。

结果 一、关于医学信息教育，医学信息教育首先是指对于“教育的教育”，即培养师生的情报意识；其次是培养师生读者利用文献情报的技能。二、关于医学信息研究，文献资源建设既要强调医学类文献信息的馆藏，也需强调哲学社会科学及其它自然科学的馆藏；鼓励馆员结合岗位职责和涉医学工作实际开展学术研究；积极鼓励馆员走出去，向前辈取经，与同行进行学术交流。三、关于医学师生管理与服务，医学图书馆需要有针对性地面向医学师生群体开展相关管理和服务工作，对于图书馆而言，管理和服务是一体两面的，即管理是为了更好地服务，而服务又是管理水平高低的最终体现。

结论 武汉大学图书馆医学分馆在学校总馆的领导和指导下，并通过向兄弟院校医学图书馆的学习，在医学信息教育、医学信息研究、医学师生管理与服务三个方面进行了一定的理论探讨和一些有益的尝试，取得了一些经验，也有进一步的规划，认为在当前疫情形势下，就教育、科研、管理、服务四项职能而言，它们自身能够形成一个整体，因此从基本原则上看能够有效推进图书馆职能建设。

2019 年至 2020 年冬春之交，一场由新型冠状病毒引起的肺炎疫情在中华大地的腹部湖北武汉爆发，病毒无情，疫情最终在整个中国大地相继蔓延。在疫情的影响下，各行各业纷纷陷入停滞、关闭、转向的严峻局面。直至国家取得疫情防控阶段性和决定性胜利，整个社会形势才得以缓慢恢复元气并逐渐勃发出生机。然而，在疫情影响下，整个经济社会秩序已然与过去不同，如果说当前在国家长效防控机制下最主要的危险已经过去，那么如何面对机遇和挑战也就成为了我们新的课题。作为一名图书馆人，笔者在这里想谈一谈关于新冠肺炎疫情影响下高校医学图书馆职能建设的问题。

1 医学信息教育

传统上，我们会根据行政建制、部门设立，比如流通部、典藏部、阅览部、系统部、阅读推广部、学科服务部、办公室等来思考图书馆的部门建设问题，也会从馆员年龄构成、学历分布、性别比例，甚至智商和情商等方面来反思图书馆的可持续性发展问题，但鉴于医学图书馆的特性，我们在此选择从职能建设出发，即从医学信息教育、医学信息研究、医学师生管理与服务三个方面来讲，因为教育、研究、管理、服务是高校图书馆的共有职能，而立足于医学，则突出高校医学图书馆的专业个性。

医学信息教育，首先是指对于“教育的教育”，也就是教育师生读者应该接受医学信息教育，应该正确、有目的和系统性地获取医学信息，即培养师生的情报意识。其次才是怎样快捷、方便、高效地去获取医学信息，即培养师生读者利用文献情报的技能。人类历史上传承下来的文化知识和技能总量浩如烟海，只有会有效获取有效信息的人，才会在知识的海洋里沉浮自如、任意东西；否则，面对未知的信息大海，轻则频频呛水、顾此失彼，查不到所需信息或查不到前沿信息，重则欺水出祸，酿出学术不端问题。当然不同于医学教育是对于医学某专业知识的系统性传授和解惑的是，医学信息教育是对于医学专业领域情报图书资料知识的系统教学。具体到医学图书馆而言，主要就是对读者进行医学数据库检索、医学信息专题讲座、医学图书期刊推介及医学新生培训等各种教学、拓展活动。

武汉大学图书馆医学分馆凝聚了一批具有高级职称或具有硕士学位或长期在学科服务一线的馆员负责推进医学信息教育，具体承担医学数据库比如中华医学会数字化期刊数据库、PubMed，甚至包括 JCR、WOS 核心合集等的检索课教学，在考虑到新冠肺炎疫情的影响，还重点讲授新冠病毒及疫情的前沿研究，从而成为师生充分研读医学领域研究成果的快捷键；适时举办医学信息专题讲座，如新冠疫情阶段研究网络课程等，让读者得以在较短时间了解或理解某一医学领域的兴起、发展及研究前沿；在分馆负责人及办公室的具体实施下举办医学类外文原版教材及学术专著展示，在协助采编部门的同时，提升了医学信息专题教育工作，今后还计划举办关于灾难主题及医学人文主题的图书展示及教育活动。

2 医学信息研究

《普通高等学校图书馆规程》（教育部 2015 年修订印发）规定：“高等学校图书馆是学校的文献信息资源中心，是为人才培养和科学研究服务的学术性机构，是学校信息化建设的重要组成

部分，是校园文化和社会文化建设的重要基地。”规程有一点很明确，即高校图书馆是一个学术性机构，高校越要向前发展，图书馆越要体现其学术性。这里仍需强调的是，高校图书馆的学术性不应仅仅体现在它馆藏有多少的纸质或电子学术资源，更重要的是，专业技术岗位的馆员应该积极承担起学术研究任务，一方面学术研究能够对图书馆的发展和建设水平起到积极的评估作用，另一方面也有利于馆员自身的业务水平和专业能力的提升和发展。

武汉大学医学图书馆在医学信息研究方面具体做到三点：一、在文献信息资源长期的建设过程中，既强调医学类文献信息的馆藏，也强调哲学社会科学及其它自然科学的馆藏，当然前者为主，后者为辅。在新冠肺炎疫情影响下，为了促进对于医学师生的长期关怀和教育，为了促进从人文出发对于医学的研究，以及促进医学专业的人文情怀培养和温度散发，医学馆特别向校总馆申请新的馆藏地，开辟专门的医学人文借阅区，目前已精心挑选并入馆藏图书约 4，000 册，并计划后期不断更新、充实、提高，希冀建成湖北武汉乃至华中地区的医学人文教学与研究资源中心。二、鼓励馆员结合岗位职责和涉医学工作实际开展学术研究，对于馆员的研究思路、研究框架和研究成果予以推介，比如在馆内召开工作经验总结会、研究情况分享会等等，在对馆员予以充分肯定的同时，也提出一些可望可及的要求，鼓励他们刊发自己的综述性、评论性、研究性文章。目前，武大医学馆在高级职称人员、本科以上学历人员，尤其是中级职称人员占比上占有比较大的优势，医学馆正在形成一种能够让学术研究可持续性发展的的氛围。三、积极鼓励馆员走出去，向前辈取经，与同行交流。武大医学馆在强调常规的工作规范之外，非常强调对外学术交流，积极争取或利用各种机会总结、交流，尤其是积极参加各级图书馆学会及各级医学会医学信息领域的学术会议，比如馆内负责阅读推广的馆员一边长期从事实践，一边尝试提出自己的观点和立论；学科服务业务组的馆员利用各种场合分享自己工作岗位的研究心得；其它人员也在疫情期间积极参与校总馆书评活动，宣传环保，宣传天人合一、天人相生的理念。

3 医学师生管理与服务

关于“医学师生管理与服务”，涉及三个概念：医学师生、管理、服务。

图书馆的使用者或服务对象是读者，对于高校医学图书馆而言，主要就是医学师生。在以上论述中，我们也反复提到医学师生，在这里，我们需要对此概念进行明确。医学师生不同于一般的社会读者，也不同于高校其他专业的师生，他们具有自身独特的需要需求和性格心理特征。出于学制和临床实践等因素，医学师生时间紧张、对专业类文献资料和管理服务工作有更为精准的需求，同时忙碌的工作和学习也导致他们承受更多的身心压力。因此医学图书馆的管理和服务需要多考虑其特别对象。

管理，本义即主持或负责某项工作，使之顺利进行，其英文 manage 的拉丁语词源意为“操纵”，因此不难看出，管理是指一定组织中的管理者，通过实施计划、组织、领导、协调、控制等职能来协调他人的活动，使别人同自己一起实现既定目标的过程。简言之，如果先行假定服务是“人做事”，管理即“通过人做事”。然而，不管是“人做事”还是“通过人做事”，二者的目的和结果都是一致的，即“做成事”，而且二者的实施者都是一致的，即都是“人”。当然这

里的“人”有分别，但谁知道一处的管理者不是另一处的服务者呢，或者一处的服务者不是另一处的管理者呢？显而易见，管理和服务其实是一体两面，本质是相同的，它们都是通过自己或直接或间接的劳动，来达成一个相对共同的目标。这一层意义也包含在青年马克思对共产主义的规定里面：“共产主义……是通过人并且为了人而对人的本质的真正占有。”[1]而人的本质，并不是所谓先天、永恒不变的的，即通俗而言它并非“人之初，性本善”，亦非“人之初，性本恶”，而是通过后天的劳动、实践、斗争，所显现出来的一切社会关系的总和。对于这样的一切社会关系的总和，我们可以概括为“为人民服务”。

“为人民服务”是一个总纲，在医学图书馆，它即“为读者服务”，“为医学师生服务”。在这里，我们非但没有忽视管理，因为管理正式为了服务，而真正的管理就是服好务；同时，我们也极为重视并尊重服务，因为具有服务的意识、能力和热心才真正体现出高质量的管理，才是图书馆人的初心。

在校图书总馆的领导和指导下，武汉大学医学分馆——尤其在新冠肺炎疫情爆发以来——在常规管理和服务外推出了一些举措。比如，组织原版外文医学图书展，针对医学师生需求尽力购买并登记上网配套电子资源；办公室牵头启动疫情期间紧急事项预案，针对医学师生集中评职称、求职、评奖需要纸质论文收录引用检索报告问题进行学科服务人员有效调配，针对疫情通过腾讯会议在全校范围开展线上疫情防控知识竞赛，针对医学师生特别开设新的医学人文专题馆藏等等；流通业务组趁半开馆期间组织集中消杀、理架、倒架和进行馆藏约 40 万册图书的数据维护；阅览组、宣传推广组也结合疫情积极创新业务。

一言以蔽之，在依法依规搞好管理的同时，心里头的念想就是千方百计为医学师生服好务，并努力形成长效机制。

4 结语

图书馆作为社会教育、研究、服务的一部分，在事业单位中处于重要的地位，高校图书馆更是高校教育、研究的重要组成部分，同时承担着管理育人、服务育人的双重责任，因此，冠高校图书馆以“大学的心脏”毫不为过。而高校医学图书馆作为专门图书馆，既有公共图书馆和高校图书馆的共性，又具体承担着医学信息教育、医学信息研究、医学读者的管理和服务工作，因此医学图书馆的职能建设——尤其在新冠肺炎疫情影响下——面临着自身较为个性化的问题。我们学习兄弟院校医学图书馆的经验，同时尽力弥补客观上存在的不足[2]。顺应国家关于制定国民经济和社会发展第十四个五年规划的计划，武汉大学图书馆医学分馆也在校总馆的领导和指导下，初步提出本馆的“十四五”规划纲要。为了顺应延续性和量化标准，保持了总体框架不变；同时，也从理念和思路上对以上三个方面进行了一些初步的探讨和有益的尝试，并认为在当前疫情形势下，相关思考和举措能够有效推进医学图书馆全方位建设。当然，以上三个方面还远非高校医学图书馆建设的全部，比如除此之外还有社会服务、文化传承等等，但是就教育、科研、管理、服务四项职能而言，它们自身能够形成一个整体，因此从基本原则上看能够有效推进图书馆职能建设。

一孔之见，以就教于方家和同仁。

PU-025

基于微信小程序护理员管理平台的开发与设计

李俊[1]、郭秋月[1]、贺庆[3]、杨小霞[2]、陈鑫凯[2]、崔娟莲[2]、张江洁[4]、熊力[2,1]
1. 中南大学湘雅护理学院 2. 中南大学湘雅二医院
3. 中南大学湘雅医学院 4. 浏阳市仁杰电子科技有限公司

目的 在“互联网+移动医疗”快速发展的背景下，本文研究和设计基于微信小程序的护理员移动教学、管理平台，为护理员提供丰富的护理专业理论和技能的学习内容，改变传统的课堂教学模式。以打破传统教学中时间、空间的限制，让护理员充分利用碎片化时间进行学习，推动其专业化的发展以满足患者个性化需求和医院发展的需要。同时，通过平台为护理员提供更多的就业信息、渠道以及国家颁布的政策法规，进一步把提升护理员职业认同感，降低护理员队伍的离职率。

PU-026

国外全科医生发展经验对我国全科医生发展的启示

王凡、程艳敏
山东省医药卫生科技信息研究所

目的 近年来，全科医学越来越受到患者的重视，但我国全科医生仍然面临数量不足和质量不高的问题。全本文通过探讨研究我国全科医学的发展现状，同时借鉴国外全科医学发展经验，为全科医学的发展提出建议。

方法 通过文献查阅法、实地考察法、问卷调查法等方法研究我国全科医学发展现状，分析全科医学发展中的不足之处，借鉴国外全科医学发展经验，为全科医学的发展提出建议。

结果 导致我国全科医学发展缓慢的原因如下：一是全科医生待遇低，工作强度大，难以吸引优秀人才；二是全科医生职称晋升难，三是全科医生的培养力度不够，职业培训不到位。

结论 对我国全科医学发展提出来如下几条建议：切实提高全科医生待遇；完善全科医生考核机制，建立富有竞争力的职称评审体系；大力发展“5+3”培养模式。通过不断改革进步，使我国全科医学事业得到快速高效的发展。

近年来，全科医学越来越受到患者的重视，但我国全科医生仍然面临数量不足和质量不高的问题。本文从三个方面分析了产生这种现象的原因，一是全科医生待遇低，工作强度大，难以吸引优秀人才；二是全科医生职称晋升难，三是全科医生的培养力度不够，职业培训不到位。同时学习国外发达国家全科医生发展经验，对我国全科医学发展提出来如下几条建议：切实提高全科

医生待遇；完善全科医生考核机制，建立富有竞争力的职称评审体系；大力发展“5+3”培养模式。通过不断改革进步，使我国全科医学事业得到快速高效的发展。

关键字：全科医学、“5+3”培养模式

近年来，以需求为导向，全科医生越来越受到患者的重视，尤其是在分级诊疗和家庭医生签约制度实施以后，全科医生的地位在逐步提升。全科医生是综合性医学人才，需要掌握广泛的医学知识和综合的医疗技术，主要承担预防保健、常见病的诊疗、疾病初诊、慢性病的调养及康复、健康管理等综合化的服务，在基层医疗服务中发挥着十分重要的作用。因此，全科医生被称为居民健康的守护人[1]。

一、我国全科医生发展现状

全科医生的概念最早由中华医学会于 1988 年引入到中国，全科医生的形式也由国外主要服务于家庭的家庭医生转化成我国的社区医院医生和乡镇卫生院医生，持有资格证的全科医生能够独立开展临床工作，能够医治 80%-90%的常见疾病，对待重症疾病具备正确及时的转诊能力。截至到 2018 年底，据中国医师协会统计，我国已有大约 30 万名全科医生。2017 年底，这一数字为 25.3 万人，也就是说，在 2018 年一年的时间里，我国全科医生的数目增加了将近 5 万人，涨幅较前几年有所提升。但仍然不可忽视的是，这一数据距离实现 2030 年城乡居民实现每万人就有 5 名全科医生的目标还相差很大，在贫困地区，全科医生的缺乏问题则更为突出。

目前我国在岗的这些全科医生，有很大一部分比例是在“十一五”和“十二五”期间，国家支持基层医疗建设而对在职医生开展了一些培训，包括针对社区卫生服务机构医生的在岗培训以及乡镇卫生院医生的转岗培训，并不是通过 3 年的住院医师规范化培训产生，这种在职培训时间较短，培训内容覆盖面小。在农村以及西部偏远地区，情况就更为严重，许多乡镇诊所的医生很少或没有接受过系统正规的教育培训。

由此可见，我国全科医学发展面临着数量不足和质量不高两个大问题。导致这些问题产生的原因主要有如下几个方面：

1、全科医生待遇低，工作强度大，难以吸引优秀人才

我国现行的全科医生薪酬支付方式主要为工资制度，一般包括基本工资和绩效或奖金。全科医生岗位多集中在社区医院和乡镇卫生院，和大型医院相比，这些地方的医生所获得基本工资更低，绩效或奖金更少。同时和专科医生相比，全科医生所需要掌握的医学知识和技能更为广泛，全科医生所面临的病人种类繁杂，急诊、夜班更多，工作强度大。因此，在当前的制度下，全科医生的工作业绩不能得到应有的回报，不利于吸引优质人才。很多全科医生在更高等级的医院接受培训后不愿意回到原有的工作岗位进行工作[2]，待遇差已经成为阻碍全科医生发展的一块绊脚石。

2、全科医生职称晋升难。

当前我国事业单位的职称评聘制度以科研项目立项、发表论文数目、发表著作以及科研获奖等指标作为评价依据，这些条件容易统计，易于对比。殊不知，对于基层全科医生来说，项目也好，论文也罢，这些都不能全部体现一名医生的专业技术水平，基层全科医生工作十分忙碌繁琐，

并且全科医生不是专攻某种病情，见到的病患大多不是疑难杂症，申请项目、发表论文对于全科医生来说就是一件非常难的事情，在职称晋升上全科医生也不占据优势。

3、全科医生的培养力度不够，职业培训不到位。

全科医生培养力度不够，教学体系不完善，很多大医院都没有全科医学个科室，这就造成了很多全科医生无法进入到一些医院进行实习。虽然最近几年，我国逐渐加大全科医生的培养力度，培养模式转变为以“五年本科临床医学教育+三年规范化培训”（简称 5+3）为主，3+2 培养模式为辅的全科医生培养模式[3]，通过扎实的基础知识教育和规范化的实习培养，来提高全科医生的医疗水平。但我国全科医学起步较晚，基础相对比较薄弱，培养制度和体系不够完善。目前我国全科医生来源比较广泛，包括大专生、本科生和研究生，就业主要面向村镇定点医疗机构、社区医院和综合医院，针对不同的生源和不同的就业方向，当前的培养模式仍然比较单一，没有针对不同的生源和就业方向进行差别化的有针对的培养。

二、国外全科医生发展经验介绍

在全科医生待遇方面，国内全科医生待遇和国外相比有较大差距，在欧美等发达国家，医生是非常受人尊敬的职业，拥有较高的社会地位和劳动报酬，全科医生承担着保护居民健康的社会责任，给予全科医生和工作量相对应的劳动报酬是激励他们认真工作的保证。全科医生的报酬支付方式主要有以下几种：按服务支付、按病种支付、按人头支付、工资支付以及混合支付制度[4]。在美国的医生是高收入人群，美国全科医生待遇支付方式为商业医疗保险模式下的多种支付方式，全科医生和专科医生收入相差不大，2015 年美国全科医生年平均工资为 192120 美元，是人均收入的 3.98 倍[5]。德国为总额预算下的按服务支付制度[6]，政府按照工作量预算向卫生部门提供资金，医生根据服务量获取劳动报酬。这种报酬支付模式激发了全科医生向患者提供服务的动力。

在全科医生职称晋升方面，英国以 QOF 协议[7]（质量与效果协议）最为著名，该协议是一项通过量化医生的服务质量，制定一系列的指标，对这些指标赋予一定的分值，根据得分对医生服务质量进行评价。QOF 协议最初设置四大领域的指标，全年满分为 1000 分，包含组织领域指标(181 分)、临床领域指标(655 分)、附加服务指标(36 分)、患者体验领域指标(108 分)及整体护理方面(20 分)。在 2016 年，QOF 协议的指标又缩减为了两个，全年满分修改为 559 分，包含公共卫生领域指标（124 分）和临床领域指标（435 分）。通过对全科医生每项指标进行打分，获得全年总得分，该得分作为评价全科医生工作量和服务质量的依据。

在全科医生培养方面，北美等发达国家的全科医学发展较早，培养制度比较完善。美国家庭全科医生的培养生源从医学院本科毕业生中选取，培训包含三年的基础培训和一到两年的高级培训，基础培训主要在医院的各个科室进行轮岗实习，广泛的掌握各种病症相关的医疗知识。高级培训则主要是针对与家庭医生相关的内容进行培训，主要包括康复治疗、老年人保健、妇幼保健等项目。完成培训后，学员必须参加由行业协会组织的统一考试，才能够获得家庭医学专科医师资格。

三、我国全科医生发展的改进建议

1、切实提高全科医生待遇，吸引优秀人才。

引入混合支付制度，在现有工资制基础上，引入按服务支付、按病种支付和按人头支付制度。全科医生多分布于乡镇卫生机构和社区医疗机构，注重提高全科医生的基本工资待遇，在总额预算的前提下，实行按照病种、人头以及服务项目相结合的收费制度，激发全科医生的工作积极性。同时政府应当加大全科医生待遇投入，避免过多的经济负担转嫁到患者身上，让全科医生得到与工作量相匹配的劳动报酬。同等条件下，优先给与在一线村镇卫生所、社区医疗点的全科医生给与晋升机会，对于愿意到偏远地区，乡村地区服务的全科医生有政策上的倾斜，尤其是在收入和子女教育方面。

2、完善全科医生考核机制，建立富有竞争力的职称评审体系。

建立科学的富有竞争力的职称评审体系，使真正踏踏实实为百姓看病的医生得到应有的劳动报酬，建立科学、有效、公平的评聘指标体系是提高全科医生职业吸引力的基本保障[8]。不能简单的把门诊治疗次数、操作次数当做唯一的考核指标，而应当综合考虑服务区内居民的实际健康状况、实际工作量，建立能够衡量全科医生专业技术水平的综合评价指标体系.参照 QOF 协议，结合我国全科医生的实际状况，设置综合评价指标体系，包括数量指标和质量指标，不仅要评价医疗服务的结果，更要注重医疗服务的过程，以提高全科医生医疗服务质量为目标，真正衡量出全科医生的专业技术水平。具体说来，将全科医生日常的工作积累作为量化指标，包括接待门诊次数、看病种类、诊治疗效、患者满意度、签约居民数量等指标作为业绩考核的重要依据，切实提高全科医生的队伍质量。

3、大力发展“5+3”培养模式。

我国近些年开始开展“5+3”培养模式，即五年的基础医学理论学习+三年规范化全科培训[9]，这个培养模式和美国的全科医生培养模式有些类似，但由于开展时间较短，很多方面还不成熟，需要医学院校和综合医院的积极配合。医学院校应当重视全科医生专业的设立，在本科教学阶段加强学生的全科医学基础知识，注重实践教学和临床应用，在本科毕业之后，学校应当为毕业生规培及就业提供帮助和指导，帮助学生联系规培医院，在规培期间落实全科医生的规培待遇，保证“5+3”培养模式的连续性，保证全科医生招的来，留的住，逐步取代原有的学历较低、培训不完整的全科医生。

PU-027

数据分析和患者细分在临床科室管理中的应用研究

牛承志[1]、赵丹[2]
1. 郑州大学第一附属医院 2. 郑州大学

目的 为了实现互联网+时代医院临床科室精准医疗服务，有必要对临床科室进行数据分析和患者进行细分并实施有效的干预管理。本文以国内某大型三甲医院为研究背景，抽取某临床科室的相关数据，提出了 ARF3T 模型（基于 RFM 模型）患者细分模型。对提出的 ARF3T 模型和聚类算法进行了验证和分析，对患者分组数据做了较为详尽的分析，针对不同的分组提出了相应的管理建议。

方法 本文使用了机器学习中的聚类算法，使用密度峰值作为聚类的初始中心，把密度峰值得到的初始的聚类中心作为 K-means 算法的初始聚类中心，并使用这种改进的 K-means 聚类算法对甲状腺患者进行细分。在数据整理中使用了层次分析法对各项指标进行权重分析，使用手肘法确定了聚类算法的 K 值，最后使用了计算了轮廓系数对聚类进行了校验。

结果 数据分析和患者的细分对于实现互联网+时代下的精准医疗服务和个性化治疗及临床科室的管理起到辅助决策作用，文中所做的研究具有一定的理论和实践意义。以国内某大型三甲医院两年的甲状腺疾病患者为研究背景，通过 ARF3T 模型和聚类算法将患者细分为 4 组，经过相应的实验检验了甲状腺疾病患者细分模型和聚类的有效性。最后通过数据分析，为诊治科室和治疗医师针对不同的患者群制定相应的医疗服务策略，实现患者的精准管理。

结论 本文探索临床科室运营指标的建设和患者细分管理的方法，其目的把医护人员从繁琐的运营管理中解放出来，把有限工作时间还给医护人员，让专业的人做专业的事，从而实现临床科室的运营效益的最大化。

1 背景

国务院办公厅日前印发《深化医药卫生体制改革 2020 年下半年重点工作任务》指出公立医院综合改革是医改的一项重要内容，自 2009 年医改启动以来，国内大型综合医院的发展模式急需从粗放扩张型向集约效益型的转变。积极践行现代医院管理制度，提升医院运营管理精益化水平，是公立医院实现转型发展的必由之路。[1]本文以郑州大学某临床科室为研究对象，采用数据分析和机器学习的技术手段，探索临床科室运营指标的建设和患者细分管理的方法，其目的把医护人员从繁琐的运营管理中解放出来，把有限工作时间还给医护人员，让专业的人做专业的事，从而实现临床科室的运营效益的最大化。[2-3]

2 数据分析

2.1 指标建立

临床科室的业务指标主要包括四个维度，即财务维度指标、工作量维度指标、考核维度指标、流程维度指标。财务维度指标主要包括住院人均费用 、住院患者费用比例、住院患者耗材比例、

科室各项费别占比 、住院患者药占比、住院患者抗菌药物占比、住院收入增长率、门诊收入增长率、门诊患者次均费用等；工作量维度指标包括出院人次、门诊诊察人次、入院收治人次、重症患者人次、一二级手术人次、三四级手术人次、收治患者地域分析、医保患者分类人次等；考核维度指标主要包括合理用药临床路径相关指标、抗菌药物临床路径相关指标、临床路径相关指标、单病种相关指标、重症相关指标等；流程维度指标主要包括病房周转率、术前平均天数、平均住院日、床位占用率、不良事件上报率等。

2.2 数据的获取

临床科室的运营数据的获取主要依托于医院的 HIS、EMRS、LIS 和 PACS 四大系统，经过数据的采集、计算、存储等步骤，这些步骤可以依托医院的临床数据中心来完成。通过医院的临床数据中心实现医院各系统的数据的集成和加工，并提供相应的对外数据调用服务，通过相应的技术手段，实时进行临床科室的针对性采集。此外，可以通过临床数据中心形成专题数据库实现临床科室的数据的回溯性采集，保证了数据采集的实时性、准确性和完整性。

2.3 数据的加工

对从临床数据中心抽取的数据，可以通过医院的辅助决策系统实现对抽取数据的二次数据加工，例如临床科室的抗菌药物、临床药物、临床路径、单病种的相关数据的专题抽取，将相关的计算算法写入预设的辅助决策系统中，把不同的计算请求通过不同的节点进行计算，从而完成临床科室运营指标的运算，以流式数据的方式对临床运营数据实现实时计算能力。

2.4 数据的可视化

临床科室的运营指标分为科室级、治疗组级、个人三个层次，通过可视化的图标来支持临床科室的精细化管理、医疗质控及财务指标等多层级的需求，以临床科室整体运营指标为中心，从不同的角度，针对不同的内容实现数据的可视化展示。临床科室的可视化展示囊括了医院及临床科室关注的各项数据：门诊运营指标、住院运营指标、抗菌药物检测指标、感染类指标、治愈率指标、临床路径指标等其他数据。

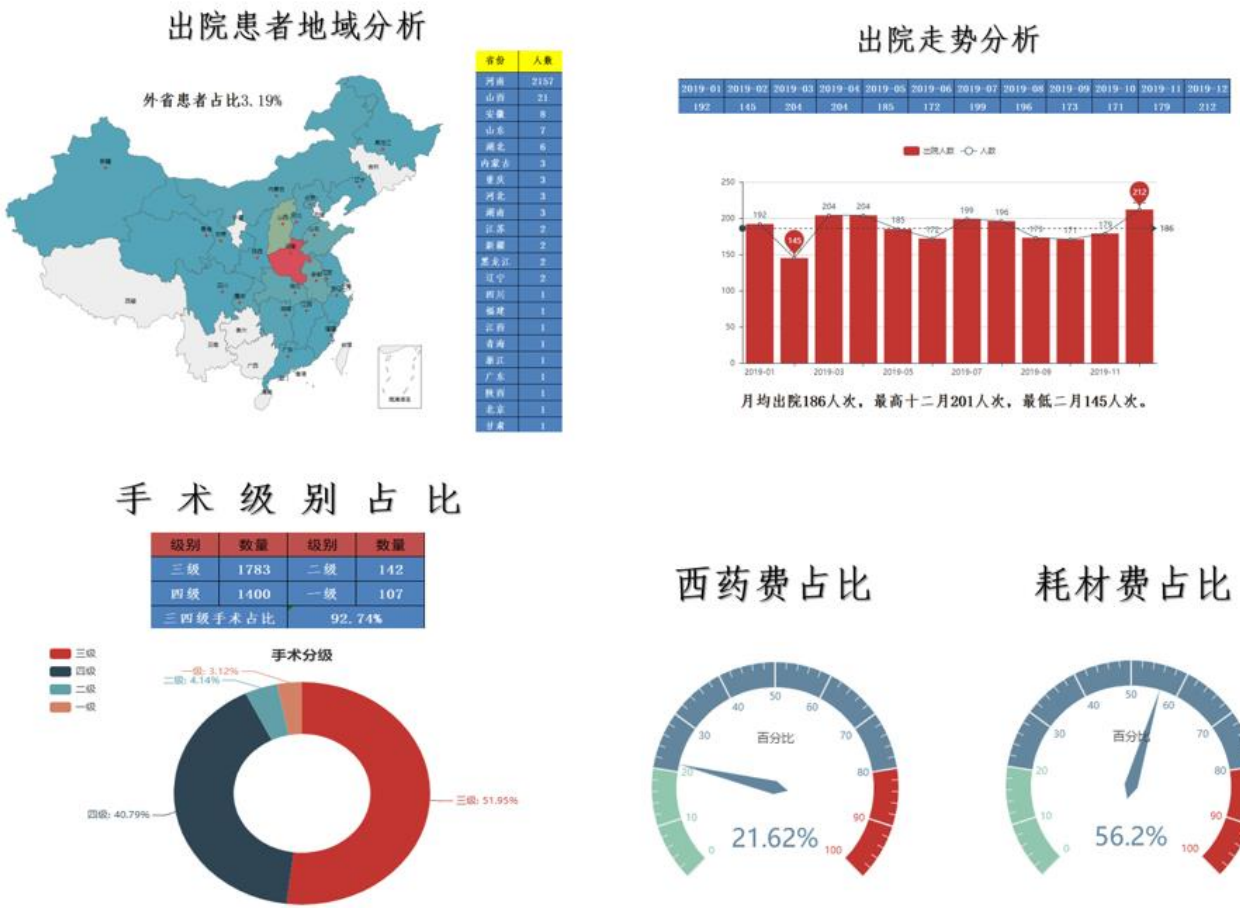


图 1 临床科室部分指标

3 临床科室患者细分

随着互联网医疗的不断深入，患者的就医习惯也在不断的发生变化，临床科室应当逐步树立品牌意识，实现患者的精细化管理，为不同的患者提供不同的个性化诊疗方案，对患者的进行有效的管理。本文以甲状腺外科为研究基础，借用商业领域的客户细分方法，对甲状腺外科患者进行患者聚类，依据患者的甲状腺检验指标，将具有相似特征的患者聚集为一簇，形成相应的患者分类，实现对甲状腺患者的精准管理。

3.1 患者细分模型的建立

参考商业领域的 RFM 模型，针对甲状腺患者的特点，本文提出了甲状腺患者管理的 ARF3T 模型。其中，A（Age）表示甲状腺疾病患者的年龄；R（Recency）表示甲状腺疾病患者最近一次就诊时间和当前时间的时间间隔（以月为单位）；F（Frequency）表示甲状腺患者在某个时间段内的就诊频率；3T 表示甲状腺患者最近就诊的三项检验项目，是反映甲状腺生理功能状态的指标，分别是三碘甲状腺原氨酸（Triiodothyronine, T3）、甲状腺素（Thyroxine,T4）、促甲状腺激素（Thyroid stimulating hormone,TSH）。这三项指标是临床上判断甲状腺功能状态的常用实验室指标，对甲状腺疾病的类型判断、疗效监测、病情评估都具有重要价值，直接影响临床医生的诊断及用药。有研究表明，游离甲状腺素水平的降低与慢性肾脏病的严重程度有关[5]，甲状腺激素及其代谢物与肝病的治疗密切相关，高水平的游离甲状腺素可以预测晚期肝癌患者的预后，游离

三碘甲状腺原氨酸与游离甲状腺素的比值是晚期转移性结直肠癌的一个新的预测指标，同时说明甲状腺激素代谢与肿瘤进展及癌症患者生存之间有着重要的联系[6]。因此，作为临床诊断和鉴别疾病的重要指标，这三项指标对实现甲状腺患者的精准管理有着重要的意义。ARF3T 模型中各项指标计算公式如下：

A = now(t) - birthday

R = now(t) - last(t)

F = count(the time of visiting doctors)

T3 = last(T3)

T4 = last(T4)

TSH = last(TSH)

其中 now(t)表示现在研究的时间节点，birthday 表示患者的出生年月，now(t) - birthday 求出患者的年龄；last(t)表示甲状腺疾病患者的最后一次就诊时间；count(the time of visiting doctors)表示甲状腺疾病患者的在某一时间段内就诊次数；last(T3)、last(T4)、last(TSH)甲状腺疾病患者最后一次三碘甲状腺原氨酸、甲状腺素、促甲状腺激素检验数值。

3.2 患者细分模型的建模流程

ARF3T 模型是对甲状腺疾病患者实现精准管理的模型，ARF3T 模型的建模过程主要包括三个步骤：

⑴数据的预处理和特征的抽取

抽取甲状腺疾病患者的就诊记录和检验数据进行抽取并进行清洗，提取 ARF3T 模型计算指标；设定计算指标参考数值，进行数据筛选，以患者卡号为主键组织数据子集，得到甲状腺疾病患者 ARF3T 模型的各项数据。

⑵甲状腺疾病患者的聚类

对甲状腺疾病患者数据集进行聚类分析，得到甲状腺疾病患者的聚类结果。

⑶甲状腺疾病患者的细分

通过对甲状腺患者的聚类，结合临床医生的相关临床经验，制定相应的细分规则，根据细分规则计算得到最终的甲状腺患者的分层管理结果。

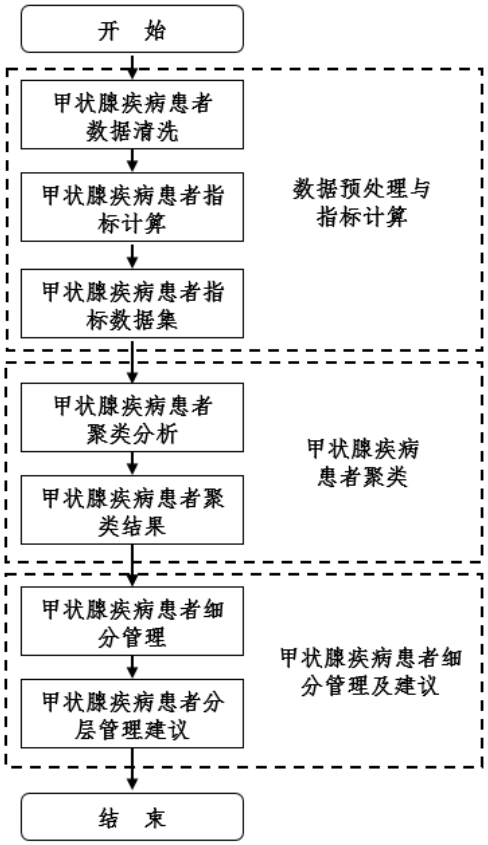


图 2 患者细分模型的建模流程图

3.3 应用实例

3.3.1 甲状腺疾病患者细分

通过医院信息系统获取到甲状腺外科在 2016 年～2018 年之间甲状腺外科门诊所有就诊甲状腺疾病患者就诊记录，主要包括患者挂号表、患者诊断表、患者检验记录表、患者基本信息表等。为了实现甲状腺疾病患者的精准细分，需要对甲状腺疾病患者数据进行如下处理：

3.3.1.1 数据清洗

2016～2018 年甲状腺疾病患者人数有 5864 人次，就诊数据和医学检验数据有 50 万条记录，需要对抽取的数据进行清洗：对存在异常值和缺失值的数据首先要做清洗，如患者无检验数据、出生年月为空、就诊记录为退费记录、只有挂号记录没有处方记录的数据；第二步对重复数据的处理，如一些患者存在多条同一时间内多条就诊记录，需要核对患者的诊断是否为甲状腺疾病。同一患者一天内做多次检验，需要抽取最后一次的结果；第三步对抽取的数据进行整合，把患者的诊断数据、就诊数据、医学检验数据按照患者的就诊卡号进行整合，抽取整合到一张数据表中。整个数据的清洗过程使用数据结构化查询语言（Structured Query Language，SQL）进行数据的清洗工作。

3.3.1.2 数据整理

依据本文提出的 ARF3T 模型，经过数据清洗，共抽取了 5888 条甲状腺疾病患者的就诊数据、基本信息、医学检验数据，整合数据之后，共得到 5682 名甲状腺疾病患者的数据，具体的数据格式如表 5 所示：

患者卡号	F	FT3	FT4	TSH	AGE	R
**14598950	4	6.06	14.14	0.01	18	96
**14442879	15	4.74	13.11	0.14	49	112
**08617037	4	5.91	15.26	0.11	57	15
**12845493	37	5.5	14.74	0.005	49	2
**13029594	100	5.61	9.68	23.85	46	7
…	…	…	…	…	…	…
**12789037	17	4.7	9.55	11.37	54	29

表 1 ARF3T 模型部分数据

3.3.1.3 甲状腺疾病患者细分

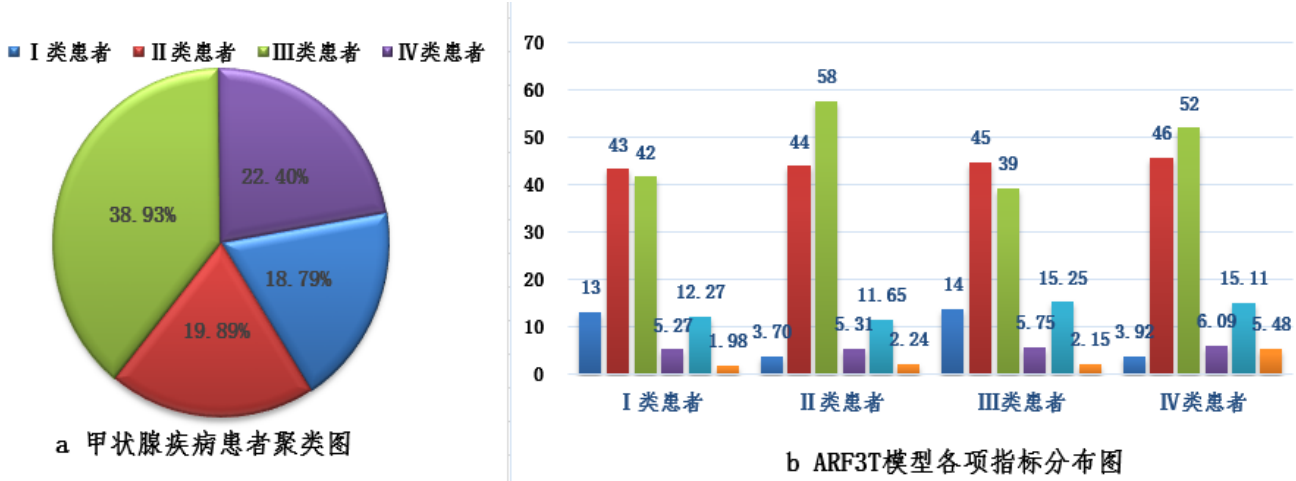
对甲状腺疾病患者数据进行标准化和加权处理后，使用改进的 HKNNDP_K-means 算法，通过手肘法的核心指标 SSE(sum of the squared errors，误差平方和)来确定聚类的 K 值。SSE 的计算公式如下：

$$SSE = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2$$

公式中C_i是第 i 个簇，p 是C_i中的样本点，m_i是C的质心（C中所有样本的均值），SSE 是所有样本的聚类误差，代表了聚类效果的好坏。

使用整理过数据，结合甲状腺外科和内分泌医学专家意思，本次实验聚类 K 值在预定在 3～5 之间取值，使用手肘法来计算 SSE 值，具体做法是让 K 从 1 开始取值直到上限 8，对每一个 k 值进行聚类并且记下对应的 SSE，然后画出 K 和 SSE 的关系图，最后选取肘部对应的 K 作为最佳聚类数。通过计算并结合专家意见，确定最佳聚类数目 K=4。

在确定 K 值为 4 之后，将经过数据整理之后的 ARF3T 模型数据，通过 K-means 算法进行聚类，得到 4 组甲状腺疾病患者数据，如图所示。



a 甲状腺疾病患者聚类图　　b ARF3T模型各项指标分布图

图 3　患者细分模型结果

3.4 聚类结果分析及患者管理建议

将 ARF3T 模型数据聚类后，得的Ⅰ类患者群、Ⅱ类患者群、Ⅲ类患者群、Ⅳ患者群四类甲状腺疾病患者群。

⑴Ⅰ类患者群 将该类甲状腺疾病患者定义为依从性较好患者群，这类患者关注自身健康，年龄在 20～40 之间，三项实验室指标都正常值范围，遵从医嘱按时复诊，及时主动和医生沟通病情。Ⅰ类患者群有患者 1140 名，约占甲状腺疾病患者的 18.79%，该类患者病情稳定，关注自身健康。在提供预约就诊的提前与患者进行精准预约，实行按时段就诊为Ⅰ类患者群提供精准医疗服务，如定时推送患者相关疾病的健康知识，通过小程序、APP 等患者及时推送检查检验结果并给出相应指导意见，定时针对所患疾病进行健康宣教

⑵Ⅱ类患者群 将该类甲状腺疾病患者定义为依从性较差病情稳定患者群，这类患者最近的一次三项实验室指标都正常值范围，但不遵从医嘱按时复诊，未按照诊疗计划及时与就诊医生进行有效沟通。Ⅱ类患者群有患者 1207 名，约占甲状腺疾病患者的 19.89%，此类患者病情较为稳定，对自身健康关注不够。对此类患者在约定复诊时间内提醒患者按时复诊，就诊时登记相关信息着重登记个人信息，方便推送针对患者个人的健康知识和不同的专家的坐诊时间、专业方向，方便患者根据自身情况选择就诊时间和相关专家。

⑶Ⅲ类患者群 将该类甲状腺疾病患者定义为依从性较好病情波动患者群，这类患者对自身健康较为关注，遵从医嘱及时复诊，及时主动和就诊医生沟通病情，但三项实验室指标至少有一项不在正常值范围内。Ⅲ类患者群有患者 2362 名，约占甲状腺疾病患者的 38.93%，是四种患者分类中占比最大的患者，此类患者关注自身健康，三项检验指标存在异常。此类患者群是占比最大的患者群，对此类患者可以提供精准的预约服务，实施按时段精准的就诊，对患者就诊数据进行数据集成，形成诊断患者的个人健康档案，生成个人健康报告，形成多学科多专业的患者健康数据分析报告，以小程序、手机短信、APP 等多种形式为患者提供精准的医疗服务。

⑷Ⅳ类患者群 将该类甲状腺疾病患者定义为依从性较差病情波动患者群，这类患者最近一次的三项实验室指标至少有一项异常，不遵从医嘱及时就诊复查病情，未按照诊疗计划及时与就诊医生进行有效沟通。Ⅳ类患者群有患者 1359 名，约占甲状腺疾病患者的 22.40%%，此类患者对自身健康关注不投，三项检验指标存在异常。通过记录患者微信、手机号码或采取人工随访的方式了解患者情况，可以采用问卷调查方式获取患者体征情况，提醒患者按时复诊。以小程序、手机短信、APP 等多种形式推送专家就诊时间和专业领域，使患者及时了解自身病情并提升复诊率。

4 结束语

数据分析和患者的细分对于实现互联网+时代下的精准医疗服务和个性化治疗及临床科室的管理起到辅助决策作用，文中所做的研究具有一定的理论和实践意义。以国内某大型三甲医院两年的甲状腺疾病患者为研究背景，通过 ARF3T 模型和聚类算法将患者细分为 4 组，经过相应的实验检验了甲状腺疾病患者细分模型和聚类的有效性。最后通过数据分析，为诊治科室和治疗医师针对不同的患者群制定相应的医疗服务策略，实现患者的精准管理。

PU-028

健康信息素养在新型冠状病毒肺炎疫情防控中的作用

黄尤江、郑利荣、 贺莲、苏焕群
广东省医学学术交流中心（省医学情报研究所）

目的 探讨健康信息素养在新型冠状病毒肺炎疫情防控中的作用。

方法 从信息健康素养内涵，新冠肺炎信息传播特点及遇到的问题进行了讨论。

结果 信息传播在新冠肺炎防控中起到了重要作用，而个人健康信息素养对疾病预防控制和自身健康受益有着积极影响。

结论 强调了健康信息素养对疫情防控的重要性，指出应加强弱势群体的健康信息素养能力的提升，同时增强基层情报单位的信息服务建设。

新型冠状病毒肺炎（COVID-19）是由 2019 年发现的新型冠状病毒（SARS-CoV-2）引起的一种肺炎，具有传播力强、大多患者预后良好等特性。COVID-19 疫情已成为国家重大公共卫生安全突发事件，我国新冠防控工作已取得成效，但国际上仍呈大流行趋势。随着经济和信息技术的发展，我国居民对健康的要求越来越高，国家层面也提出了“健康中国战略”，而信息技术的应用，使得医疗服务简易化、人性化、智能化，获取健康信息途径多样化，加速了医疗卫生的发展，对健康促进意义重大。

1 健康信息素养的内涵与作用

1.1 健康信息素养的内涵

1974 年 Simonds SK 首次提出了健康素养的概念[1]，目前广泛定义为：个体具有获取理解和处理基本的健康信息或服务，并运用些信息和服务做出合理的健康决策的能力，是影响健康的重要因素之一[2]。健康素养包含基本的健康知识与技能和健康信息素养（Health Information Literacy，HIL）两个层面，而健康信息素养是健康素养的核心。信息素养是指认识到信息需求，并有效地获取、评价和利用所需信息的能力[3]。HIL 是健康素养和信息素养两个概念的渗透与融合，其内涵是指了解和认识健康信息需求，熟悉可能的信息资源并以此来检索有价值的信息，评价信息的质量以及在具体情况下的适用性，分析、理解并利用信息做出合理健康决策的一系列能力。美国医学图书馆学会（MLA）于 2003 年成立了健康信息素养专项组，这一行动使得卫生健康领域信息素养的研究对象发生了改变，由医学生、卫生保健提供者、卫生管理决策者向公众和患者过渡[4]。

1.2 健康信息素养的作用

世界卫生组织（WHO）预言，信息是通往健康的必经之路，健康信息素养成为公众健康促进的一个关键点。健康信息素养影响公众对健康和疾病的了解和认识，以及对健康促进和预防保健服务的参与度，也影响着疾病治疗的临床决策以及自我疾病管理效果。随着循证医学的发展，人们对临床决策的参与度越来越高，而不仅仅是治疗和护理的被动接受者，未来健康信息素养对健康素养的促进作用将越来越重要[5]。已有研究证实，获取和利用健康信息的能力与患者疾病负担密切相关，健康信息素养能够对公众健康维持和患者的治疗护理起到积极作用，并可作为解决健康不公平问题、医疗资源不均衡的有效途径，在发展中国家甚至还可能是缓解贫困的一项重要战略。

2 新冠肺炎信息传播特征

2.1 信息传播途径多样化

信息传播途径的迭代是新型冠状病毒肺炎疫情健康资讯的突出特点。随着信息技术的发展，人类社会进入全媒体时代，自媒体技术发展多样化，社交媒体广泛使用，为疫情信息传递奠定了基础。相比 2003 年 SARS 疫情爆发，疫情信息的扩散程度完全处于不同的数量级，曾经是电视、报纸为主要传播方式，而当前则是各种网络社交媒体，成为了各类疫情信息传播的“主流”平台，其中微信及其平台公众号、微博、抖音等各类热门短视频平台，几乎在全程“直播”疫情。调查显示[6]，民众获取疫情相关信息和疾病知识来源主要为微信、网站和电视。各种社交媒体为公众提供了一个全面的疫情防控信息传播平台，加速了新冠肺炎健康信息和知识的传播效率，为新冠肺炎的防治提供了信息支撑。

除了大众媒体对疫情信息的传播外，国家及各省市官网、医疗机构、医务人员、医疗专业媒体、医学信息研究机构也对新冠肺炎疫情信息和相关研究信息进行了推送。国家科技图书文献中心（NSTL）、丁香园、生物谷等机构均设立了新冠肺炎疫情专题，对疫情研究信息进行监测。我中心亦对新冠肺炎信息进行了监测，监测内容包括诊疗方案、机制研究、疫苗研发、国内外重大

事件及动态等多个方面，但较上述机构其浏览量有一定差距。此外，中心开通了新冠肺炎相关立项查新及成果鉴定绿色通道，以及期刊论文发表快速通道，在疫情情报工作中发挥了作用。

2.2 信息需求与生活需要相一致

信息的存在本身就应为生活服务，疫情期间大众对 COVID-19 相关信息的需求与其生活状态息息相关。一项关于 COVID-19 的信息流行病学研究显示[7]，在 COVID-19 早期，百度搜索中相关主题主要有“医疗救助”、“防护物资”、“流行病学进展”，随着复工复产统筹推进，高频检索词有“返工”、“返校”、“远程办公”、“在线教育”、“在线医疗”，反映了大众对 COVID-19 健康素养的需求以及针对疫情防控政策变化的应对方式。

2.3 健康信息素养能力决定信息价值

一个健康素养高的人能够准确评估自身健康状况、主动探寻健康知识和掌握相关健康技能，充分合理地利用当前医疗卫生资源、主动践行良好的个人行为与生活方式并进行自我健康管理，进而能够保护和促进自己及身边他人的健康[8]。单纯地讲，新冠肺炎健康信息的网络化传播本身是双刃剑：一方面，能够帮助公众获得疫情知识，提高对疫情认识，降低感染发生风险；另一方面，长时间暴露在单一的疫情信息中可能导致公众发生焦虑等心理健康问题[9]。一项关于 COVID-19 爆发期间心理健康问题和社交媒体暴露（SME）之间的关联性研究显示[10]，武汉 COVID-19 爆发期间抑郁、焦虑及两病共存的患病率分别为 48.3%、22.6%、19.4%，参与者心理问题与 SME 频率呈正相关。但需要明确的是，具备获取信息能力不代表着有较强的健康信息素养，相反，高健康素养水平有助于提高公众心理求助态度和求助意愿的形成，从而预防心理行为问题的发生[11]。

对于数字资源的利用，不同人群之间存在数字化程度不均现象，年龄/老年人、教育或文化程度、互联网应用能力与之息息相关，能更有效利用互联网信息者则更有可能使自己的健康在疫情中获益[12]。关于湖北省武汉市居民 COVID-19 跨部门经验数据显示[13]，农村与城市居民在行为意图，主观规范，对预防行为的了解，以及人际关系和媒体资源应用的多样性方面不存差异性，但不同的信息评估能力可能导致其执行预防行为的差异性，农村居民可能对疾病预防持消极态度。因此，有必要加强老年人、儿童及低教育水平等弱势群体的健康信息素养能力。

在自媒体、全媒体时代，用户拥有充分的自由去发布和传播各种信息，但部分用户信息素养不高，进而导致信息泄露、信息侵权、信息造谣等现象层出不穷，其原因不仅在于该类用户对利益的追求，以及虚拟身份的掩饰和信息监管力度不够[14]，而且接受信息的用户也存在信息素养不足的情况。为了解决这些问题，有必要加强政府监管和引导、深化自媒体的改革与发展、提升自媒体用户自身的信息素养水平。疫情期间出现了各种造谣或虚假信息，对应也有媒体、专家和官方进行辟谣。所谓谣言止于智者，这种“智”则体现了高水平健康信息素养的积极效应及重要性。

3 讨论与看法

公共健康信息素养、健康素养是信息素养在医学领域发展的例证，信息素养研究已开始向专业层次开拓[15]。新冠肺炎疫情来袭，不仅需要有过硬的国家应急救助能力和医疗水平，民众的抗

疫参与度与疫情防控密切相关，民众卫生习惯、生活方式及健康素养对疫情防疫工作有着重要影响[16]。民众健康素养不过硬，对新冠肺炎认识不够深，将对疫情防控工作会起到负性作用，疫情期间不乏一些人不配合防疫工作，往往需要道德束缚、媒体谴责或法律追责才认识到自己的错误和不足。因此提升公众健康信息素养尤为重要。

新冠肺炎疫情发生于互联网+、大数据、中国社交媒体快速发展的背景下，各媒体提供了一个实时、海量，但质量有待评估的新冠肺炎健康信息平台，而良好的健康信息素养将对个体健康行为产生积极作用。提升公众健康信息素养，首先应强调个体，“健康中国战略”意见指出：“每个人是自己健康第一责任人”，再者要强调“全民参与、共建共享、全民健康素养”，在抗疫过程中，应强化跨部门协作，鼓励和引导单位、社区（村）、家庭和个人行动起来，形成政府积极主导、社会广泛动员、人人尽责尽力的良好局面，实现健康中国行动齐参与[17]。在信息传播方面，政府机构是疫情处置主体，须信息公开与科普并重；对于医疗系统，包括医院及信息研究机构在内，虽然专业性强而非通俗性信息，但可以为有较高健康信息素养水平的人群或特定群体提供信息支持；而社交媒体也需明确在提高公众有关新冠肺炎疫情的健康信息素养的重要职责。对于老、幼、低学历等弱势群体则应重点关注，根据其人口学特征采用有特色的健康教育模式。

总的来说，社交媒体对新冠肺炎疫情信息的传播为疫情防控提供了海量的信息基础，而良好的健康信息素养为疫情期间个人及其周围人群健康提供了深层次健康保障。面对疫情形势的严峻性和复杂性，信息源方和信息接收方都应对各种信息进行评估，确保信息的真实性和有效性，即使是基层信息研究单位，也应建立疫情信息专题，为有需要的用户提供客观数据。不同人群健康信息素养水平是有差异性的，调查显示，仍有较大比例居民健康素养水平较低，应该加大宣传力度且有针对性地对健康素养水平薄弱群体的知识盲点和短板进行健康教育[18]。从全民健康素养促进的角度出发，中国公众健康信息素养仍有很大的提升空间。在新冠疫情及其常态化期间，公众应提升个人健康信息素养，理性看待各种疫情信息，调整好形态，配合疫情工作，以乐观、自信的态度抗击疫情并战胜疫情。

PU-029

在线健康社区知识交流效率研究

张亚飞、贺培凤
山西医科大学

目的 在构建评价指标体系的过程中考虑知识本身特征、点赞、收藏等因素，以期准确测度在线健康社区知识交流效率，完善在线健康社区知识交流研究理论体系。

方法 以丁香园论坛的选定的 18 个版块 2010-2019 年的知识交流数据为例，在已有知识交流效率评价指标体系基础上，将反映知识特征的知识密度和知识粘性纳入投入指标，点赞数和收藏数纳入产出指标，采用 SBM 模型对其知识交流效率进行测度。

结果 在线健康社区 18 个版块的知识交流效率均值为 0.71，10 年间变化平稳，仅有科技动态、心情驿站的知识交流效率达到决策单元有效，其余版块均未达到决策单元有效；纯技术效率较低是影响在线健康社区知识交流效率的主要因素。

结论 新的知识交流效率投入产出评价指标体系可以全面准确的测度在线健康社区各版块的知识交流效率，完善了在线健康社区知识交流研究的理论体系。社区管理者应注重社区制度建设以及内部管理水平等技术层面的提高，进一步发挥其规模优势，努力实现技术有效才有可能达到整体效率有效。

随着互联网和信息技术的迅速普及，网络信息呈现出爆炸式的增长趋势，基于互联网的社交媒体快速流行使得信息的传播变得更加高效和便捷。根据中国互联网信息中心（CNNIC）于 2020 年 4 月发布的第 45 次《中国互联网络发展状况统计报告》显示，截止 2020 年 3 月，我国网民规模达 9.04 亿，较 2018 年底增长 7508 万，互联网普及率达 64.5%，较 2018 年底提升 4.9 个百分点[1]。我国网民搜索内容中的 73.81%和“健康与医疗”相关，位于所有搜索主题之首[2]（2019 年第一季度），我国网民对于医疗健康信息的需求不断上升。在互联网与健康医疗互相渗透的背景下，人们获取健康信息的方式不再局限于传统的面对面医患交流，越来越多的人群开始在新兴的社会化媒体平台上获取、分享、交流与健康相关的信息，一大批在线健康社区（online health communities, OHCs）应运而生[4]。

在线健康社区的核心功能之一就是知识交流，知识交流过程包含四要素：分别是知识来源、参与主体、知识本身、知识交流的环境影响[5]。知识的流动促进了知识的转移、共享和创新，探索知识交流效率的影响因素可以揭示知识在用户间流动的规律，为知识交流效率提升提供最基本的依据[6]。通过评价在线健康社区的知识交流效率，分析影响在线健康社区知识交流效率提升的因素，有利于推动更加深度的知识交流，对促进健康信息共享和知识创新具有深刻的意义。

一 数据与方法

为了全面真实的了解在线健康社区中的知识交流效率情况，本文以国内最大的专业医疗互联网平台—丁香园为例，基于 Python 的 Scrapy 框架搭建网络爬虫程序，从临床医学、药学等 10 个区中根据发帖量和回帖量交流体量相对较高的 18 个板块（具体选取版块见表 1）2010-2019 年的发帖、回帖等知识交流相关数据作为数据源，并对获取到的数据进行预处理，形成在线健康社区知识交流效率研究的基础数据。

本文采用数据包络分析（DEA，Data Envelopment Analysis）方法对在线健康社区知识交流效率进行测算，首先在前人构建的虚拟学术社区知识交流效率评价指标体系的基础上，将反映知识特征的知识密度和知识粘性纳入投入指标，点赞数和收藏数纳入产出指标，采用 SBM 模型对在线健康社区知识交流效率进行测度，以期深入了解在线健康社区的知识交流现状,为促进在线健康社区中的知识交流提升提供参考依据和合理化建议。

表 1 在线健康社区选取版块表

社区/学科	选取版块
人文医学	卫生法律人文、医学哲学和医学史、卫生管理
临床医学	心血管、呼吸胸外、危重急救、中医
基础医学	形态学与生理生化、生物信息学
公共卫生与预防医学	预防医学、循证医学
药学	新药信息
实验技术	动物与组织学技术
科研与学习交流	论文写作与投稿
考试交流	临床医师考试
信息发布区	科技动态
检索知识与求助	检索知识
休闲区	心情驿站

二 指标体系构建

通过对现有文献进行梳理，众多学者对在线社区的知识交流效果展开研究，并建立了对应的评价指标体系，比如万莉[7]和吴佳玲等[8]以虚拟学术社区的为研究对象，以小木虫论坛和人大经济论坛中的用户数量、发帖数量作为投入指标，浏览数量和回帖数量作为产出指标；晋升[9]在前人研究的基础上，将小木虫论坛的用户数、发帖数、讨论时间作为投入指标，浏览数、回复数、再回复数作为产出指标。综上所述，现有研究对于在线社区构建的评价指标体系并未考虑知识本身的特征对于知识交流效率的影响，而王俭[10]对在线评论的知识交流效率评价过程中，将知识密度、知识距离和知识粘性纳入指标指标体系，并且随着在线健康社区的发展，除浏览和回复之外，点赞和收藏逐渐成为知识交流产出的主要形式。本文在参考上述研究的基础上，将反映知识本身特征的知识密度和知识粘性纳入投入指标，点赞数和回答数纳入产出指标，结合在线健康社区的特点以及构建指标体系的全面性的原则，从知识交流广度、深度和质量三个维度建立评价指标体系，如表 2 所示：

表 2 在线健康社区知识交流效率评价指标体系

一级指标	二级指标	三级指标	指标含义
产出指标	知识交流广度	Y1（浏览数）	反映知识传播的广度
	知识交流深度	Y2（回复数）	反映知识在用户间交流的程度
		Y3（再回复数）	反映知识在用户间交流的深度
	知识交流质量	Y4（点赞数）	反映用户对知识内容的认可程度
		Y5（收藏数）	反映知识对用户的有用程度
投入指标	知识交流广度	X1（用户数）	反映知识交流中的参与主体投入
	知识交流深度	X2（发帖数）	反映知识交流中的知识源投入

知识交流质量	X3（知识粘性）	反映知识交流中知识转移的难度
	X4（知识密度）	反映知识交流中帖子必备知识量与总知识量的比例

在采用 SBM 模型进行知识交流效率计算时，各个投入和产出指标间需要满足同向性条件[3]。为检验所选指标是否合理，本文采用 SPSS 软件对收集到的投入和产出指标数据进行 Pearson 双侧相关性检验，变量检验结果如表 3 所示。由表 3 可知，本文选择的投入产出指标之间的相关系数均为正，均通过了 1%水平的双侧检验，投入和产出指标之间呈现显著的正相关关系，符合 SBM 模型的要求。

表 3 在线健康社区投入产出指标相关性检验

投入变量	产出变量				
	Y1（浏览数）	Y2（回复数）	Y3（再回复数）	Y4（点赞数）	Y5（收藏数）
X1（用户数）	.853***	.821***	.548***	.736***	.656***
X2（发帖数）	.571***	.529***	.362***	.359***	.353***
X3（知识粘性）	.682***	.715***	.573***	.490***	.455***
X4（知识密度）	.596***	.592***	.486***	.349***	.341***

注：***、**、*分别代表 1%、%5、%10 水平（双侧）下显著相关

三 实证研究

3.1 知识交流效率计算结果

表 4 在线健康社区 2010-2019 年知识交流效率

年份	综合技术效率	纯技术效率	规模效率
2010	0.704003862	0.781995716	0.882771607
2011	0.762278891	0.815111756	0.923281518
2012	0.684154826	0.771030484	0.87853878
2013	0.711464327	0.788949202	0.883334665
2014	0.787767885	0.84265199	0.904975418
2015	0.744706902	0.814713321	0.890583147
2016	0.728568608	0.802472202	0.914255097
2017	0.671556036	0.757687476	0.904761962

2018	0.675552357	0.687647189	0.965504439
2019	0.662733833	0.707549132	0.931423455
均值	0.713278753	0.776980847	0.907943009

本文使用 DEA-Solver Pro 5.0 软件对在线健康社区的知识交流效率进行测算，首先选择投入导向的 SBM 模型，其中的 CRS 模型测算在线健康社区知识交流的综合技术效率，VRS 模型测算在线健康社区的纯技术效率，二者比值为规模效率值。随后将丁香园论坛选取的 18 个版块 2010-2019 年的知识交流投入产出数据分别输入到软件中，导出其 score 表中的值，测算出的在线健康社区知识交流效率结果如表 4 所示。

3.2 在线健康社区知识交流效率分析

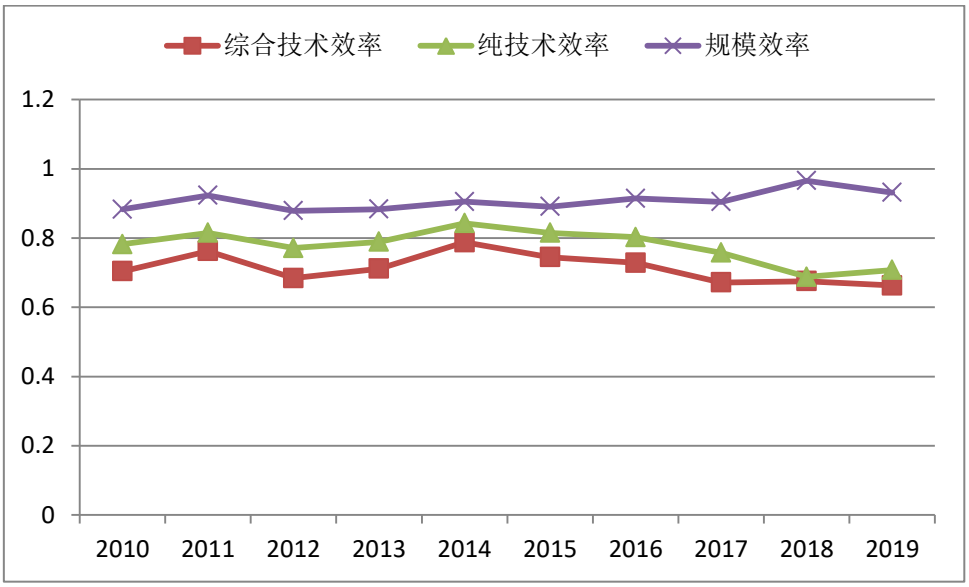


图 1 丁香园论坛 2010-2019 年知识交流效率均值

根据丁香园论坛中 18 个版块 2010-2019 年的知识交流效率计算结果，统计出 2010-2019 年的知识交流综合技术效率、纯技术效率和规模效率的均值变化如图 1 所示。

综合技术效率是对各版块知识交流效率的综合衡量，反映了丁香园论坛中用户知识交流的规模以及对投入要素的使用是否处于最佳状态。综合技术效率为 1 表示投入和产出是综合有效的，处于 DEA 有效状态，即技术和规模同时有效；如果综合技术效率小于 1，表示非 DEA 有效状态，知识交流投入产出结构未达到最佳状态，即产出水平较低，存在投入冗余和浪费，需要对制约版块知识交流效率提升的因素进行改善。由表 4 和图 1 可知，丁香园论坛各版块知识交流效率均值为 0.71，且 10 年间变化较平稳，2010-2014 年由于纯技术效率和规模效率的同时上升、下降和上升导致综合技术效率的同向变化，至 2014 年到达最高效率值 0.79 之后，规模效率仍呈现平稳上升的趋势，而受制于纯技术效率的下降，综合技术效率持续缓慢降低，2019 年达到了最低值 0.66。

纯技术效率是对各版块社区制度和管理水平的衡量，在当前投入产出规模下，反映了其对投入要素的使用是否达到最佳状态。纯技术效率为 1 时，表明在当前的制度、管理和技术水平下，投入要素得到了有效的利用；纯技术效率小于 1 时，表明对投入要素的使用存在冗余和浪费，需要通过改善制度和内部管理等方式来提升知识交流效率。由图 1 可以发现，10 年间的纯技术效率均低于规模效率，表明制约在线健康社区知识交流综合技术效率提升的因素主要是社区制度和内部管理水平、技术等；2010 到 2011 年纯技术效率小幅上升之后，2012 年又出现下降，此后两年，由于丁香园论坛在发展过程中制度和内部管理水平的提升，纯技术效率持续上升，2014 年达到 10 年间的最高值 0.84，但 2014 之后纯技术效率却持续缓慢下降，2018 年达到最低值 0.68，2019 年出现小幅度的上升，说明 2014 年之后丁香园论坛的内部管理持续出现问题，对于用户知识交流投入要素无法有效利用，社区管理者必须重视制度建设和内部管理水平等技术层面的提升，才有可能提升社区知识交流综合技术效率。

规模效率是对各版块投入产出规模的衡量，反映了在当前技术水平下，投入产出规模是否处于一个最优的状态。规模效率为 1 时，表明投入产出规模结构合理，产出效率高，投入要素得到了充分的利用；规模效率小于 1 时，表明规模结构不合理，实际规模与最优规模存在差距，需要对规模结构进行改善。由图 1 可知，丁香园论坛 18 个版块的规模效率总体水平较高，10 年间均值达到了 0.91，且每年均高于纯技术效率，整体趋势呈现平稳波动，近些年缓慢上升，并于 2018 年达到最大值 0.97，表明社区规模结构较合理，规模效益得到了充分的发挥。

PU-030

战储卫生器材信息系统的设计研究

杨举、胡红波
桂林联勤保障中心药品仪器监督检验站

目的 为提高战储卫生器材的维护保养效率和信息化程度，加强行业建设，设计研发一套战储卫生器材信息系统。

方法 研究分析当前卫生器材维护保养方式、常见故障及其维护办法、器材数据后处理、行业建设等现状，综合评估系统开发维护成本、操作难度、数据后处理分析等因素，利用成熟的通讯技术、框架和数据库设计开发战储卫生器材信息系统。

结果 通过综合研究分析，设计了一套基于浏览器为端口的战储卫生器材信息系统，该系统通过扫码枪扫描二维码并通过网页将器材信息、维护状况等录入数据库，实现器材维护业务信息化。

结论 系统建立后，在卫生器材维护信息化水平、数据分析能力、工作效率方面将有较大提高，后续不断优化后可扩展至药品制剂抽验等其他应用中。

1 引言

我军在各储供基地（综合仓库）均储备有一定种类和数量的卫生器材，也有专门力量对其维护保养，但整个维护保养过程相对落后，过多依靠人工手写记录，方法原始，大量时间花费在卫生器材登记和核验上，存在数据不完善、信息录入重复劳动、分析难度大等缺点，信息化程度较低，大多数器材维护保养后，数据仅在维护报告中体现，整体利用率极低。为提高新体制下战储物资维护保养的质量和效益，加快卫生器材监管业务规范化建设，本研究拟设计一套以战区为单位的战储卫生器材维护保养信息系统。

2 卫生器材维护保养现状

目前我军使用的是第 2 代卫生器材，从新品到待报废年限约为 10-12 年，不同设备维护保养周期存在差异，一般卫生器材维护保养次数约为 10 次，部分易耗损器材可达 20 次以上。卫生器材的日常维护保养依托于各级药品仪器监督检验站，同时该单位也承担入库检验和出库检验任务[1]，这为各项数据的标准统一、后续处理等提供较大便利。

目前，卫生器材维护保养分为两类：一是逐台检验，适用于体型大、价值高、数量少的卫生器材以及需紧急调拨的卫生器材等；二是抽样检验，适用于日常维护中数量较大的卫生器材或不易出现故障的卫生器材等。工作步骤为：①查询、调出需维护保养的卫生器材批次；②正常维护保养后将器材信息、故障状况、维护事项等誊写在表格中；③将记录后的表格转录至计算机中。整个过程中信息存在多次转化，增加了工作的繁琐程度，进一步增加了数据出错风险。

针对卫生器材维护保养现有工作模式，有不少研究提出相应的解决方案。一是联通各厂家、全军采购部门、各仓库（储供基地）等多个单位机构，从生产端到购买端到储备端实行统一标准，器材统一编号，采购系统、存储系统、维护系统等均实现统一，达到高度信息化。这种方式站位高，立足长远，是较为理想的方式，但需要较高级别单位协调，并在全军大力推广，周期较长难度较大。二是探索市场化维护保养模式，但后勤物资器材大多价值低、存储地分布广，导致规模效益低，很多厂家企业参与性不高[2]。三是放弃协调各厂家之间统一标准，转而建立自己的标准，一般是射频识别等技术，为每台件器材加装通讯芯片从而作业过程中实现信息交互[3, 4]。这种方式能实现较高的信息化程度，但存在操作使用门槛、开发维护成本较高等问题。

3 系统设计

在综合考虑工作性质、各项成本、难易程度等因素后，设计研发一套以二维码为媒介的卫生器材信息系统。

3.1 总体结构

一是现场系统，即在现场作业时使用的系统。该部分以二维码作为信息媒介，在卫生器材和数据库之间交换信息，实现卫生器材信息直观读取和在数据库中快速查找，其工作流程图如图 1 所示，箭头所指方向为数据流动方向。系统分为三个单元：

（1）卫生器材基本信息单元。其载体为二维码，粘贴于卫生器材主机或外包装表面，用于存储设备的型号、编号、历史维护记录等基本信息；

（2）数据库单元。其载体为移动平板或电脑，用于基本信息数据及维护保养详细内容的录入、存储和管理，后续数据库完善后，可以分析不同类别卫生器材的维护保养规律，给出具体维护保养建议；

（3）信息交互单元。其载体为二维码打印机和扫码枪，可实现卫生器材相关信息数字化和与主机之间的交流，用于连接基本信息单元和数据库单元。

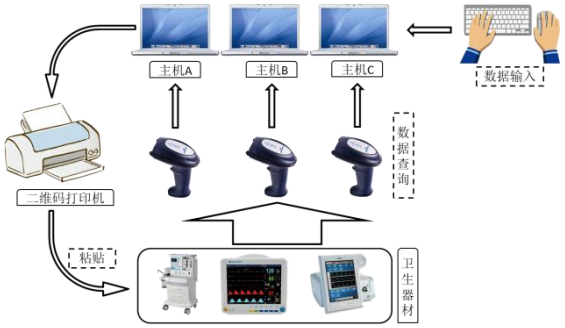


图 1 系统工作流程图

3.2 卫生器材基本信息

为方便维护人员、保管人员、运输人员、使用人员等读取有关信息和注意事项，在卫生器材表面以二维码贴纸形式（如图 2）标注有器材相关信息：二维码部分为器材唯一身份识别号，由器材名称、型号、出厂编号等组合而成，具有唯一性，技术人员通过扫码枪扫描二维码即可在数据库中调出该设备基本信息和历次维护信息；检验时间为本次维护保养日期，由系统自动生成，有效期至为检验时间加上维护保养周期自动生成，提醒保管人员及时自行维护或寻求专业人员进行维护；故障描述为此次检修查找出的问题，维修意见则为所在单位标明维修方式，维护建议为在有效期内应注意事项。

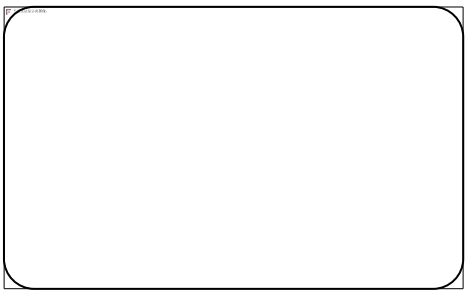


图 2 二维码贴纸包含信息

3.3 数据库

系统录入方式选择浏览器作为交互界面，有如下优势：一是有较低的维护成本，单独开发软件代码量较大，更新难度和成本较高，依托于浏览器可以大幅减少系统维护的工作量和难度；二是无需考虑用户电脑配置、操作系统等因素，兼容性极强，面对复杂的工作环境稳定性较高，系统使用成本明显降低；三是数据查阅更加方便，考虑下步可能区域、军兵种、各流程通联，数据交互更为频繁，面临更加复杂的使用环境，上级领导、基层官兵等想要正确迅速查阅数据，通过浏览器界面访问数据库更加合理高效。

我军战储卫生器材在更新换代周期内，通常由固定的厂家提供，造成特定器材较高频次出现同样故障，维护保养的办法也大致相同，因此在每一代卫生器材数据库中可以设置标准化字段以描述常见故障、维护办法等，这些字段可结合对以往数据分析和技术规范[5]得到。表 1 展示了部分卫生器材常见故障描述、维护建议和维修意见。后期针对数据库的使用状况可适当调整字段排序、字段增删等，不断优化完善。

表 1 部分战储卫生器材的常见故障

名称	故障描述	维护建议	维修意见
心电图机	电池损坏	每三个月充电维护	更换电池
	导联线损坏	日常使用不要弯折导联线	更换导联线
	屏幕损坏	不要摔打，保持干燥	紧固接触点、更换屏幕
便携式喷雾器	喷头堵塞	用后充分擦干或晾干	除锈、更换喷头
	发动机无法启动	严格按比例调配汽油，使用后油箱与发动机充分清洁	清洁油箱与发动机管路，点火器、化油器故障
野战手术灯	灯架断裂	使用时轻拿轻放	维修/更换支架
	灯泡不亮，但电池指示灯亮	轻拿轻放	灯泡损坏、线路故障
	电池指示灯不亮	对电池定时充放电	更换电池
特定电磁波治疗仪	发热管无法正常工作	轻拿轻放	更换发热管
高压蒸汽消毒锅	漏气	避免超负荷工作	更换密封圈
	锅体无法发热	无特别维护注意事项	更换发热管、更换保险、维修线路
除颤监护仪	无法开机	对电池定期充放电	更换电池
	开机后报警	\	更换电池、电源模块；打磨或更换继电器

3.3 信息交互

信息交互存在于卫生器材、维护状况、主机之间，是整个维保任务核心和花费时间最多的步骤。现有模式下，对器材的维保工作主要依赖于历次维保经验，缺乏大量数据支撑，其次信息电子化程度低、转场次数多，导致犯错几率增大。新模式下进行维护保养作业时，简单器材可以直接从二维码贴纸读取信息，复杂器材通过扫描二维码从数据库调取详细信息用于快速判断故障，随后将检验结果及维修状况等信息录入数据库中，使用二维码打印机将既定信息打印粘贴到器材表面或包装箱上，实现信息双备份和相互交流，较好弥补传统模式效率和正确率的不足。

3.4 数据分析处理

每年战区内各医疗机构、卫生队维护保养数据量巨大，这些数据经一线技术人员录入后存储到数据库中，逐步建立起卫生器材维护保养信息库，除日常研究分析外，还需考虑两方面用途：

一是代为保管处理。我军大多需维护保养卫生器材位于基层卫生队，他们对卫生器材的管理能力和精力有限，因此考虑在战区范围内由药品仪器监督检验站负责代管基层单位卫生器材数据实力，并在适当时机对接服务保障。

二是与其他数据库对接。各仓库、储供基地是卫生器材储备的重要场所，为实现物资随时可调拨随时可使用的目标，需要将维护保养系统数据与仓库、储供基地等数据库做好对接，确保卫生器材最新数据能及时有效更新到存储单位，为备战打仗做好准备。

4 系统优势

相比于现有工作模式，战储卫生器材信息系统具有如下优势：

一是直观性更强。一方面平时维护保养的卫生器材大多需拆解包装供电，每一台件设备都需要精准确定，直观读取有关数据便于维护人员和保管使用人员操作；另一方面战时条件复杂，电磁等干扰较大，直观读取抗干扰性更强，有利于加快维修速度。

二是使用简易。该系统的录入、读取等步骤快捷方便，可直接观察读取，可通过扫描二维码查询，涉及硬件设备较为简单，不需额外培训或仅简单培训就能快速上手，可弥补基层单位人员流动大、专业不对口等缺陷。

三是成本低廉。该系统硬件除服务器、客户端设备外，主要为扫码枪和二维码打印机，耗材为二维码贴纸，总体使用成本低廉；日常维护中，仅需更换二维码贴纸，系统操作简洁明了，不额外增加人工成本。

5 结论

本研究从我军卫生器材维护保养现状入手，通过分析比对现有工作模式，综合考虑工作实际、未来发展等因素，设计了一套战储卫生器材信息系统，用以提高平时战储卫生器材的维护保养信息化程度和行业建设，为以后扩展至高值卫生器材的维修记录查询、全军药品制剂检验信息化等工作打下基础。

系统建立后，可以预见战区战储卫生器材维保工作会有较大提升：一是提高信息化水平，大大提高任务效率，减轻数据采集、统计的工作量，减少工作中机械化、重复式劳动，适应我站提升应急作战能力，满足紧急、大批量器材出库检验的需要；二是增强数据分析能力，及时掌握各仓库卫生器材维护状况，自动划分卫生器材维护保养的紧急程度，紧前铺排维护保养任务，提高科学管理水平，满足工作统筹，不等不靠的需要；三是相关数据可以提供上级业务管理部门参考，在紧急情况下可以借助该数据库，调拨新近维护过的物资，提高下发卫生器材的可靠性和稳定性。

PU-031

互联网医院实现医保在线支付的设计与实践

曹凯迪、张小亮、朱佳、戴作雷、王忠民
江苏省人民医院

目的 全力保障广大患者在互联网医院门诊顺利就诊并且能够实时、准确地享受医保待遇。

方法 HIS 开放互联网门诊号源给我的南京，我的南京增加互联网问诊模块，在其中嵌入微信小程序作为患者图文咨询和视频问诊的工具，同时在江苏省人民医院企业微信中开发小程序作为医生和与患者交流的的工具，将两个小程序打通后医生可以与患者图文、视频问诊。 问诊结束后医生使用门诊电子病历写病历、开立处方，处方明细推送到患者的我的南京，市医保患者在开通我的南京中的医保在线支付的功能后只需要支付医保统筹后的自费部分，若个人医保账户内余额充足，可直接扣抵；若余额不足，可使用支付宝支付，由此实现互联网医院的医保在线支付。

结果 借助我的南京 APP 打通了互联网医院医保在线支付的通道，医保患者可在线进行医保结算，定点在我院的门统患者可以实时享受统筹支付的待遇，定点在我院的门慢患者也可以在交费时在线享受门慢待遇，结算病种信息准确无误。

结论 本项目响应国家政策，适应了当前疫情防控形势的需要，互联网医院开通医保在线支付后，解决了广大医保患者的线上就诊费用结算之忧，很大程度上减少了慢性病患者、复诊患者尤其是高血压、糖尿病、冠心病等慢性病患者来院次数，有效降低了交叉感染的风险。患者享受到了足不出户，线上问诊，在线结算，药品配送到家的服务。

PU-032

国内外医疗数据可视化研究的知识图谱比较分析

刘春鹤[1]、张晗[1]、赵玉虹[1,2]
1. 中国医科大学医学信息学院
2. 中国医科大学附属盛京医院临床研究中心

目的 通过对比国内外医疗数据可视化研究的文献分布及特点，分析国内外相关研究的共性和差异，希望二者互相借鉴和学习，促进医疗数据可视化研究的发展及医疗大数据资源的高效利用。

方法 利用科技文本挖掘工具 CiteSpace 对 CNKI、万方、维普和 Web of Science 数据库中 2010-2019 年医疗数据可视化研究的相关文献进行了文献计量、主题共现和热点趋势分析。

结果 国内外医疗数据可视化领域文献年发文量呈逐年增长趋势，国外论文主要来源于发达国家；国内外医疗数据可视化相关研究主题集中于时空数据可视化、计算机辅助诊疗、三维可视化、科学计量学等方向；国内相关研究主题突现词集中于 2010-2013 年，近期年来突现主题较少且强度较低；国外研究近期主题突现性较强，且内容更丰富。

结论 医疗数据可视化领域的国外研究论文的主题粒度更加细化且内容相对深入，国内相关研究科研创新增长程度相对疲软，可借鉴既有国际研究经验并积极丰富和拓展。

1 引言

大数据时代各行各业在运行和发展的过程中都涌现出海量的信息数据，健康医疗行业产生的数据是其中非常重要的组成部分。国务院发布的《“健康中国 2030”规划纲要》中明确提出要“推进医疗健康大数据开放共享、深度挖掘和广泛应用”[1]。当前，健康医疗相关的数据体量庞大、分布分散、复杂多变，而健康数据分析、决策支持、科普等方面对医疗信息可视化又有着巨大的需求。即使在发达国家也只有不到 15%的人能够凭借自身知识基础和健康认知力理解复杂的健康数据集所表达和传递的讯息[2]。而数据可视化可以作为一种媒介，连接在晦涩数据和读者之间，促进更加有效的信息传递交流。因此健康医疗数据可视化的研究在推进医学知识传递、成果交流、技术优化等方面就体现的尤为重要。

本研究利用科技文本挖掘可视化工具 CiteSpace 对国内外医疗数据可视化研究进行科学计量学分析，系统地对两数据集的主要内容、集中趋势予以挖掘和对比，目前国内类似研究很少，尚未有文献报道，本文意在互通有无、取长补短，希望能对我国医疗数据可视化研究的发展提供参考。

2 研究方法：

2.1 数据来源：

外文文献数据：选取 Web of Science 核心数据库。检索时间为 2020-4-2，检索策略：以“visualization”为主题查询 article 文献，从中选取了医药卫生、生物医学相关学科的文献，时间跨度为 2010-2019 年。

中文文献数据：检索 CNKI、维普网，万方数据数据库中 2010-2019 年的文献，检索策略：在 CNKI 的学术期刊库中选择医药卫生科技分类，在主题字段检索“可视化”；在万方期刊数据库中选取医药卫生分类，在主题字段检索可视化；在维普期刊服务平台中选择医药卫生分类，在题名或关键词字段检索“可视化”。最后对三个数据库来源的文献进行合并、去重、清洗。

2.2 研究工具与方法：

本研究使用了美国 Drexel University 陈超美教授开发的 Java 应用程序 CiteSpace（5.6.R3 版本）做文献分析。CiteSpace 基于 Price 的科学前沿理论，结合共引分析和引证分析,创建出由知识基础到研究前沿的理论模型[3, 4]。它能够分析科学文献中的趋势和模式[5]。

设置文献处理时间窗切片为 1 年，提取节点类型为 Keyword，筛选阈值选择了 g-index，k=25，节点连接之间运用 Cosine 算法计算网络节点的关联强度。最后对每个切片和合并后的网络使用 Pathfinder 算法进行裁剪，以简化网络并突出其重要的结构特征，使结果具备更高的可读性[6]。

3 研究结果

3.1 中外医疗数据可视化研究的分布趋势

共收集到相关外文文献 32401 篇，相关中文文献 4197 篇。图 1 展示了 2010-2019 年间国内外医疗数据可视化文献的发表情况。从累计发文数量看，国外相关文献总体上呈现逐年增加的态势，这表明科研人员对于医疗数据可视化研究的关注程度在逐渐提高，该主题一直是国际科学研究的焦点之一。2015 年之后年国内文献增长幅度有明显提高，说明了我国有越来越多的学者将目光投向医疗数据可视化领域。

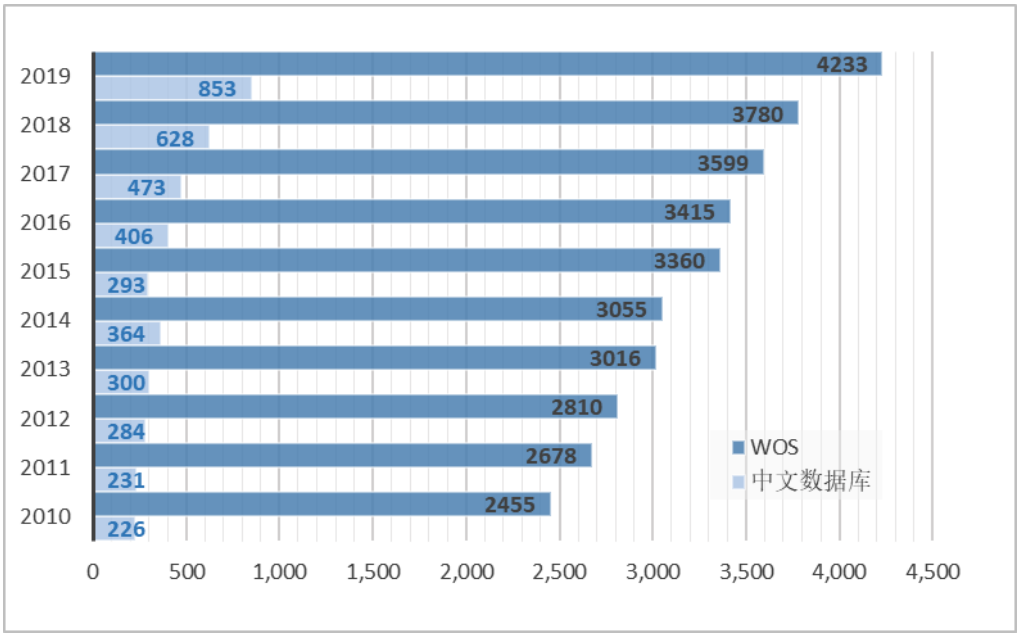


图 1 中外医疗可视化研究文献分布情况

另外，外文文献中来自美国的科研论文比例最高，约占 35%，其次是德国，约占 12%，有大约 11%的论文是来自中国（PEOPLES R CHINA ）学者。一方面可见国内医疗数据可视化的研究成果总数已经在相关国际学术领域占据了举足轻重的位置，另一方面，结合国家人口和科研人员的数量来考虑，我国在该领域的研究还有极大的发展空间。

3.2 国内外医疗数据可视化研究的主题网络的对比

研究以 keyword 作为节点，选择 log-likelihood ratio 算法对网络进行聚类，所得图 2、图 3 分别是国内外医疗数据可视化研究的主题聚类图谱。两组图谱中有相当一部分主题具有极高的相似性，如#8 spatio-temporal data fusion 和#0 时空扫描、#16 wall shear stress 和#12 壁面剪应力/#13 生物力学、#9bibliometric analysis 和#11 社会网络分析/#19 共词分析等。这些相似的聚类结果一方面证实了本研究方法的科学性，对于不同数据集的处理结果具有较好的集中趋势结果，另一方面也展示了国内外可视化领域的一些共性热点。

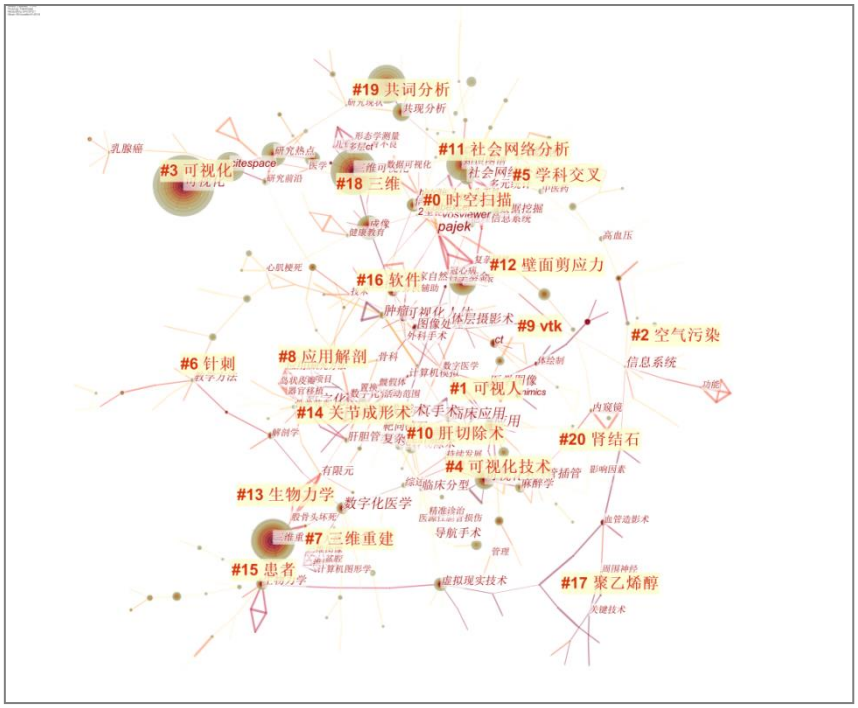


图 2 国内医疗数据可视化文献研究主题聚类图

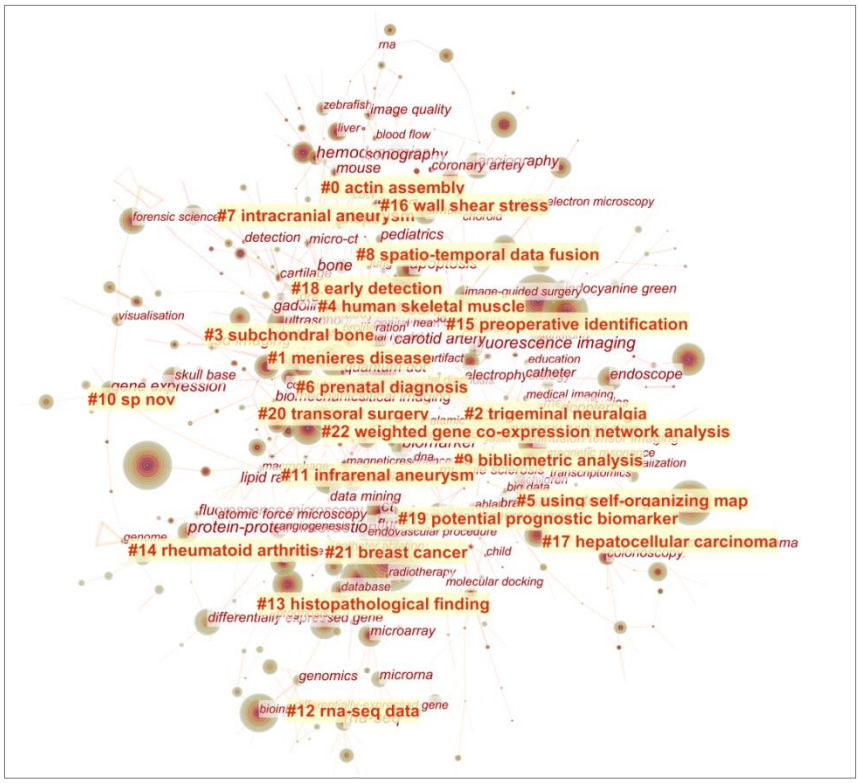


图 3 国外医疗数据可视化文献研究主题聚类图

近十年国内的医疗数据可视化研究的主题主要集中于如下几个方面：时空数据分析（#0）、人体生理病理组织结构的成像（#1、#7、#8、#18）、可视化软件和技术开发（#4、#9、#16）、计算机可视化辅助治疗（#10、#14、#20 等）生物力学（#12、#13）、文献计量学研究（#11、#19）、交叉学科应用（#5、#2、#17）、中医学相关的可视化（#6）。国外的医疗数据可视化研

究的主题比较具体，主要集中于：人体生理组织的可视化（#3、#4、#5）、时空数据可视化（#8）、生物信息学（#0、#12、#22）、计算机辅助诊断（#6、15、#18）、计算机辅助治疗（#7、#11、#17、#21）、生物力学（#16）、文献计量学（#9）。相对于国内的可视化相关研究而言，国外研究主题的粒度更加细化，专业也更加深入。除了上述二者相似的研究主题之外，国内研究也有其独有的关注焦点，比如中医学治疗方法的可视化研究（#6），可视化工具软件的开发和探索（#4、#9、#16）。

3.3 国内外文献主题突现情况

突现词是在某段时间内相对增长率(词在某时刻的发展速度相对它自身平均的发展速度)突然增长的词[7]。如果突现词为关键词，则它反映文献集中某一时段关注度突然增长的研究主题和内容。突现强度越高，说明突现年份中该关键词的共现频次骤增的程度越强。经过 Burstness 计算，国内相关文献中共测出 64 个突现词，国外文献中共测出 181 个突现词。

表 1~3 为国内外两数据集得出的突现词分布情况。突现年份每行的 10 个点，分别对应 2010-2019 的 10 个自然年，实心点代表该突现词在对应年份出现“突发”现象，●表示该突现词虽然在 2010-2019 年之间出现过“突发”现象，但当前已过了突发阶段，■表示直到 2019 年该词仍处于突现“热期”，虽然可能突现期截止于 2019 年，但很可能“热期”继续向后期延续。表 1 为突现强度降序排列的前 20 个突现词，由图可见国内相关文献中突现强度最高的几个的关键词“三维重建”“可视化人体”等均分布 2013 年之前，只有“3D 打印”和“生物信息学”在当前有爆发增长趋势。外文文献中“microRNA”和“3D printing”既是当前的研究热门更具有很高的突现强度，可以说是当前在医疗数据可视化领域关注度最强的主题概念之一。

表 1　国内外医疗数据可视化研究关键词突现词表（Strength TOP20）

外文突现词	突现强度	突现年份	中文突现词	突现强度	突现年份
microRNA	13.334	○○○○○○○○■■■	三维重建	38.7439	●●●●○○○○○○
3D printing	10.095	○○○○○○■■■■	可视化人体	10.2774	●●●●○○○○○○
prenatal diagnosis	9.9587	●●●○○○○○○○	血管造影术	8.1985	●●●●○○○○○○
coronary artery	9.625	●●●●○○○○○○	vtk	8.11	●●●○○○○○○○
robotics	9.2541	●●●●○○○○○○	计算机辅助	7.2226	●●●●○○○○○○
oncology	8.9216	○○○○○○●●○○	图像处理	6.6186	●●●●○○○○○○

appendiciti	7.7696	○○○○●●●●○○	解剖学	6.2687	○●●●○○○○○○
SNP	7.5767	○○○○●●○○○○	体绘制	6.0338	●●●●○○○○○○
quantification	7.3497	○○○○○○■■■■	3D 打印	5.8513	○○○○○○■■■■
lymph node	7.3366	●●●○○○○○○○	数字化	5.6797	○●●●○○○○○○
glioblastoma	7.2483	○○○○○○○■■■	医学图像	5.6596	●●●○○○○○○
brachytherapy	7.1447	○○○○○○●●○○	穿支皮瓣	5.6109	○●●○○○○○○
metagenomics	6.9975	○○○○○○■■■■	模型	5.5073	●●●●○○○○○○
quantum dot	6.894	●●●●○○○○○○	CT	5.0338	●●●●○○○○○○
prostate	6.845	●●●○○○○○○○	histcite	4.9003	○○○●●○○○○○
radiotherapy	6.772	○○○●●●●●○○	应用	4.8095	○○●●●○○○○○
genetics	6.7422	○○○●●○○○○○	体层摄影术	4.0806	○●●●●●○○○○
molecular visualization	6.6679	○○○○○○●●○○	生物信息学	4.0087	○○○○○○○■■■
radiography	6.6395	○●●○○○○○○○	计算机模拟	3.9906	●●●○○○○○○○
protein-protein interaction network	6.5974	○○○○○■■■■■	虚拟手术	3.9906	●●●○○○○○○○

表 2、表 3 显示的是国外、国内突现词分布年持续至 2019 年的词表，即当前仍具备突发性的关键词。经过对比发现，表 2 中突现词的突现强度分布比较均匀，在 3.9906-13.334 之间，而全部外文突现词的突现强度跨度在 0.207-13.334 之间，因此表 2 中突现词强度总体偏向高值。

表 2　国外医疗数据可视化研究关键词突现词表（~2019）

外文关键词	突发强度	突发年份
microRNA	13.3337	2015-2020
3D printing	10.0945	2016-2020
quantification	7.3497	2016-2020
glioblastoma	7.2483	2015-2020
metagenomics	6.9975	2016-2020
protein-protein interaction network	6.5974	2015-2020
differentially expressed gene	6.5519	2015-2020
compressed sensing	6.4118	2015-2020
pathway	6.3014	2015-2020
workflow	6.2984	2016-2020
gastric cancer	5.7604	2016-2020
vestibular schwannoma	5.5099	2015-2020
oct	5.3552	2015-2020
network	5.204	2015-2020
mRNA	5.0504	2015-2020
cryo-electron	5.0504	2015-2020
glaucoma	4.785	2015-2020
genome	4.7619	2015-2020
menieres disease	4.5181	2015-2020
spinal cord injury	4.3385	2015-2020
endolymphatic hydrop	4.0208	2015-2020
cornea	3.6497	2015-2020
spine	3.5658	2015-2020
neuroimaging	3.3159	2015-2020
data mining	3.1568	2015-2020
4D flow MRI	2.8466	2016-2020
registration	2.7065	2015-2020
autofluorescence	2.7035	2015-2020
image registration	2.6875	2015-2020
network analysis	2.0817	2016-2020
3D visualization	1.8756	2015-2020
pregnancy	1.2788	2015-2020

tomography		○○■■■			○○■■■
trauma	5.0145	○○○○○ ○○■■■	laparoscopic surgery	0.8189	○○○○○ ○○■■■

表 3　国内医疗数据可视化研究关键词突现词表（~2019）

外文关键词	突发强度	突发年份
3D 打印	5.8513	○○○○○○ ■■■■
生物信息学	4.0087	○○○○○○ ○■■■
可视化角膜生物力学分析仪	3.4423	○○○○○○ ■■■■
术前评估	3.0613	○○○○○○ ○■■■
脑卒中	3.0432	○○○○○■ ■■■■
高血压	2.743	○○○○○○ ■■■■
数据可视化	2.7352	○○○○○○ ■■■■
麻醉学	2.7157	○○○○○○ ○■■■
计量分析	2.6683	○○○○○○ ○■■■
时空聚集	2.6683	○○○○○○ ○■■■
流行特征	2.6318	○○○○○○ ○■■■
术前规划	2.6318	○○○○○○ ○■■■
专利分析	2.4127	○○○○○○ ■■■■
数据分析	2.4127	○○○○○○

■■■■

表 3 中词的突现强度分布在 2.4127-5.8513 之间，全部中文突现词的突现强度在 2.4127-38.7439 之间，可见表 3 中突现词强度总体严重偏低。这一现象说明国内相关研究立题角度新增的爆发性主题与国外相比较少，且爆发的强度较小，并没有出现像本时段初期类似“三维重建”或“人体可视化”这样集中趋势明显且增幅数量极大的热门研究主题。另外，通过观察具体突现词内容发现近年来国外的研究趋势倾向于更细粒度的内容如“protein-protein interaction network”、“genome”、“pathway”“metagenomics”，对应国内的研究则表达为上述主题的上一层次概念“生物信息学”。国外研究的集中性增长已经明显体现在具体病症的可视化分析上，如“glioblastoma”、“gastric cancer”、“glaucoma”等，而这种特点在近期的国内热门主题上并没有显现。

4 讨论

无论是国内还是国外，医疗数据可视化研究的年文献增长幅度都在持续走高。随着全球医疗数据数量的迅猛增长，如何有效高效利用这些数据迫在眉睫，可视化的方法能够集中表达批量的数据信息，也可以提高用户对医疗数据统计分析的效率和协助信息理解，这也是有越来越多的相关研究的原因。

纵观近十年国内外医疗数据可视化研究的主题分布，共同点主要集中于以下几个突出的方面：医疗数据时间空间维度的可视化研究、计算机辅助临床诊断治疗、生物力学应用的可视化和科学计量学。时间空间维度的可视化多应用于疾病流行病学分布，疾病病程发展等；科学计量学则包括应用较多的社会网络分析、文献计量学分析等。国内的研究在人体组织结构和生理病理成像的研究上有很大侧重，另外对中医药医疗数据的可视化的研究也是国内研究的特色之一。国外研究中比较突出的重点主题就是“生物信息学”，这一主题不仅占据很大比例而且在研究粒度上也更加精细，基因和蛋白生物特性、它们对生理病理过程的作用机制、以及它们之间的互作关系都是当前国际医疗数据可视化研究中的热点。虽然国内近年来此类研究也有所起步，主要集中在 2015 年之后，但在研究深度和粒度上还需要进一步扩展。

国内外相关研究热点趋势的分析结果也有明显差异，国外研究中最近 3-4 年新兴主题的突现强度分布比较均匀，且值偏高，充分体现了其“百家争鸣”“活力充沛”的良好研究发展趋势。其集中在生物信息学、具体部位疾病的诊断治疗、三维可视化和数据挖掘这几个方面，其特点是对象明确，部位准确。国内研究的高值突现词主题则分布在 2010-2019 年区间的前半部，近 3 年的新研究点突现值均处于较低值部分，总的来讲有些“活力欠缺”。由此可以看出虽然我国在医疗数据可视化领域的总文献数量逐年递增，但在开发科研创新点这个层面上还需要更加深入的思考和探索。在突现主题内容上可以借鉴国际研究的丰富经验将主题细化，如将大主题“生物信息学”更进一步，也继续向 protein-protein interaction network、mRNA、pathway、microRNA 等子主题深入发展；将广义的“数据分析”细化到具体某个层面的、某种性质或某个维度的数据来分析，

以此来丰富国内相关研究的深度和广度。希望上述的分析结果能够对国内外医疗数据可视化研究领域的学者提供帮助和借鉴。

PU-033

智慧教室在高等学校教学中的应用研究

杨苏彬、朴杰
哈尔滨医科大学

目的 随着信息技术的不断发展，新媒体，人工智能得到了广泛应用，智慧环境成为未来教育教学活动的新趋势,智慧教室的建设也成为高校教育教学发展的重要目标。目前许多高校的教室设备及教学模式依旧偏向传统,传统的教学方式仍旧占据主要地位，智慧教学应用发展滞缓其原因可能与智慧教室环境建设不无关系。文章从教育信息化及新媒体教室模式的现状及展望出发,对智慧教室的建设方式进行探讨,以期为寻求高校课堂模式的改革和发展路径提供参考。

“十二五”以来，教育信息化已经上升为国家战略。《国家中长期教育改革和发展规划纲要（2010-2020）》明确指出“信息技术对教育发展具有革命性的影响，必须予以高度重视”，要“加快教育信息化进程”。教育部在《教育信息化十年发展规划（2011-2020 年）》中指出“以教育信息化带动教育现代化”,并明确了现阶段基础教育的几个代表性的阶段性工作任务之一即为建设智能化教学环境，提高信息化教学水平[1]。高校作为实现教育信息化的主体，教学场所的信息化水平特别是智慧教室的研究、建设与运行关系到人才培养的质量，是衡量高等院校教育教学水平以及信息素养现代化水平的重要指标。

1 智慧教室的特征及分类

1.1 智慧教室的定义：智慧教室既是物联网、大数据、云计算等新一代信息技术与现代教室相结合的代名词，也是现代教室设计运行的高级阶段[2]。智慧教室的设计功能不但融合教学互动、教学管理为一体的高度集成化学习环境，还包含了教学空间的设计，各种媒体软件和硬件的集成，无感知的大数据采集系统，教室环境管控等多种为教学活动提供服务的支撑设备和系统。可为教学任务调度变更、教务管理、考试等教学活动提供一个更为高效的网络化平台，对教育教学、教务管理、考务管理及学生管理提供更为智能、高效和有力的物理环境支撑和信息传递交互环境支撑，更可基于动态学习数据分析和“云+端”的运用，实现教学环境自动化、评价反馈即时化、交流互动立体化、资源推送智能化，构建大数据时代的信息化课堂教学模式[3]。

1.2 智慧教室的分类：随着“探究性学习”、“小组合作学习”和“项目式学习”等多种教学模式的不断提出，智慧教室从“内容呈现”、“资源获取”和“及时交互”多个维度来增强教室的设计，智慧教室主要有“灵活互动讨论型”、“讨论型”和“智能物联型”三大类型[4]。①灵活互动讨论型指教室桌椅均为可移动，可以根据教学需要灵活转换为大组学习模式、小组研讨模式或不

同人数成组，小型或标准教室可适配 12-36 人，大班可适配 96 人左右班容量。该教室应提供全场景高清录播，多小组多屏幕互动，多设备接入（手机、平板、笔记本等），分小组记录等多种支撑系统，为就一个或者多个课题进行分组研讨和分享提供支持。②讨论型智慧教室：传统的黑板变成了纳米智慧黑板，教师在讲课时可直接显示课件，也可使用普通粉笔进行板书书写。教室是自由讨论式课桌，每六到八张课桌可环绕形成小组，每小组之间可由独立的交互智能教学一体机组成讨论书写板，学生即可实时观看板面内容，又可在课堂讨论式将本小组的内容书投屏出来。③智能物联型智慧教室：指教室内不设置课桌，只配备可移动的智能旋转桌椅。教师可通过中央控制系统来调节教室内的多种电子设备，在交互式纳米智能黑板上演示课件，使课堂氛围更加轻松活泼[5]。

2 高校课堂模式的现状分析

教室作为高校师生开展教学活动的核心场所，传统教室和多媒体教室中，教学设施以相对陈旧，对教学活动的支撑也比较基础。常见的方式还是老师在台上讲、学生在台下听，管理者对教学过程的监控和管理也缺乏有效手段。主要存在以下问题①集成度低,稳定性差：多媒体教室的多种设备和系统归属于不同的厂家，之间无法互联互动，老师在使用教室过程中需要学习多种设备。多种设备又导致教室环境非常复杂，经常由于某一个小部件的问题导致整个教学过程受阻。②终端兼容性差：很多的多媒体室或者多功能室只能采用教室内公用 PC 授课，对于自行携带电脑的支持往往存在接口和兼容性问题，学生的手机、Pad、笔记本等设备难以用于课堂互动。③分组研讨支持度低：由于多屏幕互动技术的约束，更多的教室只能由教师大屏幕和部分学生终端组成，无法进行学生分组讨论，也不提供小组讨论的屏幕，老师无法组织分主题探究教学活动。④跨校区授课支持度差：受限于高清音视频的传输和压缩技术，跨校区、甚至跨区域内外的互动无法完成，更多的录播教室只能提供单向的视频流，更谈不上将远端的学习情况即时同步到主讲教室[6]。

3 高校智慧教室建设模式的发展探究

高校课堂的交互参与主体、教学媒介平台与学科教学差别，可以搭建促进互动交流的教学环境，从而实现一个从设备到数据全面打通，从教学活动到数据评价全面融合的整体框架，提供便捷的开展课程教学的环境，激发学生的学习能动性，通过融合网络教学知识资源，丰富教学资源。3.1 灵活的教学课程与班级设计：传统的教学知识传授基本都是用语言描述的，为了满足不同学科、不同教师个性化教学需求，智慧教室可是实现教师可根据自身需要进行课程与班级设计，为广大老师探索新型教学模式，开展新型教学活动提供技术手段，为老师细致掌握学生的学习情况和学习效果提供技术手段；为老师逐步实现“因材施教”提供技术手段[7]。随着新信息技术的快速发展，教师在知识教授过程中可以利用多种方式进行演示，包括图像、动画、声音、视频等方式来制作课堂教学课件，以多种新兴技术实现实验模拟仿真，更加直观的展示实验成果和教学内容。同时，全体学生可以共同观察一个参考对象，师生之间可以互相对话，能够有效促进师生之间的双向交流互动，架起学科服务和教学支持的桥梁引导学生通过互联网络对丰富的知识资源进行分析，利用便捷的网络进行交流，培养学生信息利用能力[8]。

3.2 精准的多形式测评和学情分析：在对所学课程进行测评时，教师可以设置学生可通过移动终端进行测验答题，支持单选、多选、判断、解答等多种题型；在学生答题过程中，教师可实时查看学生的答题进度，还可根据需要设置答题截止时间或手动结束答题。所有测试结果和数据分析均在第一时间反馈给教师，帮助教师收集课堂数据进行针对性讲评[9]。

3.3 海量课堂大数据及优质的教学视频资源保存与共享：教师与学生在教师内的教学活动，包括课堂互动数据、视音频等，系统能自动记录并形成课堂大数据及优质的教学视频资源，并上传至课堂管理平台，管理人员可以随时进行课堂数据分析、视频点播和直播等功能，将教学资源充分利用。在区域内，通过管理平台，将教学资源整合应用，达到资源最优化处理[10]。基于此，智慧教室还可将提供教育云架构的资源服务、教研应用等内容，并能满足不同客户端的体验，包括 PC、笔记本、手机移动端等，能够帮助学生有效进行自主学习，弥补学科体系庞大、知识点丰富等因素造成的课堂授课和教辅材料不能有效衔接与融会贯通的弊端。

4 结语

教育是关系国计民生的重要领域，教育现代化是关系到国民素养和创新源动力的内在要求，而教育信息化是实现教育现代化的必备手段，智慧教育作为教育信息化的高端形态,目前在全球范围内已受到越来越多的关注[11]。推动信息化与教育教学的深度融合，在提高教学效果的基础上，实现数据化诊断、精细化管理、科学化决策，为教师、学生、管理者提供切实可靠的信息化教学环境是当前高校教育信息化的重要目标。

作者简介：杨苏彬，男，副教授，硕士研究生，主要研究方向：计算机网络系统集成、多媒体教室管理

朴杰，男，研究员，博士研究生，主要研究方向：卫生信息管理、教务系统管理

PU-034

代写代发论文的识别、防范及学术失范的治理

杨春华、姚敏、刘娜、黄思敏、吴芳怡
军事科学信息研究中心

目的 近年频发的撤稿、抄袭、造假等学术不端事件被社会广泛关注和质疑。学术期刊是论文发表环节维护学术诚信、抵制学术不端的一道重要防线。本文结合编辑工作中发现的代写代发论文为例，总结了这类论文的共同特征，从严格初审、规范流程、完善平台、提升能力等方面提出该类论文的具体防范措施，并进一步探讨学术不端论文的成因，建议完善科研评价机制，降低功利性论文发表需求，削减低水平期刊，从政策层面优化学术交流环境，遏制学术不端行为，通过“中国科技期刊卓越行动计划”等，培育适合优秀科研成果发表的具有高公信力的学术交流沃土，提升我国的科研竞争力。

PU-035

基于 5G 和人工智能技术的院前院内急救管理平台建设

张小亮
江苏省人民医院（南京医科大学第一附属医院）

目的 随着 5G、物联网、人工智能、大数据等前沿技术的成熟和发展，为解决院前与院内急救数据传输稳定性，提高急救患者的救治效率

方法 本文以江苏省人民医院在院前急救中的应用为例，基于 5G 和人工智能技术，在院内集成平台的基础上，对院内急诊系统进行改造，打通院前急救和院内系统。

结果 院内医生可提前查看院前急救患者的生命体征、心电等诊疗信息，并可通过 5G 网络与救护车患者或医师实时视频沟通或指导。

结论 借助于 5G 高速率、低时延、大连接的特性，大大缩短抢救响应时间，提高患者救治效率，实现院前院内息资源共享、系统互联互通，为院前急救、智慧医疗提供强大技术支撑。

院前急救需要在短时间内对威胁人类生命安全的意外灾伤和疾病采取紧急救护措施，在现场没有专科医生或全科医生的情况下，基于 5G 高速率、低时延、大连接的特性，能够在高速移动环境中依然提供稳定的通信保障，急救医生将通过基于 5G 的院前急救系统将患者生命体征和危急报警信息传输至远端专家侧，并获得专家远程指导，对挽救患者生命至关重要。通过 5G 网络实时传输医疗设备监测信息、车辆实时定位信息、车内外视频画面，便于实施远程会诊和远程指导，对院前急救信息进行采集、处理、存储、传输、共享可充分提升管理救治效率，基于人工智能技术对急救病人进行智能诊断和辅助决策。实现上“上车即入院”，抢占黄金时间，大大缩短抢救响应时间，为病人争取更大生机，最终实现院前院内息资源共享、系统互联互通，为院前急救、智慧医疗提供强大技术支撑[1-3]。

1 院前院内急救现状分析

1.1 5G 技术医疗现状

第五代移动通信技术（Fifth Generation of Mobile Phone Mobile Communications Standards，5G），也称为第五代移动电话行动通信标准，它是 4G 之后的延伸，理论下行速度为 10Gitb/s。5G 不仅能实现三维图像的高质量传输，为高速移动的用户提供高质量视频的服务，还能提供通信信息之外的数据采集、实时定位、远程诊疗等融合功能。2018 年国家卫健委出台的《关于促进“互联网+医疗健康”发展的意见》，鼓励医疗机构积极应用智能辅助诊断系统，提升基层医疗服务能力和效率，截至目前我国已有近百家医院开展了 5G+医疗场景的试点，从院内移动医疗到院间远程医疗，从移动信息化到远程手术协同，为 5G 临床应用提供了实践基础。2019 年国家卫生健康委发布《基于 5G 技术的医院网络建设标准》，为中国急救医疗体系的新一代网络基础设施建设提供了指导原则[4]。

1.2 人工智能技术医疗现状

人工智能（Artificial Intelligence，AI）是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。近年来，人工智能技术与医疗健康领域的融合不断加深，随着人工智能领域，语音交互、计算机视觉和认知计算等技术的逐渐成熟，人工智能的应用场景越发丰富，人工智能技术也逐渐成为影响医疗行业发展，提升医疗服务水平的重要因素。其应用技术主要包括：语音录入病历、医疗影像辅助诊断、药物研发、医疗机器人、个人健康大数据的智能分析等。

1.3 5G 人工智能院前急救现状

急救医学是一门处理和研究各种急性病变和急性创伤的一门多专业的综合科学，需要在短时间内对威胁人类生命安全的意外灾伤和疾病采取紧急救护措施，包括循环功能引起的体克，急性创伤，急性心脑血管疾病，多器官功能的衰竭，急性中毒等。对于危重患者实时数据传输非常重要，部分检查产生数据量极大，如超声基于时间序列的动态影像，单一患者将会产生高达 2GB 的超声影像数据，而且这些动态影响对远距离传输的图像的连贯性和延时控制有着极高要求，5G 技术出现解决了在远程超声检测过程由于网络技术局限性而对影像数据传输的稳定性和实时性无法满足的问题，5G 的高带宽、低时延和大连接的技术能力促进人工智能发展，5G 网络中的人工智能实现更智能的广泛应用模式、更快的响应速度以及更直观的用户体验。基于 5G 和人工智能的边缘云可以提升计算效率，以去中心化的工作方式极大提升了人工智能运算效率和可靠性，国内以浙江大学附属第二医院搭建的基于 5G 的智慧急救指挥中心平台为典型案例[5,6]。

2 平台设计

2.1 平台网络架构

基于 5G 和人工智能的院前院内急救管理平台架构分为院内业务和院外业务，见图 1，院内和院外基于 5G 基站或室分接入 5G 网络，依托运营商现有的 5G 承载网及核心网实现互联互通。为保障医疗通信安全及服务质量保障，通过端到端切片技术，从无线网到承载网到核心网，为医疗业务预留专用物理/逻辑通道，构建 5G 卫生医疗专网。

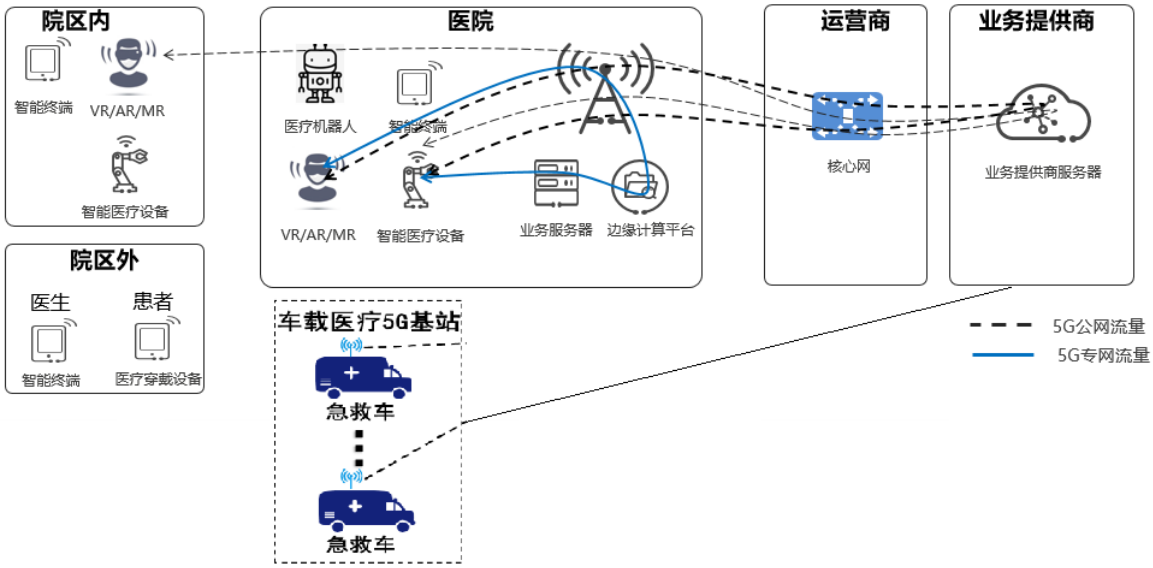


图 1 基于 5G 和人工智能的院前院内急救管理平台网络架构图

2.2 平台功能架构

基于 5G 和人工智能的院前院内急救管理平台具备院前和院内的信息实时共享功能，院内远程急救，智能诊断等功能，见图 2。通过 5G 将院前患者的生命体征、病程记录、院前心电监测、超声等数据实时传送到院内分诊和电子病历系统，供院内医生快速了解院前患者的状况。院内基于人工智能和大数据技术，实时将院内急诊的床位状况、患者就诊等待状况实时反馈到院前系统中，智能辅助院前准确决策。整个数据传输过程基于院内集成平台和标准的数据传输格式，以保证院内系统间的互联互通，详细功能如下。

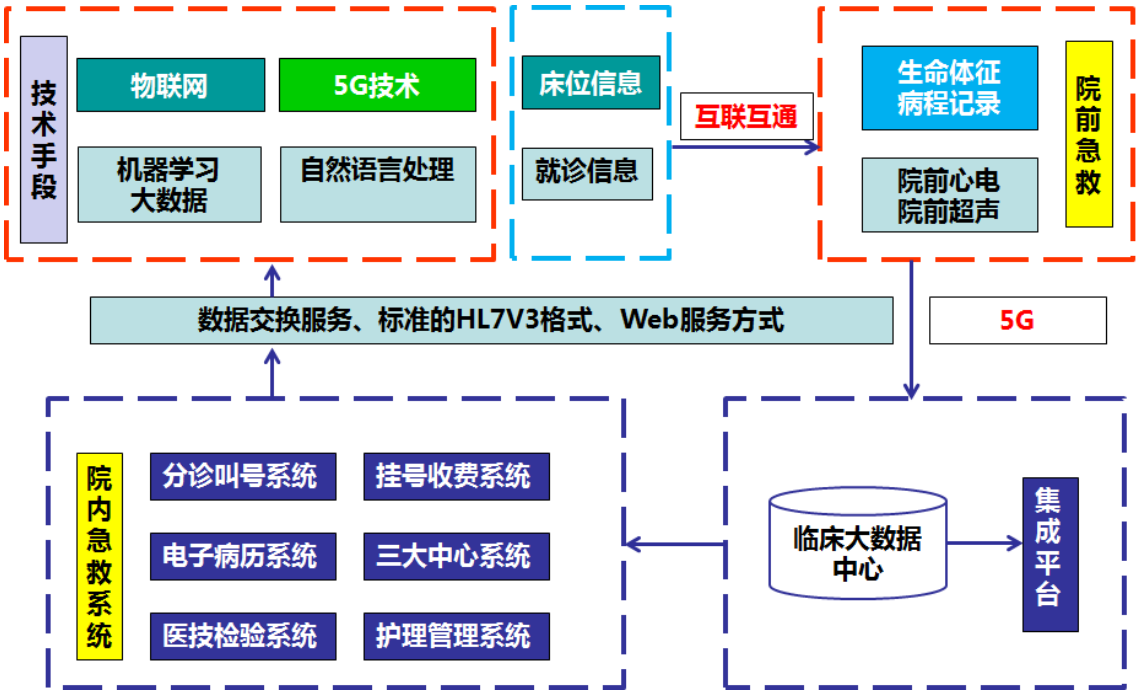


图 2 基于 5G 和人工智能的院前院内急救管理平台功能架构图

2.2.1 院前 120 和院内系统的互联互通

正式对接院前急救系统之前，患者到达医院后，分诊台的护士需要手工录入患者的姓名、年龄、性别、症状等信息，耗费患者的宝贵时间。在正式对接之后，分诊护士、急诊医生在院内急诊信息系统中可查看院前120患者在到达急诊之前的患者主诉、生命体征、心电图等数据，并可与120随车医生进行音视频交流，120患者到达医院后，救护车随车人员通过120提供的pad上的患者二维码进行交接，预诊台护士只需扫码即可快速为患者打印分诊条。

医院作为南京市远程心电中心，具有强大的心电诊断能力，随着人工智能心电诊断服务的成熟，基于5G传输为南京市120急救车提供了人工智能心电诊断服务，目前正常/房颤模型整体准确率99.15%，正常/节律/波形/节律+波形4分类模型整体准确率接近90%，大大提高了院前120患者的心电诊断能力。

2.2.2 三大中心先诊疗后付费流程

三大中心脑卒中患者到达急诊后，分诊护士在系统中进行分级并分诊，打印分诊条，发放RFID手环给患者，并发送脑卒中短信至卒中小组医师。卒中小组医师打开随身PDA上的脑卒中单病种管理系统APP，根据患者病情选择黄区或红区接诊，并点击“抽血/静脉留置针”、“呼叫CT室”，后台一键开立检验检查医嘱，实现急诊三大中心患者“先诊疗后付费”的绿色通道模式。

2.2.3 基于物联网技术的患者轨迹追踪

对于三大中心患者，通过RFID技术，对患者进出各个诊疗区的时间和操作人员进行记录，患者溶栓结束后，后台自动记录各个节点的时间节点和质控指标。

2.2.4 基于人工智能的语音呼叫技术

对于多中心的创伤中心由于涉及到多学科联合会诊，传统的人工电话通知耗时耗力，创新性的采用 “人工智能语音”应用于急会诊，通过自定义问题表单，实现APP一键AI智能语音电话通知相关人员，提高医生的会诊效率。

2.2.5 基于5G技术的医联体院前急救模式

以5G救护车为载体，构建省人民医院5G异地急救信息服务平台，配合人工智能（AI）、无人机、4K超高清和虚拟现实（VR）视频等技术应用，增强急救及应急救援能力，通过与浦口人民医院、溧阳人民医院、常熟人民医院的深度合作，实现优质的急救医疗资源下沉，构建全方位的异地急救体系。

3 总结

基于 5G 通讯技术和人工智能对院前院内急救系统进行一体化构建，进一步提升医院信息化诊疗、信息化服务、信息化管理水平，未来5G应在 5G 远程会诊、5G 远程超声、5G 远程手术、5G 远程示教、5G 应急救援、5G 远程监护、5G 移动医护等领域进行延伸应用。

通过大数据+人工智能为临床、下级医疗机构提供辅助诊疗支持、为运营管理提供分析预测、为科研提供数据分析，实现智能临床、科研、管理一体化平台。目前江苏省信息化建设整体架构已达到国内较先进水平，各项医疗服务实现闭环管理，并应用人工智能、物联网技术、大数据分析提

升患者就诊体验，在此基础上利用 5G 等新技术实现区域化医疗、远程医疗协作，远程手术、高水平专家远程平台，将成为信息化建设的重点工作。

PU-036

基于微信小程序的移动医疗数字化多媒体/短视频医患沟通系统 “摇摇灵”的开发与实现

郭秋月[1]、李俊[1]、贺庆[2]、杨小霞[2]、陈鑫凯[2]、崔娟莲[2]、张江杰[3]、熊力[1,2]
1. 中南大学湘雅护理学院
2. 中南大学湘雅二医院
3. 浏阳市仁杰电子科技有限公司

目的 在移动医疗的快速发展背景下，为实现医务人员与患者的高效沟通，我们开发了基于微信小程序的数字化多媒体/短视频医患沟通系统。旨在促进医患实时高效沟通，实现医务人员对患者需求的快速响应，减少由于非医疗过失导致的医患纠纷，提高患者就医体验，提升医务人员与患者满意度，同时可为今后的智能医疗提供可靠的原始大数据。

方法 我们设计了一个集医护呼叫系统、智能手机、信息技术和云系统于一体的基于微信小程序的移动医疗数字化多媒体/短视频医患沟通系统“摇摇灵”。该医患沟通系统创建了包含医生、护士、患者在内的群组，患者可随时在聊天框发送文字、图片、音频和短视频信息表达自己的需求，所有在岗医务人员都会即时收到信息提醒，可实现患者与医务人员之间一对一向一对多沟通方式的转变。

结果 该系统可实现医务人员对患者需求的快速响应，保证患者的得到及时的个性化医疗和护理服务，提高患者满意度。同时，可跨越空间障碍，促进医务人员之间的实时交流，增强跨学科团队之间的沟通与合作，为病人提供优质服务。多媒体信息的记录是不存在失真的原始资料，可为今后智能医疗的发展提供原始大数据。

结论 基于微信小程序的“摇摇灵”实时交互系统以互联网为载体，利用信息通信技术与传统病房呼叫系统进行深度融合与创新。该系统的发展可提升服务质量、提高管理水平、简化作业流程、树立行业标杆，有望推动传统医疗模式的实质性变革。

PU-037

面向电子病历的胃癌术后复发风险预测研究

李郁鸿、程铭
郑州大学第一附属医院

目的 在我国胃癌的发病率和死亡率较高，是影响居民健康的主要疾病。胃癌患者治疗以手术为主，由于早期诊断和癌症治疗手段的提高，患者生存期有明显的改善，但术后复发率仍较高，严重威胁患者生存率。以往胃癌术后复发的研究主要集中于实验室指标的统计模型，未能利用病历文本中的大量临床描述信息。本文拟同时抽取病历文本中的临床描述信息和实验室指标信息，构建胃癌术后复发预测模型。

方法 本文提出了一种基于注意力机制的混合神经网络模型，采用 BiLSTM 和自动编码网络分别捕获病历文本中的临床描述特征和实验室特征，构建模型预测胃癌术后患者的复发风险。

结果 实验结果表明了我们方法的有效性，优于传统统计模型。

结论 同时抽取病历文本中的临床描述信息和实验室指标信息，有利于提高预测模型的性能

PU-038

中文在线健康网站质量评估

孙伟伟、赵斌
蚌埠医学院

随着互联网的不断发展，全球信息化的共享，公众对在线健康信息有着越来越强烈的需求。通过 Alexa 指标和网页级别（PageRank, PR），筛选出 30 个中文在线健康网站，利用国际上常用的网站信息质量评价量表，对中文在线健康网站的相关质量进行评价，发现： 30 个中文在线健康网站得分的平均分是 58.30 ± 6.41（总分 81 分），总体质量水平较为适中，无严重不足；网站的可用性得分均值为 40.03 ±15.28（54 分），功能性得分普遍较低；网站可靠性得分均值为 14.27±4.43（27 分）；网站内容得分均值为 17.53±8.60（53 分）；网站的易用性方面，页面图片加载慢及访问服务器速度较慢仍是大多数网站面临的问题。通过对中文在线健康网站的可用性、可靠性、内容和易用性进行评价，找出中文在线健康网站做的不足之处，为今后提升网络健康信息质量提供相应的参考。

1 引言

近十年来我国卫生事业蓬勃发展，信息化建设也在飞速发展，卫生事业与信息化建设的结合发展使得卫生信息的传递也越发迅速，越来越多的群体选择从网络上获取健康信息[1]，以此来满足自己对健康信息的需求。美国皮尤研究中心 (Pew Research Center)调查显示，美国成年用户中

有 72%在网上搜索健康信息[2]。在我国，截至 2015 年 12 月，中国现有互联网医疗用户规模达 1.52 亿，占网络用户总数的 22.1%，其中“在线医疗保健信息查询”占据最高使用率（10.4%）[3]。而第 43 次《中国互联网网络发展状况统计报告》显示，20%非网民为方便获取医疗健康信息而使用网络，国家卫生健康委员会主办的“健康中国”阅读量排名位于国家机关头条号阅读量第十位[4]。由此可见，互联网越来越成为社会公众方便、快捷获取健康信息的重要来源。

而随着信息化的不断发展，相关网络健康问题也不断出现。尤其是以“魏则西事件”等为代表的系类事件，引发社会公众对网络健康信息的担忧，这也使得越来越多的学者开始关注对健康网站相关的研究。网络的飞速发展使得健康信息可以通过网络的传递特性来传递，健康信息与网络的结合不仅是网络进一步的全面发展，也是健康信息更进一步的普及，依靠网络传递健康信息可以使更多的人在网络上较为方便的获取健康信息，从而满足自己日益增长的对于医疗知识的需求，缓解我国国内普通群众医疗信息匮乏的现状，从侧面改善我国国内医患关系紧张的局面。因此，近年来网络健康信息质量评价受到社会各界人士的关注[5]。

2 研究对象及评价指标

2.1 研究对象

本文主要选取中文在线健康网站对其进行质量评估，依据 Alexa 周排名，选择中文医疗健康网站排行榜前 50 名的网站，在此基础上，以 PR 值大小为标准，选取 PR 值大于 4 的共 38 个健康网站。PR 值是谷歌公司用来对网页的等级及重要性的一种方法，级别分为 0 到 10 级，PR 值在 4 以上就是一个相对不错的网站了，而 PR 值达到 7 以上就说明这个网站很受欢迎或极其重要。在这 38 个网站中，根据其网址链接，删除 6 个无法打开的网站，最后剔除纯医疗广告网站 2 个，最终确定符合条件的中文在线健康网站 30 个。见表 1.

表 1 纳入评价网站

序号	网站名称	LIDA 得分	Checklist 得分
1	39 健康网	64	27
2	家庭医生在线	64	7
3	MedSci-临床研究与学术平台	65	38
4	丁香园	68	47
5	寻医问药	54	9
6	有问必答网	60	38
7	好大夫在线	61	52
8	播种网	61	35
9	闻康资讯网	54	29
10	丁香通	57	44
11	99 健康网	63	33
12	生物谷	59	36

13	飞华健康网	60	41
14	爱爱医	61	30
15	中华康网	53	41
16	三九养生堂	58	31
17	薄荷网	50	47
18	医脉通	64	54
19	无忧爱美网	43	38
20	挂号网	61	32
21	大众养生网	56	35
22	健康界	63	55
23	中医世家	45	44
24	养生之道网	59	43
25	丁香医生	65	40
26	天天营养网	61	32
27	华医网	60	36
28	中国医药联盟	60	33
29	春雨医生	58	40
30	**医学百科**	**42**	**56**

注：Checklist 问卷得分删除 10 个无打分项题目

2.2 评价指标

本文主要选取 Michigan Checklist 问卷和 LIDA 量表对健康网站进行评价，并进行数据的统计。见表 2 评价工具[6-9]

表 2 评价工具

评价工具	创建时间	创建机构	类型	评价指标	指标数量	评分方式
Checklist	1999	密歇根大学	问卷	内容、易用性	43	Yes / No
LIDA	2007	英国 Minervation 咨询公司	量表	可访问性、易用性、可靠性	41	4 级 Likert 量表

3 结果

3.1 LIDA 量表分析

3.1.1 可用性

可用性对于网站来说十分重要，如果可用性很低，会增加用户的使用成本。而良好的可用性会增加网站用户的粘性和网站收入。利用 LIDA 量表对样本量网站从清晰度、一致性、功能性、可结合性等 4 个维度 18 个评价条目，通过平均值及标准差，分析样本量网站的可用性。具体数据见表 3 LIDA 量表可用性评价

表 3 LIDA 量表可用性评价

LIDA 量表可用性评价	题号	总分	均值	标准差
清晰度	2.1.1—2.1.6	18	14.00	2.10
一致性	2.2.1—2.2.3	9	8.57	0.68
功能性	2.3.1—2.3.5	15	11.60	1.43
可结合性	2.4.1—2.4.4	12	9.87	1.66

通过 LIDA 量表可用性进行评价，通过 SPSS 25 分析样本量网站的可用性总体得分均值为 40.03 ± 15.28（54 分），由表 3 数据可知：30 个网站质量评价各维度指标均数及标准差分别为清晰度 14.00±2.10（18 分）、一致性 8.57±0.68（9 分）、功能性 11.60±1.43（15 分）以及可结合性的 9.87±1.66（12 分）。

从功能性方面来看，半数以上的网站未达到均值，而此方面要求网站必须为用户提供合适的工具，以方便用户找到他们所需要的信息。通过分析数据我们发现，在功能性 5 个评价指标当中，对于浏览工具这一指标，样本量网站中有 10 个网站仅得到 1 分或是 0 分，这也说明了部分健康网站需要在浏览工具这一方面进一步优化。相比功能性方面，其他三方面半数以上的网站均超过均值，但是具体分析来看，在清晰度和可结合性方面尤其是明确使用对象、色彩以及非文本媒体三个评价指标部分网站得分不高，这也说明健康网站需要进一步明确使用对象，优化色彩，增加非文本媒体内容。

3.1.2 可靠性

70%的研究者发现信息质量是一个问题。对于健康网站来说，如果网站没有提供全面、相关、没有偏见的信息，这样就说明该网站不可靠且可能会对用户造成伤害。我们通过 LIDA 量表从网站的实效性、利益冲突、内容制作 3 个方面 9 个评价条目，对样本量网站进行评价并分析其均值及标准差，从而判断该网站的可靠性。

对于可靠性的评价条目中，内容制作程序、内容发布程序为附加条目，对网站整体无实际影响，因此删除这两部分条目后，剩余 9 个条目共计 27 分。通过数据分析，网站得分可靠性均值为 14.27 ± 4.43，实效性、利益冲突、内容制作三方面评价指标的均数及标准差为 4.77±2.78、3.07 ±1.98、6.43 ±1.74。具体数据见表 4LIDA 量表可靠性评价

表 4 LIDA 量表可靠性评价

LIDA 量表可靠性评价	题号	总分	均值	标准差
实效性	3.1.1—3.1.3	9	4.77	2.78

利益冲突	3.2.1—3.2.3	9	3.07	1.98
内容制作	3.3.1—3.3.3	9	6.43	1.74

通过上表数据我们发现，在内容制作方面样本量网站质量相对要比实效性和利益冲突这两个评价指标好一些。具体对是实效性和利益冲突问题单独进行分析，我们发现在实效性这一指标内约 53%的网站不能够及时对最新的事件做出报道；在 3.1.2 条目中，仅有 37%的网站能够拿开通评论功能，供用户进行交流；相对于对最新事件做出报道来说，57%的网站更新频率在三个月以内，其余的网站基本上在 6—12 月以内进行一次更新。据相关调查显示，如果一个网站没有定期更新内容，那么新的认知就会造成该网站信息质量下降。而对于健康网站用户来说，公开网站或文章的赞助商也是一个非常关键的问题，因此利益冲突这项指标针对该项内容进行评价，通过数据我们可以看到，多数网站通过其网站导航我们能够发现其负责人或者是公司；针对 3.2.2 条目，83%的网站都不能够披露其资金来源，虽然仅有 17%的网站对其网站运营进行融资做出了介绍，但也没有明确指出其运行资金来源；健康网站的运营团队对用户来说也很重要，但 57%的网站也没有对运营团队有明确说明，仅有 13%的网站能够明确指出运营团队，这或许也影响着用户对一个网站的判断及好感度。具体数据见下表 5 实效性、利益冲突评价指标分析

表 5 实效性、利益冲突评价指标分析

			N=30（%）	
LIDA 量表可靠性评价	**0**	**1**	**2**	**3**
3.1 实效性				
3.1.1 最新事件	16（53）	1（3）	0（0）	13（44）
3.1.2 评论功能	18（60）	0（0）	1（3）	11（37）
3.1.3 更新频率	0（0）	9（30）	4（13）	17（57）
3.2 利益冲突				
3.2.1 网站负责人	6（20）	4（13）	0（0）	20（67）
3.2.2 网站资金来源	25（83）	5（17）	0（0）	0（0）
3.2.3 网站运营团队	17（57）	7（23）	2（7）	4（13）

3.2 Michigan Checklist 量表分析

利用 Checklist 问卷对样本量网站从内容和易用性价值出发，通过是否列出作者的姓名和学历或者是提供了出版机构的联系方式、是否有清楚的网站介绍、是否包括一张目录或网站地图、是否包含不必要的大型图片等 43 个评价指标对样本量网站进行全面的评估并给出相对应的得分。

3.2.1 内容

从用户的角度来说，健康类网站中的内容会直接影响到他们的判断及日常行为。而大部分用户并非专业人员，他们对于网站内信息的内容无法做出非常准确的判断，只能依靠文章的作者及出处进行相对比较片面的辨别。所以，对于健康网站的内容进行评估可以说是重中之重。我们利

用 Checklist 问卷从内容角度（27 项评价指标），对选取的样本量网站进行评估，而 27 项指标中有 6 项无分数，只做是否或 A、B、C 选择。根据 SPSS 25 数据分析，我们发现网站内容部分得分均值为 17.53±8.60（53 分）。

从 Checklist 问卷内容维度，我们划分为权限、流通、信息、范围/选择、观众、价值、准确性、广告等 8 部分。而对于健康类网站来说，其准确性至关重要，我们对准确性部分 6 个指标进行评估。

表 6 准确性评价条目

	N=30（%）			
题目	+3	-3	+2	-2
18.文章来源是否原创	28（93）	2（7）		
19.文章有无参考文献			7（30）	23（77）
20.文章信息有无错误或遗漏	30（100）	0（0）		
21.文章有无主观意见或误导偏见	30（100）	0（0）		
22.文章内容信息是否准确			29（97）	1（3）
23.是否存在利益冲突			23（77）	7（30）

注：得分标准 18 题是=+3、否=-3；19/22 题是=+2、否=-2；20/21 题是=-3、否=+3；23 题是=-2、否=+2

根据表 6 准确性评价条目数据分析我们可以看出，健康类网站在信息准确性方面，样本量网站中 93%的网站文章来源都是自己原创，少数网站的文章来源可能借鉴与其他地方，但网站中存在参考文献的相关文章亦较少，仅占 30%；网站在其文章信息有无错误遗漏、有无主观意见或误导偏见方面质量相对不错，有 97%的网站其内容信息较为准确。在利益冲突方面样本量网站做的也相对较好，约 77%的网站不存在商业广告嫌疑，但少部分网站在这一方面做的较差。

3.2.2 易用性

对于普通人来讲，评判一个网站的好与坏通常都是打开网站的首页，然后查看该首页的美观程度或者布局等。而以上的这些在 Checklist 问卷中相对应的也就是网站的易用性。

在该维度中共有 16 项指标，分别从导航（8 项）、速度（2 项）、访问性（6 项）三个方面对网站进行评判。这三方面的总分分别是 21 分、3 分、7 分，其中在导航和访问性中分别有 1 项和 3 项指标只做判断不进行打分。具体数据见表 7 易用性评价指标

表 7 易用性评价指标

易用性评价	题号	均值	标准差	超均值量（N（%））
导航	28-35	12.47	4.46	14（47）
速度	36-37	2.27	1.53	7（23）
访问性	38-43	5.13	2.10	14（47）

注：题号包含判断指标，结果不影响均值及标准差

依据表结果我们发现这三方面的均值分别是导航 12.47 ± 4.46 分、速度 2.27 ± 1.53 分以及访问性 5.13 ± 2.10 分。整体来看，样本量网站中在导航及访问性方面，接近半数网站都能够达到均值以上水平，但部分网站在这些方面仍有待提高。但在速度方面，仅有 7 个网站能够达到均值以上水平，所以说多数网站在一些页面图片及服务器速度快慢上需要有很大的改进。

4 结语

通过对健康类网站的质量进行研究与评价发现，我国国内大多数健康类网站做的还有许多不足，仅有个别网站能够得到较多认同。当然，对于健康类网站的评价工具，我国应尽快出台是适合我国国情的评价指标、相关法律法规政策。国内部分学者对于健康类网站的评价目前正处于开始研究与摸索的阶段，对此国家及政府应加大支持力度，推动我国健康类网站的建设及评价系统的建立与完善，以促进我国医学界的研究发展，进一步推进我国的医疗卫生事业的发展。

PU-039

护理人员护理信息系统使用行为与工作绩效关系研究

刘智勇、春兰、周建、郭唯佳
华中科技大学同济医学院

目的 本研究应用理论研究和实证研究相结合的方法，旨在对护理人员使用护理信息系统行为与其工作绩效的关系进行探讨，为提升护理人员工作绩效及护理信息系统使用效能提供参考。

方法 一是文献分析法。通过检索国内外权威数据库，系统梳理国内外护理人员使用护理信息系统的相关研究，确定护理信息系统使用情况和护理人员绩效评价的各个维度，并以系统使用情况及工作绩效的关系作为研究内容，构建体现两者之间关系的理论模型。二是实证方法，根据对护理信息系统使用情况及工作绩效维度的确定，编制并发放调查问卷，采用 SPSS22.0 与 AMOS21.0 软件对获取到的数据进行整理并验证模型中变量间关系。

结果 提出绩效、个人使用、护理信息系统因素之间存在的关系，并对此进行验证。验证结果显示提出的 10 个假设均得到了数据支持。

结论 在提高护理人员绩效方面，可基于对系统易用性及功能的改善与优化来达成目标，使用系统的实际情况、满意度则直接对提升业务能力、工作效率、护理质量产生作用，因采取各项措施而改善的使用情况、提升的满意度等，均有利于提升护理人员业务能力、工作效率以及护理质量。

1 引言：

护理信息系统（nursing information system，NIS）是指将网络通信、计算机、信息技术等运用于护理工作中，以便采集护理管理信息、护理业务信息，并对此进行保存、查询的信息系统[1]，其作为护理信息化中的重要部分，对医院护理服务有重要影响。《2017-2018 中国医院信息化发展研究报告》显示，截至 2018 年底，我国临床护理信息系统的应用比例已达到 42.36%。在管理方面，

许多医院开始陆续使用护理信息系统来进行护理方面的各项事务管理[2]。近年来，许多研究明确运用信息技术对绩效有着正向作用，但在使用行为及绩效的路径以及相关因素方面，尚未形成统一框架。目前我国针对护理信息系统的研究主要集中在应用现状研究，针对临床应用效果的研究主要是经验型研究而实证性研究较少，护理人员使用护理信息系统的行为与其工作绩效之间的关系研究更是几乎空白。因此本研究决定从细节上验证护理信息系统使用行为与护理人员工作绩效之间的关系，为系统供应商开发护理信息系统、护理人员有效使用系统提供理论依据及数据支持。

2 资料与方法

2.1 研究对象

研究组向郑州大学第一附属医院和第二附属医院的临床科室护理人员发放问卷，问卷发放总量为400份，两所医院各200份，通过检验，共有372份属于有效问卷，回收率高达93%。问卷选用五点量表测量方法，量表共包括五种回答，即非常同意、同意、中立、不同意以及非常不同意，并以1到5分对此进行赋值。

2.1 研究方法

研究采用文献分析法及实证研究法。首先通过查阅中外数据库掌握护理信息系统及护理绩效的最新研究情况，再结合技术接受模型、信息系统成功模型的基础上，提出本研究的理论框架。其次，根据提出的理论框架，设置并发放问卷。最后，在实证研究部分采用SPSS22.0和AMOS21.0对问卷数据进行信度、效度分析，验证提出的理论模型及假设。

3 假设与模型

3.1 假设

针对目前护理人员护理信息系统使用行为和护理人员绩效之间的关系，通过文献阅读，本研究从信息技术接受模型、创新扩散理论、任务技术匹配理论中借鉴发展了“使用情况”、“满意度”等概念来构建了模型。

信息技术接受模型第一次提出了“使用情况”概念，这一概念表示用户为了完成工作目标而使用信息系统的想法，并受信息系统“感知易用性”和“感知有用性”的影响。在修改后的TAM2模型中，在TAM模型提出的指标上，又增加了主观规范、工作相关性、产出质量、结果明确性、经验、自愿性等调节变量，明确指出外部环境和感知工具是判断使用意愿的指标。TAM2指出，在一定时间段内即使用户对某项新技术有消极抵触的心理，仍然会因为新技术所提供的功能及便利程度而选择使用该项新技术，此时用户新技术使用行为受其使用态度的影响就很小，因此TAM2模型中删除了使用态度这一指标。在任务基础匹配理论中提出，只有当工作任务跟新技术相匹配的时候才会有使用行为发生，从而影响绩效。结合信息系统成功模型来看，系统、信息质量将对使用者的使用行为产生决定作用，使用行为对绩效产生直接影响。通过总结相关理论会发现不管是强调用户的使用态度或者使用意愿还是强调系统功能和工作的匹配程度，都是需要通过用户使用情况来达成对改变绩效的目的。Dubois等指出，信息技术在护理实践中被应用才能对护理结果产生影响[3]。因此，本研究的前三个基本假设为：

H1：护理信息系统的使用情况和护理质量绩效有正向影响关系

H2：护理信息系统的使用情况和工作效率绩效有正向影响关系

H3：护理信息系统的使用情况和和业务能力绩效有正向影响关系

本研究的前提之一是信息系统带来的绩效提高由用户的角度来测量的[4]。所以，用户对系统的满意度比用户的使用习惯更容易对系统造成影响[5]。Cardozo 首次明确顾客满意度这一概念，提出客户满意某一产品的程度将对二次购买意愿产生直接影响[6]。许多学者都肯定满意度对用户工作绩效的直接影响，用户满意度越高用户工作绩效也会越高，随着绩效改进用户会继续使用信息系统[7-13]。因此，本研究提出以下假设：

H4：护理信息系统的用户满意度和护理质量绩效有正向影响关系

H5：护理信息系统的用户满意度和工作效率绩效有正向影响关系

H6：护理信息系统的用户满意度和业务能力绩效有正向影响关系

感知易用性，即在信息系统的运用方面，用户明确系统操作的难易程度，结合此前提及理论来看，表明感知易用性将对用户运用信息系统的态度产生影响。信息系统的技术模块化、记录结构化、页面简洁化等设计会使用户在使用该信息系统的时候更加简单有效，从而会影响用户对信息系统的使用行为。除此之外，有研究表明信息系统的易用性会直接影响用户满意度。Jaspers 通过研究内科医生满意电子病历的程度以及二次设计电子病历之间的关系，提出内科医生更为满意电子病历的关键原因，即电子病历在二次设计后具有更高的易用性[14]。何其慧等通过文献回顾发现系统易用性是影响智能信息系统满意度的重要因素[15]。从而，提出以下两个假设：

H7：护理信息系统感知易用性正向影响系统的使用情况

H8：护理信息系统感知易用性正向作用于用户满意度

用户的满意度以及实际使用行为将会因信息系统功能受到一定影响，而用户的行为态度受系统功能的影响程度则取决于“系统功能的覆盖范围和用户的工作任务”[16]。任务技术匹配理论提出，信息系统的功能模块越多，越为匹配于工作任务，用户在完成工作任务时选用此种系统的意愿将更高，诸多研究已对上述方面有所证实。通过研究，Sicotte 提出，若 PACS 系统具备更为健全的功能，医生将对此产生更高满意度，当一个系统所提供的功能跟医生的临床工作任务很匹配时，医生就会更满意[17]。李森等研究指出，采取改进 PDA 软硬件、提高无线网络支持力度等措施来用于改善 PDA 的性能会提高护士对移动护理信息系统的使用满意度[18]。实证研究表明，随着系统提高系统软硬件条件，用户满意度会提高，用户会更加有效使用系统从而提高工作绩效[19-21]。因此，本研究提出以下假设：

H9：护理信息系统的系统功能正向作用于系统使用情况

H10：护理信息系统的系统功能正向作用于用户满意度

本研究的模型认为护理信息系统的功能和易用性特征会影响护理人员对护理信息系统的使用情况和满意度，而使用情况和满意度最终影响护理人员个人绩效。

3.2 模型

根据上面提出的假设，本研究初步构建了护理信息系统使用行为与护理人员绩效关系的结构方程模型，见图 1。测量指标见表 1。

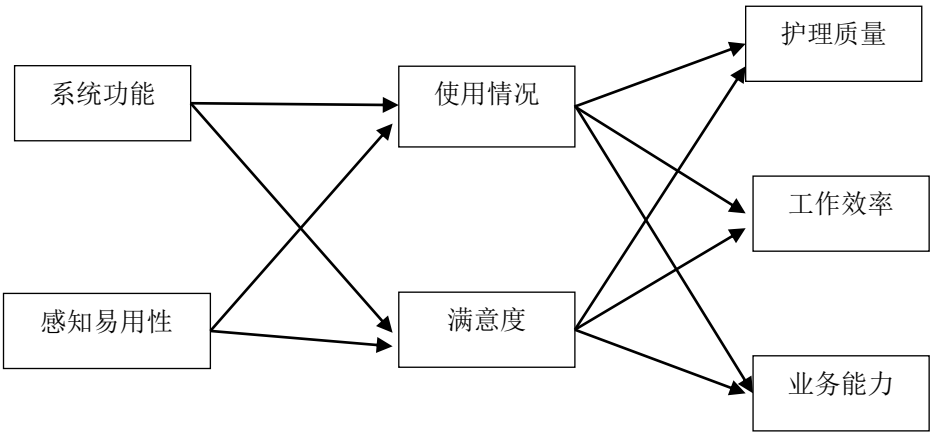


图 1 模型图

表 1 测量指标

因子	编码	观察变量
系统功能（SF）	SF1	使用护理信息系统后医嘱可以及时准确下达至护士工作站
	SF2	使用护理信息系统可以准确记录医嘱的实际执行者及执行时间
	SF3	使用移动护理信息系统可以对患者身份进行识别
	SF4	使用护理信息系统可以对患者的护理信息进行记录、更新和存储
	SF5	使用移动护理信息系统可筛选出各时间点需测量生命体征的病人
	SF6	护理信息系统可以设置巡视时间段，系统会按点提醒护士进行巡视
	SF7	使用护理信息系统可以查询患者的电子病历
	SF8	用护理信息系统可以对各类护理电子病历进行质量控制
感知易用性（PEU）	PEU1	护理信息系统对我来说模块化清晰，易操作
	PEU2	我录入信息的时候，记录病人信息的结构是相似的
	PEU3	便于统计以及修改数据，可自动针对损坏数据进行修复
	PEU4	在过去的几个月我都在使用护理信息系统
使用情况（USE）	USE1	我使用护理信息系统的方法是正确有效的
	USE2	我使用护理信息系统的方法是正确有效的
	USE3	我已运用系统的全部功能

	SAT1	我认为护理信息系统的使用很不错
	SAT2	系统运用进程极为愉悦
满意度	SAT3	愿意尝试系统之上存在的各类模块
（SAT）	SAT4	使用护理信息系统保证了护理工作的准确性及及时性
	SAT5	使用护理信息系统很好的保障了数据的安全
	NQ1	使用护理信息系统提高了我的护理服务质量
	NQ2	使用护理信息系统提高了我的护理文书书写质量
	NQ3	使用护理信息系统提高了医嘱执行的准确率
	NQ4	使用护理信息系统提高了基础护理质量
护理质量	NQ5	使用护理信息系统提高了专科护理质量
（NQ）	NQ6	使用护理信息系统提高了危重患者护理质量
	NQ7	使用护理信息系统提高了健康教育知晓率
	NQ8	使用护理信息系统减少了护患矛盾
	NQ9	使用护理信息系统后降低了压疮、坠床跌倒等护理不良事件的发生率

续表 1 测量指标

因子	编码	观察变量
	NQ10	使用护理信息系统增加了患者满意度
工作效率	WE1	使用护理信息系统减少了我的护理工作量
（WE）	WE2	使用护理信息系统让我有效管理更多的患者
	WE3	使用护理信息系统增加了我的护理操作项目数
业务能力	BA1	使用护理信息系统提升了我的应急抢救能力
（BA）	BA2	使用护理信息系统提升了我的护理文书书写能力
	BA3	护理信息系统提升了我正确评估患者的能力
	BA4	使用护理信息系统提高了我的理论操作考核分数

4 结果

4.1 信度和效度检验

本研究采用 SPSS22.0 对问卷的信度和效度进行检验。得出 7 各因子对应的 CR 值和 Cronbach’s α 值均>0.7，表明问卷信度较好；同时各因子的标准化因子负载和平均方差提取值（AVE）均>0.5，表明样本数据的收敛效度和区分效度较好。

4.2 探索性因子分析

4.2.1 KMO 和 Bartlett 检验

样本数据的 KMO 值为 0.943，表明样本适宜于开展因子分析。Barttllet 球形检验最终获得卡方值为 10885.981，P 值相对显著，说明公因子存在于样本数据相关矩阵内，可开展因子分析工作。

4.2.2 探索性因子分析结果

以 Kaiser 准则为依据，以 1 作为特征值开展探索性因子分析，再结合主成分分析法对因子负荷量进行估计，再通过最大方差法进行旋转并解释因子。本研究运用上述方法提取了 7 个公因子，累计解释方差对应于 73.374%，因子结构清晰，各个测度项在相应因子上的负载大于 0.6。

4.3 结构方程模型分析

通过 AMOS21.0 分析可知，本研究提出的模型可识别，且拟合度较高，可以进一步做分析，拟合结果见表 2。图 2 显示了因子间的路径系数，本研究的假设验证结果见表 3。

表 2 拟合优度检验

拟合优度指标	指标结果	理想值	说明
CMIN / DF	2.341	2-5	理想
RMSEA	0.060	>0.05	理想
GFI	0.828	>0.80	理想
AGFI	0.804	>0.80	理想
IFI	0.922	>0.90	理想
CFI	0.922	>0.90	理想
TLI	0.916	>0.90	理想
NFI	0.872	>0.90	可接受

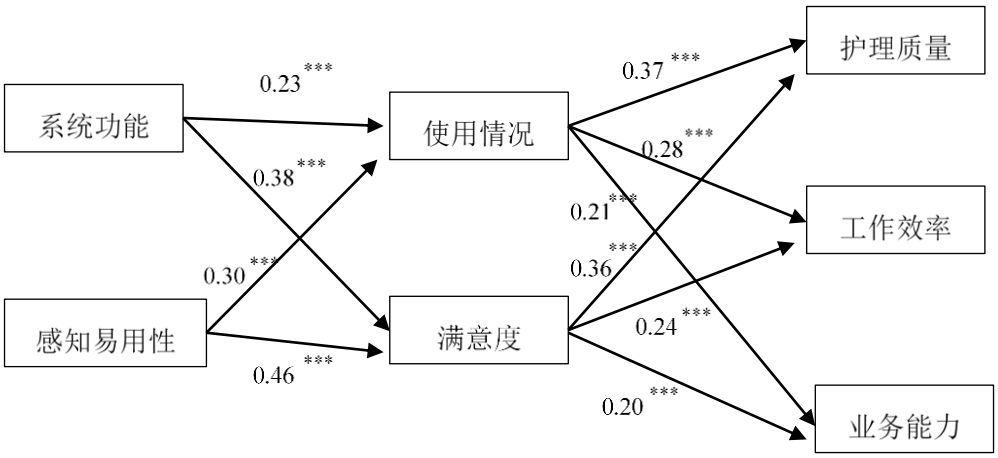


图 2 路径传导系数图（***：p <0.001）

表 3 假设验证情况

假设	假设内容	是否支持
H1	护理信息系统的使用情况和护理质量绩效有正向影响关系	支持
H2	护理信息系统的使用情况和工作效率绩效有正向影响关系	支持
H3	护理信息系统的使用情况和和业务能力绩效有正向影响关系	支持
H4	护理信息系统的用户满意度和护理质量绩效有正向影响关系	支持
H5	护理信息系统的用户满意度和工作效率绩效有正向影响关系	支持
H6	护理信息系统的用户满意度和业务能力绩效有正向影响关系	支持
H7	护理信息系统感知易用性正向影响系统的使用情况	支持
H8	护理信息系统感知易用性正向作用于用户满意度	支持
H9	护理信息系统的系统功能正向作用于系统使用情况	支持
H10	护理信息系统的系统功能正向作用于用户满意度	支持

5 讨论与建议

本研究以技术接受模型和信息系统成功模型为基础，按照护理信息系统使用行为和临床护理人员工作绩效的特点，构建结构方程模型，并通过实证分析验证了假设模型。结果显示，提出的所有假设均获得支持。

使用情况正向作用于护理质量、护理人员工作效率及业务能力。即，临床护理人员越多使用护理信息系统就越有利于个人绩效的提升。

满意度对护理质量、护理人员工作效率及业务能力具有显著的正向影响。满意度可用护理人员运用系统的态度加以体现，满意度越高越能改善护理服务质量、护理人员工作效率及业务能力。至于如何提高满意度，本研究运用的问卷则为此提供相应思路，可结合系统功能的改进、数据安全保障等诸多方面，对此加以改进。

感知易用性显著影响护理信息系统使用情况及用户满意度。感知易用性可对个人运用系统时的难易程度有所体现，外部变量将对此产生决定作用。如系统、用户特征、工作特征、开发或执行过程中的本质、组织环境、政策影响等均属于外部变量。实证研究表明，提高护理信息系统的感知易用性将利于护理人员绩效的提升，表明可结合系统感知易用性的提升，以便提高绩效，如自动生成相关记录单、加强录入信息的模块化等，以便护理人员实际工作量得以减少，但需要关

注如下方面：即不可为对结构化加以无限追求而结构化，若设计过于结构化时，将使得护理人员难以有效运用，无法在短时间内找到合适的选项，从而对护理人员的临床工作造成负担，需首先对护理人员的实际运用需求进行细致、广泛的调研，此后确保系统控件以及界面设计的合理性。

系统功能显著正向作用于使用情况及用户满意度。这证明系统功能越完善和护理人员工作任务越匹配就会有更多的护理人员使用护理信息系统，证实了任务技术匹配理论中的观点。但是，目前还是存在护士使用系统不熟练，不能发挥护理信息系统应有的功能的情况。为了最大程度的减少这种现象，建议软件公司定期对护理人员进行培训，在医院对新系统使用进行培训，往往多见于最初投入运行新系统时，后期几乎不会再开展培训工作。软件公司需要以培训护理人员作为定期项目，并通过将专人派遣致医院，以便护理人员能够通过专业人员的解答、熟练掌握系统的操作技巧，而不仅仅只是在系统刚投入应用的时候才有。更重要的是，护理信息系统的软件商应该根据护理人员需求设计出系统，在系统开发和设计过程中注重护理人员与信息技术人员的沟通，将护理人员的期望和需求传达给系统开发人员，并且随着国内护理信息系统应用的普及，有必要在系统的性能及操作细节方面展开研究，保证护理人员对系统每个部分都非常熟悉，避免出现对护理人员运用系统产生负面影响的状况。上述结果凸显系统对信息需求满足方面的关键作用，而就系统方面的软件供应商而言，可通过对此次成果的有效运用，明确满意度偏低的软件功能，以及哪部分功能将有效提升绩效。此外，基于此次研究，护理人员也可确定绩效如何依托于系统运用得到改进，进而鼓励相关人员能够更为广泛的运用此系统。

目前的研究显示，临床上护理信息系统的应用并不普及[22]，这可能是因为在临床实践中信息系统所提供的功能无法满足临床工作需要，但本研究认为在信息技术突飞猛进的情况下，信息系统的开发与设计是完全可以满足临床工作需要的，更深层次的原因还是与护理人员自身的信息技术接受及创造能力相对薄弱有关。研究表明[23]，学历越高、年资越低的护士对护理信息的重视程度越高，这应该与年轻护士在个人学习生活中接触更多的信息技术有关。因此，我们应该加大推行护理信息学教育来不断提升护理人员的信息接受及创造能力，进一步发挥年轻护士的信息素养优势，定期给予护理信息技术使用培训，促进护理信息系统的使用。

本研究的不足之处在于，样本选择的局限性以及未对护理团队绩效进行分析。

从样本的选择来看，本研究的样本仅来源于郑州地区的两家三甲医院，存在区域及数量限制，后续的研究选取的样本量应该覆盖全国东、中、西地区，如此才能得出具有更高层面指导意义的分析结果。同样由于样本量不足，本研究只是从护理人员个人绩效角度进行研究，护理信息系统的使用应该对护理团队绩效也有所影响，而本研究并未对此进行评价。

PU-040

基于电力载波技术的智慧医院配电网监测系统设计

贾志刚
郑州大学第一附属医院

目的 为了监测智慧医院电网的运行状态，快速定位医院内配电网的故障位置。

方法 首先，利用电力载波技术将实时采集到的电网运行状态信号通过院内配电网传输到院内电网监测中心，其次鉴于院内配电网为树形结构，电网监测中心上位机对配电网各节点进行遍历，最后对比父节点与子节点的信号数据，从而快速定位配电网的故障位置。

结果 结果表明该配电网监测系统监测精度高、故障诊断准确率高、反馈速度快

结论 验证了系统的实用性和科学性。

PU-041

基于大数据背景下的病案统计管理分析

郭斌
郑州大学第一附属医院

目的 在“互联网+医疗”的大数据背景下，病案统计信息化建设在医院现代化管理中越来越重要，医院信息系统每天都会产生大量数据，病案管理和信息统计工作繁杂琐碎，需协调临床、护理、医技、行政等多部门共同协作。只有通过信息化手段，对产生的数据进行深度挖掘以及通过对比分析后才能真正实现数据的价值。本文旨在提高病案统计信息化管理水平。

方法 本文基于大数据背景下，利用现代化信息技术对病案统计发展模式、数字化病案系统建设价值、大数据在病案统计分析中的应用以及大数据环境下病案管理的发展策略进行探讨。

结果 数字化病案系统建设一方面是发展的必然趋势，是“互联网+医疗”背景下医院参与激烈医疗市场竞争的必经之路，另一方面它也极大地提高了病案统计的工作效率和工作质量，能够为医院决策提供更多有价值的信息。大数据在病案统计分析中的应用包括运用高速扫描仪使信息存储更方便、快捷，利用数据库检索使统计查询更准确、快速；在首页中设置必填项和填写范围，规范和约束病案首页书写行为，提高规范性；通过信息化手段提高病案首页填写质量，从而提高准确性，实现更有效的医疗管理；通过系统集成自动抽取病历数据转换生成 PDF 文件，促进无纸化办公，等等。

结论 在当前医疗卫生事业不断发展和完善的背景下，医疗卫生病案统计具有重要的社会意义和价值。在建设数字化病案的过程中要注重统计学与信息系统的融合度，积极适应医疗工作的流程，为医院临床医务人员以及医院管理人员提供更好的服务。医疗卫生机构应当迎接信息化系统建设发展的困难，创新发展形式，整合经营流程，促进医疗卫生建设的良性、高效、稳定发展。本文

建议一是规范病案数据的表现形式，适应国家的制度要求，使用标准编码；二是提高病案统计人员的专业程度，建立完整的病案管理队伍，重视医疗统计人才；三是健全病案统计信息系统，将电子病历、HIS、LIS、PACS 等信息系统之间形成一个稳定的信息共享机制，进而形成具有整体性和系统性的数据库；四是建立健全病案统计的法律法规，提升工作人员的法律意识，加强对临床诊疗医生的监督和监管。

在“互联网+医疗”的大数据背景下，病案统计信息化建设在医院现代化管理中越来越重要，医院信息系统每天都会产生大量数据，病案管理和信息统计工作繁杂琐碎，需协调临床、护理、医技、行政等多部门共同协作。只有通过信息化手段，对产生的数据进行深度挖掘以及通过对比分析后才能真正实现数据的价值。为了提高病案统计信息化管理水平，本文基于大数据背景下，利用现代化信息技术对病案统计发展模式、数字化病案系统建设价值、大数据在病案统计分析中的应用以及大数据环境下病案管理的发展策略进行探讨。

病案统计工作在现阶段的医院经营管理中具有重要作用，能够体现医疗卫生体系改革的相关进程，是现代医学发展中具有重要意义的环节。病案统计指的是根据患者详细资料和诊疗过程进行信息记录，将详细的数据转化成为具有直观性的信息，进而提炼成为专业化的知识。简言之，就是对病案数据进行收集、整理、加工、分析，按照相关部门要求做出数据统计表格。随着各大医院信息化建设进程不断加深，病案统计的方式方法也显示出其自身的独特性，大量的患者诊疗信息被整合成具有临床研究意义的知识，对于医疗卫生事业的发展具有重要意义。但是，医院之间经营管理上存在一定差异，并且病案记录人员的专业化水平参差不齐，造成我国现阶段病案统计工作推进的局限性。

一、 病案统计发展模式

病案统计工作的产生和发展经历了从手工记录和统计到信息化的统计两个阶段，这种记录模式的转变有效提升了病案统计的工作效率，并且将信息化的病案统计工作糅合进医院的经营和管理中。

（1）手工统计

在医院病案记录早期，主要是依靠人工进行记录，记录的载体是纸质文本，基本上没有使用计算机进行记录。这种记录方式较为陈旧简单，并且难以满足医院管理的需求。即便是在后期进行了一定的计算机信息整合，大多数仅用于数据计算，缺少有效的信息化支持。

（2）信息化为核心

随着信息化的不断深化发展，病案记录也逐渐趋于信息化和电子化，医院中也设立了专门的病案统计部门整合相关资料和信息，有效提升了医院病案统计工作的效率。在大多数医院管理机制中，病案统计已经成为了成熟稳定的医院管理部门，在整合记录信息的过程中发挥了重要作用。

（3）大数据环境下

大数据环境下的病案统计转变了传统统计模式中提取门诊诊疗信息的形式，将患者在医院的所有诊疗信息进行有效提取，能够进行多项研究。利用医院建立的信息数据库，实现对患者大数据的搜集和整理工作，为医院的经营奠定良好的基础。通过信息化技术的使用，医院建立了 HIS 系统、LIS 系统、PACS 系统等，并对系统中包含的相关信息进行有效整合。在病案操作中能够便捷高效地进行查找和比对，实现了病案统计的重要研究价值。

二、数字化病案系统建设价值

病案统计学的完善和发展得到了信息化的高度支持，数字化病案管理是发展的必然趋势。近年来随着互联网信息技术的不断发展，实现互联网与传统产业联合，“互联网+医疗”应运而生，其核心内容是医疗大数据，是医疗的数字化。病案统计信息系统是医疗信息化建设的重要组成部分，传统模式的病案管理流程在当今互联网信息时代中已经显现出不足，医院如果要想在竞争激烈的医疗市场中取得长足发展，必须依靠信息化建设，利用信息化技术取缔传统的人工统计，数字化病案管理是发展的必然趋势。

数字化病案就是利用信息化技术协助完成病案统计工作，患者从门诊挂号到检查、检验，再到入院治疗等，全部通过计算机信息技术记录每一个诊疗过程，确保了采集数据的真实性和实效性。系统通过采集医嘱、病历、护理信息，进行归档、统计、分析、生成各种所需的报表，与传统的人工统计相比，极大地提高了病案统计的工作效率和工作质量，并且降低了出错率，病案管理人员由复印病历改为打印病历，复印窗口工作人员数量大幅减少，病案的传递及流转无需通过手工处理，减少了中间的流转环节，能够让病案管理人员有更多时间用于病案数据的统计和挖掘，为医院决策提供更多有价值的信息。

病案信息的统计与分析，对于现代化医院管理有非常高的价值，能够为医院的经营管理服务，为医院管理者及时、准确地提供医院业务管理数据，确保数据的实效性，还可根据需求对病案数据进行分析；能够为临床科研、教学以及管理服务，通过数字化病案管理系统的建设，信息统计工作由被动变主动；能够为医疗质量管理服务，通过数字化病案建设，可以实时监测病案首页质量、病历书写质量以及归档率，为医疗监督评估提供依据。

三、大数据在病案统计分析中的应用

（1）提高应用效率

在病案管理中，利用信息化技术，通过高速扫描仪将归档的病历存储到数据库中，使病案信息存储工作更加方便、快捷，提高病案存储的周期。利用数据库检索，在信息系统前台设计多条件检索查询功能，使病案信息的统计与查询更加准确、快速。

（2）提高规范性

在电子病历首页中，利用信息化技术，按照电子病案首页填写规范要求，在首页中设置必填项，并通过逻辑关系设置填写范围，例如损伤、中毒的外部因素编码范围应为：V、W、X、Y 开头的编码；确诊日期不得大于出院日期或小于入院日期；入院日期不得大于出院日期；手术日期必

须在入、出院日期之间等等。通过以上设置，对较为随意的病案首页书写行为进行规范和约束，使信息统计数据更加规范、准确。

（3）提高准确性

利用信息化技术，能够有效提高病案首页填写质量，使信息统计以及数据更加准确，更能够客观地反映医院医疗数据，对医疗质量管理有积极推动作用，实现了更有效的医疗管理。通过对医疗数据整合、实时分析、报表制作、数据挖掘，将数据变为可用的信息，使信息系统为公共卫生、行政管理、医疗服务提供更可靠的支持。

（4）促进无纸化办公

通过与院内 HIS、EMR、LIS、PACS 系统做数据对接自动抽取数据转换生成 PDF 文件，无电子模板的纸质医疗文书通过高速扫描转换生成 PDF 文件，对采集的数据进行审核，合格后进行自动归档管理及离线存储，并能够实现病案打印、病案借阅、病案检索、病案示踪、统计查询等功能。

四、大数据环境下病案管理的发展策略

（1）规范病案数据的表现形式

病案统计中大数据的建立应当奠定在病案数据的标准化基础上，2016 年我国已提出了手术分类编码相关标准，其中重要的意义是为病案大数据的建立奠定基础。医疗卫生部门应当不断迎合国家推出的相关政策，完善自身的管理和规划，适应国家的制度要求，尽早熟悉相关的规定和标准，使用标准编码。

（2）提高病案统计人员的专业程度

医院应当建立完整的病案管理队伍，重视医疗统计人才，保证医院在日常管理中能够有效解决统计相关的问题。在病案统计人员队伍中不可缺少的是具有临床经验的工作人员，保证病案管理中具有熟悉临床诊疗方式的专业人员，能够明确病案中的数据情况，提升病案质量。

（3）健全病案统计信息系统

医院在选择病案统计系统时应当确保系统的选择与自身的经营发展需求能够高度融合。将电子病历、HIS、医学影像、输血等信息系统之间形成一个稳定的信息共享机制，病案管理人员能够通过管理将患者的相关数据进行整合和提升，进而形成具有整体性和系统性的数据库，对于相关数据应当进行有效的分析，整理成为统一的报表，便于日后的审核和利用。卫生行政部门应当对医院进行相应的监督和监管，接受医院阶段性的上报数据，并且建立一定的数据审核标准，对医院上报的数据进行有效评估，督促医院在日常的管理和经营过程中不断完善。

（4）建立健全病案统计的法律法规

病案统计进行电子化监管有效提升了病案数据的整合和提取，但是一定程度上也增大了患者个人信息和隐私的泄漏风险。医疗卫生行业的大数据建立，应当强调相关的规章制度，提升工作人员的法律意识，保证工作人员在合理合法的范围内加强工作效率和质量。同时，加强对临床诊疗医生的监督和监管，杜绝随意修改病案的现象，对于违规医生进行严格的处理，保证医院病案的准确性和合理性。

在当前医疗卫生事业不断发展和完善的背景下，医疗卫生病案统计具有重要的社会意义和价值。在建设数字化病案的过程中要注重统计学与信息系统的融合度，积极适应医疗工作的流程，为医院临床医务人员以及医院管理人员提供更好的服务。医疗卫生机构应当迎接信息化系统建设发展的困难，创新发展形式，整合经营流程，促进医疗卫生建设的良性、高效、稳定发展。

PU-042

医疗健康中的药物物联网与大数据

徐勐
郑州大学第一附属医院

目的 许多科技可以降低预防或管理慢性病的总成本。这些包括持续监控健康指标的设备、自动执行治疗的设备、或在患者自行实施治疗时跟踪实时健康数据的设备。因为他们增加了对高速互联网和智能手机的访问，许多患者开始使用移动应用程序（apps）来管理各种健康需求。现在越来越多的设备和移动应用程序通过医疗物联网（mIoT）与远程医疗和远程健康集成。本文综述了医疗领域的 mIoT 和大数据。

方法 mIoT 是医疗保健数字化转型的一个关键部分，因为它允许出现新的商业模式，并能够改变工作流程、提高生产力、控制成本和增强客户体验。

结果 如今，可穿戴设备和移动应用程序支持健身、健康教育、症状跟踪以及协作性疾病管理和护理协调。所有这些平台分析都可以提高数据解释的相关性，减少最终用户拼凑数据输出的时间。从大数据分析中获得的见解将推动医疗保健世界、业务流程和实时决策的数字化颠覆。

结论 一种新的“个性化预防性健康教练”（数字健康顾问）将出现。这些工作人员将具备解释和理解健康和福利数据的技能和能力。他们将帮助他们的客户避免慢性疾病和饮食相关的疾病，改善认知功能，改善心理健康，全面改善生活方式。随着全球人口老龄化，这些角色将变得越来越重要。

PU-043

基于 Spark 的电子病历多维检索系统研究与应用

李育申、葛晓伟
郑州大学第一附属医院

目的 对海量非结构化电子病历数据的检索与分析需求，有效改善传统电子病历非结构化存储难以扩展及检索分析效率低下的问题。

方法 利用 Hadoop 及 Spark 大数据存储与处理技术，设计并实现一种基于 Spark 的电子病历多维检索系统。

结果 采用基于 JavaWeb 开发实现电子病历结构化多维检索与分析系统，用户通过系统界面查看分析结果，并将数据库中的计算结果返回给前端，利用开源可视化库 Echarts 插件对数据结果进行图形化展现。对现有病历资料从病种人数、患者出院科室、患者来源地区等几个方面进行分析并形成图表，方便用户对数据有一个更直观、全面地掌握。

结论 基于 Spark 的医疗大数据处理模型通过利用 Hadoop 及 Spark 大数据存储与处理技术，能够满足海量医疗数据下细粒度的统计与计算模式，并在存储或处理的数据量激增时，通过增加节点的方式提升分析处理和存储的能力，具有良好的扩展性。

PU-044

双向聚类方法在医学文本挖掘中的实现过程及应用意义

杨颖[1]、李子政[2]、许丹[1]、徐爽[1]
1. 中国医科大学 2. 中国医科大学附属第四医院

目的 探索双向聚类方法在医学文本挖掘中的实现过程及应用意义，以期促进双聚类在数据挖掘中广泛应用开展。

方法 介绍双聚类的概念和基本原理，梳理国内外双聚类发展的状况。列举双聚类的使用工具和方法，并通过案例展现双聚类在医学文本挖掘中的具体实现过程。

结果 通过案例得出双向聚类方法在医学文本挖掘中的意义，最终提出双聚类的应用展望。

结论 双向聚类在医学领域尤其是生物学领域具有广大发展前景，帮助医学科研人员有效地分析海量数据，并以更高的效率生成有价值的知识和新颖见解。

1、双向聚类的概念及原理

当今大数据驱动的信息发展时代，积累大量的数据，如何利用数据挖掘技术从海量数据中抓取有价值的信息具有重要意义。书目文献数据库中可用于文本挖掘共现分析的字段主要有论文的作者、机构、引文、关键词或主题词，其中引文的共现分析又可以进一步分解为作者同被引、期刊同被引等等。最早开展共现分析的是高被引论文的同被引聚类分析。另一条研究主线则是对主题词或关键词的共现分析。根据这些共现数据，很多学者开展了对高被引论文、高被引作者和高频主题词的聚类分析，以此将这些分析对象分类。常用到的传统聚类方法——系统聚类、快速（K-means）聚类，都是单向的聚类，即只对行（对象）或者列（属性）进行聚类，有一定的局限性。比如，有的类是扩散且互相渗透的，那么不同算法得出的结果不同，传统的聚类分析不适用于类群界限相互渗透的数据分析。另外，单向聚类分析很容易忽视系统多样性以及多元因素的非

线性因素[1]。因此，面向高维数据的挖掘方法尤为迫切，双聚类算法应运而生。其基本思想是：直观简单地确定数据矩阵的几个双聚类，同一类别中的个体有较大的相似性，不同类别中的个体差异很大，于是根据多个观测指标，找出能够度量样本(变量)之间相似度的统计量，并以此为依据，采用各种聚类算法，将所有的样本(变量)分别聚到不同的类别中。其原理是：简单确定数据矩阵的几个双聚类，矩阵中的对象根据值标上相应颜色，形成一幅矩阵图像。再重新排序行和列以聚集相似的行和列，进而形成具有相似颜色块的图像。这些块具有相似的表达值，它们是行的子集和列的子集，因此是双聚类[2]。双向聚类可以有效抓取高维数据中隐藏的局部特征信息。

2、双向聚类在数据挖掘中的应用综述

双聚类的思想最早由Hartigan于1972年提出的：将对象在不同属性下的取值看作一个矩阵，根据数据矩阵的对象和属性分别从行和列两个维度同时聚类，这样可以找出其中满足条件的各个小矩阵，即由对象子集和属性子集组成的聚类。2000年，文献[3]应用双聚类算法模型挖掘基因表达数据，之后Mirkin在文献[4]提出“盒式聚类”、“双向聚类”等类似的概念。通过调研国内CNKI数据库，发现有420篇关于双聚类的中文文献报道。学科领域以计算机领域为主，其次是医药领域(包括临床医学，基础医学，公共卫生，口腔，护理，生物，中医，药学)，再次是图书情报领域，教育，商务领域。大部分文献是对双聚类方法的应用，如，在医学数据挖掘中，有文献选择全球医学研究领域自然指数期刊为研究对象展开科学计量分析，利用gCLUTO对其词篇矩阵进行双向聚类,探索其研究前沿结构构成[5]；有文献利用共词分析挖掘国际护理信息学、乳腺癌相关酶、h指数研究领域的研究热点[6-9]。甚至在临床医学的博士论文中有所应用，利用双聚类研究SIRT6的计量学特征，与临床基础相结合[10]。gCLUTO双聚类不仅在图情领域的文献计量学中的简单应用，也有临床医生在临床实践中的应用，比如文献[11-13]分别从人类神经干细胞的主题趋势和知识结构、支架植入在胰腺疾病治疗中的应用研究趋势、在绝经后骨质疏松研究中定位主题趋势和热点识别，均是临床医生应用双聚类挖掘其感兴趣的医学领域相关知识。

3、双向聚类工具介绍

经历几十年的发展已经开发了数十种双聚类算法和工具，尤其在生物医学领域有着广泛的应用，以增强从高通量组学技术产生的大型数据集中进行分析的能力。这些算法和工具已应用于多种数据类型，包括但不限于基因组，转录组，外显子组，表观基因组，表型和药理基因组。一些通用的流行算法有SAMBA，ISA，BIMAX，QUBIC和FABIA等。另外，已经开发了几种工具（gCluto，R包，Web服务器等）来促进具有有限计算机背景的用户。 GEMS 是基于Gibbs采样范例的用于基因表达挖掘的Web服务器，而biclust和QUBICR是两个R包，集成了多种现有算法，数据预处理功能以及结果的解释和可视化。本文主要介绍应用较多的gCluto和新开发的基于R包的PubMR两种资源的获取和应用。

3.1 gCluto。聚类分析通常应用SPSS、SAS等统计分析软件来实现。目前针对双聚类分析较常用的软件是gCLUTO（Graphical Clustering Toolkit，图形聚类工具包）。它以用户友好的图形方式进行聚类，并提供交互式聚类结果可视化的方法。通过网址

http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/download可下载gCluto，并安装使用。实际的可执行文件必须在其所在的文件夹内才可保证程序的执行。

3.2 PubMR。中国医科大学崔雷教授和周晓北教授近期联合开发了新的数据挖掘共现系统——pubMR，是R平台下一个高效的PubMed文本挖掘工具，集合了：检索下载、解析抽取、基本统计、多维矩阵、论文相似、热点分析、概念识别和网络分析等多种功能。利用R语言嵌入双聚类的程序包，可以通过简单的R语言进行复杂的双聚类操作，实现聚类矩阵的可视化（https://github.com/xizhou/pubMR）。

4、双聚类在医学数据挖掘中的实现过程

实现双聚类在数据挖掘中的步骤需要四步：1、收集数据，选择数据库和研究方向，检索信息，下载数据，对于共词分析选取XML格式；对于引文分析，下载引文记录。2、数据预处理，对下载数据进行处理，利书目共现分析系统[14]（Bibliographic Item Co-Occurrence Matrix Builder，BICOMB）抽取高频词或高频引文，再生成词篇矩阵（引文-引证文献矩阵）与共词矩阵（高被引文献的共引文矩阵）；3、双聚类分析及可视化，在gCLUTO里输入数据矩阵，对结果可视化分析，结果更加直观、可读性强。选择合适的类别，进行双向聚类分析，根据聚类报告中参数对照表、山峰图和可视化聚类结果选取聚类效果最佳聚类；4、结果解读：根据聚类参数对照表和可视化聚类结果以及山峰图进行结果分析。通过以下几个案例展示双聚类在文献计量分析中的应用意义及实现过程。

案例1：主题词-高被引文献矩阵的双聚类

选择ESI免疫学高被引论文为数据源，在Pubmed数据库下载文献XML数据，利用BICOMB生成高频主题词-来源文献的词篇矩阵，通过gCluto软件进行双聚类分析，制作山峰图，挖掘出该学科研究热点结构和研究前沿。并在此基础上绘制战略坐标，探索各个主题在整个领域所属地位和发展趋势。利用UCINET和NetDraw进行社会网络分析，从微观角度探索主题内部之间的联系程度。再将社会网络分析图谱和战略坐标融合，展示出该领域的宏观结构中微观主题的贡献度。

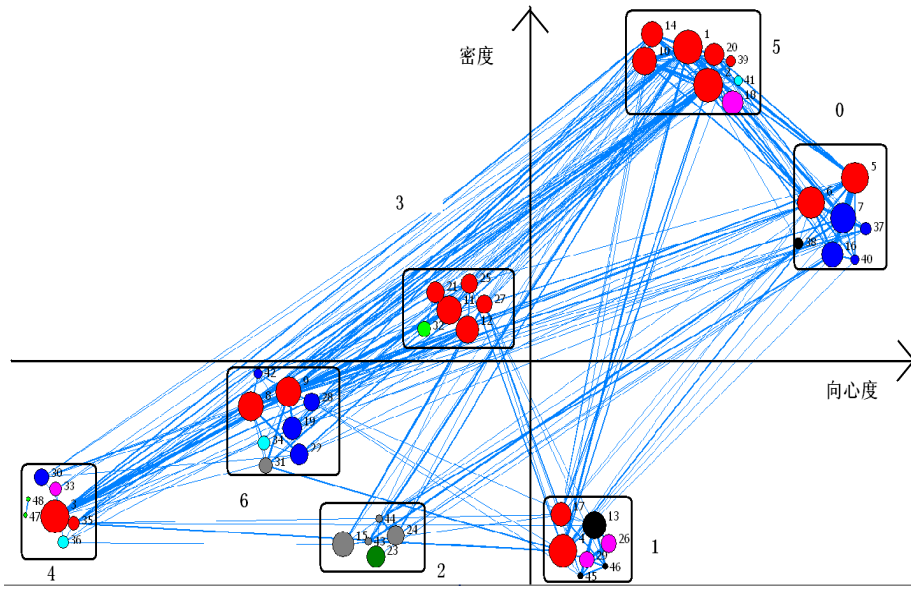


图1 共词网络分析与战略坐标的相互嵌入

案例2：引文-被引矩阵为例

以SCIE数据库中某时间段的医学信息学领域的论文作为数据源，文本挖掘软件BICOMB统计得到共被引矩阵，利用gCluto进行引文的耦合分析和被引论文的共被引分析，得到知识基础结构和相应的研究前沿。根据时间序列，计算各个类别的新颖度和关注度，绘制引文战略坐标图谱[15]。

文献计量学中常用的两种统计方法——引文耦合分析和同被引分析，常用于探索领域研究结构和前沿知识的方法。两者的区别在于：同被引方法关注被引用文献，展示本学科领域学科知识基础；而引文耦合方法则关注施引文献，代表本学科的研究前沿；针对这两种方法的研究都是单独进行的，并没有整合的分析。双聚类恰好从同被引分析和引文耦合的角度探索研究结构和前沿知识。双聚类用于引文分析与Persson[16]的观点相符，同被引文献集的引证论文集表示研究前沿，同被引文献集表示知识基础，不同之处在于，双向聚类引入同被引分析，同时实现同被引聚类分析和基于高被引论文的引文耦合分析。

案例3：作者-高频词矩阵为例

假设我们对某领域学者根据发表论文的主题进行分类，我们要计算每两位作者在所有主题上的相似性，而实际上两位作者只是在某个局部的主题上一致就可以说他们比较相似，不需要在所有主题上都完全一致。以某学科领域中部分高产作者作为聚类对象，以其高频主题词作为分类属性，运用双向聚类的方法，同时对高产作者和论文关键词进行聚类，并可视化表达，同时对每个类别的作者及对应关键词类别的详细分析，总结出这些作者在该研究领域的特点，可以探索该领域高产作者及其研究特色。

案例4：以疾病-物质名共现矩阵为例

利用PubMR进行双聚类展示：在PubMR环境下输入以下语句，可以展示异质共现矩阵疾病-物质名的共现关系。

```
library(data.table)
library(tidyr)
options(datatable.prettyprint.char=10L)
obj1=data.table(PMID=obj@PMID,MH=obj@MH)
obj1 = obj1 %>% unnest(MH)
obj1[,n:=.N,by=.(MH)]
obj1 <- obj1[n>=5,]
V <- crossprod(table(obj1[,1:2]))
V <- crossprod(table(obj3[,1:2]))
V[1:2,1:2]
idr <- which(rownames(V) %in% meshtree[class=="D",mesh])
idc <- which(rownames(V) %in% meshtree[class=="C",mesh])
```

```
V1 <- V[idr,idc]
x <- V1[1:40,1:40] x <- sweep(x,1L,rowMeans(x,na.rm=TRUE),check.margin=FALSE)
sx <- apply(x,1L,sd,na.rm=TRUE) x <- sweep(x,1L,sx,"/",check.margin=FALSE)
x[is.na(x)] <- 0
library(dendextend)
kr <- 6;kc <- 4
Rowv <- x %>% dist %>% hclust %>% as.dendrogram %>%
set("branches_k_color", k = kr) %>%
set("branches_lwd", 1.2)
Colv <- t(x) %>% dist %>%
hclust %>% as.dendrogram %>%
set("branches_k_color",k=kc,value=c("orange","blue","green","red")) %>%
set("branches_lwd",1.2)
cluster <- cutree(as.hclust(Rowv),k=kr)
clustab <- table(cluster)[unique(cluster[as.hclust(Rowv)$order])]
m <- cumsum(clustab)
m <- m[-length(m)]
heatmap(x, Rowv = owv,Colv=Colv,scale="none",add.expr=abline(h=m+0.5,lwd=3))
```

4、双向聚类在数据挖掘中的应用展望

对于数据挖掘中的异质矩阵：高被引-关键词、同被引-引文、机构-关键词、作者-关键词、疾病-药物、疾病-基因、药物-靶向位点之间这种二维关系，无法用传统的聚类方法来实现，因此受到限制，需要用双向聚类来挖掘科研研究中更多的信息与知识。

双聚类在数据挖掘方法中应用的展望：双聚类应用于引文分析中，是同被引分析和引文耦合分析的共现过程，可同时挖掘学科的基础结构及学术前沿；双聚类应用于共词分析中，可以探索双聚类在词篇分析中的应用，挖掘主题词的关联性以及在文献对比中挖掘新的知识；双聚类在关键词-作者分析中的应用可以解释学科研究者的主要分布与研究方向；双聚类在机构-关键词分析中的应用可以揭示不同机构、地区的科研活动及研究方向。应用于疾病-药物的分析中可以发现治疗疾病的有效药物；应用于疾病-基因的分析中可以解释疾病与基因的关系；药物-靶向位点的双聚类分析可以挖掘某药物作用的靶向位点。了解基因组信息与特定疾病之间的关联可用于发现新药。因此，双聚类可以从不同角度阐释科研活动的进展与趋势。

5、小结

双聚类方法在文献计量分析与评价中的应用，揭示科学研究者及科研活动的主要分布与研究方向，提出观测科学活动的新角度。双聚类渗透各个学科领域，深入挖掘高维数据中的有效信息，在数

据挖掘中具有非常重要的理论意义及实践价值。随着大数据生成技术在生物医学和卫生保健信息学中的应用日益增加，成本逐渐降低，在公共领域已经可以使用大量的生物学和临床数据集。 一方面，双聚类在医学及生物学领域为确定新的治疗靶标，药物指示和药物反应生物标志物提供了材料。 另一方面，它也给数据挖掘方法带来了更多挑战，包括检测疾病亚型，识别疾病的生物标志物和基因信号性质以及基因药物关联。

PU-045

基于本体的化学药物知识表示模型构建

秦璐[1,2,3]、徐倩[1,2,3]、罗爱静[1,3,2]
1. 中南大学湘雅二医院，湖南长沙
2. 医学信息研究湖南省普通高等学校重点实验室（中南大学）,湖南长沙
3. 中南大学生命科学学院，湖南长沙

目的 构建基于本体的化学药物知识表示模型。

方法 针对中国应用场景提出一种基于本体的化学药物知识表示方法，引入药物不良反应发生的频率，药物间的相互作用和配伍禁忌的详细关系，并对患者群体进行划分，较为细致地揭示了药物与药物、药物与疾病、药物与群体之间的关系。

结果 提出的化学药物本体知识表示模型包含 5 个语义类型、20 个对象属性和 23 个数据属性，并通过 protégé 软件实现本体的构建。

结论 基于本体的方法可以有效地组织化学药物信息，帮助化学药物知识共享。同时，也有利于对本体相关实体间的关系进行分析，促进化学药物的合理使用。

1 前言

在诊断和治疗环节中，药物信息是一个重要要素。而随着药物信息的规模急速增长、内容纷繁复杂，为促进专业人员的进一步研究，对药物信息的知识表示研究显得十分重要。

本体（ontology）作为一种语义网技术，常用三元组表示两实体间的关系[1]，可用于对特定领域中的概念及其相互关系进行形式化、规范化的描述[2-3]，其明确性、共享性、形式化、重用性等特点，有利于知识推理[4]、提高异构系统互操作性[5]等，因而被广泛用于医疗领域的知识表示。

目前，国外著名的医药本体知识表示研究，如 SNOMED CT[6]、RxNorm[7]、MedDRA[8]、DrugBank[9]、SIDER[10]等知识库，促进了生物信息系统的互操作性，但受语种、数据源等限制，对国内药物数据映射困难。国内开始主要以字典的方式对药物信息进行表示，如：《中国药典》[11]、《中国药典临床用药须知》[12]、《中国药学大辞典》[13]、《当代药品商品名与别名辞典》[14]等。随着对本体研究的不断深入，崔雷等[15]、王梅文[16]等人开始利用本体的思想对药学领域相关概念进行了分析和规范化表示，但药物知识表示的研究多集中于药物不良反应[3、5、17-18]、分子和基因层面药物学[19-20]等方向的研究，且对药物与药物、药物与疾病之间的关系表示研究较少或较为简单。

因此，本文拟基于本体的思想和方法构建符合中国应用场景的化学药物信息的知识表示模型，通过引入药物不良反应发生的频率，药物间的相互作用和配伍禁忌的详细关系，对患者群体进行划分等，较为细致地揭示了语义关系，用以支持药物数据的管理和整合。

2 药物本体构建过程

本体已经得到广泛应用，常见的构建方法有七步法、骨架法、METHONTOLOGY 法、企业建模法等方法[21]。综合分析已有的药物领域本体构建思路，本文主要借鉴了斯坦福大学医学院开发的七步法，并利用其开发的 protégé 工具构建化学药物本体。

构建药物本体的关键步骤就是确定类和属性。为确定本体相关概念，需要对药物相关数据进行选取、分析。本文数据主要来源于《中国药典》和《中国药典临床用药须知》。《中国药典》通常是以药物化学成分为基本单元描述其相关信息，包括药物的中文名字、英文名、拼音、别名、性状、规格、贮藏、化学成分及组成等；《中国药典临床用药须知》记录了合理用药以及药物相互作用、使用禁忌等方面的内容。除此还利用药品说明书、抗菌药物临床应用相关指南、不同人群的用药指南、相关文献、百科知识等进行补充，并通过咨询专家对药物本体进行完善。具体流程如图 1 所示。

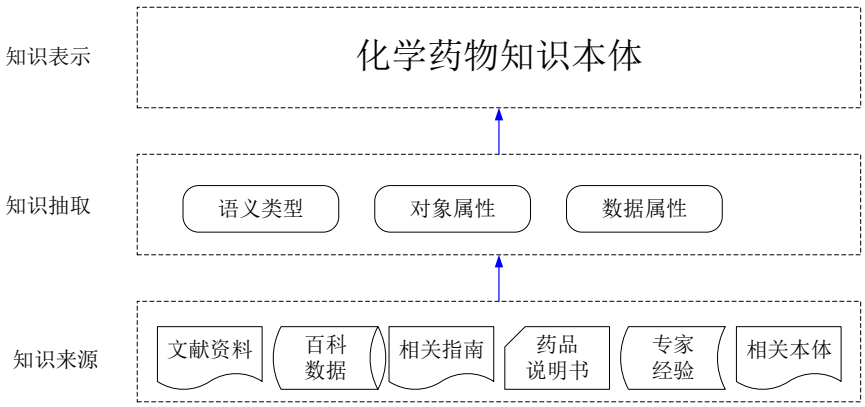


图 1 化学药物知识本体构建过程

Fig. 1 The construction process of chemical drugs knowledge ontology

2.1 确定本体语义类型

2015 年版《中国药典临床用药须知》化学药和生物制品卷（以下简称《临床用药须知》）通常包括化学药物的适应症/禁忌证、不良反应的相关信息、药物的用法用量、药物与群体的适用关系。从以往的研究中也可以看出，对药物知识表示的语义类型通常涉及药物、疾病、用法用量、特殊群体，因此在这个基础上抽取 6 个相关概念作为一级语义类型，分别为药物、化学成分、用法用量、疾病与临床表现、群体。

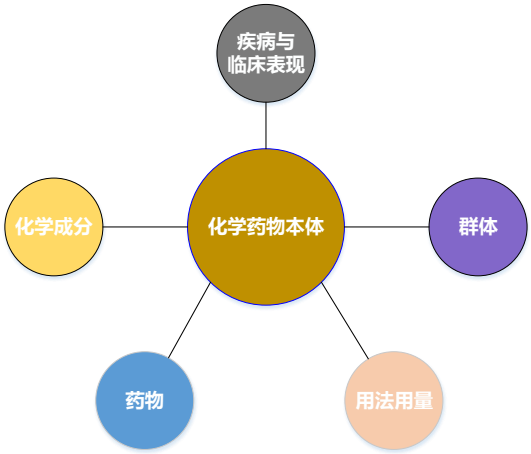


图 2 化学药物本体一级语义类型

Fig. 2 First order semantic type of chemical drugs ontology

通过对《临床用药须知》以及药品说明书的内容进行分析，【用法用量】通常包含剂型、剂量、给药途径、用药频次、度量范围、饮食等信息。【适应症】、【禁忌】、【注意事项】中有对包含对疾病、症状、过敏患者用药等信息的描述。【孕妇及哺乳期妇女用药】、【老年用药】、【儿童用药】等反映了特殊群体的用药，除此之外，还有对肝功能或肾功能不全的患者的用药调整。结合相关文献，分析我国药物常用指南，确定了 28 个二级语义类型将最后确定的一级语义类型和二级语义类型在 protégé 软件中构建，最后结果如图 2 所示：

图 3 protégé 构建化学药物本体语义类型

Fig. 3 The semantic types of chemical drugs ontology constructed by protégé

2.2 构建本体属性

在明确药物本体语义类型后需定义类的属性关系，即本体类的对象属性（object property）和数值属性（datatype property）。对象属性常用于描述类之间的关系，而数值属性是对概念的固有属性的一种描述[5]。

通过对《中国药典》（2015 年版）及相关资料分析，确定“药物”的数据属性包括通用名、别名、英文名、拼音、性状、制剂、贮藏、国家药管平台 YPID、ATC 编码、医保通用名库编码，“化学成分”的数据属性包括化学名、分子式、化学结构。通过相关文献以及专家咨询，确定“群体”的数据类型包括体质、生活习性、职业、年龄范围、性别、情绪、生理状态、指标、单位、取值范围。对象属性的确定参考 RxNorm、ULMS、TCMLs 等药物知识表达体系，确定了 9“用法用量”、“成分”、“不良反应”、“相互作用”、“配伍禁忌”、“禁忌证”、“适应证”、“禁忌人群”、“适应人群”9 个主要对象属性。其中，在《临床用药须知》中对【不良反应】的相关描述通常还记录了不良反应的发生频率；对于药物间的“相互作用”、“配伍禁忌”、 “禁忌证”、“适应证”等关系进行详细扩充，最终构建的 20 个对象属性和 23 个数据属性相关信息如表 1 所示：

表 2 化学药物本体的属性

Table 1 Properties of the chemical drugs ontology

定义域	属性	属性类型	上位类	值域
药物	用法用量	object property	-	用法用量
	成分	object property	-	化学成分
	不良反应_主	object property	不良反应	疾病与临床表现
	不良反应_次	object property	不良反应	疾病与临床表现
	不良反应_少	object property	不良反应	疾病与临床表现
	相互作用_不良反应	object property	相互作用	药物
	相互作用_药效增强	object property	相互作用	药物
	相互作用_药效降低	object property	相互作用	药物
	配伍禁忌_不良反应增强	object property	配伍禁忌	药物
	配伍禁忌_疗效降低	object property	配伍禁忌	药物
	配伍禁忌_禁忌	object property	配伍禁忌	药物
	配伍禁忌_药效增强	object property	配伍禁忌	药物
	配伍禁忌_需稀释	object property	配伍禁忌	药物
	禁忌证_相对禁忌	object property	禁忌证	疾病与临床表现
	禁忌证_绝对禁忌	object property	禁忌证	疾病与临床表现
	适应证_治疗	object property	适应证	疾病与临床表现
	适应证_降低风险	object property	适应证	疾病与临床表现
	适应证_预防	object property	适应证	疾病与临床表现
	禁忌人群	object property	-	群体

	通用名	datatype property	-	-
	别名	datatype property	-	-
	英文名	datatype property	-	-
	拼音	datatype property	-	-
	性状	datatype property	-	-
	制剂	datatype property	-	-
	贮藏	datatype property	-	-
	国家药管平台 YPID	datatype property	-	-
	ATC 编码	datatype property	-	-
	医保通用名库编码	datatype property	-	-
用法用量	适应人群	object property	-	群体
	化学名	datatype property	-	-
化学成分	分子式	datatype property	-	-
	化学结构	datatype property	-	-
	体质	datatype property	-	-
	生活习性	datatype property	-	-
	职业	datatype property	-	-
	年龄范围	datatype property	-	-
	性别	datatype property	-	-
群体	情绪	datatype property	-	-
	生理状态	datatype property	-	-
	指标	datatype property	指标约束	-
	单位	datatype property	指标约束	-
	取值范围	datatype property	指标约束	-

3 添加实例

在 protégé 中完成化学药物本体的构建后，可通过“individual”选项卡，添加实例。受篇幅限制，本研究拟通过添加抗菌药“阿莫西林胶囊(Amoxicillin Capsules)”的部分知识对该知识表示模型的重要部分进行演示。

“Data property assertions”是实例的数据属性。通过查询《中国药典》(2015 年版)获取“阿莫西林胶囊”的数据属性，具体设置结果如图 4 所示：

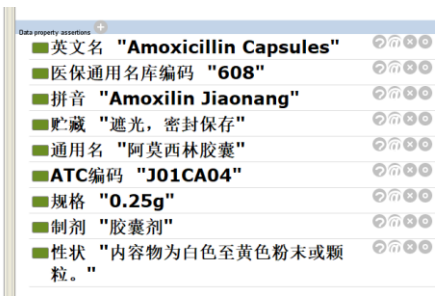


图 4 “Amoxicillin_Capsules”的“Data property assertions”设置

Fig. 4 "Data property enabling" settings in "Amoxicillin_Capsules"

“Types”指这个个体属于的类。通过对抗菌药“阿莫西林胶囊”相关信息的检索，获取其“FDA 孕妇危险等级”为“B 类”，即妊娠期患者有明确指征时需要慎用（未见到药物对胎儿的不良影响）；“医保分类”为“甲类”，即全保类药物；“哺乳期用药危险等级”为“L1”，即该药哺乳期患者最安全，不会显著增加婴儿的副作用，对婴儿的危害甚微；“处方类别”为“处方药”，即必须由执业医师或执业助理医师处方后,才能够调配购买的药物；“抗菌药物临床应用分级”为“非限制使用级”，说明该药物在长期临床应用中证明安全、有效，对病原菌耐药性影响较小，价格相对较低的抗菌药物；“药物代谢类别”为“肝、肾代谢”。在“Types”中进行设置，具体情况如图 5 所示：

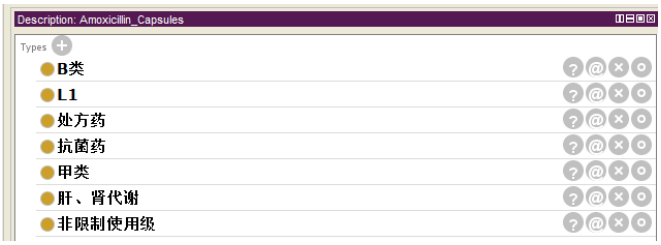


图 5“Amoxicillin_Capsules”的“Types”设置

Fig. 5 "Types" settings in "Amoxicillin_Capsules"

“Object property assertion”是对象属性声明，通过对象属性连接另一个个体，刻画两者之间的关系。图 6 展示了部分“阿莫西林胶囊”及其相关实体间的关系设置，包括“适应症”、“禁忌证”、“不良反应”、“相互作用”、“配伍禁忌”“适应人群”以及“成分”等对象属性声明：

图 6“Amoxicillin_Capsules”的“Object property assertion”设置

Fig. 6 "Object property enabling" settings in "Amoxicillin_Capsules"

由于不同“群体”的“用法用量”可能不一样，本文通过设置“患者 1”和“患者 2”两个个体对其进行揭示。图 7 为“患者 1”的属性设置。治疗年龄大于 3 个月且体重<40kg 的儿童，用法用量为 45mg/kg/日，每 12 小时 1 次，在患者无症状或已经获得细菌根除证据的时间之后，应继续治疗至少 48 至 72 小时。

图 7“患者 1”属性设置

Fig. 7 Attribute Settings for "Patient 1"

图 8 说明“患者 2”是“肾功能不全”患者，若肾小球过滤率为 10-30 毫升/分钟，其用法用量为 0.5 或 0.25g/次，每 12 小时 1 次能达到治疗目的，在患者无症状或已经获得细菌根除证据的时间之后，应继续治疗至少 48 至 72 小时。

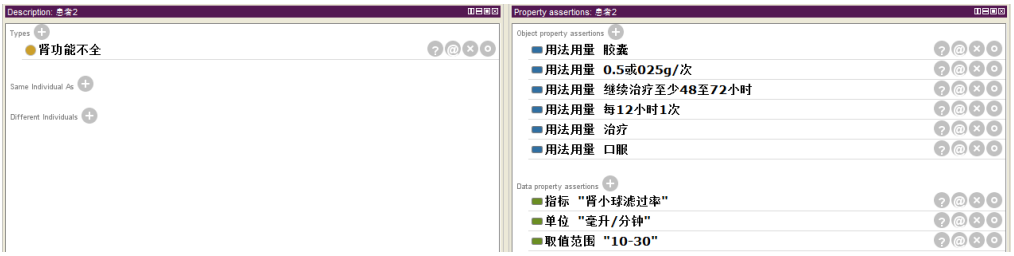


图 8“患者 2”属性设置

Fig. 8 Attribute Settings for "Patient 2"

通过 protégé 中的“OntoGraf”选项卡显示“阿莫西林胶囊”对不同群体“用法用量”的知识表示如图 9 所示：

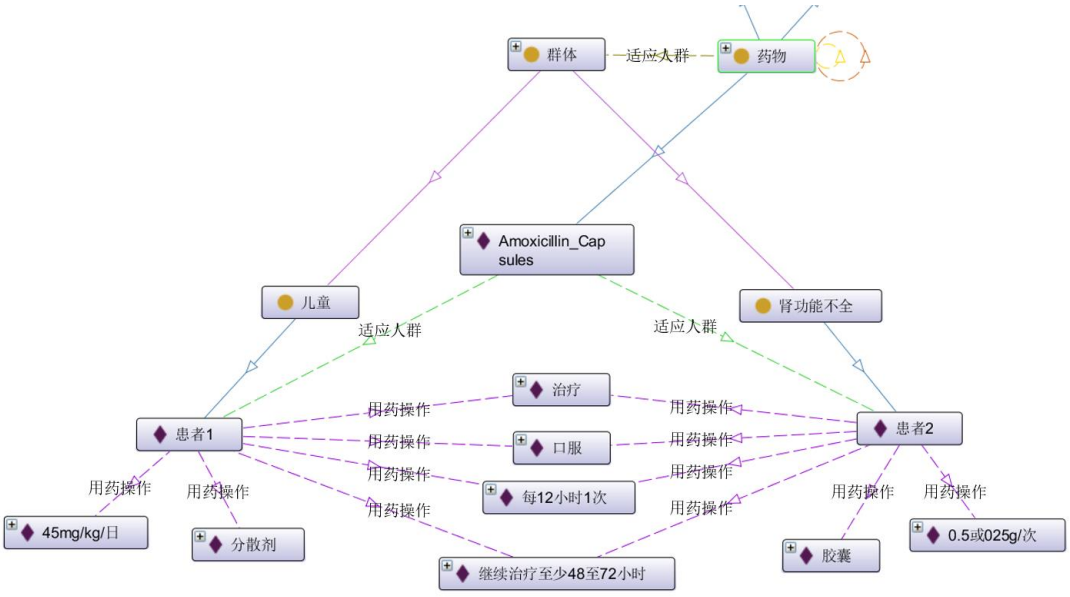


图 9 不同群体“用法用量”的知识表示

Fig. 9 Knowledge representation of "usage and dosage" in different groups

4 结论

本研究针对我国药物数据信息特点，采用本体的思想和方法构建了化学药物知识表示模型，通过 protégé 软件实现本体，并以“阿莫西林胶囊”为例，填充了部分相关信息，展示了构建的知识表示模型。本研究构建的化学药物知识表示模型引入了药物不良反应发生的频率，药物间的相互作用和配伍禁忌的详细关系，对患者群体进行划分等，较为细致地揭示了药物相关概念的语义关系。基于本体的方法可以用于共享和推理，帮助医生辅助用药和安全用药，为医生、患者等提供药物知识；还可以用于分析药物及药物相关实体间的信息，如分析药物的常用人群、常用用法用量的构成比、发生最多的不良反应以及不同人群、用法用量等因素与不良反应发生频率之间的关系。

在研究过程中还存在着一些不足，对一些权威的、应用广泛的药物及相关概念本体的映射不够，兼容性和协同性有待增强；语义类型及属性的提取受主观影响可能存在一些不准确的问题，模型的一致性、完整性和可操作性有待验证；对知识模型的展示不够完全，对“阿莫西林胶囊”的知识表示有部分预设的语义类型或属性不适用或不存在。今后的研究将针对以上的问题进行更多改进，要加强对药物及相关概念本体的映射，方便数据共享和系统的互操作性；通过专家咨询等方法，对模型进行多维度的检验，不断完善和改进，增加知识本体的覆盖面。

PU-046

从宏观结构及微观分布揭示全球免疫学科战略情报

杨颖、陈斯斯、徐爽、许丹、韩爽
中国医科大学

目的 探索将社会网络分析与战略坐标嵌入融合，以从学科宏观结构及微观分布的角度揭示学科发展的战略情报。

方法 选择ESI免疫学高被引论文为数据源，对高频主题词与来源文献进行双聚类分析，绘制战略坐标图谱和社会网络图谱，并将两者结果相互融合嵌入。

结果 结合免疫学知识对聚类结果进行分析解释，总结该领域的学科热点结构和研究前沿。

结论 社会网络分析嵌入战略坐标，从宏观结构和微观分布揭示了全球免疫学领域的热点结构与内部主题分布状况，为该领域发展方向提供参考。

1、背景

信息计量学中一种重要的分析方法——共词分析，早在20世纪70年代已经提出，经过近50年的发展，已经趋于成熟。共词分析方法的基本思路是通过找到表达某研究领域主题的词对间的共现关系，通过统计学及可视化等方法展现学科的知识结构、演变历程及趋势走向。常用的统计方法有聚类分析、因子分析及多维尺度分析等，可视化方法主要有战略坐标、社会网络分析及Citespace、MedlineR等方法。共词分析目前已经广泛应用于图情学、管理学、医学、教育、计算机、体育、农业等各个学科领域，且很多可视化研究也不局限于一种可视化图谱，而是结合多个可视化结果从多角度探索学科分布及发展。

最常用到的共词可视化——战略坐标，用于揭示某学科领域内部联系及领域间的相互影响情况[1]。其原理[2]如下：在共词矩阵和聚类的基础上描述某学科领域内部关联关系和领域间相互影响的关系。战略坐标图中，横轴代表向心度，表示领域间相互影响的强度，用来衡量这一研究领域与其它领域相互影响的程度；纵轴代表密度，表示某领域内部联系强度，用来衡量使字词聚合成一类的这种联系的强度，也就是该类的内部强度。它表示该类维持自己和发展自己的能力。战略坐标分析已在图情学、管理、教育、医学、信息等多学科中得到广泛应用，有研究通过战略坐标分析文献计量学的热点[3]、知识管理领域的学科主题发展状况[4]，以及对战略坐标改进的方法学研究[5]。

社会网络分析引入共词分析，展现词对间的共现网络，即网络在顶点的相互作用下的分布状况，以及网络在时间序列下的动态演变。k核分析是社会网络分析常用手段，即满足一个条件的子图，即子图中的点都至少与该子图中的k个其它点邻接。随着k的增加，k核子图成员逐渐减少，成员间的关系更加紧密。社会网络分析通过Citespace、UciNet、Pajek、NetDraw等软件进行可视化，常用于探索科学研究中的合作关系网络[6]、引用被引关系网络[7]以及共词关系网络[8]。

不同的共词分析图谱，是从不同的角度揭示学科的发展状况：战略坐标是从宏观结构阐释学科的核心热点结构，而社会网络图谱是从微观角度揭示核心主题。目前关于共词分析图谱的研究大多构建独立的图谱揭示不同的问题，而忽略了图谱之间的相互关联问题，比如，战略坐标宏观结构的内部微观状况，即每个核心主题在热点结构中的贡献度如何，每个核心主题在热点结构之间的互通关系如何，均未能体现出来。本研究拟将社会网络分析嵌入战略坐标分析，揭示学科领域的宏观结构及微观分布。

本文拟将免疫学科的 ESI 高被引论文作为数据源进行共词分析可视化分析，从方法上，一方面，打破传统的以高频词对为研究对象的共词聚类分析，拟对高频词-来源文献的词篇矩阵进行双向聚类；另一方面，突破社会网络分析和战略坐标的单一图谱分析，拟将社会网络分析嵌入战略坐标分析，捕捉内部潜在的知识内容，分析免疫学科的宏观结构及其内部主要贡献的核心主题构成。

2 材料与方法

2.1 数据收集

选择 WOS 核心合集数据库，检索式为 WC=immunology，收集到免疫学研究论文，再通过“ESI 精炼”，发表年份限定为 2015-2018，文献类型限定为 article 和 review，即该领域 4 年来的 ESI 高被引论文 824 篇，并在 PubMed 下载 xml 格式文件作数据源。

2.2 数据处理

利用书目共现分析系统（Bibliographic Item Co-Occurrence ，BICOMB），分析 PubMed 数据库下载的免疫学文献，统计其主要主题词出现频次，抽取频次大于等于 8 的 48 个高频主题词，并生成这些高频词与来源文献的词篇矩阵。利用 gCLUTO 软件的重复二分法完成对词篇矩阵的双向聚类，生成聚类图谱和山丘图谱。利用聚类结果和共词矩阵在 Excel 中绘制共词战略坐标。利用 Ucinet 和 Netdraw 绘制社会网络分析图谱，最终将社会网络分析结果嵌入战略坐标图中，实现共词可视化图谱的整合。

3 结果与讨论

3.1 双聚类结果

3.1.1 可视化山丘

根据每个类内数据分布，得到山峰图（图 1）。其高度与类内相似度成正比，类内相似度大则山丘陡峭。其体积与类内对象（高频词）数量成正比。红色代表低类内相似度标准差，蓝色代表高类内相似度标准差。图 1 中 cluster 0 类内高频词相似度最高，表明类内高频词具有高度一致性。图中七个山峰各自独立，高度与体积适中，聚类效果较理想。

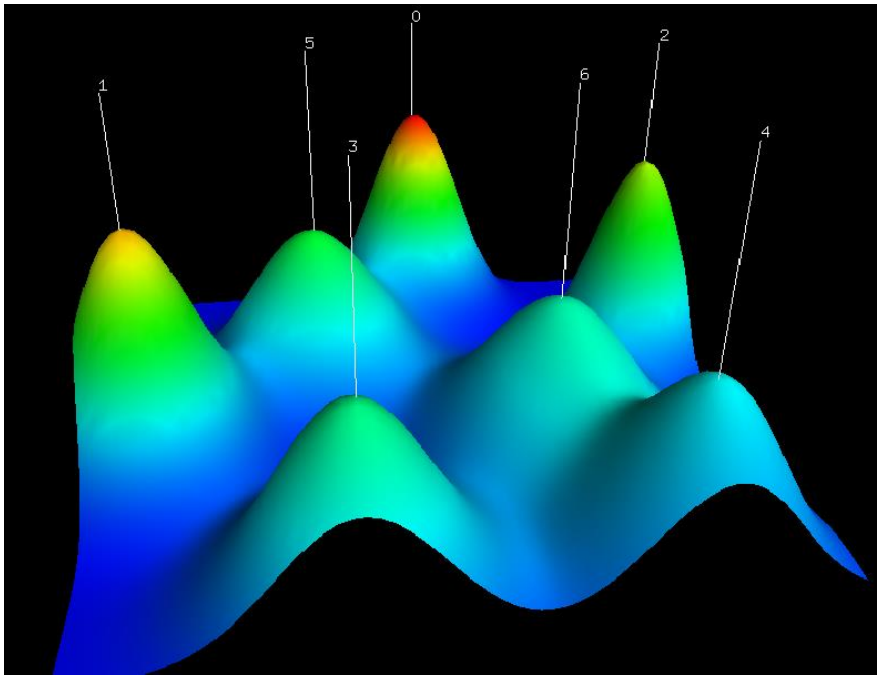


图 1 可视化山丘图

3.1.2 双聚类可视化矩阵

可视化矩阵（图 2）分别从行和列两个维度展示了高频词和来源文献的聚类结果，聚类图形的行聚类（图的左侧）表示高频词的分类，图的右侧列出相对应的高频词，横线将每个类隔开，一共分为七大类；聚类树图的列聚类（图的上方）表示来源文献的聚类，图的下方对应列出所代表的来源文献。可视化矩阵中颜色代表矩阵原始数据的值，白色表示近零值，红色表示较大的值，图 2 中颜色的深浅表示高频词出现的频次。矩阵的行重新排序使得同一组的行聚在一起。

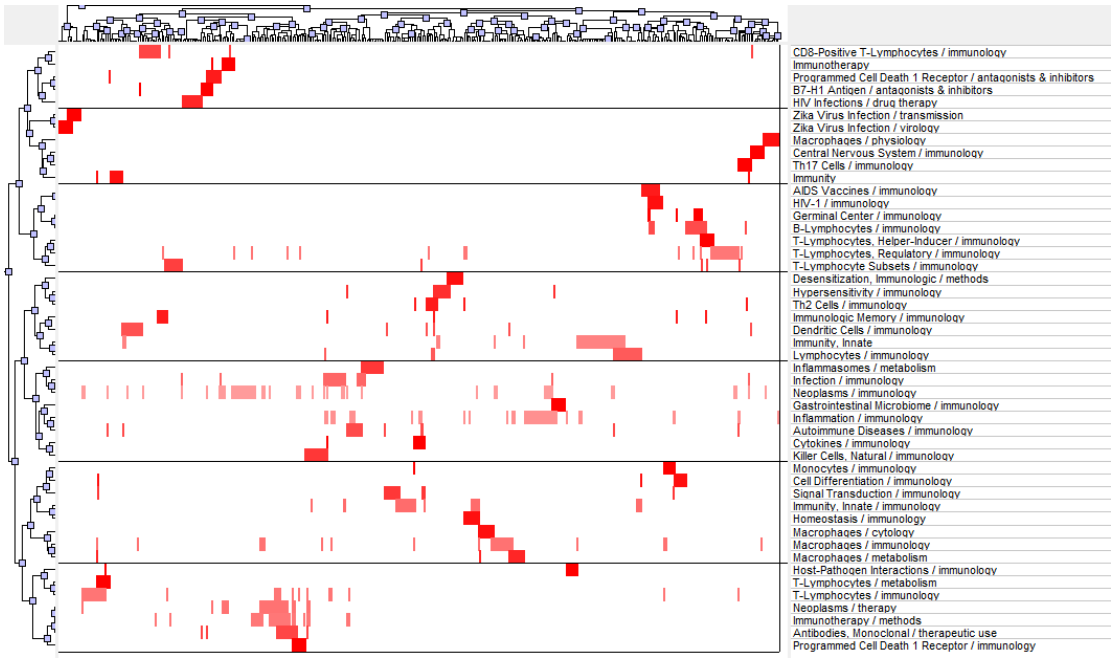


图 2 ESI 高被引论文可视化矩阵

3.1.3 高频词构成的热点结构分析

通过对 ESI 免疫学领域高被引论文的二分法聚类，得到聚类树图，图 1 显示该学科领域的热点结构分为七大类，其热点结构及其主题词构成（见表 1）如下：

一类（Cluter 0）：肿瘤的免疫治疗。肿瘤的免疫治疗是免疫学的一个重要分支，肿瘤免疫疗法已成为肿瘤治疗的重要手段。PD-1 是免疫检查点研究的热点，2018 年诺贝尔生理学或医学奖更是将其推向研究浪头，确立了肿瘤免疫治疗时代的到来。该类主要涵盖了探索新的肿瘤免疫检查点、T 细胞耗竭、新的肿瘤免疫抑制细胞亚群以及肿瘤免疫治疗新策略等方面的内容。

二类（Cluter 1）：内容涉及艾滋病疫苗的免疫学。艾滋病疫苗开发 30 余年仍未成功，从体液免疫到细胞免疫再到两者结合，历尽艰辛。主要集中在光谱中和抗体对 HIV 感染后疾病进展的意义，抗体如何经胚系 B 细胞筛选、分化产生，如何刺激有效持久的光谱中和抗体，如何突破 HIV 感染的免疫耐受和免疫逃逸产生广谱保护性 T 细胞免疫，这些机制尚待解决。

三类（Cluter 2）：内容涉及 B7-H1/程序性细胞死亡 1 受体抑制途径与效应 T 细胞功能关系，治疗 HIV 感染的机制。B7-H1 与其受体 PD-1 结合，降低肿瘤细胞免疫原性，诱导 CTL 凋亡，促使免疫细胞发生逃逸。PD-1 是肿瘤免疫抑制的明星分子，也有其与 HIV 感染的免疫机制研究。

四类（Cluter 3）：脱敏方法及免疫学机理。脱敏疗法能通过改变过度活跃的固有免疫应答，产生免疫特异性的记忆性 Th2 细胞，使外周血中高表达的细胞因子水平恢复至正常，降低高敏反应和炎性反应,达到缓解和治疗的目的。

五类（Cluter 4）：寨卡病毒感染致中枢神经系统（CNS）损伤的免疫机制。寨卡(ZIKA)病毒感染相关神经系统损伤主要有吉兰-巴雷综合征和小头畸形。通过逆转录-聚合酶链反应可诊断寨卡病毒感染，影像学检查可提高合并 CNS 并发症诊断的准确率。然而,目前尚无针对寨卡病毒感染的特效药物及疫苗。

六类（Cluter 5）：有自身免疫性疾病、肿瘤的免疫学机理以及炎症和感染的免疫学。炎性小体和细胞因子参与炎症和自身免疫相关的许多先天免疫过程。

七类（Cluter 6）：免疫调控中的细胞分化与信号转导过程。免疫细胞的发育涉及转录因子，细胞信号传导蛋白和生长因子的复杂相互作用。研究发现 microRNA 是先天免疫细胞发育、功能以及免疫稳态维持的关键调节因子[9]。

类别	主题	词频秩序	英文主题词
Cluster 0	对肿瘤的免疫治疗	5	5Immunotherapy / methods
		6	6T-Lymphocytes / immunology
		7	Neoplasms / therapy
		16	Antibodies, Monoclonal / therapeutic use

		37	T-Lymphocytes / metabolism
		38	Host-Pathogen Interactions / immunology
		40	Programmed Cell Death 1 Receptor / immunology
Cluster 1	艾滋病疫苗的免疫学	4	4T-Lymphocytes, Regulatory / immunology
		13	B-Lymphocytes / immunology
		17	T-Lymphocyte Subsets / immunology
		26	AIDS Vaccines / immunology
		29	HIV-1 / immunology
		45	Germinal Center / immunology
		46	T-Lymphocytes, Helper-Inducer / immunology
Cluster 2	B7-H1/PD1 受体的免疫学	15	CD8-Positive T-Lymphocytes / immunology
		23	HIV Infections / drug therapy
		24	Programmed Cell Death 1 Receptor / antagonists & inhibitors
		43	Immunotherapy
		44	B7-H1 Antigen / antagonists & inhibitors
Cluster 3	脱敏与脱敏免疫学机理	11	Lymphocytes / immunology
		12	Dendritic Cells / immunology
		21	Hypersensitivity / immunology
		25	Th2 Cells / immunology
		27	Immunologic Memory / immunology
		32	Desensitization, Immunologic / methods
Cluster 4	寨卡病毒感染致中枢神经系统损伤的免疫机制	3	3Immunity, Innate
		30	Immunity
		33	Macrophages / physiology
		35	Th17 Cells / immunology

		36	Central Nervous System / immunology
		47	Zika Virus Infection / transmission
		48	Zika Virus Infection / virology
Cluster 5	自身免疫性疾病、肿瘤的免疫学机理炎症和感染的免疫学	1	1Neoplasms / immunology
		2	2Inflammation / immunology
		10	Infection / immunology
		14	Autoimmune Diseases / immunology
		18	Inflammasomes / metabolism
		20	Killer Cells, Natural / immunology
		39	Cytokines / immunology
		41	Gastrointestinal Microbiome / immunology
Cluster 6	免疫调控中的细胞分化与信号转导过程	8	Macrophages / immunology
		9	Immunity, Innate / immunology
		19	Signal Transduction / immunology
		22	Macrophages / metabolism
		28	Cell Differentiation / immunology
		31	Homeostasis / immunology
		34	Macrophages / cytology
		42	Monocytes / immunology

表 1　免疫学热点结构及其主题词构成

3.1.4 代表性论文构成的前沿文献分析

在 gCLUTO 软件中导入共词矩阵，进行双向聚类分析，得到该聚类的相关数据，如表 2 所示，得到七大类代表性文章（Descriptive，代表这一类前沿特征的文献）和区分性文章（Descriminating，代表区分于其他类的文献）编号，通过文献编号在 BICOMB 查询对应的文献。Size 表示类的构成数量，Isim 为类内对象间的平均相似度，Esim 为类间的平均相似度。Isim 越大，表明聚类结果将相似的类聚到一起，Esim 越小，表明不同类间相似度较小，该聚类效果较好。根据表 2，找出有代表性的前沿文献，研究其内容所展示的前沿领域如下：

（1）、癌症免疫的共抑制途径，其中，文献[10]表明肿瘤微环境是免疫抑制的，文献表明阻断 PD-1 和 CTLA-4 检查点是有效且持久的癌症免疫疗法，且联合抑制进一步改善应答率。文献[11]介绍免疫检查点抑制剂：检查癌症的新策略。

（2）、艾滋病疫苗免疫学，文献[12]针对将具有种系 VRC01 B 细胞受体的 B 细胞转移到同类受体中，阐明前体频率、抗原亲和力和亲合力对免疫后 B 细胞应答的作用。文献[13]介绍滤泡辅助 T 细胞（Tfh 细胞）引发 B 细胞启动卵泡外和生发中心抗体反应，对亲和力成熟和体液记忆的维持至关重要。Tfh 细胞在抗菌防御、癌症和 HIV 储库及预防自身免疫方面发挥重要作用。

（3）、抗 PD-1 和抗 PD-L1 抗体：文献[14]介绍了两者抗肿瘤的治疗用途。将抗-CTLA-4 抗体和抗-PD-1 抗体组合用于晚期黑素瘤患者可抑制毒性。未来的研究将倾向于此方法的生物标志物以及如何单独或与其他免疫疗法、放射疗法、化学疗法和小分子抑制剂联合使用。

（4）、致敏与脱敏免疫学：文献[15]证明第 2 组先天淋巴（ILC2）细胞在记忆 TH2 细胞反应中具有关键作用。 文献[16]综述了 Th2 细胞因子（IL-4 和 IL-13）和先天免疫细胞在先天性和适应性炎症反应之间的交叉作用中的作用。

（5）、寨卡病毒的流行病学以及其传播途径。

（6）、感染、炎症和癌症中的自然杀伤细胞记忆：从研究 NK 细胞记忆中获得的基本概念提供关于先天免疫的新见解，可能为传染病和癌症治疗提供新策略[17]。文献[18]概述炎症小体研究的最新发展，及其如何控制健康和疾病。

（7）、稳态免疫调控：文献[19]报道了局部环境和表观遗传学在塑造巨噬细胞身份及其对组织稳态的作用。文献[20]报道了白血球能量代谢的广泛缺陷是脓毒症免疫麻痹的基础。

Descriptive & Descriminating Features

Cluster 0	**Size:** 7		**ISim:** 0.298			**ESim:** 0.015		
Descriptive:	26927206	9.7%	29139554	6.8%	27108521	5.5%	27192570	4.8%
Descriminating:	26927206	5.6%	29139554	3.9%	27108521	3.2%	27216414	2.7%
Cluster 1	**Size:** 7		**ISim:** 0.287			**ESim:** 0.008		
Descriptive:	29287996	11.2%	26907215	10.7%	27088919	9.3%	28636956	7.2%
Descriminating:	29287996	6.0%	26907215	5.7%	27088919	5.0%	28636956	3.8%
Cluster 2	**Size:** 5		**ISim:** 0.278			**ESim:** 0.010		
Descriptive:	25323844	15.1%	25415283	6.5%	25621841	6.5%	26928461	6.5%
Descriminating:	25323844	7.4%	25415283	3.5%	26928461	3.5%	26944201	2.9%
Cluster 3	**Size:** 7		**ISim:** 0.217			**ESim:** 0.013		
Descriptive:	26523868	18.3%	28658559	5.7%	28771830	3.6%	26853128	3.6%
Descriminating:	26523868	10.4%	28658559	3.3%	28771830	2.1%	26853128	2.1%
Cluster 4	**Size:** 6		**ISim:** 0.208			**ESim:** 0.007		
Descriptive:	27893700	6.7%	27070380	6.7%	25625872	6.7%	28287375	6.7%
Descriminating:	25625872	3.6%	28287375	3.6%	27070380	3.6%	27135602	3.4%

Cluster 5	Size: 8		ISim: 0.206			ESim: 0.019		
Descriptive:	26806484	9.8%	25879280	5.9%	28722709	5.6%	25458968	4.4%
Descriminating:	26806484	4.8%	25879280	3.6%	28722709	3.5%	25458968	2.0%
Cluster 6	Size: 8		ISim: 0.190			ESim: 0.013		
Descriptive:	26681458	9.4%	26950237	4.7%	28606987	3.9%	26642357	3.3%
Descriminating:	26681458	5.4%	26642357	1.9%	28606987	1.9%	26950237	1.8%

表 2 ESI 高被引论文聚类结果的相关数据

3.2 战略坐标分析

本研究得到共词战略坐标，如图 3 所示：第一象限中的词团（5，炎症和感染免疫机理）密度高，表明词团内部联系最紧密，词团（0，肿瘤免疫治疗）向心度最高，表明该词团与其余各词团有广泛的联系，这两个词团处于所有研究主题的核心。第二象限中的词团（cluster3，致敏与脱敏）处于近原点位置，表明其虽位于边缘位置，但已经受到关注，有望发展进入第一象限。第三象限的词团密度和向心度都低，处于整个研究领域的边缘，研究尚不成熟；第四象限的词团（1 肿瘤免疫治疗）中心度高，但密度低，说明该词团构成的主题领域也是核心，但不成熟。

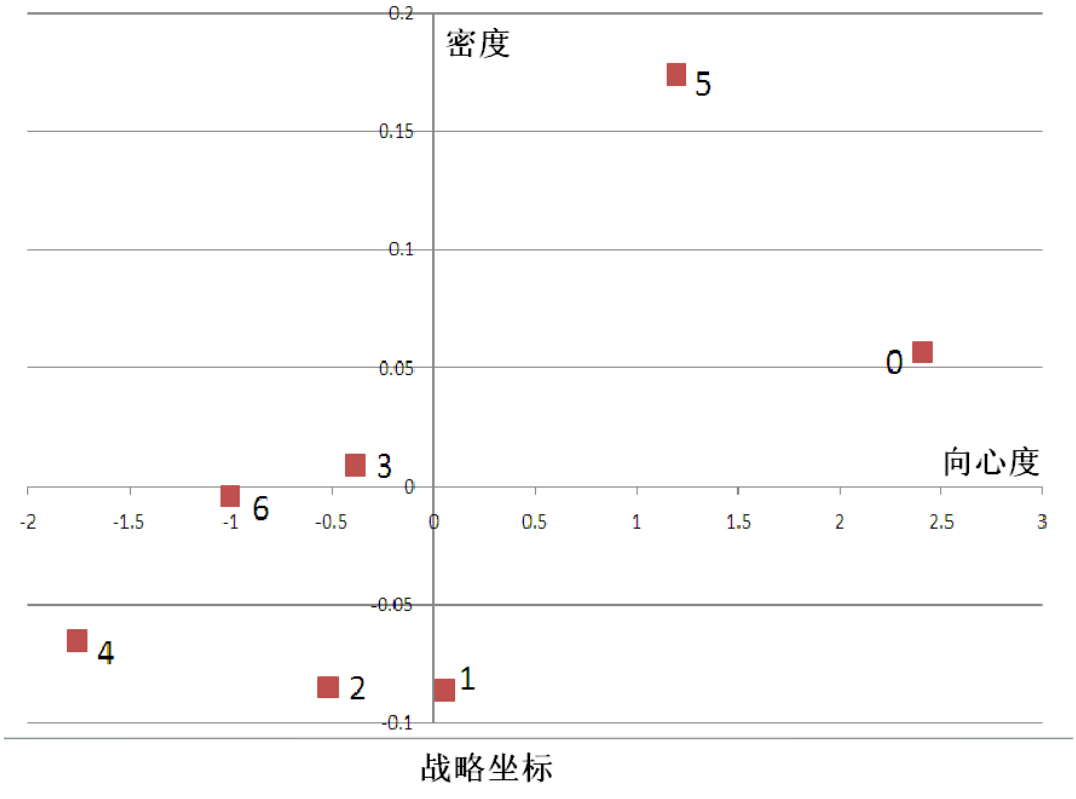


图 3 共词网络的战略坐标

3.3 社会网络分析

3.3.1 原始共词矩阵图谱

原始共词矩阵图谱展现了高频主题词的共现关系：每个节点代表高频主题词，大小与其在共词网络中出现的频次成正比；主题词之间连线表示有共现关系，粗细表示联系的紧密程度，直观地体现出词对之间的关系。如图 4 可见：节点 1（Neoplasms / immunology）最大，说明肿瘤免疫出现的次数最多，节点 5（Immunotherapy / methods）、节点 7（Neoplasms / therapy）、节点 16（Antibodies, Monoclonal / therapeutic use）之间连线粗，说明其共同出现的频率大，联系密切。

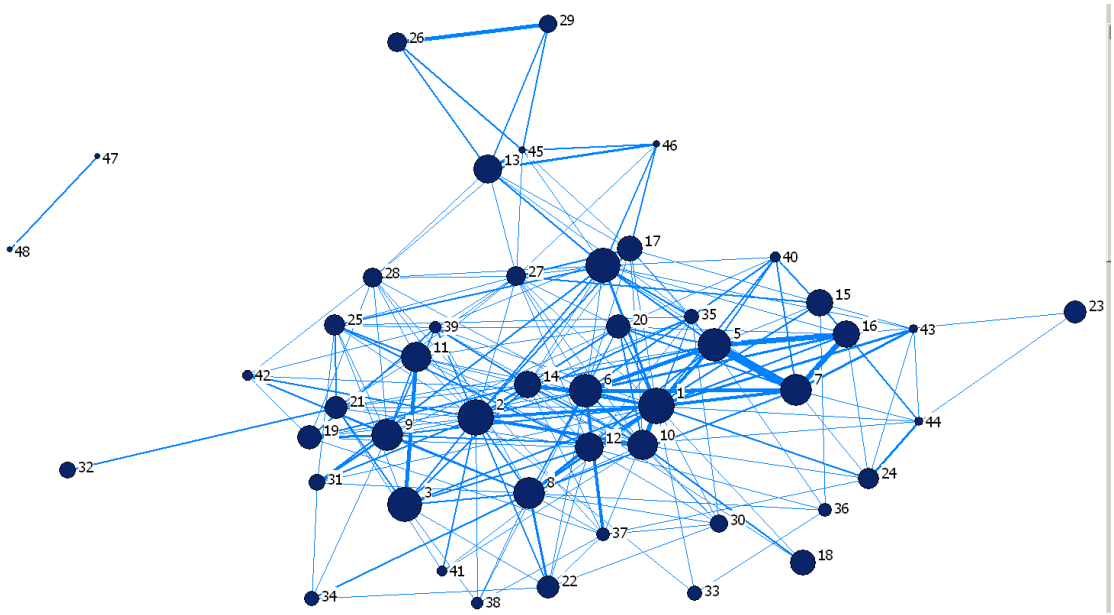


图 4　原始共现网络

3.3.2K-core 分析

图 5 中展现免疫学高被引论文共词网络的核心边缘主题。得到 8 个 K 核等级， K 的最高级数为 8，即在 K 等于 8 的核中每个词都最少和同一个核中其他词共同出现的次数大于等于 8，是整个共词网络中连接最强，关系最为紧密的一个部分，包括位于中间的红色节点，代表了最核心的主题： 1、2、3、4、5、6、8、9、10、11、12、14、17、20、21、25、27、35、39 号词。K=7 的核代表次核心主题的蓝色节点，K=6、5 的核分别为灰色、黑色，代表中间主题节点，连接核心主题词团和边缘主题词团；K=4、3 分别代表湖蓝、粉色，为次边缘主题，K=2 和 1 的核分别为草绿色（23）和翠绿色（32，47，48）节点，代表边缘主题。

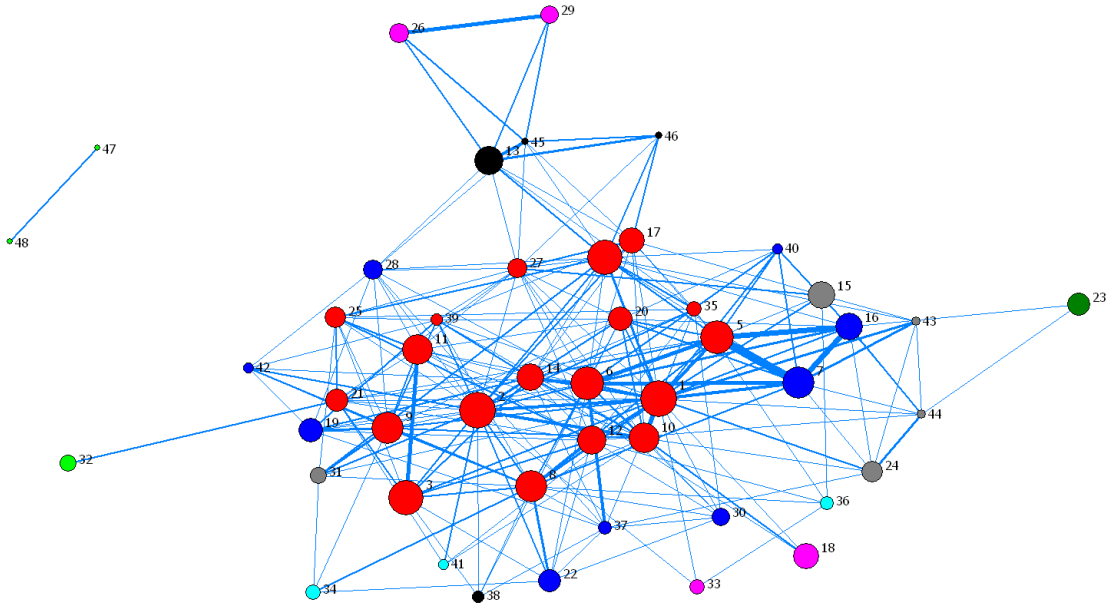


图 5 原始共词网络的核心边缘主题分布

3.4 两种可视化方法的相互嵌入融合

3.4.1 原始共词网络与战略坐标相互嵌入

依照战略坐标分布，在社会网络图中引入坐标轴，将双聚类结果中相同类别的词放置一起，得到高频主题词原始共词矩阵按照战略坐标排列的网络图（图 6），可以清晰地显示出战略坐标中类间与类内的各个主题之间的关系。

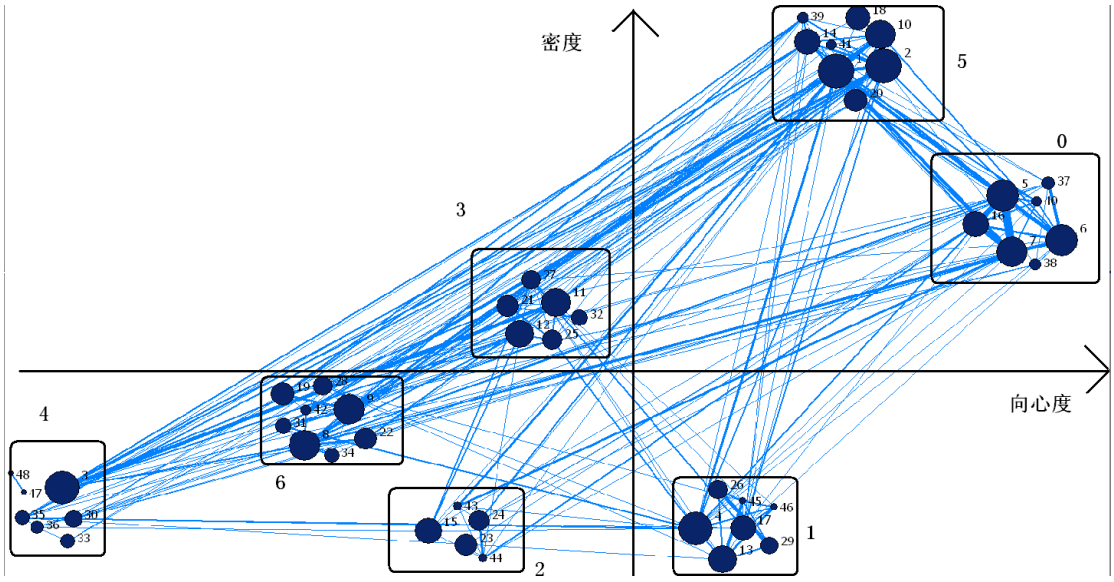


图 6 原始共词网络与战略坐标相互嵌入

3.4.2 战略坐标结果与 k 核分析结果相结合

将战略坐标结果（成熟度-核心度）与社会网络分析中的 k 核分析结果（核心-边缘）相结合（图 7）：用不同的颜色标识核心边缘程度，在战略坐标中的位置标识核心成熟程度，结合后的 k

核分析可以看到核心边缘主题词在每个聚类的类别中，而每个类别的成熟程度在战略坐标中已经体现出来。

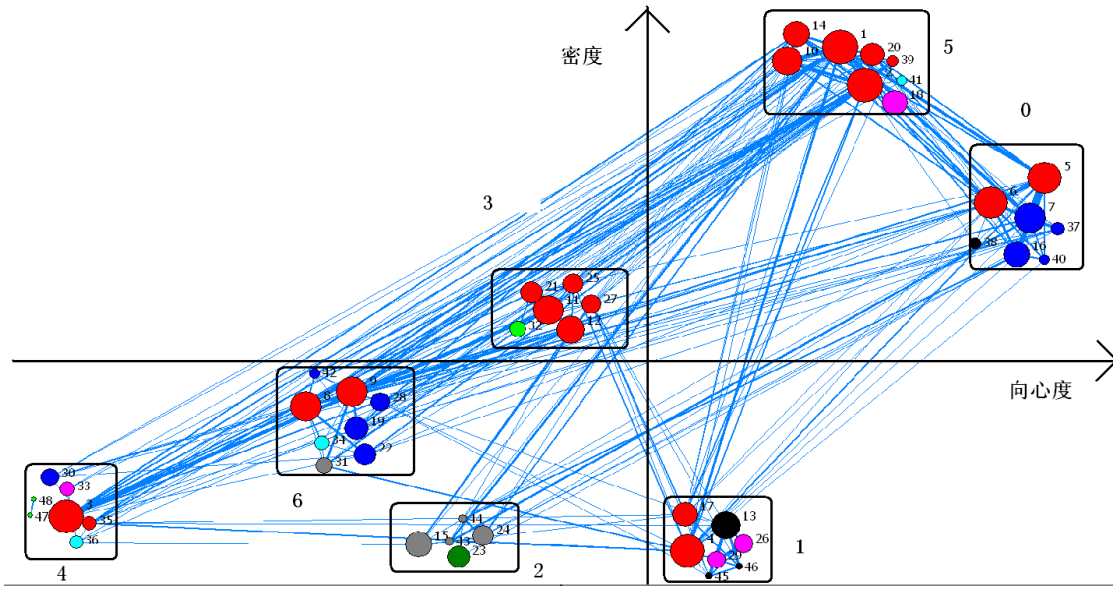


图 7 K-core 共词网络与战略坐标相互嵌入

4 讨论

4.1 数据源的选择

在探索学科结构及趋势演变计量分析中，数据选择对分析结果至关重要。通常情况下，原始数据选择有几种情况：1、全面选择数据[21]，比如 CNKI、万方、维普数据库，外文选择期刊或出版社全文数据库，如 Elsevier、Springer 等，文摘数据库 CBM、PubMed 等；2、核心数据[22]，如 CSSCI、CSCD、CSTPC、WOS 核心合集等；3、热点数据[23]，如 ESI 高被引论文、ESI 热点论文、Altmetrics 高分论文、F1000 等。目前，对热点数据隐藏知识的挖掘越来越受到科研人员广泛关注。ESI 高被引论文是被引用量排在前 1%的研究论文，反映近年来某学科高质量、高水平的重要科研成果。目前，已有文章针对高被引论文的统计分析，文献[24-27]分别对高被引论文进行学术特征、国际合作情况、文献计量及知识图谱的研究。未见对 ESI 高被引论文进行高频词对和来源文献同时进行聚类分析热点结构和前沿内容的文献报道。

4.2 词篇双聚类的特点

数据矩阵中，每一列对行的聚类结果的影响不同。挑选部分列进行聚类，其结果与选择所有列聚类的结果是有区别的。双聚类分析方法将行和列结合，行聚类的同时充分考虑列属性的特点，在行和列两个维度上同时聚类来发现稳定的类簇。以往的共词分析统计高频词之间的共现关系，只是对高频词对共现矩阵的聚类，而忽略对高频词来源文献特性的思考，因此结果有失偏差。本文综合考虑高频词与来源文献的共同特性，将双聚类的方法引入共词分析，实现词与篇的共同聚类。

行的维度是对高频主题词的聚类，结果代表了研究热点结构，列的维度是对来源文献的聚类，结果表示该领域的研究前沿。

4.3 网络图谱与战略坐标图谱相结合的优势

4.3.1 原始矩阵的共词网络图与聚类及战略坐标相结合

此种结合清晰展现出免疫学领域学科结构类间与类内的各个主题之间的关系。（1）第一象限中各个类内、类间的连线明显比第三象限中类内、类间的连线多且粗。（2）战略坐标中的向心度越高的类别，在社会网络图谱中该类别与其它类别的连线越多、越粗，战略坐标中的密度越高的类别，类内各个主题间的连线越多、越粗。例如图 6 中的第 5 类，主要由 1 号主题词贡献了向心度，同时，1 号主题词与本类中其它主题词的连线最粗，贡献了密度；第 0 类，主要由 5 号主题词（Internet）贡献了向心度，其与 7 号、16 号主题词最密切，连线最粗，贡献了此类的密度。整合的结果更有利于解读，能够直观地看出类内、类间每个词的关系以及它们对向心度、密度所起的作用。

4.3.2 战略坐标结果（成熟度-核心度）与社会网络分析中的 k 核分析结果（核心-边缘）相结合

（1）战略坐标可以看出整个领域研究的结构：即哪些类别是核心主题，哪些类别是成熟主题，一目了然。如从图 7 中可见，第一象限的 cluster0、5 为研究成熟的核心主题，第三象限的 4、6、2 类研究不够深入，属于非核心主题。

（2）k 核分析的可视化结果可以看出每个类别中的核心研究热点是什么，如图 7，聚类第 5 类的 1、2、14、39 主题词，聚类第 0 类的 5、6 号主题词为核心研究热点，显示出每个词的地位及与其它词的关系，与战略坐标结果得到的核心热点结构有所不同，战略坐标突出了整体。这种用颜色深度标识不同的 K 值，节点大小表示该词的频次，连线表示词对之间的共现关系，将几种可视化方法相结合可以清晰地展现出该学科研究热点的层次及热点词的关系。

5 结论

双聚类分析可以反映出高频词与来源文献的对应关系，可用于某学科领域的学科热点结构和近期前沿研究。基于以免疫学为样本的共词分析，绘制聚类树图展现出某学科领域的主题结构、战略坐标展现出各个研究主题在整个学科结构上的重要性或特性、社会网络图谱展现出主题内部关系。不同可视化相互嵌入，可从多方位展现学科结构的发展过程，宏观结构与微观分布一目了然，可以指导学科专业人员了解学科结构及前沿分布，帮助决策层规划学科布局，调整学科方向，促进学科发展。

PU-047

Data Mining of The Knowledge Structure in Medical Informatics by Co-citation Biclustering Analysis and Strategic Diagram

杨颖、徐爽、韩爽、许丹、邱宇红
中国医科大学

Materials and Methods Co-citation biclustering analysis was performed using the papers on medical informatics covered in the Web of Science core collection as a source from 2017 to 2018 by gCluto. And citation strategic diagram was made based on co-citation biclustering analysis.

Results Six intellectual bases and corresponding research fronts of medical informatics were got by solution reports and matrix visualization for the citation-citing matrix. The results were visualized directly by gCluto. According the novelty and attention of each cluster calculated by time sorting analysis, strategic diagram of citations in the field of medical informatics was drawn.

Conclusion Combination use of biclustering and strategic diagram in the co-citation analysis could be useful to provide a decision-making reference for professionals and managers and to promote the development of medical informatics.

PU-048

基于 FFCA 的电子病历命名实体领域本体构建研究

王凯
蚌埠医学院

目的 将模糊形式概念分析（FFCA）方法引入电子病历（EMRs）知识挖掘领域，利用本体技术构建医学领域命名实体及其依赖关系，建立机器可读的结构化电子病历。

方法 首先在 EMRs 文档集合的识别和预处理的基础上，获取命名实体并分析其层次关系，建立医学领域概念形式化描述模型，然后通过分析实体间的蕴含依赖关系，提出属性最简蕴含决策规则判定定理，将 EMRs 语义知识转换成决策形式背景，最后利用命名实体概念格生成领域本体。

结果 用文本通用工具（GATE）对本体结果的实体信息抽取与统计分析，进一步验证了基于 FFCA 的中文 EMRs 实体本体构建方法的正确性和有效性。

结论 本文针对 EMRs 中命名实体及其层次关系的抽取及语义表示问题，提出基于 FFCA 的领域本体构建方法。

引言

大数据背景下的医疗数字化和信息化产生了大量的医疗健康数据，海量医疗数据为面向临床诊疗决策的信息抽取技术提供了理想的信息来源。如何从文本语料库中抽取疾病诊断、症状及诊疗信息，实现医学信息的存储、管理及传输等，成为医学领域知识管理与知识发现的重要问题[1]。医院信息系统的电子病历(EMRs)蕴含大量的医疗健康数据集，通过分析 EMRs 的语义关联信息，挖掘与患者相关的医疗知识，能够提高 EMRs 各类命名实体的语义表达能力。

在大数据背景下，基于结构化的 EMRs 本体语义建模，构建具有层次关联关系的领域术语体系，具有十分重要的现实意义[2]。国内外学者针对领域本体建模开展了广泛而深入的研究，提出了多种本体构建方法（七步法、骨架法、TOVE 法等），开发了大量的本体表示构建工具（WebOnto、OntoEdit 以及 Protégé等）[3]。然而领域本体的构建往往需要领域专家的深入参与，构建标注信息往往费时费力，知识表达冗余低效，严重降低了本体构建的质量与效率。

模糊形式概念分析理论（Fuzzy Formal Concept Analysis,FFCA）是融合模糊集和形式概念分析理论的数据分析与层次关联规则提取方法，被广泛用于机器学习、数据挖掘等领域[4]。FFCA 以概念格（Concept Lattice）为基础，通过引入模糊隶属度，刻画领域内不确定性知识，形式化表达领域概念及其层次关系，挖掘知识间关联规则。研究基于 FFCA 的领域本体构建方法能够有效识别 EMRs 中的命名概念实体，依据上下文语境抽取概念特征，综合考虑医学概念间的类关联关系、上下文关系以及等价类关系等，获取具有深层语义关联关系的术语集。但从现有研究来看，基于 FFCA 理论的概念实体挖掘技术用于 EMRs 的命名实体抽取的研究较少：丁卫平[5]等将粗糙集理论与形式概念分析理论相结合，结合 EMRs 数据库中医学数据特征，采用条件熵对病历大量属性进行约简，构造粗糙决策规则格算法，提高了决策规则的挖掘效率；Rehman[6]等利用中图法建立医学领域形式背景,探索基于 FCA 的医学文本分类模型；牟冬梅[7]等结合临床术语（SNOMED CT）的概念表、关系表等外部信息，构建甲状腺疾病本体，并结果本体进行语义标注。

基于以上分析，本文将 FFCA 理论引入 EMRs 领域本体建模过程，利用医学语料库丰富的语义关系，建立医学概念形式化描述模型，提取领域隐含二元关系的概念层次结构，以概念格的形式

展示概念间泛化与特化关系，在识别中文医学命名实体的基础上，提出并证明了最简蕴含决策规则判定定理，并以此为基础获取决策形式背景，构建命名实体概念格，最后通过概念格映射领域本体，并进行可视化表示。

1 相关理论

1.1 经典形式概念分析理论

形式概念分析理论（Formal Concept Analysis，FCA）理论是一种概念层次分类理论，通过形式化表达概念及其层次关系，构建概念间的分类关系及其层次结构[11]。FCA 核心概念定义如下。

定义 1　形式背景

形式背景由 $K=(G, M, I)$，其中 $G=\{g_1, g_2, \cdots, g_n\}$是一组对象集合,$M=\{m_1, m_2, \cdots, m_y\}$是一组属性集合，对于任意 I，若存在 $I\subseteq G\times M$，表示 G与 M之间的二元关系，即$(g, m)\in I$，则称对象 g包含属性 m，记为 ${}_gI_m$。

定义 2　形式概念

对于任意集合 $Z=(X, Y)$，其中 $X\in G$，$Y\in M$，若满足双射关系：

$$X^{\alpha*}=\{x\in G\mid \forall y\in M, xIy\},\quad Y^{\beta*}=\{y\in M\mid \forall x\in G, xIy\}$$

则称 Z是形式背景 K上的形式概念，其中 X是形式概念的外延对象，Y是形式概念的内涵属性。

1.2 模糊形式概念分析理论（FFCA）

决策形式背景是 FFCA 中知识库的表示形式，引入决策属性集，将经典形式背景内对象属性间的二元关系扩展成具有决策属性的多元关系[8,9]。决策形式背景通常表示成五元组的形式，通过使用信息函数 f_P建立对象与决策属性上的映射关系，取代传统的二元关系 I。

定义 3 概念蕴含规则[10]

设形式概念 $Z_1(X_1, Y_1)$ 、$Z_2(X_2, Y_2)$，满足$(X_1, Y_1)\in L_1(O_1, A_1, I_1)$、$(X_2, Y_2)\in L_2(O_2, A_2, I_2)$ 。若概念非空外延对象满足 $X_1\subseteq X_2$，则称形式概念 Z_1蕴含 Z_2，记为$(X_1, Y_1)\Rightarrow(X_2, Y_2)$，称形式概念间的蕴含规则 $Y_1\rightarrow Y_2$存在。

定义 4 概念蕴含决策规则

在决策形式背景 $K_d=(G, C\cup D, f_P, V)$ 中，对任意形式概念 $Z_1(X_1, Y_1)\in L_1(O_1, A_1, I_1)$ ，存在 $d\in D$ 且满足映射关系 $d^{\alpha*}$，则称关系 $R_{Y_1\to d}$ 为决策形式背景 K_d 上的一条蕴含决策规则。

2 EMRs 领域本体构建流程

本文将 FFCA 理论引入 EMRs 的命名实体领域本体构建模型，在识别医学命名实体的基础上，分析其层次关系；同时给出实体间的蕴含依赖关系，据此获取决策形式背景，构建 EMRs 模糊概念格；最后通过概念格映射领域本体，并进行可视化表示。

2.1 预处理

EMRs 是患者接受诊疗经过的文本记录，包含大量与患者健康状况密切相关的文字、符号、影像等数字化信息。EMRs 文本的信息挖掘需要抽取记录患者治疗的各种概念命名实体以及实体间的各类关联关系，因而需对其书写规范、语言描述、文本结构等特点开展有针对性的分析，获取核心实体及其语义层次关系。EMRs 的语言特征分析要以医疗领域字典为指导，综合运用分词、词语标注、句法分析等自然语言处理技术（Natural Language Processing, NLP），抽取半结构化的文本信息。

2.1.1 词性特征处理

EMRs 的语言还具有语句模式化表达、语句表达简短精炼的结构特点，语句中的概念实体识别对词性的边界较敏感，如“具有”、“疑似”、“患有”等谓语动词的左边界往往包含与患者疾病相关的命名实体；同时概念间上下文关系的修饰语通常是动词。本文采用词性边界标注的方法，以宾州中文树库[11]（PennChinese Treebank，PCTB）的词特征标注规范为基础，借助领域专家抽取的住院病历文本，构建 3 个外部语句词典语料库，获取带有标注信息的语法结构树，抽取出的心内科相关症状或检查手段的标注信息，见表 1 所示。

表 1 命名实体词性特征标注信息

症状或检查手段	字段名	字段类型	标注	语义描述
年龄	AG	String Variable	a_1	对象年龄的数字表示

性别	SX	Boolean Variable	a_2	对象性别
胸口疼痛	CP	String Variable	a_3	四种常见胸疼症状
血压	BP	String Variable	a_4	血压数值
胆固醇	CH	String Variable	a_5	胆固醇数值
血糖	BS	Boolean Variable	a_6	空腹血糖含量大于 120mg/dl 为 true
心电图	ECG	String Variable	a_7	是否有 T 波段
最大心率	MHR	String Variable	a_8	最大心跳数（单位时间）
运动状况	EX	Boolean Variable	a_9	运动是否伴随心绞痛，true 为是，反之为否
运动相对于休息的 st 段压数值	OP	String Variable	a_{10}	ST 段压数值

2.1.2 多值化归一表示

由于 EMRs 具有半结构化文本特点，需要根据字段的语义描述信息，对命名实体进行缺值语义填充、同义化替换等操作，以保持语义的完整性，如 “冠心病”是“冠状动脉粥样硬化性心脏病”的简称，需要对其进行语义填充，同时将二者进行同义化标注，合并成相同的命名实体。

针对命名实体特征标注过程中出现的数据类型多值化问题，采用基于逻辑回归二分类模型[12]，单值化表示实体数据，去除数据量纲，多值归一化后的命名实体如表 2 所示。

表 2 命名实体特征多值归一化表示

属性	划分标准	分类	属性	划分标准	分类
AG	≤35	You(1)	BS	<126	Normal(1)
	36-60	Mid(2)		>126	Very high(2)
	61-75	Old(3)			
	≥76	Very old(4)			
SX	——	Male(1)	ECG	[-0.5,0.4]	Normal(1)
		Female(2)		[2.45,1.8]	ST-T abnormal(2)
				[1.4,2.5]	Hypertrophy(3)
CP	——	Typical angina(1)	MHR	<60	Medium(1)
		Atypical angina(2)		60-100	Normal(2)
		Non-angina pain(3)		>100	High(3)
		Asymptomatic(4)			
BP	to 139/80 to 90	Normal(1)	EX	——	False(1)
	140 to 159/90 to 99	Medium(2)			True(2)
		High(3)			
	160 to 179/100	Very high(4)			

	to 109				
	≥180/110				
CH	<160	Low(1)	OP	<2	Low(1)
	160-190	Medium(2)		2-3	Risk(2)
	190-250	High(3)		>3	High(3)
	>250	Very high(4)			

2.2 领域本体建模

命名实体概念格采用 Navigli[13]提出的本体层次构建方法，以 OWL 本体描述语言为基础，将概念格转换成领域本体，通过语义化的逻辑表示，对 EMRs 中的命名实体进行存储和语义关系解释，提升实体间的语义互操作能力。由于 EMRs 中包含大量命名实体，且实体间存在隐语义信息，需要借助领域内结构化主题词表等资源，进一步清晰化概念间的分类关系。

在领域专家的指导下，将概念格内的节点概念视为本体概念，相关术语定义成本体的类，术语间的层次关系表示成本体类之间的上下位关系，使用 OWL 语言的关系标签 SubClassOf 表示。具体做法是，①采用自顶向下的方式，将概念格顶层节点定义为本体概念的根层概念；②依次将下层节点间的偏序关系连线转换成本体概念间的上下位关系；③参考转化医学本体 TMO 推荐的医学本体库中的术语本体构建指南[14]，修正或删除不含概念或对象的底层节点。相应的映射转换规则以及 OWL 术语层次关系表示如图 1 所示。

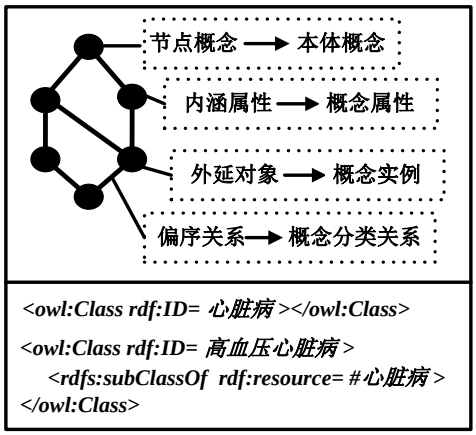


图 1 基于命名实体概念格的本体映射

3 本体构建实例

3.1 数据预处理

EMRs 中语言符号包括字和词语两种类型，面向中文文本的通用领域分词软件由于缺乏医学领域语料库的完整受控词汇信息，对 EMRs 的中文语句分词准确率较低。本文扩展了中科院 ICTCLAS 分词工具的用户词典和停用词词典，添加医学主题词表 MeSH（Medical Subject Headings）。采用以词典为基础的双向字符串最大匹配分词算法进行文本语言符号切分。

原始数据选取蚌埠医学院附属第一医院住院部心内科病案室 2014 年到 2016 年间的住院患者 EMRs 文本 338 份（包括医生查房记录与出院小结各 125 份），利用 ICTCLAS 分词工具切分数据词语，抽取语言符号特征、词性特征划分等。为了避免不必要的复杂性，在删除无意义符号、停用词等后，筛选出 35 个与心脏诊断相关的疾病专业术语，属性词数量为 1305。

3.2 FFCA 构建及分析

抽取出表 1 中的心内科相关症状的标注信息，将归一化后的患者症状作为概念属性。设 G={1，2，3…，10}，表示外延对象，C={a_1，a_2，a_3，a_5，a_6，a_9，a_{10}}，表示内涵属性，D={a_4，a_7，a_8}，表示由领域专家推荐的决策属性(本例以标记词频最高的三个属性为决策属性)。

分别求出 D 中决策属性的”α*”映射，即 $a_4^{\alpha*}$={1,3.4,6,7,8,9,10}，$a_7^{\alpha*}$={1,2,3,4,8,9,10}，$a_8^{\alpha*}$={1,2,3,4,5,6,7,8,9,10}。遍历出形式概念中外延对象是 $a_4^{\alpha*}$、$a_7^{\alpha*}$和 $a_8^{\alpha*}$子集的所有概念集合，依据外延对象的数量，按降序排列，可得到决策属性的最简蕴含决策规则集。利用决策形式背景构建 EMRs 概念格，通过 Java 应用程序分析命名实体概念和 EMRs 文本间的关联关系，获取背景知识。为便于理解和分析实验结果，在将背景知识导入 ConExp 分析软件之前，将表 1 中症状的标注符号还原成字段名属性，生成“病历文本×命名实体”语义矩阵。

本文使用英国谢菲尔德大学开发的 GATE[16]（General Architecture for Text Engineering）自然语言文本处理工具对构建的 EMRs 本体进行测试和验证。通过加载中文识别工具、OWLIM Ontology Tools 插件，将构建的命名实体蕴含决策规则集、EMRs 本体导入 GATE，进行可视化本体浏览，如图 2 所示。

将 EMRs 文本进行语义标注与验证后，获取基于 GATE 的 EMRs 命名实体语义标注信息，并以结构化的方式输出统计分析，如图 3 所示。结果证实了 EMRs 本体在 GATE 中具有较高的语义标注准确性，命名实体的查准率和查全率均有显著提高。

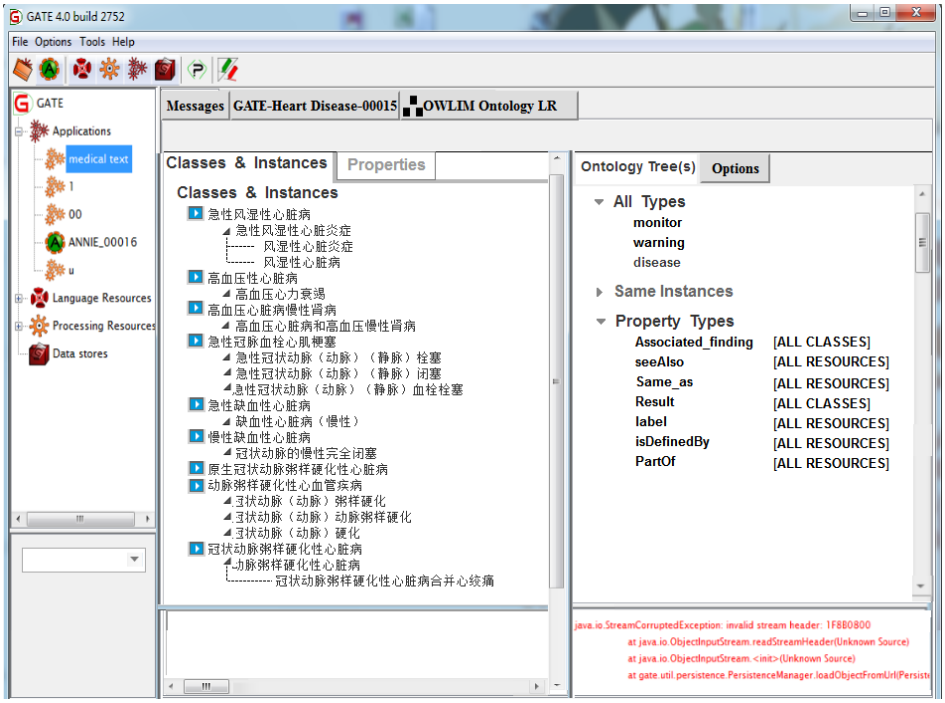


图 2 GATE 对 EMRs 本体的部分概念关系的可视化

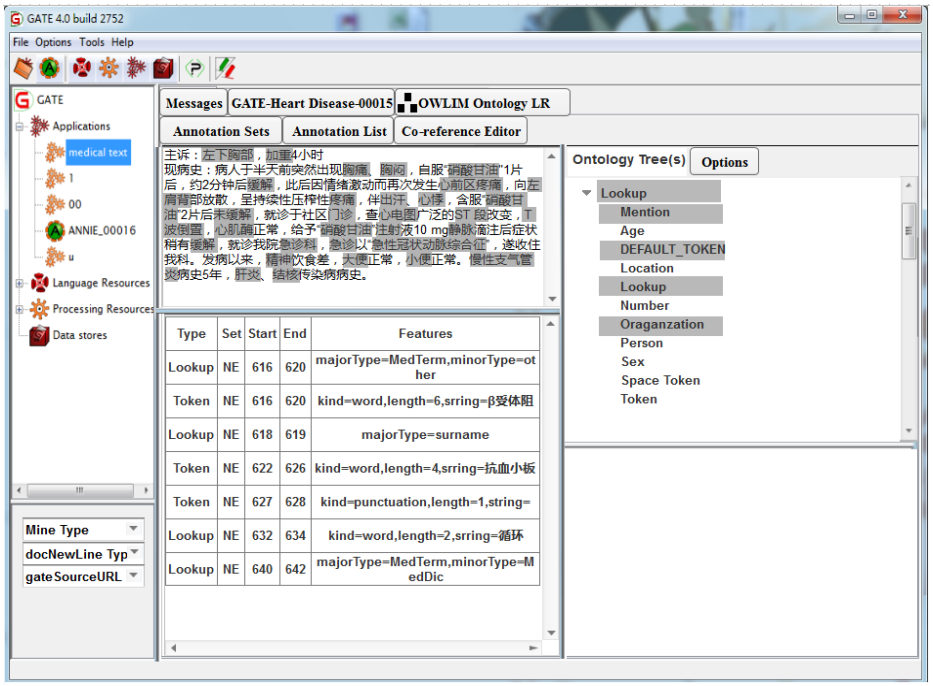


图 3 GATE 对 EMRs 本体的语义标注

4 结语

本文针对 EMRs 中命名实体及其层次关系的抽取及语义表示问题，提出基于 FFCA 的领域本体构建方法，首先通过识别中文医学命名实体，建立医学领域概念形式化描述模型，然后在分析实体间的蕴含依赖关系的基础上，将 EMRs 语义知识转换成决策形式背景，定义最简蕴含决策判定规则，得到最简属性，并依此构建命名实体概念格，最后通过概念格映射领域本体，后续研究将围绕以下三个方面展开：1）探索构建更大规模的 EMRs 语料库，研究命名实体及其关系的复杂语义抽取；2）增加 EMRs 数据来源的多样性，将医生查房记录、患者回访记录、出院小结等自由文本形式的非结构化数据作为研究对象，提高医学领域术语挖掘的深度；3）探索基于 FCA 理论的大型决策形式背景的快速约简算法，实现本体建模与本体学习的并行化操作。

PU-049

近 10 年来神经网络辅助诊断乳腺肿瘤技术研究的可视化分析

邱煜炎、吴福生
蚌埠医学院

目的 分析近 10 年来神经网络辅助诊断乳腺肿瘤技术研究的发展趋势与热点，为计算机辅助诊断的研究领域提供参考和指导。

方法 以 2009 至 2018 年 Web of Science 核心合集中收录与该领域研究相关的文献作为数据来源，运用可视化软件 Citespace 绘制了该领域的科学知识图谱，并从关键词引用频次、突现性、中心度、聚类、时间线视图等途径进行可视化分析。

结果 该领域研究热点总体呈上升的趋势，以神经网络、深度学习、乳腺癌症和肿瘤、病灶分割、特征分析以及算法优化等作为研究重点。

结论 该领域研究热点突出，研究目标聚集，具有很强的新颖性和实用性，国内学者应加强该领域的合作研究。

PU-050

基于专利计量的医院专利管理对策分析

吕鸣[1]、邹立君[2]、侯胜超[3]
1. 华中科技大学同济医学院医药卫生管理学院 2. 华中科技大学图书馆医学分馆
3. 华中科技大学同济医学院附属同济医院图书馆

目的 以湖北省的全部医院为例分析医院专利申请及授权状况，对医院专利管理工作提出建议。

方法 利用 Patentics 平台，检索湖北省全部医院的国内专利申请情况，并对专利的数量、类型等进行分类统计分析。

结果 该省医院专利申请量 7694 件，授权量 6491 件，专利申请呈现增长态势；专利类型以实用新型为主，发明专利次之；专利 IPC 分类集中在 A61 以及 G01 等技术领域；有效专利占比较低，后期维护工作欠佳。

结论 应着力提升医务人员创新能力，调动专利申请积极性；加强专利申请的过程管理，提高专利质量意识；加大优势技术领域投入，重点关注高价值专利的识别及转化；提升医院专利管理部门的服务质量，培养专利管理人才。

PU-051

高校规范化使用 ESI、Incites 指标助力双一流学科建设的知识情报服务实证研究

杨颖、韩爽、许丹、徐爽、陈斯斯
中国医科大学

目的 主要从高校规范化使用 ESI、Incites 指标的视角，挖掘学科总体发展态势、动态演进以及学科竞争力，并探讨助力双一流学科建设的科研决策建议。

方法 通过检索 ESI 和 Incites 数据库，以中国医科大学为例，探索各学科发展态势，从科研产出、科研影响力、并将通过其中各项指标绝对值转变为相对指标，探寻高质量论文产出能力及高影响力学科表现及相对表现。

结果 各学科的相对指标与绝对指标相差甚远，将相对指标进行 SWOT 分析，找到优势学科、弱势学科、潜力学科和威胁学科。

结论 规范使用评价指标，将绝对指标转化为相对指标，为学科发展做出科研决策建议，为其他同行提供学科服务参考。

1、 前沿

学科发展状况是高校整体实力评价的重要因素，也是体现高校执教能力、办学水平的重要标志，加强学科建设是高校可持续发展的必由之路，同时也是提高教育质量的重要途径。2014 年 5 月 4 日，习近平总书记指出：“办好中国的世界一流大学，必须有中国特色。2015 年 10 月，国务院公布的《统筹推进世界一流大学和一流学科建设总体方案》（国发〔2015〕64 号）提出加快建成世界一流大学和一流学科[1]。2016 年 2 月，教育部印发文件，明确要求加快“双一流”建设。2017 年 9 月 21 日，教育部、财政部、国家发展改革委联合发布《关于公布世界一流大学和一流学科建设高校及建设学科名单的通知》[2]，公布首批双一流建设高校及学科名单。从此，“双一流”建设取代 985、211，作为新时代国家对高等教育新的改革是当下国家发展教育强国的重要战略任务。根据国家的部署，我国高校以建设国际一流大学和一流学科为目标进行规划，主动参与国内学科评估的同时，将学校自身或学科置于国际环境中进行评估，了解本校在国际高校中的位置，及本校各个学科的发展情况，为明确学科发展方向提供更多的决策支撑[3]。

在学科评估中常用 Web of Science（SCIE/SSCI）、基本科学指标（Essential Science Indicators，简称 ESI）、InCites 和自然指数(Nature index，NI)平台的科研评价指标。ESI 是衡量科学研究绩效、跟踪科学发展趋势的基本分析评价工具，已成为当今世界范围内普遍用以评价高校、学术机构、国家/地区国际学术水平及影响力的重要评价指标工具之一[4]。ESI 指标成为 2012 年教育部学位与研究生教育发展中心的重要评估指标，纳入协同创新项目的评定。InCites 则是在汇集和分析 Web of Science 核心合集七大索引数据库基础上建立起来的综合性科研评价工具[5]。该数据库提供更全面的数据资源和多元化的分析指标，不仅包含全球各个国家、机构和学科的论文统计结果，而且提供如学科规范化的引文影响力、期刊规范化的引文影响力、平均百分位等独特的计量指标。代表另外一种评价指向的 NI，它遴选全球在基础研究领域最顶级的 68 种学术期刊，统计其论文的发表情况，即顶级期刊,力图在文献计量学上定位于全球优质研究综合指标[6]。目前各高校利用几个数据库的基本指标进行学科评价的研究如火如荼地开展，有学者对 ESI 学科的动态研究[7-8]，有学者对 ESI 潜力学科识别及预测的分析[9-10],有学者利用其研究高校的科研实力[11]。有学者比较 InCites 平台及 SciVal 的科研影响力评价差异[12]，以及自然指数和基本科学指标的差异[6]。

2020 年 2 月，教育部、科技部联合重磅发文，指出当前科研评价中存在 SCI 论文相关指标片面、过度、扭曲使用等现象，要求大学和研究机构规范 SCI 论文以及相关指标（如 JIF，ESI 等）的使用[13]，在科研学术界引起轩然大波。高校图书馆如何利用 SCI、ESI、Incites 进行双一流建设服务面临巨大挑战。已有学者介绍相对引用率等新的影响力评价指标[14]，但利用相对指标进行 SWOT 分析少有报道。

本文主要从高校规范化使用 ESI、Incites 指标的视角，探讨助力双一流学科建设提供科研决策建议。本文通过检索 ESI 和 Incites 数据库，以中国医科大学为例，探索其科研产出、科研影响力，并将通过其中各项指标绝对值转变为相对指标，从相对科研产出、相对科研影响力以及利用相对值作为基础数据进行各学科的 SWOT 分析，找到优势学科、弱势学科、潜力学科和威胁学科，并为学科建设提供科研决策建议。

2、我校 ESI 前 1%学科及潜力学科发展状况

通过检索 ESI 数据库，截止至 2020 年 3 月 12 日更新数据，数据范围为 2009 年 1 月-2019 年 12 月 31 日，我校共有 5 个学科进入了世界排名前 1%的行列，分别是：临床医学、神经科学与行为学、药理学与毒理学、生物学与生物化学，分子生物学和遗传学。国际排名 1031，国内排名 69，深圳大学超过我校排名 68 位。在国内同类高校入选 ESI 的学科数排名中，我校排名第 10 位，位居北京协和医学院、首都医科大学、南京医科大学、海军军医大学、空军军医大学、南方医科大学、天津医科大学、陆军军医大学、哈尔滨医科大学之后。表 1 列出我校从国际排名及百分位、国内排名及百分位、发文数量、被引次数、篇均被引、TOP 论文来展现我校各个学科表现。

表 1　我校进入世界 1%排名学科情况

学科名称	国际排名	机构总数	国际百分位	国内排名	国内高校总数	国内百分位	论文数量	被引次数	篇均被引	TOP 论文
所有学科	1031	6581	0.16	69	288	0.24	14285	147134	10.30	80
临床医学	568	4599	0.12	18	101	0.18	7299	72045	9.87	51
神经科学与行为学	544	920	0.59	17	38	0.45	1175	14517	12.35	5
药理学及毒理学	401	938	0.43	29	63	0.46	966	9445	9.78	6
生物学与生物	698	1104	0.63	44	69	0.64	1417	12431	8.77	0

化学										
分子生物学与遗传学	678	841	0.81	31	39	0.79	1605	19838	12.36	11

表 2 分析了我校潜力学科发展状况，包括免疫学、环境科学与生态学、微生物学、材料科学、化学、社会科学总论、精神病学与心理学，计算出潜力学科进入 ESI 前 1%的潜力值。我校未进入 ESI 的各学科中，免疫学潜力值为 0.97，社会科学总论的潜力值为 0.88，是最靠近 ESI 门槛值的两个学科。

表 2 我校潜力学科情况

学科名称	ESI 机构总数	ESI 国内高校总数	国内排名	论文数量	被引次数	阈值	潜力值	TOP 论文
免疫学	789	24	25	476	5235	5419	0.97	1
环境科学与生态学	1093	52	111	143	1811	4544	0.40	0
微生物学	488	15	57	165	1552	5774	0.27	0
材料科学	941	136	194	185	2747	6965	0.39	1
化学	1328	168	231	355	2900	8502	0.34	0
社会科学总论	1605	38	47	142	1410	1610	0.88	1
精神病学与心理学	736	7	26	157	1384	4398	0.31	1

3、我校近五年论文产出趋势

本文欲探索近五年的学科竞争分析，选取 Incites 及 WOS 数据库，检索策略：Incites 选中研究方向，设置筛选条件：ESI 学科分类，研究方向，机构名称”china medical university”，文献类型选择 article 和 review 两种，出版年 2015-2019，共发文 8942 篇。表 3 中从 Web of Science 论文数、被引次数的绝对值，以及相对规范化的指标，如学科规范化引文影响力、被引次数排名前 1% 的论文百分比、Q1 期刊中论文的百分比以及国际合作论文百分比等多维度指标了解我校 2015-2019 五年间各学科的差异表现。

表 3 2015-2019 五年间我校各学科表现

名称	排名	Web of Science 论文数	被引频次	学科规范化的引文影响力	被引次数排名前 1%的论文百分比	论文被引百分比	Q1 期刊中论文的百分比	国际合作论文百分比
Clinical Medicine	1	4481	26693	0.99	0.89	75.03	20.64	17.25
Molecular Biology & Genetics	2	1126	10432	1	0.62	83.3	48.87	19.63
Biology & Biochemistry	3	929	4788	0.71	0.11	72.87	17.8	17.76
Neuroscience & Behavior	4	671	5058	0.85	0.6	82.41	28.91	23.25
Pharmacology & Toxicology	5	629	3640	1.11	1.27	78.22	29.4	16.53
Immunology	6	275	1664	0.76	0.36	78.91	24.4	32.36
Chemistry	7	213	1199	0.59	0	72.77	22.73	13.62
Materials Science	8	129	1016	0.97	0	81.4	58.54	27.13
Psychiatry/Psychology	9	100	421	1.06	1	75	50	33
Environment/Ecology	10	89	527	0.76	0	85.39	28.77	15.73
Microbiology	11	85	354	0.61	0	78.82	35	38.82

Social Sciences, general	12	79	365	1.15	1.27	75.95	37.29	27.85
Agricultural Sciences	13	45	327	1.06	0	77.78	56.25	24.44
Engineering	14	24	115	1.36	4.17	45.83	42.86	25
Physics	15	21	227	2.03	4.76	76.19	83.33	52.38
Multidisciplinary	16	15	85	0.62	0	100	86.67	33.33
Plant & Animal Science	17	12	40	1.09	0	66.67	37.5	33.33
Mathematics	18	9	19	0.84	0	77.78	14.29	44.44
Computer Science	19	7	6	0.7	0	57.14	50	71.43
Geosciences	20	3	18	0.51	0	100	66.67	33.33

3.1 总体发文趋势：我校在 2015-2019 年期间论文产出如图 1 所示，2016 年稍有下滑，之后迅速增长。

图 1 2015-2019 年论文产出趋势

3.2 各学科论文产出趋势

图 2 可以得出我校产出最好的五个学科为临床医学、分子生物与遗传学、生物化学、神经科学与行为学、药理与毒理学。

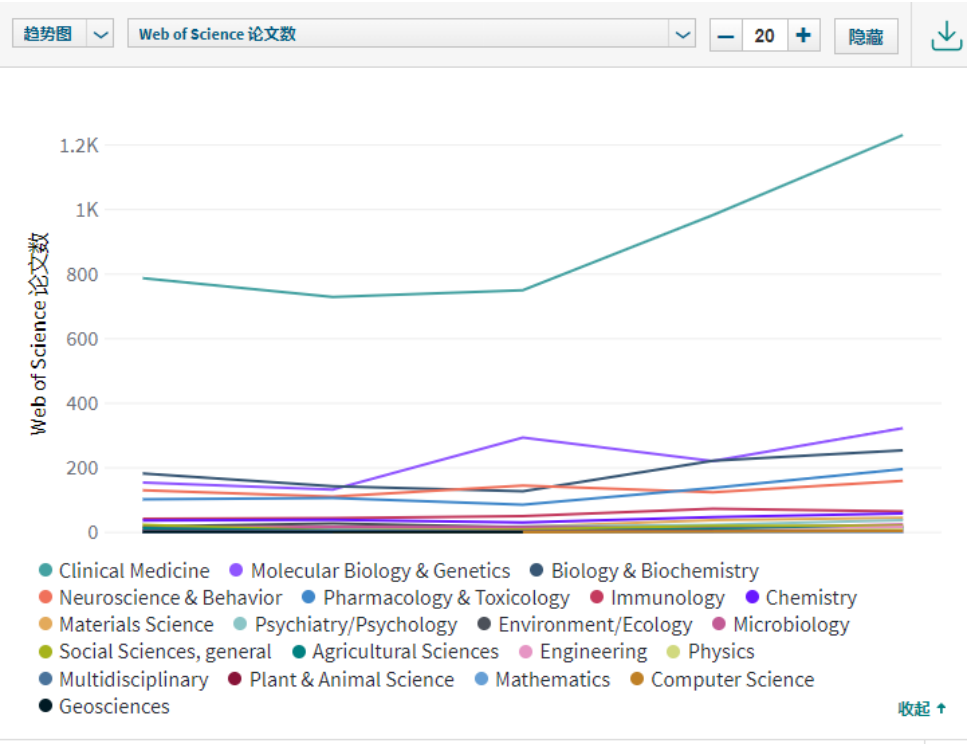


图 2　2015-2019 年各学科的论文产出趋势

4、我校近五年论文被引趋势

4.1 我校总体被引趋势

学科规范化的引文影响力（CNCI）是 Incites 平台中考察机构、国家、个人等的论文影响力的指标，也是排除了出版年、学科领域与文献类型作用的无偏影响力指标，可以用它进行不同规模、不同学科的论文集的比较。一篇文献的 CNCI 是通过其实际被引次数除以同文献类型、同出版年、同学科领域文献的期望被引次数获得的。如果 CNCI 的值为 1，则说明该组论文的被引表现与全球平均水平相当；CNCI 值大于 1，表明该组论文的被引表现高于全球平均水平；小于 1，则低于全球平均水平。对于一篇被划归至多个学科领域的论文，其 CNCI 为每个学科领域实际被引次数与期望被引次数比值的平均值。一组文献的 CNCI，是该组中每篇文献 CNCI 的平均值。如果 CNCI 等于 1，说明该组论文的被引表现与全球平均水平相当；CNCI>1 表明该组论文的被引表现高于全球平均水平；CNCI<1 说明该组论文的引文影响力低于全球平均水平。

我校总体被引表现整体是上升趋势，只有 2018 年 CNCI>1，我校论文的引文影响力超出全球平均水平，2019 年较 2018 年稍有下滑，但仍接近全球平均水平。

图 3 我校总体被引趋势（学科规范化的引文影响力 CNCI）

4.2 我校各学科的平均学科规范化影响力

虽然从被引次数的绝对值来看，临床医学是最多的，从学科规范化影响力来看，图 4 可知我校被引表现最好的几个学科，物理学、工程科学、社会科学、药理与毒理、植物与动物科学、农业科学、精神病学/心理学 CNCI 大于 1，说明被引表现高于全球平均水平；

我校临床医学、分子生物与遗传学、材料科学、数学在 0.9 与 1 之间，表明我校这几个学科被引表现稍微低于全球平均水平；其他学科被引表现较全球水平低。

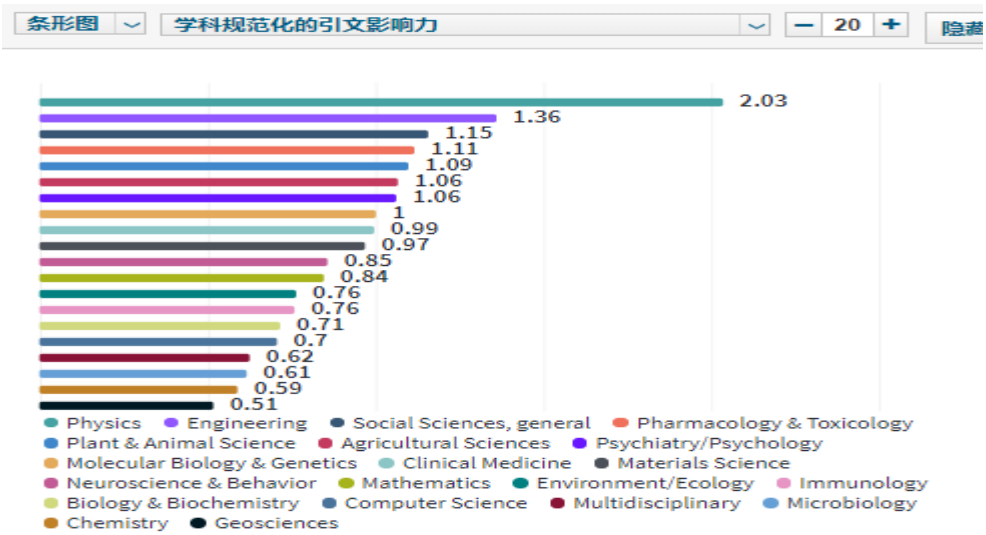


图 4 我校各学科的平均学科规范化影响力

4.3 进入 ESI 1%学科及最具潜力学科 CNCI 趋势图

我校进入 ESI 1%的学科有临床医学、分子生物学与遗传学、生物学与生物化学、神经科学与行为学、药理学与毒理学，这几个学科的 CNCI 近几年比较稳定，都是 1 附近震荡；最具潜力的免疫学基本在 1 附近，比进入 ESI1%学科 CNCI 值稍微低一些，社会科学总论在 2018 年大幅下降，但 2019 年表现非常突出，为其晋升 ESI1%做充分准备；而化学在 2019 年下降明显。

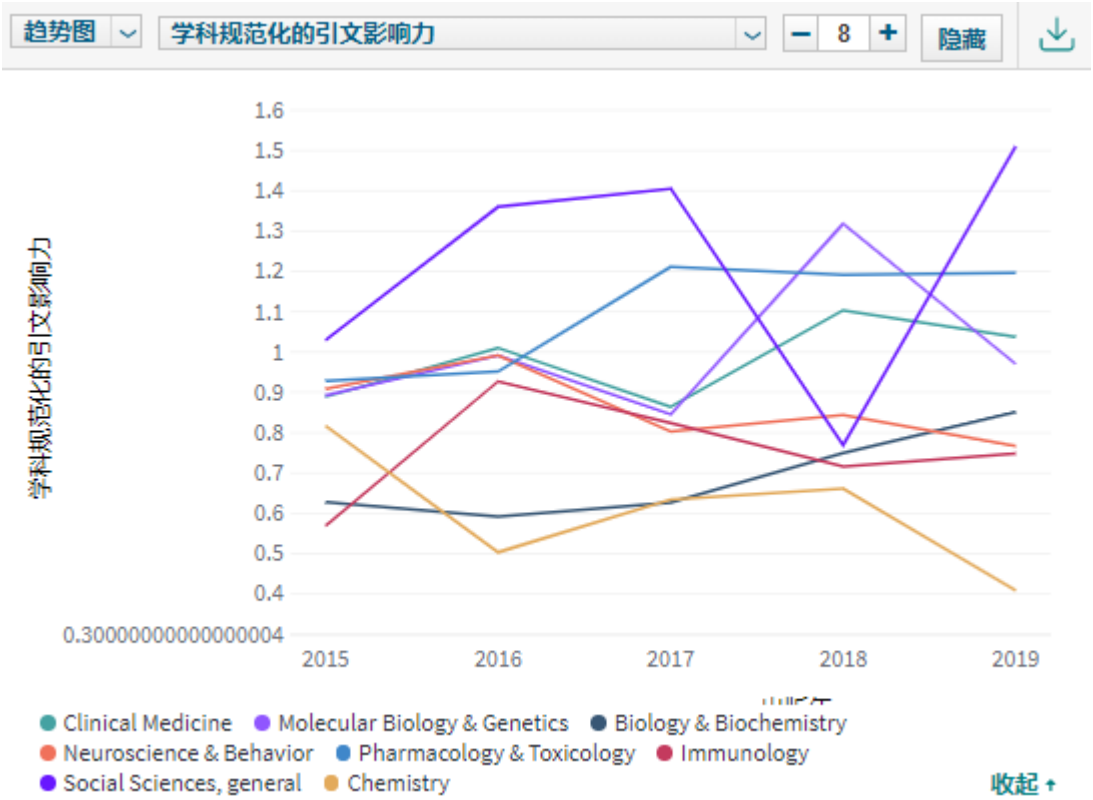


图 5 进入 ESI 1%学科及最具潜力学科被引表现趋势

5、我校各学科高质量论文产出能力及高影响力学科表现

5.1 高质量论文产出

5.1.1 我校 Q1 区论文数

Q1 区期刊是体现学术影响力的重要指标，Q1 区期刊发文多，代表有更多的机会被其他文章引用。临床医学 Q1 区论文最多，其次是分子生物学与遗传学、生物化学、神经科学与行为、药理毒理、免疫、化学等依次递减。

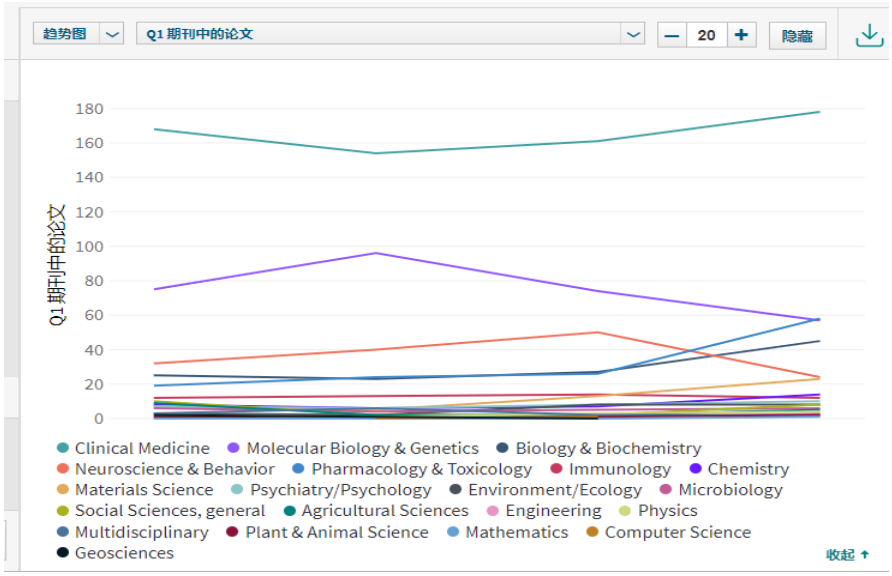


图 6 我校 Q1 区论文数趋势图

5.1.2 Q1 区期刊论文的百分比

虽然我校临床医学发文的 Q1 区期刊最多，但临床医学的占比是相对较少的，生物与生物化学也是在绝对值突出，但是相对 Q1 区占比最低。

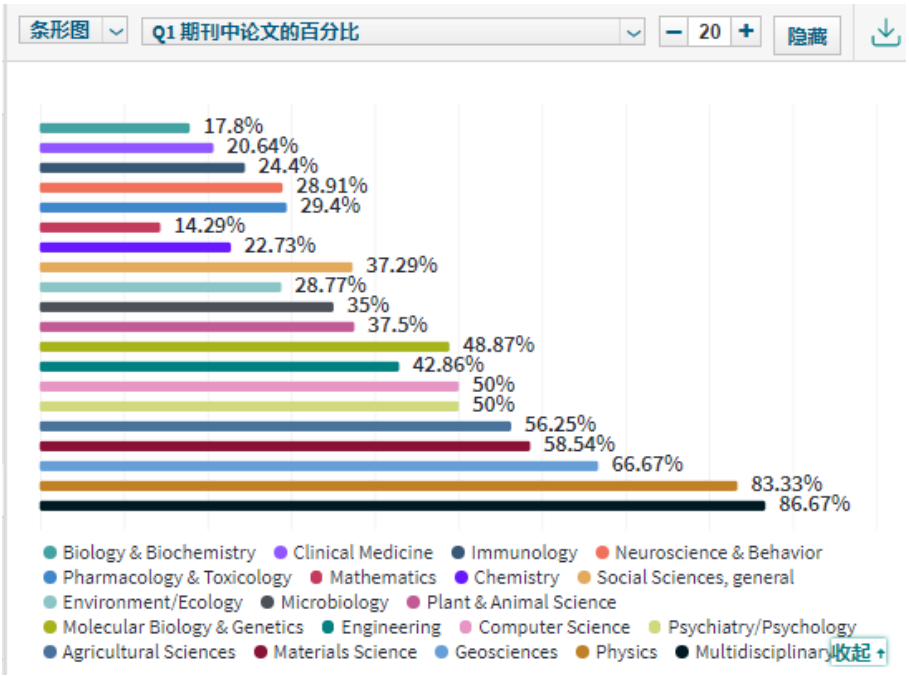


图 7 我校各学科发文 Q1 区论文百分比

5.2 高影响力论文产出能力学科表现

指标：被引频次排名前 10%的论文。如果一篇论文的被引频次在同出版年、同学科、同文献类型的论文中位于前 10%，则该论文被称为“被引频次排名前 10%的论文”。

某论文集合（如某高校）被引频次排名前 10%的论文百分比：集合中被引频次排名前 10%的论文数除以集合中的论文总数。

从我校各学科的被引频次排名前 10%论文百分比图中可知，高影响力论文产出能力学科从高到低排序为：物理，工程学，农业科学，数学，药理学，社会科学分子生物与遗传学、临床医学、动植物科学、生化、神经科学与行为学、心理学、材料科学、免疫学、化学环境科学。

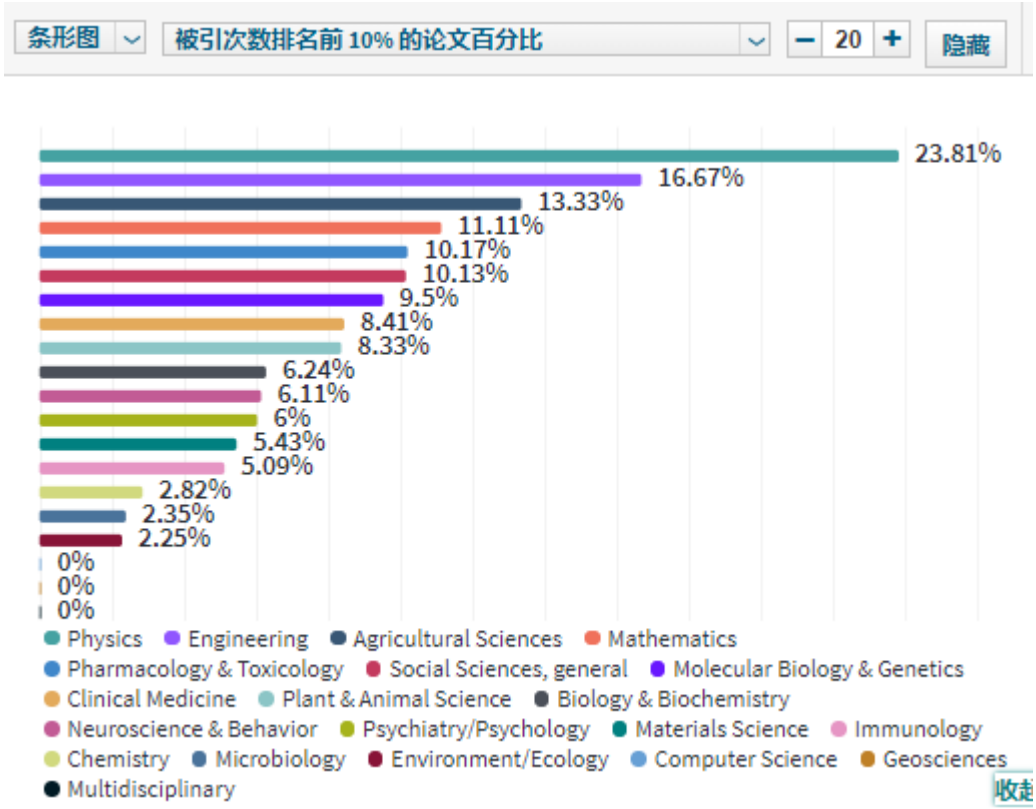


图 8　高影响力论文产出能力学科表现

5.3 我校被引频次排名前 10%的相对表现

我校与全球学科基准值的比较，相对学科表现：物理，工程学、农业科学、数学、社会科学、药理与毒理学大于 1,因此，体现出高影响力水平，分子生物与遗传学、临床医学与动植物科学临近 1，其高影响力水平不及但接近全球范围。其中全球基准值不是理想值 10%，由于被引频次排名前 10%附近论文的被引频次相同，因此通常低于理想值 10%。

表 3 我校被引频次排名前 10%的相对表现

名称	我校学科表现	全球基准值	相对学科表现
Physics	23.80952	9.415748	2.528692
Engineering	16.66667	9.605676	1.735085
Agricultural Sciences	13.33333	9.071043	1.469879

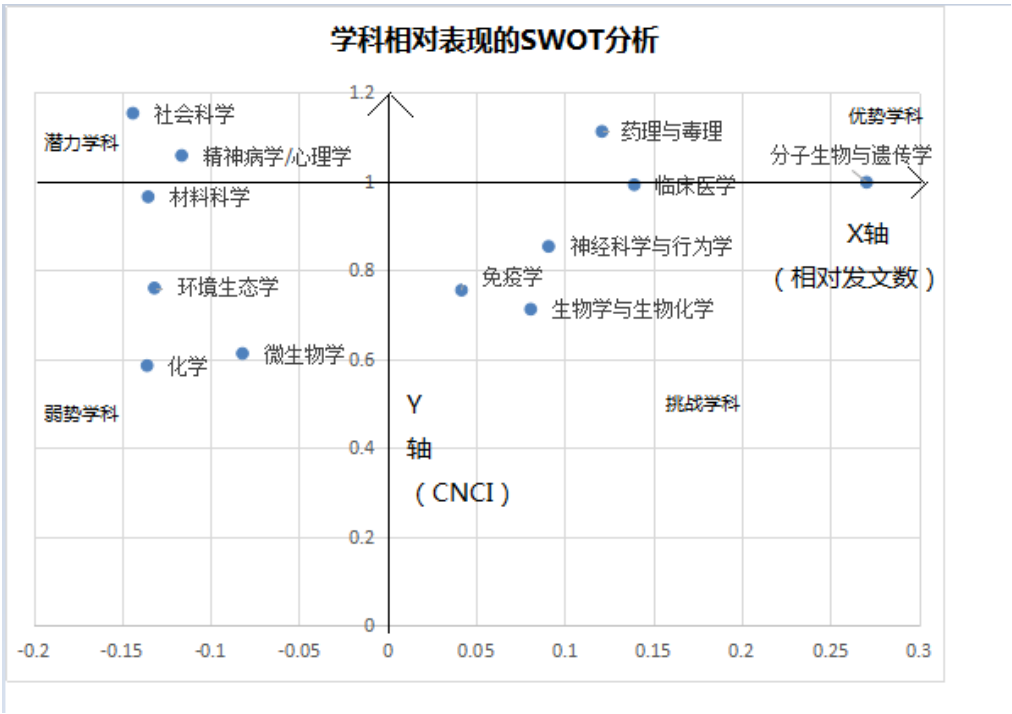
Mathematics	11.11111	8.24044	1.348364
Social Sciences, general	10.12658	8.658691	1.169528
Pharmacology & Toxicology	10.17488	9.098346	1.118322
Molecular Biology & Genetics	9.502664	9.574042	0.992545
Clinical Medicine	8.413301	9.049209	0.929728
Plant & Animal Science	8.333333	9.247102	0.901183
Psychiatry/Psychology	6	8.54907	0.701831
Biology & Biochemistry	6.243272	9.065627	0.688675
Neuroscience & Behavior	6.110283	9.557722	0.639303
Materials Science	5.426357	9.391492	0.577795
Immunology	5.090909	9.183711	0.554341
Chemistry	2.816901	9.501216	0.296478
Microbiology	2.352941	9.141739	0.257384
Environment/Ecology	2.247191	8.797629	0.255431
Computer Science	0	9.315673	0
Geosciences	0	9.481142	0
Multidisciplinary	0	9.446331	0

6、我校各学科 SWOT 分析

传统的 SWOT 分析，是以发表的 WOS 论文数为横坐标，期刊规范化引文影响力为纵坐标，原点处以论文发文均值为次要坐标轴基点，以 CNCI 值 1.0 为次要纵坐标轴基点，划分出 4 个象限。本文将对我校进入 ESI 前 1%学科及潜力学科共 12 个学科进行改良 SWOT 分析：将 SWOT 分析的指标进行规范化处理（表 4），将横轴的绝对 WOS 发文值转化为学科相对发文值，抵消各个领域研究不同广度的影响，相对发文值为每个领域的绝对 WOS 发文值除以该领域全球 WOS 发文值。纵轴依旧是期刊规范化引文影响力（CNCI），原点处以论文相对发文均值为次要坐标轴基点，以 CNCI 值 1.0 为次要纵坐标轴基点，划分出 4 个象限，如图 9 所示。第一象限为优势学科，其次是 CNCI>1，相对发文数在均值以上，最具优势为药理与毒理学，应引起足够重视；第二象限为潜力学科，也可称为机会学科，即 CNCI 值超过 1、发表论文数相对较少的学科，包括社会科学、精神病学与心理学；第三象限是弱势学科，CNCI 小于 1，并且相对发文数较少，包括材料科学、环境科学与生态学、微生物学、化学。第四象限为挑战学科，CNCI 小于 1，但相对发文量在均值以上，如果提升其被引，有机会进入优势学科，包括免疫学、神经科学与行为学、生物学与生物化学；分子生物与遗传学和临床医学这两个学科比较特殊，CNCI 接近 1，已经靠近第一象限，这两个学科处于挑战学科和优势学科的交界处。

表 4 进入 ESI 前 1%学科及潜力学科相对发文及 CNCI 表现

排名	名称	我校 WOS 论文数	全球 WOS 论文数	相对论文数	相对发文数与均值的差距	CNCI
1	临床医学	4481	1504231	0.297893076	0.139323264	0.992398349
2	分子生物学与遗传学	1126	262397	0.42912076	0.270550948	0.997862522
3	生物学与生物化学	929	387993	0.239437309	0.080867498	0.712270506
4	神经科学与行为学	671	268903	0.249532359	0.090962547	0.853559463
5	药理学及毒理学	629	224931	0.279641312	0.1210715	1.112230207
6	免疫学	275	137243	0.200374518	0.041804706	0.755115273
7	化学	213	934533	0.022792133	-0.135777679	0.585420188
8	材料科学	129	551776	0.023379052	-0.13519076	0.965277519
9	精神病学与心理学	100	236283	0.042322131	-0.116247681	1.05812
10	环境科学与生态学	89	330680	0.026914237	-0.131655575	0.759904494
11	微生物学	85	110887	0.076654612	-0.0819152	0.612665882
12	社会科学总论	79	534642	0.014776243	-0.143793569	1.153318987



7、基于学科分析的科研决策建议

7.1 总体发文数量及质量均有待提升：从绝对发文数量和被引频次的角度，进入 ESI 前 1%学科及潜力学科有绝对优势；但从高质量产出占比，以及学科规范化影响力角度，小学科以少而精的优势表现突出，进入 ESI 前 1%学科及潜力学科仅与国际水平持平或低于国际水平。因此各学科在追求数量的同时，更要注重质量，努力提高质量。

7.2 国际合作不稳定，应加强国际合作，与高水平机构创建长期稳定的合作模式，重点选择产出量及学术影响力高的机构作为合作目标。

7.3 优势学科：持续重点投入，提升数量和质量。从 SWOT 分析来看，临床医学以绝对的发文数量优势压倒其他学科，但是其 CNCI 的学科表现并不高，应大力提升其质量，提升其影响力，建议多发表 1 区论文。生物化学、分子生物与遗传学应该从数量和质量两个维度来加大建设力度。

7.4 对于潜力学科：稳质量，扩规模。我校潜力学科药理与毒理学、材料科学、心理学，应在保持原有的 CNCI 基础上，通过各种激励措施提高论文数量、扩大规模。向潜力学科过渡的社会科学、免疫学、神经科学与行为学这三个学科的 CNCI 接近 1，更要加大投入力度，鼓励科研人员向这几个方向的期刊投稿，同时提高影响力，充分利用科研网络平台对近期发表文章进行宣传，增加引用，使其尽快进入 ESI 前 1%，向优势学科进军。

7.5 弱势学科：引进领军人才，加强国际合作。环境、微生物、化学三个学科，需引进学科带头人，并加强国际合作。

7.6 威胁学科：神经科学与行为学、免疫学、生物学与生物化学这几个学科具有发文占比优势，但引文影响力相对较弱。因此，在继续保持优势基础上，应加强引用，通过利用不同社会化媒体，新闻报纸、博客、微博、论坛等国内外学术交流平台，如科学网、Researchgate、Twitter、Facebook 分享科研成果，提高自身的关注度，逐渐扩大影响力，提高引用率。

PU-052

医疗健康中的大数据分析

张伟祎
郑州大学第一附属医院

目的 大数据分析领域快速发展并已开始在医疗健康临床和研究方面发挥作用。医疗过程产生了大量完全不同、结构化、半结构化或非结构化的数据，大数据分析用来当作管理、处理、分析医疗相关数据的有效工具。文章主要着眼于大数据在医疗领域中的图像、生理信号以及基本组方向的应用来做以整理和分析。分析医疗大数据在上述三个领域中的现状、难点和作用。

方法 查阅文献分别介绍了图像、生理信号和基因组三方面目前大数据的发展和应用。例如图像方面采用 MapReduce 分析和转换数据，采用 iDASH 用于收集、压缩、共享和匿名化医疗数据；基

因组则采用生物大数据推断网络模型的技术；信号处理方面，采用监控技术实时获得生理设备的数据，使用 MongoDB 存储数据，用 Mapreduce 或者 Spark 做处理。

结果 利用大量不同、结构化和非结构化数据源的大数据分析对帮助医疗人员决策和患者的治疗有积极的促进而且目前实际中已经有所应用，但也有很多困难阻碍，例如医学图像量的指数级增长，穿戴式设备信号的采集、整合和存储，基因组尺度上重建网络是的不适定性。

结论 医疗数据具有复杂性的特点，但在此领域开发和实施大数据解决方案仍具有潜力和收益，随着大数据分析的深入应用，能够收集、整合、处理海量医疗数据，从而有机会改进和开发用于临床预测、处方和诊断的稳健 CDSS，协助探索和发现医疗问题、改善护理质量、帮助设计医疗健康政策。更重要的是，采用从大数据分析中获得的知识有可能挽救生命、改善医疗服务、扩宽医疗服务的形式，使支付与绩效保持一致，并有助于遏制医疗费用的急剧增长。

PU-053

医院图书情报服务中区块链技术的应用

刘明信、崔婷婷、张延一、王璞琳
聊城市人民医院

目的 区块链技术具有去中心化低成本、数据可溯源共享、防篡改可信任等优势，为提升医院图书馆服务效能提供一种新思路。

方法 文献调研法，分析方法与综合方法

结果 区块链技术可以在医院图书馆中健康存档、身份认证、跨区域协同、科学研究及个性化服务的五大应用场景中发挥作用。

结论 区块链技术逐步完善与成熟，区块链技术快速应用到医院图书馆进行创新和转型，对提升医院图书馆服务质量具有重大的意义。

自中本聪（Satoshi Nakamoto）[1]于 2008 年首次提出比特币及其底层支持技术区块链、2009 年创造出区块链的第一个区块“创世区块”以来，区块链技术逐渐走进大众视野，成为社会的关注焦点。据德勤发布的《2019 年全球区块链调查报告》，大多数受访者（53%）认为区块链技术将会是未来公司的战略发展重点。2019 年 10 月 24 日，习近平总书记在主持学习时强调，区块链技术作为自主创新的突破口，应用已延伸到教育、医疗健康等领域。

近年来，国内外已有很多将区块链应用到图书馆服务的案例。如 2017 年美国政府向美国博物馆和图书馆服务研究所拨款 10 万美元，用于探索区块链技术在公共图书馆数字管理方面的应用[2]；宫启生探讨了如何利用区块链技术创新解决当前制约高校图书馆机构知识库建设的问题[3]；赵伟娜提出了区块链技术下图书馆的流通平台、电子资源平台、阅读平台、元数据控制方式[4]；王庆探究了基于区块链技术的智慧图书馆创新服务策略[5]；陈小平进行了区块链技术在图书馆智慧服务中的

应用研究，挖掘管理体制、机构库建设、知识交易服务模式上的转变[6]；黄敏聪提出了基于区块链的分散型借还书服务模式，读者之间的借还书服务无需通过图书馆[7]等。就专业图书馆而言，医院图书馆作为医院的职能部门和服务窗口，加强图书馆的管理与服务能够更好为广大医务工作者以及住院患者提供更多的数据支持以及信息查询，Matthew Beyet Hoy 介绍了将区块链技术应用于医院图书馆[8]，Liu Xiaoguang 等介绍了一种基于区块链的电子健康档案数据共享和隐私保护方案[9]。

区块链技术的发展经历了数字货币的区块链 1.0 时代和智能合约可编程金融的区块链 2.0 时代，正处在各行业大规模应用的区块链 3.0 时代。由此可见，随着区块链技术逐步完善与成熟，研究区块链技术如何快速应用医院图书馆进行创新和转型，对提升医院图书馆服务质量具有重大的意义。

1 区块链技术特征分析

区块链是一种分布式记账技术，采用分布式节点共识算法进行数据生成和更新、采用块链式数据结构进行数据验证和存储、采用加密技术进行数据传输和访问、采用智能合约进行编程和操作数据的一种去中心化基础架构与分布式计算方式[10]。去中心化、共享信息以及智能合约三个特性并非区块链技术专有，区块链技术是目前唯一能将上述三个特性高效率融合的技术：

（1）点对点传输的去中心化数据库。基于点对点（Peer to Peer, P2P）网络的去中心化或分布式是区块链的最大特征。区块链网络是由分布在互联网上的数以万计的节点组成，不需要像银行一样的中心化机构或中介机构来保存数据，所有节点均存储了完整版总账，访问任一节点均能获得所有交易信息。在交易信息更新后，链上每个节点完成同步更新以实现去中心化。

（2）共享可信任的信息。区块链存储在多个节点组成的分布式网络系统上，对新生成的包含了全网交易数据的数据块进行加密，而新增加的信息需要通过其他节点验证，才可以依据时间顺序添加、永久存储（除非达到全网 51%算力）。所有节点均存储了全部区块链的副本，配合“时间戳”记录下全网历史交易，短时间内就可对全网历史交易记录追踪查询，所以其数据具有高可靠性和稳定性。

（3）执行可编码合约。区块链的智能合约机制，是一种公开透明且协商一致的计算机协议或算法，由程序自动判断触发条件并自动强制执行，整个过程不受双方主观意愿、违约惩罚等各种因素的影响，节点可以在“去信任”的情形下完成数据安全交换。

2 区块链技术在医院图书馆服务中的应用

医院图书馆面向医务人员和社会公众，拥有专业信息资源、文献数据库、领域技术人员等资源，提供生物医学、营养健康、预防保健、医疗咨询、查新查引等知识、技术服务。区块链技术能够助力医院图书馆实现服务创新和转型，进而优化医院健康管理服务。

2.1 电子健康档案

电子健康档案围绕居民个人健康，贯穿诊疗、体检、康复等整个生命过程，是满足居民健康保健管理和决策需求的科学记录[11]。在大数据时代，如何保护患者隐私权、提高数据安全性、解决医疗信息不对称等问题，成为新的研究课题。由于区块链具有加密、防篡改等功能，可保护患

者个人健康信息的安全，而且，由于存储于公共区块链而非某个医院或第三方机构内，患者通过私钥访问各自的健康信息也能改善医患矛盾。在通过区块链技术建立居民健康档案时，医务人员进行电子健康数据的录入与上传，患者能随意下载和上传交互信息。

2.2 数字身份认证

数字身份不仅包含出生信息、个体描述、生物特征等身份编码信息，也涉及多种属性的个人行为信息。区块链技术的特征，非常适合医院图书馆的用户（医务人员、患者等）数字身份认证，解决传统身份认证方式的部分弊端。如身在服务范围外的用户，图书馆可以通过区块链技术进行身份认证，授权他们不受 IP 范围限制开放内网数字资源访问、信息服务、电子书刊借阅等，及方便医生多点执业。

2.3 跨区域协同服务

协同区域内医院、图书馆、科研院所之间的关系，积极打造医学信息的地域性资源共享服务平台，保障医院图书平台资源共享的成功进行，可以让越来越多的医务人员和患者开展跨地区资源检索，享受到更有效的信息资源服务。在组建地区性医院图书馆共享资源区块链联盟（即联盟链）时，每个图书馆代表对应节点，通过 P2P 传播和智能合约的方式共享信息。如孙回等将查询逻辑分离技术与医疗联盟链技术紧密结合，满足了不同医院之间医疗数据共享的需求，保障了患者的数据隐私和医护人员查阅数据的权限[12]。虽然区块链可以防止数据篡改，但是其效率却有待提升，为此芦晓红等介绍了一种融合区块链技术的图书馆云服务平台，构建联盟链，元数据存入区块链中，方便联盟内成员调用数量庞大的数字资源[13]。

2.4 科研服务

科研服务是图书馆，特别是医院图书馆的重要服务内容之一。诸如论文发文、引证检索等服务是医院图书馆服务于科研人员的重要价值体现。随着区块链技术的发展，科研学术区块链化正在全球范围蔓延，论文区块链化后，论文与作者、机构、研究主题、资助资金、科研项目等信息构建了区块链的信息节点，全球范围内的机构、个人或者其他组织都可以通过区块链快速实时获知每个人每个机构的论文情况，论文一经录用或刊发即实现全链认证。而如果科研区块链化成为全球共识，将直接威胁查新查引等服务，需要顺应趋势拓展服务内容，或提升图书馆其他服务的价值。

2.5 个性化精准服务

只有明白图书馆用户的需求，想要阅读的书，才可以给予用户更贴切的服务。医院图书馆用户既包括有着医学专业理论知识、需要阅读专业期刊文献的医务人员，也包括对医学知识了解少、需要获取生物医学、健康保健等信息的用户。个性化精准服务以用户为中心，将区块链技术引入其中，可有效实现用户生产内容、用户定制化、用户主导化等中心思想，根据用户确定相应的授权或者合约，智能合约会自动提供个性化服务。如姚瑞卿等通过应用层的客户端记录用户账户信息，并利用区块链公链强大的互联性收集用户在其他平台的使用、浏览情况，准确还原用户画像，为用户提供直接、有价值的解决方案[14]。

3 结语

区块链技术具有数据可溯源共享等先天优势，能够用于医院图书馆的健康存档、身份认证、跨区域协同、科学研究及个性化服务等业务中，未来可落地的应用场景广阔。在将区块链技术应用到医院图书馆的具体过程中，需要更多地交互和结合大数据、物联网、人工智能、云计算和边缘计算等前沿信息技术，以改善传统图书馆的信息组织体系、知识存储体系、用户交互体系及技术应用体系等。

PU-054

基于聚类分析的疾病与基因关联关系研究

任雪菁、安新颖、范少萍
中国医学科学院医学信息研究所

目的 明确与糖尿病眼病相关联的基因，探索糖尿病眼病与基因多样性的关联关系，希望能提供糖尿病眼病的预防和治疗的参考依据，进而提高糖尿病患者的生活质量。

方法 从 PubMed 获取 2015-2017 年糖尿病眼病相关文献信息，使用 Pubtator 进行命名实体识别和信息抽取，构建疾病-基因关系对。使用 SPSS 进行聚类分析，使用 DAVID 进行基因富集分析，最后通过文献回溯来验证分析结果。

结果 提取糖尿病眼病相关基因 645 个，文献回溯有 451 个基因具有文献支持。通过聚类分析将这 645 个基因分为 6 类。通过基因富集分析，发现相关基因发生作用的可能生物过程和信号通路，富集结果与文献回溯结果一致。

结论 通过聚类分析较准确的找到了糖尿病眼病的相关基因，以及其作用可能的生物过程和信号通路，但仍存在一定的问题，需要进一步深入研究。

PU-055

基于 FDA 不良事件报告系统的二甲双胍不良事件信号研究

张旭
德昌县人民医院
四川大学华西第二医院

目的 基于 FDA 不良事件报告系统（FDA Adverse Event Reporting System，FAERS）数据库研究二甲双胍在真实世界中的药物不良事件信号，发现个系统中不良反应的特点，以期为临床安全用药提供参考。

方法 提取 FAERS 数据库 2010 年第一季度至 2019 年第四季度共 40 个季度的不良反应报告数据，检索并提取出以二甲双胍为首要怀疑药品的报告，筛除重复报告后采用比例失衡法中的 PRR 法和 ROR 法进行数据挖掘，得到有效信号，再删去药物无关不良反应（产品问题等），最后将筛除后的不良反应信号进行排序和系统归类。

结果 得以二甲双胍为首要怀疑药品的不良事件报告共 126607 例，挖掘后共得到二甲双胍的有效信号 853 个，报告数 82148 个，不良事件集中在代谢及营养类疾病、胃肠疾病系统和各类检查类等系统器官分类（SOC）。不良事件报告数较多的有乳酸酸中毒、急性肾损伤、腹泻、低血糖等。

结论 二甲双胍不良反应信号涉及广泛，尤其应注意对代谢及营养类疾病和胃肠疾病系统的监测。

随着人们饮食习惯的不断改变，糖尿病逐渐成为影响人们健康的常见病、常见病。目前，糖尿病已成为威胁人类健康和生命的疾病之一。糖尿病的主要临床表现有：暴饮暴食、多饮多尿、体质量减轻，即三多一少，同时出现视力减退、疲劳、抑郁等症状，严重威胁生命健康，要加强对患者的有效治疗[1]。新近国内外糖尿病指南[2-4]均建议，二甲双胍是治疗 2 型糖尿病（T2DM）的首选药物和联合治疗方案中的基础治疗药物，且应一直保留在糖尿病治疗方案中。 且其是目前全球应用最广泛的口服降糖药之一，为全球防控糖尿病的核心药物。二甲双胍具有良好的单药/联合治疗疗效和安全性证据、卫生经济学效益 证据、明确的预防心血管并发症等临床证据[5]。其应用于临床已有 60 余年历史，国内批准的适应症为首选用于单纯饮食及体育锻炼控制血糖无效的二型糖尿病，可单用，也可与磺酰脲类或胰岛素联用。此外，也用于预防 T2DM（国内未批准）、治疗多囊卵巢综合征伴无排卵性不孕，治疗多囊卵巢综合征伴月经不规则、预防卵巢过度刺激综合征伴多囊卵巢综合征[6-7]。

尽管二甲双胍应用应用已有 60 余年历史，但其常见的不良反应不容忽视，包括腹泻、恶心、呕吐、胃胀、乏力、消化不良、腹部不适及头痛等[5]。张味娜，熊亚丹，张罡等先后报道了二甲双胍致乳酸性酸中毒[8-10]。陈正言报道了口服二甲双胍缓释片致急性药物性肝损伤[11]。刘振杰等报道了中医药治疗服用二甲双胍而胃肠反应不耐受者 19 例[12]。《柳叶刀血液学杂志》于 2015 年 1 月第 2 卷第 1 期发布了一项关于使用二甲双胍与意义不明的单克隆免疫球蛋白病（MGUS）进展为多发性骨髓瘤关联的研究内容，指出：长期使用二甲双胍可降低意义不明的单克隆免疫球蛋白病进展为多发性骨髓瘤风险。美国 FDA 于 2016 年 4 月 8 日发布安全通报，含二甲双胍的抗糖尿病药物可用于肾功能下降患者。乳酸酸中毒是极少见的副作用。但是，由于病死率高，其仍是一个令人担忧的问题。二甲双胍诱导的乳酸酸中毒可发生于肾功能和肝功能正常的患者。

基于临床发现的一系列二甲双胍引起的药物不良反应报道，如何全面客观评估其不良反应，以期待在临床中更合理的使用，我们进行了二甲双胍不良反应的信号挖掘。

美国不良事件报告系统(FDA Adverse Events Reporting System，FAERS)是美国收集上市后药品不良反应的自发呈报系统，是药品安全性研究的重要数据来源。现多采用基于 FAERS 大数据的不良事件信号挖掘方法，应用于药品上市后安全性监测。因此我们将基于 FDA 不良反应报告系统（FAERS）对二甲双胍的药物不良反应（Adverse Reaction,，ADR）进行信号挖掘，以期为二甲双胍在临床的安全使用提供参考。

1.数据与方法

1.1 数据来源

本研究数据来源于 FAERS 数据库，该数据以 ASCII 或 XML 形式存储[13]，本研究选择了自 2010 年第一季度至 2019 年第四季度共 40 个季度的数据，将首要怀疑药物引起的不良事件共 126607 个，进行去重处理后，导入 MySQL 5.7 进行分析。选择个人信息记录（DEMO）、不良事件记录（REAC）、药物使用记录（DRUG）和事件结局（OUTC）数据表进行后续分析。

1.2 数据筛选

以 FDA 批准药品及不良反应公众数据库（public dashboard）中药品名称为标准，将二甲双胍的通用名“METFORMIN”和商品名“GLUCOPHAGE”、“FORTAMET ”、“GLUMETZA ”、“RIOMET”为关键检索词在系统中进行模糊查询，筛选出其中用药名称为二甲双胍且为首要怀疑药物的报告。报告数据包括人口资料、药物使用信息、不良事件名称、结果、报告来源、日期和原发疾病等内容。将报告的文本文件导入数据库后，删除重复和存疑数据，最终得到 37183 份报告共计 126607 个药物不良事件（Adverse Events ，ADE）作为研究原始数据。

1.3 数据处理

DE 的分类和表达采用国际医学科学组织理事会《国际医学用语词典》（Medical Dictionary for Regulatory Activities，MedDRA）中药物不良反应术语集的首选系统器官分类（system organ class，SOC）和首选术语（preferred term，PT）。在 MedDRA 官方网站（https://www.meddra.org）输入用户名与密码，下载并运行网站提供的 MedDRA Desktop Browsers，将需要汉化的 SOC 和 PT 英文名导入，则系统自动输出中英文映射表，本文中 SOC 和 PT 的中英文表达均按照该映射表。

1.4 数据分析

本研究采用比例失衡法（disproportionality methods）中的报告比值比法（reporting odds ratio，ROR）和比例报告比值法(proportional peporting ratio，PRR)进行信号检测[14]。该方法基于四格表 ，通过比较目标药品（IFN-α ）的目标事件发生比例与其他所有药品的目标事件的发生比例（背景数据），以检测潜在 ADE 信号。PRR 法报告数≥3，PRR≥2，PRR×2＞4 提示信号；ROR 法双侧检验 95%CI 下限＞1 提示信号[8]，即药品和事件之间潜在关联性。计数资料采用例数和构成比描述。该方法基于比例失衡法四格表（表 1），获取首要怀疑药物为二甲双胍的目标 ADE 报告数及 ADE 发生的背景数等数据，根据公式（表 2）计算 ROR 和 PRR 的值，满足阈值条件的

即生成一个信号[15]，有信号则提示 ADR 与二甲双胍之间存在统计学上的关联性。统计分析采用 SPSS 20.0 软件和 Microsoft Excel 2016 软件。

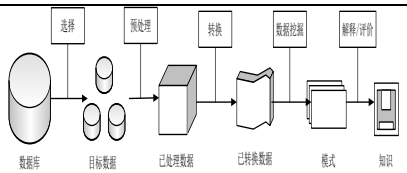
表 1 比例失衡法四格表

Tab 1 Fourfold table of Measures of Disproportionality

药品 (drug)	目标不良反应报告数 (the reported number of targeted adverse drug reactions)	其他不良反应报告数 (the reported number of other adverse drug reactions)	合计 (total)
目标药物	a	b	a+b
其他药物	c	d	c+d
合计	a+c	b+d	N=a+b+c+d

表 2 ROR 法与 PRR 法公式及阈值

Tab 2 Formulas and threshold values of ROR and PRR

方法 (means)	公式 (formula)	阈值 (The threshold value)
ROR		a≥3; ROR 的 95%Cl 下限>1，则提示生成一个信号
PRR	$PRR=\frac{a/(a+b)}{c/(c+d)}$ $SE(\ln PRR)=\sqrt{(\frac{1}{a}-\frac{1}{a+b}+\frac{1}{c}-\frac{1}{c+d})}$ $95\%Cl=e^{\ln(PRR)\pm1.96\sqrt{(\frac{1}{a}-\frac{1}{a+b}+\frac{1}{c}-\frac{1}{c+d})}}$	a≥3；PRR 值≥3； PRR 的 95%Cl 下限>1； x^2≥4； 则提示生成一个信号

2. 结果

2.1 **药品不良事件报告的基本情况**

2010 年至 2019 年的药物不良事件背景信号共 126007 个，其中首要怀疑药物为二甲双胍的 ADE 报告共 37183 份。报告中患者性别来看，男女比例为 0.77，女性占多数。在已知年龄的患者中，平均年龄为 47.06 岁，中位年龄为 50 岁。报告者大多数为其他健康专家及消费者，占 67.60%。报告国家前五位分别是美国、意大利、大不列颠及北爱尔兰联合王国、德国和法国，美国报告数占 20.75%。2010 年至 2019 年分别报告 5145、7907、7862、8644、9082、6208、13088、12594、29353、26410 例，大致呈下降趋势。报告基本信息见表 3。

表 3 2010-2019 年二甲双胍相关不良事件报告基本信息

Tab 3 Primary information of Metformin related events reported from 2010 to 2019

信息 (information)	分类 (category)	报告例数/N (case no.)	构成比/% (proportion)
性别	男性	10154	8.06
	女性	13177	10.46
	不清楚	103276	81.96
年龄段	＜18 岁	748	0.59
	18-64 岁	11779	9.35
	≥65 岁	12693	10.07
	不清楚	101387	80.46
报告者	其他健康专家	54484	43.03
	消费者	31106	24.57
	医师	28700	22.67
	药师	6598	5.21
	不清楚	5481	4.33
	护士	140	0.11
	律师	98	0.08
报告国家（前五）	美国	26275	20.75
	意大利	13718	10.84
	英国	9803	7.74
	德国	9523	7.52
	法国	8601	6.79
年份（2010-2019）	2010	5145	4.06
	2011	7907	6.25
	2012	7862	6.21
	2013	8644	6.83
	2014	9082	7.17
	2015	6208	4.90
	2016	13088	10.34
	2017	12594	9.95
	2018	29353	23.18
	2019	26410	20.86

	不清楚	314	0.25

2.2 各 SOC 不良事件报告及信号情况

对以二甲双胍为首要怀疑药品的不良反应信号进行筛除，经过 ROR 和 PPR 法筛选后最终获得二甲双胍信号数为 853，PT 报告总数为 82148。在有信号的 PT 中，报告数前 20 和以 PRR 和 ROR 值大小排序的前 20 位如表 4、表 5 所示。PRR 和 ROR 数值越大，信号越强[15]。较为常见的不良反应表现为乳酸酸中毒、急性肾损伤、腹泻、低血糖等；相关性较高的不良反应为皮肤褶皱测量、维生素 B12 吸收试验异常、大血管矫正性错位等，ADR 分散在各个系统。本研究还将检出的信号和说明书进行了对比，结果显示新的不良反应较多，为进一步研究新的不良反应提供可能。

表 4 报告数前 50 的首选术语

Tab 4 TOP 50 ADR signals of Metformin in descending order of signal reports

排序(No.)	PT	报告例数(case no.)	占比(proportion)	PRR	PRR x 2	95%CI 下限 PRR (lower limit)	ROR	95%CI 下限 ROR (lower limit)
1	乳酸酸中毒	8252	6.52	319.24	1197000.04	309.36	341.43	330.54
2	急性肾损伤*	5443	4.30	12.74	41714.11	12.35	13.12	12.71
3	腹泻	3431	2.71	2.65	3519.35	2.56	2.69	2.60
4	低血糖	2515	1.99	26.35	55899.81	25.30	26.86	25.77
5	各种制剂毒性*	2336	1.85	6.85	11404.58	6.57	6.96	6.67
6	呕吐	2164	1.71	2.23	1466.92	2.14	2.25	2.16
7	代谢性酸中毒	2115	1.67	41.02	71676.34	39.20	41.70	39.82
8	血葡萄糖升高	1969	1.56	5.23	6634.78	5.01	5.30	5.07
9	恶心	1882	1.49	1.15	38.47	1.10	1.16	1.10
11	低血压	1344	1.06	3.27	2096.32	3.10	3.29	3.12
12	肾衰*	1233	0.97	3.71	2412.38	3.51	3.74	3.53
13	高钾血症*	1111	0.88	15.87	14625.58	14.94	16.00	15.06
14	自杀既遂	1035	0.82	6.34	4556.32	5.96	6.39	6.00
15	有意过量用药*	999	0.79	9.00	6884.62	8.46	9.07	8.51
16	药物相互作用*	914	0.72	2.87	1104.84	2.69	2.88	2.70
17	脱水*	894	0.71	2.97	1160.35	2.78	2.99	2.80
18	用药过量*	873	0.69	2.67	908.69	2.50	2.69	2.51

19	腹痛	862	0.68	1.69	242.62	1.58	1.70	1.59
20	休克*	780	0.62	19.49	12761.47	18.13	19.60	18.23
21	妊娠过程中暴露*（10073513）	705	0.56	5.65	2645.68	5.25	5.68	5.27
22	体重降低	688	0.54	1.18	19.49	1.10	1.18	1.10
23	肾功能损害*	687	0.54	3.94	1485.37	3.65	3.95	3.67
24	血肌酐升高*	676	0.53	4.94	2086.51	4.58	4.96	4.59
25	糖尿病控制不佳	636	0.50	18.83	10039.82	17.38	18.92	17.46
26	心脏停搏*	633	0.50	3.66	1211.25	3.39	3.68	3.40
27	意识模糊状态*	630	0.50	1.87	255.04	1.73	1.88	1.74
28	自杀未遂	611	0.48	5.08	1965.21	4.69	5.10	4.70
29	糖化血红蛋白升高	593	0.47	12.93	6227.75	11.91	12.98	11.95
30	高血糖症	582	0.46	7.60	3245.04	7.00	7.63	7.02
31	妊娠过程中胎儿暴露*	560	0.44	3.35	914.50	3.09	3.36	3.09
32	上腹痛	549	0.43	1.26	29.74	1.16	1.26	1.16
33	腹部不适	524	0.41	1.53	96.39	1.41	1.53	1.41
34	药物水平升高*	521	0.41	19.78	8655.88	18.10	19.86	18.17
35	无尿*	483	0.38	27.41	11155.01	24.97	27.51	25.05
36	嗜睡*	477	0.38	1.16	10.50	1.06	1.16	1.06
37	胃肠疾病	440	0.35	3.01	585.09	2.74	3.02	2.75
38	血液透析*	435	0.34	31.53	11511.27	28.56	31.64	28.64
39	血葡萄糖降低	432	0.34	4.60	1197.53	4.18	4.61	4.19
40	体温过低*	426	0.34	21.58	7739.99	19.55	21.65	19.61
41	孕产妇在妊娠过程中的暴露*（10071408）	419	0.33	2.82	486.46	2.56	2.82	2.56
42	昏迷*	401	0.32	4.37	1025.68	3.96	4.38	3.97
43	高乳酸血症	382	0.30	138.53	34416.15	122.46	138.94	122.79
44	意识丧失*	378	0.30	1.40	42.85	1.26	1.40	1.27
45	糖尿病酮症酸中毒	377	0.30	7.51	2072.14	6.78	7.53	6.80
46	急性胰腺炎	363	0.29	7.74	2072.61	6.98	7.76	6.99
47	脓毒症*	349	0.28	1.39	38.73	1.25	1.40	1.26
48	类天疱疮*	345	0.27	29.19	8472.91	26.12	29.27	26.19
49	呼吸衰竭*	343	0.27	2.15	210.72	1.94	2.16	1.94

50	多汗	331	0.26	1.21	12.11	1.09	1.21	1.09

表 5 信号强度前 20 的首选术语

Tab 5 TOP 20 ADR signals of Metformin in descending order of signal strength

排序(No.)	pt	报告例数(case no.)	占比(proportion)	PRR	PRR x 2	95%CI 下限 PRR (lower limit)	ROR	95%CI 下限 ROR (lower limit)
1	皮肤褶皱测量	8	0.01	2149.73	1909.09	268.87	2149.87	268.88
2	维生素 B12 吸收试验异常	5	0.00	1343.58	1117.99	156.97	1343.63	156.97
3	大血管矫正性错位*	9	0.01	806.15	1809.34	218.24	806.21	218.25
4	碱剩余阴性*	14	0.01	627.00	2625.03	240.95	627.07	240.97
5	尿白蛋白/肌酐比率异常*	33	0.03	521.63	5830.23	290.57	521.76	290.63
6	先天性乳头异常*	7	0.01	376.20	1091.43	119.40	376.22	119.40
7	原发性家族性低镁血症*	4	0.00	358.29	610.78	80.19	358.30	80.19
8	凝血酶原消耗时间延长*	10	0.01	335.90	1483.99	132.57	335.92	132.58
9	乳酸酸中毒	8252	6.52	319.24	1197000.04	309.36	341.43	330.54
10	先天性卵巢异常*	3	0.00	268.72	400.08	54.24	268.72	54.24
11	胎位不稳*	3	0.00	268.72	400.08	54.24	268.72	54.24
12	碱过剩降低*	80	0.06	249.97	10277.47	184.37	250.13	184.47
13	碱剩余*	11	0.01	173.88	1147.92	81.45	173.89	81.45
14	血乳酸	30	0.02	167.95	3063.77	106.43	167.99	106.45
15	套细胞淋巴瘤 III 期*	3	0.00	161.23	298.57	38.53	161.23	38.53
16	黄瘤病*	3	0.00	161.23	298.57	38.53	161.23	38.53
17	高乳酸血症	382	0.30	138.53	34416.15	122.46	138.94	122.79
18	广泛性特发性骨质增生*	3	0.00	134.36	264.73	33.60	134.36	33.60
19	成人隐匿性自身免疫性糖尿病*	31	0.02	134.36	2735.56	87.30	134.39	87.32

20	温度觉测试异常*	4	0.00	134.36	352.97	40.46	134.36	40.46

注：表中标有“*”为说明书中未提及的不良反应

1.3 各系统器官分类不良事件报告及信号分类

根据 MedDRA 的系统器官分类（System Organ Class, SOC）对有信号的 PT 进行 SOC 分类排序，有信号的 SOC 仅有 27 个，表 6 为分类后各个 SOC 的信号数和报告数情况。82148 份报告中，“代谢及营养类疾病”SOC 的 ADR 报告数最多，共 19571 份，占 23.82%，其中乳酸酸中毒报告数最多，共 8252 份，占总报告数的 10.05%。除此以外，报告数较多的 SOC 还有胃肠疾病系统（腹泻等，14.04%）、各类检查（血葡萄糖升高等，12.31%）、肾脏及泌尿系统疾病（急性肾损伤等，10.90%）。853 个信号中，代谢及营养类疾病、各类检查类 SOC 中的 ADR 信号数最多。

表 6 SOC 项下二甲双胍 ADR 信号检测

Tab 6 ADR signals of Metformin involved systems and organs

器官系统分类(SOC)	信号数(PT)	不良事件报告数(ADE)	报告数构成比（%）
代谢及营养类疾病	62	19571	23.82
胃肠系统疾病	51	11532	14.04
各类检查	141	10114	12.31
肾脏及泌尿系统疾病	40	8952	10.90
各类损伤、中毒及手术并发症	43	7191	8.75
各类神经系统疾病	63	3515	4.28
精神病类	17	2984	3.63
血管与淋巴管类疾病	15	2748	3.35
心脏器官疾病	38	2669	3.25
全身性疾病及给药部位各种反应	25	2389	2.91
肝胆系统疾病	36	1491	1.82
妊娠期、产褥期及围产期状况	47	1381	1.68
呼吸系统、胸及纵隔疾病	25	1356	1.65
感染及侵染类疾病	34	1138	1.39
皮肤及皮下组织类疾病	26	1126	1.37
各种手术及医疗操作	14	1047	1.27
各种先天性家族性遗传性疾病	77	786	0.96
血液及淋巴系统疾病	24	714	0.87
产品问题	10	552	0.67
眼器官疾病	14	274	0.33
各种肌肉骨骼及结缔组织疾病	7	230	0.28

良性、恶性及性质不明的肿瘤	22	195	0.24
生殖系统及乳腺疾病	10	61	0.07
内分泌系统疾病	5	57	0.07
免疫系统疾病	3	34	0.04
耳及迷路类疾病	3	29	0.04
社会环境	1	12	0.01

3 讨论

二甲双胍是为双胍类药，可减少肝脏葡萄糖的生成量及肠道对葡萄糖的吸收，并通过增加外周葡萄糖的摄取和利用提高胰岛素的敏感性，改善 T2DM 患者的葡萄糖耐量，降低基础和餐后血糖。

3.1 不良反应发生的人群特点

本研究共纳入以二甲双胍为首要怀疑药品的 ADE 共 126607 份报告。主要来自美国（20.75%）和意大利（10.84%），英国、德国、法国差距不大（6%~7%）因此 ADE 的发生数据更多集中于这几个国家，其他地域和种族的药品不良反应可能有所差别。患者性别上差异不大，在使用量大致相等的前提下，二甲双胍 ADR 的发生可能没有性别特异性，目前还没有性别相关研究，仍需进一步证明。值得一提的是，大部分病例年龄偏大，18 岁以下的不良反应报告仅占 0.59%，但患者出现糖代谢问题即发病用药也可能不是在 18 岁以前。本次研究的结果给 18 岁以下患者的不良反应提供了更多研究资料，提示其使用二甲双胍的安全性可能较好。由于尚未获得二甲双胍各年龄和不同性别群体的使用量数据，因此难以确定不良反应发生率的人群差异，目前仅反映 ADE 的构成比数据而无发生率数据。从 2010 年到 2019 年，数据的占比逐年上涨，可能随着医学技术的发展，以及大众对健康的认知改变，更多的患者被诊断。

3.2 不良反应发生的器官及系统特点

二甲双胍报告数多的 ADR，如恶心、呕吐、腹泻、腹痛、食欲差等，符合双胍类常见的不良反应。说明书中提及到的可能引起严重不良反应，如乳酸性酸中毒，主要表现为嗜睡、肌肉痛、肌肉痉挛、呼吸困难、腹痛、衰弱和体温降低，进而昏迷，基本在不良反应信号中均有所体现。本研究结果提示，二甲双胍的 ADR 集中在代谢及营养类疾病（23.82%），胃肠系统疾病（14.04%）和各类检查（12.31%）三个 SOC。糖尿病肾病可发生于 1 型糖尿病、2 型糖尿病，以及其他继发性糖尿病中。糖尿病肾病患者有 4 种主要的肾小球组织学改变：系膜扩张、肾小球基底膜增厚、足细胞损伤和肾小球硬化[15,16]。除此包括代谢及营养类疾病和胃肠疾病系统不良反应较多，可能是均由于原发疾病的影响，较容易出现腹痛、恶心、食欲差等不良反应，需要进一步确定其与药物使用的关系。

3.3 特殊人群中的安全性

二甲双胍不推荐 10 岁以下儿童使用。老年人使用本药更易发生低血糖，且患者本身可能出现肾功能减退，应定期检查肾功能，并根据肾功能调节本药的剂量。65 岁以上的老人是慎用。

在所有 SOC 中均有响应信号，提示该药的安全性相对较低。虽然报告的代谢及营养类疾病和胃肠疾病系统不良反应较多。二甲双胍主要不良反应是胃肠道反应，多出现在治疗后 10 周。随着治疗时间延长，大多数患者可逐渐耐受或症状消失。小剂量起始，逐渐加量，适时调整剂量，非缓释制剂分次随餐服用，或改成 1 次/d 的缓释制剂，可减少胃肠道反应二甲双胍本身无肝、肾毒性,目前尚无确切证据证明二甲双胍与乳酸酸中毒有关。在掌握好禁忌证的前提下，长期应用二甲双胍不增加乳酸酸中毒风险。二甲双胍具有明确的心血管保护作用，其可减少新诊断及已经存在心血管疾病的 T2DM 患者的心血管事件发生风险。

3.4 药物相互作用

本研究中未挖掘到药物相互作用相关信号，但二甲双胍与多种药物联合使用时仍可能产生不良反应。其普通片口服后的绝对生物利用度为 50%-60%，达峰时间为 2.5 小时；缓释片和缓释胶囊平均达峰时间为 7 小时。本药几乎不与血浆蛋白结合，部分进入红细胞。以原型随尿液排出，肾脏清除率大于 400mL/min，终末血浆消除半衰期约为 3.6 小时。其与碳酸酐酶抑制药联用可增加发生代谢性酸中毒的风险，合用时应谨慎。与西米替丁联用可增加本药的血浆和全血曲线下面积。与磺酰脲类联用，可增加发生低血糖的风险，但与格列苯脲的联用，会减少格列苯脲的血药峰浓度和 AUC，但具有高度变异性，临床意义尚不明确。与呋塞米联用，增加本药 AUC，但本药的肾清除无变化，呋塞米的最大血药浓度与 AUC 均减少，终末半衰期缩短，肾清除无变化。与硝苯地平联用，增加本药的最大血药浓度与 AUC，且本药的尿排泄量增加，达峰时间和半衰期无变化。与华法林联用，可能增加华法林的抗凝作用。经肾小管排泌的阳离子药物联用，理论上可能与本药在肾小管竞争转运，建议合用时密切监测，调整本药和（或）此类药物的剂量。与可导致高血糖症的药物（噻嗪类或其他利尿药、皮质类固醇、口服避孕药、苯妥英、烟酸、拟交感神经药、钙离子通道阻滞药、异烟肼）联用，可能导致血糖控制失败，合用时应密切观察是否出现血糖控制失败，停用此类药物后，应密切监测是否出现低血糖症。与树脂类药联用，减少本药的吸收。与血浆蛋白高度结合的药物联用（水杨酸类药、磺胺类药、氯霉素、丙磺舒），基本不发生相互作用[17]。

3.5 研究局限

本研究仍存在一定的局限性。由于 FAERS 是自发呈报系统，并不能收集二甲双胍不良反应的全部信息，存在漏报的可能性，而且呈报的不良事件和药品之间因果关系未进行考量，同时报告数据可能存在模糊、不规范等问题。此外，由于药物剂量根据个人的血糖水平进行调整，而无法获得二甲双胍使用量的信息，并不能计算不良事件在总体中的发生率，仅能获得构成比而无法获知发生率。尽管提高了信号探查的阈值，但也不能完全排除假阳性的信号，同时还可能漏查部分信号。最后，对于信号挖掘结果，仅表示不良反应与药物存在统计学上的关联，为药物的安全性提出假说，最终结果仍需要进一步研究证实。

4 结论

对服用二甲双胍的患者，我们每 3-6 个月检测 1 次 A1C，每年检测 1 次血清肌酐，并且在诊断时以及发生提示贫血、神经病变或肾功能恶化的症状时检测血红蛋白、血细胞比容和红细胞指数。本次研究对 FAERS 数据库 2010 年至 2019 年非索非那定的不良反应报告进行了较为全面的分析。结果显示，监测到二甲双胍在 27 个系统器官有 ADE 信号，为临床安全用药提供保障。在用药过程中要注意对代谢及营养类疾病和胃肠疾病系统进行不良反应监测。

PU-056

基于大数据的新型冠状病毒肺炎疫情防控策略的探讨

秦三利[1]、齐江华[2]、张仲男[1]、石昕[1]
1. 甘肃省医学科学研究院
2. 甘肃省肿瘤医院

目前，新型冠状病毒肺炎疫情在全球蔓延，严重威胁人类生命。每个疫情防控部门都应详细了解疫情病毒的基本情况、传播特征、防治方法等，做好相应防控措施，控制疫情进一步蔓延扩散。本文就新型冠状病毒-19（COVID-19）基本特征、致病机理、传播途径、疾病主要临床症状、实验室检测指标、主要治疗药物以及基于大数据的新型防控措施做简单论述，为更好地防控新型冠状病毒肺炎疫情进展提供参考。

1 引言

2019 年 12 月，湖北省武汉市发现了一种尚未在人类中发现过的新型冠状病毒，命名新型冠状病毒-19（COVID-19）。这种新型冠状病毒可引起严重肺炎疫情，并且于 2020 年 1 月 30 日世界卫生组织（WHO）宣布此次疫情构成国际关注的突发公共卫生事件[1]。

新型冠状病毒疫情疫情在全球蔓延，形势较为严峻，截止 2020 年 7 月 10 日，全球已有新型冠状病毒肺炎疫情国家 212 个，现有确诊 4654871 人，累计确诊 12403597 人，死亡 557749 人，治愈 7190977 人（国外数据来源：WHO、各国官方通报和权威媒体报道）；中国现有确诊病例 531 人，境外输入确诊 1962 人，累计确诊 85447 人，死亡 4648 人，治愈 80268 人（国内数据来源：根据国家卫健委、各省市区政府卫健委，港澳台和海外权威官方渠道和媒体的公开信息整理而成）[2]。目前，国内疫情相对稳定，但境外及输入性病例仍然较多，因此，了解新型冠状病毒肺炎基本情况以及如何做好相应防控措施，阻止疫情进展尤为重要。

2 新型冠状病毒肺炎疫情概况

2.1 新型冠状病毒COVID-19 的基本特征及致病机理

冠状病毒是针对人类呼吸系统的主要病原体之一，先前的冠状病毒暴发主要包括严重急性呼吸系统综合症（SARS-CoV）和中东呼吸系统综合症（MERS-CoV）。2019 年武汉暴发的肺炎疫情已证实为COVID-19 感染所致，此次感染与当地海鲜和野生动物批发市场有关，可实现人际传播[1]。

COVID-19 属于β属冠状病毒，有包膜，颗粒呈圆形或椭圆形，常为多形性，直径 60～140nm[3]。系统发育分析结果显示：COVID-19 的遗传序列与SARS-CoV的同源性超过 80%，与MERS-CoV的同源性超过 50%，最有可能从蝙蝠起源的冠状病毒发展而来[4]。COVID-19 是一个很长的单股RNA(+)基因组，长度为 29 kb，其刺突蛋白受体结合域与血管紧张素转化酶 2（ACE2）受体结合而发生人际感染，更重要的是，COVID-19 受体结合域的序列与SARS-CoV序列相似，提示COVID-19 最有可能通过 ACE2 受体进入宿主细胞[5,6]。

2.2 新型冠状病毒肺炎的主要临床症状及实验室检测指标

COVID-19 通常在经过约 5.2 天的潜伏期后出现相关临床感染症状。不同患者的年龄和免疫状态不同，从出现症状到死亡为 6-41 天，中位时间 14 天[7]。感染COVID-19 时常见临床症状是发烧、干咳、喘、疲劳等，有些患者可出现肌肉酸痛、咯血、腹泻、淋巴细胞减少等症状[7,8]。患者胸部CT扫描显示为肺炎，但常伴有急性呼吸窘迫综合征、急性心脏损伤和毛玻璃样变等容易引起死亡的异常特征[8]。在两肺胸膜下同时观察到多个外周毛玻璃样变时，很可能会引起全身性和局部性免疫反应，导致炎症反应加重。

研究显示，感染COVID-19 的患者体内血液中白细胞数量、细胞因子和趋化因子水平明显升高；对ICU的一些患者观察显示：IL2、IL7、IL10、G-CSF、IP10、MCP1、MIP1α及TNFα等促炎性细胞因子较高，这与疾病严重程度相关[8]。

2.3 新型冠状病毒肺炎的主要传播途径及消毒措施

目前已经确定的COVID-19 的感染途径主要是呼吸道飞沫传播和接触传播，前者主要是近距离接触经患者，通过直接吸入患者喷嚏、咳嗽、说话的飞沫而感染；后者主要是通过接触沉积在物品表面的飞沫后，再接触口腔、鼻腔、眼睛等粘膜导致感染[9]。

现有的对冠状病毒理化特性的认识多来自对SARS-CoV和MERS-CoV的研究，认为病毒对紫外线和热敏感，乙醚、75%乙醇、含氯消毒剂、过氧乙酸及氯仿等脂溶剂均可有效灭活冠状病毒[9]。

2.4 新型冠状病毒肺炎的主要治疗药物

目前，还没有针对抗COVID-19 感染的特异性药物或疫苗，临床上主要使用奥司他韦、更昔洛韦、洛匹那韦、利托那韦及等广谱抗病毒药物进行治疗[10]。有报告显示，广谱抗病毒药物瑞姆昔韦和氯喹在体外可有效控制COVID-19 感染，可考虑将其用于临床治疗COVID-19 感染[11]。但目前仍然迫切需要进行更多研究，以研发并鉴定敏感性或特异性的用于治疗COVID-19 感染的新型药物。为此，也迫切需要建立一种动物模型以复制目前在人类中观察到的这种严重的疾病，以进行动物实验研究。此外，还需要建立快速追踪并评价新型药物疗效的方法，用于测试潜在的新型化合物和疫苗的治疗效果[12]。

3 新型冠状病毒肺炎疫情的防控策略

3.1 基于大数据的新型冠状病毒肺炎疫情的监测

目前，全球新冠肺炎疫情发展迅速，平均每天都新增十几万病例，国内疫情控制相对稳定，但平均每天仍新增病例几十例，多为境外输入性病例，因此建议所有国家都应作好常态化控制疫

情的准备，增强防护、主动监测、早期发现、严格隔离、积极追踪密切接触者，防止新冠病毒肺炎进一步传播，并与世卫组织积极合作，共享全部相关疫情数据。

医院作为抗击新型冠状病毒肺炎的主战场，是有效控制疫情蔓延的关键环节。在对疫情全面分析的基础上，有些医院结合自身实际情况，对全院所有工作人员进行网格化管理，搭建信息化平台，要求各科室工作人员严格执行防护措施，并由科室负责人每日监测身体健康状况并统一上报主管部门；特别安排专人动态追踪从外地返岗、体温异常及有咳嗽、胸闷、乏力、头痛、腹泻等不适症状人员，并要求其做到及早隔离，同时做好进一步处置预案，逐步制定出一套全覆盖、网格化、信息化健康监测管理策略，提升医院人员管理的效率和质量[13]。同时，医院也通过对就诊患者和家属进行体温监测，并要求其提供健康二维码和出行码，以及早发现可疑患者，并对其进行处理。

同时，随着互联网+的推广，利用互联网数字化医疗进行新型冠状病毒肺炎疫情的防控取得一定成绩和效果。

相较2003年严重急性呼吸综合征(SARS)的情况，现在的医疗资源和处置水平已大有进步，但面对我国庞大的人口基数和存在地区差异的线下医疗资源分布情况，传统线下公共卫生和医疗健康服务体系依然存在着短板和不足之处。互联网数字化医疗借助其高效、便捷、普惠等方面的优势，紧密结合国家防控工作，运用互联网+、大数据、人工智能等技术，快速投入到战“疫”中，以互联网医院、互联网诊疗等形式构建出线上线下一体化的公共卫生和医疗健康服务模式，降低了医疗机构线下交叉感染的风险和医务工作者的劳动强度，减少了医务工作者的工作压力，为患者提供及时有效的诊疗咨询服务，并能保障医疗信息的安全和准确，为疫情防控提供了有效的支撑[14]。

另有研究基于2020.1.20-2020.2.25期间中央和各省区市以及地市州官方疫情指挥部以及各类官方媒体、自媒体都在发布、释放与疫情相关的各种各样海量信息（全国各省区市网络直报的新型冠状病毒疫情大数据），选用累计确诊人数、累计死亡人数、累计治愈人数作为特征变量，建立了深度学习的长短期记忆模型，通过PYTHON程序实现了模型高精度的拟合和对疫情的动态监测，严防新冠肺炎疫情拐点的反复，确保早日取得新冠肺炎疫情阻击战的全面胜利[15]。

3.2 基于信息化平台及大数据的新型冠状病毒肺炎疫情的预警系统

研究发现，构建并使用基于信息化平台的远程移动中央综合监护与自动预警系统可实时自动监测并评估COVID-19患者病情，第一时间将重要病情变化通知相关医护人员，使患者第一时间获得有效处置，还可以进行异地远程会诊，对患者的远程实时监护与精准施治发挥了重要作用[16]。穆亚梅设计了基于ZigBee无线传感网络监控疫情预警系统，主要通过温度传感器对人体体温进行定点、连续或随机测量，并将数据发送到接收中心进行数据分析，能在第一时间获得有效数据，防止人口密集群体的大规模爆发，并且能降低防疫人员的测量风险，有效控制疫情发展[17]。Ping Yang通过对133例COVID-19患者的临床资料进行回顾性分析，调查COVID-19患者的危险因素，构

建了基于COVID-19危险因素的预警模型，利用该模型，可及早准确地识别危重COVID-19患者，并对其进行早期干预，可有效防止严重COVID-19的发生[18]。

自2019年12月以来，武汉暴发的COVID-19疫情由于春节人口流动快速蔓延。自2020年1月23日起全国大范围实施围堵缓疫策略，并不断提高病毒检测和检出率，有效地抑制了疫情快速蔓延的趋势。在COVID-19爆发的早期，如何利用数学模型并结合少量和实时更新的多源数据，对疫情进行风险分析，评估防控策略的有效性和时效性等具有非常重要的现实意义。有研究结合前期研究基础，系统介绍如何依据疫情发展的不同阶段和数据的完善，逐步建立符合我国防控策略的COVID-19 传播动力学模型，给出模型由自治到非自治，风险评估指标由基本再生数到有效再生数，疫情发展与评估由早期的SEIHR传播动力学决定到最终取决于隔离人群和疑似人群规模的演变等的重要研究思路[19]。

3.3 基于新型冠状病毒肺炎疫情应急管理体系的构建及实践探索

新型冠状病毒肺炎疫情暴发以来，一些疫情防控定点医院、航空、铁路部门等围绕《突发公共卫生应急处置预案》，构建了相对完善的应急管理体系，并积极地应用于实践。如浙江大学医学院附属第一医院针对疫情情况，迅速成立了COVID-19专项应急指挥系统，通过双融合、双联动的战略路径，从源头上确保组织全覆盖、学科全融合、管理全过程，有序开展COVID-19患者救治工作，并提出了进一步加强和完善大型医院应对突发重大公共卫生事件应急机制的相关建议[20]。南京大学医学院附属鼓楼医院也建立了应对突发公共卫生事件的应急管理体系，包含组织指挥体系、专家救治体系、培训防控体系、综合管理体系、人员调度体系、物质保障体系、信息保障体系、关怀激励体系、科研攻关体系、宣教引导体系等10个方面[21]。许东松从航空应急救援基地建设、航空应急救援装备建设、航空应急救援队伍建设、航空应急救援支持保障体系建设几个方面探讨了航空应急救援体系建设的方向[22]。铁路防疫应急体系是国家防疫应急体系的重要组成部分，既为国家防疫提供运输保障，也是控制疫情扩散的重要环节，潘键从新冠肺炎疫情防控方面思考了铁路防疫应急管理体系建设的对策[23]。

3.4 基于医疗大数据，整合资源，多部门联防联控共同应对新型冠状病毒肺炎疫情

有研究通过健全突发公共事件紧急医疗救援指挥体系，建立省、市（州）紧急医疗救援中心及急救网络，制定和完善紧急医疗救援预案，加强应急救援队伍建设管理及培训和演练，重视应急医疗救援设备装备和救援物资储备，与当地政府、公安、消防、军队等多系统密切协调合作，急诊医学专家指导伤员的分类急救和重症监护治疗，可以较好地完成突发公共事件的紧急医疗紧急救援，提高救治成功率[24]。桐乡市采用志愿者模式整合社会救援力量，组建专业化要求、规范化管理的高素质应急救援队伍，以街道人防办为依托，构建应急救援机构、应急指挥网络、应急预案保障全覆盖的应急救援体系，完善涵盖意外伤害保险、应急救援经费及应急救援演练的全方位应急保障机制，有效提升了应急救援队伍的整体应急救援能力[25]。

4 结束语

目前新型冠状病毒肺炎疫情在全球蔓延，情况不容乐观，如何有效预防控制疫情刻不容缓。在抗击新型冠状病毒肺炎疫情的关键时刻，全面整合各地区官方通报和主流权威媒体报道的信息数据，依托全民健康信息平台，动态监测疫情情况，动用并整合各方资源，高效实现各部门协作，常态化防控疫情，最大限度控制肺炎疫情进展是各级单位的基本职责。基于大数据的新型冠状病毒肺炎相关研究，可有效为各个相关部门提供必要的帮助，提高疫情防控能力。

PU-057

5G+智慧医疗应用场景探索与研究

陈军
郑州大学第一附属医院

目的 结合 5G 技术在智慧医疗领域的应用场景进行探索与研究，为智慧医疗发展提供参考。

方法 结合各种参考文献，以及我院目前已经实施的智慧医疗系统，对 5G 技术在智慧医疗领域的应用场景进行探索与研究，为智慧医疗发展提供参考。

结果 随着科技水平的发展，医疗行业也跟随着它的步伐努力的向现代化、智能化的方向进行发展，借助 5G 技术的助力，智慧医疗也最终将会进入医疗资源互联互通的新时代，不仅改变传统的医疗方式，也能进一步的提高医疗质量、提高医疗服务效率、降低医护人员工作强度、合理分配医疗资源，为患者带来便利和实惠。

结论 随着科技水平的发展，医疗行业也跟随着它的步伐努力的向现代化、智能化的方向进行发展，借助 5G 技术的助力，智慧医疗也最终将会进入医疗资源互联互通的新时代，不仅改变传统的医疗方式，也能进一步的提高医疗质量、提高医疗服务效率、降低医护人员工作强度、合理分配医疗资源，为患者带来便利和实惠。

PU-058

数据挖掘在病案信息中应用

王春杰
郑州大学第一附属医院

目的 本文分析了医院当前病案信息管理系统中数据的特点，将目前病案信息系统现状大体分为三类：病案信息化系统不完善、病案信息化系统非常完善和病案信息化系统比较完善三种情况，然后针对病案信息系统非常完善的情况介绍了病案数据数据挖掘的部分理论和数据挖掘中数据仓库的设计、ETL（Extract-Load-Transform，即抽取、转置、加载）抽取和决策分析等的设计，介绍

了一下各部分的设计内容及方法，并设计构造了一个病案信息系统的数据挖掘流程，并根据流程构造一个医院病案信息系统数据挖掘的模型。

方法 本文以近二十余年来某大型医院病案信息系统中的数据为基础，基于病案数据的挖掘，构建数据挖掘模块进行分析研究，通过介绍几种典型常用的数据挖掘算法，根据这几种数据挖掘算法的特点，选取关联规则的算法来做为本系统数据挖掘的算法， 数据挖掘首先从资料集合中找出所有的高频项目组，然后再由这些高频项目组中产生的关联规则，并以患者病种统计查询分析为例，介绍了如何使用关联规则数据挖掘技术，根据实际应用需求进行系统设计，对患者的所在地、费用、登记时间等信息进行抽取,

结果 分析了病案系统中患者病种结构，并简单列举了如何利用数据挖掘系统进行相关的分析研究，最终以图形的方式展现出来，从而对管理者提供有力的决策支持。

结论 分析了病案系统中患者病种结构，并简单列举了如何利用数据挖掘系统进行相关的分析研究，最终以图形的方式展现出来，从而对管理者提供有力的决策支持。

一、引 言

近年来，随着信息技术的发展，我国各大医院信息化的也进入了高速发展时期，医护人员对信息化的依赖程度也越来越高，医院各个信息化子系统也越来越多，各系统所产生的医疗数据也越来越多，这为后面的数据挖掘分析提供了良好的数据基础。数据挖掘是一个跨多个领域的交叉学科，受数据库系统、统计、人工智能、信息技术、临床应用等多个学科影响。数据挖掘是为临床应用和生产实践服务的，数据分析的目的是根据医院的业务模式，通过对数据分析发现目前业务模式中的瓶颈，提出改进优化方式和流程。如何充分利用医院现有多年沉淀的各种数据，搭建合理先进的数据挖掘平台，成为摆在医院管理者面前的主要问题。

基于病案数据的挖掘分析系统是为了服务临床、服务管理决策而开发的。需要加强各个临床业务应用系统的分析工作。比如，病案数据挖掘模块，主要功能是根据现有数据协助服务管理决策，将数据挖掘技术引入病案管理系统，通过分析出院患者费用及医保支付方式的构成比例,可以全面剖析出医院住院收入的的主要构成；通过分析门诊病历数据中挂号就诊时间，从中分析出门诊病人的就诊流量集中时间段；通过分析出院患者病历中患者家庭地址所在地，从中分析出某个疾病的地区分布；通过分析出院患者病种结构，可以分析出院主要收治病种结构，这样一来，就可以改进和加强诊疗流程、人员配置、资金统筹等方面的工作。

二、病案数据的现状和价值

病案数据分为广义病案数据和狭义病案数据。无论狭义的病案还是广义的病案，如果医院的病案数据不同，产生的病案数据质量也完全不一样。目前病案数据的现状总体分为三类：

信息化系统完善，希望利用数据来提高医院的运营、管理。这类医院规模化、信息化程度相对较高，病案数据质量相对较高。他们一般分布在各个省会城市，经济比较发达的地区。医院医疗质量控制相对较高，信息化程度完善，人员比较重视病案基本数据内容的获取和整理。

信息化较为完善，管理层不重视病案、统计的管理。病案数据质量一般，缺项比较严重。病案数据质量很多医院并不重视，只能满足国家规定的基本的统计报表。这类医院一般在各地级市，仅仅满足了国家政策性要求，信息化程度较高，但管理理念跟不上，对数据的认知程度比较保守。

信息化程度一般，信息化管理只是局限于费用管理、医嘱管理、药品管理部分。针对综合管理与统计分析，数据量不是很大，沉淀抽取的数据意义也不是很大。这类医院一般为县级医院以及非三甲类医院。

针对病案数据挖掘，本文主要指广义病案的数据挖掘。它适用于信息化系统非常完善，已经沉淀了大量的数据信息，并且管理人员希望利用数据来提高医院的运营管理。

二、基于病案系统的数据挖掘

根据病案系统的特点与业务需求，我们基于病案管理系统的数据挖掘流程，设计构造一个病案系统的数据挖掘流程，如下所示:

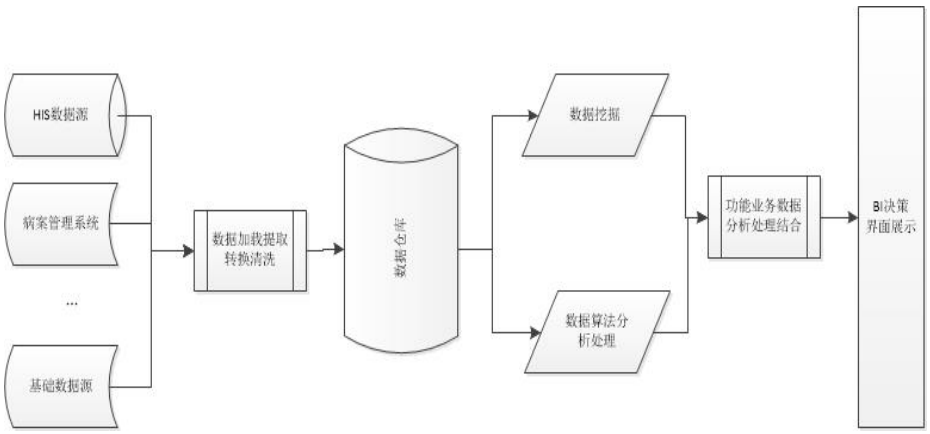


图 1 基于病案管理系统的数据挖掘流程图

(1)数据仓库设计

该系统模块的实现由多个数据仓库设计组成， 数据仓库表结构的设计是很重要的，一个系统能否正常高效的运行同数据仓库的设计有很大的关系。设计数据仓库，要注意的就是和实际相联系，设计时要充分考虑到用户使用和实现功能时的效率。

(2) ETL 数据抽取转换分析设计

完成数据仓库的设计，重要的是完成数据从各个异构系统中提取的过程。通过事先定义好的格式规则，将患者的基本信息、费用信息、结算信息、科室信息、诊断信息等分析数据，经过清洗、转换等过程，变成统一的、标准的数据格式，储存在数据仓库当中。

(3)决策分析展现界面设计

数据挖掘的用户展示界面是设计需求的最终部分，它展现的好与坏是和临床应用体验息息相关的。所以，必须实现好系统界面设计这一板块。在系统开发之初首先要进行的是页面的评审，结合页面详细介绍一下用户界面设计的流程。该项目主要涉及到的用户界面包括用户角色管理界面、患者费用分布比例界面、门诊挂号分布情况界面、地区就诊分布界面、收入分析界面、住院证开立时间分析界面等。

三、数据挖掘的流程

数据挖掘的流程对整个数据挖掘的设计起着至关重要的作用，虽然针对不同的行业特点和需求，数据挖掘的流程也不尽相同，但比较公认的是由 Fayyad 等人提出的数据库知识发现（Knowledge Discovery in Databases ，KDD）过程。它是从数据集中识别出新颖的、有效的、有潜在价值的，以及最终可解释的模式的非平凡过程。它包含了如下图所示的五个阶段：

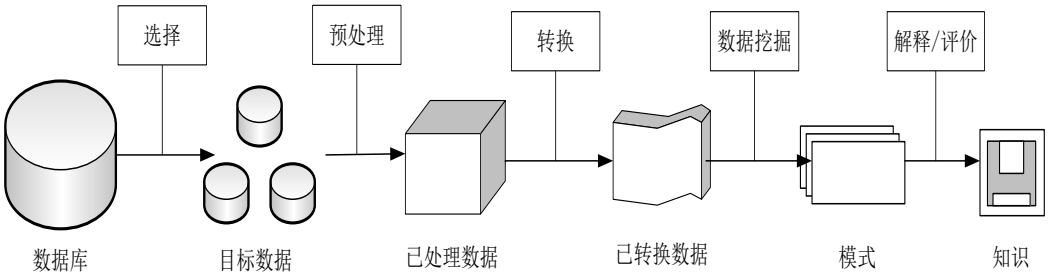


图 2 Fayyad 数据挖掘流程图

数据选择：检索数据库中与业务对象有关的数据信息，并从中选择适用于数据挖掘的数据。比如，将所有预约挂号的患者信息筛选出来，形成一个特定的数据子集。

数据预处理：预处理是数据挖掘之前必不可少的操作过程，目的在于检查数据的一致性和完整性，消除干扰数据，过滤与挖掘无关的冗余数据，并能够根据实际变化情况补充新的数据。

数据转换：对已经预处理的数据进行再次处理，目的是根据数据挖掘工具所需的数据值和数据类型把不同数据源的数据进行格式统一。比如对患者生日信息进行转换，统一存储类型及格式。

数据挖掘：这是整个 KDD 过程中最重要的一个步骤，利用上面提到的发掘技术和方法对数据进行计算分析，找到有价值的规律和关系。

解释和评价：对挖掘后的模式进行解释，转化成能被利用的有价值的知识，并将它以更容易理解的方式展现给用户，比如做成表格或图形等。

三、基于病案数据利用关联规则算法的结果分析

关联规则挖掘过程主要包含两个阶段：第一阶段必须先从资料集合中找出所有的高频项目组，第二阶段再由这些高频项目组中产生关联规则。

关联规则挖掘的第一阶段必须从原始资料集合中，找出所有高频项目组。高频的意思是指某一项目组出现的频率相对于所有记录而言，必须达到某一水平。一项目组出现的频率称为支持度(Support)，以一个包含 A 与 B 两个项目的 2-itemset 为例，我们可以经由公式 Support(A,B)=P(A and B)求得包含{A,B}项目组的支持度，若支持度大于等于所设定的最小支持度(Minimum Support)门槛值时，则{A,B}称为高频项目组。一个满足最小支持度的 k-itemset，则称为高频 k-项目组，一般表示为 Large k 或 Frequent k。算法并从 Large k 的项目组中再产生 Large k+1，直到无法再找到更长的高频项目组为止。

关联规则挖掘的第二阶段是要产生关联规则。从高频项目组产生关联规则，是利用前一步骤的高频 k-项目组来产生规则，在最小信赖度的条件门槛下，若一规则所求得的信赖度满足最小信

赖度，称此规则为关联规则。例如：经由高频 k-项目组{A,B}所产生的规则 AB，信赖度可经由公式 Confidence=p(A and B)/p(A)求得，若信赖度大于等于最小信赖度，则称 AB 为关联规则。

以患者病种统计查询分析，使用关联规则数据挖掘技术，对在某一时间段出院的患者病案中的疾病进行资料挖掘。首先必须要设定最小支持度与最小信赖度两个门槛值，在此假设最小支持度为 min_support=3% 且最小信赖度为 min_confidence=8%。因此符合该病种需求的关联规则将必须同时满足以上两个条件，并经过挖掘过程找到关联规则。

[试管受精,郑州]=17.9%，[试管受精,商丘]=9.7%，[试管受精,驻马店]=8.4%

来医院看病的患者中，需要试管内受精(ICD 码 Z31.200)的患者中地区出现比例较高的是郑州市，其他地区的还有商丘和驻马店。

若有患者来生殖中心进行住院，就可根据患者的所在地的不同，对相同疾病进行相同治疗方案。

AVG[疾病名称]平均费用、疾病科室诊断排序、住院日进行分析，预先告知患者初步住院时间长短，平均每天费用，平均费用信息。科室常见诊断疾病等从而支持临床入路径治疗。

出院病种统计查询

查询出院时间：2015-8-1至2015-8-31　　共3046条记录

编号	疾病名称	ICD-10	出院病人数							治愈率(%)	好转率(%)	死亡率(%)	平均住院日数(日)	术前平均住院日数(日)	平均每人住院医疗费					药品比例(%)
			合计	小计	治愈	好转	未愈	死亡	其他						总费用	其中				
																床位	药费	手术	检查	
1	手术后恶性肿瘤化学治疗	Z51.102	2533	10	5	4	0	1	2523	99.8	0.2	0	8.3	3.80	11637.65	212.28	6801.89	200.29	3284.24	58.4
2	为肿瘤化学治疗疗程	Z51.100	1564	1	1	0	0	0	1563	100	0	0	10.8	2.50	15917.69	307.53	9559.47	273.78	3999.55	60.1
3	试管内授精	Z31.200	636	0	0	0	0	0	636	100	0	0	6.9	1.50	11796.6	214.05	1336.54	3953.43	2959.49	11.3
4	冠状动脉粥样硬化性心脏病	I25.103	590	586	451	133	2	0	4	77.1	22.5	0	8.8	2.70	33363.44	322.01	5969.19	3195.92	4531.77	17.9
5	肺恶性肿瘤	C34.900	537	517	98	395	22	2	20	22	73.6	0.4	12.4	4.60	19790.19	328.71	10130.64	377.31	7001.54	51.2
6	手术后随诊检查	Z09.000	466	0	0	0	0	0	466	100	0	0	8.4	3.40	15685.58	280.16	5209.04	1121.82	3382.6	33.2
7	脑梗死	I63.900	440	436	136	294	5	1	4	31.8	66.8	0.2	11.9	5.40	20235.04	317.65	11412.73	295.24	5828.73	56.4
8	冻融胚胎移植	Z31.202	400	0	0	0	0	0	400	100	0	0	1.3	0.70	3980.3	35.38	404.95	1267	1641.98	10.2
9	肺部感染	J98.414	389	384	180	191	12	1	5	47.6	49.1	0.3	13.4	4.40	30450.21	833.85	16259.7	432.51	8612.03	53.4
10	甲状腺恶性肿瘤	C73.x00	298	294	278	15	1	0	4	94.6	5	0	9.8	2.90	22290.04	309.36	6461.18	3678.82	7375.89	29
11	乳腺恶性肿瘤	C50.900	267	262	235	25	2	0	5	89.9	9.4	0	12.8	3.10	17885.38	436.02	6365.55	1050.43	6424.38	35.6

四、总结与展望

本文主要介绍了医院病案数据挖掘系统的可行性、数据库建模、ETL 数据抽取等数据挖掘系统各业务模块功能分析，根据病案系统的特点 ，构造一个病案数据挖掘系统的流程，利用关联规则，对病案数据挖掘的结果进行了分析。基于病案系统的数据挖掘，不仅仅能够对出院患者的医保类型进行挖掘，还能够对病人来源、病人就诊时间分布等挖掘。当然不同的挖掘算法可能会得到稍有不同的结果。

PU-059

区域健康医疗大数据应用发展政策内容分析

牟燕、吴敏、何有琴
山东省医药卫生科技信息研究所，山东第一医科大学（山东省医学科学院）

目的 健康医疗大数据是国家重要的基础性战略资源，2016 年 6 月，国务院办公厅颁布了《关于促进和规范健康医疗大数据应用发展的指导意见》，对健康医疗大数据的应用发展作出战略指导。

地方政府以《意见》为参照，结合区域特色，对健康医疗大数据应用发展战略进行细化和深化，制定切实有效的实施方案，是地方政府落实国家健康医疗大数据政策的核心与关键。本研究尝试对国家与区域健康医疗大数据应用发展政策进行分析，梳理政策内容与典型案例，以探寻健康医疗大数据政策在实施过程中的重点、实施路径和区域特色，为政府的健康医疗大数据政策的制定、修订与实践提供参考。

方法 将13个省出台的健康医疗大数据应用发展相关政策文件纳入研究，运用Nvivo12软件对以上政策进行文本内容的编码分析。

结果 针对夯实健康医疗大数据应用基础、全面深化健康医疗大数据应用、规范和推动“互联网+健康医疗”服务和加强健康医疗大数据保障体系建设等任务，分析不同区域间在政策制定和实施过程中的共同点和特色之处。

结论 从整体来看，区域政策大都以国家《意见》为参照，政策内容和结构上存在趋同化现象，对健康医疗大数据应用发展重点任务十分明确，从基础层铺设，到应用体系和产业体系的建设，并以保障体系建设为后盾，从而实现健康医疗大数据应用与发展。健康医疗大数据应用基础是基本任务。目前国家级平台共享交换系统等投入使用，32个省级平台已初步建成，四级平台基本实现连通全覆盖。在数据交换方面，国家卫健委印发了一系列标准文件来规范医疗机构数据报送和统一接口，以打破医疗信息机构间的壁垒。健康医疗大数据保障体系建设仍需加强。我国健康医疗大数据安全管控措施仍需加强，需要加快对个人信息保护的立法，并在立法时综合考虑和论证个人健康医疗信息的特殊性，结合时代特色出台相关法律法规，维护大数据的健康有序发展。目前，在法规建设区域政策的推动落实过程中，只有贵州和山东省有所动作，还需进一步加快立法步伐。健康医疗大数据的应用与服务凸显区域特色。由于各地健康医疗大数据发展水平不一，其应用与服务相关政策都凸显了区域特色，推动健康产业与养老、旅游、互联网、健身休闲等的深度融合，尤其在培育健康医疗大数据应用新业态、研制推广数字化健康医疗智能设备和推动健康医疗大数据产业发展方面，各区域都基于自己省域的经济优势和特色，来布局产业发展。

2015年10月，党的十八届五中全会正式提出“实施国家大数据战略，推进数据资源开放共享”，这表明我国已将大数据视作战略资源并上升为国家战略。健康医疗大数据是国家重要的基础性战略资源，其应用发展将带来健康医疗模式的深刻变化，有利于激发深化医药卫生体制改革的动力和活力，提升健康医疗服务效率和质量，扩大资源供给，不断满足人民群众多层次、多样化的健康需求，有利于培育新的业态和经济增长点[1]。2016年6月，国务院办公厅颁布了《关于促进和规范健康医疗大数据应用发展的指导意见》（以下简称《意见》），对健康医疗大数据的应用发展作出战略指导。地方政府以《意见》为参照，结合区域特色，对健康医疗大数据应用发展战略进行细化和深化，制定切实有效的实施方案，是地方政府落实国家健康医疗大数据政策的核心与关键。

目前我国健康医疗大数据领域的相关研究主要集中在健康医疗大数据的概念与发展[2,3,4]，健康医疗大数据应用与实践[3,5]、健康医疗大数据标准和资源体系建设研究[6,7]和健康医疗大数据的信息安全与保护[8]等。健康医疗大数据政策的相关研究很少，李后卿等[9]分析了我国发布的健康医疗大数据国家战略文件及实施现状，于琦等[10]从政策文献量化的视角分析了我国国家层面健康医疗大数据政策，尚无对国家和省级政府出台的健康医疗大数据应用发展政策进行内容分析的研究。本文尝试对国家与区域健康医疗大数据应用发展政策进行分析，梳理政策内容与典型案例，以探寻健康医疗大数据政策在实施过程中的重点、实施路径和区域特色，为政府的健康医疗大数据政策的制定、修订与实践提供参考。

1 数据来源及分析方法

1.1 数据来源

自2016年6月《意见》颁布以来，截至2020年6月，经查询各省、直辖市人民政府官网及卫生健康委员会官网，全国（不含港澳台）已有13个省相继出台了健康医疗大数据应用发展的实施方案；另外，从媒体相关报道中查询到北京市发布《北京健康信息互联互通与大数据应用行动计划工作方案》、甘肃省印发《甘肃省促进和规范健康医疗大数据应用发展的实施方案》的信息，但并未从政府公开渠道获得全文；查询到浙江省发布的《关于促进和规范浙江省健康医疗大数据应用发展的实施意见（征求意见稿）》，未查询到正式发文。因此最终将13个省出台的政策文件纳入研究（统称为“区域政策”），政策文本由各省市人民政府官网及卫生健康委员会官网下载获得，具体情况见表1。

表1 各省、直辖市颁布的健康医疗大数据应用发展政策一览

颁布时间	省份	发文部门	政策名称
2016.10	湖北	湖北省人民政府办公厅	湖北省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见
2016.10	云南	云南省人民政府办公厅	云南省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见
2016.11	河北	河北省人民政府办公厅	河北省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见
2016.11	广西	广西壮族自治区人民政府办公厅	广西壮族自治区人民政府办公厅关于印发广西促进和规范健康医疗大数据应用发展工作实施方案的通知
2016.12	四川	四川省人民政府办公厅	四川省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见
2016.12	安徽	安徽省人民政府办公厅	安徽省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见
2016.12	重庆	重庆市人民政府办公厅	重庆市人民政府办公厅关于印发重庆市健

			康医疗大数据应用发展行动方案(2016—2020 年)的通知
2017.02	广东	广东省人民政府办公厅	广东省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见
2017.05	青海	青海省人民政府办公厅	青海省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见
2017.07	贵州	贵州省人民政府办公厅	贵州省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见
2017.07	山东	山东省人民政府办公厅	山东省人民政府办公厅关于贯彻国办发(2016)47 号文件促进和规范健康医疗大数据应用发展的实施意见
2017.10	天津	天津市卫生计生委	市卫生计生委印发关于促进和规范健康医疗大数据应用发展的实施方案的函
2017.12	陕西	陕西省人民政府办公厅	陕西省人民政府办公厅关于印发促进和规范健康医疗大数据应用发展实施方案的通知

1.2 分析方法

本研究采用内容分析方法，运用 Nvivo12[11]对以上 13 项健康医疗大数据应用发展政策进行文本内容的编码分析，并分析不同区域间在政策制定和实施过程中的共同点和特色之处。

2 健康医疗大数据应用发展政策的总体情况

2.1 政策主要特征分析

13 项政策文本时间跨度为 2016 年 10 月到 2017 年 12 月，显示健康医疗大数据作为新兴发展战略，在国家相关政策发布后，各地政府十分重视，跟进推动政策的制定和落实，为健康医疗大数据的应用发展营造了良好的政策环境。政策文种类型是指颁布政策采用的形式，通过分析可以看出现有政策约束性和可操作性的强弱[12]。根据政策标题内容对 13 项政策文本进行政策文种分类，有 9 项政策文本以“实施意见”为题，占 69.23%，3 项政策文本以“通知”为题，占 23.08%，另外还有 1 项政策文本以“函”为题，占 7.69%。从文件出台部门的管理层级看，12 项政策为省（直辖市）人民政府办公厅发布，1 项政策为原卫生计生委发布。有 9 项政策明确了重点任务和重大工程的牵头部门和责任部门，这样能更有效地部署和推动政策的落地落实。

综上，从政策力度来看，区域健康医疗大数据应用发展相关政策层次较高，具有较强的规范性和约束性，并且具有很强的指导性和可操作性。这充分体现了各地省级政府结合当地实际，出台相应政策来有效承接国家战略。

2.2 政策内容总体分析

将《意见》与 13 项区域健康医疗大数据应用发展政策导入 Nvivo12 软件。首先进行聚类分析，发现区域政策与国家政策高度正相关，Pearson 相关系数最高值达 0.97（国家政策与湖北政策），最低值为 0.84（国家政策与四川政策），一致程度较高。从政策文本结构来看，大都遵循《意见》的结构，分为指导思想、基本原则和发展目标、重点任务和重大工程以及加强组织实施三大部分。因此，在对区域政策进行编码分析时，将《意见》中的标题作为父节点，二级标题作为子节点，三级标题作为二级子节点，形成编码框架，父节点编码情况见表 2。对于重点任务和重大工程部分，在遵循原有的二级子节点的前提下，将具有区域特色的部分编码为新的二级子节点。将区域政策内容按照编码框架进行编码，符合节点的内容编码为一个参考点，参考点数值指的是某一政策节点下编码的次数，若某政策节点中的内容分散在文本的不同地方，则进行多次编码，形成相应的参考点数值。对参考点进行统计，可以看出区域在健康医疗大数据应用发展中的政策分布情况。

“指导思想、基本原则和发展目标”和“加强组织实施”这两部分参考点数较低，从内容来看区域政策与《意见》具有很高的一致性，因此不再展开讨论。

表 2 父节点整体编码情况表

父节点	政策来源数	参考点数
指导思想、基本原则和发展目标	12	26
重点任务和重大工程	13	225
加强组织实施	13	55

3 健康医疗大数据应用发展政策的重点任务和重大工程分析

根据《意见》的框架，重点任务和重大工程一级子节点分别为夯实健康医疗大数据应用基础、全面深化健康医疗大数据应用、规范和推动“互联网+健康医疗”服务和加强健康医疗大数据保障体系建设，下面结合二级子节点展开分析。

运用 NetDraw 对区域政策二级子节点绘制网络图谱，如图 1 所示。方块节点代表的是地区，圆圈节点代表的是健康医疗大数据应用发展的主要任务，连线的粗细代表的是在该项任务下区域政策参考点数的多少。从图中可以看出，除省份节点中间的圆圈部分代表的是《意见》中提出的十四项重点任务和重大工程外，省份两旁的节点代表的是各区域根据自己的特色实施的一些新的重点任务，不同的任务每个省市的参考点数也不同。从图 1 可以明显看出，安徽、广东、广西、山东、四川、天津、云南和重庆在大数据应用基础方面政策参考点较多；陕西、山东、重庆和贵州在健康医疗大数据应用和产业发展方面的政策参考点较多；广东、贵州、青海、山东和天津在加强健康医疗大数据保障体系建设方面的政策参考点较多。

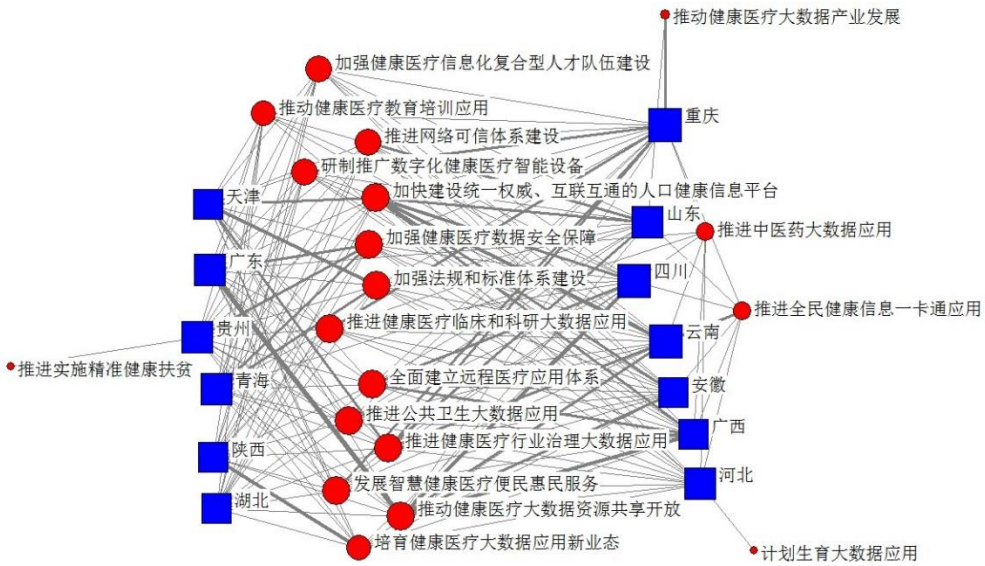


图 1 区域重点任务网络图

3.1 夯实健康医疗大数据应用基础

“夯实健康医疗大数据应用基础”子节点编码见表 3。

表 3 “夯实健康医疗大数据应用基础”子节点编码情况

二级子节点	政策来源数	参考点数
加快建设统一权威、互联互通的人口健康信息平台	13	20
推动健康医疗大数据资源共享开放	13	22

区域政策的重点任务都提到了要全面建成互通共享的省、市、县三级人口健康信息平台，健全全员人口、居民电子健康档案和电子病历三大基础数据库，建立健康医疗大数据共享机制，建立省级健康医疗数据资源目录体系，制定健康医疗大数据开放应用政策规范，消除数据壁垒，实现卫生与其他行业数据的共享共用、业务的高效协同。区域在健康医疗大数据应用基础建设中的特色政策见表 4。

表 4 “夯实健康医疗大数据应用基础”区域特色

特色	区域	举措
“健康云”	广东、广西、河北、山东、山西、四川和云南	建设“健康云”，依托三级人口健康信息平台，整合三大数据库，建设健康医疗大数据中心。
健康医疗大数据共享交换平台	山东、广西、贵州、湖北、青海、陕西和重庆	建立健康医疗大数据共享交换平台，将可穿戴设备、智能健康电子产品、健康医疗移动应用等产生的数据资源规范接入健康医疗大数据集合。

3.2 全面深化健康医疗大数据应用

“全面深化健康医疗大数据应用”子节点编码见表5。从表5中可见，除了《意见》中提出的推进健康医疗行业治理大数据应用、推进健康医疗临床和科研大数据应用、推进公共卫生大数据应用、培育健康医疗大数据应用新业态和研制推广数字化健康医疗智能设备等指导意见外，还有一些区域将推进中医药大数据应用、推动健康医疗大数据产业发展、推进实施精准健康扶贫和推进计划生育管理大数据应用作为区域的实施重点。区域在全面深化健康医疗大数据应用中的特色政策见表6。

表5 “全面深化健康医疗大数据应用”子节点编码情况

二级子节点	政策来源数	参考点数
推进健康医疗行业治理大数据应用	13	18
推进健康医疗临床和科研大数据应用	13	14
推进公共卫生大数据应用	13	13
培育健康医疗大数据应用新业态	11	13
研制推广数字化健康医疗智能设备	12	13
推进中医药大数据应用	7	7
推动健康医疗大数据产业发展	2	3
推进实施精准健康扶贫	1	1
推进计划生育管理大数据应用	1	1

表6 “全面深化健康医疗大数据应用”区域特色

特色	区域	举措
健康医疗大数据监管	广西、云南、重庆和天津	广西和云南要依托人口健康信息平台，建立医疗、医保、医药信息联动机制；重庆要建立全市医药大数据中心，为药品监管及政策制定提供支撑；天津提出要通过家庭医生签约服务信息管理平台信息的整合、连续记录和共享，实现健康医疗大数据对医药卫生体制改革的有效支撑。
临床医学数据示范中心	广东、陕西、天津和重庆	利用现有资源、立足专科优势建立国家级区域临床医学数据示范中心；重庆提出建立生物医学基因数据资源中心和国家级数字人资源中心来推动医学科研方面的创新。
公共卫生机构合作	广东	广东省提出要加强与世界卫生组织、国际及港澳台地区公共卫生机构的合作。
血液管理	河北	河北省提出要建立血液管理和安全预警系统。
健康医疗大数	贵州、陕西、	贵州省提出要发展个性化体育健服务、发展智慧健

据产业	山东、重庆	康养生养老服务和发展健康医疗旅游服务；陕西省提出要培育发展数字化健康医疗企业和产业，进一步健全政府购买体育公共服务体制机制；山东省提出要促进大数据支撑下的健康维护、健康保障、健康产业三大体系融合发展，搭建山东省健康医疗大数据科技创新平台，积极争取承建国家健康医疗大数据中心；重庆市提出要培育家庭健康服务和妇幼保健服务等新业态，打造新型健康医疗金融保险服务业，发展以健康医疗大数据位支撑的生物基因与医疗设备产业，发展健康医疗新技术及智能软硬件产业，建设健康医疗产业孵化园及健康示范小镇。
中医药大数据发展与应用	广东、广西、河北、青海、四川、云南和重庆	通过建立中医药数据资源库、大数据中心等措施来指引患者“治未病”，发展中医药创新服务。
精准健康扶贫	贵州	运用大数据来推进精准健康扶贫，建立贵州省城乡居民因病致贫、因病返贫管理数据库。
计划生育管理大数据应用	河北	建立动态管理的全员人口数据库及共享机制，为人口监测预警提供依据。

3.3 规范和推动“互联网+健康医疗”服务

“规范和推动互联网+健康医疗服务”子节点编码见表 7。规范和推动“互联网+健康医疗”服务是依托互联网技术，建立和完善远程医疗信息系统，运用远程医疗技术，推进三级医院的数据共享和业务协同，推动落实分级诊疗制度，方便群众获取健康服务。区域在规范和推动“互联网+健康医疗”服务中的特色政策见表 8。

表 7 “规范和推动互联网+健康医疗服务”子节点编码情况

二级子节点	政策来源数	参考点数
发展智慧健康医疗便民惠民服务	13	15
全面建立远程医疗应用体系	13	14
推动健康医疗教育培训应用	11	11
推进全民健康信息一卡通应用	7	7

表 8 “规范和推动互联网+健康医疗服务”区域特色

特色	区域	举措
智慧医疗	重庆	建立人体活动体征监测服务平台、慢病及特殊人群健康监护服务平台和健康医疗服务平台，基于可穿

		戴及智能传感设备实现对特殊人群体征数据的监测。
第三方服务	广东	发展第三方在线健康调查、咨询评价、健康管理、健康宣教等应用服务，促进健康医疗服务质量的提升。
智慧医保	广东、贵州和云南	整合现有城镇居民基本医保等医保管理信息系统，实现统一平台管理，支持跨地区费用核查和即时结算
省际、国际远程医疗	广东、广西	广东提出要发展省际、国际远程医疗；广西省提出要加强与东盟国家的健康医疗合作与交流。
区域诊断中心	广东、陕西和云南	建设区域病理、影像、心电诊断中心，实现检查检验结果的互连共享。
远程医疗服务平台	山东	山东提出建设省级多学科整合型疑难病综合诊疗远程医疗服务平台，设立 17 个市级远程医学中心。
智能辅助医疗工程	广西和四川	推进人工智能、物联网等技术在健康医疗中的应用，运用智能辅助医疗工程来提高健康服务质量与效率。
全科医生教育培训	广东和重庆	探索建立基于互联网的全科医生培养模式
居民健康卡建设工程	安徽、广西、湖北、山东、四川、云南和重庆	将居民健康卡作为覆盖居民全生命周期、跨地域的健康医疗服务体系的载体，实现个人健康信息的共享和推动覆盖全生命周期的一体化电子健康服务。

3.4 加强健康医疗大数据保障体系建设

“加强健康医疗大数据保障体系建设”子节点编码见表 9，这一部分的区域政策相似程度较高，区域特色见表 10。

表 9 “加强健康医疗大数据保障体系建设”子节点编码情况

二级子节点	政策来源数	参考点数
加强法规和标准体系建设	13	16
推进网络可信体系建设	12	13
加强健康医疗数据安全保障	13	15
加强健康医疗信息化复合型人才队伍建设	12	12

表 10 “加强健康医疗大数据保障体系建设”区域特色

特色	区域	举措

健康信息安全规划	山东	制定山东省全民健康信息安全规划，强化责任追究和通报制度，落实国家密码政策法规和技术标准，制定完善密码应用规划和建设方案，加强密码应用的统筹协调，规范密码应用。
引进和培养人才	贵州、青海、山东和天津	依托健康医疗大数据的科研和产业，引进和培养领军人才和高端人才；贵州省还提出要“大力发展健康医疗大数据发展应用行业组织，落实全民健康信息化人才职称晋升机制”，以培养人才、留住人才。

4 结语

本文运用 Nvivo12 软件对国家及区域健康医疗大数据应用发展政策进行编码分析，从整体来看，区域政策大都以《意见》为参照，政策内容和结构上存在趋同化现象，对健康医疗大数据应用发展重点任务十分明确，从基础层铺设，到应用体系和产业体系的建设，并以保障体系建设为后盾，从而实现健康医疗大数据应用与发展。共性之外，不同省份的政策内容也随省情存在不同程度的差异性。

4.1 健康医疗大数据应用基础是基本任务

平台的建设、数据的共享是健康医疗大数据的基础任务，只有畅通部门、区域、行业之间的数据共享通道，才能促进全民健康信息共享应用。国家和区域政策都非常重视平台建设的统一权威、互联互通工作，国家 2017 年印发了《省统筹区域全民健康信息平台建设总体方案》、《省统筹全民健康信息平台建设方案编制指引》等文件来推进全民健康信息平台建设，目前国家级平台共享交换系统等投入使用，32 个省级平台已初步建成，四级平台基本实现连通全覆盖。在数据交换方面，国家卫健委印发了《国家健康医疗大数据标准、安全和服务管理办法（试行）》、《全国医院信息化建设标准与规范（试行）》、《全国基层医疗卫生机构信息化建设标准与规范（试行）》等文件来规范医疗机构数据报送和统一接口，以打破医疗信息机构间的壁垒。

4.2 健康医疗大数据保障体系建设仍需加强

健康医疗大数据是国家重要的基础性战略资源，健康医疗大数据的安全关系到国家战略安全、国家生物安全、人民生命安全和公民个人隐私安全。尽管我国个人信息保护制度和法律体系尚未健全，国家对全民健康信息采集、管理、利用、安全和隐私保护等工作十分重视。《国家健康医疗大数据标准、安全和服务管理办法（试行）》明确了健康医疗大数据安全管理的范畴，分别对健康医疗大数据管理责任单位和监管单位的责任进行了明确。《基本医疗卫生与健康促进法》第四十九条、第九十二条也明确要保障健康医疗信息的安全，意味着个人健康医疗信息的安全保护逐步走向法制化道路。我国健康医疗大数据安全管控措施仍需加强，需要加快对个人信息保护的立法，并在立法时综合考虑和论证个人健康医疗信息的特殊性，结合时代特色出台相关法律法规，维护大数据的健康有序发展[13]。目前，在法规建设区域政策的推动落实过程中，只有贵州和山东省有所动作，贵州出台了《贵州大数据安全保障条例》，但并非针对健康医疗大数据；山东省起

草了《山东省健康医疗大数据管理办法（草案征求意见稿）》，现已公开征求意见，相信不久这一法律法规就会落地，健康医疗大数据应用发展会更加规范。

4.3 健康医疗大数据的应用与服务凸显区域特色

区域经济水平、产业结构是健康医疗大数据政策制定的重要依据，由于各地健康医疗大数据发展水平不一，其应用与服务相关政策都凸显了区域特色，推动健康产业与养老、旅游、互联网、健身休闲等的深度融合，尤其在培育健康医疗大数据应用新业态、研制推广数字化健康医疗智能设备和推动健康医疗大数据产业发展方面，各区域都基于自己省域的经济优势和特色，来布局产业发展。如贵州省健康医疗大数据应用发展规划特色就很鲜明，如提出运用大数据来推进精准健康扶贫；发展健康医疗旅游服务，将医疗护理、健康体验、康复疗养与贵州多梯度山地运动、避暑纳凉气候条件相结合，深度开发中医药美容美体旅游、长寿养生旅游、医疗疗养旅游等高端特色旅游产品；在教育培训方面，在充分利用好省内医学教育资源的同时也提出了与国内知名医院间的对口帮扶机制。而山东省依托省会济南丰富的医疗资源优势，提出建设健康医疗大数据研究成果示范推广和产业化平台，积极争取城建国家健康医疗大数据中心等，并积极推动政策的落地。2018 年 4 月，国家卫生健康委与山东省政府、济南市政府签订《关于共建国家健康医疗大数据北方中心合作框架协议》，标志着国家健康医疗大数据北方中心正式落户济南。贵州、陕西、山东和重庆都提出健康医疗大数据中心或产业园建设，引进或培育相关企业和高端人才，推动健康医疗大数据新产业、新业态的形成。广东和广西立足地理位置和优势，在公共卫生和远程医疗方面分别提到了与国际及港澳台地区的合作，与东盟合作，河北政策则体现了京津冀跨区域的协作。遵循区域特色来制定健康医疗大数据的应用和服务相关政策，必将会推动区域卫生事业和健康产业的发展。

PU-060

中文临床自然语言理解评估基准构建

杨飞洪[1,2]、王序文[1,2]、李姣[1,2]
1. 中国医学科学院医学信息研究所 2. 北京协和医学院

目的 为评估不同模型对中文临床文本的理解能力，构建面向中文临床领域的多任务自然语言理解算法评估基准(CCLUE Benchmark)。

方法 梳理整合中文临床自然语言处理领域相关资源，包括电子病历命名实体识别数据集，糖尿病中文临床术语标注数据集，临床试验文本分类数据集，中文健康问答数据集。针对四个有代表性的中文临床自然语言理解任务，（即，命名实体识别，术语标注，问答，文本分类任务），分别对其数据集进行整理并设计评价指标。以代表性的 BERT 模型为基线方法，评估其在 CCLUE Benchmark 上的表现。

结果 获取电子病历命名实体标注样本 2674 份，糖尿病文献及临床指南命名实体标注样本 9556 份，健康问句匹配标注样本 20000 份，临床试验短文本分类标注样本 38341 份，分别对每个数据集进行整合优化处理，并以 7:1:2 的比例将这四份标注数据集分别划分为训练集、验证集和测试集。中文 BERT 模型在 4 个任务中验证集的平均 F1 为 0.7577，测试集的平均 F1 为 0.7707。其中，电子病历命名实体识别 F1 为 0.5873，糖尿病术语标注 F1 为 0.7789，健康问答 F1 为 0.8900，临床试验文本分类 F1 为 0.8268.

结论 本文构建的 CCLUE Benchmark 评估基准提供了一个算法性能评价平台，便于研究人员针对不同模型在中文临床自然语言理解任务中的表现开展测评和研究，有助于促进中文临床自然语言理解能力的快速提升。

PU-061

美国 9 所医学院（校）图书馆信息素养教育现状调查与分析

刘靓靓
中国医学科学院医学信息研究所/图书馆

目的 本文以医学院校图书馆信息素养教育为研究主题。

方法 采用网络调研法，选取美国 9 所医学院（校）图书馆作为调查对象，调查其信息素养教育资源网页组织方式、信息素养教育内容、信息素养教育资源开放程度等内容。

结果 发现被调查的美国医学院（校）图书馆培训形式以线上和线下相结合，培训课件资源很少提供下载渠道，信息素养教育内容除了传统的文献检索课程及培训、图书馆资源与服务利用指导、工具软件培训，还包括数据管理、信息道德与知识产权、教师信息素养相关的培训。

结论 结合调查结果，本文对医学院校图书馆信息素养教育提出以下建议：把全文获取能力作为信息素养教育的切入点、进一步开展信息意识与信息道德教育、开展健康信息素养教育。

1 引言

信息素养是获取信息、管理信息、分析信息的必备素养，对个人的学习能力、科研能力、创新能力及终身发展能力有着深远影响。医学院校图书馆是信息素养教育的重要阵地，经过 30 多年的建设和发展，我国医学院校图书馆信息素养教育已取得一定成绩，但是随着医学信息资源的快速增长，信息获取渠道日益复杂，传统的医学院校图书馆信息素养教育面临极大挑战，需进一步总结新环境下医学生对信息素养教育的需求特点，在分析传统医学院校图书馆信息素养教育存在问题的基础上，探讨医学院校图书馆信息素养教育的发展方向。

他山之石，可以攻玉。自信息素养教育的作用在个体发展、机构建设、国家规划中的价值愈发显现，国外医学院图书馆纷纷开展了不同形式和不同内容的信息素养教育活动。本文对美国 9

所医学院（校）图书馆信息素养教育案例进行梳理、归纳，为我国信息素养教育进一步优化提供借鉴参考。

2 数据来源与调查方法

在国外调查对象的选取上，为总结国外医学院校图书馆信息素养教育的经典案例，系统、集中揭示信息素养教育的成功经验，考虑到美国是较早开展信息素养教育的国家，积累了丰富的信息素养教育经验，因此选取 2019 年 U.S.News 医学院排名前 10 名的医学院（校）图书馆作为调查对象[1]，排除网页无法打开的 1 所医学院（校）图书馆，最终确定 9 所医学院（校）图书馆作为调查对象（见表 2.1），调查时间为 2020 年 3 月 17 日-3 月 27 日。在调查方法上，本文主要采用网站调查法与内容分析法，通过浏览图书馆网站各版块内容，提取、总结、分析与信息素养教育主题相关的内容。

表 2.1 医学院（校）图书馆信息素养教育调查对象名单

机构名称	网址
哈佛大学图书馆	https://library.harvard.edu/
约翰霍普金斯大学图书馆	https://www.library.jhu.edu/
纽约大学图书馆	http://library.nyu.edu/
斯坦福大学图书馆	https://library.stanford.edu/
加州大学旧金山分校图书馆	https://www.library.ucsf.edu/
宾夕法尼亚大学图书馆	https://www.library.upenn.edu/
加州大学洛杉矶分校图书馆	https://www.library.ucla.edu/
圣路易斯华盛顿大学图书馆	https://library.wustl.edu/
杜克大学图书馆	https://library.duke.edu/

3 调查结果

3.1 信息素养教育资源网页组织方式

本文对信息素养教育资源网页组织方式的调查主要通过浏览被调查对象网站，查看信息素养教育资源的链接目录层次和所属版块。链接目录层次指的是信息素养教育在图书馆网站中的链接深度，所属版块指的是信息素养教育资源在图书馆网站中的所属版块，这些内容可以在一定程度上反映出被调查的医学院（校）图书馆对于健康信息素养教育的重视程度与开展状况。

根据调查，只有圣路易斯华盛顿大学图书馆网站拥有信息素养教育二级专题页面[2]（The Instruction and Information Literacy Program），其余 8 所医学院（校）图书馆都采用了分散式的信息素养教育网页信息组织方式，即图书馆网页没有以信息素养教育命名的专题栏目，而是分散在网页各个版块，将信息素养教育融入了图书馆资源与服务中。例如，哈佛大学图书馆网页的信息素养教育资源分布在用户指南（How to）、分馆链接（Library）、服务（Service）、工具（Tool）四个一级版块，内容分别是：①用户指南版块以简洁浅显的语言指导用户如何利用图书

馆的资源与服务；②分馆链接版块以各个分馆链接为文本组织单元，信息素养教育内容蕴含在分馆服务中。例如，点击医学分馆（Francis A.Countway Library of Medicine）网站的链接，在分馆服务版块中可以看到分馆培训的相关信息，包括培训日期、培训内容和培训形式；③服务版块汇聚了哈佛大学图书馆提供的所有服务内容，如面向教职工的科研协助（Research Help for Harvard Offices）、研究咨询（Research Consultations）等；④工具版块发布了哈佛大学图书馆面向用户提供的相关工具，如用来创建和共享数据管理计划的在线工具 DMPTool，对工具功能、使用方法进行详细说明。除此之外，调查对象的信息素养教育网页文本内容组织方式还具有以下几种特征：

（1）根据研究主题组织相关信息

被调查的 9 所医学院（校）图书馆网站全部设有研究主题（Subject）索引，将图书馆的信息素养教育资源依据主题进行分类组织。研究主题索引通常位于研究指南（Research Guide）目录下，按照字母顺序排列，主题内容包括学科专业、学习与科研周期的信息需求。以圣路易斯华盛顿大学图书馆网站的研究主题索引[3]为例，共包含 83 个研究主题，覆盖人文社科、管理学、理学、工学、医学等领域的 71 个专业学科，以及论文写作、版权与出版、数据管理等 12 个学习与科研主题。每个研究主题含有二级资源链接、资源描述、主题标引馆员信息、标引日期、浏览量、标签等详细信息。

（2）根据用户类型细分信息层次

一些被调查图书馆还根据用户类型，细分信息素养教育网页信息的层次深度。例如，加州大学洛杉矶分校图书馆首页有研究指南（Research Guide）和研究与教学支持（Research &Teaching Support）版块，两者都含有信息素养教育相关内容，但是在文字表达、内容深度上都有所不同。研究指南版块以一站式检索和初级指南（Get Started:Information Research Tips）为主。用户在一站式检索框输入资源或服务的关键词，快速定位到相关指导。搜索框旁边列有搜索提示，引导用户使用简单词汇进行检索。初级指南以简单句介绍电子资源、纸本资源的检索方法、如何缩小检索范围等基本知识，满足图书馆新用户的信息需要。研究与教学支持版块则针对教职工和学生的不同信息需求，分别列举相应的图书馆服务内容。这样差异化的文本层次有助于不同信息素养水平的用户快速找到匹配自身信息能力与信息需求的资源和服务。

（3）根据总-分馆模式组织相关信息

被调查的美国医学院（校）图书馆都设有分馆，在信息素养教育资源的内容组织上也使用了总-分馆模式，具体可以划分为两种形式。第一种是总馆和分馆都具有独立页面。点击分馆网页链接可以查看该馆的资源与服务，能够充分揭示分馆的线下资源和特色资源，有利于用户针对性掌握图书馆讲座与培训信息。例如哈佛大学图书馆、约翰霍普金斯大学图书馆、杜克大学图书馆都设有医学分馆，医学生可以专门访问分馆网站，直接获取医学领域的信息素养教育资源。第二种是总馆页面下设分馆资源的指向链接。例如哈佛大学图书馆和纽约大学图书馆网站的研究指南版

块可选择根据分馆进行分面查看，分馆资源与服务指导按照字母顺序排列，方便用户快速定位，见表 3.1。

表 3.1 调查对象信息素养教育资源网页组织方式

机构名称	链接目录层次	所属版块
哈佛大学图书馆	2 级	用户指南（How to）、分馆链接（Library）、服务（Service）、工具（Tool）
约翰霍普金斯大学图书馆	2 级	帮助（Help&Support）、服务（Services）
纽约大学图书馆	2 级	分馆链接（Locations）、服务（Service）、图书馆活动（Library Workshops & Events）、研究指南（Research Guides）
斯坦福大学图书馆	2 级	关于图书馆（About）、分馆链接（Libraries）、利用图书馆（Using the Libraries）、研究支持（Research Support）
加州大学旧金山分校图书馆	2 级	资源发现（Find Materials）、利用图书馆（Use the Library）、咨询服务（Ask an Expert）
宾夕法尼亚大学图书馆	2 级	帮助（Help with）、分馆链接（Libraries）、图书馆新闻与活动（News&Events）
加州大学洛杉矶分校图书馆	2 级	研究指南（Research Guide）、研究与教学支持（Research &Teaching Support）、分馆链接（Locations）、图书馆新闻与活动（News&Events）、利用图书馆（Using the Libraries）
圣路易斯华盛顿大学图书馆	2 级	分馆链接（Libraries）、研究与指导（Research &Instruction）、服务（Service）
杜克大学图书馆	2 级	利用图书馆（Use the Library）、研究支持（Research Support）、课程支持（Course Support）、分馆链接（Libraries）

3.2 信息素养教育内容

3.2.1 健康信息资源整合与利用指导

如表 3.2 所示，被调查的美国医学院（校）图书馆均提供了健康信息资源整合与利用指导，用户可以在网站的医学学科资源导航、资源常见问题汇总的位置获取健康信息资源及利用方法。每个医学院（校）图书馆的医学学科资源导航分面维度不同，例如约翰霍普金斯大学图书馆在研究指南中列出了医学、药学和生命科学研究学科相关的馆藏文献资源，每个学科下设二级学科专题资源页面，提供在线健康信息资源导航、相关医学数据库导航、药物指南、公众健康信息、论文撰写和馆藏指引等分面导航条目；加州大学旧金山分校图书馆的医学学科资源导航页面包含了获取印本资源、学科研究指导、远程资源访问、论文撰写等分面导航条目。杜克大学图书馆导航页面则是从资源入门、文献获取、数据获取、图书获取、查找国家研究背景、引用与研究工具、二级与三级学科分类等维度进行医学学科资源组织的。

表 3.2 调查对象网站健康信息资源整合与利用指导

机构名称	医学学科资源导航	常见问题汇总
哈佛大学图书馆	生物与化学科学，生物科学，生物学，人口统计学，营养学，健康学，医学科学，药学、口腔医学与公共卫生，心理学，公共卫生，	有
约翰霍普金斯大学图书馆	生物科学，医学与预防保健，心理学	有
纽约大学图书馆	生物学，健康、护理、保健与医学，神经科学，营养与食品健康，公众健康	有
斯坦福大学图书馆	细胞生物学，发展生物学，心理学，生物学（一般问题），实验室安全，药物和生物化学，微生物学，生物学移动应用，医疗器械设计与规范，神经生物学，神经科学	无
加州大学旧金山分校图书馆	基础科学，生物信息学和统计学，临床医学，公众健康信息素养，口腔医学，循证医学，公共卫生，医学史，LGBT 人群健康，医学（一般问题），护理学，药学，物理疗法，医学行政管理，系统评价	有
宾夕法尼亚大学图书馆	生物学，生物医学，临床医学，口腔医学，健康与生命科学，健康与社会，卫生保健与管理，护理学，心理学，人口研究与人口统计，退伍军人用药	无
加州大学洛杉矶分校图书馆	生物学与生命科学，健康科学，心理学	无
圣路易斯华盛顿大	生物学，医学史，生命科学，医学，哲学、神	有

学图书馆	经科学与心理学，公共卫生，	
杜克大学图书馆	健康与医学科学，	有

3.2.2 健康信息素养培训

经过调查，这 9 所医学院（校）图书馆的健康信息素养培训方式主要有新生教育和参考咨询。新生教育方面，哈佛大学图书馆设有新生馆员（Librarians for First-Year Students）岗位，馆员与新生结对，负责新生入学第一学年的图书馆利用指导、论文指导和研究指导。圣路易斯华盛顿大学图书馆在其实体馆内区域开辟了新生探索中心（Newman Exploration Center），供集中开展新生教育活动。

读者培训规模一般有 Workshops、Events、Classes 三类，主要区别在于：①Workshops 通常聚焦于某一培训主题，举办时间、次数比较随机，具有讲座的特征，如约翰霍普金斯大学图书馆的 Workshops 时长从 30 分钟-2 小时不等，一些专题会举办连续性讲座；②Events 和 Classes 更倾向于通常意义上的课程，授课时间更加固定，一些课程还是某些专业学生的必修课。读者培训内容主要包括了医学文献与医学数据库检索培训、医学论文写作与研究指导、医学数据管理培训、工具软件培训、信息道德与知识产权相关培训、教师信息素养培训和信息素养专题培训。

（1）医学文献与医学数据库检索培训

数据库培训方面，依据数据库类型，可以将培训内容分为医学文献资源数据库培训、循证医学数据库培训、临床信息资源数据库培训、药学资源数据库培训。例如，宾夕法尼亚大学图书馆提供生物医学学科的一对一的面对面指和网络研讨会，培训内容包括医学文献数据库 PubMed、临床信息资源数据库 BLAST、ClinVar、MedGen 等，用户通过培训后可以掌握文献信息的检索策略，使用和分析核苷酸、蛋白质序列等生物学数据。圣路易斯华盛顿大学图书馆为全校学生和教职工提供培训服务，特色之处是对美国国立卫生研究院（National Institutes of Health，NIH）的一系列医学信息产品（如 PubMed， My Bibliography，NIH Biosketch，NCBI SciENcv 及 NIH 产品的开放获取政策）开展定制化培训服务，用户填写培训申请表，即可以小组或班级形式接受医学文献与医学数据库检索定制培训。

（2）医学论文写作与研究指导

被调查的图书馆都开展了医学写作与研究指导。如斯坦福大学图书馆开设了科研专题培训班（Gear Up for Research），旨在为斯坦福大学的研究生、博士后、教职员工和研究人员增强研究经验。课程每年三次，冬季和春季课程侧重于提升科学和工程研究人员的技能，而秋季课程侧重于社会科学和人文科学中的数据管理。哈佛大学图书馆设有“朋辈研究员”（Peer Research Fellows）项目。该项目由本科生担任研究指导员，解答学生在论文写作与科研方面的问题，还可以把问题转交给专业学科馆员。纽约大学图书馆开设参考文献格式指导专题页面，整合汇聚了 AMA、APA、MLA 等各类参考文献格式的标注方法，配有参考书目、工具和相关网页链接。此外，斯坦福大学图书馆提供研究提醒（Research Alerts）插件，用户可以设置检索条件与数据库范围，当符

合提醒条件的新文献出现时，会自动向用户发送电子邮件或更新的 RSS 提要，使用户及时了解相关研究进展。加州大学洛杉矶分校图书馆开发了一款研究论文计划助手（Research Paper Planner），用户可以创建或者复制写作计划，工具会将写作任务分解为可管理的阶段性内容，可以随时跟踪进度并关联使用图书馆资源。杜克大学图书馆网页在列举常见参考文献引用格式的同时，还列出了杜克大学课程论文与毕业论文的参考文献要求与指南。

除了直接提供论文写作与研究指导，一些图书馆也从辩证角度指出用户需要在提出图书馆服务与指导的需求前，首先充分评估研究中的问题，以及思考图书馆指导是否适用于研究与写作的当前环节。哈佛大学图书馆针对用户文献检索结果不理想的情况，开设了检索问题诊断页面[4]。该页面引导用户寻找符合自身检索问题的描述（检索结果过多、过少、结果不相关），查看原因分析与解决建议，参考提供的检索案例重新制定检索策略。系统综述是医学研究与论文撰写的重要方法之一，加州大学旧金山分校图书馆设有系统综述专题页面[5]。值得注意的是，该页面没有直接介绍系统综述的写作方法，而是首先指出系统综述并非适合所有研究和所有环节，需要经过充分评估。许多用户不了解系统评估，往往在研究过程中才提交系统评估指导需求，例如咨询某个检索词是否需要调整，浪费了大量的时间和精力。因此，图书馆目前不接受研究过程中提交的系统综述指导申请，而是要求用户在研究开始前邀请图书馆员参与其中。图书馆员在系统综述中的角色包括：参与选题、根据 PICO 原则设计方案、制定检索策略、筛选文献、文献管理软件使用帮助、指导研究方法章节撰写等。该专题网页还详细阐述了系统综述的定义、适用范围、研究方法和参考资料，帮助用户加深对系统综述的了解，更明智地选择科研与写作方法。

（3）医学数据管理培训

被调查的所有医学院（校）图书馆全部设有医学数据管理的专题页面，页面内容包括：数据管理的内涵，如何进行数据管理，包括数据获取、数据管理、数据存储、数据可视化、数据出版、数据存档等方面，以及图书馆在此方面的培训与帮助。所有图书馆都会开展的一项工作是对于撰写数据管理计划（Data Managing Plan）和使用撰写工具的指导。在美国，申请大多数基金项目都需要提交一份数据管理计划，内容包括制定有竞争力的研究计划，能够对研究数据管理和基金管理方要求提出有效解决方案，常用的数据管理计划撰写工具是 DMPTool。哈佛大学图书馆 DMPS 样例（Sample DMPS）服务含有哈佛 Dataverse 样本数据管理计划、DMPTool 提供的公共数据管理计划以及收录 NIH 等常用机构的数据共享计划示例。约翰霍普金斯设有 8 个模块的线上培训，对查询、撰写和复制 DMP、下载、共享、审核 DMP 进行了指导。宾夕法尼亚大学图书馆对 DMP Tools 的数据安全问题进行了详细解答，包括使用工具创建数据管理计划的版权归属、使用政策、隐私政策等。

除了提供常规培训，许多图书馆也通过项目（projects）形式探索数据管理培训的新方向。斯坦福大学图书馆的 Bibliopedia 是一个可视化的文献管理平台，支持文献资源的上传、检索、共享和可视化展示。可视化网络揭示了文献之间的关联，帮助用户了解研究发展历程、研究合作等。该平台是一个开源平台，可供所有研究人员、图书馆和其他机构使用。通过使共享、分析和

保存研究成果变得更加容易来增强数字学术可见度。图书馆的新兴技术团队（the Emerging Technologies Team ,ETT）由图书馆管理人员和技术部门人员组成，负责识别、测试、评估、推广新技术在图书馆内的使用，为图书馆提供新技术咨询和评估调研，他们撰写了 Stanford Box 等工具软件的使用测评，发布信息技术相关的博客、期刊资源，帮助用户更好地掌握工具软件的使用。R 语言培训是斯坦福大学图书馆面向全校师生的培训项目，图书馆拥有 R 语言实验室（R Open Lab）供用户咨询、培训。开设讲习班，咨询，线上教程包。宾夕法尼亚大学图书馆每年举办数字技能研讨会（Digital skills workshops），围绕信息素养主题开展数字出版、数字研究方法、沉浸式技术以及工具软件的研讨和实际操作。

（4）工具软件培训

调查发现，按照工具软件类型分类，培训内容可以分为参考文献管理软件培训、论文写作工具培训、文件与数据管理工具培训、数据分析与统计工具培训、多媒体工具培训和其他工具软件培训。为了帮助用户更好地依据使用场景判断所需工具，约翰霍普金斯大学图书馆、纽约大学图书馆、斯坦福大学图书馆都设有文档管理解决方案网页。页面以表格形式体现，将各种用户需求与各类工具软件对应，用图标或文字表示图书馆推荐在此需求场景下使用该款工具软件，用户可以在表格中查找适合自身需求的工具软件。例如，斯坦福大学图书馆网页上，用户需求是与多人在线同时编辑同一个文档，图书馆推荐使用的工具软件是 Google Drive for Stanford、Google Shared drives for Stanford、OneDrive for Business。每个工具软件名称都附有链接，用户可以点击链接进一步学习该工具软件的使用方法。宾夕法尼亚大学图书馆在常规软件工具的培训基础上，还提供了网页制作、数字人文主题相关的工具软件培训。

此外，许多被调查对象结合图书馆空间服务、特色服务和优势学科开展培训。例如哈佛大学图书馆和斯坦福大学图书馆内设有多媒体工作室，为学生提供加州大学旧金山分校图书馆和加州大学洛杉矶分校图书馆拥有实体创客空间和丰富的 3D 打印与创客空间培训，因此在工具软件的使用培训上也突出了对于相关技术的指导。加州大学旧金山分校图书馆开展了 Bioconductor 、Bowtie、Galaxy、GATK、Samtools 等生物信息学软件的培训。

按照工具软件来源分类，被调查图书馆会对所购买、获取的第三方工具软件进行培训，还会对学校及图书馆自行研发的工具软件进行培训。Perma.cc 是哈佛大学图书馆开发的一款针对网页类型参考文献的引用工具，通过网页快照和链接能够防止网页丢失无法引用的情况。WashU ELN 是圣路易斯华盛顿大学图书馆开发的面向实验数据的电子笔记本，是基于云的研究数据协作和管理工具，图书馆用户可以在该平台上组织、存储和共享实验室数据，操作快速简单。

（5）信息道德与知识产权相关培训

被调查的 9 所医学院（校）图书馆均开展了信息道德与知识产权内容的培训（见表 3.3）。培训形式分为线上网页资源和线下讲座课程。培训内容方面可分为常规知识普及和专题项目。常规知识普及方面，如约翰霍普金斯大学图书馆介绍了科研诚信概念、开放获取资源概念与使用规范、区分抄袭和剽窃区别。加州大学旧金山分校图书馆对“孤儿作品”（Orphan Works）、文献属于

公有领域，无明确知识产权说明（Public Domain）等特殊的知识产权情况进行了阐述。纽约大学图书馆网页重点讲解了文字、图像等不同类型资源的知识产权与许可申请。加州大学洛杉矶分校图书馆从知识产权使用方与版权拥有方的角度分别阐述了相关知识，为论文作者提供维与咨询工作。

专题项目方面，哈佛大学在学校图书馆、档案馆和博物馆都实施了版权第一响应者（Copyright First Responder）项目。项目专员具有版权法律知识专业背景，及时响应用户知识产权相关问题。 斯坦福大学图书馆主办了“数字时代伦理学”（Ethics in the Digital Age Series）线下研讨会项目，会议主题覆盖数字健康、数据科学与信息学、新闻学、社交媒体等，旨在搭建跨学科、跨行业的研讨平台，呼吁人们关注数字时代的学术道德问题。2018 年 7 月和 11 月举办的两次研讨会主题是医学健康信息伦理。与会专家探讨了数据时代个人隐私保护、健康信息共享的潜在风险等议题，形成了《斯坦福大学数字医疗道德指导原则声明》（Stanford Statement of Guiding Principles for Ethics in Digital Health）[6]。圣路易斯华盛顿大学图书馆的版权知识专题（Learning about Copyright）设有免费的知识产权线上课程，每个课程由 3-10 分钟的视频及小测验组成，方便学生和教职员工开展知识产权自测。

表 3.3 调查对象信息道德与知识产权培训情况

机构名称	常规项目	专题项目名称	专题项目面向人群
哈佛大学图书馆	有	版权第一响应者（Copyright First Responder）	教职工、学生
约翰霍普金斯大学图书馆	有	无	无
纽约大学图书馆	有	无	无
斯坦福大学图书馆	有	数字时代伦理学（Ethics in the Digital Age Series ）	教职工、学生、公众
加州大学旧金山分校图书馆	有	无	无
宾夕法尼亚大学图书馆	有	无	无
加州大学洛杉矶分校图书馆	有	无	无
圣路易斯华盛顿大学图书馆	有	版权知识专题（Learning about Copyright）	教职工、学生、公众
杜克大学图书馆	有	无	无

（6）教师信息素养教育

许多图书馆通过教师培训和嵌入式信息素养教育课程，为学校教师提供信息素养教育支持。内容包括对教师自身信息素养能力的培训和协助教师开展信息素养教育，见表 3.4。

教师自身信息素养能力的培训有线上课程和线下培训形式，培训内容包括对教师的远程授课和线下授课提供指导，对教师利用图书馆资源与服务、科研与数据管理进行培训。哈佛大学图书馆为本校教师提供远程教学线上线下培训，培训内容覆盖远程资源访问、电子教参、研究支持、馆际互借、设备支持等。约翰霍普金斯大学图书馆的教育与学习中心（the teaching and learning center）[7]面向教师开展了教学与研究能力培训，包括提供 Blackboard、turnitin、Panopto 等教学工具的使用、基金项目撰写申报培训、教师午餐会交流等。还设有“教育学院”（Teaching Academy）、“创新导师”（The Innovative Instructor）、“教学支持”（Blackboard Support）三个教师提升专项计划，旨在帮助教师获得学术职业技能提升。杜克大学学习创新中心（Duke Learning Innovation）包含了图书馆提供的多项教师信息素养教育内容，教师可通过线上或者线下培训课程掌握教学设计、实施、评估等环节知识，学习使用教学辅助工具，增强科研能力，从而帮助教师拥有更高水平的信息知识与信息能力。

嵌入式信息素养教育课程的调查方面，约翰霍普金斯大学图书馆提供的定制课程内容覆盖。研究设计、信息检索、信息组织、信息评估、学术交流和其他等 6 个版块，教师在定制课程申请表上勾选所需涉及的课程领域，由图书馆专人联系进行定制课程沟通。圣路易斯华盛顿大学图书馆开设了线上形式的教学与信息素养教育项目（The Instruction and Information Literacy Program），通过电子邮件和线上线下课程平台为教师和学生提供信息素养教育支持。教师可以提出申请，邀请图书馆员加入课程教学计划，提供一般性信息素养指导、针对特定学科的信息素养指导、专题馆藏利用指导、数据管理等内容的教学；邀请馆员为学生设计和指导具有信息素养教育内容相关的作业、研究指南和线上学习管理系统模块，帮助教师开展信息素养测评等。宾夕法尼亚大学图书馆的教学设计与教学协作服务计划（Instructional Design & Teaching Collaborations）通过图书馆员协助教师制定课程大纲、课程活动和课程评价，帮助教师培养学生的基本信息素养和创新技术技能，已经拥有将 3D 建模和打印纳入戏剧布景设计、在协作式 Wikipedia 编辑作业中引入知识产权讨论等成功案例。

表 3.4 调查对象教师信息素养教育情况

机构名称	教师培训	嵌入式信息素养教育课程
哈佛大学图书馆	教学培训、图书馆利用培训、文献综述写作培训	信息素养通识教育、信息素养专题教育、数据可视化课程、定制课程
约翰霍普金斯大学图书馆	教学培训、图书馆利用培训、文献综述写作培训、教育与学习中心、教师提升专项计划	信息素养通识教育、信息素养专题教育、定制课程
纽约大学图书馆	开放教育资源利用指导、图书馆利用培训、多媒体资源的使	信息素养通识教育、信息素养专题教

	用指导	育、嵌入数字校园视频流媒体
斯坦福大学图书馆	课程顾问、图书馆利用培训、论文写作与科研指导	信息素养专题教育
加州大学旧金山分校图书馆	教学培训、图书馆利用培训、论文写作与科研指导、多媒体资源的使用指导	课程设计指导、信息素养通识教育、信息素养专题教育、与创客空间结合
宾夕法尼亚大学图书馆	教学培训、图书馆利用培训、论文写作与科研指导	课程设计指导、信息素养通识教育
加州大学洛杉矶分校图书馆	教学培训、图书馆利用培训、教学中的版权问题指导	课程设计指导、信息素养通识教育、 图书馆教学工具包
圣路易斯华盛顿大学图书馆	图书馆利用培训、论文写作与科研指导	信息素养通识教育、信息素养专题教育、定制课程、教学与信息素养教育项目
杜克大学图书馆	教学培训、图书馆利用培训、论文写作、出版与科研指导、学习创新中心	课程设计指导、信息素养通识教育、信息素养专题教育

3.2.3 健康信息参考咨询

被调查的国外医学院（校）图书馆开展了健康信息参考咨询服务，主要渠道是通过学科馆员指导、线上咨询、到馆咨询的方式提供该服务（见表 3.5）。所有被调查对象都提供了线上和到馆咨询健康信息服务的方式。学科馆员的咨询领域划分方式有所不同，例如哈佛大学从本科生常用学科、统计学划分标准、院系设置和专题项目四个维度细分了学科馆员负责领域。一级学科划分最详细的是加州大学洛杉矶分校图书馆，共 30 个医学细分学科馆员，最宏观的是斯坦福大学图书馆，仅有 1 人为生物学学科馆员。学科馆员具有专职和兼职两种状态，例如纽约大学图书馆的健康科学和生命科学馆员各 1 人，专门负责相应领域的健康信息参考咨询，约翰霍普金斯大学图书馆则采用兼职馆员政策，负责生物技术学科馆员同时兼职了环境学、教育学和工程学等 3 个学科的咨询馆员。

表 3.5 调查对象健康信息参考咨询服务

机构名称	医学学科馆员一级学科领域划分	医学学科馆员人数	线上咨询	到馆咨询

哈佛大学图书馆	生物工程，生物与进化生物学，分子和细胞生物学，干细胞与再生生物学	4 人	有	有
约翰霍普金斯大学图书馆	生物技术，化学与生物分子工程，环境与健康工程，生物伦理学，生物学，生物医学工程，公共卫生研究	4 人	有	有
纽约大学图书馆	健康科学，生命科学	2 人	有	有
斯坦福大学图书馆	生物学	1 人	有	有
加州大学旧金山分校图书馆	基础科学，生物信息学和统计学，临床医学，公众健康信息素养，口腔医学，循证医学，公共卫生，医学史，LGBT 人群健康，医学（一般问题），护理学，药学，物理疗法，医学行政管理，系统评价	7 人	有	有
宾夕法尼亚大学图书馆	健康科学	10 人	有	有
加州大学洛杉矶分校图书馆	生物工程，生物化学，生物数学，生物统计学，化学和生物分子工程，化学和生物化学，临床医学，社区卫生科学，生态与进化生物学，环境卫生科学，流行病学，表观遗传学，基因调控，遗传学和基因组学，卫生政策与管理，医学史，人类遗传学，整合生物学和生理学，微生物学、免疫学和分子遗传学，分子和医学药理学，神经生物学，神经科学，护理学，生理学，精神病学和生物行为科学，心	11 人	有	有

	理学，病毒学			
圣路易斯华盛顿大学图书馆	生物学，医学史，药学，心理与脑科学，公共卫生	4 人	有	有
杜克大学图书馆	生物学，进化人类学，全球健康，健康科学，医学史	3 人	有	有

3.3 信息素养教育资源开放程度

本文对 9 所美国医学院（校）图书馆信息素养教育资源的开放程度进行了调查，结果表明，被调查医学院（校）图书馆的线上培训与讲座大多数面向校内师生群体，在进入资源获取页面时需要身份验证，这也就使得相关类型的信息素养教育资源基本是部分校外访问状态。在梳理信息素养教育资源的开放程度的同时，提取出网页明确标注面向公众开放的信息素养教育资源，共有 3 处。

宾夕法尼亚大学图书馆接受宾州社区成员的公共研讨会、定制会议、小组培训和一对一咨询的申请，培训内容包括工具软件使用培训（音视频软件、Excel、网页设计、文献管理软件等）和教师信息素养教育（教学设计与作业设计）[8]。加州大学洛杉矶分校图书馆面向公众提供即时研究指导（Drop-in research assistance）[9]，各个分馆都有咨询专业的划分，用户可以通过电子邮件或者到馆完成咨询。圣路易斯华盛顿大学图书馆为经过授权的公众提供除法律、医学和社会工作之外的所有学科的技术支持服务[10]。社会公共捐助 100 美元及以上，就可以在一年期限内使用杜克大学图书馆的在线研究指导服务[11]，通过即时咨询、邮件咨询、电话咨询的方式向学科馆员寻求帮助。

4 总结与启示

4.1 把全文获取能力作为信息素养教育的切入点

Web2.0 时代，用户获取信息的渠道更加多元，图书馆作为信息资源提供者的竞争力受到挑战。医学院（校）图书馆应当及时调整思路，一方面除了继续提供优质资源与服务吸引用户使用图书馆，另一方面为用户获取全文指导途径，包括文献传递、搜索引擎、开放获取资源等。例如约翰霍普金斯大学图书馆就专门对谷歌学术的使用进行了专题页面介绍，从全文获取视角提升学生的信息素养。

4.2 增强信息意识与信息道德教育

信息意识、信息知识、信息能力、信息道德都是信息素养的组成元素，医学院校图书馆应当结合教学、培训、宣传等方式，普及信息素养概念与内涵，增强读者信息意识，提高他们对信息的关注度和敏感度，促进读者对于信息素养教育的关注程度与参与程度，从而扩大信息素养教育受众群体、改善信息素养教育效果。

在调查中，9 所美国医学（院）校图书馆均开展了这方面的培训。加州大学旧金山分校图书馆围绕科研全流程的信息道德问题，开展了参考文献引用、避免抄袭剽窃、数据引用与数据分享等主题的信息道德培训。宾夕法尼亚大学图书馆面向教师、学生和数据管理提供不同的信息道德培训。医学院校图书馆应当通过读者培训等方式加强信息道德教育，开展包括知识产权、合理引用、公有领域信息引用、“孤儿作品”使用等相关内容的教学与普及与政府部门、科研机构、知识产权与专利机构展开合作，邀请专家进行讲座培训，进一步提升读者的信息道德意识与信息道德知识。

4.3 开展健康信息素养教育

作为汇聚医学信息资源与服务的图书馆，医学院校图书馆有能力也有责任开展好健康信息素养教育。调查结果发现，被调查的美国医学院（校）图书馆信息素养教育除了传统的文献检索课程及培训、图书馆资源与服务利用指导、工具软件使用培训，还开展数据管理、信息道德与知识产权、教师信息素养相关的培训，满足医学生、教职工在学习、教学、科研方面的各类培训需求，围绕生物信息检索、药学检索、医学专利检索等主题开展了健康信息素养教育培养。我国医学院校图书馆也应当面向医学生、医护人员和公众开展不同层次的健康信息素养教育，培育个人对健康信息的关注度与敏感度，以及获取、分析、传播健康信息能力的提升。

PU-062

基于舆情视角下的重大突发公共卫生事件引发的信息疫情防控模式研究——以新型冠状病毒肺炎疫情舆情防控为例

刘帅
安徽省医学情报研究所（安徽省医学会）

目的 为深入贯彻习近平总书记关于疫情防控及舆情治理的重要论述精神，从舆情视角探讨重大突发公共卫生事件引发的信息疫情防控模式，开拓医学信息学应用新思路。

方法 论述了相关基础理论，以新型冠状病毒肺炎疫情期间的舆情防控情况为例，分析了信息疫情产生的过程及当前存在的问题，据此提出了舆情防控模式和相应运行保障机制。

结果 该模式重点聚焦信息疫情，及时预警原生舆情，着力避免次生舆情。

结论 对于重大突发公共卫生事件引发的信息疫情舆情防控有一定参考意义。

习近平总书记多次就疫情防控中的舆情治理强调，要增强舆情引导的针对性和有效性，加强舆情跟踪研判，主动发声，正面引导，强化融合传播和交流互动，让正能量始终充盈网络空间。当前正处于新型冠状病毒肺炎防控的关键时期，在此过程中，“互联网+”环境下各类疫情信息在畅通公众信息了解渠道、助力疫情防控阻击战的同时，也应注意到海量信息中存在虚假信息、错误信息等误导群众，一定程度上会引发“信息疫情”，进而诱发一系列“次生舆情”。今年 2 月，

在世界卫生组织召开的新冠疫情创新研究会上，“信息疫情”成为会议聚焦的一个新名词，其具有全球传染性，传播速度极快，内容参差复杂，受众面广、危害性大，现已被世卫组织和各国政府重点关注。《麻省理工科技评论》指出，新冠病毒是第一个真正的社交媒体“信息疫情”。互联网是一个没有硝烟的战场，舆情的引导和防控深刻影响到公众的认知、政府公信力以及国家意识形态领域的阵地建设等方面。因此，如何通过舆情防控此类重大突发公共卫生事件引发的信息疫情值得深入思考，此类研究既是对现有医学信息学、情报学理论的拓展和补充，同时能助力健全公共卫生应急管理体系、完善重大疫情防控体制机制，具有理论和实践双重意义。

1 相关理论

1.1 信息疫情

在2003年SARS爆发期间，David Rothkopf首次提出“信息疫情”(Infodemic)的概念，它是指一些事实，加上恐惧、猜测和谣言，被现代信息技术在世界范围内迅速放大和传递，以与根本现实完全不相称的方式影响了国家和国际的经济、政治甚至安全的现象。世卫组织全球传染病防范主任Sylvie Brian博士在此次新冠肺炎疫情防控中再次提及“信息疫情”，她认为过多的信息（有的正确，有的错误）反而导致人们难于发现值得信任的信息来源和可以依靠的指导，甚至可能对人们的健康产生危害。

1.2 次生舆情

目前学界尚未统一定义次生舆情，与“原生舆情”“原始舆情”的定义概念相对应，与“舆情次生灾害”“二次舆情”“衍生舆情”等定义概念相类似，通常是指在原有网络事件、网络舆情的基础上被诱发衍生而来的新议题。与原生舆情相比，次生舆情同样具有危害，且针对性更强、爆发性更剧烈，揭丑、放大、施压等各种舆论功能也更突出。一旦负面次生舆情大量扩散，公众极易出现情绪化、非理性的问题，可能会引发围观起哄、造谣传谣、舆论审判、谩骂污蔑和侵犯隐私等甚至更极端的情况，从而导致社会流瀑效应和群体极化效应，更不利于事件或问题的真正解决。

1.3 沉默的螺旋和塔西佗陷阱

沉默的螺旋是指，为舆论主流的观点会产生放大效应，非主流观点则相对“沉默”，而“沉默”的一方会促使另一方更加强势，如此循环往复，形成一方越来越壮大而另一方越来越沉默的两极螺旋发展。塔西佗陷阱是指，指当政府部门或某一组织失去公信力时，无论说真话还是假话，做好事还是坏事，都会被认为是说假话、做坏事。

2 过程分析

从2019年12月至2020年6月底，新冠疫情爆发的信息洪流中存在疫情恐慌、热搜体、“污名化”、“正能量”谣言、“政治化”等多种形式的信息疫情，并形成了原生的新冠肺炎疫情舆情和一系列次生舆情，比如“韩红基金会被举报”及“红十字会信任危机”等网络舆情事件就属

于次生舆情。大致可分为三个阶段加以分析：

2.1 积聚期：信息饥渴

在疫情传播较早时期，公众在事态初期处于“信息饥渴”状态，权威部门和专家未及时发布“人传人”等危机预警信息，政府部门、疾控机构以及医疗机构没有更早重视病毒的传播危机，未及时发布权威信息，公开透明回应群众关切。随后，社交媒体各种关于疫情的言论、推测逐渐增多，出现了谣言肆掠、信息真假难辨、公众无所适从的现象。信息严重不对称，真实信息和虚假信息失衡，此时容易出现沉默的螺旋效应，相信虚假信息的人群多于相信官方报道的人群。

2.2 爆发期：信息爆炸

病毒广泛传播后，社交媒体的信息芜杂，微信、微博、各类 App 等移动传媒“信息爆炸”，谣传疫情严重造成民众恐慌。夸大药品偏方效用，物价遭到哄抬，形成抢购风波，物资诈骗也层出不穷。同时，由于权威疫情信息发布平台的话语权主导功能缺失、信息发布弱势，专家学者被诟病质疑，政府未能及早树立公信力和号召力。网民对武汉市政府防控措施质疑，对湖北省红十字会的信任危机和抗疫形式主义等事件上爆发不满。政府开始采取高规格防控措施之后，通过官方和其他可靠渠道公布疫情信息、辟谣、释疑，但此时如果处置不当，可能陷入塔西佗陷阱。

2.3 稳定期：信息过载

随着疫情防控推进，全国确诊病例逐步减少,清零城市逐渐增多，经过政府与社会的干预与共同努力，舆情防控逐步进入有序化，舆情治理起到了正向的引导作用，中国式抗疫和制度优势得到认可。这一时期，原生舆情和次生舆情力量此消彼长,沉默的螺旋与塔西佗陷阱可能同时存在，且在引导舆情方面可能是有利的，官方占据主流观点，民众起到监督作用。同时，病患收治、医疗服务、志愿服务、慈善捐助、交通出行、生活用品等各类信息充斥网络空间，防疫物资交流、志愿者服务、捐赠信息服务等各种交流平台争相出现。民众看到了如原黄冈卫健委领导在媒体见面会中一问三不知的状况，也看到了全程脱稿、数据明确的市长受到广泛赞扬的情况。之后，负面舆情逐渐呈现退潮趋势，正能量事件和评论呈现有利态势。

纵观全程，信息疫情出现了如下特征：**一是短期集聚，迅即生成。**官方媒体、权威专家、微博“大 V”、网民同时发声，微博、微信、抖音、知乎等多个平台实时推送，疫情信息呈现出“信息饥渴—信息爆炸—信息过载”的动态生成过程。**二是辐射扩散，二次叠加。**疫情事件一旦触发公众关注，各个网络平台的传播扩散速度极快。网络舆情呈现出裂变式的传播态势和辐射式的扩散效应，进而可能引发次生舆情，不断地被再解读、再传播、再扩散。一些情况下，辟谣速度远不及谣言的生产速度，且有时辟谣甚至还会被二次定义为“谣言式辟谣”，小道消息病毒式传播，被过度解读、无限放大，导致公众判断力下降，恐慌焦虑情绪蔓延。**三是波及面广，影响力大。**疫情舆情一经形成，波及范围由此迅速扩大，线上线下双向联动，事件影响力不仅在网络空间蔓延、扩散和叠加，也会波及现实环境，进一步扩大疫情事件的影响力。

同时，也反映了如下问题：**一是信息公开不及时。**官方媒体发布信息与民众需求期待存在滞后或错位的问题,部分主流媒体被动追随舆情走向。**二是处置不到位。**未能及时发现产生信息疫情

的苗头，在预警监测、快速反应上存在不足。**三是监管不力。**多次出现监管过度或缺失的现象，一些相关部门尚不能合理地监管网络舆情,在应急处理、危机公关方面有待进一步改善。

3 模式构建

3.1 舆情防控模式

根据此次新冠肺炎疫情相关情况，笔者认为在重大突发公共卫生事件引发的信息疫情舆情防控方面，需要保持整个舆情防控流程的衔接，包括在快速反应、应急处理、危机公关、分析总结等四个环节层层递进，有效做到全程发力（如图 1 所示）。

3.1.1 快速反应

即在舆情积聚期重点关注网络舆情、媒体舆情、社会舆情，做到实时监测预警，持续动态追踪，有效处理原生舆情，防止事态扩大。通常来说，需要建立舆情监测平台，广泛收集网络舆情信息，建立预警等级和敏感信息库。针对敏感信息在网络上的出现频率及搜索搜索和评论热点，注意划分地域和群体。全方位舆情监测平台的建立，也需要多方协同，政府部门，公益性信息机构，营利性信息机构共同参与建设，相互补充。保持媒体通联，收集媒体信息，注意媒体反馈。搜集整理各种投诉、举报、意见反馈。

3.1.2 应急处理

及时信息公开，公布权威消息，释疑解惑辟谣，规避次生舆情。梳理需求，收集整理群众最关心最迫切的问题，回应群众关切。及时发声、主动发声,抢占舆论制高点,抢占沉默的螺旋的先机，引导社会舆论处理社会负面影响和社会恐慌的危机。主流媒体应及时介入，将事件产生、发展和状态等相关信息尽量在第一时间发布，增强信息的公开透明。此外，要加强与传统媒体、网络意见领袖等联系，全面协同和整合信息获取渠道和信息发声平台，使权威真实信息能够广泛传播和全方位无死角覆盖。提前做出防控预案，并根据预案解决此类问题。

3.1.3 危机公关

重视塔西佗陷阱，谨慎处理政府信任危机，消除负面影响，解决次生舆情。应及时识别区域的潜在风险，分析关于信息疫情事件爆发和扩散的相关情况和统计数据，有关部门和主流媒体进一步加强对相关信息的发布力度，及时回应民众的需求要点和现实关切。对政府部门和官员一些失职现象，政府部门积极面对和承担责任，坚决处理。并团结主流媒体、意见领袖等，引导众多网民的言论朝着有利于政府形象重塑的方向转化，助力扭转舆情危机。

3.1.4 分析总结

开展总结评估，修复舆情创伤。组织对网络舆情进行全面排查，发现潜在问题并消除舆情隐患。在监测预警基础上，持续跟踪网民的评议、关注和意见，及时反馈，设法与民众开展良性互动，及时满足民众的合理诉求，提升政府公信力和民众的信任度。此外，还应组织专家学者和专业技术人员对网络舆情治理工作开展全方位多层次的系统评估，对整个舆情事件的产生原因、特点、状态和防控的经验教训等进行总结梳理，从而为类似重大突发公共卫生事件的网络舆情治理

和应急管理体系的建立与完善提供借鉴和参考，提升政府和公众应对重大突发公共卫生事件的防御力。

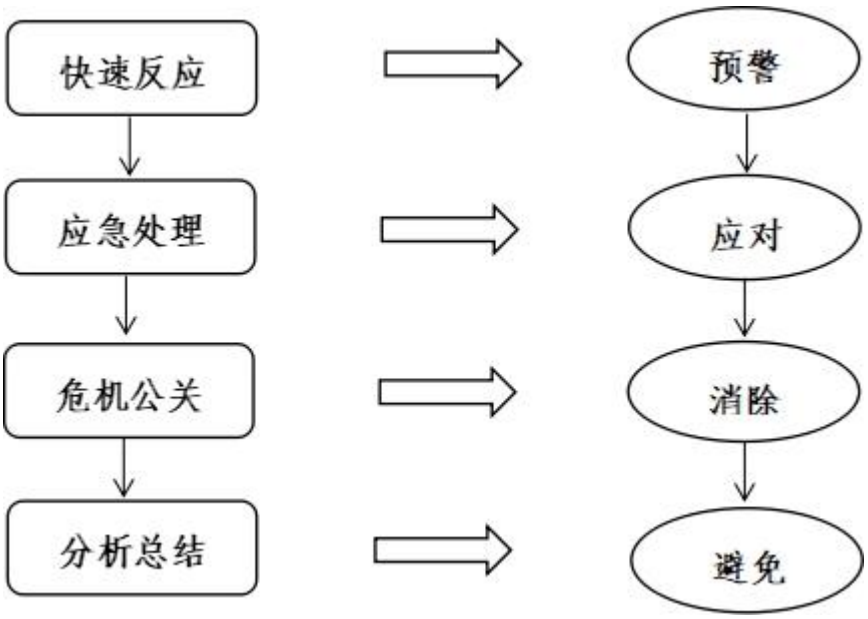


图 1 舆情防控模式图

3.2 运行保障机制

3.2.1 宣传教育

此次信息疫情爆发的根源还是在于公众的信息鉴别能力不足,容易引发恶性舆论连锁反应。“互联网+”时代下，信息传播速度非常快、信息庞大复杂、涉及面广、受众人群众多，需要着重加强全社会舆情防控教育，提高公众不传谣、不信谣的意识，主动地判断信息的真实性以及注意传播源，对于身边的信息保持警惕，注意辨别。政府部门、媒体、意见领袖、网络“大 V”等能够发挥正向作用，传播真实信息，弘扬正能量。

3.2.2 立法支持

2020 年 3 月 1 日，已实施《网络信息内容生态治理规定》。但应对重大突发性网络舆情，特别是根据此次疫情，重大突发公共卫生事件引发的舆情，其相关配套法律法规仍有待健全,舆情引导有法可依尚存在漏洞。各级政府部门舆情应对机制标准不够统一清晰，地方相关法规也不健全，各方在具体应对舆情时存在相互掣肘问题。因此，保障舆情防控模式有效运行，需要进一步健全舆情法律法规，特别是完善网络安全和网络生态治理法律法规。

3.2.3 深化监管

目前，政府部门在应对网络舆情危机采取措施的过程中，缺少成体系的机制进行严格监督。所以，建立舆情危机治理监督体系，有利于政府在治理过程中不断进行改进和修正。监督体系的建立可考虑从内部监督和外部监督进行，内部监督即为政府部门对网络舆情危机治理的自我监督，以上级主管部门为监督主体，对下级网络舆论危机治理部门进行监督和考核，同时要建立独立的信息反馈系统。外部监督可以考虑从加强网络舆论的司法监督、媒体监督和公众监督几个方面着

手，建立专门的监督平台，在舆情事件爆发时，立即启动监督机制，至舆情危机进入恢复期结束。

今年是“十三五”收官之年，同时也是全面建成小康社会、决战决胜脱贫攻坚的关键之年。新型冠状病毒肺炎疫情是一次“大考”，也一定程度上反映出我国在应对重大突发公共卫生事件方面存在不足。以此次新型冠状病毒肺炎疫情舆情防控为例，说明打赢疫情防控阻击战不仅是与病毒的正面搏斗，还需要警惕各类信息病毒的传播。应对重大突发公共卫生事件要格外注意信息疫情，需进一步健全舆情防控体系，完善舆情防控模式，配套运行保障机制，真正在快速反应、应急处理、危机公关、分析总结等环节取得良好效果，有效监测预警，从容应对，消除影响，避免预防，从而提高信息疫情免疫力，及时消除信息病毒危害。

PU-063

基于 CiteSpace 的我国订单定向医学生研究热点和趋势分析

魏艳君、牟燕、宋奎勐
山东省医药卫生科技信息研究所，山东第一医科大学（山东省医学科学院）

目的 基于文献计量分析方法系统地梳理我国订单定向医学生相关的研究热点及研究趋势。

方法 以中国知网引文索引（CNKI）数据库收录的“订单定向医学生”主题论文为数据源，利用 CiteSpace 分析软件，从时间分布、作者与机构分布、研究热点分布与演进趋势等方面归纳总结我国订单定向医学生研究领域的研究现状。

结果 我国订单定向医学生领域的研究呈现发文量增长较快、学术集群现象明显、研究对象相对单一、研究由理论向实证转变的演进趋势。

结论 订单定向医学生作为为基层培养的急需的全科医学人才，对均衡医疗资源分布、提升卫生服务的公平可及、促进全民健康具有重要意义。未来应加强该领域学者和研究机构间的交流合作、丰富研究主题、提升学术影响力，为促进我国培养更加优质高效的基层卫生人才提供依据。

PU-064

贵阳市第一人民医院基于电子签名的医院无纸化应用探讨

邵建华
贵阳市第一人民医院

目的 数字化医院是我国现代医疗发展的新趋势，随着医院电子签名系统的实施，信息系统的不断完善，医院无纸化时代已经到来。医院电子病历实施无纸化，涉及的科室较多，业务场景复杂、技术难度大，本文以贵阳市第一人民医院为例，详细阐述了医院基于电子签名的无纸化应用实践，

从医院的检验检查、处方、住院病历、患者及家属的知情同意书等不同维度实现电子签名的流程为例进行阐述，探索医院无纸化建设的难点和重点。

方法 本文以贵阳市第一人民医院为例，详细阐述了医院基于电子签名的无纸化应用实践，从医院的检验检查、处方、住院病历、患者及家属的知情同意书等不同维度实现电子签名的流程为例进行阐述

结果 我院从 2014 年开始实施电子签名系统，初步实现了全院无纸化，电子签名系统的实施，从技术角度解决了电子病历的合法性问题，同时优化了患者的就医流程，有效的解决了医院看病难的问题，提升了患者就医满意度，同时病历无纸化的实施在一定程度上降低了医院运行成本，简化了病历管理流程，提高了医院运行效率，很好的提升了医院的整体形象。

结论 医院实现“无纸化”本身并非目的，我们从最深层次挖掘，医院无纸化其实是一所医院信息化成熟到一定程度的副产品和重要标志，是规范应用电子病历系统、电子签名等多种信息技术之后，水到渠成的结果。无纸化，表面上看，是医务人员桌面上纸张文件的减少、诊疗过程中的纸质开单消失、以及医院纸张、打印耗材用量的减少；无纸化的背后，实则是一所医院管理水准所达到的新高度。无纸化，对于一所医院的最大产出结果应该是：规范诊疗行为、提高医疗质量和患者安全。

引言

随着医院的发展和就医人数的增加，纸质病历给医院的管理带来诸多不便，如库房存储空间不足，借阅查阅困难等，这些问题均可以使用无纸化来解决，使用无纸化可以取消病案的纸张打印，解决了病案储放空间越来越紧张的问题，并节省大量的纸张和相关耗材，病人查阅资料，无需纸质复印，现场直接打印，医生可以利用网络直接进行查阅病案相关信息，减少了病案收回、整理、装订、复印和借阅等环节的流程。无纸化已成为医院信息化发展的必然趋势。当前大多数医院病历档案管理采用“双轨制”管理模式，即电子病历档案与纸质病历档案并存的一种档案归档管理模式。该模式导致病历管理成本的增加，纸张、硒鼓、病历架等耗材的浪费，医护、医技人员工作的冗余。

医院无纸化，必须解决各类医疗文书的来源可靠、数据真实、信息完整等问题。国家为了促进电子签名的发展，相继颁布了一些列法律法规，2005 年颁布了《电子签名法》，其中第十四条规定可靠的电子签名与手写签名或者盖章具有同等的法律效应。2010 年卫生部颁布了《卫生系统电子认证服务管理办法（试行）》，2017 年卫生健康委颁布了《电子病历应用管理规范》，其中第十条强调有条件的医疗机构电子病历系统可以使用电子签名进行身份认证，可靠的电子签名与手写签名或盖章具有同等的法律效力。 国家及省级电子认证相关法律法规的陆续出台，为医院无纸化提供了有力的法律支持;卫生行业电子认证规范文件的逐步制定，为医院无纸化提供了明确的的业务指导，医院无纸化时代已经到来。

一、医院背景

贵阳市第一人民医院始建于 1919 年 7 月，经过百年的奋斗与发展，现已成为集医疗、科研、教学、预防、保健、康复和急救于一体的市属首家三级甲等综合医院，院年门诊量约 68 万人次，年出院 3.2 万人次，医院信息化起步较早，在 2005 年就已经建立了全院的信息系统，到 2013 年实现了全院信息系统的互联互通，同时我院也是贵阳市首家最早实现全院电子签名的医院，医院在 2014 年就上线了全院电子签名系统，随着电子签名系统在全院的铺开，医院领导为了打造数字化医院，节约医院成本，改善患者就医体验，决定推进医院无纸化建设。

二、医院无纸化整体需求和总体方案

2.1 整体需求

2.1.1 过程签名电子化

实现全部医疗过程记录的电子化，包括电子化医嘱、电子化病程记录、电子化检查报告、电子化的知情同意书等。

2.1.2 归档举证电子化

实现病案归档的电子化，包括已有纸质文件的电子化、电子化查阅、电子化封存、电子化举证等。

2.1.3 电子签名关键点

电子病历是指医务人员在医疗活动过程中,使用医疗机构信息系统生成的文字、符号、图表、图形、数据、影像等数字化信息，在整个医疗过程中需要实现电子签名的地方较多，其中比较关键的地方有：病案首页、入院记录、手术记录、出院记录、知情同意书、会诊记录、影像报告、检验报告、常规心电图、其他检验单、入院证、门诊病历、病程记录、长期医嘱、临时医嘱都需要进行电子签名，另外还包括患者及家属的知情同意书类、护理文书类等。

2.2 总体方案

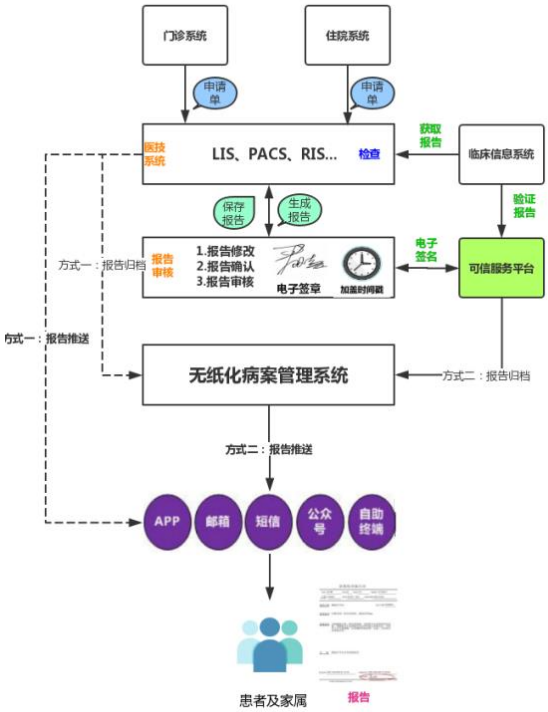
医院应用系统较多，涉及需要电子签名的部门也较多，大致可以分为医护、医技科室签名、患者及家属签名、病案管理签名、以及调查取证管理等，所以医院总体方案图如下：

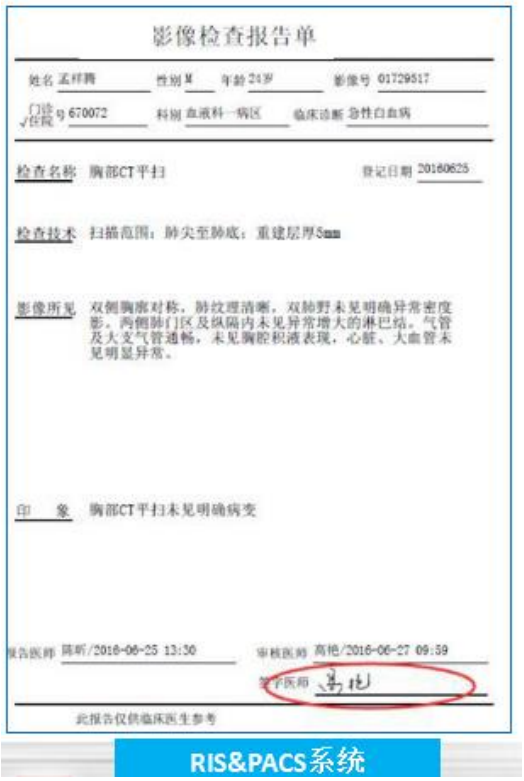
电子签名总体方案图

三、医院各应用场景及实现方案

3.1 医技科室无纸化实现

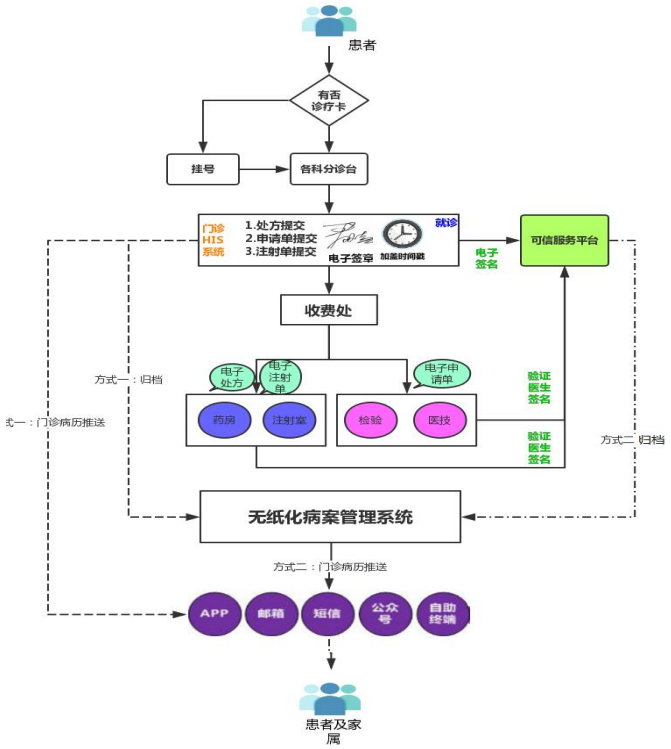
医技科室电子签名是实现门诊无纸化的关键，主要实现流程是医生（包括住院和门诊）在医生工作站填写检查申请单（医生实现电子签名）→患者缴费→检查科室检查→报告书写（电子签名）→报告审核（电子签名）发布→患者在 APP 或公众号上查看报告，医生也可在医生工作站查看报告或影像，具体实现流程图如下：

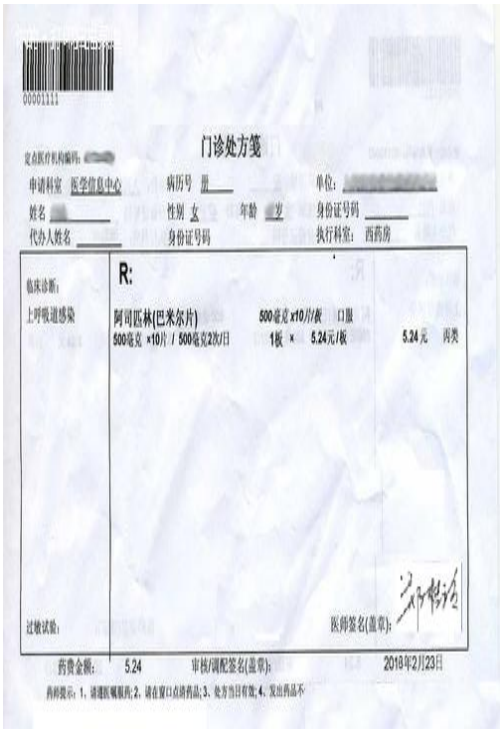




3.2 门诊处方及门诊病历无纸化实现

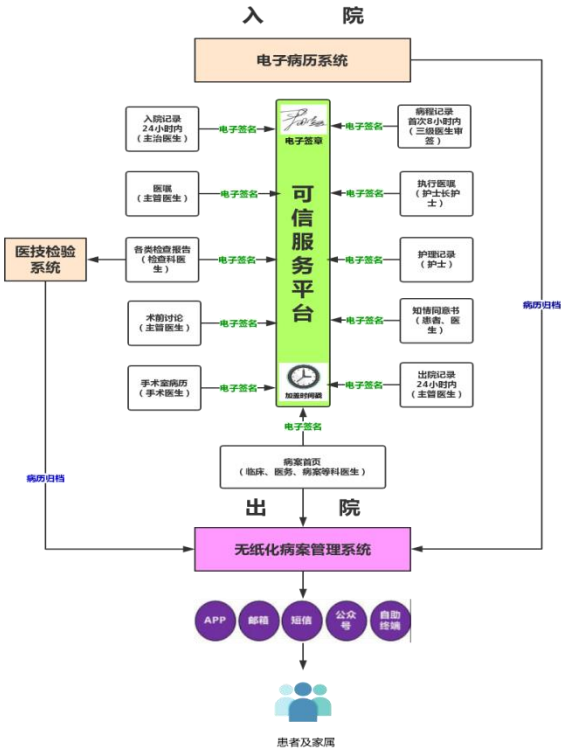
门诊医生电子签名是门诊无纸化的关键，主要实现流程是门诊医生开具处方（完成电子签名）→患者缴费→药房刷卡取药。医生在门诊医生工作站书写门诊病历后完成电子签名，患者可以在 APP 或者医院公众号上查看门诊病历和电子处方。

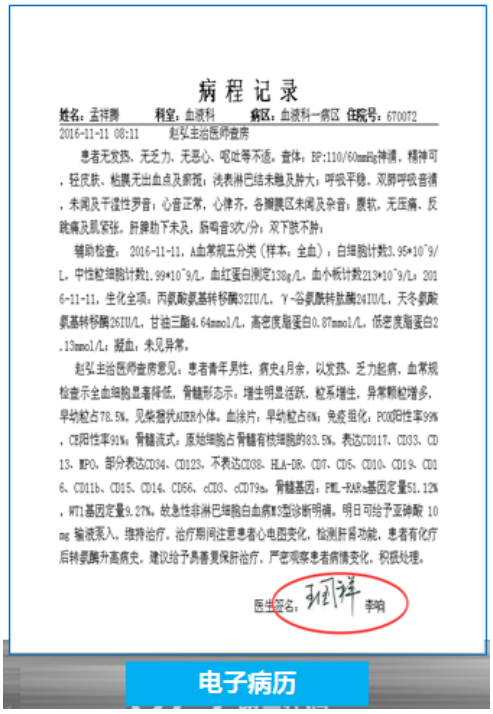




3.3 住院电子病历无纸化实现

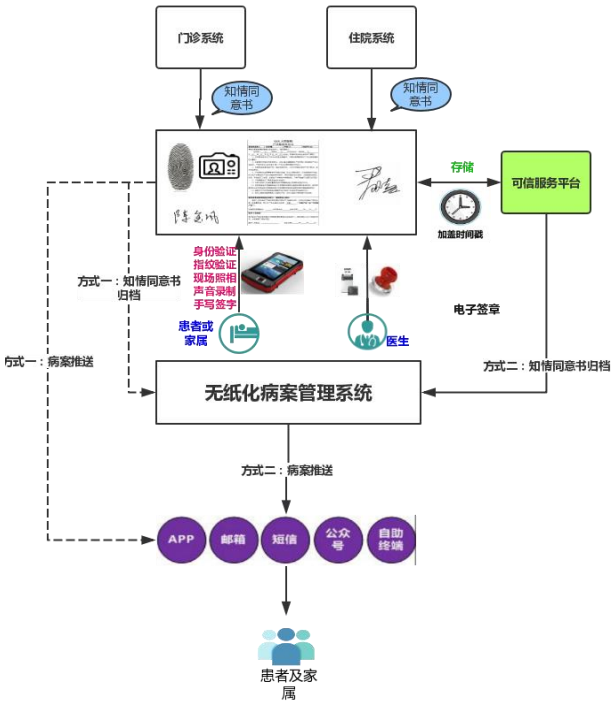
住院病历无纸化是医院无纸化实施的重点，由于很多病历具有双签名或者三签名的情况，这就要求我们在签名权限上进行控制，医院所有的病历具有修改留痕可追溯功能，当上级医师签名之后下级医师不能更改。具体实现流程如下：





3.4 医患沟通类无纸化实现

医患沟通类无纸化必须实现患者电子签名。这是全院无纸化的难点，患者和医护人员相比具有不特定性，这就要求在每个病区以及手术科室提供手写签名设备指纹采集设备，结合 CA 证书签名机制，实现一种创新模式的患者知情同意书的无纸化解决方案，并使得电子知情同意书具有法律效应，具体实现流程如下：



四、应用结论

我院从 2014 年开始实施电子签名系统，初步实现了全院无纸化，电子签名系统的实施，从技术角度解决了电子病历的合法性问题，同时优化了患者的就医流程，有效的解决了医院看病难的问题，提升了患者就医满意度，同时病历无纸化的实施在一定程度上降低了医院运行成本，简化了病历管理流程，提高了医院运行效率，很好的提升了医院的整体形象。

五、思考

医院实现“无纸化”本身并非目的，我们从最深层次挖掘，医院无纸化其实是一所医院信息化成熟到一定程度的副产品和重要标志，是规范应用电子病历系统、电子签名等多种信息技术之后，水到渠成的结果。无纸化，表面上看，是医务人员桌面上纸张文件的减少、诊疗过程中的纸质开单消失、以及医院纸张、打印耗材用量的减少；无纸化的背后，实则是一所医院管理水准所达到的新高度。无纸化，对于一所医院的最大产出结果应该是：规范诊疗行为、提高医疗质量和患者安全。

PU-065

山东省病原微生物实验室及实验活动备案管理系统的研发与应用

席敏[1]、段永璇[2]
1. 山东省医药卫生科技信息研究所
2. 山东第一医科大学现代教育信息中心

目的 目的：根据国家病原微生物实验室生物安全管理规定，结合山东省病原微生物实验室的实际情况，开发此山东省生物安全实验室备案管理系统，通过该系统将病原微生物实验室备案工作进行网络化集成，实现病原微生物实验室备案活动的数字化和流程化管理，实现了在线远程管理，提高备案工作效率。

方法 方法：采用 B/S（Browser/Server，浏览器/服务器）模式，利用 ASP.NET、Access 数据库等技术，开发山东省生物安全实验室备案管理平台

结果 结果：通过试用和实际应用，证明了该平台能满足备案工作的实际需求，且易于维护，在相关政策支持下，目前该平台已推广至山东省 17 地市，实现了 100%覆盖率。

结论 结论：该平台的研发意义重大，平台的成功推广及普及促进了全省病原微生物实验室安全管理工作模式转变，为生物安全信息化管理奠定了基础。

一、 引入

新冠肺炎疫情的发生对国际公共卫生造成严重危机，生物安全这一概念又一次引发关注。作为当今世界快速发展的新兴经济体和生物多样性大国，我国面临的生物安全形势尤为严峻。实验室安全作为生物安全的重要组成部分，它是指避免实验室危险生物因子造成实验室工作人员暴露，向实验室外扩散并导致危害的综合措施[错误!未找到引用源。]。科学高效的实验室安全管理对规范生物技术活动、保障公众健康乃至经济发展、社会稳定都有十分重要的意义。

早在 1983 年，WHO（世界卫生组织）就提出了生物安全问题在国际事务中的重要性，提出了“实验室生物安全手册”（第一版），将传染性微生物根据其致病能力和传染的危险程度等划分为四类，即 BSL 1-4。在该手册中，WHO 鼓励各国针对自己国内的微生物实验室制定相应的生物安全条例，该手册目前已有第 2、3 版**错误!未找到引用源。**。我国在实验室生物安全管理领域起步比较晚，直到 2004 年 11 月份才有首个相关管理条例《病原微生物实验室生物安全管理条例[错误!未找到引用源。]。直到 SARS 爆发后，国家逐步认识到生物安全管理的重要性，陆续出台了《实验室生物安全通用要求》（GB19489-2008）（以下简称《要求》）、《病原微生物实验室生物安全管理条例》（国务院令第 424 号）（以下简称《条例》）、《可感染人类的高致病性病原微生物菌（毒）种或样本运输管理规定》及《人间传染的病原微生物名录》（以下简称《名录》）等规范、标准和法律法规，以加强实验室生物安全管理。按照规定，所有涉及病原微生物的检验、教学、科研机构都要加强实验室生物安全管理。为响应国家政策，全国各省市也积极开展病原微生物实验室及

实验活动备案工作，但是传统的实验室备案需要提供大量纸质材料，然后上交给审核部门，相关部门审核后再反馈……整个过程耗费人力物力，效率低下，无法及时实现信息的共享。随着信息技术的发展，为实验室安全备案工作提供了助力，全国多个省市和地区根据各自的工作特点和实际情况自主开发了生物安全实验室备案管理系统[错误!未找到引用源。]，极大地提高了实验室备案效率。本系统旨在自主研发山东省生物安全实验室备案管理系统规范实验室管理，提高备案时效和生物安全实验室信息化管理水平。

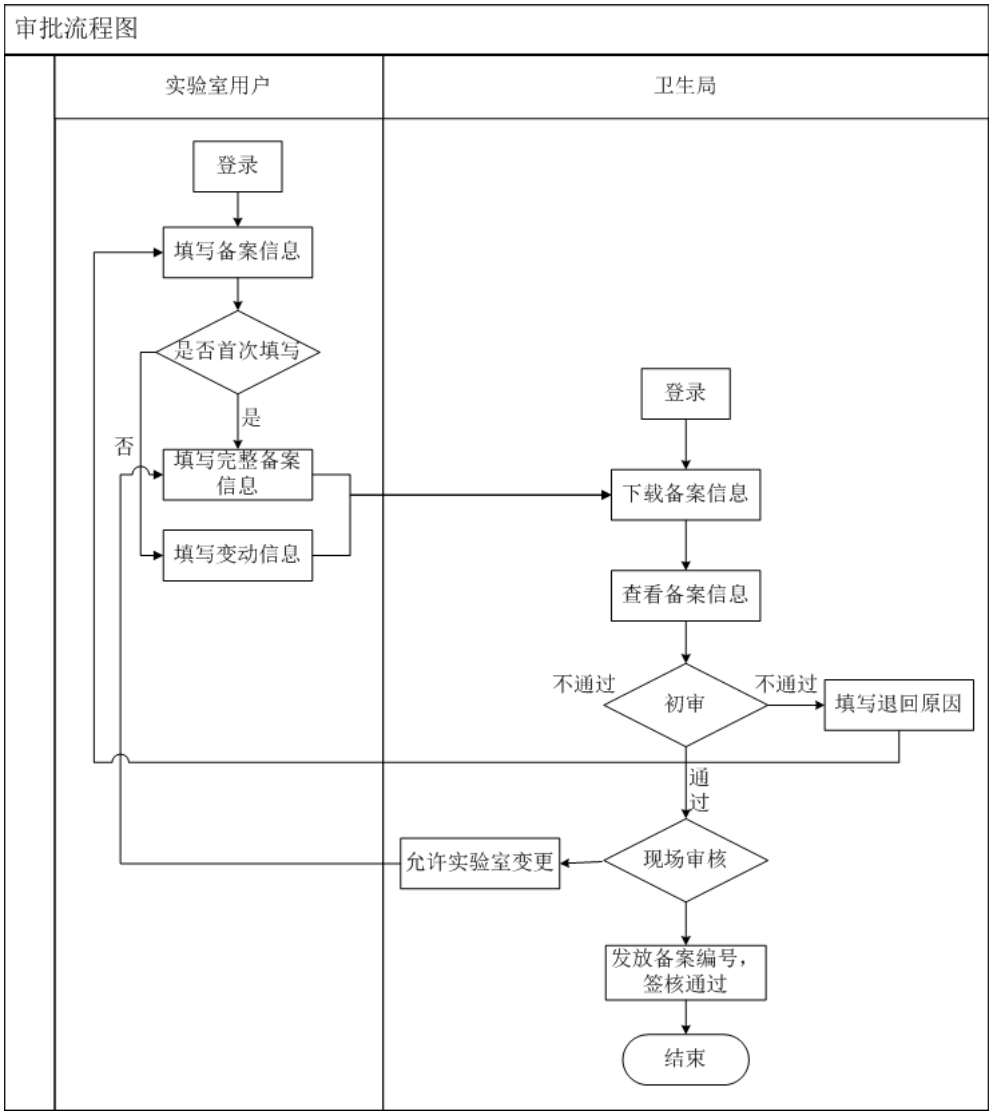
二、 平台功能分析

本系统主要针对以下四类用户，包括医疗机构微生物实验室、地市卫生主管部门、省级主管部门和平台管理员。实验室申请单位通过平台填报单位和实验室信息、在线填报实验室及实验室活动备案申请，各级卫生行政部门对实验室申报单位提交的备案信息进行审核，省卫生局负责完成最终审核并可以查询统计全省实验室备案信息。主要功能涵盖了微生物实验室备案申报、属地化审批、分级管理等业务流程，还有信息自动检测、审核意见及进度反馈、多文件上传、自动归档、多条件综合查询、格式审查、批量处理以及信息统计等功能。

2.1 备案流程设计

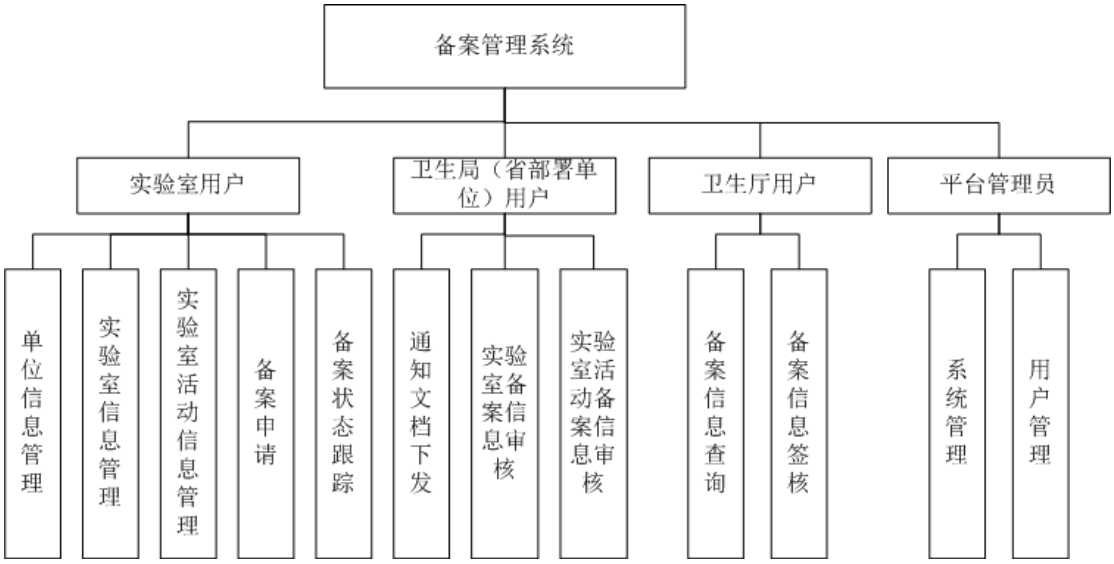
实验室所在的医疗卫生机构注册、登录平台后，第一次备案需要填写完整的实验室信息和实验活动信息，然后提交备案申请，如果是在原来申报的基础上更改或者更新，只需提交变动信息。备案信息提交后，医疗卫生机构只能查看当前状态，无法对其修改。地市卫生主管部门对已提交的备案信息进行审核，审核未通过的实验室，将退回其备案申请，并填写退回原因及整改意见，实验室用户整改后重新提交。

对不符合规定的申请材料，卫生行政部门及时通知申请单位，要求在 10 个工作日内补正，逾期未补正或者经补正仍不符合规定的，不予备案。实验室的基本信息、负责人、实验活动等与生物安全相关的重大事项发生变更（含新增）时，应于变更之日起 30 个工作日内向原备案卫生行政部门提交备案变更说明书。审核通过的实验室，审核单位将组织专家对其进行现场查验，签核通过后发放备案编号。对于不符合备案要求的实验室，不予备案。山东省病原微生物实验室及实验活动备案证明备案编号有效期 5 年。需要继续从事病原微生物实验活动的，应当在有效期满前 3 个月按照本规定重新申请备案。



2.2 平台功能模块分析

实验室用户作为申请单位，可以进行单位信息管理、实验室信息管理、实验室人员管理、委员会成员管理、实验室及实验活动备案以及备案结果查询等操作。审核单位可以通过系统对生物安全实验室用户填报的备案信息进行审核，在线填写审核意见并反馈于实验室申报单位进行相关情况的整改。主管单位可以查看山东省内的生物安全实验室及实验活动信息以及实验室所在单位信息，并对以上信息进行分类统计；另外，还可以查看到备案实验室的历史备案信息。平台管理员除了具有以上用户的所有权限外，还可以进行系统管理、用户管理、文件管理等。



实验室基本信息管理

各注册单位登录系统后，可在信息系统实验室列表页面维护各自单位实验室的基本信息，点击实验室相关信息可编辑相关信息，新增、删除和修改实验室相关信息内容，其中实验室的基本信息包括实验室名称、负责人基本信息、所属单位、所属卫生局等。通过该系统可实时将实验室信息提交给相应的上级管理部门，方便省内各卫生局以及卫生厅及时掌握其辖区范围微生物实验室的基本信息，方便查询、审核、监督。

病原微生物种类及实验活动类别的规范管理

根据生物安全防护等级的相关规定，建立病毒、原生菌、真菌的病原微生物种类库，各注册单位需填写清楚病原微生物名称、危害程度，勾选各自单位实验室涉及的病原微生物种类及实验活动类别(包括病毒培养或大量活菌操作、动物实验、临床标本检测、灭火材料操作和无感染性材料操作五类)，根据实验活动和《人间传染的病原微生物名录》填写详细的实验室活动信息。通过该信息系统可规范实验室日常活动，对病原微生物种类库中的标本种类进行跟踪，高效地管理与监督病原微生物种类及实验活动。

生物安全管理体系文件的规范管理

病原微生物实验室生物安全规范，包括对病原微生物实验的设施和要求、运行规范和人员管理。申请备案登记的实验室，需要具备一些必要条件与资质，例如，实验室需要具备与所从事的实验活动相适应的实验设施、设备及个体防护措施，满足《实验室生物安全通用要求》等国家标准规定；从事实验活动的人员应当参加生物安全培训，并通过考核取得上岗证书；设立单位有完善的生物安全管理体系，对所从事的病原微生物进行危害评估，制订生物安全防护方案、实验方法及相应标准操作规程、意外事件应急处理预案及感染监测方案等。实验室用户需要提交相关的电子材料，如实验室人员管理组织结构图、人员资质证书、平面结构图、实验室安全管理规定等。通过该系统将实验室现行生物安全管理体系信息的上传，一方面对规范省内病原实验室生物安全

管理起到督促作用，另一方面提高病原微生物实验室备案系统资料的实时性，通过系统在线填写、上传文件的功能，省时省力，提高了备案效率。

应用效果

截至 2020 年 4 月底，共有 5527 家医疗卫生机构二级生物安全实验室的资料审核和实验室备案工作，目前有 7600 多项实验室及实验室活动备案文档记录。

图 10 系统用户

图 11 备案文档

用户通过生物安全实验室备案管理网络平台，可以极大地提高工作效率，实现在线申请备案、审核等功能，简化了实验室生物安全管理的工作流程。此外，通过此备案系统，用户可以及时更新实验室生物及活动信息，将减少由于信息更新滞后而存在的潜在实验室生物安全问题。通过建立全省病原微生物实验室备案数据库，利用信息统计功能，上级管理部门不仅可以实时掌握全省病原微生物实验室的数量、实验室病原微生物活动类别以及实验室和菌（毒）种的分布情况，还能及时全面地了解病原微生物实验室更加详细的资料，实现了信息资源的共享，更方便上级管理部门监督与管理病原微生物实验室生物安全工作。

结语

新冠疫情再次为生物安全的防范敲响了警钟，其中病原微生物实验室的科学高效管理起着至关重要的作用。本系统实现了病原微生物实验室备案活动及文件的在线管理功能，能在很大程度上提高备案效率，后期需要加强数据备份与数据安全的保障工作。期待未来全省病原微生物实验室的备案数据库可以加入数据整合的队伍，融入互联网的大数据、运营商大数据、医疗大数据等高维、多源的数据来源，切实有效地为公共卫生安全提供基础防控与保障。

PU-066

结核病专项管理信息系统应用研究

张现辉、宋浩
河南省传染病医院 郑州市第六人民医院 郑州市结核病防治所

目的 我院作为河南省传染病医院、郑州市结核病防治所，主要任务为防治防治一体。在《中华人民共和国传染病防治法》中结核病被明确归为乙类传染病。该文件中明确要求对结核病进行网络上报便于预防、控制和消除。我院根据《结核病信息报告管理工作规程》、《传染病防治》、《结核病监测信息网络直报管理制度》开发结核病专项管理信息系统，完善结核病人就诊、登记流程，完善工作程序。

方法 我们将应用软件系统分为患者初定诊登记、传染病上报、结核电子病历三大部分，为了优化、完善结核患者的初诊、上报、定诊的流程，将医嘱、病历信息进行有效的串联，通过对结核患者进行建档管理，建立符合我院实际流程的结核患者电子病历管理系统，实现患者唯一初诊号、定诊号，将整体就医流程形成初诊登记、确诊上报、定诊登记、结核电子病历书写、整合等闭环管理模式。通过患者首次来院进行初诊登记，完成检查检验筛查；通过确诊上报，医生对确诊患者进行上报管理；通过定诊填写，医生对确诊患者进行统一管理，患者复诊信息整合；通过结核电子病历书写、整合对结核患者就诊记录有及时、准确、完整的信息资料，为结核病防治的决策提供科学依据。

结果 提高了医生的工作效率。以前需要医生手工完成登记、上报、病历书写等工作，上报不及时、容易漏报、错报，重报等，现全流程患者管理可在系统实现闭环，减少医生工作时间，提高了工作效率；结核患者上报管理有效的提高及时性、准确性。取消主管部门安排专人收集纸质传染病

报卡。系统审核可与结核专报系统完成实时对接，互联互通。结核电子病历保障了结核患者就诊的连续性，并对患者治疗、复查提供参考，具有可追溯性。结核病患者治愈后病历可进行整合并归档，具备结核患者管理的完整性、规范性。

结论 结核患者专项管理系统通过梳理、优化结核患者初诊、上报传染病、定诊、治疗、复查的各个流程，形成结核患者就诊流程的闭环管理。本系统的应用改变了确诊结核病患者病例漏报、错报和重复上报等现象，通过初定诊登记，并监测患者的查痰日期，对医生进行智能提醒示；通过对患者历史就诊信息、病历进行有效的串联，提高了医生的工作效率，大大地发挥了信息化优势，结核患者专项管理系统使结核患者的上报管理工作实现了规范化、制度化和程序化，提高了对结核病患者的管理水平及信息化处理的速度与精度，为改变结核病上报管理及结核病人病历管理的工作效率和模式开辟了新思路。

一、 研究背景

结核病是严重危害人类健康的重大传染病疾病之一，是对人类身体健康危害巨大的一种慢性呼吸道传染病，在全球广泛流行，我国结核病疫情仍很严重，实现联合国千年发展目标仍面临着很多困难和挑战。因此，要全面贯彻我国结核病控制策略，进一步加强结核病防治服务体系建设，提高结核病防治规划是是治疗，实现联合国千年发展目标，降低结核病疫情，提高全民健康水平，确保社会经济发展。我院作为河南省传染病医院、郑州市结核病防治所，主要任务为防治防治一体。在《中华人民共和国传染病防治法》中结核病被明确归为乙类传染病。该文件中明确要求对结核病进行网络上报便于预防、控制和消除。我院根据《结核病信息报告管理工作规程》、《传染病防治》、《结核病监测信息网络直报管理制度》开发结核病专项管理信息系统，完善结核病人就诊、登记流程，完善工作程序。

二、现状与发展

传统的结核病上报方式存在很多不足:①增加医生及医务人员的负担，由医生进行手写登记比较麻烦，并会占用患者和医生大量的时间，造成就诊时间过长，容易诱发医患矛盾。②由于收集数据方式比较传统，传染病上报需手工填写传染病报告卡，医生无法核实病人是否进行过上报。③未形成闭环管理，导致确诊患者再次就诊时，对患者历史数据的调取和应用存在“信息孤岛”，无法对信息进行有效的串联。传统结核病防治信息管理不适性凸显，信息重复、丢失、遗漏等现象客观存在，与结核病防治现代化技术发展需求不符，有悖信息共享、联动防治发展规律，为此需加大结核病防治管理系统研究力度。只有收集及时、准确、完整的信息资料，才能为结核病防治的决策提供科学依据。

结核病专项管理信息系统，包括结核病监测信息管理和结核病防治规划信息管理两部分。结核病信息管理在现代结核病控制策略中起着重要的作用，是制定结核病防治策略和措施、评价结核病防治工作效果与质量、预测结核病流行趋势的重要依据。对信息资料进行分析前，首先要对信息资料的质量进行评价。主要从完整性、准确性和及时性等方面对原始信息资料填写（包括初诊登记、定诊登记、病历记录）。因此，结核病专项管理信息系统是我院结核病防治规划的重要组成部分。

三、结核病专项管理信息系统的应用

3.1 系统设计

通过对河南省结核病专项信息管理系统需求的分析和理解，我们将应用软件系统分为患者初定诊登记、传染病上报、结核电子病历三大部分，为了优化、完善结核患者的初诊、上报、定诊的流程，将医嘱、病历信息进行有效的串联，我院设计了结核初诊、定诊、耐多药患者的上报系统，实现初诊、定诊无纸化。通过对结核患者进行建档管理，建立符合我院实际流程的结核患者电子病历管理系统，实现患者唯一初诊号、定诊号，将整体就医流程形成初诊登记、确诊上报、定诊登记、结核电子病历书写、整合等闭环管理模式。通过患者首次来院进行初诊登记，完成检查检验筛查；通过确诊上报，医生对确诊患者进行上报管理；通过定诊填写，医生对确诊患者进行统一管理，患者复诊信息整合；通过结核电子病历书写、整合对结核患者就诊记录有及时、准确、完整的信息资料，为结核病防治的决策提供科学依据。

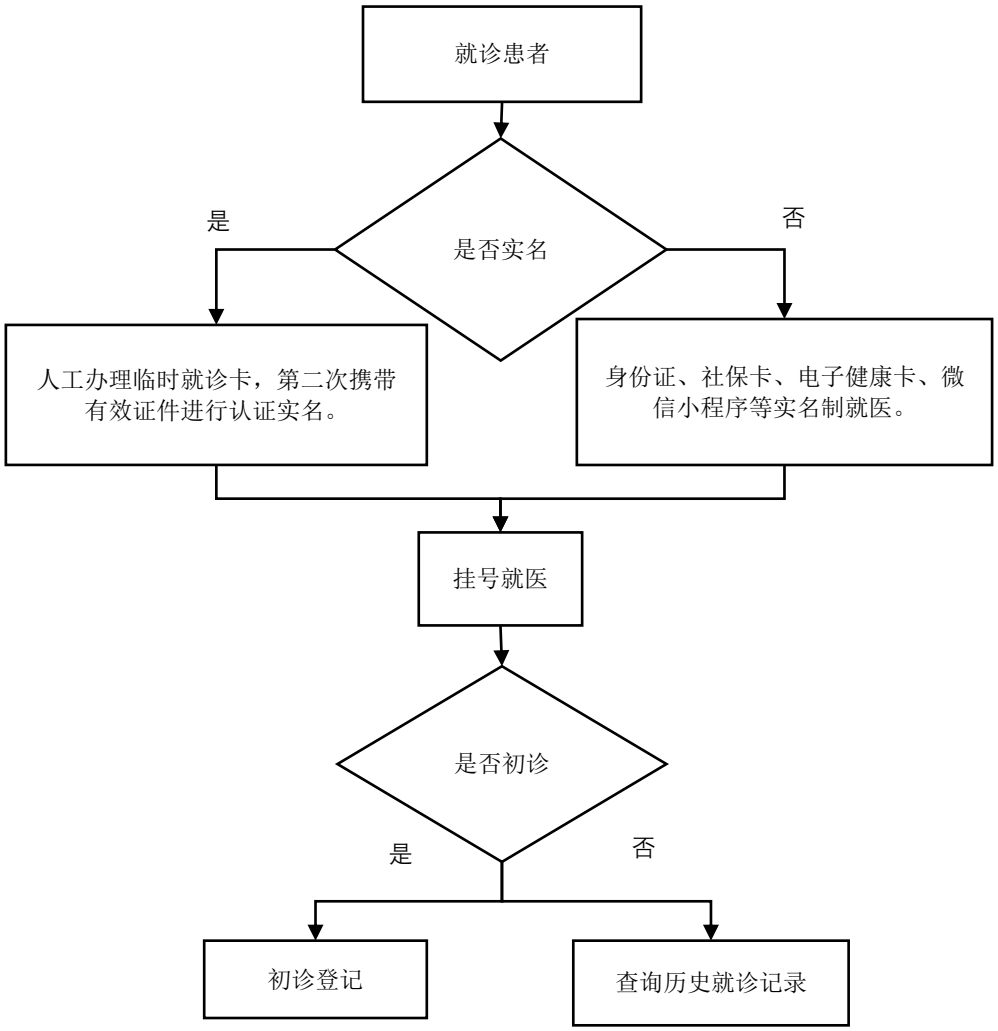
软件设计遵循医院结核病管理相关制度。采用客户机/服务器（Client/Server,C/S）架构，使用 PowerBuilder 编程语言进行开发，实现 HIS 的分流管理、门诊管理、预防管理、门诊医生站、门诊病历系统、住院医生站和 EMR 的电子病历等多个模块中的相关数据进行分析，挖掘和整合后，通过初诊登记、疾病上报模块、定诊登记、门诊病历管理等模块实现系统预期功能。本系统将项目管理、业务建模、分析与设计等统一起来，贯穿整个开发过程，支持扩展和结构化定制，能够满足结核患者专项管理的个性化需求；具有完备的数据安全与系统备份机制，确保数据在采集、传输、存贮过程中的完整性、保密性。同时系统支持 IE8.0 及以上版本的浏览器 WindowsXP/Windows2008/Windows 7/ Windows2008/ Windows10 等操作系统和浏览器的 PC 机

3.2 系统功能

3.2.1 初诊登记及信息采集

为了确保初诊患者进行信息登记，我院实行实名制就诊，就诊方式为：身份证、社保卡、电子健康卡、微信小程序等就诊方式，使用身份证号作为唯一索引标识，身份证索引标识是符合唯一性的条件，保证了患者索引信息的准确性，结核患者专项系统中每一项功能都以此为基础，从

而保证了数据的一致性，从初诊结核病人上报，结核专项管理系统进行初诊患者限制，如对初次来院就诊的患者自动弹出初诊上报。及时提醒医生填写电子初诊登记。

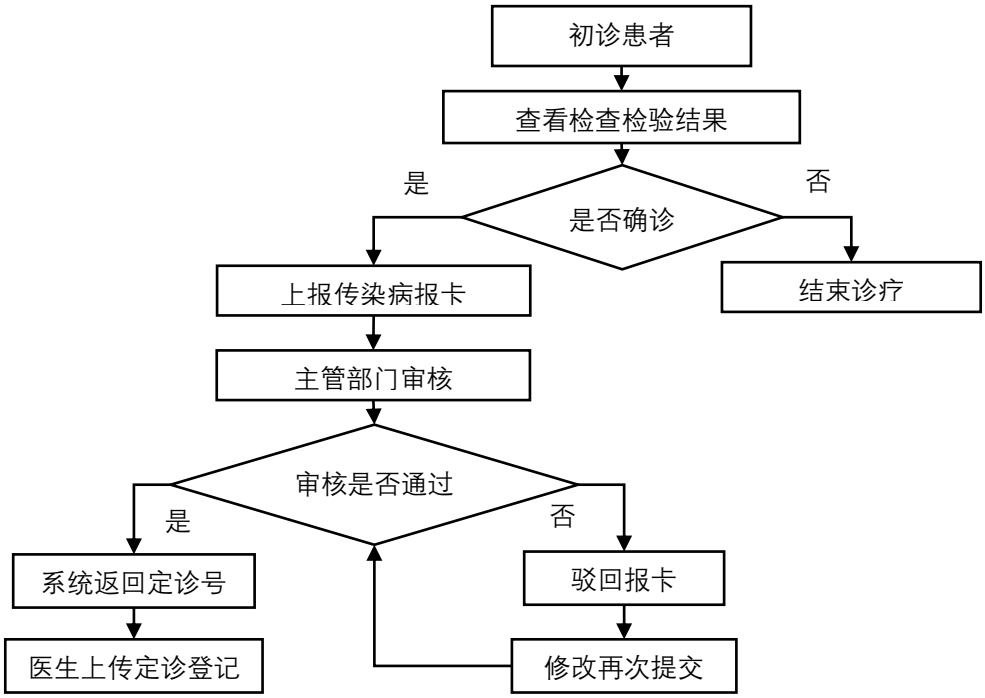


3.2.2 传染病上报及定诊登记

医师根据就诊患者实际情况进行判定是否确诊，确诊患者进行传染病上报，选择肺结核类型：利福平耐药、涂片阳性、仅培阳、仅分子生物学阳性、病原学检查阴性、无病原学结果、单纯性结核性胸膜炎、其他肺外结核、仅病理学阳性等进行上报，上报主管部门后，主管部门可通过系统进行审核，审核通过后自动生成定诊号，并生成确诊日期。医生站可根据提示书写定诊登记，定诊患者系统自动对信息自动匹配，将历史就诊病历进行整合串联。系统可根据定诊号可开立结核患者免费项目及药品。

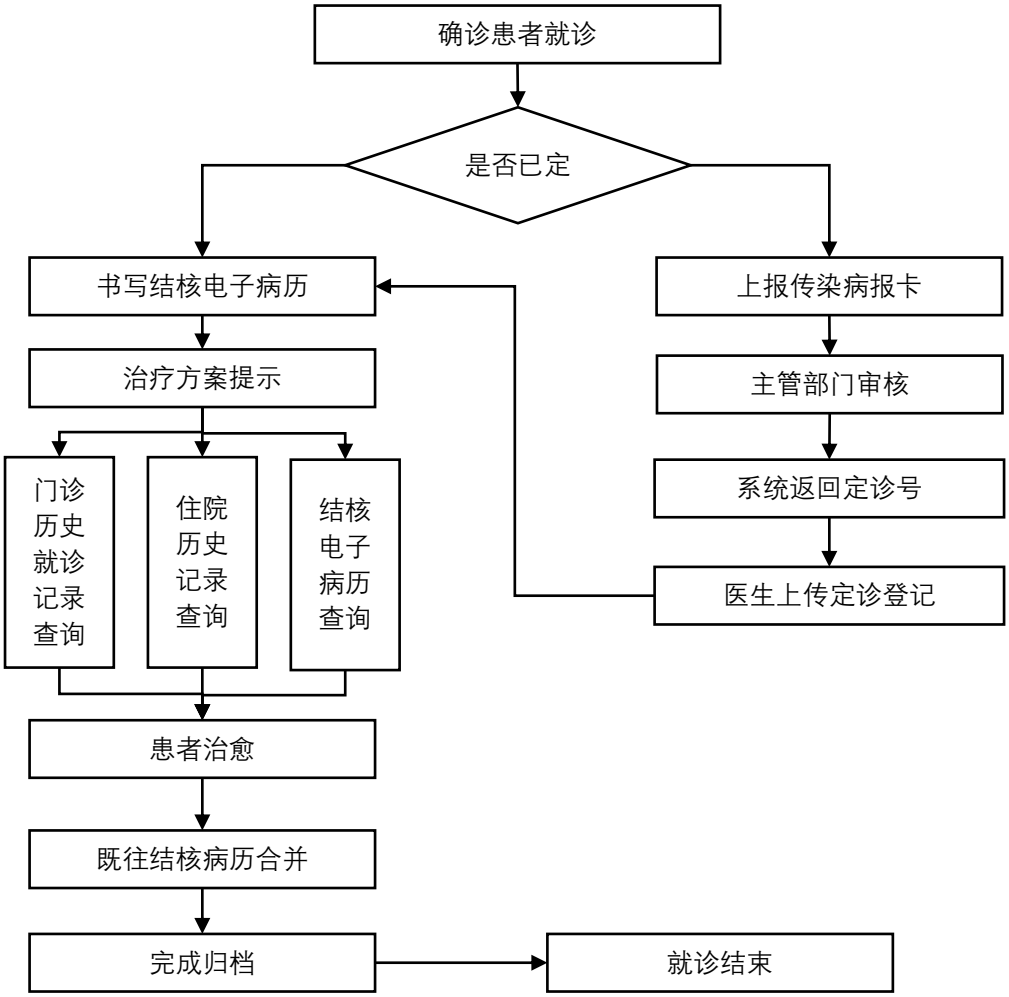
主管部门通过登记、查询对患者信息审核后进行审核确定，不符合的传染病报卡可被主管部门进行驳回处理；符合要求的传染病报卡审核通过后与结核专报系统进行实时上报，审核上报后医生端将无法进行修改，以保证网上申报与 HIS 中的信息保持一致。对已确诊的患者，系统在医生端界面显示中有患者类型进行标识，定诊患者信息中患者类型为“定诊患者”，并有定诊号显

示。系统可自动标注出各阶段时间查痰日期。同时确诊计算的痰检日期进行只能判定，对距离查痰前后一周的患者进行弹窗提醒。系统对定诊结核患者进行信息串联，通过患者的定诊号及病历号将患者的历史就诊信息进行串联，确保门诊病历、住院医嘱系统、电子病历系统的一致性，方便医生进行查阅。



3.2.3 门诊病历管理

已进行过定诊登记的结核病患者前往医生站进行就诊，医生接诊时提示病人类型为“定诊患者”的患者进行治疗及复查，并书写结核病历。对多次就诊的患者，医生可通过历史就诊信息查询查询功能随时查阅患者既往就诊记录、门诊处方、历次检查检验结果等书写此次就诊结核患者病历。



3.3 系统应用效果

结核病患者专项管理系统中初定诊患者登记、传染病上报、病历管理等功能运行稳定，医患反应良好，主要表现在以下几个方面：

（1）提高了医生的工作效率。以前需要医生手工完成登记、上报、病历书写等工作，上报不及时、容易漏报、错报，重报等，现全流程患者管理可在系统实现闭环，减少医生工作时间，提高了工作效率。

（2）结核患者上报管理有效的提高及时性、准确性。取消主管部门安排专人收集纸质传染病报卡。系统审核可与结核专报系统完成实时对接，互联互通。

（3）结核电子病历保障了结核患者就诊的连续性，并对患者治疗、复查提供参考，具有可追溯性。结核病患者治愈后病历可进行整合并归档，具备结核患者管理的完整性、规范性。

四、结语

结核患者专项管理系统通过梳理、优化结核患者初诊、上报传染病、定诊、治疗、复查的各个流程，形成结核患者就诊流程的闭环管理。本系统的应用改变了确诊结核病患者病例漏报、错报和

重复上报等现象，通过初定诊登记，并监测患者的查痰日期，对医生进行智能提醒示；通过对患者历史就诊信息、病历进行有效的串联，提高了医生的工作效率，大大地发挥了信息化优势，促进了我院结核病管理水平的提升，获得了院方及同行单位的高度评级。结核患者专项管理系统使结核患者的上报管理工作实现了规范化、制度化和程序化，提高了对结核病患者的管理水平及信息化处理的速度与精度，为改变结核病上报管理及结核病人病历管理的工作效率和模式开辟了新思路。

PU-067

区块链技术在医疗卫生领域现状研究

钱旦敏
南通大学医学院

目的 国际清算银行的一份研究报告显示，到 2025 年全球医疗保健市场在区块链上的支出预计将达到 56.1 亿美元，采用区块链技术每年可为医疗行业节省高达 1000 亿至 1500 亿美元的成本，医疗领域逐渐成为区块链技术的一大应用场景。

方法 笔者从药品供应链、突发事件应急管理、冷链配送和人工智能等方面阐述国内区块链技术研究现状，从健康链、重症医学、数据安全共享和智能医疗等方面阐述国外区块链研究现状。同时具体介绍区块链技术在公共卫生领域、药品供应链、医疗机构以及医疗保险等方面的实际应用。

结果 总结应用过程中数据的所有权和存储问题、隐私保密问题、系统安全问题以及监管问题。

结论 针对医疗机构监管、诚信、财务和安防系统，提出应用区块链技术的建议，以期为相关研究者提供参考。

习近平在中央政治局第十八次集体学习时强调，把区块链作为核心技术自主创新重要突破口，加快推动区块链技术和产业创新发展[1]。区块链是基于分布式节点共识算法、密码学和智能合约等技术的一种全新的分布式基础架构与计算方式，由于它具备去中心化、开放性、匿名性和不可篡改等特征，区块链在医疗卫生领域迅速发展，成为仅次于金融领域的第二大应用场景[2、3]。早在 2017 年，全球最大的信息技术和业务解决方案公司 IBM 就曾预言：全球 56%的医疗机构将在 2020 年前投资区块链技术[4]。

自区块链技术在 2008 年正式提出后，国内外学者对其展开深入研究，不断将其应用于医疗卫生领域，世界各国也都相继出台了相关政策，全力支持区块链技术的发展。国内方面，我国现已出台有关“区块链+医疗”的相关政策，并将区块链技术投入医疗保健等领域。2018 年，国家卫生

健康委员会表示正与工业和信息化部积极沟通，共同探索区块链在健康医疗大数据安全保障方面的应用[5]。2019 年 11 月，第八届国家卫生健康标准委员会卫生健康信息标准专业委员会召开的第一次全体委员会议上，胡建平副主任对互联网+医疗健康、健康医疗大数据、人工智能、区块链等新技术标准研发应用工作提出了要求，保证健康中国战略的实施和深化医改工作任务的贯彻落实[6]。同年 12 月，在北京召开的《基于分级诊疗的区块链研究与实践项目》课题中期汇报会指出“区块链+医疗健康”可以在数据隐私保护、信息可追溯等方面发挥优势，促进行业应用发展[7]。

国外方面，2016 年，欧盟开始资助经营医疗卫生安全领域的企业与领先研究型大学间的跨国合作，旨在建立一个可以汇总并共享整个欧盟各医疗卫生机构与患者之间生物医学信息的区块链系统[8]，同年 6 月美国公共卫生服务部(HHS)向医疗健康行业征集区块链技术应用研究报告。2018 年，美国 5 家规模最大的医疗卫生企业、2 家保险公司、1 个国家医疗测试实验室和 1 家医疗索赔机构通过区块链系统收集医疗服务提供者的相关数据[9]。2018 年 5 月，日本医疗机构 Medley 正在申请区块链电子处方管理专利，探索更为高效的电子处方管理方案，推动电子处方的普及[10]。2019 年，柬埔寨政策开放，瑞银华宝应用“区块链技术”在柬打造医疗健康大数据[11]。

1 国内外研究现状

随着科技水平的提高和医疗技术的迅速发展，国内外学者对医疗卫生领域的区块链技术研究更加深入广泛，并且发表了众多研究成果，为区块链技术在医疗卫生领域的应用提供了新思路。

1.1 国内研究现状

（1）区块链与药品供应链：如周倩（2018 年）研究了区块链在药品供应链中的应用，区块链技术不仅能有效解决产品供应链中产业协同、数据可行度、利益分配和生产过程的监管情况的制约，而且有助于医疗机构将药物生产、加工、运输和销售等多个环节全部管控，解决传统供应链信息不透明和交互困难等问题，使药品质量得到保障[12]。

（2）区块链与突发事件应急管理：如石菲等（2020 年）研究了区块链在健全公共卫生应急管理体系中的应用，区块链的应用不仅能够完善舆情管控手段，有效正面引导舆论，还能实时追溯应急物资，提升跨区域整体统筹调控能力[13]。

（3）区块链与冷链配送：如徐俊等（2020 年）研究将区块链在物流领域的应用和药物冷链配送相结合，便于实时监控药品配送状况，提高药品质量与安全性，倘若药品毁坏，可通过区块链技术迅速定位出错环节，便于追责[14]。

（4）区块链与人工智能：如欧阳丽炜等（2020 年）利用区块链信息数据的分布式计算及智能合约与人工智能等技术结合，研发了更为先进的传染病应急系统，人工智能技术能有效辨别有用信息，减少人工筛查负担，区块链则用于追溯、监管上报信息，两种技术结合使传染病应急系统具有精准定位和快速预警的优点[15]。

（5）区块链与药材发展和认证系统：如张小波等（2020 年）针对药材产业链复杂且涉及环节多的情况，利用区块链和大数据等信息技术，建设道地药材高质量发展和认证系统尤为重要。该

系统基于区块链技术规范和统一产品链，将药材的相关信息上传至区块链进行保存，为药材的认证提供数据来源[16]。

（6）区块链与疫情防控措施：如赵超（2020 年）利用区块链的不可篡改和信息加密等特点，实现疫情期间相关人员的全程监测和追踪，快速定位疑似患病人员，有效遏制疫情扩散。同时将区块链融入物资调配系统，提高相关信息获取速度和资源调配效率[17]。

（7）区块链与档案信息保护：如查佳凌和张渊（2020 年）鉴于区块链的去中心化特点，研究将区块链技术应用于患者档案信息的加密保护，防止患者信息被恶意篡改，在患者转院过程中，相关信息也可流通共享，较为充分的实现了患者信息的跨区域、跨机构的保存[18]。

（8）区块链与处方药流通管理：如杨怡（2020 年）利用区块链建立可靠的信息平台，可对处方药的流通进行合理管控，记录医院处方信息，避免处方造假的现象。对处方药的供应与销售和开具处方整个过程进行闭环式管理，使处方药的去向有据可循，实现处方药的有效监管[19]。

（9）区块链与医疗数据共享：如高梦婕和王化群（2019 年）研究基于区块链实现对称性可搜索的医疗数据信息共享，通过密钥和智能合约等完成信息的存储与共享，进而保障研发企业所需医疗共享数据的安全[20]。

1.2 国外研究现状

（1）区块链与健康链：如 Ray Hales Hylock 和 Xiaoming Zeng（2019 年）基于区块链技术提出了健康链概念，一个新的以病人为中心的区块链框架，研究目的是在一个安全、可互操作的环境中加强患者参与、数据监管和累积信息的规范传播[21]。

（2）区块链与重症医学：如 Tiago Guimarães 等（2020 年）将区块链技术应用到重症医学智能决策支持系统，其目标是通过确保其不变性和监督来提高来自重要传感器和监视器的数据的准确性和价值，以及隐私和不充分数据管理的责任[22]。

（3）区块链与数据安全共享：如 Rob Currie 等（2019 年）利用区块链创建出一个认证系统，有效地共享标准护理过程中产生的健康数据，能够在不侵犯患者的隐私的情况下，实现利用患者关键信息对疾病进行分析，极大地加速疾病如癌症治疗的进展[23]。

（4）区块链与智能医疗：如 Faisal Jamil 等（2020 年）设计开发了基于区块链的智能合约的新型平台，区块链技术受益于诸如隐私和数据透明性等独特功能，使新型平台不仅可以监视患者生命体征，而且可以随时随地从全球访问医疗信息[24]。

（5）区块链与药品数据流：如 Jen-Hung Tseng 等（2018 年）提出以 Gcoin 区块链作为药品数据流的基础，创建透明的药品交易数据。药品供应链的监管模式可以由单纯的检验和检验模式转变为监管网络模式，药品供应链中的每个单位都可以同时参与，以防止假药的传播，保护公众健康[25]。

（6）区块链与急救医疗数据源：如 Shirin Hasavari 和 Yeong-Tae Song（2019 年）研究一种基于区块链的应用方案，用于在患者从一个医疗机构走到另一个医疗机构时记录与患者急救相关的医疗数据，从而将患者作为一个安全且可扩展的数据源进行连续记录[26]。

（7）区块链与移动应用程序：Park SungJun 等（2020 年）利用区块链记录的透明性和不变性是区块链技术的关键特征，开发了一个应用程序，使用区块链来监控病历的变化[27]。

（8）区块链与传染病报告系统：Min Cheol Chang 和 Donghwi Park（2020 年）研究了区块链如何改善当前传染病报告系统的缺陷，从而在当前的 COVID-19 大流行期间或在将来发生其他传染病时为人们提供帮助[28]。

2 国内外应用现状

2.1 区块链应用于公共卫生领域

公共卫生是关系到一国或一个地区大众健康的公共事业。公共卫生包括对重大疾病的预防和控制、对公共环境卫生的监督管制以及相关的卫生宣传和健康教育等。迄今为止，应对公共卫生突发事件就是一项复杂的工程[29]。然而，由于区块链技术正逐渐融入公共卫生领域，为应对公共卫生突发事件提供技术支持，一定程度上推动公共卫生事业的发展。

2019 年 12 月以来，由新型冠状病毒（COVID-19）引起的肺炎（NCP）在多国迅速蔓延，成为全球关注和讨论的话题。2020 年 2 月 14 日，习近平总书记在中央全面深化改革委员会第十二次会议指出要鼓励运用大数据、人工智能和云计算等数字技术，在疫情防控措施中发挥支撑作用[30]。尽管通过国民的不懈努力，疫情防控已经接近尾声，但与此同时，在疫情防控过程中也暴露出了许多问题，例如武汉市红十字会在防护物资调用出现纰漏以及武汉市在对疫情防控预警信息的滞后、不准确等[31]。

针对上述问题，虽然区块链技术并未深入使用，但不可否认的是区块链技术以独特的技术特点在这次疫情中取得了良好收益。由于区块链去中心化的特点，各节点平等参与交易，任意节点的交易都将传递到所有节点验证，有助于捐赠资金的监管，保证捐赠数据安全可信，提高公信度。此外，利用区块链去中心化特性，还可实现点对点的精准慈善，有望重塑慈善公益模式。

疫情期间存在的信息不对称问题可能会造成一定的社会动荡。社会高层对当地疫情信息的不准确、不及时的更新，将会造成民众的误听误信，对政府失去应有的信任。为了更好地控制疫情发展、稳定民心，对疫情信息及时地公开是必要的。基于区块链所建立的疫情信息共享平台有广州市南沙市的疫情防控协同系统，该系统保证了各个部门的数据互通，有利于政府更好地工作。

区块链在医养方面也具有一定的应用场景。人口老龄化是中国目前面临的难题之一，医养结合是很多学者认为的应对人口老龄化问题现实需要的一种方式[32]。利用区块链技术的分布式记账以及联盟链与公有链的分类特征创造老龄人口医养服务信息共享平台，填补了老龄人口医养服务领域存在的信息鸿沟，解决信息传递不畅导致的医养服务供需不对等、不平衡的问题[33]。我国关于医养结合的研究正处于探索期，仍需要加强多方面技术，但可以肯定的是“区块链+医养”的研究仍具有较大发展空间，在今后的工作中加深区块链与医养的融合度，以实现“区块链+医养”养老模式的同时为老人们提供更好的公共卫生服务。

2.2 区块链应用于药品供应链

药品的溯源一直是我国重要的民生问题，药物造假的事件更是层出不穷[34]。2018 年 7 月 15 日，国家药品监督管理局指出长春长生公司存在严重的疫苗造假行为，该行为已严重危害到国民的生命健康。对于此类伪造事件，传统溯源系统采用的中心化账本记账，不能保证数据准确性，且传统溯源系统一般流程长，溯源耗时长，增加了数据在溯源过程中被篡改的可能性。因此，针对当前医药行业溯源流程长、环节复杂的特点，基于区块链的分布式储存技术及其不可篡改、去中心化等特点所建立的区块链溯源服务系统将产品从生产、销售到运输整个全过程的信息均写入区块链中，同时，区块链可通过多方进行监管，提高数据信任，解决传统溯源系统的信息孤岛问题，提高了产品信息的透明度[35][36]。

此次新冠肺炎疫情中，我国传统中药材发挥了重要作用，使中药材市场开始升温，随之而来的是各种药物造假事件。《国务院办公厅关于加快推进重要产品追溯体系建设的意见》中明确要求加快推进利用信息技术对药物等产品的溯源体系的建立[37]。利用区块链和大数据计算等技术建立药物溯源体系，避免药材在流通环节中出现造假和过度抬高价格等现象。中药溯源系统采用了“一物一码”的形式，为每一株药材生成对应的二维码或条形码，通过第三方系统或相应扫码设备，将相关信息上传至溯源服务平台，对信息进行上链，通过点对点储存、去中心化管理和加密处理，使其具有不可篡改可共享的特性，提高追溯的可信度，便于监管人员对药材信息的查询。区块链去中心化、不可篡改及数据加密等特性有助于中药材溯源体系形成一条信息共享链，可以实时对每个节点的信息进行回溯和定位。中药溯源系统不但能起到监管、规范的作用，还能惠及大众保障用药安全，控制药材最终定价，防止中间商恶意抬高价格，是实现安全药、放心药、有效药的重要民生工程。

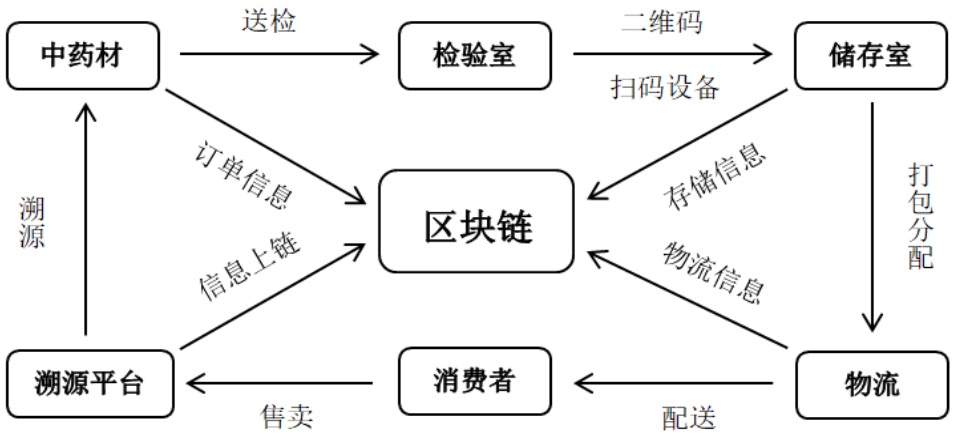


图 1 中药材溯源系统结构图

区块链的信息追踪不仅可以用于药品溯源，还可以用于其他行业商品的溯源和防伪，如农产品、食品及进口商品。在信息时代，食品防伪和溯源仍面临着重大问题，监管范围大，耗时长，中间环节复杂等，均为食品监管的“痛点问题”，而新兴的“区块链+物联网”则可为此类问题带来全新的解决之道。2017 年 8 月 8 日，天猫国际宣布全面启动全球溯源计划，利用区块链技术以及大数据跟踪进口商品全链路，汇集生产、运输、通关、报检、第三方检验等信息，给每个进口

商品打上“身份证”，能精准管控每一件跨境进口商品，让消费者可以通过扫描“身份证”了解该商品的真伪，及其原材料信息、生产地信息和物流信息等，实时查询商品信息，做到来源可控，质量有保障[38]。进口商品生产、加工、运输的每一个环节的信息均记录在区块链上，由于区块链数据不可篡改的特性，确保了商品的唯一性，若出现产品纠纷，可快速定位出问题环节，并进行相应追责。

同年 12 月 14 日，沃尔玛、京东、IBM、清华大学电子商务交易技术国家工程实验室宣布成立中国首个安全食品区块链溯源联盟，旨在通过区块链技术进一步加强食品溯源工作，提高食品的安全性，提升食品供应链的信息透明度，进一步保障消费者的饮食安全[39]。

2.3 区块链应用于医疗机构

国内医疗资源的配置呈现出不平衡的“倒三角”形式。一些高级的医疗机构及高素质医务人员主要集中在大城市，而农村、边远地区和贫困地区的配置相对较低，医疗水平相对落后。医疗资源的不平衡配置迫使一些农村重症患者离开家到市级或省级医院寻求治疗。

然而，在一些高级的医疗机构中，“看病难”已是常态。由于知名医院的号源在数量上往往有一定的限制，患者即使到了医院也不一定能挂上号。据华山医院信息科副主任张琪介绍，复诊患者占据了大部分号源，致使一些初诊患者挂不上号。而在复诊患者中，慢病患者居多，他们通常需要长期的药物供应，但处方药一次只能开一到两周的药量，所以他们只能不断地往返医院。针对这一类型的患者，华山医院以内分泌科为试点，与蚂蚁金服合作推出我国首个区块链电子处方，百度超级链与重庆市联合打造电子处方流转平台，此外，还有我国首个具有自主知识产权的“区域慢病智能管理与药品配送平台”[40]。这些平台不仅解决了复诊患者的麻烦，而且为初诊患者节约了大量的号源。平台运转的基本流程就是将医生的诊断记录、处方、取药信息、送药信息以及支付信息等盖上“戳”，记录在电子处方的流转链上。根据区块链不可篡改的特性，患者不用担心处方在流转过程中被篡改，且该处方全程可追溯，方便出现意外后的追责。在患者购买药品时，为了实现购买过程透明化，患者的个人数据会被上传，符合监管部门的要求，避免处方滥用保证了处方的唯一性。

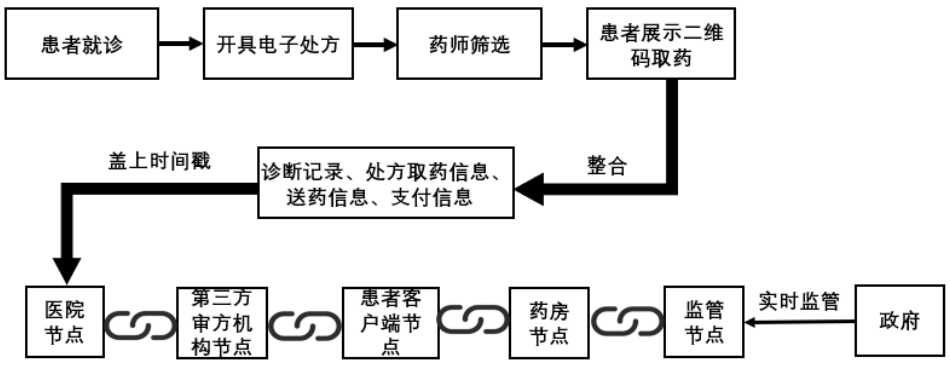


图 2 电子处方流程图

除了“看病难”的问题，医院间也存在一种普遍的现象：一位患者在一个医院检查的结果，在转诊的过程中，需要再次检查，使患者花了大量的金钱和时间。这一现象的存在暗示了医院间的“数据竖井”问题，目前，我国越来越多的公司和研究机构正在使用区块链实现医院间从信息孤岛到互联互通，如阿里健康与常州市的“区块链+医联体”试点项目、长三角地区的“互联网+医联体”、上海第一人民医院和安徽省立医院间的电子病历等[41、42]。北京聚农科技联合统捷集团基于区块链技术建立分布式银行，让医院、厂家、用户在平台点对点的进行交易，不仅彻底的改变行业原有的经营模式和消费模式，还推动医疗行业的进步以及经济水平的提高。

传统的医疗信息由中心化数据库存储，容易受到黑客攻击。国外医疗公司 Anthem 曾被黑客攻击，造成了 8000 万患者和雇员的信息被泄露，美国 UCLA Health 公司也曾泄露 450 万患者的数据[43]。与传统的医疗信息存储相对比，区块链的去中心化技术可以更好地保障数据安全，患者在不同医疗机构的就医记录将打包加密存储在区块上，若调用非本机构产生的用户数据时，需经患者的授权许可，通过密钥对比获取相关的医疗信息。在数据共享的同时，有一个问题将会显现，不同医院使用的术语可能不同。对这一问题的解决，仁济医院搭建了一个基于区块链与人工智能搜索引擎的云平台，将所有检验报告通过搜索引擎，统一转化成标准化报告进行输出，供医联体大平台下各家医院医生统计分析与病理研究[44]。医联体的发展可以促进上层医疗机构与下层医疗机构的衔接，类似于“先富带动后富”的思想。

在患者信息数据的采集、存储和共享的方面研究，国外的发展比我国更早且更为成熟。2016 年，Gem 公司的 Gem Health 项目，旨在打造一个将互连整个医疗保健生态系统的区块链网络[45]。2017 年，美国食品和药物管理局(FDA) 就已经与 IBM 沃森健康部门签署研究计划，将区块链用在医疗记录、临床试验、基因数据和可穿戴设备等多个方面；同年 7 月，日本 GMO Internet 集团推出的医疗记录共享系统[46]。这些研究将区块链的分布式账本技术运用在电子医疗数据上，在保护病人隐私的同时实现了数据共享，使医生能够更加了解病人的过往病史、用药史及过敏史等，更好地为病人服务。此外，第三方医疗机构可以通过医院共享的患者数据对特定类型的疾病进行建模分析，从而达到更好的辅助决策和治疗的目的，或利用大量的患者数据来研制新药，更好地助力智慧医疗的发展。

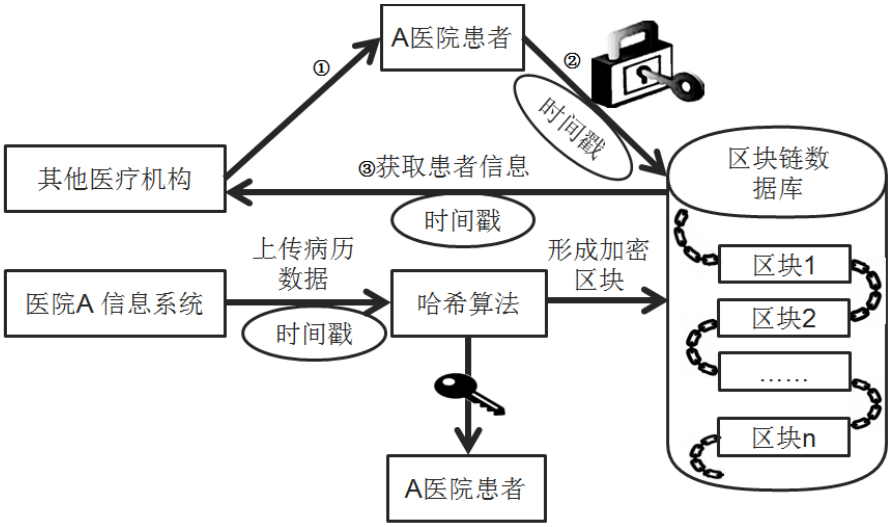


图 3 基于区块链的电子病历共享平台

大数据时代利用好医疗数据不仅会推动医疗产业的发展，还能提高医疗效率与准确性。浙大一院与阿里巴巴共同建设医疗人工智能工程实验室，这个实验室将建设基于区块链技术的医联体数据平台、医疗人工智能影像辅助诊断平台、医疗大数据智能科研平台、医学人工智能临床决策支持平台等四大公共平台产品[47]。通过机器人可以医学影像结果中对特征进行提取，对病灶进行数量评估与标记，医生根据机器人得出的结果作出相应的影像报告。国外方面，科技巨头 Xerox 将区块链应用于电子健康档案的保护，由于区块链的分布式记账系统具有开放性、透明性的特点，既可以使交易分散进行，也能有效保证所存储的电子健康记录的安全，实现数据隐私和公开之间的平衡[48]。英国萨里大学将区块链和机器学习算法将可穿戴式的数据结合，为公共和私人医疗保健提供者提供安全的存储和传播方式，帮助他们设计“有针对性”的治疗方案，是区块链在个性化医疗护理活动中的应用体现。数据安全公司 Guardtime 在 2017 年与爱沙尼亚政府合作创建了一个基于区块链的框架来验证患者身份，确保 EHR 中的审核记录不能被恶意修改并且该系统对医疗记录的任何修改都是安全且可审核[49]。

2.4 区块链应用于医疗保险

目前，我国患者的医疗信息基本由医院掌握，医保经办机构与医疗机构存在消息不对称问题，为了防止医保诈骗，参保人需要提供各种证明才能拿到医保。对异地就医的患者，他们报销手续更为复杂，花费的时间更长。这主要是由于各地区的医保目录范围和报销比例不一以及医保经办机构与医疗机构消息不互通。针对这一方面，区块链技术将会起到一定的作用。由于上链的信息不可篡改，信息录入员必须确保信息真实准确，医疗档案的记录上链存储，满足监管要求，保证审查机构的追溯、追责权利。可以说，区块链在参保人、医院和医保经办机构之间建立了信任机制。2019 年，浙江省推出的“浙里办”APP，实现了群众看病到报销“最多跑一次”，即市民看完病，可通过“浙里办”APP 查看医疗票据，并进行网上报销[50]。近日，福建省福州市第二医院也推出了“区块链电子票据平台”，真正解决群众纸质医疗票据难保存、数据读取慢等问题，为医保报销提供了更为方便的通道[51]。

除了解决医保问题的电子票据，国内外的保险企业也在抓紧区块链技术发展的势头，正不断探索区块链在保险业务的应用。据公开的资料显示，保险业务应用于区块链技术最早可以追溯到2015 年众安科技组建的区块链团队，该团队主要是对区块链与保险产业的融合进行探索。近几年，“区块链+保险”也在逐步发展，上海保交所推出“保交链”和“共保体+区块链”，中国人保开展的保险营销和养殖保险等领域的区块链，中再集团联合众安科技等企业推出的国内首个区块链再保险实验平台等[52]。截止到现在，区块链主要运用于保险业的再保险、年金管理、健康医疗等11 个应用场景。

3 区块链技术在医疗健康领域应用问题及建议

3.1 存在的问题

3.1.1 数据所有权问题

由于区块链中除各方私有信息会以密码学技术加密外，其他数据对所有人公开，任何人都可以通过公开的接口查询区块链数据和开发相关应用，因此难以界定数据的所有权。国内目前缺乏相关法律提供指导方针，各种利益相关者对数据所有权持有不同看法，导致个人健康数据的所有权存在争议[53]。

3.1.2 数据存储问题

一方面随着大数据在医疗健康领域的应用，数据量呈现出“爆炸式”增长，区块链存储的信息量将越来越大，使终端用户进行数据同步的时间延长，从而导致区块链的阻塞与延迟。这一问题将直接影响医疗机构更新信息和获取患者数据的效率[54]。另一方面区块链存储的数据无法删除，一旦出现错误或不良信息，将无法纠正，从而对区块链造成永久性伤害。

3.1.3 隐私保密问题

尽管区块链上信息传递可以匿名进行，在一定程度上保护了交易者的身份隐私和交易数据的隐私。但是通过检查分析区块链上共享的信息，仍然可以推导出该用户的相关信息，用户的健康数据将会被泄露[55]。

3.1.4 系统安全问题

区块链并不是绝对安全的。虽然现实中很难实现 51%的攻击或分布式拒绝服务攻击或智能合约的攻击，但并非完全不可能，一旦有人恶意攻击，其后果不堪设想。此外，Fedorov Kiktenko 提出：在 10 年内，量子计算将会破解区块链的加密算法，这对于使用加密算法保护链上数据的区块链来说是一大风险[56]。

3.1.5 监管问题

去中心化是区块链最本质的特征。由于没有中心化机构的存在，完全通过预先设定的程序自动运行，在降低成本和提高效率的同时，监管问题也随之而来。区块链应用领域繁杂，目前关于区块链的相关法律体系及其应用标准尚未确立，导致相关的医疗机构持保守态度，限制区块链的推广及应用。

3.2 建议

3.2.1 医疗机构监管系统

基于“监管沙盒”模式的医疗机构监管系统是基于区块链技术创造一个安全空间，在监管有效、用户受保护的前提下，放宽规定，减少束缚，将新型医疗技术和政策加入监管系统进行实验，从而验证方案可行性，提高医疗监管系统的完善性。“监管沙盒”模式实施流程需要医疗机构提交申请，审核结束后形成测试方案，系统及时跟进监控，方案实行完成后由监管系统审核，报告一经审核通过，医疗机构即可向整个市场推广新型医疗技术和政策。将金融科技领域里的区块链“监管沙盒”模式应用到医疗卫生领域，使得在风险可控的前提下，推动医疗卫生领域的发展创新[57]。

3.2.2 医疗机构诚信系统

当今社会中，患者恶意拖欠医药费或辱骂殴打医务人员等恶劣行为时有发生，为了防止此类现象发生，医疗机构可以利用区块链中信息公开透明且可追溯的特性，将病患就诊过程中所有治疗和缴费记录都将保存在区块链上，当患者存在不良行为时，可将此类事件记入区块链中，并将患者纳入医院失信名单，区块链信息的共享使链上各大医院均了解患者失信情况，逐步建立完善医院诚信系统[58]。

3.2.3 医疗机构财务系统

医疗机构财务管理部门可借鉴区块链在会计行业的应用。一方面区块链的分布式记账和加密技术不仅能有效保证财务数据的安全与准确而且还能解决财务信息失真、不对称、传输不及时的状况。另一方面，区块链的链式记录能够按照时间节点存储信息，便于信息的查找和管理，从而降低了信息造假的可能性[59]。

3.2.4 医疗机构安防系统

医疗机构的安全问题一直是医疗领域的重中之重，尤其当互联网普及之后，任何一个网络环节出现问题，都会影响患者的安全以及医疗机构的利益。将区块链技术应用于医疗机构的安防系统，将监控视频和医疗数据等内容加密并上传至区块链，提高医疗机构的安全系数[60]。

4 结语

区块链由于其去中心化、数据不可篡改和开放性等特性，受到国内外医疗机构和相关企业的青睐。近年来，国内外众多医疗机构和企业已陆续将区块链技术投入医疗卫生相关领域的项目应用，不仅为消费者提供了便利，同时也促进了医疗卫生领域的快速发展。针对现阶段区块链的应用分析，可见区块链在医疗领域的发展尚处于早期，随着区块链技术的不断发展，区块链本身的技术问题将会得到解决，未来将不是区块链寻找应用，而是应用找上区块链。因此，各医疗机构要跟上时代的浪潮，了解区块链技术在实际医院管理工作中所能解决的问题，并着手实践。

PU-068

基于 GM（1,1）灰色预测模型的山东省公立医院医疗费用分析及预测

戴涛明[1]、宋奎勐[2]、甄天民[2]
1. 济南大学 山东省医学科学院医学与生命科学学院
2. 山东省医药卫生科技信息研究所，山东第一医科大学（山东省医学科学院）

目的 分析和预测山东省公立医院人均医疗费用的变化趋势，为山东省进一步完善医疗服务价格动态调整机制提供政策建议。

方法 收集 2012-2018 山东省公立医院医疗费用的有关数据，采用描述性统计、GM（1，1）灰色预测模型对数据进行分析与预测。

结果 山东省公立医院医疗费用将持续增长，医疗费用中药占比逐年下降，然而卫生材料费占比逐年上升，体现医务人员技术劳务价值的费用项目占比较低且变化趋势不明显。

结论 建议尽快建立医疗服务价格动态调整机制，加强对医疗机构医用耗材管理工作的监督与管理，建立符合医疗卫生行业特点的薪酬分配机制。

PU-069

基于真实世界数据的奥利司他不良事件信号挖掘研究与分析

袁婷[1,3]、王丽[1]、张杰[2]、陈力[1]、宿怀予[3]、钱灿[3]
1. 四川大学华西第二医院 2. 成都市药品不良反应监测中心
3. 德阳市人民医院

目的 通过美国食品药品监督管理局（Food and Drug Administration，FDA）不良事件报告系统（FDA Adverse Event Reporting System，FAERS）挖掘奥利司他不良事件信号并进行分析，为临床提供安全用药参考。

方法 采用报告比值比（reporting odds ratio，ROR）法和比例报告比值（proportional reporting ratio，PRR）法，对 FAERS 数据库中 2015 年至 2019 年共 5 年奥利司他导致的不良事件报告进行挖掘与分析。

结果 获得以奥利司他为首要怀疑药品的不良事件报告共 9277 例，男女比例为 0.16，平均年龄为 48.24 岁，中位年龄为 48 岁。使用 ROR 和 PRR 法共得到 109 个信号，新的不良反应累及 7 个系统；还检索到药物无效、治疗不服从、疑似假冒产品等与产品问题相关的不良事件信号。

结论 奥利司他药品说明书未收录的不良反应累及器官广泛，产品问题值得密切关注，临床用药时应严格保障药品质量，提高患者依从性，警惕新的不良反应信号。

肥胖症已成为危害人类健康的重要疾病，全球患病率显著升高，特别是在中国等发展中国家，肥胖症与心脑血管疾病、高血压、2 型糖尿病、血脂紊乱、睡眠呼吸暂停综合征等疾病的发生有密切的关系[1]。肥胖是人类健康的隐患，减肥意识也逐渐被人们所认识。2015 年国家卫生计生委发布了《中国居民营养与慢性病状况报告》，报告指出无论成人还是儿童青少年，超重肥胖率均呈现上升趋势[2]。

奥利司他作为新型的减重药物，于 1999 年获得美国食品药品监督管理局（Food and Drug Administration，FDA）批准上市。该药为可逆性胃肠道脂肪酶抑制药，通过与胃和小肠腔内的胃、胰脂肪酶的火星丝氨酸共价结合使酶失活而发挥治疗作用，失活的酶不能将食物中的脂肪（主要是甘油三酯）水解为可吸收的游离脂肪酸和单酰基甘油，从而减少热量摄入，控制体重。有研究数据分析 2013 年-2016 年减肥产品零售市场，国内市场奥利司他药物数量逐年增加[3]。

随着人们对肥胖的重视，作为非处方减肥药的奥利司他使用量逐年增加，在奥利司他的广泛应用下，对于不良反应报道的文献研究资料较少[4]。本文检索了美国食品药品监督管理局不良事件报告系统（FDA Adverse Event Reporting System，FAERS）数据库中奥利司他的不良事件报告，并对数据进行整理分析，为临床提供更多的用药参考。

1 数据与方法

1.1 数据来源

研究数据为美国食品药品监督管理局不良事件报告系统中 2015 年至 2019 年共 5 年的不良事件报告。该研究数据以 ORLISTAT、ALLI、XENICAL 为关键词，从 FAERS 数据库进行检索，共检索到 9277 例药物不良事件(Adverse Events，ADE)报告，报告数据包括患者基本信息、药品信息、不良事件名称、结果、日期和报告国家等内容。

1.2 数据处理

1.2.1 标准化

本文采用国际医学科学组织理事会《国际医学用语词典》（Medical Dictionary for Regulatory Activities，MedDRA）中药物不良反应术语集的首选系统器官分类（System Organ Class，SOC）和首选术语（Preferred Term，PT）。本研究下载 MedDRA 23.0 版，用于首选语(PT)、高位语(High Level Term, HLT)、高位组语(High Level Group Term, HLGT)、系统器官分类(SOC)编码和语言汉化[5]。

1.2.2 数据挖掘方法

本文采用信号检测方法为报告比值比（reporting odds ratio，ROR）法和比例报告比值（proportional reporting ratio，PRR）法，主要计算目标药品不良事件与数据库中对应的药品不良事件等发生之间的关系，若两者之间的关系达到失衡，则提示信号生成，该研究方法具有较高的可信度和灵敏性。

1.2.3 数据筛选和计算

对标准化后的数据中药品不良事件报告不完全、信息不确定报告及其他包含异常信息报告进行排除。筛选出报告数≥3 的 PT，按照比例失衡法四格表（表 1），计算出相应 PT 值及其对应的 ROR 值、PRR 值和相应的 95%CI 下限（表 2）并统计信号数，再进一步剔除 ROR 与 PRR 比值的 95%CI≤1 的报告[6-8]。

表 1 比例失衡法四格表

药品种类	目标不良反应报告数	其他不良反应报告数	合计
目标药物	a	b	a+b
其他药物	c	d	c+d
合计	a+c	b+d	a+b+c+d

表 2 ROR 法与 PRR 法公式及阈值

方法	公式	阈值
ROR	$ROR=\frac{(a/c)}{(b/d)}=\frac{ad}{bc}$ $SE(\ln ROR)=\sqrt{(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d})}$ $95\%Cl=e^{ln(ROR)\pm1.96\sqrt{(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d})}}$	a≥3; ROR 的 95%CI 下限>1，则提示生成一个信号
PRR	$PRR=\frac{a/(a+b)}{c/(c+d)}$ $SE(\ln PRR)=\sqrt{(\frac{1}{a}-\frac{1}{a+b}+\frac{1}{c}-\frac{1}{c+d})}$ $95\%Cl=e^{ln(PRR)\pm1.96\sqrt{(\frac{1}{a}-\frac{1}{a+b}+\frac{1}{c}-\frac{1}{c+d})}}$	a≥3; PRR 的 95%CI 下限>1，则提示生成一个信号

2 结果

2.1 药品不良事件的基本信息

2015 年至 2019 年，由奥利司他导致的不良事件报告数共计 9277 例，共 4290 个不良事件。从报告中患者的性别来看，男女比例为 0.16，女性占大多数。在已知年龄的患者中，平均年龄为 48.24 岁，中位年龄为 48 岁。2015 年至 2019 年分别报告 1572、726、873、623、496 例，呈下降趋势。报告国家前五位分别是美国、英国、西班牙、巴西、意大利，美国报告数占 91.72%。报告者大多数为消费者，占 93.38%。报告基本信息见表 3。

表 3 2015-2019 年奥利司他相关不良事件报告基本信息

信息	分类	报告例数/N	构成比/%
总计	/	4290	100
性别	男性	422	9.84
	女性	2599	60.58
	不清楚	1269	29.58
年龄	＜18 岁	16	0.37
	18-64 岁	1180	27.51
	≥65 岁	225	5.24
	不清楚	2869	66.88
年份（2015-2019）	2015	1572	36.64
	2016	726	16.92
	2017	873	20.35
	2018	623	14.52
	2019	496	11.56
报告国家（前五位）	美国	3935	91.72
	英国	127	2.96
	西班牙	27	0.63
	巴西	14	0.33
	意大利	13	0.30
报告者	消费者	4006	93.38
	律师	119	2.77
	医师	70	1.63
	药师	44	1.03
	其他卫生专业人员	42	0.98
	不清楚	9	0.21

2.2 药品不良事件信号挖掘结果

将筛选后得到 9277 例药品不良事件报告，使用 ROR 和 PRR 法进行计算，合并同名并删除非 ADE 的 PT，再以 a≥3，95%CI（ROR）下限>1 为筛选条件，共得到 109 个信号。奥利司他致药品不良事件报告数和信号强度居前 50 位的信号详见表 4。

表 4 奥利司他不良事件的发生频次 PT 排序和信号强度前 50 位排序

发生频次排序	信号强度排序

PT	报告数/例	95%CI(ROR)	95%CI(PRR)
药物无效	1600	8.56	7.31
脂肪痢	571	3321.93	3127.73
治疗不服从	480	60.46	57.64
产品质量问题	412	14.67	14.12
直肠分泌物	412	1930.08	1850.26
腹泻	367	3.24	3.16
体重增加	218	5.40	5.31
上腹痛	176	4.86	4.80
便秘	166	4.33	4.28
胃肠疾病	161	11.91	11.75
粪便变色	157	39.71	39.15
肠胃气胀	150	15.30	15.11
不适	140	1.40	1.39
产品给予错误	126	16.00	15.83
未明确定义的疾病	100	13.62	13.51
肛门失禁	97	45.73	45.35
排便频率增加	92	19.61	19.46
腹痛	86	2.01	2.00
不良药物反应	78	5.01	4.99
意外的治疗反应	68	5.50	5.47
腹胀	67	3.29	3.27
腹部不适	51	1.41	1.41
不良事件	49	2.26	2.25
粪便异常	44	25.63	25.54
产品投诉	44	14.75	14.70
肌痉挛	42	1.00	1.00
疑似假冒产品	41	46.77	46.63

PT	报告数/例	95%CI(ROR)	95%CI(PRR)
脂肪痢	571	3321.93	3127.73
直肠分泌物	412	1930.08	1850.26
脂尿	5	1347.76	1347.29
肛门失禁	38	102.39	102.10
治疗不服从	480	60.46	57.64
疑似假冒产品	41	46.77	46.63
白色便	24	46.19	46.12
肛门失禁	97	45.73	45.35
粪便变色	157	39.71	39.15
排便习惯改变	17	38.01	37.98
治疗产品效果不稳	11	31.02	31.01
粪便异常	44	25.63	25.54
急诊检查	8	24.70	24.69
紧急排便感	30	24.61	24.57
产品伪造	9	24.51	24.50
肾小管萎缩	5	20.86	20.86
排便频率增加	92	19.61	19.46
产品给予错误	126	16.00	15.83
肠胃气胀	150	15.30	15.11
产品投诉	44	14.75	14.70
产品质量问题	412	14.67	14.12
未明确定义的疾病	100	13.62	13.51
维生素缺乏症	6	12.71	12.71
胃肠疾病	161	11.91	11.75
绝经期症状	6	9.02	9.02
丘脑梗死	3	8.63	8.63
药物无效	1600	8.56	7.31

给用了过期产品	40	3.90	3.90	胰腺炎复发	3	8.07	8.07
肛门失禁	38	102.39	102.10	胆囊切除术	15	7.32	7.32
紧急排便感	30	24.61	24.57	大便不规律	15	7.31	7.31
消化不良	26	1.21	1.21	胆囊病变	21	6.67	6.66
给用了额外剂量	26	3.62	3.61	急性肝炎	10	6.58	6.58
肝脏疾病	26	2.87	2.86	药物对未经批准的指征起效	14	6.50	6.50
胃部疾病	25	2.30	2.29	体重减轻不足	6	5.99	5.99
白色便	24	46.19	46.12	硬便	7	5.50	5.50
色素尿	22	3.97	3.97	意外的治疗反应	68	5.50	5.47
不舒服	22	1.81	1.81	体重增加	218	5.40	5.31
肝酶升高	22	1.60	1.60	粪便松软	11	5.29	5.29
胆囊病变	21	6.67	6.66	不良药物反应	78	5.01	4.99
便血	20	1.62	1.62	上腹痛	176	4.86	4.80
胆石症	19	2.73	2.73	肢体畸形	4	4.59	4.59
排便习惯改变	17	38.01	37.98	便秘	166	4.33	4.28
大便不规律	15	7.31	7.31	眼黄疸	6	4.31	4.31
胆囊切除术	15	7.32	7.32	胃肠鸣音异常	6	4.27	4.27
用药错误	15	1.51	1.51	暴发性肝炎	4	4.14	4.14
肾石症	15	1.28	1.28	色素尿	22	3.97	3.97
药物对未经批准的指征起效	14	6.50	6.50	给用了过期产品	40	3.90	3.90
憩室炎	13	1.45	1.45	肝坏死	4	3.83	3.83
肝损伤	12	2.94	2.93	血胆固醇降低	4	3.74	3.74
粪便松软	11	5.29	5.29	大便次数少	4	3.67	3.67

2.2 药品不良反应累计系统

根据 MedDRA 的系统器官分类（SOC）对有信号的 PT 进行 SOC 分类排序，有信号的 SOC 为 15 个，不良事件信号累及系统与新的 ADR 信号详见表 5。

表 5 奥利司他导致药品不良事件信号与累及器官和（或）系统

累及器官、系统	有信号的 PT	不良反应种类/种	报告数/例	构成比/%

胃肠系统疾病	腹部不适、腹胀、腹痛、上腹痛、粪便异常、肛门失禁、大便不规律、排便习惯改变、便秘*、紧急排便感、腹泻、消化不良、上腹不适、粪便变色、硬便、粪便松软、肠胃气胀、排便频率增加、功能性胃肠紊乱、胃部疾病、胃肠疾病、胃肠痛、胃肠鸣音异常、牙龈疾病、牙龈疼痛、便血、大便次数少、吸收不良、胰腺疾病、胰腺炎复发、直肠分泌物、脂肪痢	34	2851	30.73%
全身性疾病及给药部位各种反应	不良药物反应*、不良事件*、不良反应*、假死*、活动减少*、不舒服*、治疗产品效果不稳*、药物对未批准的指征起效*、药物无效*、饿*、未明确定义的疾病*、不适*、治疗产品效果不稳*、意外的治疗反应*、治疗不服从*	15	2589	27.91%
产品问题	产品投诉、产品伪造、产品质量问题、疑似假冒产品	4	506	5.45%
各类检查	血胆固醇降低、急诊检查、肝酶升高、体重增加*	4	252	2.72%
各类损伤、中毒及手术并发症	儿童意外暴露于产品、产品给予错误、给了过期产品、给了额外剂量、用药错误	5	213	2.30%
肝胆系统疾病	急性肝衰竭、自身免疫性肝炎*、胆石症*、胆囊病变*、肝坏死、肝痛、肝脂肪变性、肝炎、急性肝炎、胆汁淤积性肝炎*、暴发性肝炎、黄疸*、肝脏疾病、肝损伤、眼黄疸*	15	147	1.58%
肾脏及泌尿系统疾病	色素尿*、脂尿*、肾石症*、肾病、少尿、肾疼痛、肾小管萎缩	7	68	0.73%
肌肉骨骼及结缔组织疾病	肢体畸形*、肌肉萎缩、肌痉挛	3	51	0.55%
代谢及营养类疾病	贪食*、维生素缺乏症*、食欲增加*、肥胖*、体重减轻不足*	5	29	0.31%
各种手术及医疗操作	胆囊切除术*、胆囊手术*、肝移植*	3	22	0.24%
感染及侵染类疾病	憩室炎*、疖	2	19	0.20%
神经系统疾病	嗅觉丧失*、脑静脉血栓形成*、协调功能失常*、丘脑梗死*	4	18	0.19%
皮肤及皮下组织类疾病	皮肤萎缩、皮肤气味异常*、皮肤紧缩感、黄色皮肤	4	14	0.15%
血管与淋巴	低血容量性休克、凝血障碍	2	12	0.13%

管类疾病				
生殖系统及乳腺疾病	绝经期症状	1	6	0.06%
精神病类	自我伤害意念*	1	4	0.04%

注：“*”代表未在奥利司他说明书中出现的 PT

2.3 产品问题

本研究还检索到 506 例（占比 5.45%）产品问题的 ADE，涉及产品投诉、产品伪造、产品质量问题、疑似假冒产品，详见表 6。

表 6 奥利司他导致产品问题的不良事件

类别	报告例数/N	构成比/%
产品投诉	44	8.70%
产品伪造	9	1.78%
产品质量问题	412	81.42%
疑似假冒产品	41	8.10%

3 讨论

3.1ADE 人群特点

本研究共检索到以奥利司他为首要怀疑药品的不良事件共计 9277 例。91.72%的报告来自美国，93.38%的报告者为消费者，可以看出 ADE 的大部分数据从美国获取，其他国家或种族的不良事件可能有所差别，报告主体与中国有所差异，中国目前的上报主体为药品生产企业、药品经营企业、医疗机构，患者报告途径的建立是提高数据可利用性的一个方向[9]。患者性别上看，女性发生率较高，可能与女性使用奥利司他的样本量较大有关，目前暂无性别相关研究，需要数据的支持。18 岁以下不良事件的发生率为 0.37%，提示儿童使用奥利司他的安全性相对较好，为儿童的不良事件提供了研究资料。

3.2ADR 信号分析

本研究得到的奥利司他不良反应与药品说明书、文献研究等相关安全性信息基本一致，主要累及的主要器官系统是胃肠道系统、肝胆系统、泌尿系统、肌肉骨骼系统等，表现为脂肪痢、直肠分泌物、脂尿、胰腺炎复发、急性肝衰竭。美国食品药品监督管理局于 2010 年 5 月警告其存在可能引起严重肝损害的风险，2011 年 3 月，国家药品不良反应监测中心发布了第 36 期《药品不良反应信息通报》，提醒医务人员关注奥利司他安全性问题，提出了上市后监测到严重肝损伤和胰腺炎的报道，与本文检索结果一致[10-15]。提示临床在使用奥利司他期间应该密切监测患者肝功能及胰腺情况。

通过比例失衡法计算可以看出，信号越强表明药物与不良反应之间的相关性越大[16]，在不良事件信号中，脂肪痢、直肠分泌物的信号强度和发生频次均相对较高，该提示临床医生在诊疗过程中，如果患者出现脂肪痢的情况，优先考虑是由奥利司他所致的药品不良反应。五年的研究数据中，脂尿的报告例数只有 5 例，但其信号强度极强，脂尿是由奥利司他所致的不良反应同样作为临床优先考虑项。脂肪痢、直肠分泌物、脂尿均属于胃肠道系统的不良反应，在系统器官分类中，奥利司他导致胃肠道系统的不良反应，共计 2851 例，占比 30.73%，该类不良反应的发生与奥利司他阻止脂肪吸收的药理作用相关[17-18]，提示临床在使用奥利司他应注意密切监测胃肠道系统的不良反应。

3.3 新的药品不良反应

奥利司他所致新的不良反应主要涉及的 7 个系统为：胃肠道系统、肝胆系统、泌尿系统、肌肉骨骼及结缔组织疾病、代谢及营养类疾病、神经系统、皮肤及皮下组织类。还检索到全身性疾病及给药部位各种反应，提醒临床在使用奥利司他期间出现以上不良反应时应进行密切监测。

3.4 与药品本身相关的 ADE

本研究除了检索到奥利司他所致的药品不良反应外，也检索到了全身性疾病及给药部位各种反应、产品问题等，药物无效发生的频次最高，治疗不服从、产品质量问题依次位于发生频次的第三、第四。在信号强度排序中，治疗不服从、疑似假冒产品位于前十位中。通过以上研究数据的检索，我们发现市面上奥利司他存在药物治疗效果不佳、不合格药品的现象，奥利司他治疗的有效性上值得医务人员引起重视，对于服用奥利司他的患者应对其进行随访，关注药物的治疗效果，同时提醒药品采购工作上务必核对药品的药品生产企业营业执照、药品生产许可证、药品 GMP 证书、药品注册批件、省级药检所药品质量检验报告书等相关资质证件，以确保药品的安全性。处方开具上，医生应严格参照药品说明书和指南推荐给药，提高患者用药依从性，保障药物的安全有效。

3.5 研究局限性

本研究利用美国 FAERS 数据库为自发呈报系统，数据库容量大，但是仍具有局限性，该数据库无法提供使用奥利司他未发生不良反应的例数，无法计算其总体发生率[19]。该数据无法对患者的原患疾病、合并用药情况，及发生 ADE 的时间逻辑关系进行分析，且大部分由消费者进行上报，可能存在不准确的情况。本研究结果主要反应了欧美国家的研究结果，但是可为中国人群提供用药参考。

4 结论

本文检索美国 FEARS 数据库，通过 ROR 法、PRR 法得到奥利司他的不良事件信号，其中不良反应信号与药品说明书基本一致，胃肠道系统不良反应发生率最高，肝胆系统、泌尿系统、肌肉骨骼及结缔组织疾病等系统方面新的不良反应值得临床关注。奥利司他出现药品无效、产品质量问题等信号需要密切关注，在药品购进时应严格把控，保障药品安全性。

PU-070

基于 VGG16 模型的图像情感分类方法

王瑶、马梦楠、周毅
中山大学中山医学院

目的 为了对图像进行情感分类，提出了一种基于 VGG16 模型的图像情感分类方法，并实现对图像情感所属类别进行预测的功能。

方法 首先加载预训练好的 VGG16 模型并对全连接层进行微调，构建情感分类器，然后将情感数据集输入到网络中进行训练，不断更新网络的参数与权重。最后，构建预测模块，使用构建的网络模型对输入图像进行情感类别预测。

结果 对图像进行正负极性分类，分类精度达到 91.7%，正向情感预测精度为 98.9%，负向情感预测精度为 95.5%。

结论 基于 VGG16 模型实现图像情感分类是可行且合理的，能够为实现图像情感分类提供决策支持。

随着移动通信技术以及在线社交网络的飞速发展，人们可以访问越来越多的信息资源。互联网用户往往希望通过发表图像来分享他们的经历，表达他们对某事物的看法，这就为图像情感分类提供了很多潜在的领域[1]。然而，人类对感情的理解与机器的解释之间存在着巨大的差距，因此，跨越人机感情机制之间的鸿沟，将深度学习中的卷积神经网络与图像情感分类结合起来具有十分重要的现实意义[2]。

对图像进行情感分类，可以及时了解图像发布者的情感表达意图，进行暴躁症、精神分裂症等心理疾病的筛查[3]，帮助用户及时寻找专业咨询；也可以获悉用户对当前热点话题、政策方针的态度[4]，进行网络舆论监督，提升互联网的稳定性与安全性，促进网络健康稳定发展；同时能够用于艺术广告设计领域，预测消费者对产品的反应，或者帮助杂志编辑与专栏作家更快地寻找合适的配图或插图[5]。

1 数据与方法

1.1 数据与处理 图像处理的目的是滤除图像中的干扰信息，增强数据量，以便后期更清晰的提取到数据集的特征[6]。本文使用 artphoto 情感数据集。数据集共包含 806 张图像，图像名称即为所属情感类别。因实验需要，本文将其分为正负极性两种情感类别，其中正向情感图像 378 张、负向情感图像 428 张。并将每种类别的数据集分为训练集、测试集和验证集，每种集合中不同类型的情感图像分配情况如表 1 所示。使用 transform 函数对图像依次进行旋转、裁剪、翻转、归一化以及标准化来达到数据增强的目的，帮助模型正确的分类图像。接着，用数据集对模型进行训练，首先利用训练集计算卷积神经网络的网络参数，不断更新验证集上的损失和准确性，最后利

用测试集对训练好的模型的准确率进行验证，以确定最佳的超参数。本文采用监督学习的方法，将图像情感标签映射到整数编码文件中以便读取，其中，“1”代表正向情感“optimistic”，“2”代表负向情感“pessimistic”。

表 1 训练集、测试集和验证集中情感图像的分类情况

Emotion	train_dataset	valid_dataset	test_dataset
optimistic	318	30	30
pessimistic	368	30	30

1.2 训练分类器 本文使用微调的 VGG16 模型[7]，模型的整体框架如图 1 所示。模型中包含 13 个卷积层，并且在每个卷积层后使用 Relu 激活函数以及最大池化层处理，以防止过拟合、梯度消失等现象的产生。对已经预训练好的 VGG16 模型进行微调，定义 3 个新的全连接层为情感分类器。因为仅需进行正负两类情感的分类，所以前 2 个全连接层有 4096 个神经元，第 3 个全连接层输出为 2。且最后一个全连接层使用 Log-Softmax 函数进行分类，能够解决函数 overflow 和 underflow，加快运算速度，提高数据稳定性，对网络参数进行优化[8]。所有卷积层均采用 3x3 卷积核，使得训练过程中各层的特征图保持不变。训练过程中，只更新前馈网络的权重，人工设置不同的超参数（学习率、batchsize、epoch 等等）来寻找最优的模型[9]。

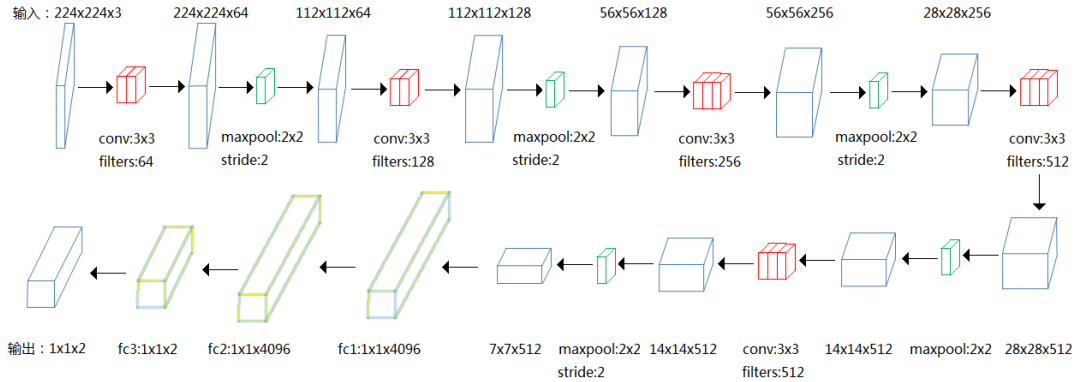


图 1 微调的 VGG16 模型结构图

对图像数据处理完成后，便可构建和训练分类器了。首先加载预训练好的 VGG16 模型，提取出输入图像的特征信息。使用预训练的网络能够减小工作量，拓展模型的适用范围，同时也能更专注于用例细节，提高实验精度[10]。特征提取主要在卷积层进行，卷积层将输入图像放进一组卷积过滤器，每个卷积器激活图像中的某些特征[11]。卷积器的参数一般初始化为固定值，之后通过不断地学习更新，调整为最适合的参数，从不同的角度提取图像特征并输出。

之后使用新构建的情感分类器对模型进行反向训练，理解网络模型提供的特征，不断改进参数，对网络进行微调，得到最好的训练效果。训练分类器是对大量标注好的样本数据进行观察，使用这些样本数据来学习每一个类长什么样。在训练过程中，先训练低层次的简单网络，再用得到的权重来初始化后面的高层次模型，提取出更加抽象、更加复杂的特征，减少参数，加速网络

的收敛[12]。在匹配过程中，将图像规模化为统一的尺寸，再输入到卷积网络之中。使用滑窗在图像中不断的进行滑动，然后根据子窗口内的区域特征计算结果，对各个不一样的窗口的归类效果进行均衡化，接下来平均各个尺寸所得出的数值结果并得出最后的理论数值结果。

1.3 分类预测 编写函数来使用经过训练的网络对输入图像进行预测，将需要进行预测的图像传递到网络中，并在图像中预测情感的种类。该函数接受图像和模型，然后返回最有可能的类以及概率。首先需要处理输入图像，以便它可以在网络中使用。使用 PIL 加载图像，这一功能应与训练所用同样的方式处理图像。预测并计算类的概率，然后找到最大值。为了要获得张量中的最大值，使用 topk()函数，该方法既返回最高 k 概率，又返回对应类的概率的指数。使用 class_to_idx 将这些整数编码索引转换为实际的情感类别标签，并添加到模型中，或者从用于加载数据的 ImageFold 转换。最后，将图像标准情感与预测的情感进行对比，并通过条形图的方式准确直观地展现出实验的准确度。

2 实验结果与分析

2.1 模型建立与实验 使用搭建好的 VGG16 网络对分类器进行反向传播训练，每一次训练结束后使用验证集对模型进行评估，输出每次 epoch 结束后的损失度以及精确度。可以从实验结果中看到，训练损失度以及验证损失度逐渐减小，验证精度则处于不断增大的状态，最大达到 85%，说明模型效果较好，基本能满足要求。如图 2 所示为模型训练结果。

```
Starting training
Training epoch # 1
Epoch: 1/5..  Training Loss: 2.955..  Validation Loss: 1.112..  Validation Accuracy: 0.500
Saving model
Best accuracy so far: 0.500
Epoch: 1/5..  Training Loss: 0.354..  Validation Loss: 0.725..  Validation Accuracy: 0.500
Training epoch # 2
Epoch: 2/5..  Training Loss: 0.310..  Validation Loss: 0.594..  Validation Accuracy: 0.617
Saving model
Best accuracy so far: 0.617
Epoch: 2/5..  Training Loss: 0.286..  Validation Loss: 0.489..  Validation Accuracy: 0.783
Saving model
Best accuracy so far: 0.783
Training epoch # 3
Epoch: 3/5..  Training Loss: 0.270..  Validation Loss: 0.389..  Validation Accuracy: 0.817
Saving model
Best accuracy so far: 0.817
Epoch: 3/5..  Training Loss: 0.272..  Validation Loss: 0.380..  Validation Accuracy: 0.817
Training epoch # 4
Epoch: 4/5..  Training Loss: 0.260..  Validation Loss: 0.398..  Validation Accuracy: 0.850
Saving model
Best accuracy so far: 0.850
Epoch: 4/5..  Training Loss: 0.244..  Validation Loss: 0.489..  Validation Accuracy: 0.850
Training epoch # 5
Epoch: 5/5..  Training Loss: 0.249..  Validation Loss: 0.398..  Validation Accuracy: 0.817
Epoch: 5/5..  Training Loss: 0.228..  Validation Loss: 0.488..  Validation Accuracy: 0.817
Epoch: 5/5..  Training Loss: 0.247..  Validation Loss: 0.398..  Validation Accuracy: 0.817
Done training
```

```
Loading checkpoint from  vgg16_emotion_classifier.pth
Model vgg16. Validation accuracy: 0.850. Achieved on epoch: 3
```

图 2 模型训练结果

测试并输出最终的网络参数。通过网络运行测试集并测量准确性，判断所训练的网络的准确度并输出，模型精度达到 91.7%，取得了很好的效果。

2.2 模型验证与结果 对需要进行预测的图像进行预处理，调整图像的大小与长宽比，对其进行与测试数据集同样的操作，最终提取并输出图像最能表示情感特征的区域。图 3 是正负两种情感极性图像的原图像与预处理图像对比图。

图 3 原图像与预处理图像对比

对预测图像的情感进行预测，并输出每种情感可能的概率，正向情感类别标签为 1，负向情感类别标签为 2，最终用直方图直观展示出来。如图 4 和图 5 所示为预测精度及直方图展示。可以看到，对正向情感图像预测为 optimistic 的概率达到 98.9%，对负向情感图像预测为 pessimistic 的概率达到 95.5%，准确度非常高，说明模型取得非常显著的效果，证实了本实验的可行性。

	optimistic	pessimistic
probability	[0.989 0.011]	[0.955 0.045]
catagory	[1 2]	[2 1]
	(a)	(b)

图 4 情感图像预测

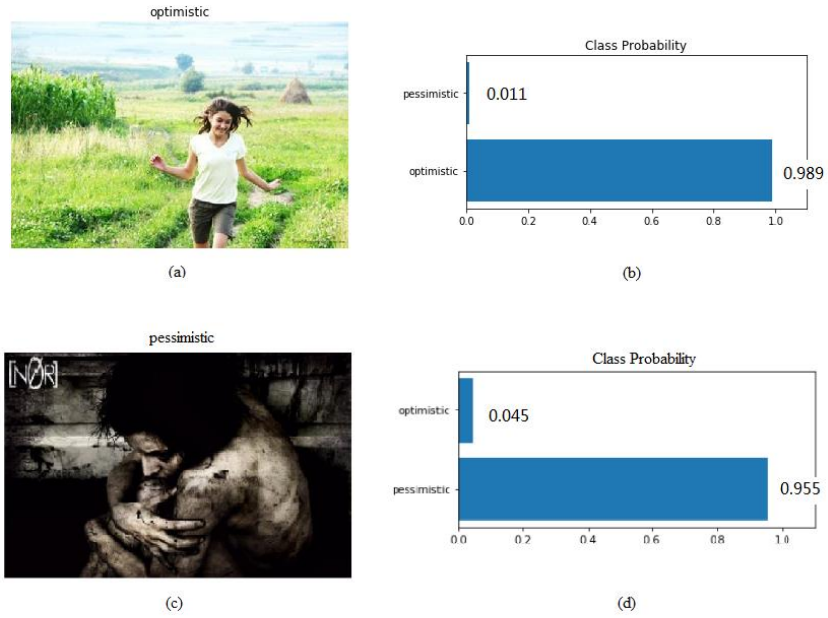


图 5　情感预测精度直方图

3 结论

本文主要对图像情感语义进行高维抽象性特征的选择研究，结合图像情感语义，对图像进行正负感情极性分类和预测。一方面基于 VGG16 网络构建用于图像情感分类的模型，另一方面构建预测模块，精准预测出图像的情感内容，最终证实模型的可行性与有效性。但是本文使用的数据集与网络模型较为简单，对图像进行更深层次的分类时能否保持优良的分类效果还有待研究。同时，视频信息比图像蕴含着更加丰富的内容，进行视频情感分类也是一个比较有前景的研究方向，在下一步工作里，将进行更加深入的研究，进一步拓展算法的实用性，达到更好的效果。

PU-071

国际生物信息学领域发展脉络研究及热点探析

许丹
中国医科大学图书馆

目的/意义　梳理国际上生物信息学领域发展至今的脉络态势及热点研究，为行业领域内科研人员提供文献信息支撑依据，也为我国未来生物信息学科的发展提供建议和指导。

方法/过程　以 SCIE11 种国际生物信息学期刊 1998 年-2019 年发表的文献为信息源，结合文献计量学方法对发文国家、机构等情况进行阐述。利用 Histcite、Pajek 对 1998 年-2014 年文献进行脉络梳理及可视化分析并筛选出主要路径。运用 BICOMB、gCluto 对 2015 年-2019 年文献进行关键词聚类分析来探析研究热点。

结果/结论 根据引文编年图和主要路径中的文献信息总结出 1998 年-2014 年生物信息学领域 9 个方面的研究热点并筛选出 3 条主要路径；获得 2015 年-2019 年生物信息学领域的 6 类研究热点。

1. 引言

生物信息学（Bioinformatics，简称 BI）是一门新兴的生物学科，研究生物学数据收集和处理的技术发展，并利用这些数据进行生物学探索和预测。该学科涵盖了所有适用于分子生物学的计算方法和理论，以及用于解决包括模型与数据集处理等生物学问题的计算机技术[1]。生物信息学以计算机技术为研究手段和工具，同时采用数学、统计学的模型、模拟研究手法来解决生物科学的问题，因而成为生物学、统计学、数学、计算机学甚至工程学等多学科之间的交叉领域[2]。

生物信息学是计算机科学和生物学研究的交集核心，人类基因组相关计划吸引了许多专家，带来了生物医学研究和卫生保健领域的新挑战[3]。它是当今生命科学和自然科学的重大前沿领域之一，同时也将是 21 世纪自然科学的核心领域之一。其研究重点主要体现在基因组学(Genomics)和蛋白质组学(Proteomics)，从核酸和蛋白质序列出发，分析序列中表达结构功能的生物信息，包括生物分子数据库的收集、存储、管理、挖掘，基因组信息的获取、处理、存储、分配和解释[4]。

我国有学者对国内外生物信息学研究领域进行了探索。2004 年张晗[5]应用共词聚类分析方法对 MEDLINE 中生物信息学文献高频主题词进行聚类获得学科研究热点内容。宋茂海[6]等运用共词分析结合多维尺度分析对 1998-2013 年间国内生物信息学热点研究变化趋势进行了探析。张长宏[7]采用基于短语的主题模型对 1998-2015 年国际生物信息学期刊进行主题提取及可视化主题演化研究。游鸽[8]利用 CiteSpace 对 2005-2014 年间 SCI 收录的生物信息学 5 种高影响力外文期刊论文进行统计和可视化，分析揭示生物信息学领域的研究热点、研究前沿以及知识基础。他们的研究涉及了国内外生物信息学领域论文研究热点主题及变化趋势，但存在研究时间较早、分析内容不足、不够全面等问题。

笔者利用 Histcite 等软件以 SCIE(Science Citation Index Expanded，SCI 扩展板)核心期刊为基础对国际生物信息学领域的核心文章进行脉络梳理，直观展示该研究领域的论文渊源及发展脉络，发现该领域中的沿革继承关系和发展趋势并绘制学科领域的主要路径，再对前者未涵盖时间的发文情况进行聚类分析探索研究热点。该研究为生物信息学领域科研人员提供了可靠的研究文献信息支撑，有利于信息专业人员为用户提供参考咨询服务，帮助专业研究人员跟踪专业研究动向，从而促进我国生物信息学领域的发展和进步。

2. 数据来源与研究方法

2.1 数据来源

本研究数据来源为 SCIE 数据库中 11 种生物信息学期刊 1998 年-2019 年发表的文献。根据 2018 年最新版本期刊引证报告（JCR，Journal Citation Reports）选取分类为“Mathematical& Computational Biology”，再筛选出影响因子高于 MIF（Median Impact Factor）的期刊。因 JCR 所属分类中包含其他学科会对研究产生偏移，故将筛选期刊与 PubMed 期刊中学科字段（Broad

Subject Term）相对照，最终保留学科字段中含有Computational Biology的11种高影响力期刊论文作为研究对象（见表1）。在SCIE中按照期刊名称进行检索，共检索得到文献35 594篇。

表1 生物信息学领域11种核心期刊情况

序号	期刊名称	创刊年	SCI收录年	出版国家	发文量	2018 IF值
1	BRIEFINGS IN BIOINFORMATICS	2000	2004	England	1141	9.101
2	Wiley Interdisciplinary Reviews-Computational Molecular Science	2011	2011	United States	388	8.127
3	BIOINFORMATICS	1998	1998	England	13545	4.531
4	PLoS Computational Biology	2005	2005	United States	6202	4.428
5	Database-The Journal of Biological Databases and Curation	2009	2009	England	1036	3.683
6	BMC BIOINFORMATICS	2000	2000	England	9692	2.511
7	Molecular Informatics	2010	2010	Germany	714	2.375
8	Frontiers in Computational Neuroscience	2007	2007	Switzerland	1207	2.323
9	BioData Mining	2008	2008	England	311	2.301
10	Evolutionary Bioinformatics	2005	2005	United States	439	2.203
11	Computational Intelligence and Neuroscience	2006	2011	United States	919	2.154

2.2 研究方法

2.2.1 工具介绍

Histcite对SCI、SSCI、A&HCI数据库的检索结果进行分析和重组，可用于确定某研究领域的重要机构、作者等信息，展示专业领域关键事件的年代顺序及相互引用关系，让读者快速了解某个主题的历史发展和变化趋势[9]。该软件研究本质就是对某领域内研究文献选取同行专家的推荐引

用情况进行分析。本研究使用中国科学技术大学罗昭锋开发的 Histcite Pro 2.1 软件包[10]以 LCS（Local Citation Score,本文献集内引用次数）指标为研究基础开展分析。

Pajek 是一个专门为处理大数据集而设计的网络分析和可视化程序，可以生成一系列交叉网络进行分析并考察网络的演化[11]。主要路径分析可以鉴别出整个引文流中最相关的文献，强调哪些文献是该特定主题的发展核心，在该主题发展过程中产生主要影响，是科学领域研究重建的重要部分[11]。本研究使用 Pajek1.2.6 软件进行主要路径分析，结合 Histcite Pro 对 1998 年-2014 年文献进行脉络梳理及可视化分析。

BICOMB（书目共现分析系统）是一种文本挖掘工具[12]，可对医学文献数据库中的书目信息进行快速读取、准确提取字段并归类存储、统计，并生成书目数据的共现矩阵，为进一步研究提供全面、准确、权威的基础数据。本研究利用 BICOMB 对 2015-2019 年间文献提取高频关键词，并利用 gCluto 软件进行双聚类分析获得领域研究热点主题。

2.2.2 实施过程

本研究分为三个步骤：

第一步，绘制引文编年图谱。将 SCIE 收集数据进行清洗去重等处理后，导入 Histcite Pro 中，设置阈值绘制引文编年图。

第二步，获取主要路径。运行 Pajek1.2.6 软件，导入 Hsitcite Pro 保存的.net 格式文件，运用“Critical Path Method”功能获得研究领域主要路径。

第三步，关键词聚类分析获取研究热点。用 BICOMB 软件，对关键词数据进行统计，生成高频关键词词篇矩阵，随后运用 gCluto 软件进行聚类分析。

3.结果与结论

3.1 文献计量学分析

3.1.1 生物信息学领域发文时间及引用情况分布

文献作为信息传播的主要载体方式，发文量及引用情况的时间演化规律与该学科的发生、发展情况有着密切的关联[13]。图 1 显示了生物信息学领域发文量及引用情况随时间发展的变化规律。生物信息学发展起步于 1998 年，目前仍属于一个相对新兴学科，研究论文逐步积聚至今已具一定规模。1998-2004 年是该学科发展的起步上升时期，年发文量在 100-800 不等，2005 年起是其发展壮大时期，每年发文量数以千计，得力于科学技术的飞速发展以及各国对科研人员力量和科研经费的大量投入。被引情况则由起步时期的缓慢上升，到 2005 年被引频次达到了峰值，随后呈下降趋势，在 2009 年经过一个上升拐点后继续呈下降趋势直到 2019 年。图 1 中可以看到 2005 年年发文量首次突破 1000，各国学者前期积累在这一年爆发，促进了生物信息学的发展壮大，被引量也居于最高。之后随着年发文量的增加及文献总量的积累，带来了总被引频次的降低。2009 年，可能因开发 Burrows-Wheeler 程序，实现了对人类基因组等大的参考序列的有效读取比对[14]，属于生物信息学领域的重大突破，累计被引 842 次，致使 2009 年下降趋势出现拐点。2010 年之后的文献由于研究主题的拓展和分散，使得文献总被引频次继续保持逐年下降的趋势。

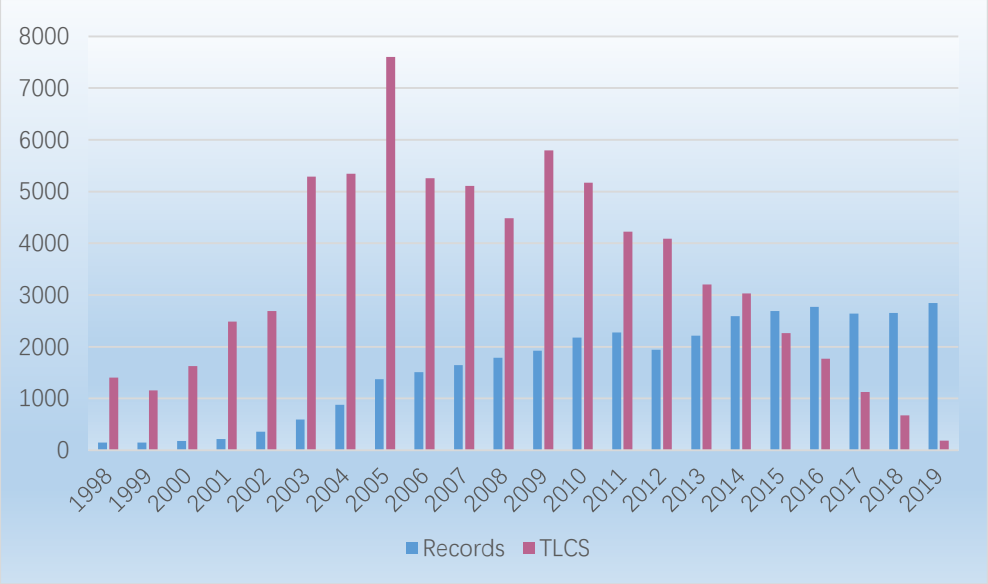


图 1　生物信息学领域发文时间及引用情况分布

注：Records 代表当年发表文献记录数，即当年发文总量，TLCS（Total Local Citation Score）代表当年发表文献在本文献集内的总被引用次数。

3.1.2 生物信息学领域发文国家、发文机构分布

通过对 35 594 篇 SCIE 发表论文的国家和机构进行统计得知，共有来自 130 个国家和 12362 个机构的著者参与发文。表 2 列出了生物信息学领域中发文数量排名 TOP10 的国家和机构。该学科领域发表论文 TOP10 的国家主要分布于欧美发达国家和地区。这些国家无论是科学技术的发展还是人员力量的投入均领先于世界其他国家。发文数量排名 TOP10 的机构中美国占 6 所，英国占 3 所，还有 1 所来自中国，可以看到美国和英国在该研究领域中绝对的世界垄断地位。中国的发文数量和机构发文排名分列第 4 位和第 5 位，说明我国在生物信息学领域中已经取得了一定的国际地位。

表 2　生物信息学领域发文国家和发文机构的分布情况（TOP10）

序号	发文国家	发文量	TLCS	TGCS	发文机构	发文量	TLCS	TGCS
1	USA	15112	35165	722555	Harvard Univ	649	1605	42562
2	UK	4255	10884	228639	Univ Calif San Diego	586	1601	48462
3	Germany	3842	8136	170031	Univ Cambridge	536	1120	21116
4	Peoples R China	3541	4679	77200	Stanford Univ	516	1596	26307
5	France	2014	3734	96411	Chinese Acad Sci	513	958	20019
6	Canada	1616	3223	56758	Univ Oxford	457	893	22595

7	Italy	1478	1714	31713	Univ Washington	417	1348	10135
8	Spain	1336	1783	69586	Univ Calif Berkeley	399	2192	20506
9	Japan	1273	3216	51457	Univ Michigan	373	1115	45016
10	Australia	1225	2698	74138	UCL	371	645	25642

注：TLCS（Total Local Citation Score）表示该国家、机构发表文献在本文献集中的总被引次数，TGCS（Total Global Citation Score）表示本文献集中该国家、机构发表文献在整个 SCIE 中的总被引次数。

3.2 可视化引文分析

3.2.1 引文编年图

Histcite Pro2.1 将文献集内部被引频次 LCS 设定 COUNT=40，即选取生物信息学领域中被引频次最高的 40 篇文献按年代顺序绘制出研究领域编年图，时间覆盖了 1998 年至 2014 年，见图 2。通过图 2 可以看出，生物信息学中 1998 年节点数较少，说明研究正处于起步阶段，文献数量有限，参与著者数量也不多；1999 到 2007 年间，节点数目显著增加，更多文献得到了广泛的应用，说明了该领域的研究进入了活跃期，大量有价值的文献开始涌现，并且形成了一定的研究主题内容的连续。2008 年开始到 2014 年，该学科的发展处于一个快速发展时期，陆续涌现出不同的引用关系图标以及各种散点图标。

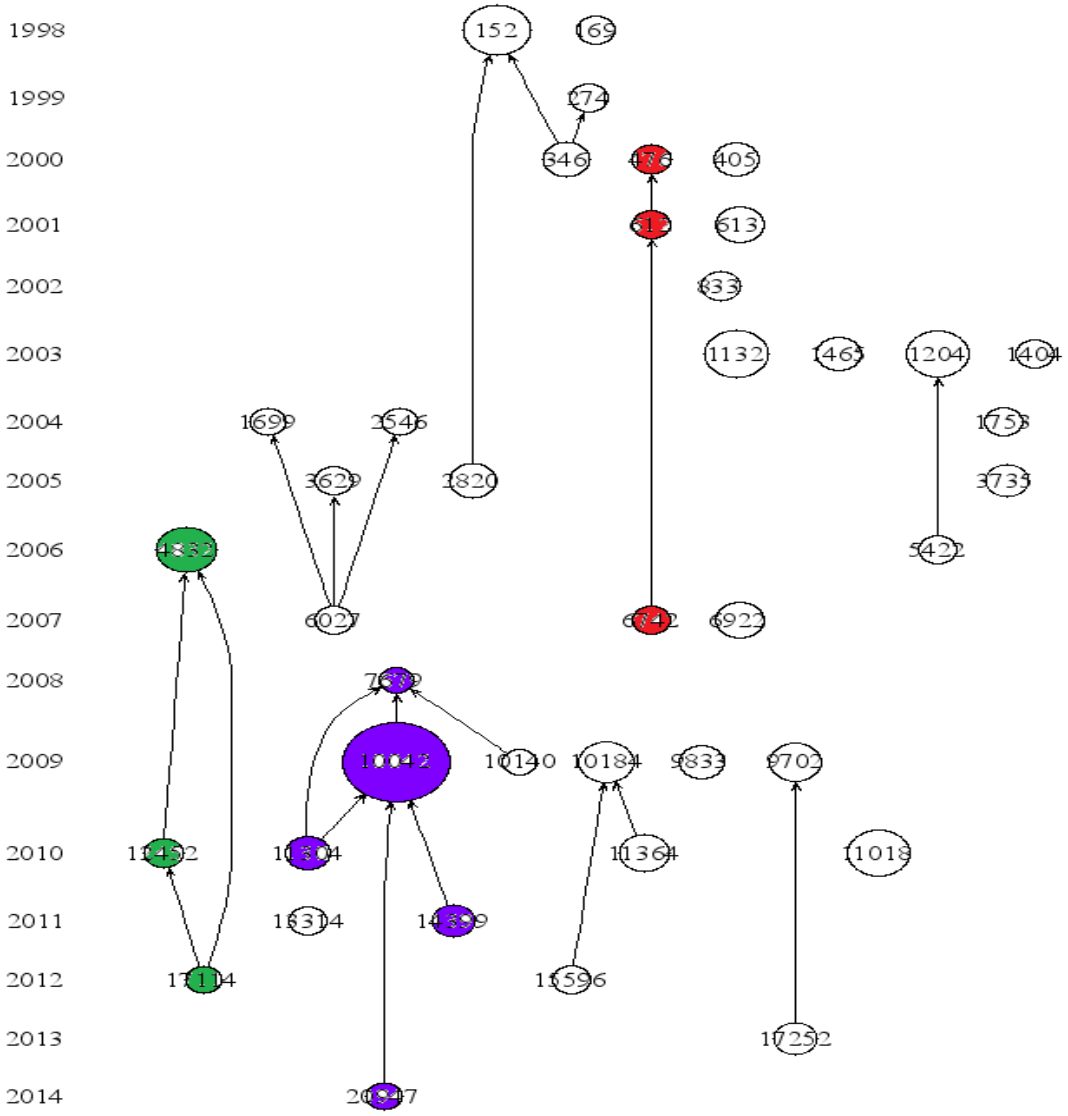


图 2 生物信息学引文编年图

通过分析引文编年图中被引频次 TOP40 的文献题目和摘要内容，以及文献之间的引用关系，总结概括出 1998 年-2014 年不同阶段呈现出的研究热点主题内容，如表 3 所示。

表 3 1998 年-2014 年生物信息学领域研究热点内容分析

研究时间范围	研究热点	来源文献序号
1998--2005	利用隐马尔可夫模型方法来检测蛋白质序列的远端同源性，包括 HMMs 方法、SAM-T98 方法以及 HMM-HMMs 对比方法[15-17]。	152、169、2820
1999--2003	DNA 序列比对及蛋白质二级结构分析及预测[18-21]。	274、346、405、1465
2000--2004	基因芯片（又称为 DNA 芯片，DNA 微阵列）检测技术的应用，包括检测基因表达水平、识别基因序列、发现新	476、612、613、1132 、1699、

	基因、检测基因突变等[22-28]。	1753
2003--2007	从基因本体论出发进行基因表达及数据分析[29-32]。	1404、2546、3629、6027
2006--2012	蛋白质、DNA、RNA 高通量序列比对和聚类[33-36]。	4832、17114、6922、12452
2008--2011	人类基因组长链、短链测序及序列比对的技术发展,包括 SOAP、SOAP2、Burrows-Wheeler、BWA-SW、SAMtools、edgeR、BEDTools、VCF[14, 37-43]。	7679、10140、10042、11304、10184、11018、11364、14399
2009--2013	RNA 测序研究，工具包括读图算法 TopHat 和 STAR[44-45]。	9702、17252
2012--2014	下一代测序技术开发，包括读取仿真模拟器 ART 及预处理工具 Trimmomatic [46-47]。	15596、20947
2002--2011	生物网络可视化、评估、统计、分析模型、规范化语言等软件工具包，包括 ROCR、Cytoscape、赖特-费雪遗传变异中性模型、COPASI 模拟平台、Python 工具库等[48-54]。	3735、13314、833、1204、5422、6742、9833

3.2.2 主要路径分析

从 Pajek 软件中获得 1998 年-2014 年生物信息学研究领域内的主要路径分为三条，如图 3 所示。作者将 Pajek 得到主要路径关键文献节点引入编年图中，即图 2 中所展示的彩色标识，直观展示出该领域内三条主要路径与其他文献的相互关系及引用关系。

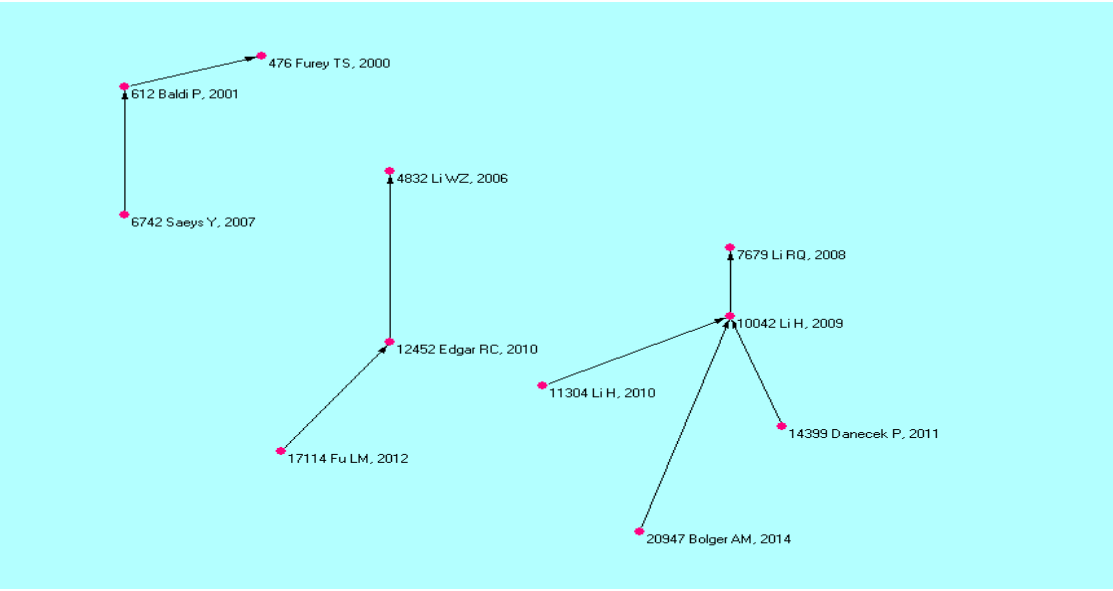


图 3 生物信息学引用关系主要路径图

第一条路径有 3 篇文献，文献序号是，“476”、“612”、“6742”，始于 2000 年截止于 2007 年，研究内容主要围绕在基因芯片检测技术进行的基因数据分析及表达预测。第二条路径有 3 篇文献，文献序号是“4832”、“12452”、“17114”，始于 2006 年截止于 2012 年，研究内容主要围绕在蛋白质、核酸高通量序列比对和聚类的各种程序及检测方法的进步。第三条路径有 5 篇文献，文献序号是“7679”、“10042”、“11304”“14399”、“20947”，始于 2008 截止于 2014 年，研究内容涉及基因组的测序技术、序列比对、表达、可视化等各类检测技术、工具的发展等。三条路径所展示的都是不同时间基于蛋白质、核酸、基因组的各种检测技术、方法的更新进步，可见应用技术的发展变革对于生物信息学领域的发展起到了至关重要的推动作用。

3.3 关键词聚类分析探析学科研究热点

2015 年-2019 年期间发表文献共有 13609 篇，占文献总量的三分之一，提取关键词共有 13979 个，截取阈值为 23 的 37 个高频关键词生成词篇矩阵，经聚类调整优化后最终获得 6 大类研究热点内容，详见表 4。

表 4 2015 年-2019 年生物信息学研究领域关键词聚类分析

聚类编号	聚类名称	关键词	时间范围	代表文献
Cluster0	机器学习、脑电图、神经网络等深度学习方法在生物信息学领域的应用[55-57]。	machine learning、deep learning、bioinformatics、EEG、neural networks	2015-2017	参考文献 55、56、57
Cluster1	生物信息学领域中基因组学下一代测序技术的软件开发应用[58-59]。	Bioinformatics、Next generation sequencing、Metagenomics、Genomics、Software、NGS	2015-2016	参考文献 58、59
Cluster2	从基因本体出发的基因表达以及 RNA 测序可视化和聚类分析[60-61]。	Visualization、Gene expression、Next-generation sequencing、Clustering、RNA-Seq、Gene ontology	2016-2017	参考文献 60、61、
、Cluster3	系统生物学中蛋白相互作用及药物研发的网络分析[62-64]。	Data integration、Database、Drug discovery、Systems biology、Protein-protein interaction、Network analysis	2016-2018	参考文献 62、63、64
Cluster4	ChIP-seq、RNA-seq 技术的深入学习挖掘来预测基因表达[65-66]。	ChIP-seq、gene expression、RNA-seq、Prediction、Text mining、Deep learning	2015-2018	参考文献 65、66

Cluster5	DNA 甲基化特征选择标记物来预测原发肿瘤以及定量构效模型建立[67-69]。	Classification、Feature selection、Data mining、Random forest、Cancer、DNA methylation、QSAR	2017-2018	参考文献 67、68、69

4.讨论与结论

4.1 研究方法

通常情况下，Histcite 绘制引文编年图结合 Pajek 筛选主要路径是针对整个学科领域进行的发展脉络探索。本研究通过对生物信息学领域 1998 年-2019 年被引频次最高的 40 篇文献绘制引文编年图，探析研究热点内容并进行主要路径筛选。但是从引文编年图中可见，被引频次最高的 40 篇文献均为 1998 年-2014 年发表的核心论文，对于 2015 年之后发表论文未能显示，可能是由于文章相对较新，积累被引频次不够而没有进入引文编年图的绘制，导致未在整个学科领域的发展脉络中体现出来。

本研究中 2015 年-2019 年论文不适合再次使用同样方法，无论是以 COUNT（引文数量）或 VALUE（被引阈值）再次设定阈值都不太恰当。为降低因发表时间过新带来的被引频次阈值不够产生的偏倚，笔者运用关键词聚类方法对 2015-2019 年间发表的生物信息学领域文献的研究热点主题再次进行探析，以保证该学科领域分析的整体连贯性。

4.2 生物信息学领域 1998 年-2014 年研究热点与 2015 年-2019 年研究热点对比

经过引文编年图和主要路径分析，作者总结出了 1998 年-2014 年间生物信息学领域内的 9 类研究热点内容。通过 2015 年-2019 年间的聚类结果对比发现，这 5 年间的研究热点有先前研究的发展延续，也有新兴的内容方向。Cluster0、Cluster1、Cluster2、Cluster4 是之前生物信息学领域中研究热点内容的延续。笔者认为对于各种方法的应用、检测技术和工具的更新换代将在生物信息学领域中一直保持着热度，并且将持续推动学科领域的快速发展。Cluster3、Cluster5 分别是这几年间新兴起的研究热点，包括药物研发分析以及 DNA 甲基化预测原发肿瘤。可以看出，随时代发展各种检测技术方法在不断更新迭代，研究内容方向也得到了极大的扩展，生物信息学科研究热点正在逐步向对人类治疗疾病有帮助的领域转变，更加注重研究成果的转化。

4. 3 结论

本研究按时间顺序对生物信息学领域中 35594 篇文献进行梳理获得学科发展脉络图谱，总结了生物信息学 20 多年来的发展态势。

利用 Histcite Pro 绘制出 1998 年-2014 年生物信息学领域的引文编年图谱，分析学科发展脉络，总结出 9 类研究热点内容，集中在基因芯片，基因表达及预测，蛋白质、核酸、人类基因组及 RNA 的测序、比对、聚类、可视化等检测技术方法发展等方面。

利用 Pajek 筛选出 1998 年-2014 年生物信息学领域的三条主要路径：一是基因芯片检测技术应用；二是蛋白质、核酸高通量序列聚类和比对；三是基因组的测序技术、比对、可视化等。

利用 BICOMB 及 gCluto 获得 2015 年-2019 年生物信息学研究热点分 6 类，集中在方法论在生物信息学领域应用、测序技术发展变革、基因表达及可视化、蛋白相互作用及药物研发、DNA 甲基化肿瘤预测等方面。

可以预见，生物信息技术和工具的发展决定着生物信息学领域发展的广度和深度，同时，文本挖掘的理论方法对生物信息技术和工具的发展也起到了辅助和推动的作用。生物信息学领域的研究对象从最开始的蛋白质、核酸比对，发展到后来的人类基因组测序，再延伸到如今对人类治疗疾病有帮助的领域，可见其与转化医学结合得更为紧密。

PU-072

基于本体的膀胱癌临床诊疗决策支持知识库构建

李星雨[1]、李锴文[2]、林天歆[2]、周毅[1]
1. 中山大学中山医学院
2. 中山大学孙逸仙纪念医院泌尿外科

目的 提高膀胱癌的临床诊疗决策有效性和高效性，明确其复杂的诊断分级、治疗手段及应用指征

方法 从临床需求分析出发，进行了科学而全面的知识采集，基于本体建模原理并借助 Protégé 工具构建膀胱癌专病知识库。

结果 在此基础上绘制知识图谱以进行膀胱癌知识的可视化，实现知识的查询、管理及共享等功能。

结论 通过对知识的结构化表示，达成了辅助医疗工作者进行膀胱癌临床诊疗的目的，为膀胱癌的医疗服务水平提高提供了坚实的保障。

1 引言

信息化已成为各行各业发展的大势所趋，其中医疗信息化的发展更是兼具深度与广度。目前医疗系统面临的主要问题在于临床数据的爆发式增长，医生很难从中抽提出全部的有效信息，继而造成了决策失误的发生。如何提高诊疗效率，为临床决策的科学性及有效性提供保障，是当今医疗行业发展的重中之重。

知识库（Knowledge Base）区别于普通数据库（Data Base）的最大优势，便在于它是针对某些领域内的实际问题需要，对海量数据进行整理与分析后的知识集合。知识库以描述性方法存储和管理了知识，便于知识的检索和利用，能够为临床医生和科研学者提供可靠而全面的知识源[1]。

随着 AI 技术的发展与临床需求的日益增长，知识库不仅局限于检索文献等简单用途，能够进行知识推理与诊疗决策建议的专科专病知识库成为研究热点。针对某一专科疾病或单一病种建立的知识管理系统，即为专科专病知识库。目前，我国已在慢性病如高血压[2]、传染病如肺结核[3]、孕产疾病[4]等领域的建立了专病知识库，但是成功得到临床应用的专科疾病领域仍十分局

限，临床的急切需求对知识库的开发和完善提出了更高的要求，其中又以病因复杂、治疗手段多样的癌症为甚。

膀胱癌是泌尿系统中最常见的恶性肿瘤，其发病率和死亡率在全球均居于前列，且呈现出逐年上升之势[5]。不同类型的膀胱癌根据复发风险和预后的不同进行风险程度分级，从而采取不同的治疗策略，包括经尿道膀胱肿瘤切除术（transurethral resection of bladder tumor, TURBT）、根治性膀胱切除术等，并结合不同的腔内辅助灌注化疗及全身放化疗方案等，其治疗手段多样，选择指征复杂，仅凭临床医生的经验判断较为困难，建立临床诊疗专病知识库，有利于为医生提供决策支持和依据，提高医疗质量。

因此，本研究拟建立膀胱癌诊疗专病知识库，帮助临床医生掌握现有循证医学证据，提高诊疗效率及诊治水平，切实解决临床应用中的难点。

2 膀胱癌专病知识库构建方法

2.1 知识库构建技术路线

研究根据本体建模原理及临床决策知识库构建方法，采取了由知识采集、知识表达、知识应用、知识评估四个步骤组成的技术路线方法。

具体架构如图 1 所示。

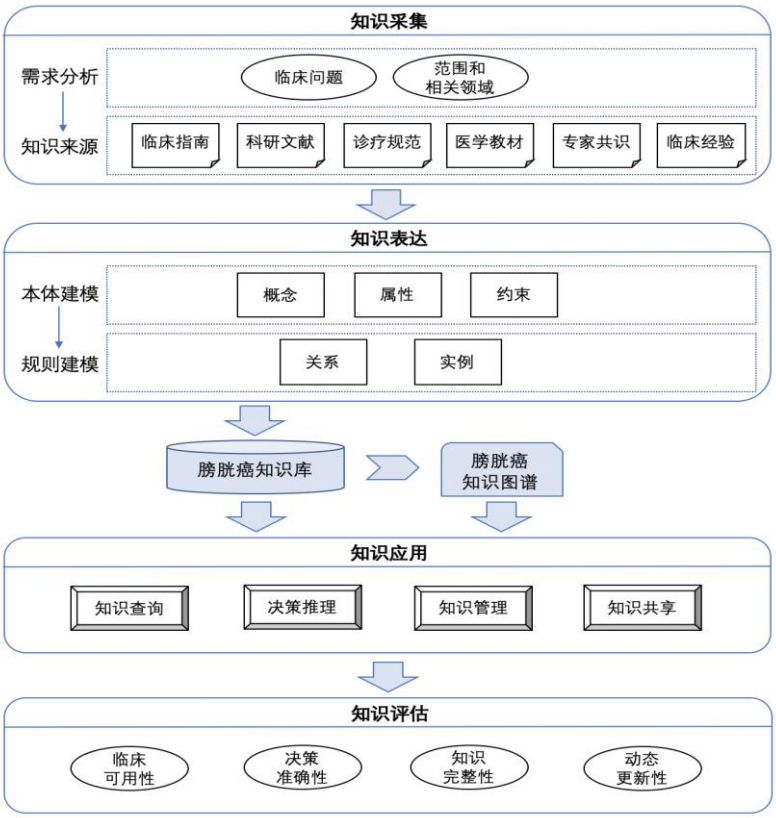


图 1 膀胱癌知识库构建技术路线

2.2 知识库构建步骤

2.2.1 知识采集

知识采集是建立在需求分析基础之上的，明确有待解决的临床问题及所涉及的知识领域，进而确定知识库的范畴，据此确立知识来源。对本研究而言，膀胱癌知识库旨在辅助医生进行诊疗决策，其范畴以膀胱癌专病知识为主，同时涉及泌尿科其他概念及药物、手术术式等知识。

基于以上需求，从多种渠道选取了知识来源，如临床指南主要参考了权威的欧洲泌尿外科学会(European Association of Urology，EAU)指南[6]、美国泌尿学会(American Urological Association，AUA)指南[7]、美国国立综合癌症网络(National Comprehensive Cancer Network，NCCN)指南[8]等；为了更切合国内情况，同样参考了国内发布的诊疗规范[9]及专家共识[10]；同时，通过在CNKI、万方、维普等中文数据库及PubMed、Elsevier等外文数据库中检索关键词的方法，搜集了近年来膀胱癌诊治相关的文献研究及循证医学证据，并参考了第九版外科学等教材；结合临床医生的建议和经验，作为知识库的知识来源。

将上述数据汇总后，进行清洗、梳理、筛选、总结、分析，从中抽取出过程模型和概念模型，构成元本体，成为知识库的数据来源。

2.2.2 知识表达

知识表达（Knowledge Representation，KR）是指采用形式化的方法描述或表达知识的过程，不仅能够方便理解，更是知识库建立的基础[11]。其中，本体是知识表达的重要方法，是对共享概念模型明确、形式化的规范说明[12]。本体的构建方法即搜集、整理并提取非结构化的知识，并以计算机可理解的方式表达出来，形成对领域知识、概念、实体及其关系的明确而规范的概念化描述。

本研究以本体表示语言表达元本体中的概念模型和过程模型，形成领域本体和语义规则，进而构建结构化的膀胱癌诊疗知识库。对比其他本体编辑工具（OilEd、OntoEdit、WebOnto、Ontolingua Server等），斯坦福大学开发的Protégé[13]使用更为方便快捷，其基于Java语言开发，能够在Windows、Linux、Mac OSX等多平台使用，不仅提供了图形化、可拓展的开发工具，也能支持本体查询及推理等功能，因此，本研究主要借助Protégé进行膀胱癌本体的构建。

具体而言，首先需要对膀胱癌领域知识的相关概念（Concepts）进行界定，明确类（Classes）的划分，搭建膀胱癌知识体系，然后补充各概念的属性（Properties）信息，并对各属性进行约束（Restrictions）。随后进行对象属性（Object Properties）定义，创建类间关系（Relations），并在此基础上为个体的属性赋值，即添加实例（Instances）。

1) 概念（类）：首先，分析并列举膀胱癌及泌尿外科领域的基本术语，如“NMIBC”、“TURBT”等，并以临床术语标准（Systematized Nomenclature of Medicine - Clinical Terms，SNOMED CT）为参考，形成基本概念体系。

2) 属性：明确各概念的属性，从而提供本体所需的逻辑分析基础。其中又可分为对象属性和数据属性，前者用于描述两个类的关系，如“血尿”是“膀胱癌”的常见症状；后者则描述数据和类的关系，如根据“肿瘤直径”对“危险程度分级”进行判断归类等。

3) 约束：确定属性中数据的类型、值域等，限制类的性质或类间关系。约束公理可分为值约束及基数约束，前者限制属性的值域，后者限制属性取值的个数。比如值约束中的allValuesFrom指被约束属性的所有取值都必须是其指定的类的实例，相当于全称量词。

4) 关系：描述了各概念之间的关联，可分为成员关系（part-of）、子类关系（kind-of）、实例关系（instance-of）及属性关系（attribute-of）四种主要类别。本研究使用了自顶向下法构建领域本体的类层次结构，例如，首先定义领域中最大的类，膀胱癌分类及分期等，随后将膀胱癌依据分期进一步划分“NMIBC”/“MIBC”并细分危险程度分级等。同时，领域中不仅存在严格的上下位关系，也包含大量的并行关系及实例关系。

5) 实例：是不能再次划分的最低逻辑层级的概念，表示某个类的具体实体。如“中国株治疗用BCG”即为“BCG”的实例，当针对某类型膀胱癌的治疗方案为“BCG膀胱灌注治疗”时，该实例即会自动匹配。

基于初步构建的本体原型，以上不同层次的构建过程往往需要迭代进行，以求不断完善知识库的基本架构，更符合实际需求。

使用本体方式表示知识能够很好地明确膀胱癌专病领域中相关的概念定义、属性与约束及概念的层次结构关系等，通过对以上五个本体的基本元素的描述，构建了结构化的膀胱癌知识库，基本实现对膀胱癌诊疗相关知识的语义规范化表达。

3 膀胱癌专病知识库构建结果

依据上述的构建技术流程，使用Protégé工具对膀胱癌相关知识进行了本体构建，成功搭建了包含类、对象属性及约束、相关关系及实例的本体结构，以展示膀胱癌知识结构层次。

3.1 膀胱癌本体类的构建

对于本体的基本结构，共创建3个大类（class），包括膀胱癌分类及分期（Staging_and_classification_of_bladder_cancer）、膀胱癌症状（Symptoms_of_bladder_cancer）及膀胱癌治疗方法（Treatment_of_bladder_cancer），并对各类分别作出定义。

3.2 膀胱癌本体对象属性及数据属性定义

对于3个大类中的概念定义属性和约束，明确相关概念。

1) is-part-of：表示包含关系，即部分属于整体；

2) have-symptom：表示膀胱癌相关症状；

3) related-disease：表示该症状能够提示膀胱癌的存在；

4) treat：表示治疗方法能够对某类膀胱癌起到积极作用；

5) need-treatment：表示某类膀胱癌适用的治疗方法。

在上述对象属性中，have-symptom 与 related-disease、treat 与 need-treatment 构成了两对相反（Inverse Of）的对象属性。

进一步的，对各类的数据类型进行了定义，比如对膀胱癌治疗方法大类的描述包括治疗类型（Type）、治疗名称（Name）、适应症（Indications）及禁忌症（Contraindications）等。

3.3 膀胱癌本体类层次结构关系

在经过迭代进行类的创建、属性和约束定义及关系描述这几个步骤后，膀胱癌本体的类及其层次结构已经逐渐清晰，图 2 为膀胱癌知识库的整体知识结构，并展示了膀胱癌治疗手段类的进一步细化结构。图 3 为膀胱癌本体结构图，其中，实线线段及箭头符号表示类间关系，不同颜色的虚线箭头表示不同的对象属性。在整体框架的基础上，再为各类添加实例。

图 2 膀胱癌本体类层次结构

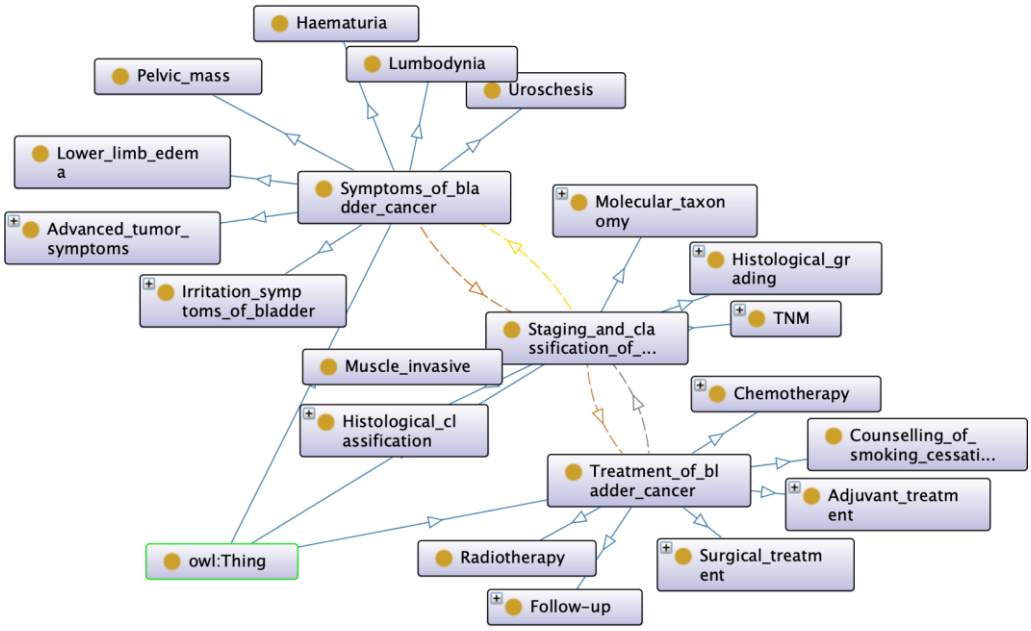


图 3 膀胱癌本体结构图

3.4 膀胱癌知识图谱绘制

知识图谱（Knowledge Graph）不仅仅是知识可视化的重要手段，而且能够更加有效地管理知识，同时对外提供便捷访问的借口，满足更加海量及精准的知识获取需求（图 4）。

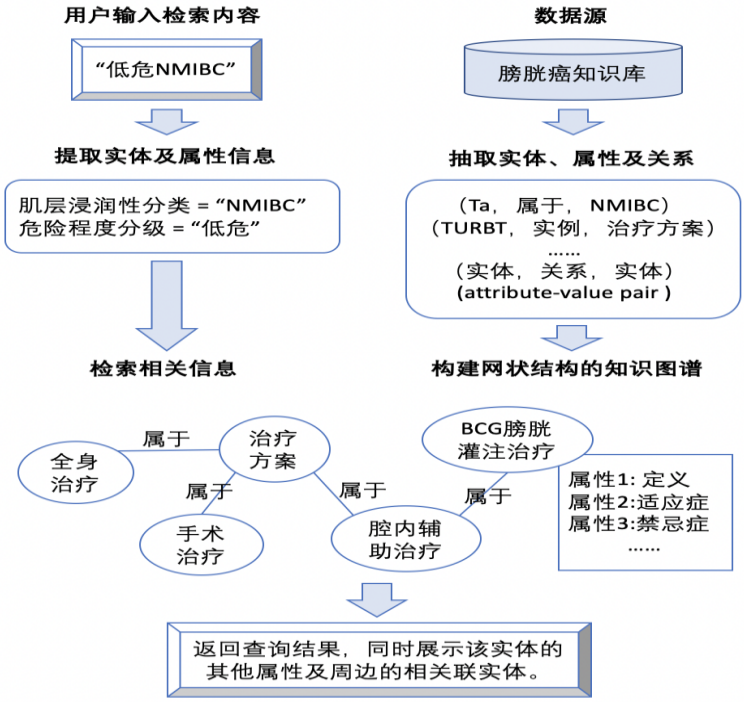


图 4 膀胱癌知识图谱工作流程

膀胱癌知识图谱中导航功能与知识查询功能的结合应用，使知识库检索结果的资源实例和其他相关类型实例形成关联，能够提供更加丰富的检索结果，完成知识的综合展现和层次梳理，方便用户理解类及类的层级关系，在临床应用中更有利于医疗工作者学习和掌握相关概念。

4 讨论与结论

4.1 知识应用

知识库建立后，可在多种场景得到广泛应用。

（1） 知识查询：由于本体表示方法的概念术语规范性、类层次结构清晰性和关系逻辑性，更有利于知识查询的准确性，同时术语间由规则进行了关联，方便检索时获取相应的其他知识，提高查询和检阅的效率，为临床及科研工作提供便捷。例如，检索某些治疗方式时，不仅能出现该疗法的具体内容及适应症和禁忌症，更能自动关联推荐相关的其他疗法和用药，提高诊疗质量。

（2） 决策推理：知识库构建完成后，进一步构成了临床决策支持系统（Clinical Decision Support System，CDSS）的核心，推理机从人机接口获取患者数据模型后，结合存储在知识库中的结构化知识对其进行比较和匹配，再将数据与知识共同传递给规则引擎，返回决策结果，获得适用于当前临床语境的决策。

（3） 知识管理：知识库作为一种成熟的数据管理系统，通过建立结构化决策知识点，集成化和体系化了多种信息资源，对知识的管理是其基本功能，而数据的标准化和平台的交互性为这一功能提供了保障：知识库提供的服务接口可与 HIS、CIS 等系统集合，方便数据等访问和交流。

（4） 知识共享：可重用性是以本体方式建立的知识库的巨大优势，对于研究已建立的膀胱癌知识库，仅需对其中的本体进行修改就可以直接用于新的知识库系统，从而避免组织知识的重复工作，提高经济效益与临床信息化速度。同时，各知识元之间的相对独立性和可拓展性也为后续内容的更新与优化奠定了基础，由于知识库表达的通用性和规范性，可以便捷地整合新知识、新领域，令开展后续拓展和维护工作的可行性大大提高。

4.2 知识评估

除了在建立知识库的过程中对模型质量进行评估，如概念准确性、逻辑一致性、架构合理性等，在知识库走向应用层面后，同样需要结合具体的使用场景对知识库的性能和效用做出度量，进一步优化其功能。

（1） 临床可用性：在医院泌尿科实际接诊膀胱癌患者的过程中应用知识库辅助诊疗决策，收集临床医生的反馈和建议，优化调整知识库架构及功能，以切实提高实际可用性为根本目的。

（2） 决策准确性：对于知识库返回的决策建议，同时记录临床医生做出的诊疗决定，计算决策的准确率，同时发现决策失误所属的类别及领域，进行针对性的语义和逻辑框架优化。

（3） 知识完整性：在实际应用中常遇到的诊断假阴性等情况，多是由于知识库中缺少相应症状体征或分类指标的编码所致，导致输入信息无法与知识库形成映射，从而影响了知识库的

检索及推理效能。因此，知识库中的各概念、属性、关系等元素需要不断完善，同时纳入更多的参考资料，已实现知识库的全面性和广泛适用性。

（4） 动态更新性：膀胱癌相关诊断标准、技术手段及治疗措施发展日新月异，一个实际可用的辅助诊疗知识库必然是具有动态性的，以便于纳入最新的诊疗指南，科研成果等知识资源，持续支撑临床工作的开展。

4.3 研究局限及展望

目前已建成的膀胱癌知识库尚存在许多局限，比如知识库的知识更新及丰富依赖于人工操作，效率较为低下，这也提示我们进一步研究医疗人工智能处理技术，使用自然语言处理、机器学习等方式自动抽取知识，建立并更新知识，以建成能够真正独立运转，履行医疗工作的智能化临床辅助工具，推进医疗行业的信息化发展。

为了更高效的获取、融合和推理知识，知识的表示学习方法是一种可行的解决方案。通过对知识库中的实体及其关系进行表示学习，语义信息被转化为稠密低维实值向量，即可在低维空间尺度实现实体和关系语义联系的高效计算，解决数据稀疏的问题。

技术的发展同样造成了用于医疗决策的数据种类和数据量的爆炸式增长，肿瘤诊疗领域也面临着极大的挑战。在知识库建成之后，可以结合机器学习模型实现知识的智能应用，利用随机森林、人工神经网络等机器学习算法，基于医疗大数据提取并分析临床相关变量之间的联系，学习并识别模式，从而对概念之间的未知关系进行预测，进一步提高知识库在临床决策中的适用领域范畴。

4.4 研究结论

本研究以辅助医疗工作者进行膀胱癌临床诊疗为目的，采用本体表示语言的方法，经历知识的采集、表达、应用与评估四个阶段，建立了膀胱癌临床诊疗决策支持知识库。知识库包含膀胱癌及泌尿专科范畴的概念、属性、约束、关系及实例，可用于初步诊断、治疗方案提示、医疗预警、知识学习等多个使用场景，减少医疗失误的发生，为膀胱癌这一常见泌尿系恶性疾病的诊治水平提高提供了保障。

PU-073

基于人工智能的肾结石评估模型研究

李星雨[1]、赖聪[2]、承垠林[1]、许可慰[2]、周毅[1]
1. 中山大学中山医学院
2. 中山大学孙逸仙纪念医院

目的 通过对肾结石患者的临床数据分析，评估肾结石患者的病程、复杂程度、并发症风险、治疗预后，实现对肾结石的智能临床决策支持。

方法 以预测经皮肾镜碎石取石术（PCNL）术后结石清除状态及并发症的发生风险为切入点，回顾性分析肾结石患者的临床资料，根据术后结石清除状态 S-ReSC 评分及术后并发症 Clavien-Dindo 分级对患者进行分组。将潜在影响因素纳入随机森林建立预测模型，依据变量重要性得分降维后，进一步使用多因素 Logistic 回归进行分析，将两种结局统一为同一模型并建立评分标准，并分析其临床意义。

结果 建立肾结石患者综合情况评估模型，输入患者初步诊断指标后，输出肾结石复杂性评分，预测术后结石清除状态及并发症发生风险。

结论 本研究以泌尿系疾病中高发的肾结石为切入点，基于医疗大数据及机器学习算法建立病情分级模型，为智能分级诊疗的实现提供了实验基础。

1 引言

尿路结石是最常见的泌尿外科疾病之一，其发病率在中国人群中可达 5.8%[1]。对于复杂上尿路结石的处理，经皮肾镜取石术(percutaneous nephrolithotomy, PCNL)是目前金标准方法，2019 年 EAU（European Association of Urology）指南推荐大负荷肾结石手术首选 PCNL[2]。对比其他疗法，PCNL 具有最高的术后结石清除率,但其并发症发生率也十分突出,可达 20%[3],包括感染、出血、尿瘘、胸腔积液等。

近年来，许多研究试图寻找影响 PCNL 术后结石清除率的因素，包括结石体积、结位置、肾脏解剖学结构等，目的在于评估肾结石的复杂程度，预测手术成功率及提供多中心比较的标准等，从而得到对病情的客观反映。例如，有研究指出结石负荷、肾造口扩张术式及感染性结石与手术预后显著相关[4]。然而，目前仍没有一种方法得到普遍认同，各评分标准仍然存在着各种问题，其中最为突出的一点就是，它们都无法准确对术后并发症的发生风险进行预测，因而缺乏一定的临床意义。

本研究基于医疗大数据及机器学习算法建立病情分级模型，以术后结石清除状态及并发症发生风险为预测指标，能够根据患者术前情况推断手术风险及预后，辅助临床决策，优化分级诊疗。实现对初诊患者的合理分流，诊治过程的标准化，以及后期随诊的规范化，实时动态监测患者疾病发生发展过程，真正做到以技术为基础的、数据导航的疾病全程管理过程。

2 现有肾结石评分模型的缺陷

目前常用的肾结石评分模型包括：Guy’s 分级、S.T.O.N.E.评分系统、CROES 列线图、S-ReSC 评分系统、SHA.LIN 协和评分等[5-9]。其中 Guy’s 分级应用简便；S.T.O.N.E.评分基于 CT 这种目前肾结石患者首选的影像学检查而建立；CROES 列线图由大型的多中心数据库统计而来，客观性更高；S-ReSC 评分仅取决于结石位置，为疾病复杂性分级提供了一种简单的方法；而 SHA.LIN 协和评分是基于我国病例数据建立的。

这些方法中，除 CROES 外，均基于文献和经验得出，缺乏基于大数据的系统性调研与统计学分析。其纳入标准严苛，许多复杂的解剖学因素未加考虑。Guy’s 和 CROES 评分除影像资料外，尚需要患者其他信息，使用不便，而且 CROES 还存在着计算复杂的问题。对于术后结石清除状态缺乏统一标准，各方法敏感性不同可能会造成偏差。

每种评分模型各有利弊，但大多对术后结石清除率的预测效能相当，目前仍没有一种评分体系得到广泛认可和引用。同时，尽管有文献认为在排除合并症患者减少混杂因素后，Guy’s 评分与术后并发症有一定相关性，但几乎所有方法都无法有效地预测并发症的发生。研究指出，这种单纯对于手术复杂性的评估很难真正影响临床决策，其意义更多在于对治疗结果的比较设立统一标准。而如果术前评分能够警示并发症的发生，那么医护人员就可以提前做好相应准备，以收获更好的临床疗效。

综上，理想化的评估模型应在具有较高的结石清除率预测能力的基础上，将术后并发症纳入考虑，同时兼具使用简便，适用范围广的优点。

3 基于机器学习的肾结石评估模型构建方法

课题总共分为三大步骤：数据采集、模型构建及评价验证。

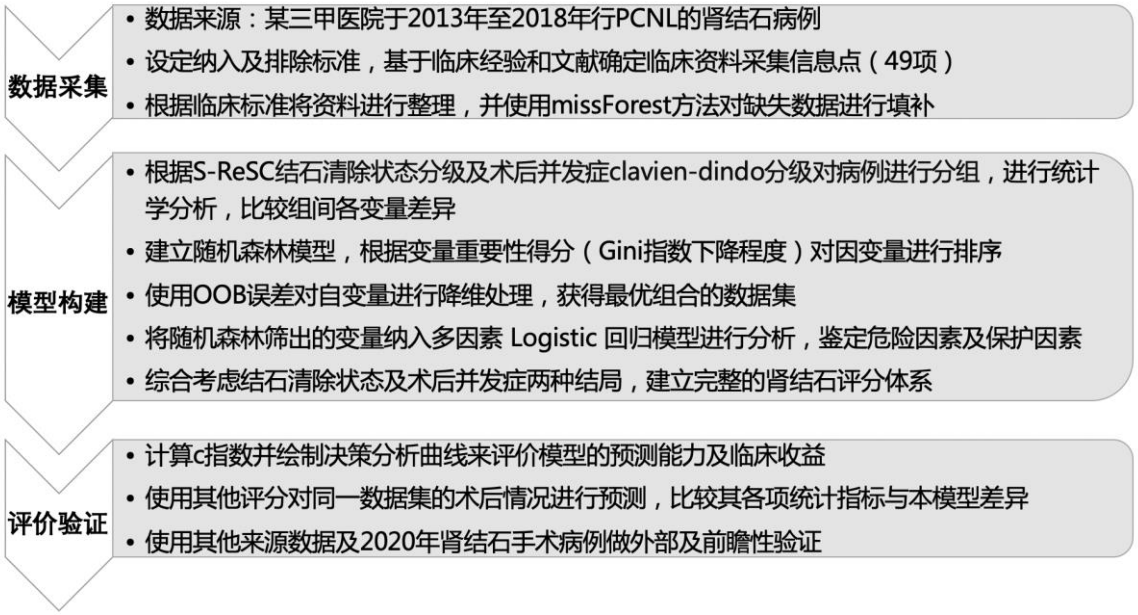


图 1 课题设计思路

3.1 数据采集

3.1.1 数据来源

某三甲医院于 2013 年至 2018 年行 PCNL 的肾结石病例，目前约 1500 例并将持续收集、更新，将数据从医院电子病历系统中导出，整理。

3.1.2 数据筛选

根据纳入及排除标准筛选病例资料。

图 2 纳入排除标准及临床资料采集指标

3.1.3 数据预处理

医学数据处理中常遇到数据缺失问题，主要解决方法包括直接删除法和数据插补法。其中直接删除法不仅造成数据的浪费，降低了样本信息，而且当删除的样本是研究人群中的一个选择性子样本时，会产生偏差估计。而数据插补法考虑利用相关其他信息，借助统计学建模方法，为缺失值插补一个合理的预测值，使得数据集完整。

missForest 算法核心思维在于利用已知的变量数据当作特征，将缺失值的变量当作标签。其中在标签中有数值的数据为训练集，缺失数据标签部分为测试集，然后通过随机森林预测去更新缺失值，相较于其它方法，它可以更好地避免插值对原始数据分布造成的偏倚[10]。

根据临床标准将资料进行整理，形成统计表格，使用 missForest 方法对缺失数据进行填补。

3.2 模型构建

3.2.1 统计学分析

应用 SQL Server 建立数据库，根据术后结石清除状态 S-ReSC 评分[10]及术后并发症 Clavien-Dindo 分级[12]将病人分组，应用 SPSS 25.0 软件对数据进行统计学分析，比较组间变量差异。连续性变量若符合正态分布，则用$\bar{x}\pm s$表示，采用方差分析；连续性变量若不符合正态分布，则采用中位数及范围区间表示，应用 Kruskal-Wallis 秩和检验进行数据分析；组间率的比较采用 R×C 列联表卡方检验，$P<0.05$ 为差异有统计学意义。

3.2.2 随机森林模型构建

使用 R 3.6.2 软件进行分析，分别以术后结石清除状态 S-ReSC 评分及术后并发症 Clavien-Dindo 分级为因变量，以各临床指标作为自变量，建立随机森林模型。

a) 根据 bootstrapping 方法，对总样本采用放回随机抽样法，每次从原始数据中抽取 2/3 的数据生成一个自主样本集作为训练集，未被抽到的 1/3 样本组成袋外（Out-of-Bag, OOB）数据作为测试集，用于估计分类误差和计算变量重要性评分。重复操作上述步骤 n_tree 次，共得到 n_tree 个包含 2700 个样本(训练集样本量为 1800，测试集样本量为 900)的数据集，并分别训练决策树模型。

b) 对于单个决策树模型，每次分裂时根据信息增益（Gini 指数）选择最好的特征进行分裂，过程中不需要剪枝。最终将生成的多颗决策树组成随机森林，按照多棵树分类器投票决定最终分类结果。

c) 进行特征重要性评估，将 46 个自变量按相对重要程度进行排序。重要程度以平均基尼(Gini)指数下降程度表示，即计算每个变量对分类树每个节点上观测值的异质性的影响，从而比较变量对于整个模型的重要性，该值越大表示该变量的重要性越大。

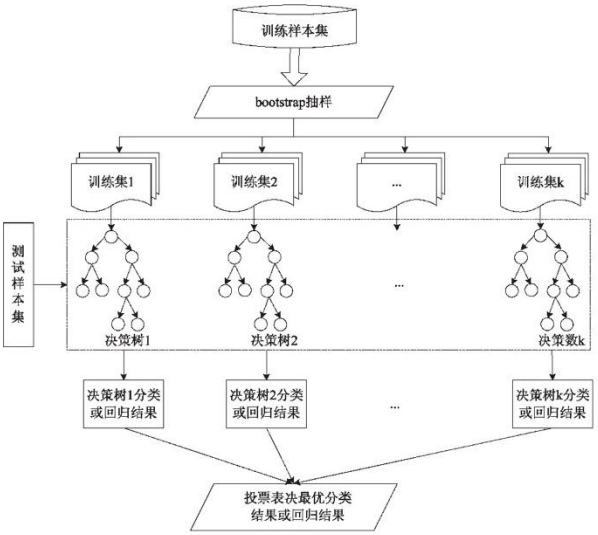


图 3 随机森林建模原理示意图

3.2.3 变量降维

以纳入模型变量数为横坐标，OOB 误差为纵坐标绘制错误率图。寻找图中变量数相对较少、错误率相对较低时对应的变量组合作为降维后的变量集。

3.2.4 模型调参

优化随机森林中的两个超参数：n_tree 与 m_KValue。默认树数目 100，默认随机特征数量为 log2(n-1)+1。为了确定最优参数进行多次实验，将树的数量分别设置为 100,200,300,400,500,600,700,800，观察平均误差结果的变化，选择最小值对应的数值作为 n_tree。在树的数量取到最优的情况下，随机特征的数量设置为 5,6,7,8,9,10,11,12，再次取最优值作为 m_KValue。

3.2.4 Logistic 回归分析

将随机森林筛选出来的因素纳入多因素 Logistic 回归分析，计算回归系数及 p 值，鉴定相关因素为风险因素或保护因素，分别建立肾结石 PCNL 术后结石清除率及术后并发症风险评分体系。

3.2.6 模型整合

观察两种评分间参数特征及联系，整合为统一模型，并使用原数据集对其预测能力进行检验。同时将术后结石清除状态及术后并发症分级同时作为结局输入随机森林，建立对两种因变量共同的决策模型，进一步对三者进行对比分析。

3.3 评价验证

3.3.1 模型评价

对于建立的模型，计算其一致性指数（Index of Concordance, C-index）以反映模型的预测能力、敏感性与特异性。绘制决策分析曲线（Decision Curve Analysis, DCA）以判断该评分体系的临床收益大小。

3.3.2 横向比较

使用 Guy’s 分级、S.T.O.N.E.评分系统、CROES 列线图对同一数据集的术前结石状况进行打分，预测术后结石清除状态及术后并发症情况，进行 Logistic 回归分析并计算一致性指数，与本研究结果进行对比分析。

3.3.3 外部验证

使用其他来源的数据输入模型，进行外部验证，计算相应准确率、特异性与敏感性，评价模型的可重复性。

3.3.4 前瞻性验证

收集 2020 年于该院行 PCNL 术的肾结石病例数据，对该模型的预测能力进行前瞻性验证。

3.3.5 临床意义分析

分析评分模型各参数临床意义，以对临床预防和诊治及下一步研究产生提示作用。

4 研究结果与讨论

泌尿系结石在全球范围都是泌尿系统的常见病，给医疗体系带来沉重负担。结石的高发病率及严重危害给我国医疗体系及临床医师提出巨大挑战，现有的外科手段在结石治疗和预防中仍力有不逮。近年来药物治疗、影像学检查、腔内技术和碎石设备技术的创新和改进，推动了先进治疗手段的应用和临床经验的积累，并催生了一系列设备改进和技术创新。目前在肾结石的综合性防治方面，急需借助类似人工智能、互联网等新兴信息技术，将患者的全程管理纳入研究重点，通过医工研究的融合，给患者提供最佳的诊疗方案及满意的医疗服务。

目前已有的肾结石评分模型(如 Guy’s 分级、S.T.O.N.E.评分系统、CROES 列线图、S-ReSC 评分系统、SHA.LIN 协和评分等)的建立主要基于 Logistic 回归分析。Logistic 回归分析属于线性概率模型中的一种，其要求研究数据缺失值少，且需符合线性、正态分布，满足方差齐等条件，

但临床数据易存在缺失值，影响 PCNL 效果的结石及肾脏解剖因素较多，因素之间的作用方式复杂，因此，基于 Logistic 回归分析建立的预测模型在临床应用中存在较大局限性。

比较而言，许多机器学习算法对数据分布无线性要求，自动发现并利用相关因素之间的交互效应及非线性关系;可以充分利用有缺失的数据，无须事前对数据进行填补，更加真实地反映数据的原貌;通过大量随机选择样本的方法平衡了样本误差的影响，由此产生的大量数据训练出的模型，较仅以单个测试样本进行拟合的 Logistic 回归模型的结果更为可靠;能够处理大样本数据，并在分析处理数据的过程中不断学习和记忆，不断提高其预测能力。理论和实践均证明，随机森林算法是对大多数分类及回归问题都有效的通用模型，具有对异常值、缺失值和噪声容忍度高，不易出现过拟合等优点，在预测分类问题上具有很高的应用价值。

本研究以泌尿系疾病中高发的肾结石为切入点，基于医疗大数据及机器学习算法建立病情分级模型，以术后结石清除状态及并发症发生风险为预测指标，能够根据患者术前情况推断手术风险及预后，辅助临床决策，优化分级诊疗。实现对初诊患者的合理分流，诊治过程的标准化，以及后期随诊的规范化，实时动态监测患者疾病发生发展过程，真正做到以技术为基础的、数据导航的疾病全程管理过程。

PU-074

基于电子病历的膀胱癌术后复发风险预测综述

毛钤镶、刘迷迷、周毅
中山大学中山医学院

目的 膀胱癌在我国男性肿瘤中排列在比较靠前的位置，当前膀胱癌在泌尿系恶性肿瘤中是最常见的疾病，且术后复发率很高，预测膀胱癌术后复发风险具有重要临床意义。

方法 国内外关于膀胱癌术后复发风险因素的分析及风险预测，大多数是利用统计软件进行传统统计分析，致力于发现各种膀胱癌分子标志物作为预测指标。随着人工智能技术的发展，机器学习方法以及深度学习方法的使用将比传统统计分析方法具有更好的准确度和泛化能力。

结果 获取电子病历数据对身体没有创伤，相对廉价，从电子病历中抽取患者的信息构造特征，运用机器学习的方法做疾病预测是当前热门的研究方向，在糖尿病以及心血管等疾病的预测上已有显著成果，然而，关于膀胱癌术后复发风险预测的研究比较有限。

结论 基于电子病历的膀胱癌术后复发风险预测将是一个具有挑战性和重要现实意义的课题。

1、引言

膀胱癌是泌尿生殖系统中最常见的恶性肿瘤，欧美国家膀胱癌发病率排在恶性肿瘤前列。随着化工产业的发展、环境污染加重、吸烟人群增多，膀胱癌在我国的发病率逐年升高。根据资料

显示，2008 年我国肿瘤登记地区膀胱癌的发病率为 7.49/10 万人，死亡率为 2.60/10 万人[1]，2015 年中国的膀胱癌新发病例数达 80.5/10 万人，死亡率达 32.9/10 万人[2]。六年时间里，膀胱癌发病率及死亡率急剧增长。根据膀胱肿瘤的浸润深度，分为非肌层浸润性膀胱癌（non-muscle invasive bladder cancer, NMIBC）和肌层浸润性膀胱癌（muscle invasive bladder cancer, MIBC），新发病人中大部分是 NMIBC。NMIBC 的标准治疗方式为经尿道膀胱肿瘤切除术（Transurethral resection of bladder tumor, TUR-Bt）联合膀胱灌注化疗，MIBC 的治疗则主要以根治性膀胱全切除术（Radical cystectomy, RC）为主。由于尿路上皮癌的本身的易复发特性以及 TUR-Bt 手术方法的特殊性，研究表明，膀胱癌患者在术后 5 年内总复发率高达 60%~70%，此外有近 16%~25%的患者出现了肿瘤的恶性升级，可能需要接受全膀胱切除，威胁患者的生存[3]。因此预测膀胱癌术后复发风险，以便对患者实行早期干预治疗具有重要临床义意。

2、膀胱癌术后复发预测模型

近年来，国内外学者在有关膀胱癌术后复发风险因素分析以及膀胱癌术后复发预测模型建立方面做了比较充分的研究。王超[4]等用 Log-rank 检验比较生存曲线，发现上尿路尿路上皮癌(UUT-UC)患者的肿瘤病灶数以及术前是否行输尿管镜检术可以作为判断根治性肾输尿管切除术(RNU)后膀胱肿瘤复发的危险因素。邓助朋[5]等研究了非肌层浸润性膀胱癌（NMIBC）术后肿瘤复发的相关因素，发现肿瘤大小、肿瘤数目、肿瘤分级、肿瘤分期、是否初发为影响非肌层浸润性膀胱癌术后肿瘤复发的独立因素。钟广正[6]等鉴别和筛选出促进膀胱癌于细胞(bladder cancer stem cells, BCSC)自我更新的长链非编码 RNA(long non-coding RNAs, lncRNAs)，发现 lncRNA-BCSC 高表达与 TURBT 术后膀胱癌复发率呈正相关,有望成为预测膀胱癌复发与精准治疗的分子标志物。林英立[7]等研究发现血清中 CDH13 基因启动子甲基化是非肌层浸润性膀胱癌行 TURBT 治疗后复发的预测指标。Bassi P F[8]等比较了人工神经网络(ANN)和标准逻辑回归(LR)在膀胱癌根治性膀胱切除术患者中的预后性能。Abbod M F[9]等采用人工智能方法和传统统计方法，利用膀胱癌患者的实验分子标志物，包括 p53 和错配修复蛋白，以及常规临床病理数据构建膀胱癌复发预测模型。

通过对国内外研究现状的分析，发现既往关于膀胱癌术后复发因素的研究，大多数是通过选取常规变量包括性别、年龄、体重、吸烟史等，利用软件进行统计分析，而医学研究者比较倾向于研究膀胱癌新型基因和生物标记物的发现，对于分析方法上并没有太多改进。随着人工智能的发展，机器学习方法在医疗数据挖掘方面相比于传统统计分析方法拥有更好的性能表现，在处理复杂的问题上有较好的准确度和泛化能力。

3、基于机器学习方法的疾病预测模型

将计算机技术应用于医疗领域是当前研究的热点，机器学习在疾病预测、疾病辅助诊断、疾病预后评估、新药研发、健康管理、医学图像识别等领域的都有重要应用价值，对于提升整个医疗行业的发展有重大义意[10]。其中利用机器学习以及深度学习方法进行疾病预测已有许多研究。杨洋[11]基于高血压流行病学调查数据，构建了 BP 人工神经网络模型用于高血压预测。马宗帅[12]结合心脑血管疾病数据的特点，构建了基于深度信念网络（DBN）的心脑血管疾病预测模型。洪烨[13]对导致糖尿病的危险因素进行分析，并采用 BP 神经网络模型、支持向量机模型和集成学习模型构建了糖尿病预测模型。苗立志[14]采用大数据分析技术并基于随机森林模型，应用多个弱分类器将多个决策树获得的结果进行集成，得到乳腺癌风险预测，模型预测精度可达 99.04%。Kusiak A[15]等基于决策树模型和粗糙集理论探讨了影响晚期肾病患者的最主要因素。Jin P[16]将糖尿病危险因素的时间敏感性特征作为预测因子构建了基于神经网络的糖尿病发病风险评估模型。

然而，用机器学习以及深度学习方法取代传统统计分析方法建立膀胱癌复发预测模型的研究比较有限，值得进一步研究。

4、基于电子病历的疾病预测模型

电子病历(Electronic Medical Records，EMRs)记录了病人的相关就诊信息，包括首次病程记录、入院记录、诊疗经过、出院小结等，是提供临床决策支持以及开展医学研究的宝贵资料。通过 EMR 可快捷地获取医疗资料，对其中数据进行挖掘与分析，可以及时有效地监测病人的身体状况，提供临床决策支持，对特定事件的发生进行监测，对疾病的早期预警起到重要作用。对大规模 EMR 的分析，有助于发现致病因素以及不同疾病间的关联，从源头上预防疾病。电子病历包含信息丰富，且相对廉价易得，目前已经有许多利用电子病历做疾病预测的尝试。

Bandyopadhyay S[17]等基于电子病历数据提出了一种贝叶斯网络模型，能够解决电子病历中存在缺失值的问题，根据他的研究结果，该模型在预测心血管事件的风险方面有着良好的性能。Eapen Z J[18]等依据电子病历中选取变量构建的模型，可对患有心力衰竭患者 30 天内的病情结果进行风险分层，并对病人入院后 30 天内的死亡率、出院后 30 天内的再入院率和出院后 30 天的死亡率进行了预测分析，为临床决策提供一个有效的工具。李季[19]等以电子病历大数据为基础，采用朴素贝叶斯理论、信息增益法等数据挖掘算法，构建了适用于多种常见严重疾病的动态危重度预测模型，并验证了模型的有效性。Nguyen[20]等提出将患者的电子病历数据表示成一串按时间先后顺序排列的序列，利用卷积神经网络（Convolution Neural Network, CNN）对其进行患者的表征提取。Ma[21]等利用双向循环神经网络（Bidirectional Recurrent Neural Network, Bi-RNN）进行电子病历的表征学习，改善了卷积神经网络只能捕获局部特征信息的缺陷，更适用于时序学

习相关任务，并利用多种关注机制（Attention Mechanism）方法来提升模型的表征学习能力和可解释性。对于进一步提升模型的预测性能，依然有大量学者在进行着不懈的探索。

5、总结与展望

获取电子病历数据对身体没有创伤，相比于基因测序等成本很低，相对廉价。从电子病历中抽取患者的信息构造特征，再运用机器学习以及深度学习的方法对患者电子病历中的风险因素进行建模，从而实现对疾病风险预测，是当前热门的研究方向。在糖尿病以及心血管等疾病的预测上已有显著成果，表明其可行性，然而，关于膀胱癌术后复发风险预测的研究却十分有限。

在基于电子病历的膀胱癌术后复发风险预测研究中，重要步骤之一就是电子病历信息的抽取，电子病历信息抽取是后续开展一系列医学研究的基础。由于电子病历中包含大量非结构化的自由文本，可能存在歧义或者包含许多不太规范的描述语言，因此要解决一系列的自然语言相关问题。通过对自由文本的分析，挖掘出其中的有用信息，通过有效的表征学习方法来构造膀胱癌患者的特征。电子病历包含丰富的如诊断、医嘱、检查检验、生命体征、人口学等患者信息，传统机器学习方法的特征工程通常需要大量的人工干预。而最近几年，随着深度学习在理论上的突破以及在众多领域的成功应用，利用深度学习来处理电子病历数据的表征学习也可以为此提供一种解决思路。总的来说，基于电子病历的膀胱癌术后复发风险预测研究将是一个具有挑战性和重要现实意义的课题。

PU-075

自然语言处理技术在临床上的应用

毛钤镶、刘迷迷、周毅
中山大学中山医学院

目的 了解自然语言处理技术在临床上的应用，提高医疗系统的运行质量，降低运行成本。

方法 对国内外关于自然语言处理在临床上的应用进行详细叙述。

结果 介绍了自然语言处理技术在电子病历信息提取、医疗知识挖掘、医疗决策支持、以及其他临床领域的应用

结论 对未来医学领域自然语言处理技术的发展进行了展望。

1、引言

自然语言处理（Natural Language Processing，NLP）利用计算机对人类语言文字进行处理，是研究人与计算机之间用自然语言进行有效通信的理论和方法。自然语言处理是计算机科学领域与人工智能领域中的一个重要方向，其目的是让计算机能够处理或“理解”自然语言，以执行文

本分类、信息检索和机器翻译等任务，用到的技术包括命名实体识别、词性标注、语义消歧、指代消解、结构分析等。自然语言处理可以分为两个流程：自然语言理解（Natural language Understanding , NLU）和自然语言生成（Natural language Generation , NLG）[1]。自然语言理解主要是使计算机理解自然语言的含义，提取出有用的信息。具体来说，它可以是使自然语言结构化，比如分词、词性标注、句法分析等；也可以是表征学习，字、词、句子的向量表示，构建文本表示的文本分类；还可以是信息提取，如信息检索（包括个性化搜索和语义搜索，文本匹配等），信息抽取（包括命名实体提取、关系抽取、事件抽取等）。自然语言生成与之相反，从知识库或逻辑形式等等机器表述系统去生成自然语言，根据提供的结构化的数据、文本、图表、音频、视频等，生成人类可以理解的文本。

临床医学信息大量的以非结构化或半结构化文本形式存储于信息系统中，通过自然语言处理，这些非结构化的医疗文本被转化为包含重要医学信息的结构化数据，科研人员才可以从这些结构化的数据中发现有用医学信息，从而提高医疗系统的运行质量，降低运行成本[2]。

早在 20 世纪 60 年代，美国就已经开始了临床医学领域 NLP 的研究，早期的研究基于有限的电子医疗文本验证了 NLP 在医疗领域的可行性。到 80 年代和 90 年代，大量的医学数据和领域知识库被逐渐建立起来，例如 SNOMED（The Systematized Nomenclature of Medicine），UMLS（The Unified Medical Language System）。随后，大量的临床医学 NLP 系统开始出现，代表性的临床医学 NLP 系统有 MedLEE[3]、cTAKES[4]、HITEx[5]等。这些临床医学 NLP 系统的应用覆盖了医学信息抽取、医学文本分类、医疗决策支持、病人信息管理、医疗信息问答、医学知识挖掘及知识库建立等诸多领域。下面将从电子病历信息提取、医疗知识挖掘、医疗决策支持等几个方面详细论述自然语言处理在临床上的应用。

2、电子病历信息提取

电子病历(electronic medical record，EMR)是在临床治疗过程中产生的，由医务人员撰写的描述患者医疗活动的记录，病历电子化使得大规模病历的自动分析成为可能[6]。临床信息如入院记录中记录的临床症状和体征、手术记录或放射学报告等，对患者的诊断、评估和治疗具有非常重要的意义。然而临床报告中存在大量的非结构化文本，阻碍了电子病历数据的二次利用，而 NLP 技术能有效地从非结构医疗文本中提取信息。前文提到的医学语言提取及编码系统 MedLEE 就是由哥伦比亚大学的 Carol Friedman 等人设计开发的英文 NLP 系统，它将文本形式的病历报告转换成编码数据以促进乳腺癌研究，有利于病人看护质量的提高[7]。由于中文两个语言单元之间缺乏空格作为自然分隔标记[8]，因此从中文电子病历中提取相关临床信息与英文有很大不同。

自然语言处理在医学影像领域表现出良好的应用前景。NLP 可以对各类影像的综合信息进行分析，将报告中的检查结果和建议等信息与全面的临床信息相互关联，对患者个体疾病处理提供帮助。使用 NLP 可高效率地分析大数量患者群体影像诊断信息，得到群体的特征性数据，构建患者

队列，从而提高流行病学研究效率，为循证影像医学研究提供帮助。NLP 还可以识别医学影像学的流程和质量指标，判断影像报告是否符合相关指南或诊断规则，评价报告的完整性和规范。

利用 NLP 进行电子病历信息提取，在肿瘤学研究中也取得了一定的进展。这些研究可以分为三大类：病例识别、分期和临床结果[9]。NLP 系统通过分析医疗记录查找癌症术语并确定它们能否与患者进行准确匹配，在获得癌症术语和确定上下文信息之后进行癌症的最终分类。癌症分期及其分期所需参数信息以非结构化形式常规记录在 EMR 中。此外，不同的癌症类型存在不同的分期系统。NLP 技术可以从临床报告的相关部分识别出在一个或多个分期系统中使用的参数。明确诊断、给出分期对患者临床结果的研究非常有意义。针对肿瘤学临床结果的自然语言处理项目主要包括癌症复发的判别或者预测癌症的进展等方面。

3、医学知识挖掘

医学文本资料中的知识是不同国家、不同时代人们智慧的结晶，展现的是大量未整理的文献资料以及诊疗记录。大量医学文本资料中含有病史、诊断、治疗方法、药物等名词，利用 NLP 将隐藏在文本中的知识挖掘出来，对医学的发展具有重要意义。由于自然人的学习能力有限，因此学者们尝试通过自然语言处理辅助完成汇总医学知识的过程，将知识提炼出来提取其中有用的诊疗信息，最终形成知识本体或者知识网络，从而为后续的各种文本挖掘任务提供标准和便利，主要涉及命名实体识别、关系提取和文本分类等任务。

命名实体识别（Named entity recognition，NER）将词汇分组并将其和已知概念进行匹配。医学文本是一种包含许多首字母缩略词、缩写词和同义词的专业语言，因此医学领域的 NER 面临更多困难。命名实体识别方法大致分为三类：基于词典的方法、基于规则的方法和基于机器学习的方法。基于字典的方法简单且实用，但是新的命名实体不断出现，很多命名实体的长度较长甚至存在变体，难以建立一个完整的生物医学命名实体字典。利用启发式信息产生识别命名实体的规则可以灵活地定义和扩展，但是对领域知识的依赖性很强，需要该领域专家参与并花费大量时间。此外由于命名实体类型多样，新类型的命名实体还在不断涌现，很难建立一套一致的规则。基于机器学习的方法是目前主流的方法，它们利用统计方法从大量数据中估算相关参数和特征进而建立识别模型，机器学习模型的训练通常需要大量标注语料。与其他领域相比，受限于病人隐私保护的要求，大规模的人工标注语料在医学领域尤其缺乏。此外，基于深度学习的命名实体识别方法也有广泛研究。关系提取是一个用于确定实体之间的相关性的任务。比如“肿块”和“右叶”可以通过诸如“位于”之类的某种关系彼此相关。文本分类为一段文本分配类别标签。然后可以使用一组相关实体、关系和标签来填充模板，该模板以结构化数据的形式呈现。

医学知识挖掘或者知识发现对医学研究分析具有重要现实意义，是提高医学工作者知识水平的关键性技术。例如利用 NLP 技术挖掘出药品说明书中所有的医疗关系和实体。自动构建基于药品说明书的合理用药知识图谱，可提高合理用药的自动化程度和准确度，降低不合理用药[10]。

4、临床决策支持

临床决策支持系统（Clinical Decision Support System, CDSS）用于对医务人员进行诊疗方面的指导，降低出错的概率，提高工作效率。CDSS 是提升医疗质量、控制医疗费用的重要手段，它可以帮助临床医生深入分析病历资料，从而做出最为恰当的诊疗决策。自然语言处理已用于临床决策支持的软件中，如自动检测药物的不良反应。

词库是自然语言处理的基础，医疗决策支持系统的建立首先应建立词库。使用医学专业词汇、频率极高的谓词、量词等词汇、医疗文书词汇的常用组合及常用语句等，加上基本的语法库，形成用于医学语言处理的知识库。接着进行语言处理。按照中文自然语言处理的一般步骤，进行分句、分词、语义分析、形成文本摘要。

语义分析、文本摘要指根据汉语基本语法，对词汇进行重组，剔除意义不大的部分，形成摘要。临床决策支持系统以临床诊疗指南、操作规范为参考，在对医疗文本进行语言处理后进行推理、分析，找出其中存在的问题。分析模型是其中的关键，在分析模型中，比照的是临床诊疗指南、操作规范，所以在建立此知识库时，所用的词汇、短语应该与语言处理所用的知识库相对应，否则会增加建立分析模型的难度和复杂性。

5、其他临床领域的应用

自然语言处理技术在其他临床相关领域的应用也取得了成功的经验，例如医疗问答系统。大数据时代到来，而医学专业有其特殊性，面对网络上质量参差不齐的医学信息，非医学专业人员利用传统的信息检索查找、理解及获取医学信息存在着诸多困难和障碍。而基于自动问答的医学信息搜寻模式作为更智能的医学信息资源获取工具，不仅对海量数据资源的有效利用具有重大意义，而且在一定程度上可以缓解医患信息不对称，提高医疗资源利用效率。此外，自然语言处理也被用于自动创建放射学报告建议和多发性硬化症性状。通过创建临床摘要和从在线文章中搜索相关信息来改进医疗记录信息的呈现方式。

6、总结与展望

自然语言处理技术具有广阔的应用领域和前景。未来自然语言处理的发展趋势将是与许多领域深度结合，专业化的服务趋势会越来越强，从而为各相关行业创造价值。医学领域的 NLP 近年来取得了巨大进展，虽然在临床的广泛应用仍然存在一些挑战，如歧义与非标准化问题、患者隐私问题、标注成本过高等，各种电子病历自然语言处理测评的发起与组织也将吸引越来越多的不同领域的工作者参与到临床数据自然语言处理当中[11]。

目前我们正处于医疗大数据时代，海量的以自然语言表示的临床数据为 NLP 提供了更加宽广的应用舞台。医学领域对自然语言处理的需要也在日益提高，随着医学信息化水平的逐步深入，利用 NLP 技术实现医疗数据的有效使用对精准医疗、疾病防控、研发新药、医疗费用控制、攻克顽疾、健康管理等工作都有着重要的意义。

PU-076

基于 EEG 脑网络的癫痫发作状态转变研究

韦晓燕[1]、马梦楠[1]、陈子怡[2]、周毅[1]
1. 中山大学中山医学院
2. 中山大学附属第一医院神经内科

目的 目前还没有足够的定量计算方法来描述脑电信号在脑网络中的传播路径。本研究的目的是探索一种新的癫痫大脑网络方法。

方法 该研究旨在通过映射脑电非线性特征的同步来构建动态脑网络。

结果 分析了复杂网络的特点，描述了癫痫放电的短期和长期传播。

结论 本研究丰富了动态脑网络同步关系领域的空白，有助于癫痫的定位。

1. 前言

过去几年表明工作，网络连接癫痫中心-脑网络概念**错误!未找到引用源。**，即更复杂的“癫痫网络”概念已经取代了经典的、简单的单一概念癫痫。在“癫痫网络”中，兴奋性增强（或抑制性降低）的“节点”的同步活动参与了病理性癫痫的发生。

了解大脑不同脑区，不同节点相互关系可能有助于更全面地了解正在进行的大脑信息交互情。导联作为基本节点记录大脑的各个部位信息，同时向其他大脑区域/节点传递并影响其他大脑区域/节点的信息。关于这种同步影响程度的计算称为同步性测量，包括线性和非线性方法。线性方法即是从时域频域线性方法计算双变量直接的相关性**错误!未找到引用源。**。非线性方法是通过上文提到的非线性方法如信息转移，非线性指标计算不同变量之间的非线性依赖关系**错误!未找到引用源。**。由于节点传播的影响会因为距离而有所不同，如果在同侧脑区节点传播快，异侧脑区节点传播慢，这就有了网络节点的同步延迟问题。然而，很少有研究考虑到这一潜在问题探索同步脑网络。

在同步指标的基础上，图论技术构建功能网络的拓扑结构**错误!未找到引用源。**。大量的实验表明其网络拓扑结构随着发作的不断推进而变化，无规律中带有规律性**错误!未找到引用源。**，将同步可能性应用于 7 例患者发作期 IEEG 记录显示网络朝着更有序的配置方向移动**错误!未找到引用源。**：如发作过程中具有较高的聚类系数，同时路径长度变短。而这在发作间期期间网络是随机和无序的。Hao **错误!未找到引用源。**表明，发作期的信号聚集程度显著高于发作间，但是信息传播的路径长度并无明显变化。采用互信息同步核心节点的方式定位致痫灶，事实证明高的互信息值和活跃的癫痫病灶区域有关**错误!未找到引用源。**。但是大多数研究都只是选择两个时期进行研究，没有涉及到更多的癫痫不同时期转变状态研究。

本研究选择代表性颞叶和额叶癫痫患者的 EEG 数据，基于非线性系统相空间重构和互信息指标计算基础上，使用复杂网络技术构建癫痫 EEG 的动态脑网络功能模型。同时分析了复杂网络的特

点，描述了癫痫放电在不同时期的指标变化和传播路径，从空间上监测发作变化过程。

2. 实验数据

为了保证脑网络模型的构建，数据的选取标准首先是局灶性新皮质癫痫患者。其次有明显发作行为和脑电信号；每一个病人的长时程脑电图必需包括最少 2 次或更多的发作。基于上述标准，我们选择了两位癫痫患者数据，分别是颞叶癫痫患者（Temporal Lobe Epilepsy，TLE）和额叶癫痫患者（Frontal Lobe Epilepsy，FLE）构建脑网络模型以便于对比。设置 24 小时视频 EEG 系统，采样频率为 500hz。然后将信号从 0.3-75Hz 进行滤波。每次的发作选择包括四个时期的发作全过程，每次发作分为 2～6 个节段。将不同大小的时间窗分割成相同的数据片段计算同步度。这里需要添加详细数据信息。

表 6-1 实验数据信息表

患者	性别	年龄	导联数目	意识状态	监测时长	发作次数	发作时长
14	女	28 岁	22 导	清醒→睡眠	24 小时	5	824s
15	男	39 岁	22 导	清醒→睡眠	24 小时	4	895s

2 同步性计算

互信息是计算同步性指标的方法，其数学原理如下**错误!未找到引用源。**：设定收集到的 EEG 信号$X_{k,i}$为 M 个通道（k=1,…..M）,N 个时间点（i=1,…..N）,对于每个导联的信号进行延迟后转化为嵌入矢量

$$X_{k,i} = (x_{k,i}\,,\ x_{k,i+l}\,, x_{k,i+2l}\,, \ldots\ldots x_{k,i+(m-1)l})$$

其中 l 是延迟量，m 是维数。该向量集可表示为

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ & & \ddots & \vdots \\ x_{n-m+1} & x_{n-m+2} & \cdots & x_n \end{bmatrix} \qquad (2\text{-}1)$$

这些向量在相空间中都有必然的几率散布，由香农熵公式可分别得导联 a 和导联 b 的信息熵

$$H(S) = -\sum_i p_s(s_i)\log_2 p_s(s_i) \qquad (2\text{-}2)$$

$$H(Q) = -\sum_j p_q(q_j)\log_2 p_q(q_j) \qquad (2\text{-}3)$$

再由联合香农熵公式得两导联之间的联合熵：

$$H(S,Q) = -\sum_{j,\ i} p_{qs}(q_j, s_i)\log_2 p_{qs}(q_j, s_i) \qquad (2\text{-}4)$$

然后将前面三个公式带入香农公式即得到两通道之间的信息量

$$MI(S,Q) = H(S) + H(Q) - H(S,Q) \qquad (2\text{-}5)$$

3 脑网络分析

3.1 网络建设规则

两个节点是否存在边取决于通道之间的同步大小，当其互信息量大于阈值时存在连接边，反之不建立连接边，即可得到 0-1 二值无权网络。肯定阈值的方式目前有几种方法如下**错误!未找到引用源。**选择：一是选择能够使得网络较为聚集无孤立点的数值作为阈值，二是采用医学统计学方法确定置信区间确定阈值，三是选择统计指标如平均水平，变异系数等。本文采用最后一种，选择全部通道的均匀互信息作为阈值完成网络模型的构建。

3.2 复杂网络测量指标

在构建成功脑网络之后，我们对脑网络的特征指标进行探索，主要有：

（1） 聚类系数

$$C_i = \frac{2e_i}{k_i(k_i-1)} \qquad (2\text{-}6)$$

此中e_i为节点 i 邻人间现实毗连边的数量，分母是最大毗连边数量。总体的平均值作为统计指标，其衡量网络模块化的程度，平均聚类系数越大，说明其模块化程度越高。

$$C=\frac{1}{N}\sum_{i=1}^{n} C_i \qquad (2\text{-}7)$$

（2） 平均路径长度

描写全局毗连，藐视全局两两节点之间平均间隔大小。若不考虑节点与自身距离：

$$L=\frac{1}{N\ (N-1)}\sum_{j,i!=j} d_{ij} \qquad (2\text{-}8)$$

N 代表收集节点总数，暗示收集中节点 I,j 之间的间隔。均匀路径越短，图的布局越紧凑，连通性越好。

（3）度及度分布

选择度分布作为衡量网络同步性的指标，节点 i 的度k_i可定义为它相连接的边的数目，。如果链接到节点 i 的节点数为 k（i），则其阶数为 k（i）。$k_i = \sum_{jN} a_{ij}$

$$\text{度分布表示 P(k)}=\frac{\textit{度等于k的节点数}}{\textit{节点总数}} \qquad (2\text{-}9)$$

3.3 统计分析

如果数据具有正态分布和齐次方差，则使用方差分析（ANOVA）模型进行统计，以检验同步的不同特征值；否则采用非参数检验。分析采用 SPSS 软件（18 版）进行，设置检验水准为 0.05。

4 实验结果与分析

4.1 同步性指标下不同时期的比较

表 5-2.不同时期的互信息指标特征分析

EEG	平均值		变异系数	
	颞叶癫痫	额叶癫痫	颞叶癫痫	额叶癫痫
清醒期	0.926±0.132	1.007±0.066	7.549±2.619	7.960±1.647
睡眠期	1.062±0.062	1.000±0.143	9.201±2.273	9.472±2.108
发作期	1.072±0.119	1.071±0.176	8.431±2.206	10.283±1.744
发作间期	1.062±0.062	1.010±0.143	10.201±2.273	9.472±2.108
发作前期	1.032±0.119	1.021±0.176	8.431±2.206	10.283±1.744

对于额叶癫痫患者，在表 5-2 的互信息指标显示发作期 EEG 同步性最高，其次是发作前期，睡眠期，发作间期和清醒期其互信息指标较为统一。从变异系数指标发现，发作期同步分布差异最不平衡，清醒期最一致。不同时期的互信息指标统计检验，发作期，发作前期和发作间期的检验统计值均小于 0.5，表明互信息熵值可以很好得区分发作间期，发作前期和发作期主要指标。

表 5-3.不同时期的互信息指标统计检验

EEG	平均值		变异系数	
	额叶癫痫	颞叶癫痫	额叶癫痫	额叶癫痫
睡眠期 vs 清醒期	P<0.05*	P>0.05	P<0.05*	P<0.05*
睡眠期 vs 发作前期	P<0.05*	P<0.05*	P>0.05	P<0.05*
发作前期 vs 发作期	P<0.05*	P<0.05*	P<0.05*	P<0.05*
发作间期 vs 发作期	P<0.05*	P<0.05*	P>0.05	P<0.05*
发作前期 vs 发作间期	P>0.05	P<0.05*	P<0.05*	P<0.05*
K-W test	P<0.05*	P<0.05*	——	P<0.05*

4.2 脑网络放电变化途径

脑网络只展现 20 对最大互信息的通道。如图所示。MI 网络的核心节点指示了异常信道的位置。类似于基于互信息的脑功能网络构建，可以方便地构建四个指标的网络拓扑结构。其含义仍不清楚。

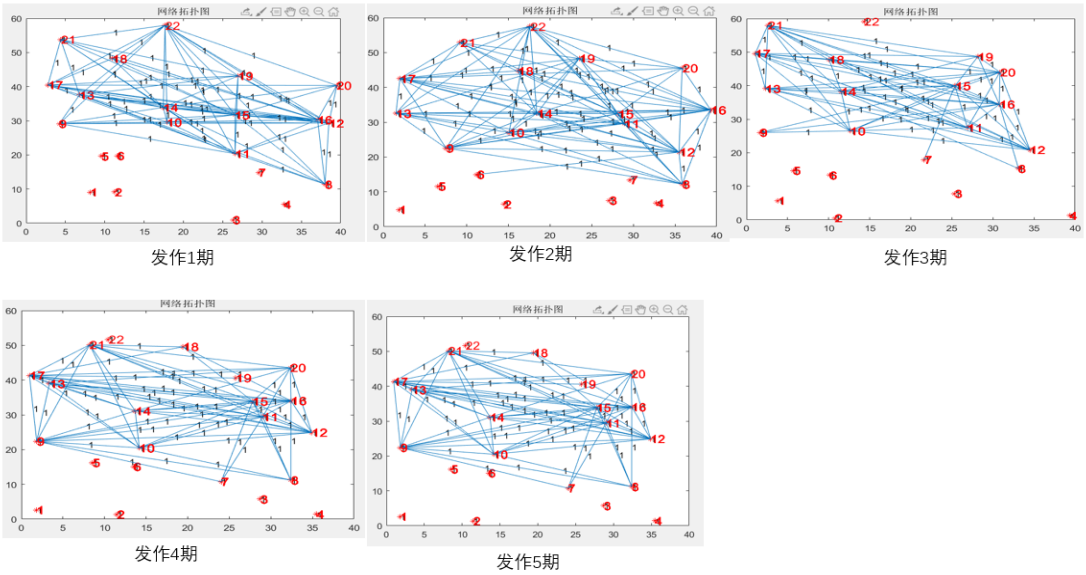


图 4-1：基于互信息熵的颞叶癫痫的大脑网络

如图 4-1 所示，每幅图显示的是脑电波片段的连接持续 12 秒。第二行的 5 个数字显示了癫痫发作的 5 个连续阶段。粗体点代表核心节点通道。癫痫发作期的棘波和慢波复合活动主要以右前段（F4 和 Fp2）为主，偶尔以左前段（F3 和 Fp1）为主。然后部分癫痫发作转移到前部主导的慢活动，涉及太阳穴（F8）和枕部（O2）。

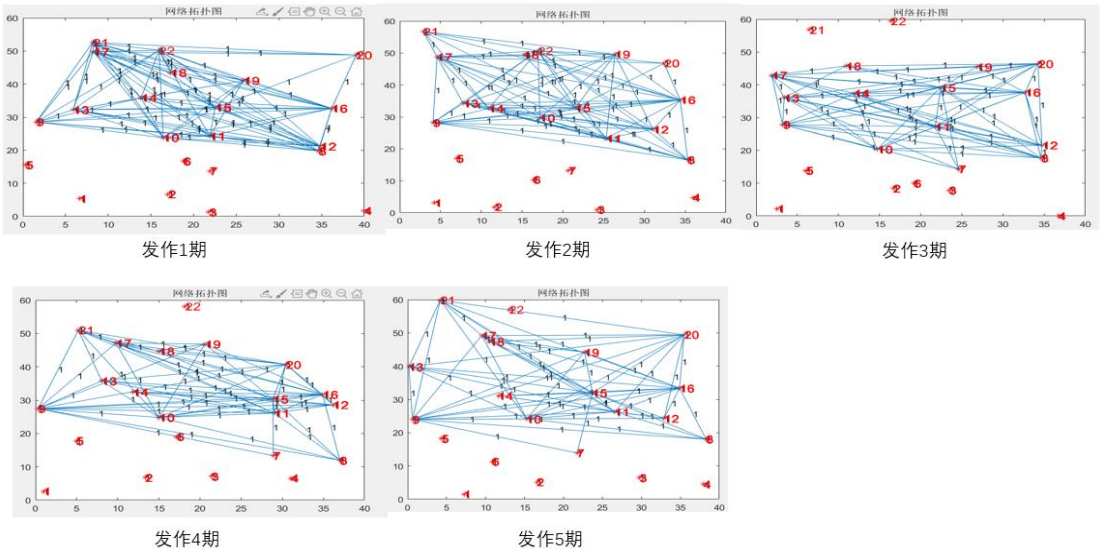
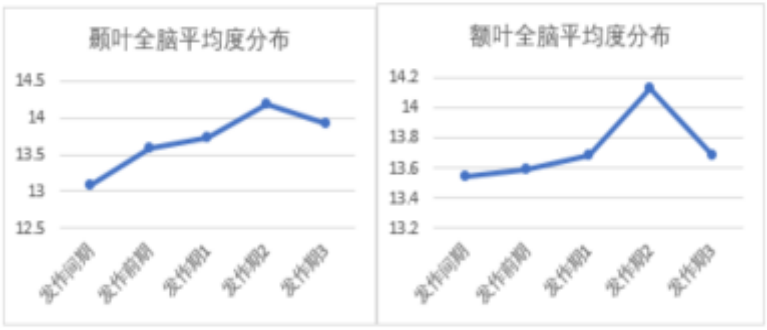


图 4-2：基于互信息熵的脑电网络，额叶癫痫

每幅图显示了 6s 的脑电部分的连接。第二行的 6 幅图表现了癫痫爆发的 5 个持续阶段。粗体点表示核心节点通道。对于颞叶癫痫患者，EEG 临床诊断显示癫痫发作期从左太阳穴开始，通过左中太阳穴和后太阳穴（T3 和 T5）可看到连续的棘波。在发作过程中，网络首先在左后脑区域被激活，然后在整个大脑中对称地活跃，最后在右前脑区域保持活跃。

4.3 脑网络指标分布



4-3. 基于互信息熵的脑网络模型度分布

当发作间期转变到发作期的过程中，整个神经网络的度数分布呈初始下降趋势。在癫痫发作过程中，脑网络的低度分布表现为轻微的上升，然后又下降。癫痫发作期度数先略有上升后下降。当癫痫发作时，大脑的信息流保持不活跃，随着癫痫发作的继续，信息交互作用增强。

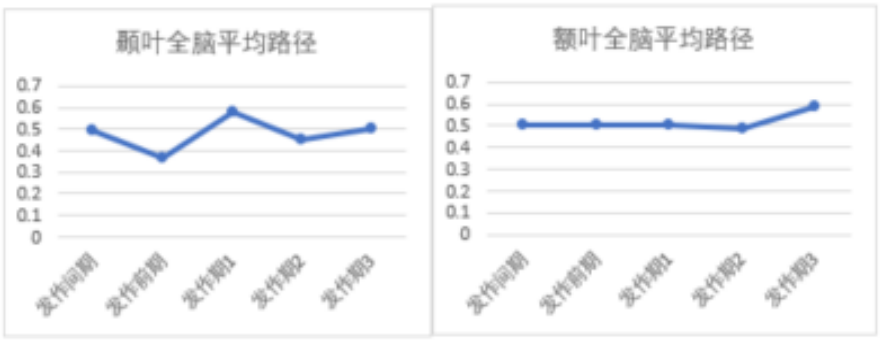


图 4-4. 基于互信息熵的脑网络模型平均路径分布

网络的平均路径衡量了信息传播的速度，从图 4-4 看出，在发作间期其信息传播路径较高，在发作前期时候由于发作即将到达，超量异常放电，传播的平均路径变短，到发作期平均路径达到最小，由于其大脑放电的传播性。到发作后期发作逐渐结束，大脑恢复正常放电，其平均路径逐渐升高。

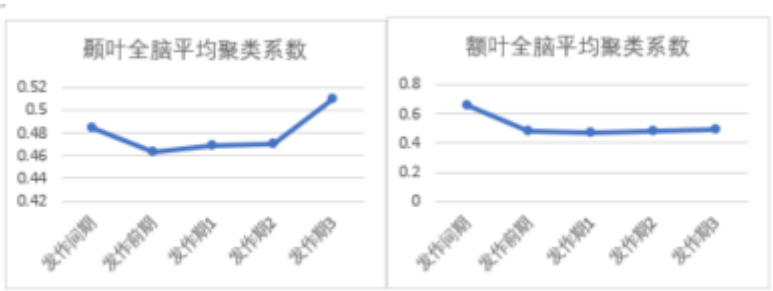


图 4-5. 基于互信息熵的脑网络模型度平均聚类系数

对于聚类系数来说，其描述了网络整体的聚集程度。其数值越大表明网络聚集程度越高。从发作间期开始，网络聚集程度较高，呈现模块化。到发作开始前一段时间，其逐渐下降，考虑是大脑过度放电导致网络较为分散，所以聚集系数下降，一直到发作过程其大脑网络都是呈现分

散的趋势。直至发作后期放电恢复正常大脑又恢复聚集。

5 总结和展望

在这项研究中，我们已经证明了基于同步的大脑网络是随着时空变化的。癫痫发作时 EEG 同步传播路径和核心节点可准确定位癫痫病灶。尤其是互信息指标，给出了信息交互作用程度的量化信息，可以与临床表现相一致。考虑到脑电信号反映了脑内神经元的放电，脑电信号在各通道之间的同步程度代表了信息交换的强度。是以，本钻研可作为摸索动态脑收集的基准。

PU-077

基于 EEG-FMRI 的多模态信息融合技术进展

韦晓燕[1]、马梦楠[1]、陈子怡[2]、周毅[1]
1. 中山大学中山医学院
2. 中山大学附属第一医院神经内科

目的 患者在诊疗过程中会产生多种类型的数据，包括脑电、心电、监护视频、影像、电子病历等等。这些信息对于了解大脑发展机制具有重要作用。长期以来的分析由单一数据类型支撑，但如果可以将多种医疗数据加以智能化融合，多源数据类型可以实现一定程度上的特征互补，达到完善数据特征的作用。

方法 本研究从心电，脑电，FMRI 三种多模态数据信息加以综合，从数据采集，数据预处理，数据融合方法，数据指导应用展开全流程的多模态信息方法论，观察大脑机制，优化信息融合方法。

结果 着重于同步脑电 fMRI 技术和相关的早期研究，处理与同步信号的采集和处理相关的问题，包括对临床和技术前景的批判性讨论。

结论 目前复杂网络领域的研究已取得极大的进展，如何在网络特征水平（如最短路径长度、小世界属性、度分布等），甚至更为复杂的高阶网络水平（如层级性网络、子网络的整合等）上进一步发展多模态信息融合理论方法，值得重点关注。

1.前言

大脑不仅仅在有刺激时候有活动，在静息时一样有神经元同步活动，静息态网络叫做（Rest state network， RSN）。国际联盟组织也定义脑异常疾病为大脑网络异常疾病**错误!未找到引用源。**。

功能磁共振成像（fMRI）具有较高的空间分辨率，越来越多地被用于评价神经和精神疾病患者的脑功能。功能磁共振成像的主要局限是通过相关的缓慢血流动力学变化间接检测神经活动**错误!未找到引用源。**。由于神经-血管耦合可由病理条件或药物局部改变，fMRI 反应可能不能真正

反映神经活动。在 fMRI 数据采集过程中，根据需要可以获得脑电（EEG）记录，以最佳的时间分辨率直接检测事件相关单位，捕捉大脑神经电活动。

因此，将这些技术结合起来以获得同时的 EEG-fMRI 记录越来越引起人们的兴趣。其整合了脑电的高时间分辨率和功能磁共振的高空间分辨率，无创技术，同步大脑电生理活动和血氧变化，**错误!未找到引用源。**全方位检测参与发作间期和发作期癫痫活动的皮质区域。其次，将诱发电位与功能磁共振成像相结合，可以为神经外科或康复计划提供一种准确的研究雄辩皮质区的方法，避免了上述功能磁共振成像的局限性。最后，使用这种方法来评估静息状态网络的功能连通性将扩展 EEG-fMRI 在不合作或无意识患者中的应用。

2.多模态信息融合技术框架

关于脑电和核磁的融合流程其大致如下，包括使用实验仪器进行数据采集，接着数据预处理，数据融合和数据应用等过程**错误!未找到引用源。**。

2.1 实验数据采集设计

关于 EEG-FMRI 的数据采集，很重要的问题是是否应该同步采集，一般根据实验条件可分为同步采集和非同步采集两种方法，受试者一般要求处于一定的条件下如静息态和减少头动。，

1)同步采集

同步采集即同时采集脑电 EEG 和 FMRI 的数据，从而可同步采集血氧代谢活动和神经电生理活动，保证了两种不同采集方式时间上的一致性，是研究脑活动从时间和空间两个维度上的最佳策略。这种对于实验条件要求较高，一般要求实验仪器自带同步采集转换模块，同时同步采集过程中 EEG 和 FMRI 设备会相互产生干扰，增大了后续数据预处理的难度**错误!未找到引用源。**。

2)非同步采集

非同步采集包括两种方式。一种是对同一实验者的 EEG 和 FMRI 分开采集，一种是交替采集，按照顺序 FMRI 采集，接着 EEG 采集再然后 FMRI 采集的方式，可以看作是一个采集周期。非同步就是不同时进行数据采集，避免信号干扰，对于一些实验任务和普通设备可以选择这种方法，但是次忽略了时间对应关系，也无法保证 EEG 与 FMRI 数据时程上的对应关系，缺失了多模态同步研究的意义。

2.2 实验数据预处理

1）EEG 数据预处理

在脑电和核磁共同采集的环境下，脑电会受到很强的磁场干扰，这个干扰的强度一般远远超过脑电信号本身，同时采集环境的不舒适增加了受试者头动心电等变化，因此在这个过程中主要产生两种类型干扰如下

1. 由核磁引起的干扰

首先是射频脉冲产生的 EEG 信号，这类伪迹和 EEG 信号的频率存在较大差异，可以直接采用低通滤波器进行过滤去除。2-梯度伪迹，是由梯度磁场的快速切换导致的，，具有周期性，是普通 EEG 的 100-1000 倍，主要采用伪迹模板相减法**错误!未找到引用源。**，计算出多个核磁扫描周期的平均脑电信号模板，从脑电信号中减去模板即可消除。另外一种是盲源分离法，即采用独立成分分析，主成分分析方法，得到信号的独立成分，识别出伪迹进行剔除**错误!未找到引用源。**。

2）由心电肌电引起的干扰

脑电中还会存在由肌肉和心电活动引起的干扰，可大可小。主要有平均伪迹相减法， 通过借助同步记录的心电波形来获得每次心电伪迹出现的时间， 然后平均各导的心电信号得到心电伪迹模板， 最后再从原始信号中减去此模板**错误!未找到引用源。**。另外是盲源分离法**错误!未找到引用源。**， 通过独立主成分分析和方法将原始 EEG 信号分解为多个成分， 然后根据每个成分的时间进程识别出与心电伪迹相关的成分， 最后从原始信号中减去此成分 去除方法。但是目前仍是难点，亟待开发适用不同心电信号质量不同头动规律的心电伪迹去除方法

2）FMRI 预处理

核磁共振数据的操作主要包含头动校正、功能像注册到结构像、空间标准化、平滑等数据预处理操作。

1.头动校正

头动是核磁共振数据中的一种重要干扰，指的是在实验过程中被试因为头部移动而对数据造成干扰的现象。头部校正，主要分析出被试六个维度的头动信息并在多维空间中通过矩阵变换进行旋转校正，目前可在核磁仪器中设置。

2.切片时间校正

对全脑的扫描是分层进行的，层与层之间扫描的时间是有时间差异的，时间差可达数秒。需要采用切片时间校正，数据操作通过在时间轴上移动信号即通过移动信号的相位来实现切片时间校正。该操作首先通过傅里叶变换使核磁共振信号转换到频域，然后通过将信号与一个滑动窗卷积以实现相位移动，使所有扫描层对应于同一个数据扫描时间点。

3.结构像和功能像的操作

核磁共振的数据分为两种，一种是基于 T1 像的结构像数据，一种是基于 T2 像的功能像数据**错误!未找到引用源。**。结构像的数据特点是分辨率很高，功能像的特点是可以对神经活动进行记录，二者需要通过算法匹配以提高功能像的数据精度，通过仪器即可完成。

2.3 EEG-FMRI 数据融合方法

EEG-FMRI 的数据融合技术经过了多年的发展，已趋于成熟，主要是包括对称融合和非对称融合融合方法，对称融合即同时分别分析 EEG 和 FMRI 的数据特征建模。非对称方法是在 EEG/FMRI 基础去去分析 FMRI/EEG 的数据建模，即利用其中一种模态的信息来指导另外一种模态的分析。

2.3.1 不对称融合

1）EEG 指导 FMRI，用特定的 EEG 特征卷积上血氧动力学响应函数对 FMRI 的波形进行约束

通常的做法是从脑电中提取神经活动的时间过程并以此来推动核磁共振的数据分析。基于神经血管的线性耦合假设，EEG 中提取到的神经活动特征序列与血氧动力学响应函数(Hemodynamic Response Function，HRF) 进行卷积**错误!未找到引用源。**。之后，利用一般线性模型(General Linear Model，GLM)构成了预测的功能磁共振信号，然后在全脑的功能磁共振信号中找出相关的神经活动激活区域。虽然融合步骤相同，但是选择不同的脑电特征，就选择了不同的融合原理，其背后的神经基础也随之不同。

2）FMRI 指导 EEG，用从 FMRI 活动的空间活动信息来约束 EEG 的源重建

空间约束, 即基于 fMRI 空间约束的 EEG 源成像[5].该融合框架是通过对 fMRI 空间信息的研究, 使用 fMRI 空间信息约束 EEG 信号中的源位置. 其方法通常是采用独立成分分析 (Independent component analysis, ICA) 算法从 fMRI 中提取大脑功能网络的感兴趣区域 (Region of interest, ROI) 并以之作为 EEG 源定位的先验信息, 其核心思想就是利用 fMRI 的空间先验信息来提高 EEG 源定位的精确度**错误!未找到引用源。**. 避免依赖 EEG 数据衍生出源空间，所以相对于 EEG－Informed fMRI 来说，fMRI－informed EEG 能减少 EEG 中容积导体的影响。

2.3.2 对称融合

建立共同的生成模型或者利用交互信息来解释两种模态的数据.该方法采用共同的构造模型来融合两种模态的数据, 通常采用 ICA 算法对 EEG 与 fMRI 数据分别在时间和空间上做独立成分分解, 之后基于 EEG 和 fMRI 的独立成分, 分别在空间模式和时间过程上建模. 时空对称融合方法结合空间约束和时间预测两种模型并进行时空的并行整合来达到时间和空间上的融合**错误!未找到引用源。**. 典型相关分析 (Canonical correlation analysis, CCA) 方法基于线性混合模型从各组数据中分别提取相关的典型相关成分 (EEG 的时间信息、 fMRI 的空间信息),通过寻找两种模

态数据的最大共变性确定数据集中成分之间的相关性，并计算相关系数. 可以通过这种方法确定与 EEG 波形相类似的 fMRI 激活区域的动力学响应. 与使用单一变量确定 EEG 一个潜伏期内特征和单个 fMRI 激活区域之间关系的方法不同。

2.3.3 方法对比

在 EEG 与 FMRI 对称融合分析中，使用这种方法只需分别对两种数据进行处理，但两种数据的关系不紧密，两者没有直接联系。在 EEG 指导 FMRI 中，对 EEG 不同信号指标与 FMRI 的 BOLD 信号做相关分析。可以利用 EEG 的不同信号指标研究不同的自发活动机制，是相对成熟且较为常用的融合方法，但是容易受 EEG 的心电和梯度伪迹影响。在 fMRI 指导 EEG 中，通过 FMRI 建立源空间，在源空间中提取该源在 EEG 中的时间进程。减少了 EEG 中容积导体的影响，减少了伪迹影响，特别是心电伪迹**错误!未找到引用源。**。

2.3.4 开源软件

目前两个较大的开源软件分享平台“1000 人功能连接体计划”（http://www.nitrc.org/projects/fcon_1000/）与 GitHub 开源共享平台（https://github.com/open-source）都提供了丰富的开源软件包下载资源。如下表也是目前国内外流行的数据融合工具**错误!未找到引用源。**。

3.EEG-FMRI 应用

01-癫痫

可用于癫痫病灶的定位，这些功能磁共振信号改变与痫样放电区域具有高度的一致性。同步脑电图结合功 能磁共振成像同时测量与癫痫电事件相关的神经和血流动 力学变化。不仅仅是在术前评估中 为外科手术候选者定位， 它还显示 出对手术结果的强 预测价值 ，并且有 助于识别局灶性或全身性癫痫综合征的特异神经元网络。但是在产生棘波的痫样放电其一致性关系不能证明，同时对于脑神经畸变的患者也不能证明**错误!未找到引用源。**。痫样放电 B 信号改变不仅在预期的致痫灶，也可 出现在附近甚至远隔脑区**错误!未找到引用源。**。

02-大脑自发节律和活动

脑电静息态为切入点研究大脑节律主要有两种**错误!未找到引用源。**：（ 1）基于节律的动态分析方法， 研究静息态下不同频段的 EEG 节律，主要关注其空间分布、谱分布、峰值位置等信息;（ 2）基于微状态的分析,主要关注其持续时间、幅度等特征探讨节律与脑网络关系，静息态脑网络与多个节律有关，表明不同频段的神经振荡共同作用于同一功能系统。

03-睡眠和认知功能

在 FMRI 中同时记录到的 EEG 数据可以用来区分不同的睡眠时相， 进而能够探究睡眠进程中血氧代谢活动的动态变化过程。同时分离加工各个认知过程，并进行区分机制. 提取单试次事件相关单位的特征如 P300, 比如成分的幅度，潜伏期，空间拓扑分布等**错误!未找到引用源。**。

未来展望

神经退行性疾病和精神疾病的发病率越来越高，需要越来越先进的诊断和监测工具。这篇综述将着重于同步脑电 fMRI 技术和相关的早期研究，处理与同步信号的采集和处理相关的问题，包括对临床和技术前景的批判性讨论。目前复杂网络领域的研究已取得极大的进展，如何在网络特征水平（如最短路径长度、小世界属性、度分布等），甚至更为复杂的高阶网络水平（如层级性网络、子网络的整合等）上进一步发展多模态信息融合理论方法，值得重点关注。**错误!未找到引用源。**随着多种成像技术手段的发展，针对海量的多模数据进行有效挖掘，如深度机器学习等等。构建多中心、多模数据的共享、整合以及融合分析平台，实现对脑信息的充分挖掘。

PU-078

基于注意力 U 型网络的眼底图像硬性渗出物检测

张亮军、承垠林、马梦楠、 周毅
中山大学中山医学院

目的 对视网膜眼底图像中的硬性渗出物进行准确检测，为未来人工智能在医学图像的发展提供新的参考。

方法 采用基于注意力机制的 U 型网络的方法进行硬性渗出物的检测。

结果 本研究使用的方法在 RIM-ONE 达到了测试集 AUC 的最终值为 0.9328，训练集 AUC 的最终值为 0.9998，AUPRC 的值为 0.5734。

结论 基于注意力机制的 U 型网络的方法，可以对硬性渗出物起到很好的检测作用，对未来人工智能在医学领域的应用有参考价值。

1.国内外研究现状

硬性渗出（hard exudates，HE）是糖尿病视网膜病变的表现之一，医学术语描述为：硬性渗出液是脂蛋白和其他蛋白通过异常的视网膜血管渗漏[1]。表现为边缘清晰的白色或黄白色沉淀物[2]，通常排列成团或圆形环，位于视网膜的外层，如图 1 所示，为视网膜眼底图像中的硬性渗出，可见渗出分布呈圆环状。

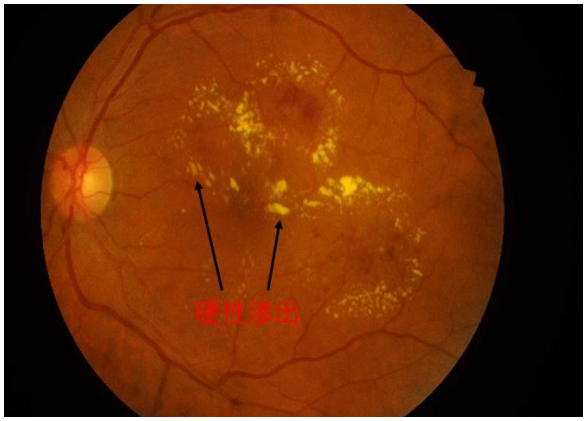


图 1 视网膜眼底图像中的硬性渗出。

硬性渗出会引起视网膜水肿，严重时会导致视力丧失，通过对视网膜眼底图像的分析处理，可以指导患者进行早期治疗，延缓病情[3]。

Osareh[4]等人提出使用模糊 C 均值聚类（FCM）与神经网络（NN）结合的方法实现硬性渗出的检测。该方法首先对眼底图像做预处理，包括选定一个特定的视网膜图像作为参考，将数据集中的数据按照参考图像进行局部对比度增强和颜色矫正，随后使用 FCM 对图像进行聚类，最后使用 NN 对聚类后的区域进行是否为渗出物的分类。Shadgar[5]等人也采用了 FCM 算法对视网膜图像进行分割。具体为首先并对图像进行颜色归一化和对比度增强等预处理，随后使用 FCM 算法对整个图像进行区域分割，建立小区域的图像数据集。为了将分割出的区域分类为渗出区和非渗出区，作者提取了颜色、大小、边缘强度和纹理等特征，接着使用遗传算法对特征进行排序，识别出分类效果最好的子集。最后使用多层神经网络分类器对所选的特征向量进行分类，该算法在视网膜渗出物检测上获得了 93.5%的准确率。颜嵩林等人提出基于多重迁移学习的糖尿病视网膜病变检测，实验结果表明,丰富的层次特征可提高识别精度,均匀背景下的二次迁移可使模型性能更佳[6]。

2.研究内容

本研究使用印度糖尿病视网膜病变图像数据集（Indian Diabetic Retinopathy Image Dataset, IDRiD），使用了其中 81 张用于进行硬性渗出检测的图像作为数据来源，其中训练集 51 张，测试集 30 张，每张图片均对应一张经医生手动标注病变区域图像的作为“金标准”。

2.1 注意力机制

通俗的讲，神经注意力机制可以让神经网络更专注于输入（或特征）的子集，该子集中包括了神经网络需要的特定输入内容。假设$x \in R^d$是输入向量，$z \in R^k$为特征向量，$a \in [0,1]^k$为注意力向量，$g \in R^k$为注意力关注的内容，$f_\phi(x)$是参数为ϕ的一个注意力网络。通常，注意力模型的实现如下：

$$a = f_\phi(x) \#(1)$$

$$g = a \odot z \#(2)$$

式中，⊙表示对应元素相乘，z是参数为θ的神经网络$f_\theta(x)$的输出。注意力可以分为软注意力（soft attention）[7]、硬注意力（hard attention）[8]、自注意力（self-attention）[9]等等。软注意力包含一个值在 0 和 1 之间的掩膜（mask），用该掩膜对特征的权重进行调整。而硬注意力则是先选中注意力区域，随后将该区域的权重调整为 1，其他区域为 0。

为了说明注意力的重要性，首先需要考虑神经网络的真正含义。简单来说，神经网络是一个函数逼近器，其逼近不同函数的能力主要取决于该神经网络的体系结构。一个典型神经网络的实现是通过矩阵乘法和逐个元素非线性的链完成，其中输入或者特征向量的元素只通过加法相互作用。注意力机制会计算一个掩膜（mask），随后利用该掩膜对特征进行处理。

2.2 Attention Gate 模型

通过在标准 CNN 架构中逐渐降低特征图的采样率，可以获取更多的语义上下文信息，从而能够从图像中获取到更多有用的信息。这样，就可以在尺度较大的空间网格中定位特征以及处理不同特征之间的关系。但是，对于显示出较大形状变化的细小物体，仍然很难降低假阳性率。为了提高准确性，本文在 CNN 模型中加入 AGs，与在不同阶段都需要对感兴趣区域重新定位的 CNN 模型相比，AGs 抑制了无关背景区域中的特征响应，而不需要重复定位感兴趣区域。

注意力系数$\alpha_i \in [0,1]$，可用于识别图像中的显著区域，以保留与特定任务相关的激活区域。AG 的输出是输入特征图和注意系数对应元素相乘：

$$\hat{x}_{i,c}^{l} = x_{i,c}^{l} \cdot \alpha_i^{l} \#(3)$$

式中，x^l表示在神经网络中卷积层通过逐层处理局部信息提取到的高维度图像，i和c分别表示空间和通道尺寸。

对图像中的每个像素向量，都需要计算它们的单标量注意力值$x_i^l \in \mathbb{R}^{F_l}$，$F_l$表示在第$l$层中特征图的数量。门控向量（gating vector）$g_i \in \mathbb{R}^{F_g}$用于判断每个像素i是否属于目标区域，门控向量包含上下文信息，可以对图像中的低级别响应特征进行剪枝操作。此外，使用注意力加法计算门控系数，其公式为：

$$q_{att}^{l} = \psi^T\left(\sigma_1\left(W_x^T x_i^l + W_g^T g_i + b_g\right)\right) + b_\psi \#(4)$$

$$\alpha_i^l = \sigma_2\left(q_{att}^l\left(x_i^l, g_i; \Theta_{att}\right)\right) \#(5)$$

式中，$\sigma_2(x_i, c) = \frac{1}{1+\exp(-x_i,c)}$对应 Sigmoid 激活函数。AG 的特征是由一组参数Θ_{att}确定，包括：线性变换$W_x \in \mathbb{R}^{F_l \times F_{int}}$，$W_g \in \mathbb{R}^{F_g \times F_{int}}$，$\psi \in \mathbb{R}^{F_{int} \times 1}$和偏置项$b_\psi \in \mathbb{R}$，$b_g \in \mathbb{R}^{F_{int}}$。这里采用$1 \times 1 \times 1$的张量对线性变换进行计算，该方法也被称为基于向量级联的注意力，级联特征x^l和g被线性映射到$\mathbb{R}^{F_{int}}$维度的空间内部。同时，采用 Sigmoid 激活函数对注意力系数σ_2进行归一化。最后，AG 的参数通过标准的反向传播算法进行训练。注意力加法计算门控系数的流程如图 2 所示，使用注意力系数α，对输入特征x^l进行缩放，使用门信号g在较粗尺度上的激活区域和上下文信息选择空间区域，使用三线性插值进行重采样。

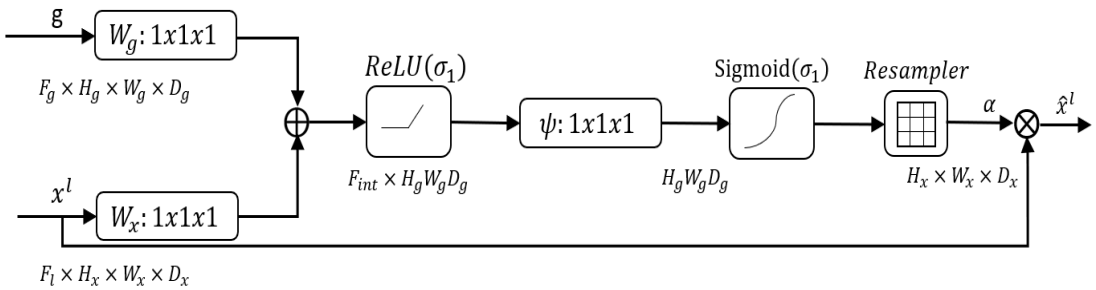


图 2 AG 示意图。

2.3 U 型网络

卷积神经网络主要应用在分类任务当中，对于一般的图像分类来说，输出的是图像的单个类标签，例如 2014 年赢得 ImageNet 大赛冠军的 AlexNet，也是利用卷积网络在庞大的数据集上对图像进行分类。但是在医学图像的视觉任务中，除了输出图片的分类信息，更需要的是输出图像中的定位信息，即每个标签应当分配给每个像素点，而不只是每张图片只有一个最终的标签。

针对缺乏数据的问题，Ciresan[10]等人提出在滑动窗口设置中训练网络，通过将像素周围的局部区域（patch）作为输入来预测每个像素的类别标签。通过对 Ciresan 提出的结构进行分析，可以发现其中有两个问题未得到完美的解决，一个是该结构将图像分成多个 patch，这样就造成了每个 Patch 都需要

运行一次程序，同时由于 patch 之间存在重叠，会造成大量的冗余计算。第二是在定位的准确性和上下文的使用上存在一个权衡问题。较大的 patch 需要更多最大池化层数，但这样会使定位精度降低。较小的 patch 会导致结构只能使用到有限的上下文。

针对上述未解决的问题，Ronneberger[11]等人提出了 U-Net 架构。U-Net 架构像一个大写的字母“U”,包含两条路径，分别为收缩路径（contracting path）和扩张路径（expansive path），如图 3 所示。

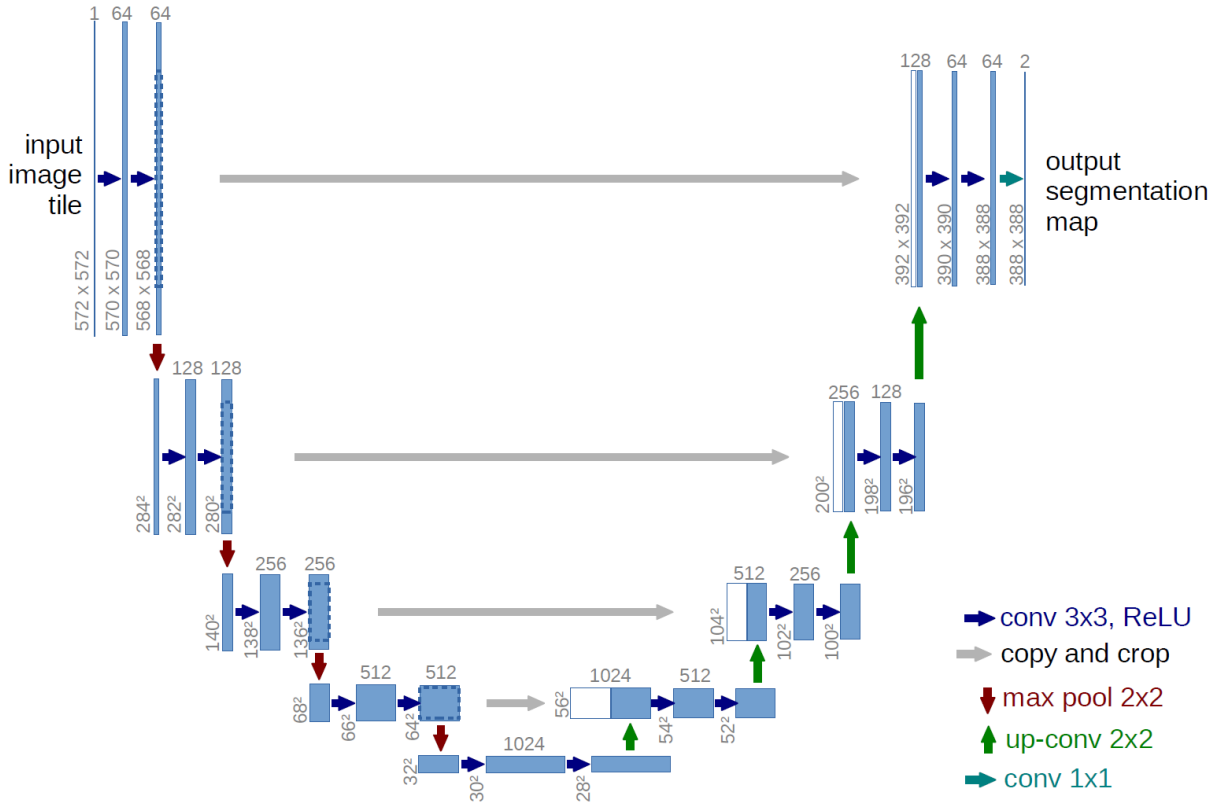


图 3 U-Net 网络架构

2.4 模型训练与测试

Attention U-Net 模型使用 2-4 个样本的小批量计算进行梯度更新。对于较大的网络，在多个向前和向后传递中使用平均梯度。模型的训练依旧使用 Adam 优化器，此外，还使用批量归一化，深度监督和数据增强技术进行训练，数据增强具体包括随机裁剪、平移、缩放等操作。同时，对图像的强度值进行线性缩放，使其符合正态分布$N(0,1)$。

此外，使用 Sorensen-Dice 损失函数[115]对模型进行训练。Dice 系数（Dice coefficient）用于比较两个样本数据的相似性，其取值范围为$[0,1]$，其数学形式为：

$$s = \frac{2|X \cap Y|}{|x| + |Y|} \#(6)$$

式中，X和Y表示两个样本数据，$|X|$表示样本X中元素个数，$s \in [0,1]$，1 表示两个样本完全一致。由于分母中对样本元素存在重复计数，系数 2 用于消除重复计数对计算结果的影响。实验证明，该损失对类别失衡较不敏感。

门控参数需要初始化定义，以使特征向量在任意空间位置都会通过注意力门。另外，注意力系数可以从训练图像中获取，在实验过程中可以观察到，AG 最初在所有位置均具有均匀分布和传递特征，随着训练的进行，它会逐渐更新并定位到目标区域边界。此外，在较粗的尺度上，AG 会提供区域的粗略轮廓，并以较高分辨率对轮廓逐步完善。通过在每个图像尺度上训练多个 AG，可以观察到每个 AG 都会专注于特定区域的某个子集。

3.实验结果

在进行眼底图像硬性渗出检测时，每个像素点被分类为渗出和背景像素，为了评估所提算法的性能，本研究采用 AUC（Area Under Curve，AUC）和 AUPRC（Area Under Precision Recall Curve）2 个衡量指标对实验结果进行分析。

AUC 被定义为 ROC（Receiver Operating Characteristic，ROC）曲线下面积，这两者常用来评估二值分类器的性能。ROC 曲线的横坐标为 FPR（False Positive Rate，FPR），纵坐标为 TPR（True Positive Rate，TPR），在进行像素分类时，每个像素点都对应一个概率输出，用于表示该像素点属于硬性渗出或者背景像素的可能性。

AUPRC 用于计算 PRC（Precision Recall Curve，PRC）曲线下面积。PRC 曲线的横坐标为召回率，纵坐标为精度，同 ROC 曲线的绘制原理相同，PRC 曲线也是利用多组精度值和召回率绘制曲线。

本研究在模型迭代 3000 次后，测试集 AUC 的最终值为 0.9328，训练集 AUC 的最终值为 0.9998，AUPRC 的值为 0.5734。最终分割结果如图 4 所示：

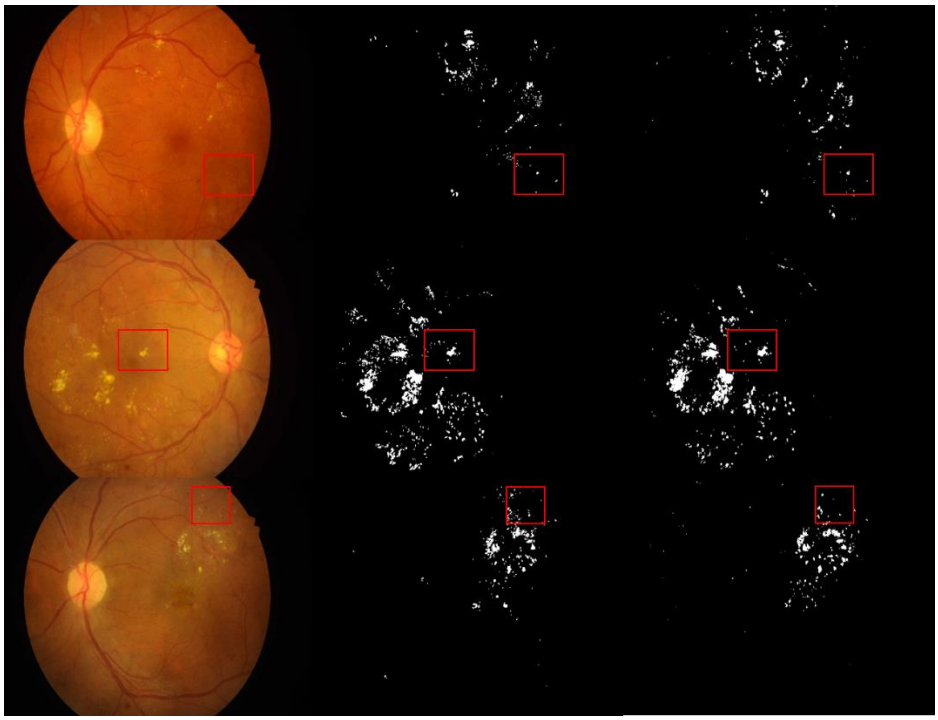


图 4 硬性渗出结果图。第一列为输入原始图像，第二列为“金标准”，第三列为算法分割结果。

PU-079

基于距离正则化水平集演化的视盘分割

张亮军、承垠林、马梦楠、周毅
中山大学中山医学院

目的 对视网膜眼底图像中的视盘进行准确分割，为未来人工智能在医学图像的发展提供新的参考。

方法 采用基于距离正则化水平集演化的方法进行视盘分割。

结果 本研究使用的方法在 IDRiD 达到了平均重叠率为 0.9801。

结论 基于距离正则化水平集演化的方法，可以对视盘起到很好的分割作用，对未来人工智能在医学领域的应用有参考价值。

1.国内外研究现状

视盘是连接视网膜和大脑的重要器官，眼球中所有的血管都汇聚到视盘，视盘也是视网膜图像中最亮的部分。目前，视盘的检测方法具体可分为基于图像强度信息、基于模板匹配和基于视盘形状。

视盘是视网膜中最亮的区域，因此可以利用图像的强度信息对视盘进行检测。Sinthanayothin 等人[1, 2]提出了一种通过定位平均强度变化最大的区域来检测视盘的定位算法。首先，在 HSI 颜色空间提取的亮度通道上应用自适应局部对比度增强对图像进行预处理；然后，使用与视盘大小近似相等的窗口，通过搜索每个像素的强度与相邻像素的强度之间平均变化最大的区域来定位视盘的中心，因为正常情况下许多深色的血管都会汇聚在亮度较强的视盘上。

模板匹配是在图像中查找目标对象的经典方法[3]。Osareh 等人[4]利用模板匹配对视盘中心进行定位。该方法通过灰度形态学将视盘区域定位，然后进行 Snake 模型拟合。作者提出并实现了 Snake 模型的自动初始化和形态学在色彩空间中的应用。具体实现为：首先在预处理中使用形态学方法去除眼底血管，接着使用 Snake 算法提取视盘边界，该算法在视盘的增强区域初始化，最后对视盘区域进行提取。

由于视盘在视网膜中呈大圆形，因此可以使用 Hough 变换检测视盘，Hough 变换能够检测图像中的几何形状[5, 6]。Abdel Ghafar 等人提出基于 Hough 变换的视盘定位方法。该算法首先利用形态学技术去除血管，然后使用 Sobel 算子提取图像中的边缘，最后将圆形 Hough 变换应用到边缘图中，对视盘的边界进行估计，将其中最大圆形定位视盘。

2.研究内容

本研究采用 RIM-ONE[9](retinal image database for optic nerve evaluation)和 DRIONS-DB(Digital Retinal Images for Optic Nerve Segmentation Database)数据集进行视盘分割。

2.1 图像灰度化

在视盘分割的图像预处理中，主要使用 PCA 实现对原始图像的灰度化。根据上述对 PCA 原理的阐述可以知道，通过 PCA 可以对数据进行降维，同时将不同维度的数据投影到一维向量上。因此，利用 PCA 进行图像灰度化的步骤为：

（1）对图像 R,G,B 三个通道数据进行归一化处理，即中心化。

（2）计算中心化后图像数据的协方差矩阵，此时图像数据实际上为三维数据：

$$Cov(R,G,B)=\begin{bmatrix} Cov(R,R) & Cov(R,G) & Cov(R,B) \\ Cov(G,R) & Cov(G,G) & Cov(G,B) \\ Cov(B,R) & Cov(B,G) & Cov(B,B) \end{bmatrix} \quad (5\text{-}5)$$

（3）计算协方差矩阵$Cov(R,G,B)$的特征值和特征向量；

（4）选择特征值最大的特征向量，将图像数据的值映射到该向量对应的坐标上，完成图像的灰度化。

2.2 形态学去除血管

本研究采用数学形态学的相关操作对视盘图像中的血管进行去除，从而减少血管对视盘分割的影响。数学形态学是由 Matheron 和 Serra[10]提出的可用于提取图像成分的一种方法。数学形态学主要应用在二值图像的处理中，目前已经推广到灰度图像，甚至是彩色图像，形态学操作都是基于简单的对图像的扩张或收缩操作。

本研究基于距离正则化水平集演化方法，提出基于图像边缘的视盘分割方法，用于对眼底图像中的视盘进行分割。

设$\phi:\Omega\to\Re$是定义在域Ω的 LSF，定义能量函数$\xi(\phi)$为：

$$\xi(\phi)=\mu R_p(\phi)+\xi_{ext}(\phi) \quad (1)$$

$$R_p(\phi)=\int_\Omega p(|\nabla\phi|)dx \quad (2)$$

其中，R_p可保持 LSF 有符号距离的属性，定义其为距离正则化项，$\mu>0$是常数，$\xi_{ext}(\phi)$是依赖于图像分割区域的外部能量，$p:[0,\infty)\to\Re$为能量密度函数。当ϕ的零水平集位于图像边界时，$\xi_{ext}(\phi)$具有最小值。

为简便计算，取能量密度函数p为：

$$p=p_1(s)\triangleq\frac{1}{2}(s-1)^2 \quad (3)$$

因此，式（2）可表示为：

$$P(\phi)=\frac{1}{2}\int_\Omega(|\nabla\phi|-1)^2dx \quad (4)$$

上式用于表征有符号距离函数到ϕ的偏差。

变分法中，计算最小化能量函数$F(\phi)$的标准方法是求梯度流（gradient flow）方程的稳态解：

$$\frac{\partial\phi}{\partial t}=-\frac{\partial F}{\partial\phi} \quad (5)$$

$\partial F/\partial\phi$为函数$F(\phi)$的加托导数，$-\partial F/\partial\phi$是函数$F(\phi)$的最快下降方向。因此，梯度流也被称为最快下降流或梯度下降流。

对式（2）求解加托导数，可得：

$$\frac{\partial R_p}{\partial \phi} = -div(d_p(|\nabla\phi|)\nabla\phi) \quad (6)$$

$$d_p(s) \triangleq \frac{p'(s)}{s} \quad (7)$$

式中，$div(\cdot)$为散度算子。

从式（1）和加托导数的线性可得：

$$\frac{\partial \xi}{\partial \phi} = \mu\frac{\partial R_p}{\partial \phi} + \frac{\partial \xi_{ext}}{\partial \phi} \quad (8)$$

式中，$\frac{\partial \xi_{ext}}{\partial \phi}$是外部能量函数$\xi_{ext}$关于$\phi$的加托导数。能量函数$\xi(\phi)$的梯度流为：

$$\frac{\partial \phi}{\partial t} = -\mu\frac{\partial R_p}{\partial \phi} - \frac{\partial \xi_{ext}}{\partial \phi} \quad (9)$$

将上式与式（8）联立，可得：

$$\frac{\partial \phi}{\partial t} = -\mu div(d_p(|\nabla\phi|)\nabla\phi) - \frac{\partial \xi_{ext}}{\partial \phi} \quad (10)$$

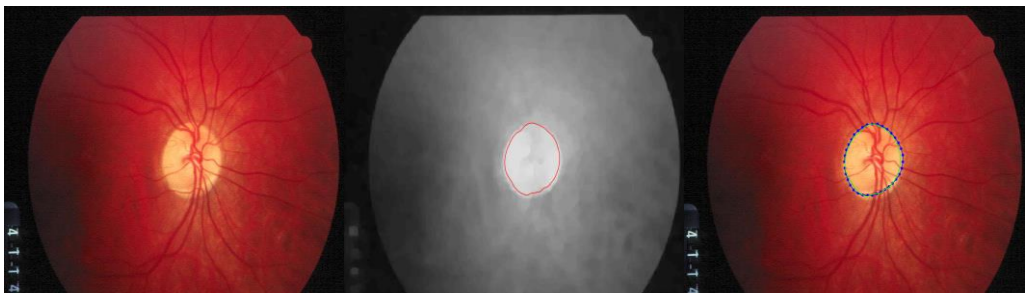
式（10）称为 DRLSE，因为它可保持 LSF 的有符号距离特性，且该式与式（1）中的距离正则化项R_p相关。通常采用诺伊曼边界条件（即第二类边值条件）和给定的初始化函数ϕ_0对该偏微分方程进行求解。

3. 实验结果

本研究通过计算重叠率的方法对视盘分割效果进行评估，该方法的数学表达为：

$$S = (Seg \cap R_{seg})/(Seg \cup R_{seg}) \quad (11)$$

其中，Seg是所提算法对视盘分割结果的像素集合，R_{seg}是“金标准”所包含的像素集合，该方法以像素点重叠的数量作为评估标准。其中，S的值位于 0 和 1 之间，值越大，表明两者的图像重叠率越高，算法精度越高，分割效果越好。如图 1 所示，为本研究视盘分割结果，（a）为原始图像，（b）为对预处理后的图像进行分割，（c）为算法分割结果与医生手动标注的进行对比。



(a) (b) (c)

图 1 视盘分割结果。（a）为原始图像，（b）为对预处理后的图像进行分割，（c）为算法分割结果与医生手动标注的进行对比，蓝色为医生手动标注，绿色为算法分割结果。

本研究使用的方法在 RIM-ONE 达到了平均重叠率为 0.9801。由于 RIM-ONE 数据集分辨率较高且仅包括视盘区域的图像，重叠率较高。从以上结果可以看出，本研究所采用的方法在视盘分割中具有良好的效果。

PU-080

新一代信息技术在 2019 新型冠状病毒肺炎防控中的研究与应用

刘迷迷、李星雨、周毅
中山大学中山医学院

目的 随着新一代信息技术革命的不断深入，社会各行各业呈现出数字化、网络化、智能化融合应用与发展的态势。新一代信息技术可以高效、智能地对海量结构化数据和非结构化的健康医疗数据进行有效处理及利用，是公共卫生疫情防控领域的新动能。本文主要目的是通过调研，了解和分析新一代信息技术在 2019 新型冠状病毒肺炎防控中的研究与应用，并对新一代信息技术应用于公共卫生疫情防控领域的未来发展进行展望。

方法 基于新一代信息技术、2019 年新型冠状病毒肺炎防控和公共卫生疫情防控等方面的关键词，在网络搜索引擎、文献数据库，以及各级政府部门、医疗健康相关事业单位和医学信息技术企业等官方网站进行调研，检索相关政策文件、重要讲话、疫情防控指南、行业报告和文献报道。对检索资料进行详细分析，总结新一代信息技术在 2019 新型冠状病毒肺炎防控中的研究与应用情况。

结果 对相关政策文件、重要讲话、疫情防控指南、行业报告和文献报道等检索资料进行详细分析，发现大数据和人工智能等新一代信息技术在临床诊断、病毒基因检测与溯源和疫情防控等方面已经有较多的研究与应用，体现了新一代信息技术在重大疫情防控的强大驱动作用。

结论 在国家领导及相关政府部门的鼓励下，大数据、物联网、云计算、5G 技术和人工智能等新一代信息技术在临床诊断、病毒基因检测与溯源和疫情防控等方面得到了广泛的应用，成果显著。但应用和研究的深度和全面性还不够，并且缺少物联网、云技术和 5G 技术的应用，有待进一步提升和发展。

引言

2019 年 12 月湖北省武汉市爆发的新型冠状病毒肺炎（简称“新冠肺炎”）疫情，是新中国成立以来在我国发生的传播速度最快、感染范围最广、防控难度最大的一次重大突发公共卫生事件，不到一个月的时间就迅速扩散至全国及其他 30 多个国家。新冠肺炎疫情是继“非典”疫情之后，对我国公共卫生疫情防控治理能力的又一次重大挑战。但与“非典”疫情防控治理不同的是，十多年来我国信息技术的创新、发展和应用持续推进，并取得显著成果，特别是互联网技术已经深度应用于人们的生产生活之中。数字化、网络化和智能化的新一代信息技术的融合应用为社会各行各业带来了全新的动力，加快了相关社会生产生活服务发展和治理管理体系建设，为提升公共

卫生疫情防控奠定了有利的技术基础和社会应用环境[1]。

新一代信息技术是具有数字化、网络化、智能化等突出特征的最新的信息技术。主要包括大数据、物联网、云计算、5G 技术和人工智能（AI）等技术[2]。我国《国务院关于加快培育和发展战略性新兴产业的决定》将新一代信息技术纳入七大新兴的战略产业之一。新一代信息技术使传统的健康医疗领域呈现崭新的面貌，充分发挥新一代信息技术在健康医疗资源配置中的优化和集成作用，将新一代信息技术的创新成果深度融合于健康医疗领域之中，可提升健康医疗领域的创新力，促进健康医疗服务效率和水平的提高，形成更广泛的以信息技术为基础设施和实现工具的健康医疗发展新趋势[3]。

在新冠肺炎疫情防控中，习总书记发言和相关国家政府部门发文，均鼓励运用大数据、人工智能、云计算等数字技术，在疫情监测分析、病毒溯源、防控救治、资源调配等方面更好发挥支撑作用，并充分运用新一代信息技术支撑服务疫情防控和企业复工复产工作[4-6]。在国家领导及相关政府部门的鼓励下，大数据、物联网、云计算、5G 技术和人工智能等新一代信息技术在临床诊断、病毒基因检测与溯源和疫情防控等方面得到了广泛的应用，成果显著。本文以 2019 年新型冠状病毒肺炎为例，详细分析新一代信息技术在公共卫生疫情防控领域的研究与应用情况。最后，将对新一代信息技术在公共卫生疫情防控领域应用的未来发展进行展望。

1 在临床诊断中的研究与应用

新冠肺炎确诊主要有两种方式，核酸检测和临床诊断。其中，核酸检测新型冠状病毒为阳性的即新冠肺炎确诊病例。若患者核酸检测为阴性，即使其影像学检查等特征异常，也只能被诊断为疑似病例，无法作为确诊病例。然而，临床诊断是疑似病例经影像学检查具有新冠肺炎特征的病例，不需要核酸检测。

核酸检测是新冠肺炎检测的“金标准”，但核酸检测存在试剂短缺、假阴性、实验室条件要求高，以及具有检测资质的机构较少等问题。因此临床一线专家提出将影像学诊断也作为新冠肺炎确诊的依据，并纳入新冠肺炎诊断指南和诊疗方案之中。但临床诊断仅限于疫情严重的湖北省，湖北省以外的其他省市自治区，核酸检测仍然是新冠肺炎确诊的唯一依据。其主要原因是，利用大数据和人工智能技术可以快速且较准确地筛选出病例，以便患者及早按照确诊病例接受规范治疗，进一步提高救治成功率。

影像学诊断在新冠肺炎确诊中逐步起主导作用，而影像人工智能是目前医学人工智能发展成熟、成果显著的领域，因此影像人工智能诊断在新冠肺炎诊断中发挥着越来越重要的作用。Shuai Wang 等利用改进的深度学习模型来辅助筛选典型的 CT 图像可能有助于早期筛查疑似病例[7]。据报道，各医疗人工智能企业推出了“CT+AI”的新冠肺炎辅助诊断系统或在原有的肺炎 AI 产品上强化了新冠肺炎的检出功能，并在武汉同济医院、武汉协和医院、武汉大学中南医院等疫区一线医院上线新冠肺炎影像辅助诊断产品，为放射科医生在没有核酸检测条件下的 CT 影像诊断提供智能化分析与预后方案，辅助决策[8]。上海市公共卫生临床中心上线新冠肺炎智能影像评价系统，采用人工智能全肺定量分析技术为临床专家提供基于 CT 影像的智能化新型冠状病毒性病灶定量分析及

疗效评价等服务，可高效、准确地为临床医生提供决策依据[9]。

此外，以色列初创公司 Vocalis Health 开发一种通过语音检测新冠病毒的方法，基于 AI 算法对人群语音样本进行分析，识别感染者语音的特点，未来将与医院合作牵头研究，通过移动应用程序对确诊患者的声音进行采样，优化算法，从而用于辅助远程诊断和监测[10]。这种创新的新冠肺炎诊断筛查方式，是基于人群健康的更全方位的数据，利用大数据和人工智能等新一代信息技术来辅助诊断筛查，目前报道较少。

最后，我国最新新冠肺炎诊疗方案根据是否有临床症状、肺炎及肺炎的严重程度，以及是否出现呼吸衰竭、休克、其他器官功能衰竭等症状，将新冠肺炎患者分为轻型、普通型、重型和危重型，以指导临床医生有针对性地分级诊疗不同类型的病例，以提高现有医疗资源利用率。但患者入院诊疗的指标数量众多，数据类型多样，且各指标间关系复杂，人为无法根据这些指标快速判断患者的临床分型，此时利用大数据和人工智能技术辅助病例的临床分型无疑成为最佳方式。然而，目前未见利用大数据和人工智能技术辅助新冠肺炎临床分型的相关报道，有待进一步跟踪研究。

2 在病毒基因检测与溯源中的研究与应用

在病毒基因检测方面，传统的核酸检测过程复杂、时间长，且病例需要向专业检测中心集中运输处理，这些痛点极大地影响了新冠肺炎患者的确诊效率，使得潜在感染者的防控措施不能及时，也就无法从根本上切断传染源。利用大数据和 AI 算法可以有效提升基因检测效率，深兰科技全资企业基于便携式第三代测序仪把整个检测过程缩短到 45 分钟之内，同时应用深度学习算法对各种病原微生物基因组和人源基因组进行快速比对，不但可以快速测定是否感染新冠病毒，而且可以同时知道病毒负荷、并发感染以及病人的多项应激反应情况，从而帮助快速了解病人的病程和治疗效果。浙江省疾控中心上线自动化的全基因组检测分析平台，利用阿里达摩院研发的 AI 算法，可将原来数小时的疑似病例基因分析缩短至半小时，大幅缩短确诊时间，并能精准检测出病毒的变异情况。此外，百度大脑基因检测，将病毒全基因序列组结构预测从近一个小时缩短至 30 秒以内。

在病毒溯源方面，利用人工智能等算法构建病毒宿主预测模型，可将提取的新型冠状病毒的基因组，与现有病毒基因数据库做数据检索和比对，实现新型冠状病毒自然宿主的寻找和预测。Qian Guo 等学者利用深度学习算法预测新型冠状病毒的宿主和传染性，以 DNA 序列为输入检测可以感染的宿主病毒，通过对模型计算的结果分析，筛选的病毒宿主有犬、猪、貂、龟和猫，经分析比较发现蝙蝠和水貂感染的冠状病毒有很强的感染人类的潜力，该研究结果对未来病毒分析和病毒性病原体的早期防治有一定的指导意义[11]。Tianmu Chen 等建立了一个“Bats-Hosts-Reservoir-People”传播网络模型，以模拟从感染源（可能是蝙蝠）到人类感染的潜在传播，并根据模型计算病毒的基本再生数（R0），以评估新型冠状病毒的传播能力[12]。

3 在疫情防控中的研究与应用

面对全国乃至全球大流行、大爆发的新冠肺炎疫情防控，新一代信息技术在疫情的数据统计

与报告、流行病学研究、药物筛选与评价等方面的重要作用凸显：

在数据统计与报告方面，加强疫情防控综合信息平台的建立和优化，充分利用大数据分析挖掘和人工智能技术，可提高新冠肺炎疫情信息报告及监测预警的智能化和工作效率。北京大学前沿计算研究中心陈宝权等学者利用热度图等方法对新冠肺炎的传播区域及其特点进行可视化分析，并建立模型仿真预测新冠肺炎的传播力及发展形势[13]。美国约翰斯• 霍普金斯大学的董恩盛和杜鸿儒汇总、整理和清洗全球新冠肺炎疫情数据，利用可视化等技术制作了全球新冠肺炎疫情实施检测的统计图表，并自动更新数据，成为全球权威新冠肺炎疫情数据图[14]。

在流行病学研究方面，应用大数据、人工智能和地理空间等信息技术，可提高新冠肺炎疫情流行病学调查、暴发流行发展势态预测评估与危害性评价、地理空间疫情分析与控制能力，同时提出预防控制对策，为社区、医院、疾病预防控制中心和卫生健康局等相关单位和部门的新冠肺炎疫情精准决策提供科学依据，辅助防范和化解疫情风险。

浙江省依托大数据优势，通过“大数据+网格化”的方法研判疫情，利用大数据分析出人群聚集热点分布、人群跨区域流动等信息，提前预测疫情发展趋势，指导医疗资源的合理调度。同时，通过数据回溯分析，相关部门可以尽早发现疑似病患、密切接触者，有助于及时隔离、切断传染源[15]。此外，Zixin Hu 等提出了基于人工智能的方法来实时预测我国各地新冠肺炎疫情的大小、长度和结束时间，发现基于人工智能的新冠肺炎发展预测方法具有较高的精度，并得出我国新冠肺炎的流行将在 4 月底基本结束，与我国新冠肺炎疫情发展的实际情况吻合[16]。Tianyu Zeng 等提出了三种无模型预测方法，以及一种多模型常微分方程组神经网络模型，来综合预测新冠肺炎疫情在我国大陆跨省传播，特别是湖北省的跨省传播，以及疫情传播结束，并发现综合数学和人工智能的方法可以对新冠肺炎疫情传播结束一致的、合理的预测[17]。Xumao Zhao 等基于百度地图的交通流量，以及离开武汉的航空旅等大数据，来量化潜在的传染人群，并建立以当地人口和航空旅客为预测变量的多元线性模型，来解释中国各城市确诊病例的差异[18]。

在药物筛选与评价方面，大数据和人工智能技术可以帮助药物研究与开发人员，充分利用现有数据来解决紧迫的公共卫生问题。与传统的药物研究与开发工作流程相比，能够非常快速地发现潜在药物，为社会选择合适的、已上市的药物提供合理的建议，也可节约新药物研究的时间和资源。伦敦帝国大学和 BenevolentAI 人工智能新药研发公司在柳叶刀（Lancet）期刊发表文章，报道利用深度学习、知识图谱的人工智能技术对可能有效阻止病毒感染的已上市药物进行了搜索，快速对接数十亿个分子结构，筛选出可治疗新冠肺炎的潜在药物巴瑞替尼（Baricitinib）[19]。Anh-Tien Ton 等开发了一种新的深度学习平台——深度对接，通过深度对接 13 亿化合物，快速识别 2019 年新型冠状病毒主要蛋白酶的潜在抑制剂。并公开提供新型冠状病毒前 1000 个潜在配体，供科学界进一步鉴定和开发，促进全球努力开发新的候选药物[20]。Zeheng Wang 等使用基于本体的副作用预测框架和深度学习评估中国官方推荐的新冠肺炎中药，得出 7 个安全性指标较高的中药方剂，建议在治疗新冠肺炎时应首先考虑这些方剂[21]。

4 总结与展望

目前新一代信息技术在新冠肺炎的临床诊断、病毒基因检测与溯源和疫情防控等方面的应用比较分散，其深度和全面性还不够，并且缺少物联网、云技术和 5G 技术的应用，有待进一步提升和发展。在国家和地区公共卫生疫情防控的宏观层面，新一代信息技术可以发挥更大的作用，比如在联防联控、应急指挥和物资保障与供应等方面。基于当前新一代信息技术在 2019 年新冠肺炎疫情防控中的研究与应用实例分析，未来，如何充分发挥新一代信息技术强大的作用，助力公共卫生疫情防控，关键在于：加强公共卫生疫情防控数据汇聚与信息协同，以及建立相关大数据协调系统平台，提高公共卫生疫情防控信息化建设；以大数据、物联网、云技术和 5G 技术等新一代信息技术为支撑，利用医疗联合体和分级诊疗等疫情防控模式，有效整合与协调区域内相关单位和部门的应急体制机制和物资保障与供应等。

在加强公共卫生疫情防控数据汇聚与信息协同和建立相关大数据协调系统平台方面，第一，加强国内外相关信息标准的建立与现有标准的使用，建立公共卫生疫情防控跨领域、跨机构的数据汇聚与信息协同发展路线，促进术语、数据元、数据交换格式、互联接口服务等相关标准体系，以及数据抽取、模式转换、术语映射等相关标准数据整合技术的研究与应用。第二，加强公共卫生疫情防控数据中心、数据平台和信息系统等信息化建设，建立跨机构且包含大数据汇聚、平台展示和人工智能辅助决策等功能的综合系统平台，以保障和促进公共卫生疫情防控跨领域、跨机构的数据互联互通，以及数据的有效利用。

在公共卫生疫情防控的病原体数据收集、疫情监测分析、病毒溯源、防控救治、信息报送、资源调度、捐赠受赠、应急指挥、综合调度等方面，需要优化已建立的大数据协调系统平台，并以大数据、物联网、云技术和 5G 技术等新一代信息技术为支撑，促进医疗联合体和分级诊疗等公共卫生疫情防控模式的应用，实现区域之间、部门之间优势互补的信息共享、物资支持与快速扩容、联合救援行动，以提高联防联控、应急指挥能力和物资保障与供应水平。

综上，大数据、物联网、云技术、5G 技术和人工智能等新一代信息技术有力地突破了传统公共卫生疫情防控在资源、可及、均等、智能、知识等方面的局限，促使资源配置更合理、防控质量更保证、防控过程更便捷、防治服务更高效、防治费用更适宜，促进了我国公共卫生事业发展，提高了人民健康水平。未来，新一代信息技术将在公共卫生疫情防控领域发挥更重要的作用。为充分发挥新一代信息技术在公共卫生疫情防控中的驱动作用，相关医疗机构、政府部门、高校、科研院所和企业应通力合作，加强数据共享、科研协作和成果转化与应用，共同应对突发重大公共卫生疫情。

摘要刊登

PA-001

健康大数据与精准健康管理

刘福荣
郑州大学第一附属医院

目的 借助大数据，加快精准健康管理及精准健康管理体系的可持续发展。

方法 通过文献综述，了解健康大数据与精准健康管理之间的关系。

结果 大数据时代的来临，给人类社会带来了生活、工作与思维的大变革。精准健康管理是医学科技发展、多学科交叉融合推动、多种医学理念跃迁、医学模式转化的创新产物。大数据挖掘技术和人工智能应用技术是推动精准健康管理不断发展的重要技术保障。解决精准健康管理如何发展的问题，关键在于如何利用大数据更加精准地揭示个体遗传特征与环境因素的相互作用对健康的影响，深度开发与健康状态相适应的、功能强大的健康风险预测模型，从而实现全生命周期精准、有效干预

结论 我国精准健康管理的发展尚处于初步阶段，未来的发展道路艰辛而漫长。因此，提高大数据管理能力、加强复杂的网络环境下动态数据的利用和隐私保护、建立统一的数据标准、加强大数据挖掘技术的动态开发能力等是加快精准健康管理及精准健康管理体系的可持续发展的重要保障。