

· 专题二:双清论坛“全维度数据与智能诊疗的前沿与挑战” ·

全维度数据与智能诊疗的前沿与挑战

——第 232 期“双清论坛”学术综述

冯建峰¹ 王拥军² 徐 涛³ 王守岩¹ 曹 森¹
谢小华¹ 孟 霞² 李 焱³ 孙洪赞⁴ 曹河圻^{4*}

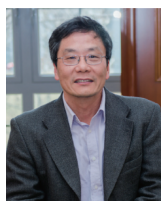
1. 复旦大学 类脑智能科学与技术研究院, 上海 200433
2. 首都医科大学 附属天坛医院, 北京 100050
3. 中国科学院大学, 北京 100049
4. 国家自然科学基金委员会 医学科学部, 北京 100085

[摘 要] 基于第 232 期“双清论坛”, 本文分析了我国全维度数据与智能诊疗发展的前沿和趋势, 从大规模人群队列建设、生物标志物挖掘和人工智能大数据分析三个方面总结了近年主要研究进展, 提出了下一步发展机遇和挑战, 探讨了未来全维度数据与智能诊疗多学科交叉发展目标和科学基金重点资助方向。

[关键词] 全维度; 大数据; 智能诊疗; 学科交叉; 临床应用

我国医疗健康事业的快速发展对医疗新技术的变革与应用提出了迫切需求, 而人工智能理论和技术与医疗健康卫生大数据、信息化系统的发展相融合, 为全维度数据疾病诊断、智能医疗决策等战略性新兴产业领域带来了巨大的发展机遇^[1]。近年来, 医疗领域的数字化进程不断向纵深方向推进, 并逐渐向智慧医疗阶段迈进^[2, 3]。智慧医疗以人工智能技术为工具, 提供基于大数据的系统化精准化医疗服务。当前, 人工智能在医疗健康领域的应用已经非常广泛, 从应用场景来看主要分成虚拟助理、医学影像、药物挖掘、生物技术、健康管理、可穿戴设备、风险管理等多个领域。人工智能在医疗领域的应用带来了诊疗模式、数据处理方式、前瞻性健康管理等诸多方面的变革, 在智能化临床诊疗、可视化数据价值提升、智能化健康管理、精准医学等领域发挥重要价值, 推动现代医疗向智慧、精准、高效发展。

国家自然科学基金委员会(以下简称“自然科学基金委”)第 232 期“双清论坛”于 2019 年 5 月 11~12 日在上海召开, 以“全维度数据与智能诊疗的前沿与挑战”为主题, 充分而深入地研讨了我国全维度



冯建峰 教授, 上海数学中心首席科学家, 复旦大学类脑智能科学与技术研究院院长、大数据学院院长、上海脑科学与类脑研究中心副主任。长期致力于发展数学、统计与计算机的理论和方法, 并原创性地将它们应用于解决神经科学和各类脑疾病中的具体问题。2011 年获英国皇家学会沃夫森研究功勋奖。



曹河圻 研究员, 博士。现任国家自然科学基金委员会医学科学部五处处长。

数据与智能诊疗发展的前沿和挑战, 总结了生物、计算、临床等多学科交叉取得的主要进展和成就, 凝练了围绕智能精准医疗的生物和数学基础研究、智能诊疗共性关键技术研究以及重大疾病诊治辅助智能决策研究的关键科学问题, 探讨了前沿研究方向和科学基金资助战略, 短期内希望在全维度数据智能精准诊疗等方面取得突破性成果, 进而通过 5~8 年

时间实现基于大数据的慢性病发病机理和干预策略研究,推动基础到应用的原创性研究突破。

1 全维度数据与智能诊疗多学科交叉前沿进展和发展趋势

1.1 全维度数据及大规模人群队列建设

20世纪中叶以来,对慢性病发病机制研究不断深入,尤其是肿瘤、心血管疾病等的发病机制复杂,不单纯由某一先天遗传特征或后天环境暴露引起,而是由基因、环境、行为等多因素联合作用导致^[4]。目前公认的大型队列研究样本量多在5万至50万之间^[5],其中以10万以上居多。这些人群队列通过整合基因组学、表观组学、蛋白组学、代谢组学等多个水平上的生物标志物,结合传统流行病学宏观研究的暴露组学,并依托大数据科学、分子影像等新技术,可以更好地理解疾病发生、发展的生物学机制,服务于疾病精准的预防、诊断和治疗。

根据大型队列的来源,可分为两类,一类是在已有较小规模(相对大型队列而言)的研究基础上,采取合并、协作、追加样本量等方式建立的队列,如欧洲营养与肿瘤前瞻性调查(European Prospective Investigation into Cancer and Nutrition, EPIC);另一类是新建的队列,如英国生物样本库(UK Biobank)、美国的All of Us Research Program和中国的China Kadoorie Biobank(CKB)。

EPIC于1992年启动,旨在研究欧洲普通人群中膳食模式、生活方式、遗传特征与肿瘤等慢性病的关系,是经典的多中心分散式大型队列^[6]。欧洲10国共23个研究中心共同参与,总样本量达52万人。EPIC由国际肿瘤研究协会(International Agency for Research on Cancer, IARC)负责,协调总部设在法国里昂。队列随访通过死因登记、主动随访和死亡记录进行信息收集,每3~4年重复获得调查对象生活方式和疾病的变化^[7]。

分散式大型队列的运行模式通常由大型医疗中心组建多地委员会,每个中心募集和随访本医疗中心覆盖的调查对象。该模式的优势在于可测量较细致的表型,且可以由课题组长通过不同途径改善技术和方法并评价发现的重现性。缺点是需要很昂贵的半永久性研究中心去维持,且标准化问题往往受到质疑。

UK Biobank由英国政府发起,于2006年启动。研究目的是寻找遗传因素、环境因素、生活习惯与重

大疾病的关联,改善各种严重和危及生命的疾病的预防、诊断和治疗,包括癌症、心脏病、中风、糖尿病、关节炎、骨质疏松症、眼科疾病、抑郁症和各种形式的痴呆症。2006—2010年,项目共采集了50万名40~69岁志愿者的血液、唾液、尿液,生活方式(包括营养、生活方式和药物使用等)和血缘数据,并将跟踪记录志愿者之后数十年的医疗健康档案信息,获得全基因组遗传数据、临床测量以及健康记录。UK Biobank实现了样本库保管全自动化、机器人智能化和精准化流程,数据向全球范围免费共享。2016年,UK Biobank宣布建立全球最大内脏器官扫描图像数据库项目,使用核磁共振成像(Magnetic Resonance Imaging, MRI)以及其他心脏、大脑、脂肪、骨骼和血管成像技术获取相关信息^[8, 9]。2018年10月,《Nature》杂志集中报道了UK Biobank的遗传数据,包括8428个个体的全基因组数据和磁共振成像脑部扫描数据,发现了与铁元素运输和储存,阿尔茨海默症、帕金森病等神经退行性疾病,突触可塑性和神经纤维修复以及抑郁症、多发性硬化或中风等有关的遗传相关性和一些重要的相关基因^[10, 11]。

UK Biobank的设计和运行优势在于:(1)设计完善,自2000年前后至2006年全面启动基线调查,研究团队耗时数年进行了周密的规划和设计,从样本量计算到实验室检测,各环节均有详尽分析,且在正式开始前进行了两次预实验;(2)该队列建立于英国国民卫生服务系统基础之上。该系统基本覆盖全体英国国民,有从出生到死亡的详细健康信息,且每个个体有唯一编码。该系统极大提高了UK Biobank的可操作性;(3)UK Biobank采用集中模式以提高工作效率、节约成本。分中心通过专门的信息传送网络和样本运送通道,每天将采集到的信息和样本发回协调中心,由后者统一保管。

美国All of Us Research Program由美国国立卫生研究院(National Institute of Health, NIH)于2016年提出,项目整体计划运行十年,预算高达14.55亿美元,预计建立100万人的研究队列,包括不同种族、年龄和性别的病人和健康人群。项目采集参与者的电子健康记录、行为和生活环境,通过可穿戴设备获取健康报告和信息。每个参与者将被采集35份血液和尿液,共计3500万个生物样本。所有参与者将进行全基因组测序。

该项目在以下方面做了重点规划:(1)为了保

证生物样本免受自然灾害的影响,所有的生物样本存放于 Mayo Clinic 在明尼苏达州的生物库中,其中 20%~25% 的样本存放在佛罗里达州备份生物库;(2) 项目创建了专门的数据研究和中心,收集、整理,并将参与者的所有信息存储到安全的云环境中,收集到的数据仅用于研究目的;(3) 通过设置不同的访问级别以保障数据安全。每位参与者都可以访问自己的健康信息;总体人口统计数据以及风险较低的一般信息向公众开放;更敏感的数据将受到严格的控制和保护,其中一些会提供给合作研究机构的科学家使用。通过端到端的加密手段和保密证书相结合的方式,保证任何数据不能下载到研究者的系统中;(4) 项目构建了大量生物医学领域分析工具,帮助研究人员梳理各类数据,寻找可能产生新发现的任何关联信息^[12-14]。

中国慢性病前瞻性研究(China Kadoorie Biobank,CKB)于 2004 年启动,目标为调查主要慢性病及相关危险因素状况,通过长期随访,探讨环境、个体生活方式、体格和生化指标、遗传等众多因素对复杂慢性病或症状的发生、发展的影响。覆盖中国的农村和城市各 5 个地区。2004—2008 年完成募集和基线调查,共采集 30~79 岁的调查对象 51.2 万人。于 2008 年和 2014 年对约 5% 的队列成员各完成一次重复调查,以研究结局影响因素的长期变化趋势。样本库集中存储了基线调查和两次重复调查的全部样本,达到 58.8 万份,并实现了样本存储管理的全程电子化和所有样本全方位监管。

CKB 具有以下特点:(1) 利用各地常规疾病死亡和发病报告体系,结合社区工作人员主动随访研究结果;(2) 与医保数据库链接,系统获取住院事件、补充常规发病、死亡监测,极大地提高了疾病监测能力;(3) 项目覆盖我国东北、西北、华东、华南和西南具有不同经济发展水平、社会文化背景以及暴露谱和疾病谱的城乡地区,为我国转型期社会人群健康状况的发展和变化、疾病谱的改变及影响因素研究提供了宝贵的人群现场,将成为我国人群复杂疾病机制及防治研究不可或缺的重要基础性资源和本土高质量的病因学证据^[15]。

国际上也在探索建立一些特定领域的队列,如脑磁共振成像相关大型队列。除前面介绍的 UK Biobank(约 1 万条脑磁共振影像数据)之外,当前的大型脑磁共振影像队列还有美国的 Adolescent

Brain Cognitive Development(ABCD) Study(已完成约 10 000 人)、中国的 CHIMAGEN Project(约 7 000 人)、欧洲的 IMAGEN Longitudinal Cohort Study(2 000 人、已完成 2 次脑核磁共振影像跟踪)、美国的 Human Connectome Projects(HCP)(1 200 人,具有 3T 以及 7T 的数据)等。此外,国际多中心的全球合作研究计划 ENIGMA Project 整体拥有大约 5 万人的脑磁共振影像数据。由复旦大学类脑智能科学与技术研究院承担的张江国际脑库队列计划将采集包括脑疾病队列(脑卒中、神经退行性疾病、精神分裂症、儿童孤独症、抑郁症)以及正常人群体队列共 15 000 人,相关采集工作已经在稳步推进之中。

大规模人群队列规范化建设及其共享机制在我国尚处于初级阶段。2017 年 7 月,国际医学期刊编辑委员会(the International Committee of Medical Journal Editors, ICMJE)同时在 *NEJM*、*Lancet*、*JAMA*、*BMJ* 等国际顶级医学期刊上刊登了临床试验数据共享声明,期望共享去标识的研究数据成为常态^[16]。许多研究资助机构或合作团体也已经就促进医学研究数据的共享原则和目标达成共识^[17]。这将促进数据共享的发展,也会推动高影响力、高成本—效益回报型研究的进展。

我国开展队列数据共享工作目前仍面临一些困难和挑战。首先,数据公开的程度不足。现有的队列缺乏足够的信息曝光度,队列研究项目之间的知晓度相对较低。其次,在实现不同来源共享数据分析之前,需要解决队列数据协调的问题,提高数据的标准化程度。第三,因涉及到研究者和研究团队的学术贡献和权益,需要提前解决好数据所有权归属和研究利益共享机制的问题,包括研究成果发表时的署名顺序和规则制定等。另外,还需要考虑数据管理方式、医学伦理等问题。

1.2 以组学为核心的生物标志物挖掘研究

组学(Omics)概念出现于 21 世纪初期,随着生物技术的不断更新与蓬勃发展,如今已演化成为系统化组学研究领域,包括基因组学(Genomics)、转录组学(Transcriptomics)、翻译组学(Translatomics)、蛋白组学(Proteinomics)、代谢组学(Metabolomics)、脑连接组学(Brain Connectomics)和影像组学(Radiomics)等。近年来,多组学技术快速发展并广泛应用于生物标志物的筛选,从 DNA、RNA、蛋白质、代谢物、神经影像、临床数据多方面监测多因素

水平的变化,根据患者存在差异的生物标志物检查结果,为理解患病机制及临床个体化用药治疗等提供参考依据。多组学帮助我们围绕中心法则全面地、实时地了解疾病发生发展,实现从“因”和“果”两个方向探究生物学问题,不同层次间相互验证,深入研究疾病的致病机制,挖掘重要的生物标志物。

多组学研究在精神类疾病领域已经取得一些进展。冯建峰团队基于逾1万人的脑影像组学和基因组学大数据,发现相比正常人群,精神分裂症高风险人群中锌转运体 SLC39A8 基因突变对壳核灰质体积的强相关显著降低,提示精神分裂症发病机制可能与神经发育过程中的离子转运和免疫反应有关^[18]。加州理工学院在关于孤独症的最新研究中,基于微生物组学及代谢组学分析结果证实了患者来源的肠道菌群移植会导致小鼠出现孤独症核心行为学表型以及大脑中孤独症相关基因的可剪接异常^[19]。在神经类疾病研究领域,Jove 等人通过对脑卒中患者(发病24小时内)的血浆样本进行代谢组学分析,发现结合模式识别法可有效提取生物标志物预测卒中复发^[20]。Aerts 等人基于肿瘤患者的高维影像组学数据,通过机器学习方法提取出可准确预测独立样本肺及头颈部肿瘤特征的影像学标志物,可用于辅助临床肿瘤的精准诊断^[21]。此外,Menden 等人基于细胞系的癌症基因组特征和化学组学特性,开发了一种机器学习模型来预测肿瘤细胞对药物治疗的反应,通过建立一个前馈感知神经网络模型和随机森林回归模型,显著提升了预测准确率^[22]。以药物化学分子结构、分子靶标、基因表达为特征,Napolitano 等人使用高斯核多分类 SVM,可准确预测药物和疾病类别之间的关系,有助于为个体化用药提供参考^[23]。

然而,目前多组学研究技术还有许多困难要克服。首先,数据的分析方法处于发展阶段,还未形成成熟的体系,亟待开发出通用的数据整合和分析方法,以充分利用已产生的多组学数据获得更多的信息。其次,筛选出的潜在生物标志物的可信度有待提高,亟需建立基于组学技术筛选生物标志物与经典标志物相结合的全维度生物信息库。再次,全维度数据具有数量多、噪声高、异构程度高的特点,需要建立相应的专业化计算分析体系。

1.3 面向疾病早期预警和精准诊疗的人工智能大数据分析

随着深度学习和大数据技术的不断革新,智能

诊疗在人工智能理论的支撑已成为人工智能和大数据技术的集大成者,主要体现在以下技术:(1) 计算机视觉技术,可实现图像的分类、分割和识别任务,因此可以实现医学影像数据的自动化、快速化处理,完成病变部位的自动标识和病变类型的分类等;(2) 自然语言处理技术,可在问诊环节的患者描述中提取临床需要的信息和指标,或与计算机视觉技术相结合,将纸质材料扫描结果转换为文字并获取关键数据,从而实现电子档案的快速建立和完善;(3) 知识图谱技术,可在临床数据收集过程中指出需要完善的信息和检查,根据知识库数据获取相关文献,指定可行的治疗方案;(4) 机器人技术,机器人技术与自然语言处理、计算机视觉等技术相结合,必会带来大量的智能诊疗终端,例如导诊机器人、手术机器人以及穿戴监测设备等。

基于人工智能大数据分析的精准诊疗将可以显著降低临床治疗费用。疾病早期预警和精准诊疗是医疗人工智能研究的重要目的,特别是针对罕见疾病和慢性疾病。智能诊疗还将推动建立新的诊疗标准。针对慢性疾病如糖尿病,已有研究通过人工智能中的聚类方法(中国人群2316例和美国人群815例),采用年龄、BMI、血糖水平、胰岛素敏感性指数及胰岛细胞功能指数等5个维度将两类人群分别分成了4个亚型^[24]。在神经疾病方面,人工智能算法也发挥着重要作用。基于行为、影像的大数据分析将有可能为神经退行性疾病、脑卒中等疾病的诊断与康复提供量化、个体化策略,通过量化评估结果对病情进行更合理有效的解读,为临床治疗策略的调整控制和病人的居家管理提供有价值的参考^[25]。在精神类疾病研究领域,冯建峰团队提出了一种整体分析大脑各个脑区之间功能性连接的方法,成功发现抑郁症患者难以控制负面情绪可能与“憎恨网络”的消失有关,并精准定位了抑郁症影响异常功能脑区^[26]。目前,该项成果正在上海市精神卫生中心开展大样本临床试验。此外,医疗人工智能的发展也为肿瘤的早期预警与精准诊疗带来前所未有的契机,例如通过在MRI图像中对肿瘤进行自动化智能分割,医生可以精确定位肿瘤的位置、快速客观地测量肿瘤体积,有助于实现个体化治疗计划和康复策略^[27, 28]。

1.4 智能诊疗临床转化应用

智能诊疗的临床转化应用场景和应用实践,可以围绕患者管理、医学影像、临床决策辅助系统、疾

病预测等方面进行场景分析和设计,全面提升医院临床诊疗能级。在国际上,2017年初,美国斯坦福大学的研究人员在 *Nature* 杂志上发表了一篇文章,利用人工智能技术对 129 450 张临床皮肤病变图片的深度学习,在角化细胞癌、恶性黑色素瘤的疾病诊断上已与 21 名认证皮肤病科学家不相伯仲^[29]。2017年,另一项研究发表在 *JAMA* 杂志上,报道利用 AI 影像识别技术对糖尿病视网膜病变、青光眼、老年性黄斑退化等眼科疾病的诊断, AI 诊断准确率可达 90% 以上,可媲美 5 年以上临床经验并经过专业培训的眼科医生^[30]。2019年 *Nature Medicine* 报道一项美国梅奥中心的研究,利用 AI 辅助 ECG-TTE 检查,帮助医生准确地筛查出无症状心功能不全患者^[31]。研究利用 9 万人数据库对 AI 进行训练、内部验证及外部测试,筛查准确率可达 93%。

目前,国外有很多基于人工智能的临床决策支持系统逐步走向实用化,而国内尚处于探索阶段。天坛医院正在开发的临床决策支持“BioMind 天医智”系统将有针对性地应用于多种脑疾病患者管理、医疗决策以及教学和科研,为临床医生提供丰富的资源和工具,一定程度上缓解医生的工作压力。2018年,复旦大学类脑智能科学与技术研究院研发了 F-STROKE 全自动影像处理软件,可实现对脑中风的智能诊断,在达到国际权威脑中风诊断软件 RAPID 性能的同时具备自主的知识产权。另外,2019年 *BioMedical Engineering* 杂志最新发表了一项来自上海新华医院团队的研究,利用人工智能辅助肺结节诊断准确率可达 89% 以上^[32]。目前,国内越来越多的人工智能辅助诊疗产品已在临床转化方面开始探索,协助医生更好地提高医疗质量,具有极大的发展前景。

需要注意的是,缺乏以病人为核心的算法越来越成为当前人工智能进入医学领域的主要限制因素。一个有缺陷的算法可能对病人造成重大伤害,导致医疗事故^[33]。另一个很重要的问题是,目前正在广泛使用的人工智能医疗工具,其核心算法大多来自在此领域有发展优势的国家,其数据集中群体标本比较单一,许多算法中都存在固有的偏见,不一定适用于我国人群。因此,开发适用我国人群的算法,建立一个真正具有中国人群代表性的人口截面,还需要做大量工作。

目前,我国智能诊疗的临床转化应用面临着五大问题:(1) 数据数量少,标注质量参差不齐,需要

花大量精力进行修正和补充;(2) 全国医院影像科仪器设备厂家规格不统一,成像质量差异大,为影像的智能判读带来困难;(3) 人工智能相关人员对医学知识的认识不够深入,需要和相关医师进行大量的沟通,需要长期的深入合作;(4) 人工智能领域算法颇多,需要多年的相关经验才能更加深入的理解不同算法在不同任务中的优势,也需要耗费大量的时间和精力来改进;(5) 人工智能某些环节的可解释性较差,医生对智能诊疗的可接受度提出了挑战。

2 全维度数据和智能诊疗的机遇和挑战

人工智能在医疗领域的应用发展方向集中在医学影像处理、辅助诊断、健康管理、药物研发和疾病预测等方面。目前,我国在卫生信息与医疗健康大数据系统建设方面取得了显著成效,制定了国家健康医疗大数据标准,多地积极推进国家健康医疗大数据中心建设。科技部及地方政府积极推动多种疾病队列研究,在上海市重大科技专项支持下,全维度脑数据库建设工作及国际人类表型组计划已经启动。这些工作推动了以基因、表型、影像等全维度数据为基础的精准医疗领域进步,在疾病诊疗、疾病管理以及疾病机制研究等方面呈现出蓬勃发展的趋势,但同时也在非完备性大数据分析、模糊标定数据集建模、辅助诊疗决策、标准化队列建设等方面对医疗大数据人工智能理论、方法及技术提出了新的挑战。加强发展全维度数据分析的理论和技术,将智能诊疗技术融入医疗改革,将为我国解决医疗资源紧张、提升医疗卫生服务能力带来新契机。智能诊疗将能从如下四个方面助力我国医疗卫生事业在普惠精准诊疗方向的发展。

第一、能够缓解医疗人力资源紧张。智能诊疗除了可以给临床医生提供决策支持外,还可能减少后勤、行政工作的劳动力,如计费、手术室和诊所预约的调度以及人员配备^[35]。就每位患者而言,如果个人的生物学、解剖学、生理学、环境、社会经济和行为数据实现全维度采集,理论上每个人可以受益于此,即从他们的相关数据中得到针对各种医疗情况的最佳预防方法、治疗方法和结果。

第二、能够推动慢性病预防。慢性病存在筛查准确度低、针对性干预难度大、健康管理工具缺失等医疗难题。目前,成人糖尿病的知晓率仅 30.1%。上海瑞金医院推出一款基于人工智能实现的糖尿病

及并发症管理产品,在公众号中输入个人的相关信息,包括性别、体重、空腹血糖等,可以预测出个人近3年患糖尿病的风险系数。当指标超过一定比例,还会建议个人尽快去医院就诊,为糖尿病的预防提供极大助力。

第三、提高疾病筛查效率。早诊早治是提高癌症治愈率的关键。2018年12月,中山大学肿瘤防治中心牵头开展上消化道肿瘤人工智能诊疗决策系统的研发及推广应用项目,根据该系统试用初期数据分析,临床试用中的恶性肿瘤识别准确率已达到95%以上。人工智能在胃癌、肺癌、乳腺癌、肝癌等早诊早治方面均有广泛应用前景。

第四、助力公共卫生科学决策。业内专家认为,人工智能通过海量的数据模拟出医疗流程、医疗诊断、医疗建议和治疗方案,将使公共卫生政策的制定更为科学。

从理论层面来看,智能医疗领域目前研究仍然是以机器学习、深度学习为基础,主要的理论探讨将集中在神经网络、数据科学、自然语言处理、以及物联网等方面,未来需要解决全维度大数据与人工智能算法相融合的挑战。

从疾病层面来看,智能诊疗目前重点关注肿瘤,如前列腺癌、乳腺癌,同时糖尿病早诊和智能干预也受到广泛关注。精神疾病、脑损伤、退行性疾病涉及人群广,社会负担呈指数增长,将是未来智能医疗的研究重点。

从实践层面来看,智能诊疗领域将开展更多的临床试验,同时如何制定符合远程医疗、电子医疗特点的医疗卫生政策也是一个重要的研究领域。

3 未来全维度数据与智能诊疗多学科交叉发展目标及资助重点

建议基于我国现有及正在构建的优质队列数据,依托国内数学、生物、医学、信息等优势力量、组成联合攻关团队,通过3~5年时间,在以下几个方面取得突破性成果:

一是发展适用于不同尺度数据的非线性关联算法,利用数据时空特征挖掘与疾病复发风险相关的多组学生物标志物,研究其生物机制,寻找治疗干预新靶点;

二是利用新的生物标志物,发展疾病尤其是我国现阶段重大慢性疾病的动态演变预测模型,发展智能诊疗,以期实现疾病风险性精准分级与病人分

类管理;

三是建立药物基因组学,实现从单一蛋白质靶点到多靶点治疗,通过分析多靶点有向网络,明确药物治疗机理,实现多靶点联合个体化治疗的精准、智能医疗;

四是应用智能算法量化社会、经济、家庭、环境等多因素对治疗预后的影响,建立符合医疗伦理的定量化评估体系,为制定适合我国国情的智能诊疗卫生政策提供理论和实践支撑。

通过5~8年时间,联合多家优势单位,建立联合研究基地,构建有疾病针对性的人群队列,围绕“实现基于大数据的慢性病发病机理和干预策略研究”,推动基础到应用的原创性研究突破:

一是基于组学大数据的新的分子分型研究;

二是高危人群的防控新策略研究;

三是疾病的发病机理及新型药物靶点研究。

4 结 语

在重大民生需求和国际前沿竞争环境下,我国亟需整合已有的队列优势研究基础,面向重大临床需求,基于“高质量、高数量”多组学临床数据,融合不同尺度、不同模态数据、发展面向因果决策、非线性关联的人工智能算法,为疾病发生与复发风险分级、药物干预机制研究和个体化治疗提供新的解决方案。

在信息技术和互联网的助推下,人工智能在医疗领域的大幕正徐徐拉开。随着人工智能技术的不断进步,其所适用的医疗应用场景将会越来越多,将为各种疾病的预防、诊断和治疗做出更大的贡献。虽然仍有问题需要解决,但是人工智能在医疗领域的应用带来了诊疗模式、数据处理方式、前瞻性健康管理等诸多方面的变革。人工智能让医疗产业链全面升级,并让医疗行业走向更高效率与更高层次,我国医疗智能化时代即将开启,将在国际上引领智能诊疗的发展。

参 考 文 献

- [1] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 2019, 25(1): 44-56.
- [2] He JX, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 2019, 25(1): 30-36.

- [3] 陈梅, 吕晓娟, 张麟, 等. 人工智能助力医疗的机遇与挑战. *中国数字医学*, 2018, 13(1): 16—18.
- [4] 熊玮仪, 吕筠, 郭彧, 等. 大型前瞻性队列研究实施现况及其特点. *中华流行病学杂志*, 2014, 35(1): 93—96.
- [5] Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nature Reviews Genetics*, 2006, 7(10): 812—820.
- [6] International Agency for Research on Cancer. About EPIC. <http://epic.iarc.fr/about/about.php>.
- [7] Bingham S, Riboli E. Diet and cancer—the European prospective investigation into cancer and nutrition. *Nature Reviews Cancer*, 2004, 4(3): 206—215.
- [8] Collins R. What makes UK Biobank special?. *The Lancet*, 2012, 379(9822): 1173—1174.
- [9] Collins R. UK biobank: protocol for a large-scale prospective epidemiological resource. UK biobank coordinating Centre, 2007.
- [10] Elliott LT, Sharp K, Alfaro-Almagro F, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*, 2018, 562(7726): 210—216.
- [11] Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 2018, 562(7726): 203—209.
- [12] National Institutes of Health. NIH announces national enrollment date for All of Us Research Program to advance precision medicine. (2018-05-01)/[2020-04-08]. <https://www.nih.gov/news-events/news-releases/nih-announces-national-enrollment-date-all-us-research-program-advance-precision-medicine>.
- [13] Facher L. NIH opens nationwide enrollment for huge precision medicine initiative. (2018-05-01)/[2020-04-08]. <https://www.statnews.com/2018/05/01/nih-precision-medicine-all-of-us-enrollment/>.
- [14] Genomeweb. All of Us Research Program Set to Open National Enrollment. (2018-05-01)/[2020-04-08]. <https://www.genomeweb.com/genetic-research/all-us-research-program-set-open-national-enrollment>.
- [15] 李立明, 吕筠. 大型前瞻性人群队列研究进展. *中华流行病学杂志*, 2015, 36(11): 1187—1189.
- [16] Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials. *British Medical Journal*, 2017, 357: j2372.
- [17] Walport M, Brest P. Sharing research data to improve public health. *The Lancet*, 2011, 377(9765): 537—539.
- [18] Luo Q, Chen Q, Wang W, et al. Association of a schizophrenia-risk nonsynonymous variant with putamen volume in adolescents: a voxelwise and genome-wide association study. *JAMA Psychiatry*, 2019, 76(4): 435—445.
- [19] Sharon G, Cruz NJ, Kang DW, et al. Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice. *Cell*, 2019, 177(6): 1600—1618. e17.
- [20] Jové M, Mauri-Capdevila G, Suárez I, et al. Metabolomics predicts stroke recurrence after transient ischemic attack. *Neurology*, 2015, 84(1): 36—45.
- [21] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 2014, 5(1): 4006.
- [22] Menden MP, Iorio F, Garnett M, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One*, 2013, 8(4): e61318.
- [23] Napolitano F, Zhao Y, Moreira VM, et al. Drug repositioning: a machine-learning approach through data integration. *Journal of Cheminformatics*, 2013, 5(1): 30.
- [24] Zou XT, Zhou XH, Zhu ZX, et al. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *The Lancet Diabetes & Endocrinology*, 2019, 7(1): 9—11.
- [25] Fereshtehnejad SM, Zeighami Y, Dagher A, et al. Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression. *Brain*, 2017, 140(7): 1959—1976.
- [26] Tao H, Guo S, Ge T, et al. Depression uncouples brain hate circuit. *Molecular Psychiatry*, 2013, 18(1): 101—111.
- [27] Konstantinos K, Enzo F, Sarah P, et al. DeepMedic for brain tumor segmentation. //International Workshop on Brain Lesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Switzerland: Springer, 2016, pp 138—149.
- [28] Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. //International MICCAI Brain Lesion Workshop. Switzerland: Springer, 2018, pp 311—320.
- [29] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, 542(7639): 115—118.

- [30] Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*, 2017, 318(22): 2211—2223.
- [31] Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*, 2019, 25(1): 70—74.
- [32] Ye WJ, Gu W, Guo XJ, et al. Detection of pulmonary ground-glass opacity based on deep learning computer artificial intelligence. *Biomedical Engineering Online*, 2019, 18(1): 6.
- [33] Debsford F. IBM-Watson delivered unsafe and inaccurate cancer recommendations. (2019-06-06)/[2020-04-08]. <https://www.massdevice.com/report-ibm-watson-delivered-unsafe-and-inaccurate-cancer-recommendations>.
- [34] 孔鸣,何前锋,李兰娟. 人工智能辅助诊疗发展现状与战略研究. *中国工程科学*, 2018, 20(2): 86—91.

Full-dimension Data and Intelligence Medical Care: Prospects and Challenges

Feng Jianfeng¹ Wang Yongjun² Xu Tao³ Wang Shouyan¹ Cao Miao¹
 Xie Xiaohua¹ Meng Xia² Li Ye³ Sun Hongzan⁴ Cao Heqi^{4*}

1. *Institute for Science and Technology in Brain-inspired Intelligence, Fudan University, Shanghai 200433*

2. *Tiantan Hospital, Capital Medical University, Beijing 100050*

3. *University of Chinese Academy of Sciences, Beijing 100049*

4. *Department of Health Sciences, National Nature Science Foundation of China, Beijing 100085*

Abstract Abstract Basing on the outputs of the 232nd Shuangqing Forum of National Natural Science Foundation of China, this paper analyzed the prospective developments of full-dimension data and intelligence medical care in China. Recent research progress was summarized from the three main aspects of MegaCohort construction, biomarker discovering and big data analysis by artificial intelligence. Concerning on these fields, future opportunities and challenges were discussed. Meanwhile, the key research funding goals and directions in the coming several years were also investigated.

Keywords full-dimension; big data; intelligent medical care; interdisciplinary; clinical application

(责任编辑 姜钧译 仲斌演)

* Corresponding Author, Email: caohq@nsfc.gov.cn