

• 行业研究 •

大数据分析在智慧医疗辅助诊断中的应用与发展趋势*

和海妍¹, 刘伟¹, 王云霞¹, 府伟灵¹, 张阳^{2△}

(1. 陆军军医大学第一附属医院检验科, 重庆 400038; 2. 重庆市人民医院检验科, 重庆 400013)

摘要: 智慧医疗以临床大数据为基础, 以物联网、云计算、人工智能等技术为手段, 是一种以患者数据为中心的医疗服务模式。该文介绍了当前智慧医疗和大数据概况, 分析总结了当前大数据分析在智慧医疗辅助诊断中的应用, 包括在医学检验、医学图像分析、临床决策支持系统以及远程诊疗中的应用, 并对大数据分析在智慧医疗诊断中的挑战及未来发展趋势进行了展望。

关键词: 大数据; 智慧医疗; 检验医学; 图像分析; 临床决策支持

DOI: 10. 3969/j. issn. 1673-4130. 2019. 13. 001

中图法分类号: R0-5, TP433

文章编号: 1673-4130(2019)13-1537-04

文献标识码: A

Application and development of the big data analysis in the intelligent medical diagnosis*

HE Haiyan¹, LIU Wei¹, WANG Yunxia¹, FU Weiling¹, ZHANG Yang^{2△}

(1. Department of Clinical Laboratory, First Affiliated Hospital, Army Medical University, Chongqing 400038, China; 2. Department of Clinical Laboratory, Chongqing People's Hospital, Chongqing 400013, China)

Abstract: The intelligent medicine is a patient data-centric medical service model, which is based on big data and uses internet of things, cloud computing, artificial intelligence and other technologies. This paper introduces the present situation of big data and intelligent medicine, analyses and summarizes the application of big data analysis in assistant diagnosis of intelligent medicine, including medical examination, medical image analysis, clinical decision support system and remote diagnosis and treatment. The challenges and future of the big data analysis in diagnosis of intelligent medicine are also prospected.

Key words: big data; intelligent medicine; laboratory medicine; image analysis; clinical decision support

物联网、云计算、人工智能等新兴信息技术的发展, 正在深刻改变着传统的医疗与健康服务模式。大数据已给智慧医疗带来无限的发展空间, 智能决策正在不断提高医疗的精准度。因此, 利用信息化手段提高医疗行业的服务范围和服务水平已迫在眉睫。目前, 以云计算和大数据为支撑的智慧医疗在疾病的早期筛查和诊疗方面发挥着日益重要的作用。实现智慧医疗的关键是对医疗大数据进行挖掘和分析, 这一模式正在不断地改变着传统的诊疗模式。

1 智慧医疗和大数据概况

智慧医疗是一种以患者数据为中心的医疗服务模式^[1]。智慧医疗利用大数据存储与处理平台, 广泛采集和深度利用数据, 应用数据挖掘关键生理特征, 对医疗历史数据进行建模与分析, 达到发现早期疾病和预测健康风险的目的, 同时, 为医务人员的诊疗提供参考。因此, 实现智慧医疗的关键是医疗大数据,

基于大数据的智慧医疗正在不断改变传统的医疗模式。大数据由海量的数据集组成, 大数据本身无价值, 大数据产生价值的关键是对其进行处理^[2], 大数据处理技术能够帮助人类更直观地了解数据, 其作为一种新型产业, 在各行业的应用甚多, 尤其在医疗领域的应用与发展已成为行业热点。

2 医疗大数据现状和大数据平台体系架构

2.1 医疗大数据现状 医疗行业每年会产生大量的医疗数据, 普通的医疗机构每年会产生 1~20 TB 相关数据, 大型医院每年能产生 300 TB 至 1 PB 的医疗数据。这些医疗大数据主要包括 6 类: (1) 个人日常生理数据; (2) 基因组学、蛋白质组学数据; (3) 医院电子病例数据、电子健康档案数据; (4) 大型队列研究、医疗科研数据; (5) 网络健康数据; (6) 公共卫生数据^[3]。当前医疗大数据的应用还处于初级阶段, 很多潜在的利用价值还需不断被揭示。

* 基金项目: 国家重点基础研究发展计划(“973”计划)项目(2015CB755402); 国家重点研发计划“精准医学研究”项目(2017YFC0909902); 西南医院智慧医疗重点项目(SWH2017ZDCX4210)。

作者简介: 和海妍, 女, 实验员, 主要从事大数据和智慧医疗研究。 △ 通信作者, E-mail: millen001@163.com。

本文引用格式: 和海妍, 刘伟, 王云霞, 等. 大数据分析在智慧医疗辅助诊断中的应用与发展趋势[J]. 国际检验医学杂志, 2019, 40(13): 1537-1540.

2.2 医疗大数据平台体系架构 医疗大数据与数据挖掘处理技术的结合,能够帮助人们从存储的大体量、高复杂的医疗数据中提取有价值的信息,加速医学成果转化。下面对医疗大数据平台的体系架构进行简要描述,见图 1,此体系架构主要包括:数据采集层、数据存储层、业务处理层、数据分析和应用层^[4]。(1)数据采集层:利用多个数据库来接收来自客户端的数据,并进行简单编码和封装处理。(2)数据存储层:将来自前端的数据导入到一个集中的大型分布式数据库中,如 Hadoop 集群服务、医学知识库中。(3)业务处理层:利用大数据的挖掘处理技术对数据进行预处理,以满足大多数常见的分析需求。(4)数据分析和应用层:在现有数据上进行基于各种算法的计算分析,达到疾病的筛查和智能分析、大数据业务分析引擎、医学知识库搜索引擎等^[4]。

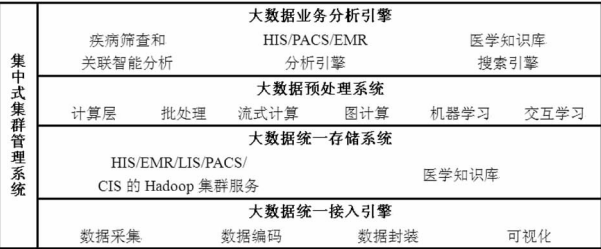


图 1 医疗大数据平台体系架构

3 大数据在智慧医疗辅助诊断中的具体应用

从实践层面看,智慧医疗正成为我国医疗信息化的重要领域和重点发展方向,基于大数据的智慧医疗解决方案在医学检验领域、医学图像分析、临床决策支持系统及远程医疗等方面具有潜在应用价值。

3.1 在医学检验领域中的应用 目前,临床检验通常只用了有限的检验指标数据对疾病进行诊断,尽管很多临床检验数据并没有明显的超过现有指标的正常阈值,但其可能在某个区间内存在上升或下降的趋势。通过大数据分析、云计算,可实现检验数据与疾病症状的逻辑关系分析,从这些容易忽视的数据中挖掘出与疾病相关的新标志物。此外,还可将某些常规的指标与多种疾病进行数据对比和计算分析,进而发现其数据层面的对应关系,再通过恰当的病理学验证,发现新的疾病预警指标。

基于检验大数据分析有望发现新的疾病预警指标。近年来,根据大数据分析研究发现,红细胞体积分布宽度(RDW)在很多非血液系统疾病中具有重要的临床价值,如国外专家在 6 年的跟踪调查中,对 21 939 起重大心血管事件和 4 287 例患心血管疾病死亡患者的大数据分析发现,在普通个体中,RDW 越高,全因死亡风险和心血管疾病发病风险也越高^[5];在恶性肿瘤和心血管疾病患者中,RDW 越高则预后越差;在肝病或自身免疫性疾病患者中,RDW 增高提示疾病病情较重或活动度较高^[6-7]。通过对多种常见指标与疾病的检验大数据高通量无差别对比分析,将对疾病预警体系的完善有极大的促进作用。

生物标志物的检测是筛查、诊断某些疾病的重要手段,检验大数据分析为众多疾病生物标志物的发现提供了新的契机。蛋白质组学技术是癌症生物标志物和治疗靶点研究的一个重要工具,肺癌临床研究表明,通过运用蛋白质组学技术,发现了大量与肺癌的发病机制、转移、早期诊断及治疗相关的蛋白,如 S100A14、STC2、ENO1、transgelin-2 等,这些肿瘤标志物的研究为肺癌的临床研究提供了可能性^[8-13]。代谢组学分析在肿瘤标志物筛选中同样起到关键作用。通过对大量肿瘤患者临床标本血清、尿液、胆汁、脑脊液等体液的分析,发现一些异常表达的潜在生物标志物,其诊断效率较常规的检测项目显著提高,并且部分标志物与肿瘤的分期、分化等预后因素密切相关^[14]。

然而,单纯研究某一生物标志物指标的变化,其灵敏度或特异度往往偏低,很难满足临床相关疾病诊断及预后的期望。因此,提倡同时检测多种标志物来达到提高灵敏度和特异度的目的,而基于多组学检验大数据分析能够更加完善地检测多种临床疾病生物标志物。通过整合各个组学筛选的生物标志物,进行数据分析,将大大提升对疾病的预警及诊断作用,以达到疾病早期诊断及预后的目标。

3.2 在医学图像分析中的应用 医学图像在疾病诊断中起很大的辅助作用。常见的医学图像有电子计算机断层扫描、核磁共振等。据统计,医院超过 90% 的医疗数据来自于医学图像,目前影像学检查需要医生结合专业知识与工作经验给出诊断意见,读片缺乏量化标准,工作量大,易产生误判。而基于医疗大数据、医学图像和人工智能相结合的智能阅片系统、图像分析系统已在国内外展开了广泛的研究。下面通过一些例子说明医学图像、大数据和人工智能相结合,在疾病临床检验和辅助诊断方面的进展。(1)染色体智能分析与识别。智能化的染色体显微数字图像扫描系统已在染色体的分析与识别中进行应用。该系统对海量的染色体涂片进行扫描、拍摄,通过图像采集模块采集一定数量的染色体核型图,进行数字处理和优化后,再通过染色体核型分析系统进行分析。此智能化系统能对人体 46 条染色体图像进行分割,将相同标号的染色体进行配对,可识别染色体数量的变化、畸变,对染色体结构上的缺失、易位、倒位、插入、重复等常见异常也可初步识别判断,其分析的准确性可达 50%~70%^[15]。(2)细胞智能分析与检测。骨髓细胞学检查是通过对骨髓中各系列细胞数量、比例的变化及形态学特征的分析来了解骨髓造血功能和病理改变的一项重要检查,在血液病等疾病的诊断、分期和临床疗效评估中有着不可替代的作用,目前在临床中已广泛运用。然而,由于生理状况、地区差异,正常骨髓图象各系列、同系列各阶段细胞比值变异范围较大,加之不同级别医院及医学检验从业人员专业知识参差不齐,阅片受主观因素的影响较

大,因此,骨髓图文报告结果难免存在偏倚。为了解决上述问题,为检验医师提供有价值的参考信息,有学者以大数据为基础,利用数据挖掘技术,通过回顾性研究,对既往明确诊断血液系统疾病及健康者大量的骨髓细胞学图文资料进行深度学习,骨髓细胞图文智能分析系统能够精准识别骨髓细胞的种类、形态、大小,通过对各系列细胞的计数、核形、核浆比、染色质形态、细胞质及增生程度等的分析,进行疾病辅助诊断。同时,该智能系统通过对大量骨髓细胞图文数据的分析,发现骨髓各系列细胞计数、比例与年龄、海拔变化之间的内在联系,从而为不同年龄、地区的患者进行骨髓细胞学检查置信区间的设置提供一定的参考^[16],为真正实现了精准医疗奠定基础。

目前,国内研发的血液细胞形态分析与检测系统已运用于医学检验流水线上,使血常规计数结果与形态学检测成为一体。此仪器采用数字图像分析技术,通过对大型细胞数据库的细胞形态进行学习,建立强大的细胞图像数据库和算法支持系统,系统分析样本时会将会所提取的细胞特征信息转化为数字信号,然后通过人工智能计算和神经网络系统进行形态学分析。该仪器可扫描分析定位白细胞,对白细胞进行形态分类及异常有核细胞识别与筛选;根据红细胞的直径、颜色、色素含量、对比度等特征信息,将红细胞划分为大小异常、颜色异常、形态异常 3 大类共 10 种形态;同时该仪器可进行外周血的异常细胞初筛、原始幼稚细胞初筛、有核红细胞识别;还可以完成对脑脊液、浆膜腔积液、肺泡灌洗液等样本的分析^[17]。

3.3 在临床决策支持系统中的应用 传统诊疗依赖于医务人员的专业知识,易出现诊断错误。而大数据的出现给临床决策提供了更准确的信息,开创了疾病诊断的新领域。运用医疗大数据的临床决策支持系统已在部分医院进行应用。该技术的运用将减少医生专业局限性导致的误诊,提高疾病诊断效率,节省医疗成本,让患者获得更优质的医疗服务,同时,在患者疾病的术前诊断、疗效评估、预后预测等方面提供指导。(1)术前临床治疗方案的选择。医疗大数据的一个重要发展趋势就是指导临床治疗方案的选择,下面用一个实例来说明综合应用不同类型医疗大数据的必要性。胶质母细胞瘤患者术前是否进行化疗的预警流程图,见图 2,该系统通过对海量的胶质母细胞瘤患者的临床数据、分子数据、医疗图像、治疗计划等进行储存记忆、深度学习和反复训练来建立预警模型,模型借助于大量的数学分析方法对胶质母细胞瘤患者的历史医疗数据进行分类、建立关联以及对各类关系进行分析,提取数据的特征和属性。术前将胶质母细胞瘤患者的病例信息、检验结果、以往医疗数据上传到预警诊断模型中,预警模型读取数据后开始搜索以往储存的相似医疗数据,进行对比分析,给出预测:如果预测肿瘤因化疗而萎缩,医生则对患者进行术前化疗,同时将化疗后的效果反馈给预警模型,使

模型进行不断的调整优化;如果预测肿瘤不因化疗而萎缩,则患者将会直接进行手术而不必忍受术前化疗的风险和痛苦^[18-19]。(2)术前临床治疗决策的制订。术前结直肠癌淋巴结阴性或阳性难以确认,淋巴结转移难以诊断,虽然一些组织病理学发现,如淋巴浸润和肿瘤分化,是已知的淋巴结转移的预测因子,但它们仅在术后可用。医生常规诊疗会对淋巴结进行清扫处理,因为通过 CT 片只能看到结直肠癌淋巴结的机构信息,其反映的信息很小,只能得到百分之二三十的概率预测结直肠癌淋巴结为阳性,切掉以后 70% 是阴性。淋巴结转移术前知识可为辅助治疗和手术切除的充分性提供有价值的信息,从而有助于治疗前的决策制订。中科院运用人工智能的方法,根据 500 余例临床病理和影像数据完整的结直肠癌患者数据,采用组学方法选择关键影像特征结合临床病理信息建模并进行分析,加之临床检验结果做辅助,提高了前端数据预测结直肠癌淋巴结阳性的准确率^[20-21],为临床治疗术前决策提供了支持。

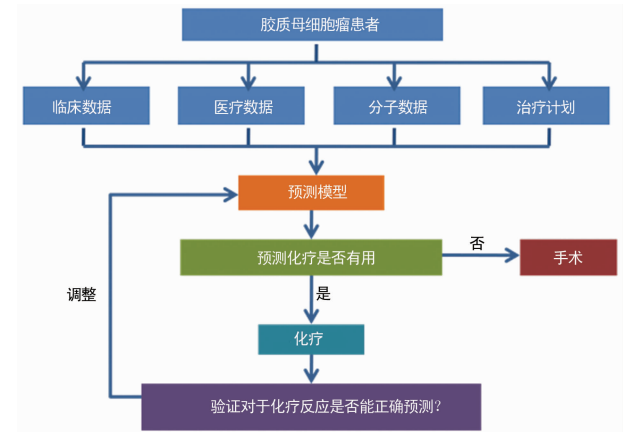


图 2 胶质母细胞瘤患者术前是否化疗预警流程图

3.4 在远程诊疗中的应用 基于大数据的智慧医疗可构建互联互通、服务便捷、反馈即时和质量可控的远程监护平台。该平台可克服地域及时空的限制,将各医院各科室的医疗数据进行整合分析,构建多方交互、智能共享的医疗大数据服务中心。各医疗设备、医院传感器、个人电子终端等采集到医疗数据时,可通过互联网与医疗大数据服务中心进行连接,从而实现远程诊疗目的。

以糖尿病、高血压、心血管疾病等“慢性疾病监护系统”为例,患者佩戴不受时空限制,高精度采集器自动采集监测到的心率、舒张压、收缩压、平均血压、呼吸率、血氧饱和度等生命体征数据,由智能手机来接收,并将监测数据实时传送给医疗大数据服务中心,数据中心再将分析出的异常报告发送给医生,实现生命体征大数据、患者及医疗机构的完整衔接。最后,医生通过手机对患者进行远程医疗指导和相关医疗建议^[22]。

4 大数据面临的挑战

国内医疗行业的大数据应用取得了一些进展,但

也存在典型的问题,严重制约了数据的利用价值。医疗大数据正面临着以下挑战:(1)数据开放性问题。即能否将医院内大量分散的医疗信息进行拆分和整合,并赋予这些数据以逻辑关系。(2)数据应用规则的合理性和可行性问题。数据本身没有价值,主要看最终的应用场景。(3)数据存储方案问题。非结构化医疗大数据的存储和处理更为复杂,但目前还没有综合、经济且安全的大数据存储解决方案。(4)数据共享问题。大数据研究主要依赖多个实验室或数据平台的资源共享,海量的数据目前已超出现有计算机和网络共享程序的能力范围。(5)数据隐私保护问题。面对未来系统流行病学研究将涉及大量人群信息暴露,如何确保数据的隐私安全至关重要^[23-24]。

5 大数据驱动的智慧医疗前景展望

基于大数据的智慧医疗已在医学检验、医学图像分析、临床诊断等领域发挥了巨大优势。大数据时代的到来为医学检验、临床诊断等提供了新的方法,推动了医疗行业的发展。但目前医疗行业对大数据的应用范围还比较局限,智慧医疗在我国发展尚处于初期阶段,还面临着很多挑战。为实现大数据的价值,医疗行业需要提高数据的标准化和互用性,促进信息的共享,建立有效的数据管理方式,改进分析技术和方法,培养数据分析专业人才等。总之,要解决这些问题仍有很长的路要走。未来,期望基于大数据的智慧医疗将会在更宽广的范围和更深入的水平上对人类健康水平的提高做出更大的贡献,研究者应该就这些具体的问题继续研究、探索,以期能在我国全面实现智慧医疗。

参考文献

- [1] 倪明选,张黔,谭浩宇,等.智慧医疗——从物联网到云计算[J].中国科学(信息科学),2013,43(4):515-528.
- [2] TOLAN N V, PARNAS M L, BAUDHUIN L M, et al. "Big data" in laboratory medicine[J]. Clin Chem, 2015, 61(12):1433-1440.
- [3] LUO X, LIU Y J. Medical big data research situation and its clinical application[J]. J Med Informat, 2015, 36(5):10-14.
- [4] 张伟.医疗大数据平台数据高并发方案设计与关键技术分析[J].信息技术与网络安全,2018,37(4):18-22.
- [5] ARBEL Y, RAZ R, WEITZMAN D, et al. Red blood cell distribution width and the risk of cardiovascular morbidity and all-cause mortality: a population-based study[J]. Eur Heart J, 2013, 34(1):300-307.
- [6] 刘伟,张阳,黄姣祺,等.检验组学在精准医疗中的应用价值[J].国际检验医学杂志,2018,39(9):1025-1028.
- [7] RIED J, KAIDER A, REITTER E M, et al. Association of mean platelet volume with risk of venous thromboembolism and mortality in patients with cancer. Results from the Vienna Cancer and Thrombosis Study (CATS)[J]. Thromb Haemost, 2014, 111(4):670-678.
- [8] ZHANG X D, LI W, HOU Y L, et al. Comparative membrane proteomic analysis between lung adenocarcinoma and normal tissue by iTRAQ labeling mass spectrometry[J]. Am J Transl Res, 2014, 6(3):267-280.
- [9] NA S S, ALDONZA M B, SUNG H J, et al. Stanniocalcin-2 (STC2): a potential lung cancer biomarker promotes lung cancer metastasis and progression[J]. Biochim Biophys Acta, 2015, 1854(6):668-676.
- [10] YU L, SHEN J, MANNOOR K, et al. Identification of ENO1 as a potential sputum biomarker for early-stage lung cancer by shotgun proteomics[J]. Clin Lung Cancer, 2014, 15(5):372-378.
- [11] JIN H W, CHENG X, PEI Y H, et al. Identification and verification of transgelin-2 as a potential biomarker of tumor-derived lung-cancer endothelial cells by comparative proteomics[J]. J Proteomics, 2016, 136(1):77-88.
- [12] CLARK D J, FONDRIE W E, YANG A, et al. Triple SILAC quantitative proteomic analysis reveals differential abundance of cell signaling proteins between normal and lung cancer-derived exosomes[J]. J Proteomics, 2016, 133(2):161-169.
- [13] YU K, ZHANG C, BERRY G J, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features[J]. Nat Commun, 2016, 7(12):12474.
- [14] FUJITA K, NONOMURA N. Urinary biomarkers of prostate cancer[J]. Int J Urol, 2018, 25(9):770-779.
- [15] 张时民.医学检验领域人工智能技术应用与展望[J].国际检验医学杂志,2018,39(5):513-516.
- [16] 周格琛,贺望娇,韦维,等.骨髓细胞学检查的全面质量管理[J].国际检验医学杂志,2017,38(4):511-513.
- [17] 郭月锋,吴强,吉冰,等.全自动血细胞分析仪与血涂片细胞形态学联合检测在血常规检验中的应用效果[J].临床医学,2018,38(5):64-65.
- [18] ITAKURA H, ACHROL A S, MITCHELL L A, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities[J]. Sci Transl Med, 2015, 7(33):303ra138.
- [19] 郭毅可,杨磊.精确医学与大数据[J].上海大学学报(自然科学版),2016,22(1):17-27.
- [20] NIE K E, SHI L M, CHEN Q, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI[J]. Clin Cancer Res, 2016, 22(21):5256-5264.
- [21] HUANG Y Q, LIANG C H, HE L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer[J]. J Clin Oncol, 2016, 34(4):2157-2164.
- [22] 夏洪图,孙静.宁夏农村基层远程诊疗系统的建设应用现状分析[J].中国数字医学,2018,13(1):85-87.
- [23] 张国明,顾兵.医学检验大数据应用的思考[J].临床检验杂志,2017,35(4):304-305.
- [24] MARX V. The big challenges of big data[J]. Nature, 2013, 498(7453):255-260.