

基于大数据分析的心力衰竭诊疗研究进展

于佳女¹, 沙悦¹, 郭树彬²

¹中国医学科学院 北京协和医学院 北京协和医院普通内科, 北京 100730

²首都医科大学北京朝阳医院急诊科, 北京 100020

通信作者: 郭树彬 电话: 010- 65295303, 电子邮件: E-mail: shubinguo@126. com

摘要: 心力衰竭是一种严重的疾病, 患病率、死亡率均较高。大数据分析是新兴的研究方法, 利用该方法进行心力衰竭研究, 能够提高心力衰竭的诊断和治疗水平。目前心力衰竭的大数据研究主要涉及诊断、分类、预测预后等方面。本文回顾总结了相关文献, 显示了基于大数据分析方法进行心力衰竭研究的有效性。

关键词: 心力衰竭; 诊断; 大数据分析

中图分类号: R541. 6 文献标志码: A 文章编号: 1000-503X(2018)06-0843-04

DOI: 10. 3881/j. issn. 1000-503X. 10242

Research Advances in Big Data-based Diagnosis and Therapy of Heart Failure

YU Jianü¹, SHA Yue¹, GUO Shubin²

¹Department of General Internal Medicine, PUMC Hospital, CAMS and PUMC, Beijing 100730, China

²Department of Emergency Medicine, Beijing Chaoyang Hospital, Capital Medical University, Beijing 100020, China

Corresponding author: GUO Shubin Tel: 010- 65295303, E-mail: shubinguo@126. com

ABSTRACT: Heart failure is a serious condition with high prevalence and mortality. The application of the novel big data analysis in heart failure can improve the management of this condition, especially in terms of diagnosis, classification, and prognostic prediction. This articles reviews relevant literature and validates the role of big data analysis for heart failure patients.

Key words: heart failure; diagnosis; big data analysis

Acta Acad Med Sin, 2018, 40(6):843–846

心力衰竭是一种重要的公共健康问题, 该病患病率、死亡率、医疗费用均较高。据 2017 年美国心脏协会发布的心脏病与脑卒中统计数据显示, 2011 至 2014 年美国共有 650 万例心力衰竭患者, 预计 2030 年患者数量将增至 800 万, 而死亡统计数据显示, 美国约 1/8 死亡医学证明上含有心力衰竭诊断, 2012 年美国因心力衰竭产生的医疗费用高达 307 亿美元^[1], 目前亟待新的研究提高心力衰竭的诊疗水平。

近年来随着医疗系统电子化与可穿戴监测设备的流行, 可利用的医疗数据呈指数增长, 与此同时数据科学和大数据分析方法快速发展, 心力衰竭的研究迎来新

的机遇。目前有关心力衰竭的大数据分析研究, 主要基于患者的病史资料、查体特征、辅助检查、治疗方案等数据, 使用算法建立模型, 进行诊断、分类和预测预后方面的研究分析。笔者使用“心力衰竭”、“机器学习”、“深度学习”、“聚类分析”、“神经网络”等关键词在 PubMed、Web of Science、Scopus、ScienceDirect、EMBASE 数据库进行文献检索并汇总如下。

大数据分析概述

大数据是指具有 5 种特征的数据, 包括数据量大

(TB级以上)、数据快速产生和更新、数据类型多样、数据来源于真实世界、数据价值高而价值密度低^[2-4]。大数据分析是发现大数据的隐藏规律、潜在价值的方法,主要包括传统机器学习、深度学习等^[2-4]。大数据分析的主要流程包括数据预处理、数据标注、数据特征选择、应用算法建立数据分析模型、应用数据检验模型、进一步反馈优化模型等^[5]。

心力衰竭大数据分析研究现状

诊断 在心力衰竭诊断方面,大数据分析方法能够利用患者的病史资料、查体特征、辅助检查、治疗方案等数据,建立自动诊断和预测心力衰竭的模型。Sudarshan等^[6]使用双树复小波变换算法分析心力衰竭患者和健康对照者的心电图数据,发现45个有差异的心电图数据特征,使用k-近邻算法进一步根据心电图数据特征,建立心力衰竭诊断模型,经测试数据验证,该模型的诊断特异度达99.94%。既往通过人工方法分析识别心电图特征诊断心力衰竭的特异度较低,而大数据分析方法能够进一步精准和深入的利用心电图数据特征,提高了应用心电图数据诊断心力衰竭的特异度,具有里程碑式意义,近年来不断有基于大数据分析方法的研究分析心电图数据特征,建立心力衰竭的诊断模型^[7-12],诊断特异度均较好。Zheng等^[13]使用最小二乘支持向量机算法分析心力衰竭患者和健康对照者的心音数据,从中提取数据特征,并根据数据特征建立心力衰竭的诊断模型,经测试数据验证,该模型的诊断特异度可达96.59%。该研究使用大数据分析方法分析心音数据,为传统心力衰竭研究开拓了新的可利用数据领域。Choi等^[14]使用回归神经网络算法分析3884例心力衰竭患者和28903名对照者的门诊医嘱数据,包括医嘱的时间顺序、空间密度等特征,建立心力衰竭的诊断模型,经测试数据验证,该模型的诊断的ROC曲线下面积(area under curve, AUC)达0.883。医嘱数据具有数据量大、数据价值密度低的经典的大数据特征,传统研究方法并不能对其进行充分利用和分析,而大数据分析方法却能够较为有效地利用和分析这些数据特征。这些研究显示了大数据分析方法在建立模型诊断心力衰竭方面的有效性,未来有待更大样本量、数据质量更高、算法性能更好的大数据研究,进一步提高心力衰竭诊断模型的准确度,最终实现计算机系统远程监测患者数据,自动分析诊断和预测心力衰竭疾病的发生。

分类 综合分析心力衰竭患者整体情况,对患者进一步分类,为各种类型的心力衰竭患者提供更精准的诊疗服务,是另一个重要的心力衰竭研究方向。传统研究中,心力衰竭患者的分类办法,主要包括根据血流动力学特征、N端-脑钠肽前体、纽约心脏协会(New York Heart Association, NYHA)分级分类,还可以根据美国心脏协会(American Heart Association, AHA)/美国心脏病学会(American College of Cardiology, ACC)分期进行分类,其中包括根据心力衰竭患者的危险因素、症状、心脏结构、治疗方案数据。这些分类方法存在一定不足,如根据血流动力学分类后的心力衰竭患者仍存在有临床特征差异,可以进一步分类, NYHA 分级评价存在一定主观性, ACC/AHA 分期过于复杂而未能临床工作中广泛应用普及。

大数据分析方法能够利用大量的多种类型的临床数据,建立心力衰竭患者的分类模型,分类更为精准和细化。Shah等^[15]使用系统聚类分析算法分析射血分数保留型心力衰竭(heart failure with preserved ejection fraction, HFpEF)患者的67种数据资料,其中包括超声心动图中的心脏结构、心脏收缩/舒张功能、血流动力学、压力-容积分析等变量,建立HFpEF患者的分类模型,最终进一步区分出在病因、病理生理学和预后方面均有统计学差异的患者,同时该研究还进行了HFpEF患者的前瞻性队列研究,验证该模型的有效性, AUC 达 0.704。这项研究的意义在于,传统研究已经根据血流动力学特征对心力衰竭患者进行了分类,而大数据分析方法能够进一步细化分类,推进心力衰竭分类研究的发展。Guidi等^[16]使用随机森林算法分析心力衰竭患者的远程监测数据,包括患者末次住院数据(身高、体重、BMI、血压、心率、血氧饱和度、心脏射血分数、NYHA 分级、心电图、合并症、治疗)、家访数据(体重、血压、血氧饱和度、颈静脉充盈程度、皮肤颜色、脚踝是否水肿、生物电阻抗等)、患者自我监测数据(二导联心电图、生物电阻抗、脉搏传导时间)等,建立心力衰竭患者疾病严重程度的分类模型,分类特异度可达95%,显示了计算机系统远程监测患者全面的大数据、自动分类和评价心力衰竭患者的可行性。Chen等^[17]使用非平衡决策的基于决策树的支持向量机算法,分析心力衰竭患者和健康对照者的24h动态心电图,作者先将心电图截成多个5分钟片段,之后根据RR间期分析心率变异性,建立心力衰竭严重程度分类模型,分类准确度可达96.61%。综上,大数据分析方法能够有效对

心力衰竭患者进行分类评价,分类评价结果较传统研究更为细化,未来有待更多的外部队列研究进一步证实该方法的准确度及成本-效益分析。

预测预后 传统研究通过生存分析研究患者预后的影响因素,再根据患者所具有影响因素预测其预后。大数据分析能够利用更大量、更多类型以及一些价值密度较低的数据,建立心力衰竭患者的预后预测模型,较传统研究所利用的数据更全面,预测的准确度更高。如传统研究的西雅图心力衰竭模型(Seattle heart failure model, SHFM)^[18]对 1125 例心力衰竭患者的 21 种数据变量进行生存分析,预测心力衰竭患者预后 AUC 可达 0.729,并通过多个外部队列的验证^[19]。而 Panahiazar 等^[20]以 SHFM 模型为基础,纳入更大样本量患者和数据变量,包括 5044 例心力衰竭患者、43 种研究变量,使用随机森林等算法建立疾病预后预测模型,AUC 提高至 0.81。同时,Panahiazar 等^[20]的研究是利用真实世界的大数据,较 SHFM 研究意义更大。Shameer 等^[21]使用朴素贝叶斯算法分析 1068 例心力衰竭患者多达 4250 项数据变量,包括诊断($n=1763$)、治疗($n=1028$)、实验室检查($n=846$)、医嘱记录($n=564$)、生命体征($n=4$)数据,建立心力衰竭患者再住院预测模型,预测 AUC 达 0.78。Shameer 等^[21]的研究利用了数据量巨大但价值密度较低的临床数据,是传统研究无法做到的。Taslimitehrani 等^[22]使用对比模式辅助 Logistic 回归算法分析 5044 例心力衰竭患者的 40 种数据变量,包括人口学资料、生命体征、身高、体质量指数、实验室检查、治疗、合并症数据,预测心力衰竭患者预后准确度达 93.70%。Gleeson 等^[23]使用机器学习方法分析心力衰竭患者的 72 种数据变量,包括心电图、超声心动图的特征等,研究发现其中 27 种变量 $P<0.02$,能有效应用于预测心力衰竭患者预后,并指出空间 QRS-T 角 $>110^\circ$ 变量与心力衰竭患者再住院事件相关性非常强(风险比 3.7)。Koulaouzidis 等^[24]使用朴素贝叶斯算法,分析心力衰竭患者末次住院和远程监测数据,包括患者一般情况、心力衰竭病因、合并症、实验室检查、心功能 NYHA 分级、治疗、远程监测数据(生命体征、体重、一般情况、治疗、饮酒量),建立心力衰竭患者再住院预测模型,经随访(286 ± 281)d,预测 AUC 达 0.82。Koulaouzidis 等^[24]研究显示了大数据分析处理患者远程监测数据,实时动态自动分析和预测心力衰竭患者再住院的可行性和有效性。Zheng 等^[25]使用支持向量机算法分析心力衰竭患者数

据,包括年龄、医保类型、敏度评估(视听与思维)、合并症、是否急诊治疗、用药风险、末次住院周期等数据,建立心力衰竭患者再住院预测模型,预测准确度达 78.4%。以上这些研究显示了大数据分析方法在建立心力衰竭预测模型方面的有效性。

总结与展望

综上,目前大数据分析方法在心力衰竭研究中应用广泛,与传统研究相比,在研究数据方面,大数据分析能够利用数据量巨大、更多种类的临床数据,能够利用医嘱的时间顺序、空间密度等价值密度较低的临床数据,能够利用患者的远程监测数据,实时动态分析;在研究成果方面,大数据分析所建立的疾病诊断模型、疾病分类模型、疾病预后预测模型,准确度均较高。目前基于大数据分析的心力衰竭研究还存在较大发展空间,如进一步提高原始数据的数量和质量,进行更多的前瞻性队列研究验证大数据分析所建的模型的有效性,大数据分析本身也有待进一步发展成熟。相信未来在计算机专业科学家、心脏病学研究者的紧密合作下,会有更多、更高质量的医学大数据研究,从而提高心力衰竭疾病诊疗水平。

参考文献

- [1] Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart disease and stroke statistics-2017 update: a report from the American Heart Association [J]. *Circulation*, 2017, 135(10):e146-e603. DOI:10.1161/cir.0000000000000485.
- [2] Tan SS, Gao G, Koch S. Big data and analytics in health-care [J]. *Methods Inf Med*, 2015, 54(6):546-547. DOI: 10.3414/me15-06-1001.
- [3] Obermeyer Z, Emanuel EJ. Predicting the future-big data, machine learning, and clinical medicine [J]. *N Engl J Med*, 2016, 375(13):1216-1219. DOI:10.1056/NEJMp1606181.
- [4] Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges [J]. *Nat Rev Cardiol*, 2016, 13(6):350. DOI: 10.1038/nrcardio.2016.42.
- [5] Drovandi CC, Holmes C, McGree JM, et al. Principles of experimental design for big data analysis [J]. *Stat Sci*, 2017, 32(3):385-404. DOI:10.1214/16-sts604.
- [6] Sudarshan VK, Acharya UR, Oh SL, et al. Automated diagnosis of congestive heart failure using dual tree complex

- wavelet transform and statistical features extracted from 2s of ECG signals [J]. *Comput Biol Med*, 2017, 83 (C): 48-58. DOI:10.1016/j.combiomed.2017.01.019.
- [7] Melillo P, Fusco R, Sansone M, et al. Discrimination power of long-term heart rate variability measures for chronic heart failure detection [J]. *Med Biol Eng Comput*, 2011, 49(1): 67-74. DOI:10.1007/s11517-010-0728-5.
- [8] işler Y, Kuntalp M. Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure [J]. *Comput Biol Med*, 2007, 37(10): 1502-1510. DOI:https://DOI.org/10.1016/j.combiomed.2007.01.012.
- [9] Thuraingham RA. A classification system to detect congestive heart failure using second-order difference plot of RR intervals [J]. *Cardiol Res Pract*, 2009, 2009: 807379. DOI: 10.4061/2009/807379.
- [10] Yu SN, Lee MY. Conditional mutual information-based feature selection for congestive heart failure recognition using heart rate variability [J]. *Comput Methods Programs Biomed*, 2012, 108(1): 299-309. DOI:10.1016/j.cmpb.2011.12.015.
- [11] Pecchia L, Melillo P, Sansone M, et al. Discrimination power of short-term heart rate variability measures for CHF assessment [J]. *IEEE Trans Inf Technol Biomed*, 2011, 15(1): 40-46. DOI:10.1109/titb.2010.2091647.
- [12] Masetic Z, Subasi A. Congestive heart failure detection using random forest classifier [J]. *Comput Methods Programs Biomed*, 2016, 130 (Supplement C): 54-64. DOI: 10.1016/j.cmpb.2016.03.020.
- [13] Zheng Y, Guo X, Qin J, et al. Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics [J]. *Comput Methods Programs Biomed*, 2015, 122 (3): 372-383. DOI: 10.1016/j.cmpb.2015.09.001.
- [14] Choi E, Schuetz A, Stewart WF, et al. Using recurrent neural network models for early detection of heart failure onset [J]. *J Am Med Inform Assoc*, 2017, 24(2): 361-370. DOI:10.1093/jamia/ocw112.
- [15] Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction [J]. *Circulation*, 2015, 131(3): 269-279. DOI:10.1161/circulationaha.114.010637.
- [16] Guidi G, Pollonini L, Dacso CC, et al. A multi-layer monitoring system for clinical management of Congestive Heart Failure [J]. *BMC Med Inform Decis Mak*, 2015, 15(Suppl 3): S5-S5. DOI:10.1186/1472-6947-15-S3-S5.
- [17] Chen W, Zheng L, Li K, et al. A novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement [J]. *PLoS One*, 2016, 11(11): e0165304. DOI:10.1371/journal.pone.0165304.
- [18] Levy WC, Mozaffarian D, Linker DT, et al. The seattle heart failure model: prediction of survival in heart failure [J]. *Circulation*, 2006, 113(11): 1424-1433. DOI:10.1161/circulationaha.105.584102.
- [19] Alba AC, Agoritsas T, Jankowski M, et al. Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review [J]. *Circ Heart Fail*, 2013, 6(5): 881-889. DOI:10.1161/circheartfailure.112.000043.
- [20] Panahiazar M, Taslimitehrani V, Pereira N, et al. Using EHRs and machine learning for heart failure survival analysis [J]. *Stud Health Technol Inform*, 2015, 216: 40-44. DOI: 10.3233/978-1-61499-564-7-40.
- [21] Shameer K, Johnson KW, Yahi A, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using mount sinai heart failure cohort [J]. *Pac Symp Biocomput*, 2016, 22: 276. DOI:10.1142/9789813207813_0027.
- [22] Taslimitehrani V, Dong G, Pereira NL, et al. Developing EHR-driven heart failure risk prediction models using CPXR (Log) with the probabilistic loss function [J]. *J Biomed Inform*, 2016, 60 (C): 260-269. DOI: 10.1016/j.jbi.2016.01.009.
- [23] Gleeson S, Liao Y, Dugo C, et al. ECG-derived spatial QRS-T angle is associated with ICD implantation, mortality and heart failure admissions in patients with LV systolic dysfunction [J]. *PLoS One*, 2017, 12(3): e0171069. DOI: 10.1371/journal.pone.0171069.
- [24] Koulaouzidis G, Iakovidis D, Clark A. Telemonitoring predicts in advance heart failure admissions [J]. *Int J Cardiol*, 2016, 216: 78-84. DOI:10.1016/j.ijcard.2016.04.149.
- [25] Zheng B, Zhang J, Sang WY, et al. Predictive modeling of hospital readmissions using metaheuristics and data mining [J]. *Expert Syst Appl*, 2015, 42(20): 7110-7120. DOI: 10.1016/j.eswa.2015.04.066.

(收稿日期: 2017-12-18)