

机器学习方法在中医药研究中的应用进展

王志杰¹, 樊薛津², 王豫蹇², 庄朋伟^{3*}, 赵婷婷^{1*}

1. 天津科技大学人工智能学院, 天津 300457

2. 天津中医药大学, 天津 301617

3. 天津中医药大学第一附属医院, 天津 300381

摘要: 随着信息技术的快速发展及数字化时代的到来, 数字化技术在多个中医药研究领域, 如中药质量控制、数据挖掘、新药发现、处方配伍优化以及中医临床诊断等方面发挥重要作用。机器学习 (ML)、深度学习 (DL) 等数字化技术的应用可优化中医药研究设计、降低临床研究时间成本、将临床研究与基础研究有机结合, 提高临床研究质量和效率, 为中医药现代化研究的科学性、客观性、规范性提供了保障。因此, 完善多种数据资源、实现多种数字化技术交叉结合使用、优化算法和模型是中医药未来发展的必然趋势。对近年来ML在中医药研究领域中的应用进展进行梳理, 分析说明聚类、支持向量机 (SVM) 和深度学习 (DL) 等ML方法在中药药性、中药配伍、中药毒性分析、中药药效研究、中药制药过程工艺优化、中药饮片质量等级分类、中医药临床用药规律挖掘、中医药治疗原理及药效机制解析等研究中的具体应用。旨在探索ML在中医药研究中的应用趋势, 并对其应用前景进行展望。

关键词: 机器学习; 中医药; 算法模型; 人工智能; 大数据; 深度学习

中图分类号: R911 **文献标志码:** A **文章编号:** 1674-6376 (2024) 08-1906-08

DOI: 10.7501/j.issn.1674-6376.2024.08.024

Application of machine learning methods in research of traditional Chinese medicine

WANG Zhijie¹, FAN Xuejin², WANG Yuqian², ZHUANG Pengwei³, ZHAO Tingting¹

1. College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China

2. Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China

3. First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin 300381, China

Abstract: With the rapid development of information technology and the arrival of the digital era, digital technology plays an important role in several Chinese medicine research areas, such as quality control of traditional Chinese medicine (TCM), data mining, discovery of new medicines, optimization of prescription compounding, and clinical diagnosis of Chinese medicine. The application of digital technologies such as machine learning (ML) and deep learning (DL) can optimize the design of TCM research, reduce the time and cost of clinical research, organically combine clinical research with basic research, improve the quality and efficiency of clinical research, and provide a guarantee for the scientificity, objectivity, and standardization of modernized TCM research. Therefore, it is an inevitable trend for the future development of TCM to improve multiple data resources, realize the cross-combined use of multiple digital technologies, and optimize algorithms and models. In recent years, the progress of ML application in the field of TCM research is sorted out, and the analysis illustrates the specific applications of ML methods such as clustering, support vector machine (SVM) and deep learning (DL) in the research of TCM medicinal properties, TCM compounding, TCM toxicity analysis, TCM efficacy research, TCM pharmaceutical process optimization, TCM tablets quality grade classification, mining of TCM clinical medication rules, and analysis of TCM therapeutic principles and pharmacological effect mechanisms. The research aims to explore the application of ML in the study of TCM. It aims to explore the trend of ML application in TCM research and to prospect its application prospect.

Key words: machine learning; traditional Chinese medicine; algorithm model; artificial intelligence; big data; deep learning

收稿日期: 2024-04-16

第一作者: 王志杰(1998—), 男, 在读硕士研究生, 从事中药配伍研究。E-mail: 19118253746@163.com

*共同通信作者: 庄朋伟(1982—), 男, 博士, 教授, 主要从事中药药理学研究。E-mail: zhuangpengwei@163.com

赵婷婷(1986—), 女, 博士, 教授, 主要从事机器学习算法、智能信息处理、机器人控制等研究。E-mail: tingting@tust.mail.cn

机器学习(ML)是通过先验知识形成训练集,识别和提取数据特征,建立预测模型,并将所学习到的规律应用于相同类型数据中进行预测或分类的一种方法,是人工智能(AI)的一个分支,也是一类算法的总称^[1-2]。1959年,“机器学习”一词被AI领域的先驱亚瑟·塞缪尔(Arthur Samuel)提出。1998年, Tom M. Mitchell为“机器学习”提供了一个正式的定义:“如果一个计算机程序在T任务中的表现(由P衡量)随着经验E的提高而提高,那么就说明它从经验E中学习了关于某类任务T和性能指标P的经验E”。ML主要分为监督学习、无监督学习、半监督学习和强化学习等类别^[3],包括支持向量机(SVM)、K近邻(KNN)算法、决策树(DT)、随机森林(RF)、AdaBoost、K-means聚类、卷积神经网络(CNN)、人工神经网络(ANN)等算法。而深度学习(DL)是基于ANN算法在机器学习领域中的一种分支,二者都是实现AI的具体方法,也是AI和数据科学的核心。

中医药药性理论、配伍理论等是传统中医临床经验的高度概括和总结,蕴含着庞大且复杂的知识系统。同样,中药物质基础的复杂性也对中医药现代化研究带来了巨大挑战。虽然现代实验方法对中医药研究有一定实用性,从化学成分研究到药效评价及药理作用机制研究取得了巨大成就,但是在深入研究中医药配伍等复杂知识体系方面,传统实验方法似乎很难获得突破性成果。随着中医药领域数据化的不断完善,ML这种处理大型、复杂和不同数据的能力正是中医药现代化研究所需要的^[4]。Zhao等^[5]构建多图卷积网络(MGCN)的处方推荐模型,模拟了中医诊治中方-症之间的“多元”(n-ary)关系,显著提高了中药处方预测的准确性,具有很大的实用价值。Wang等^[6]建立症-病-药的三者联合分析的非对称概率模型,以发现和提供中医的潜在知识。舒琛洁等^[7]构建了“气虚”证候要素-症状数据,以受试者工作曲线(ROC)曲线下面积等作为评价指标,对比了分类回归树、RF、KNN、反向传播神经网络和SVM共5种ML算法模型,优选RF最适于构建“气虚”辅助诊断模型的算法,为证候要素辅助诊断提供方法学基础。以上可得,ML在明确中医方-证关联、中药质量标准控制、精准医疗等领域具有很强的指导作用^[8-9]。本文就ML在中医药基础理论研究中的应用进行综述,以期对中医药研究及中医药现代化的客观性、科学性提供一定的依据。

1 ML在中医药研究中的一般流程

应用在中医药研究的ML主要采用监督学习方法,即通过对带有标签的数据集进行训练,学习输入和输出之间的关联模式,进行预测或分类,主要包括以下关键步骤^[10-11]。(1)建立纳排标准进行数据收集(资料来源于中医药数据库、专家辨证、临床中医病案、教材等)和数据预处理;(2)特征选择和编码:选择最相关的特征,将数据转换成ML可以处理的形式;(3)模型的构建、训练与评估:选择适合任务的ML算法(SVM、KNN、RF、BP神经网络等),使用ROC曲线下面积、准确率等指标进行评估;(4)预测与模型优化:将训练好的模型用于新数据的预测,不断迭代,优化模型性能。

2 ML在中医药基础理论研究中的应用

2.1 ML应用于中药药性理论研究

中药具有四气五味、升降浮沉和归经等特性,一般是以复方形式使用发挥治疗效果。中药药性理论在指导临床遣方用药方面具有重要价值,中药药性与临床药效的发挥关系密切,存在相关性。然而中药药性主要是根据药物作用于机体后患者的感觉,相关表述抽象、模糊,难以量化表示。如关联规则挖掘、SVM、KNN、马尔可夫决策过程(MDP)、概率神经网络(PNN)等在该领域应用广泛。王伟斌等^[12]利用古今医案云平台对《医案类聚》中364条郁证医案的439条处方进行频次统计、聚类分析、关联分析、复杂网络分析发现治疗郁证以药性温、平、微寒中药为主,温性最高;以苦味、辛味为主,以苦味药物频次最高;多归于脾、肺、心经,以归脾经最高,为郁证的临床治疗提供一定的参考。张喜科等^[13]构建XGBoost模型对61味中药在无水乙醇、蒸馏水、氯仿、石油醚4种溶剂下的中药紫外光谱数据进行寒热药性识别,通过稳定性评价和外推评价等方式评价模型的性能,为中药寒热药性识别研究提供了一种新的思路和方法。邓乐等^[14]探索了一种基于多层前馈神经网络的药向量训练模型,利用中药学教材中整理得到的474味中药的药性数据对模型进行初始化,利用相关药物对应的功效-药性数据进行训练,得到能够更好反映中药属性的BP药向量,并且药相量在欧几里得距离方面能够体现中药功效差异。刘莉萍等^[15]围绕解决中药药性和药效关联的问题,在经典关联分类算法的基础上通过垂直数据格式对其进行改进,选取补虚药和清热药2类功效药物为代表构建数据集,证实了改进关联分类算法在规则生成时间、有效性及准确率方面具有

明显提升,提高了算法的鲁棒性。

随着现代中药提取技术的不断发展,人们越来越多的关注和探索中药成分或组分的药性属性,中药药性与其物质基础的关系也成为药性理论创新发展的重要突破点。利用现代算法技术可以对海量成分结构特征信息进行特异性提取,贾荣浩等^[16]围绕中药化合物药性识别问题,基于中药化合物的结构决定其药性的假说,构建了一种基于K-近邻算法(KNN)的中药化合物寒热平药性预测模型,通过提取2 012个中药化合物的结构特征,利用KNN度量中药化合物的相似性,构建药性预测模型,并利用交叉验证的方式验证了中药化合物药性预测模型的有效性。张冰冰等^[17]利用中医药综合数据库(TCMID, <http://www.megabionet.org/tcmid/>)、中医药数据库(ETCM, <http://www.tcmip.cn/ETCM/index.php/Home/Index/index.html>)收集646味中药,10 053个化合物成分,应用图卷积神经网络实现了中药化合物分子的寒热属性预测,并通过细胞增殖-毒性测试实验检测梯度加权得分的寒性中药成分对大鼠肾上腺髓质嗜铬瘤(PC-12)细胞氧糖剥夺再灌注(OGD/R)模型的保护作用,为进一步探究中药“性-构”关系提供算法支持。ML算法技术在中药药性研究方面的应用不仅有利于中药药性理论的量化,还对拓展现代组分中药药性理论具有重要价值,更好地促进中医药传承创新发展。

2.2 ML应用于方剂配伍理论研究

中药方剂是中医临床用药的基本形式,方剂配伍规律的分析与挖掘是中医药研究的重要课题,是中医药传承创新的重要抓手之一。方剂资源主要涉及历代医家临床经典方剂的积累与总结,主要涉及方剂组成、治法治则、主治病症及其证候特征、各药味间的配伍关系等内容。方剂配伍规律主要包括方剂相似性比较、治疗某一特定病症药物的配伍规律等,明确方剂配伍规律对临床遣方用药具有重要指导价值。中药及其复方功效的相似性是中医药的明显特征之一,针对同一病症的治疗往往存在多个构成和功效相似的中药复方,通过对复方中药相似性比较,可以更好地发现中药处方的共性特征,在指导临床灵活用药和创新中药研发方面具有重要意义。

郭文龙等^[18]通过对7 710首不同类型方剂的整理,基于统计学方法原理,应用TF-IDF(term frequency-inverse document frequency)算法构建了中药方剂数据挖掘系统,可直观展示方剂之间的相

似度,并且发现相似度高的方剂在功能主治方面具有相似性,实现了相似度和方剂信息、方剂与药物间的关联关系等信息可视化展示。针对特定疾病的处方用药规律分析有利于发现优效处方,王慧敏等^[19]以治疗失眠的处方为例,利用层聚类分析方法对《方剂大辞典》中治疗失眠的处方进行了探索性分析,针对相关单味药的药性理论信息,通过单连接、全连接和平均连接3种判断相似度的度量方式评估簇间相似性,发现基于全连接的层次聚类分组符合中医理论,相对较为合理。方剂配伍规律分析多依托《方剂大辞典》等,以及名家医话医案等为数据来源,通过爬虫框架以及手工录入等方式获取一定数量的经典方剂,利用贝叶斯、TF-IDF、K均值聚类、层聚类、Louvain,搜索(Tabu)等分类算法等进行方剂配伍用药规律的挖掘^[20],对中医药传承创新发展具有重要参考价值。

2.3 ML应用于中医毒性理论研究

中医药历来重视临床用药的安全性,在临床实践过程中逐渐发展并完善了中药的大毒、有毒、小毒毒性分级理论,配伍禁忌、妊娠禁忌、证候禁忌、饮食禁忌等用药禁忌理论,然而现代实验研究在中药安全性评估方面受到化学成分复杂、物种差异等诸多因素的限制,难以满足临床需求。基于计算机算法技术的虚拟预测评价技术越来越多地被开发并应用于中药安全性预测,针对药物的不良反应预测,张文青等^[21]基于模式识别分类原理,整合ANN和SVM算法进行预测模型的构建,利用数据库筛选获得的125个具有明确药源性肾炎的药物和122个未导致药源性肾炎的药物作为预测模型的训练集,同时各选择45种药物进行验证,最终应用于对10种中药的药源性肾炎进行评价,结果发现该方法的预测准确度可达90%以上。此外,针对用药禁忌预测,姜皓等^[22]围绕妊娠期中药“禁忌慎”的判别,利用ANN、SVM、RF和朴素贝叶斯4种ML算法建立预测模型,用224种妊娠禁忌分类明确的药物作为建模药物,对442种禁忌表述不明确的中药进行了预测,对指导临床安全用药具有一定实用价值。

3 ML在中药学研究中的应用

3.1 ML应用于中药药效研究

传统中药的功效理论主要是历代医家基于临床经验总结获得,现代研究多基于实验药理学方法,利用实验动物、细胞等实验模型评价中药及其复方的药理作用,一方面尽管通过高通量筛选等技术的不断开发,对中药及其复方效应快速评价取得

了很大进展,但是由于中药多成分、多靶点、多环节的作用特点,往往会面临耗资大、耗时长等问题,并且受其“成分组合爆炸”的限制,使得一些研究的可行性存在很大问题。另一方面现代药理实验难以很好地兼顾与传统功效理论的相关性,基于实验药理学的方法已不足以满足中药现代研究的需求。快速、高效地评估中药的化学成分与传统功效理论的相关性对中医药现代化发展具有重要价值,随着现代计算机算力及算法技术的不断开发,基于虚拟算法技术预测中药及其复方药效已成为当前及未来中药及其复方研究的重要方向。

为解析中药化学成分与传统功效之间的关系,雷蕾等^[23]提出“化学功效组学”的概念,以26个止血功效的中药和38个活血化瘀功效的中药为示范,通过收集化学成分并规范功效及药理作用表述,使用KNIME 4.3.2软件CDK算法和RDKit算法包识别了止血和活血化瘀功效相关的成分Murcko骨架,结果分别发现了止血药和活血化瘀药功效代表性成分共性特征,如止血药核心成分主要包含特有的萘醌、四环三萜等Murcko骨架,而活血化瘀药核心成分主要为生物碱Murcko骨架,为中药化学成分的功效研究提供了一定借鉴。孙飞等^[24]采用偏最小二乘(PLS)算法对山楂炮制与否的化学成分与消食健脾功效的“谱-效”关系进行了较为系统的研究,通过对山楂和焦山楂不同提取部位分离制备并进行成分含量测定,在功效评价的基础上进行谱效关系研究,初步明确了山楂、焦山楂炮制前后的消食健脾功效成分变化,揭示了山楂炒焦炮制缓和药性,增强消食导滞作用的炮制机制。Ouyang等^[25]通过MBPLS建模来识别碳化蒲黄花粉(CTP)中对止血效果有贡献的关键化学成分类型,分别建立PLS和SVM模型,探讨关键化学成分与药效之间的“谱-效”关系,鉴定了CTP的潜在活性成分,并在体外进一步验证了其止血生物活性。Kong等^[26]使用相似性分析(SA)和分层聚类分析(HCA)评估UPLC指纹图谱,采用典型相关分析(CCA)研究了UPLC指纹图谱与黄连抗菌活性之间的“谱-效”关系。Chen等^[27]利用紫外-可见分光光谱法和UHPLC/Q-TOF-MS数据,通过PLS回归分析来预测了熊果叶片的总酚含量和抗氧化能力。目前,CCA、PLS、灰度相关分析(GCA)、SVM已被证明对于建立“谱-效”关系模型是可行的,为阐明中药药效物质基础提供方法。

3.2 ML与中药制药过程工艺优化

在现代中药制药工业领域,常围绕中药材提取进行生产过程工艺参数进行优化设置,可以有效保障中药材的提取效率、批次一致性、药效及安全性等。均匀设计、正交实验设计等传统实验设计方法常用于进行工艺条件的摸索,然而存在可探索的因素水平有限、仅能产生局部最优等诸多不足。通过数字建模分析,人工神经网络等现代算法技术可对中药提取制备工艺过程中的多参数进行随机非线性优化,对促进现代中药制药工业智能化发展具有重要现实意义。薛启隆等^[28]提出了一种基于深度强化学习的中药制药工艺动态优化策略,主要包含模型的离线训练和工艺的在线决策两部分,关于模型的准确性与泛用性主要取决于数据集的规模与质量,其数据可来源于中药制药工艺知识库。通过迁移学习方法可快速完成模型的训练,在实际生产过程中不断地扩充知识库并且决策精度也会逐渐提高。仲恂等^[29]以去壁灵芝孢子粉制药过程为例,将贝叶斯网络、卷积神经网络、帕累托多目标优化算法等先进计算工具与精益六西格玛分析工具相结合,提出了基于精益理念的中药制药工业数据挖掘及生产工艺持续改进策略,基于现代算法的中药制药工艺优化策略具有较高的工业应用价值,可为现代中药生产工艺的持续改进提供实用方法。

3.3 ML应用于中药饮片质量等级分类

中药饮片质量对保障临床用药的有效性及安全性至关重要,传统的中药饮片质量控制多依赖于经验,主观性强,缺乏客观统一的标准。随着计算机和影像学技术的不断发展,图像识别技术在中药饮片识别和质量控制方面得到很好的应用。有学者对不同实际应用场景下一百多种常用中药饮片进行了图像数据库的构建,在图像采集过程中通过人为增加不同光照强度、不同拍摄角度等复杂背景因素以增强算法泛化能力,在此基础上利用YOLO模型进行特征提取、模型训练,结果发现该算法对小目标检测效果较好,通过特征金字塔网络思想的引入提高了模型的泛化能力。吴冲等^[30]以山楂、半夏等饮片为代表,通过建立中药饮片图像数据库,以深度学习为手段对中药饮片外在性状特征进行智能分析,建立深度卷积神经网络模型对中药饮片进行识别,为中药饮片质量的检测提供了技术支持。

4 ML在中医药临床用药规律挖掘研究中的应用

针对特定医家或特定学术流派治疗特定疾病

的用药规律分析有利于经验的挖掘与传承,关联规则挖掘算法常用于中医药治疗疾病的组方配伍规律的研究,王利娟等^[31]以清初以来16位孟河医派代表性医家的治疗头痛的医话医案为基础,利用Apriori关联规则算法其用药规律进行了挖掘,明确了孟河医家治疗头痛的用药规律。秘红英^[32]利用贝叶斯分类算法对吴以岭教授346例治疗冠心病心绞痛的用药规律进行了挖掘,分析了其证型、药物功效、用药频次、药性特征等,有利于名医大师学术思想及临证经验传承发扬。此外,关联规则算法也常用于中西药联合用药的用药规律分析,马雪骄等^[33]以林洪生教授诊治肺癌的有效方剂为例,利用Apriori关联规则算法对其200个有效方剂进行了挖掘,发现肺癌分子靶向治疗过程中中医药使用多以扶正固本为纲领,并加以祛邪之品以增强化学药疗效。此外,现代AI算法技术也常用于中西药联合应用方案的优化,吴丽娜等^[34]利用Apriori关联规则算法分析了结肠恶性肿瘤患者的中西药联合应用特征,发现肿瘤化疗过程中中西药联合应用效果优于单独应用化学药,针对临床用药的优效方案评估预测。席俊羽等^[35]进行了临床联合用药提高临床疗效预测的探索性研究,以生脉注射液治疗心肌梗死为例,围绕心肌梗死的病因、症状以及并发症等问题的联合用药情况,利用Louvain算法对1208名心肌梗死患者的用药信息数据进行复杂网络分析,发现生脉注射液联合其他中成药和化学药治疗心肌梗死的用药规律,为临床中西药联合应用提供了一定支撑。

针对影响临床用药安全性条件的预测,蒋程等^[36]以参麦注射剂药物不良反应为选题,通过分析出现与否药物不良反应患者的个体信息和临床用药信息,构建多变量数据矩阵建立偏最小二乘-判别分析算法模型,评估了病史、药物过敏史、年龄、溶媒等因素对参麦注射液不良反应发生的影响,对指导临床安全用药具有重要意义。

临床用药规律分析多围绕名家医话医案或特定医学流派经验等为数据来源,对特定疾病的用药经验获取一定数量的经典方剂及病症特征,利用关联规则(Apriori、FP-growth)、搜索(Tabu)、社区发现算法(Louvain)等进行临床用药规律的挖掘,对指导临床用药、优化临床用药方案等具有较强的实用价值。

5 ML在中医药治疗原理及药效机制解析研究中的应用

中医药在临床治疗疾病过程中形成了其独特

的理论体系,包括药性理论、辨证论治、调和阴阳平衡等。然而由于在其理论形成及发展过程中受限于解剖、生理、病理等微观认知,其治疗多是依据临床证候现象的调理,受限临床效应的现代分子机制及治疗原理难以有效阐释,极大地限制了中医药的创新发展。利用现代科学语言对传统中医药治疗原理进行阐释可以更好地促进其与现代医学融合发展,更好地促进现代医学发展。针对中医药治疗原理,近日国内外学者联合研究报道利用蛋白质网络拓扑邻近关系解析中药与疾病症状的关联^[37],对中药“辨证论治”治疗原则的现代科学内涵进行了探索性研究,首先确认了症状关联基因在人类蛋白质网络上具有形成显著的集聚模块的现象,在此基础上进一步发现具有特定疗效的中药针对的症状也存在普遍性的网络邻近关系,最终提出人类蛋白质网络上的网络邻近关系可以作为中药疗效预测的指标,该方法极大的丰富了中医药治疗原理解析的方法学策略。针对中医药关键作用环节及药效机制,以免疫浸润关键病理环节所介导的相关疾病分析为例,王朴等^[38]利用CIBERSORT反卷积算法对皮肤免疫浸润机制进行了分析,首先对健康与皮肤炎患者的差异基因进行了分析,并将差异基因与免疫浸润细胞进行Pearson相关性分析,最终明确了皮肤炎发病过程中巨噬细胞以及CD8⁺T细胞的特征性变化,发现关键基因靶点并对潜在的中药治疗效果进行了预测,对基于病理机制的中药治疗研究有一定参考价值。刘洋等^[39]针对围绕结直肠癌“炎癌转化”过程进行了研究,整合ML和神经网络模型的算法对公共数据库中正常结肠、炎症结肠(溃疡性结肠炎)以及肠癌相关疾病信息进行了关键基因的筛选及挖掘,通过构建蛋白质-蛋白质相互作用(PPI)网络进一步筛选核心基因,运用梯度提升树、随机森林和决策树3种分类ML算法进行结局的预测学习,构建深度学习框架下的人工神经网络模型并验证,在此基础上发现具有潜在治疗作用的中药。

中医药治疗原理及作用机制解析研究多利用公共数据库,通过蛋白质网络拓扑邻近关系解析、人工神经网络等策略,进行疾病关键病理环节识别、疾病早期诊断及潜在治疗药物发现等,对中医药传承创新发展具有重要价值。

6 结语与展望

当前ML在中医药领域中的应用逐渐广泛,在安全性评价、中药质量评估、中药配伍预测、处方中

药配比、医案数据处理等方面具有显著成果。传统ML方法因其小样本适应性、可解释性已在方剂领域取得很多成果,而DL技术目前应用虽少,但其结构、学习方式适用于拟合中医辨证论治、组方配伍思想,近年来在中医面诊、中药鉴定等方面展开应用研究。传统的中医药研究方法并不智能,同时相比于ML方法效率偏低,ML模型与中医的交叉融合为突破中医药现代化的发展瓶颈、探索中医药科技创新发展提供了可能。

与传统的中医药研究方法相比,基于ML的研究方法存在数据可靠性、模型选择以及可解释性等问题。其一是深度学习需要大量的样本数据训练模型,目前通过古书、网络爬虫等方法收集的中医药样本数据质量并不能得到保证;其二是ML算法的适用性,明确算法原理,针对不同的中医药数据特征优选适用的算法。对于模型框架、模型中各种参数的选择需要根据特定中医药领域中的问题进行不断尝试与改进,具有较大的工作量;其三是深度学习针对中医药问题构建的模型多为黑盒模型,模型的可解释性较差。

随着大数据技术、AI技术在药物研究和发现中应用越来越广泛^[40],传统ML方法已无法全面阐释中医药研究的全部规律。深度学习方法具有其强大的拟合能力、良好的泛化能力,是中医药现代化研究的必然趋势。另外,随着AI领域中大型模型不断发展,基于大型模型的中医药研究也是未来的一大趋势。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Buchlak Q D, Esmaili N, Leveque J C, et al. Machine learning applications to clinical decision support in neurosurgery: An artificial intelligence augmented systematic review [J]. *Neurosurg Rev*, 2020, 43(5): 1235-1253.
- [2] Jordan M I, Mitchell T M. Machine learning: Trends, perspectives, and prospects [J]. *Science*, 2015, 349(6245): 255-260.
- [3] Choi R Y, Coyner A S, Kalpathy-Cramer J, et al. Introduction to machine learning, neural networks, and deep learning [J]. *Transl Vis Sci Technol*, 2020, 9(2): 14.
- [4] Handelman G S, Kok H K, Chandra R V, et al. eDoctor: Machine learning and the future of medicine [J]. *J Intern Med*, 2018, 284(6): 603-619.
- [5] Zhao W, Lu W K, Li Z Y, et al. TCM herbal prescription recommendation model based on multi-graph convolutional network [J]. *J Ethnopharmacol*, 2022, 297: 115109.
- [6] Wang S, Huang E W, Zhang R S, et al. A conditional probabilistic model for joint analysis of symptoms, diseases, and herbs in traditional Chinese medicine patient records [A]// 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) [C]. Shenzhen: IEEE, 2016.
- [7] 舒琛洁, 梁浩, 刘淑明, 等. 机器学习算法对证候要素 "气虚" 辅助诊断的性能评估 [J]. *北京中医药大学学报*, 2021, 44(10): 928-934.
- [8] Shu C J, Liang H, Liu S M, et al. Performance evaluation of machine learning algorithms on the pattern elements of "qi deficiency" assisted diagnosis [J]. *J Beijing Univ Tradit Chin Med*, 2021, 44(10): 928-934.
- [9] Buchlak Q D, Esmaili N, Leveque J C, et al. Machine learning applications to neuroimaging for glioma detection and classification: An artificial intelligence augmented systematic review [J]. *J Clin Neurosci*, 2021, 89: 177-198.
- [10] Rajkomar A, Dean J, Kohane I. Machine learning in medicine [J]. *N Engl J Med*, 2019, 380(14): 1347-1358.
- [11] 普布玉珍, 高屹. 机器学习与知识图谱在藏医药领域的应用研究 [J]. *西藏科技*, 2023, 45(12): 69-73.
- [12] Pu B Y Z, Gao Y. Research on the application of machine learning and knowledge graph in the field of Tibetan medicine [J]. *Xizang Sci Technol*, 2023, 45(12): 69-73.
- [13] Jiang T, Gradus J L, Rosellini A J. Supervised machine learning: A brief primer [J]. *Behav Ther*, 2020, 51(5): 675-687.
- [14] 王伟斌, 李敬华, 于琦, 等. 基于古代医案的郁证用药规律分析 [J]. *中国实验方剂学杂志*, 2020, 26(5): 162-167.
- [15] Wang W B, Li J H, Yu Q, et al. Research on medication rules of depression syndrome based on ancient medical records [J]. *Chin J Exp Tradit Med Form*, 2020, 26(5): 162-167.
- [16] 张喜科, 赵文华, 马志庆, 等. 基于XGBoost算法的中药寒热药性识别研究 [J]. *时珍国医国药*, 2022, 33(8): 2022-2024.
- [17] Zhang X K, Zhao W H, Ma Z Q, et al. Study on identification of cold and hot nature of Chinese herbal medicines based on XGBoost algorithm [J]. *Lishizhen Med Mater Med Res*, 2022, 33(8): 2022-2024.
- [18] 邓乐, 丁长松, 黄辛迪, 等. 基于多层前馈神经网络的中药药性量化研究 [J]. *中草药*, 2020, 51(16): 4277-4283.
- [19] Deng L, Ding C S, Huang X D, et al. Quantitative study on medicinal properties of traditional Chinese medicine based on BP neural network [J]. *Chin Tradit Herb Drugs*, 2020, 51(16): 4277-4283.

- [15] 刘莉萍, 李欢, 何正宏. 关联分类算法改进及其在中药 "性-效" 分析中的应用 [J]. 现代信息科技, 2023, 7(18): 150-153, 158.
Liu L P, Li H, He Z H. Improvement of association classification algorithm and its application in "performance-efficacy" analysis of traditional Chinese medicine [J]. Mod Inf Technol, 2023, 7(18): 150-153, 158.
- [16] 贾荣浩, 魏国辉, 赵文华, 等. 基于K-近邻算法的中药化合物寒热平药性预测研究 [J]. 中华中医药杂志, 2023, 38(4): 1522-1525.
Jia R H, Wei G H, Zhao W H, et al. Study on the prediction of cold-hot-neutral natures of Chinese herbs compounds based on K-nearest neighbor algorithm [J]. China J Tradit Chin Med Pharm, 2023, 38(4): 1522-1525.
- [17] 张冰冰, 朱琦, 张晶新, 等. 基于深度学习的中药寒热属性预测研究 [J]. 世界中医药, 2023, 18(21): 3047-3052, 3059.
Zhang B B, Zhu Q, Zhang J X, et al. Prediction of cold/hot properties of Chinese herbal medicines based on a deep learning approach [J]. World Chin Med, 2023, 18(21): 3047-3052, 3059.
- [18] 郭文龙, 罗熊, 姜惠娟, 等. 基于TF-IDF算法的方剂构成相似度可视化研究 [J]. 中国中医药信息杂志, 2019, 26(7): 104-108.
Guo W L, Luo X, Jiang H J, et al. Visualization study on approximate degree analysis of prescription properties based on TF-IDF algorithm [J]. Chin J Inf Tradit Chin Med, 2019, 26(7): 104-108.
- [19] 王慧敏, 龚庆悦, 胡孔法, 等. 基于层次聚类的失眠处方用药分析 [J]. 计算机时代, 2020(3): 28-31.
Wang H M, Gong Q Y, Hu K F, et al. Analyzing insomnia prescription drugs with hierarchical clustering [J]. Comput Era, 2020(3): 28-31.
- [20] 吴地尧, 章新友, 张玉娇, 等. 分类算法在中药研究中的应用及其进展 [J]. 科学技术与工程, 2019, 19(35): 1-9.
Wu D Y, Zhang X Y, Zhang Y J, et al. Application and progress of classification algorithm in traditional Chinese medicine research [J]. Sci Technol Eng, 2019, 19(35): 1-9.
- [21] 张文青, 赵珊, 钱文秀, 等. 基于人工神经网络和支持向量机算法的药源性急性间质性肾炎计算模型的建立及在中药域的应用 [J]. 中草药, 2023, 54(2): 416-424.
Zhang W Q, Zhao S, Qian W X, et al. Establishment of computational models for drug-induced acute interstitialnephritis based on algorithms of artificial neural network and supportvector machine and their applications in traditional Chinese medicine [J]. Chin Tradit Herb Drugs, 2023, 54(2): 416-424.
- [22] 姜皓, 张冰, 张晓滕, 等. 基于4种机器学习算法的妊娠期中药 "禁忌慎" 判别 [J]. 中草药, 2021, 52(24): 7596-7605.
Jiang H, Zhang B, Zhang X M, et al. Discriminant of "prohibiting, avoiding or using caution" in traditional Chinese medicine during pregnancy based on four machine learning algorithms [J]. Chin Tradit Herb Drugs, 2021, 52(24): 7596-7605.
- [23] 雷蕾, 李小阳, 刘骏, 等. 化学功效组学: 中药现代研究的新方向: 以止血药和活血化瘀药核心成分群研究为例 [J]. 中国中药杂志, 2022, 47(24): 6803-6809.
Lei L, Li X Y, Liu J, et al. Chemical funtiomics: A new direction for modern research on traditional Chinese medicine: A case study of core component group of hemostatic and blood-activating/stasis-resolving Chinese herbal drugs [J]. China J Chin Mater Med, 2022, 47(24): 6803-6809.
- [24] 孙飞, 吴相亲, 戚悦, 等. 基于偏最小二乘算法探究山楂和焦山楂消食健脾功效成分 [J]. 中国中药杂志, 2023, 48(4): 958-965.
Sun F, Wu X Q, Qi Y, et al. Application of partial least squares algorithm to explore bioactive components of crude and stir-baked hawthorn for invigorating spleen and promoting digestion [J]. China J Chin Mater Med, 2023, 48(4): 958-965.
- [25] Ouyang X J, Li J Q, Zhong Y Q, et al. Identifying the active ingredients of carbonized *Typhae Pollen* by spectrum-effect relationship combined with MBPLS, PLS, and SVM algorithms [J]. J Pharm Biomed Anal, 2023, 235: 115619.
- [26] Kong W J, Zhao Y L, Xiao X H, et al. Spectrum-effect relationships between ultra performance liquid chromatography fingerprints and anti-bacterial activities of *Rhizoma coptidis* [J]. Anal Chim Acta, 2009, 634(2): 279-285.
- [27] Chen C, Li S X, Wang S M, et al. A support vector machine based pharmacodynamic prediction model for searching active fraction and ingredients of herbal medicine: Naodesheng prescription as an example [J]. J Pharm Biomed Anal, 2011, 56(2): 443-447.
- [28] 薛启隆, 苗坤宏, 于洋, 等. 基于深度强化学习的中药制药过程自主优化决策方法研究 [J]. 中国中药杂志, 2023, 48(2): 562-568.
Xue Q L, Miao K H, Yu Y, et al. Methodology for adaptive decision: Making research on manufacturing process of traditional Chinese medicine based on deep reinforcement learning [J]. China J Chin Mater Med, 2023, 48(2): 562-568.
- [29] 仲怿, 范骁辉, 李振皓. 基于精益理念的中药制药工业

- 数据挖掘及生产工艺持续改进策略研究: 以去壁灵芝孢子粉制药过程为例 [J]. 中国中药杂志, 2023, 48(3): 829-834.
- Zhong Y, Fan X H, Li Z H. Lean strategy for data mining and continuous improvement of Chinese pharmaceutical process: A case study of sporoderm-removal *Ganoderma lucidum* spore powder [J]. China J Chin Mater Med, 2023, 48(3): 829-834.
- [30] 吴冲, 谭超群, 黄永亮, 等. 基于深度学习算法的川贝母、山楂及半夏饮片的智能鉴别 [J]. 中国实验方剂学杂志, 2020, 26(21): 195-201.
- Wu C, Tan C Q, Huang Y L, et al. Intelligent identification of *Fritillariae Cirrhosae Bulbus*, *Crataegi Fructus* and *Pinelliae Rhizoma* based on deep learning algorithms [J]. Chin J Exp Tradit Med Form, 2020, 26(21): 195-201.
- [31] 王利娟, 赵杨, 吴天晨, 等. 基于Apriori算法及无监督分类方法探析孟河医派治疗头痛用药规律 [J]. 山西中医, 2023, 39(9): 54-57.
- Wang L J, Zhao Y, Wu T C, et al. Based on Apriori algorithm and unsupervised classification method, the regularity of Menghe medical school in treating headache was analyzed [J]. Shanxi J Tradit Chin Med, 2023, 39(9): 54-57.
- [32] 秘红英, 魏聪, 李红蓉, 等. 基于贝叶斯算法的吴以岭教授治疗冠心病心绞痛病案的病案分析研究 [J]. 中国中西医结合杂志, 2018, 38(5): 534-538.
- Bei H Y, Wei C, Li H R, et al. Bayesian-algorithm-based study of coronary heart disease angina pectoris cases treated by professor Wu Yi-ling [J]. Chin J Integr Tradit West Med, 2018, 38(5): 534-538.
- [33] 马雪娇, 王学谦, 吴晋芳, 等. 基于Apriori算法和因子分析挖掘肺癌分子靶向治疗中医用药配伍规律 [J]. 西部中医药, 2022, 35(7): 53-57.
- Ma X J, Wang X Q, Wu J F, et al. Medication rules of TCM therapy combined with molecular targeted therapy for lung cancer based on apriori algorithm and factor analysis [J]. West J Tradit Chin Med, 2022, 35(7): 53-57.
- [34] 吴丽娜, 谢雁鸣, 刘垲, 等. 基于关联规则Apriori算法的真实世界中结肠恶性肿瘤患者的中西药联合应用特征研究 [J]. 中药药理与临床, 2019, 35(5): 138-143.
- Wu L N, Xie Y M, Liu H, et al. Real-World Study on the characteristics of combined application of Chinese and Western medicine in the treatment for colonic malignant tumors by association rules analysis [J]. Pharmacol Clin Chin Mater Med, 2019, 35(5): 138-143.
- [35] 席俊羽, 魏瑞丽, 谢雁鸣, 等. 基于louvain算法的生脉注射液治疗心肌梗死的真实世界优效方案研究 [J]. 中药药理与临床, 2022, 38(4): 131-136.
- Xi J Y, Wei R L, Xie Y M, et al. Real-world optimal efficacy of Shengmai (生脉) injection in the treatment of myocardial infarction: Based on Louvain algorithm [J]. Pharmacol Clin Chin Mater Med, 2022, 38(4): 131-136.
- [36] 蒋程, 丁静, 孙云峰, 等. 基于偏最小二乘-判别分析算法的参麦注射剂不良反应关键影响因素分析 [J]. 中国临床药理学杂志, 2017, 26(4): 221-225.
- Jiang C, Ding J, Sun Y F, et al. Analysis on critical effect factors of adverse drug reaction induced by Shenmai Injection based on partial least squares discriminant analysis algorithm [J]. Chin J Clin Pharm, 2017, 26(4): 221-225.
- [37] Gan X, Shu Z X, Wang X Y, et al. Network medicine framework reveals generic herb-symptom effectiveness of traditional Chinese medicine [J]. Sci Adv, 2023, 9(43): eadh0215.
- [38] 王朴, 胡敏, 潘素跃, 等. 基于CIBERSORT算法分析皮肤肌炎免疫浸润机制与中药治疗靶点预测 [J/OL]. 中国免疫学杂志, (2022-12-23)[2024-03-21]. <https://kns.cnki.net/kcms/detail/22.1126.R.20221221.1839.004.html>
- Wang P, Hu Min, Pan S Y, et al. Analysis of immune infiltration mechanism of dermatomyositis and prediction of therapeutic targets of traditional Chinese medicine based on CIBERSORT algorithm [J]. Chin J Immunol, (2022-12-23) [2024-03-21]. <https://kns.cnki.net/kcms/detail/22.1126.R.20221221.1839.004.html>.
- [39] 刘洋, 曹赛雅, 冯月娇, 等. 应用机器学习和神经网络模型识别结肠癌"炎癌转化"过程的关键基因及防治中药预测 [J]. 中草药, 2023, 54(19): 6386-6399.
- Liu Y, Cao S Y, Feng Y J, et al. Identification of key genes in "inflammation-cancer transformation" process of colorectal cancer and prediction of traditional Chinese medicines for prevention and treatment through machine learning and neural network models [J]. Chin Tradit Herb Drugs, 2023, 54(19): 6386-6399.
- [40] 杨艳伟, 胡文元, 林志, 等. 大数据和人工智能技术用于计算机辅助药物设计的研究进展 [J]. 药物评价研究, 2023, 46(6): 1369-1375.
- Yang Y W, Hu W Y, Lin Z, et al. Research progress in big data and artificial intelligence techniques for computer-aided drug design [J]. Drug Eval Res, 2023, 46(6): 1369-1375.