数字中医中药

DOI:10.13193/j.issn.1673-7717.2026.02.003

中医药领域机器学习研究的现状与发展趋势探讨

吴文玉[1,2],詹少锋[1,2],焦欣[3],蓝婉宁[1,2],吴鹏[4],王凯[1,2]
(1.广州中医药大学第一临床医学院,广东 广州 510405;2.广州中医药大学第一附属医院,广东 广州 510405;
3.广州中医药大学第二临床医学院,广东 广州 510405;4.江苏省中医院,江苏 南京 210029)

摘要:机器学习(machine learning,ML)作为人工智能的核心驱动力,在信息化迅猛发展的背景下,为中医药的发展带来了机遇与挑战。其与中医药研究的结合,显著推动了中医病证体系的标准化、高效化和智能化进程,但目前仍处于探索阶段。从古籍文献、名医经验、中医辨证诊断、疾病预测以及中药和针灸等领域,对机器学习在中医药领域的研究进展进行概述,并探讨其未来发展趋势,以期为机器学习等智能技术在中医药研究中的应用提供新的方法论思路。

关键词:机器学习;人工智能;中医药
中图分类号:R2-03 **文献标志码**:A **文章编号**:1673-7717(2026)02-0011-08

Exploring Current Status and Development Trends of Machine Learning Research in Traditional Chinese Medicine Field

WU Wenyu[1,2],ZHAN Shaofeng[1,2],JIAO Xin[3],LAN Wanning[1,2],WU Peng[4],WANG Kai[1,2]
(1. College of First Clinical Medical,Guangzhou University of Chinese Medicine,Guangzhou 510405,Guangdong,China;
2. The First Affiliated Hospital of Guangzhou University of Traditional Chinese Medicine,Guangzhou 510405,Guangdong,China;
3. The Second Clinical Medical School of Guangzhou University of Chinese Medicine,Guangzhou 510405,Guangdong,China;
4. Jiangsu Provincial Hospital of Traditional Chinese Medicine,Nanjing 210029,Jiangsu,China)

Abstract:Machine learning,as the core driving force of artificial intelligence,brings opportunities and challenges to the development of traditional Chinese medicine in the context of rapid development of information technology. Its combination with traditional Chinese medicine research has significantly promoted the standardization,efficiency and intelligence process of traditional Chinese medicine disease syndrome system,but it is still in the exploration stage. In this paper,it provided an overview of the research progress of machine learning in the field of traditional Chinese medicine from the fields of ancient literature,experience of famous doctors,syndrome-based diagnosis,disease prediction,as well as traditional Chinese medicine and acupuncture,and discussed its future development trend,with a view to providing new methodological ideas for the application of intelligent technologies,such as machine learning,to the study of traditional Chinese medicine.

Keywords:machine learning;artificial intelligence;traditional Chinese medicine

机器学习(machine learning,ML)是人工智能的核心组成部分,赋予机器高效执行特定任务的能力。其基本原理是通过对大量数据的分析和学习,使机器能够识别模式、提取信息,并将所获得的知识应用于实际问题的分析与解决[1]。近年来,随着大数据、信息化和智能化的迅速发展,数据量的激增为机器学习提供了丰富的资源。同时,多学科的交叉融合也推动了智能信息技术在各领域的应用,特别是在中医药研究方面引发了广泛关注和探索。中医药拥有悠久的历史和丰富的理论体

基金项目:国家自然科学基金项目(82574938);国家中医药传承创新中心科研专项(2022QN24,2022QN27);国家中医优势专科(广州中医药大学第一附属医院肺病科)建设项目广东省普通高校重点领域专项(2022ZDZX2014);广东省研究生教育创新计划项目(2022ANLK025);广东省中医药科研项目(20231169,20251124);广东省自然科学基金项目(2024A1515012160);"广东特支计划"省卫生健康委卫生健康人才项目(0720240224);广东省医学科学技术研究基金项目(B2023051);广东省重点科室(中西医协同科室)建设项目;深圳市宝安区医疗卫生科研项目(2023JD124,2023JD107,2023JD105,2023JD110,2024JD289,2024JD293,2024JD316);深圳市宝安区医学会医疗卫生科研项目(BAYXH2024011);深圳市中西医结合医院院内项目(YJ-2023-106,YJ-2023-104,YJ-2023-115)
作者简介:吴文玉(1996-),男,广东韶关人,主治中医师,博士在读,研究方向:中医药相关研究。
通讯作者:王凯(1988-),男,山东淄博人,主治中医师,博士,研究方向:中西医结合诊治内科相关疾病。E-mail:827108317@qq.com。

系,其诊疗方法强调个体化和整体观念。然而,传统中医在现代化进程中面临着许多挑战,如病证体系的标准化、有效性评估等。机器学习作为一种新兴技术,能够有效地整合和分析来自中医古籍、临床数据及其他相关资料的信息,帮助研究人员更好地理解和运用中医理论。通过构建智能辨证诊疗模型,机器学习不仅可以辅助医生进行疾病的诊断和治疗方案的制定,还能提升中医药的研究效率和准确性。例如,借助机器学习算法,研究人员能够从大量病例中发现潜在的关联性和规律,从而揭示不同症状与疾病之间的联系。这种数据驱动的方法使得中医药的诊断过程更加科学化和系统化,减少了主观因素的干扰。此外,机器学习还可以应用于药物研发,通过分析药物与疾病的相互作用,为新药的发现与验证提供支持[2]。尽管机器学习在中医药研究中展现出广阔的前景,但仍面临一些挑战。首先,数据的质量和数量直接影响模型的效果,中医临床数据的标准化和规范化亟待加强。其次,中医理论的复杂性和多样性使得机器学习模型的构建和解释变得困难。因此,如何平衡传统中医理论与现代技术手段之间的关系,是未来研究的重要课题。机器学习在中医药研究中的应用,标志着中医药现代化发展的新阶段。随着技术的不断进步和研究的深入,未来机器学习有望在中医药领域发挥更大的作用,促进中医药理论的创新与发展。本文将进一步探讨机器学习在中医药研究中的具体应用实例以及未来的发展趋势,期待为这一领域的研究者提供有益的启示和参考。

1 机器学习及相关算法简介

机器学习作为人工智能的重要引擎,是一种分析多维数据的技术,并从中构建具有自主学习能力的复杂系统[3]。应用ML是一个半自动化的过程——通过向计算机提供数据,基于学习框架创建复杂的分析模型,以细化和优化预测的准确性,并在不被明确编程的情况下不断提高自身准确性[4]。在ML中,有4种常用的学习算法:监督、无监督、半监督和强化学习[5],用以处理分类、回归、聚类及降维4类任务[6]。

线性回归(Linear regression)是最经典的回归算法[7],只有一个自变量的情况称为简单回归,大于一个自变量情况的叫作多元回归。K-近邻算法(K nearest neighbors,KNN)诞生于1967年,是一种用于分类和回归的非参数统计方法[8]。ID3[9]、CART[10]、C4.5[11]是决策树的3种典型实现,出现于80年代到90年代初期,因算法结构简单,模型具有较好的可解释性。之后,受大脑层次结构的启发,1986年诞生了用于训练多层神经网络的反向传播算法[12],这是现在的深度学习中仍然被使用的训练算法,奠定了神经网络走向完善和应用的基础。此后研究人员一直致力于探究多层神经网络,如LECUN Y等[13]设计出第一个卷积神经网络(Convolutional neural network,CNN),用于手写数字的识别。随后,浅层学习模型包括传统的隐马尔可夫模型(Hidden markov model,HMM)、条件随机场(Conditional random field,CRF)、最大熵模型(MaxEnt)和支持向量机(Support vector machine,SVM)等算法逐渐兴起,神经网络则黯然失色。2012年,Alex的成功[14]使得深度神经网络卷土重来,此后,伴随着循环神经网络,图神经网络等诸多深度学习算法的涌现,深度学习(Deep learning,DL)广泛应用于机器视觉、语音识别等领域,逐渐成为ML领域最热门的研究方向。

2 机器学习在中医药领域的应用

2.1 机器学习在古籍文献挖掘中的应用

医学古籍作为中医药文化的重要载体,承载着医家丰富的诊疗经验,然而其内容繁杂,单纯依靠人力处理需要付出大量时间,成本昂贵。机器学习能够对复杂高维的数据分析处理,在古籍文献挖掘、整理的研究中受到广泛应用。

罗粤铭[15]对古籍中肾系水肿深入挖掘,应用深度学习等方法,挖掘六味地黄丸的深层联系,对方中茯苓进行关联学习,并推导出药对(茯苓-山药-五味子)及推演发现推荐方茯菟丸,拓宽了对六味地黄丸的隐层知识学习。王倩[16]应用决策树、SVM等算法构建白头翁汤、葛根芩连汤、黄芩汤三方一药味分类模型,并应用人工神经网络(Artificial neural network,ANN)及线性回归算法对葛根芩连汤各主要药效组分含量进行预测建模,通过多种机器学习方法厘清三方临床用量情况及剂量规律,解决了历代医家对《伤寒论》原方用量模糊,剂量上重剂、轻剂并存等问题。孙燕[17]基于SVM原理,应用径向基函数解决中医分类识别问题,创建了中医SVM(Chinese medicine support vector machines,CMSVM)机器学习系统,就《伤寒论》方证依据八法分类建模,为中医方证分类表达提供新思路、新方法。

通过应用机器学习等智能技术手段,能够对古籍文献数据自动化处理,探寻隐层关联知识,提高古籍知识库的构建效率及应用于临床的服务水平[18]。但也面临年代久远、古籍文献质量偏低、中医术语缺乏标准化处理等诸多问题。

2.2 机器学习在名医经验总结中的应用

名老中医学术思想是中医药知识库中的宝贵财富,其经验传承多基于书籍查阅或跟师学习。然而传统学习方法存在效率低、创新性不足、易受环境影响等问题,传承什么和怎么传承成为中医药发展现代化的两大难题[19-20]。如今,在人工智能技术的推进下,机器学习方法成为名医学术思想传承的重要工具。

肖晔[21]基于方剂学巨擘陈潮祖教授"五经五纬"辨证理念,应用决策树、ANN等多种学习算法根据纳入医案建模学习,得到方证对应模型,并通过ANN、比较规则等方法,获得症状权重系数,在此基础上,应用辨证模型分析症状和方剂的相关性。庞博[22]从施今墨学术流派专家临床医案入手,采用跟师学习与机器学习相结合,群案分析与个案分析相结合,监督学习(贝叶斯网络、SVM)与无监督学习(聚类分析、关联规则)相结合,定性研究(文献研究、深度访谈、焦点组访谈)与定量研究相结合的方法,将施今墨学术流派及其代表医家诊治糖尿病经验转化成定量表示的知识,以量化施今墨学术流派辨病、辨证、辨症论治方法的规律与经验,探索中医经验向知识再向证据转化的多元化途径。

2.3 机器学习在中医辨证诊断中的应用

辨证是在中医理论指导下,基于四诊合参,对患者临床资料进行整合,从而对疾病当前阶段的本质作出判断[23]。由于

个人的主观性,对疾病证型的判别易受辨证人员自身条件影响[24],从而影响对证型的精准判断。辨证在机器学习领域属于分类问题[25],机器学习可以通过其自主学习能力,针对病症繁杂、非线性关系的中医临床数据进行科学处理,充分发挥舌诊、脉诊优势,构建智能辨证模型。

2.3.1 中医疾病辨证分型 近年来,利用机器学习和深度学习技术在中医证候分类和诊断中的应用逐渐增多。刘丽蓉等[26]基于SVM及ANN算法构建荨麻疹证候分类模型,并发现在样本含量较小的情况下,SVM模型准确率优于ANN模型。LIU G P等[27]采用深度学习和多标签学习来构建中医慢性胃炎的证候诊断模型,结果表明,深度学习可以提高证候识别的准确性。HU Q N等[25]采用7326份中医临床病历数据集中的非结构化文本作为输入,将辨证建模为文本分类任务,基于CNN算法构建中医阴虚阳虚辨证模型。这些研究表明,先进的算法在中医证候分析中具有良好的应用前景。

2.3.2 中医舌诊脉诊 机器学习在中医舌象和脉象分析中的应用取得了积极进展。刘婧玮[28]基于CNN算法,自动提取舌象图片特征信息,实现舌象分类,通过使用数据扩增技术和迁移学习,防止模型过拟合,并通过同种方法训练腻苔分类模型,建立了齿痕舌、腻苔舌象辨识智能模型。陈松晔[29]在中医脉诊原理的基础上,提出基于CNN和无阂值递归图的脉象信号分析研究方法,建立了适用于常规29种脉象的脉象信息集成分类模型,研究实现了对脉象的分析识别,促进脉诊的客观化。以上研究证明,基于机器学习方法,能够探寻证候的内在联系及本质规律,助力脉诊、舌诊标准化进程,为精准辨证提供强有力的技术支持,但建立更全面、客观的智能辨证系统,仍需更深入研究。

2.4 机器学习在中医疾病预测中的应用

随着现代医学技术的高速发展,人们对疾病的态度正经历着由治病为中心到以健康、预防为中心的转变[30-31]。“治未病”是中医防治疾病的重要原则[32],《黄帝内经》最早提到:“上工治未病,不治已病,此之谓也。”在疾病防治过程中,构建疾病风险预测模型,评估筛选高风险患者群,并采取针对性干预措施,能够大大降低疾病的发生率或病重率。目前,西医已开展大量疾病预测模型相关研究,针对中医疾病预测模型研究较少。

LIN S等[33]认为面部望诊特征与冠状动脉疾病的风险增加有关,因此开发并验证了一种基于面部照片检测冠状动脉性心脏病(Coronary heart disease,CAD)的深度学习模型,该技术有望在门诊进行CAD概率预测试或在社区进行CAD筛查。佟旭等[34]基于多标签机器学习方法,应用SVM、多标签最近邻(ML-KNN)、自组织增强等多种算法构建糖尿病肾病多标签风险评估模型。李军等[35]将糖尿病基本生理指标与舌象参数相结合,应用ANN,SVM等4种算法,建立风险预测模型,并在研究中发现糖尿病患者的舌色偏黯,舌苔偏厚,构建的糖尿病前期分类模型效能优良,能够为糖尿病的前期检测服务。王梓民[36]采用贝叶斯网络算法,建立治未病检测平台,由患者将自身情况输入数据库,通过智能模型对患者信息进行分析对比,提醒患者相关注意事项并及时就医。

借助现代智能技术手段,能够在基于中医理论的指导下,对人体健康状态进行分析评估,从而高效地对疾病早期辨识干预,促进“治未病”理念的普及[37]。

2.5 机器学习在中药中的应用

中药是立足于中医理论框架,用于预防、治疗、诊断疾病并有康复和保健作用的药物[38]。然而,其数据复杂、语意模糊、标准滞后,使中药进一步推广研究受到诸多限制。由于机器学习能够模拟人类自主学习并处理目标任务,探究多维数据的本质规律,越来越多的研究人员将机器学习方法融合应用于中药研究中,并且成绩斐然。

2.5.1 中药质量控制 中药鉴别面临着品种众多和市场真伪并存的问题,传统方法往往依赖大量人工,成本较高。为此,吴冲[39]利用中药材图像数据集,提出依据注意力机制的轻量级中药识别网络模型(Attention-TCM-Net)和应用深度学习方法建立的基于自适应注意力机制的中药特征描述网络模型,研发了一种新型智能化中药鉴别系统,为中药鉴别方法提供了一种新的技术手段,显著提升了鉴别的效率和准确性。总之,这项研究为中药鉴别提供了有力的技术支持,推动了中药行业的发展。

2.5.2 中药有效成分筛选 中药组方有效成分的筛选依然面临较大困难。为了解决这一问题,鲁路[40]在肾衰营养胶囊有效成分筛选的研究中,通过分子对接联合RF、SVM等4类算法模型,对分子—靶点活性进行分类判断,预测可能存在的有效组方活性成分,提高药物靶点预测特异性。马磊等[41]在黄芩提取物药效质量研究中,基于ANN算法,将颜色、味道、成分含量3种指标相结合与体外抑菌率构建预测模型,发现上述指标能快速准确地预测体外抑菌率,从而科学地评价黄芩提取物整体质量,这些研究为中药成分筛选和质量评价提供了新的方法和思路。

2.5.3 中药配伍及方剂分类 目前,从草药配方中挖掘核心药物的方法主要有3种:①使用草药的频率[42];②关联规则[43];③聚类分析[44]。上述方法都只关注药材本身进行药材评价,而未能在临床处方中有效地利用症状与药材的关联。为此,BU D等[45]提出FangNet平台,首先使用PageRank算法构建症状-草药网络,之后在交互式可视化中提供了3种类型的草药隐藏知识,包括草药重要性等级、草药共现和互斥性、与特定症状的关联,用于从一组经验处方中挖掘草药隐藏知识。王鹏等[46]基于决策树算法,以方剂对应的证候作为分类标签,对古代胸痹心痛方归类建模,并发现相较于朴素贝叶斯算法,决策树在分类精度上有更大潜力,这些研究为草药的挖掘和应用提供了新的思路和工具。

2.5.4 中药疗效预测 夏赫通过Light GBM、SHAP算法构建两轮预测模型,对临床数据拟合并对重要特征筛选,从而对新冠中药疗效进行预测。结果表明,当患者白蛋白、球蛋白及淋巴细胞计数处于异常值时,相较于西药,中药及中成药能够有效降低患者插管死亡率[47],这一研究提供了中药在新冠治疗中的重要参考依据。

2.5.5 中药不良反应筛查 如今,中药潜在毒性成分的早期筛查仍然是中药新药研发的一大挑战[48],有学者利用树形算法,构建能够对中药成分肝毒性具有预测能力的定量构效关系模型,为药物研发提供肝毒性评估,减轻后期药物研发成本及使用风险[49]。吴淇慧[50]在中草药肝毒性的预测模型中,应用ANN、RF、SVM及KNN 4种算法并结合不同的分子指纹,初步构建单一的分类模型,并通过整合表现最佳的单一模型,得到对应的组合模型,较单一模型,组合模型能够降低噪音并提高预测能力,为预测中草药肝毒性提供新策略。机器学习方法在中药研究领域方兴未艾,能够应用于中药配伍、鉴定、有效成分及药物毒性筛选等多个领域,但现阶段研究多应用传统机器学习算法,对深度学习算法的应用研究仍较少。

2.6 机器学习在针灸中的应用

针灸学是基于中医理论指导,研究经络腧穴、灸刺技术及治疗方法,探究针灸诊疗防病规律,并阐明其机理的一门学科[51]。将机器学习技术引入针灸领域,能够更好地发挥针灸学优势,促进临床数据的收集应用、针灸治疗手法标准化、疗效客观化分析,为精准针灸医疗提供技术支持。

2.6.1 针灸训练方法 机器学习在针灸训练亦有可用武之地。刘涛涛等[52]认为,针灸学重视临床实践操作,对针刺方法的训练,强调对穴位的准确定点、针刺角度的精准判断,进针力度的严格控制,但在实验教学中,针刺训练多基于书本知识,在纸板、动物皮革、海绵垫等模型上操作,难以对穴位准确把握,并针对性提出一种新型针灸训练系统,采用基于临床实际的硅胶手臂模型,应用ANN算法构建分类模型,对针刺手法进行标准化判断,从而提高针刺训练效率。

2.6.2 疾病辨证选穴 疾病辨证选穴在针刺治疗疾病过程中至关重要,这将直接影响到疾病的疗效。谢天宇[53]在膝骨关节炎针灸智能辨证选穴的研究中,基于患者诊疗及针灸选穴数据库,首先采用无监督学习法(因子分析、主成分分析)构建症状—证型辨证模型,再运用Apriori关联规则算法探究证型—穴位的关联规律,将二者结合,并最终得出“症状—证型—穴位”的智能辨证选穴算法模型,为针灸智能选穴提供技术参考。温川飙等[54]提出基于CNN算法,构建穴位敏化特征识别模型,并开展算法训练,得出穴位敏化的策略权重估值,为穴位敏化研究设计提供数据技术支持。基于精准的机器学习算法,将更好地测算出腧穴的选择,为患者提供更好的治疗疗效。

2.6.3 针灸疗效预测 疗效的评估可将治疗效果最大化,在实际生活中具有重要的指导意义。韦晓燕等[55]利用ANN算法处理大量非线性数据,建立针灸治疗颈椎病疗效评价模型,并在实验中得出,模型均能对针灸治疗颈椎病疗效预测,相较而言,BP神经网络(Backpropagation neural network,BP)预测能力更佳,可为针灸治疗颈椎病提供疗效评估指导。机器学习应用于针灸领域虽取得一定研究成果,但算法选择困难、模型缺乏解释性等问题[56-57],亟待解决。

3 机器学习在中医药领域研究热点及发展趋势

3.1 数据说明及软件配置

文献来源于CNKI中文数据库及Web of Science核心合集数据库,中文数据库以“(中医 + 中药 + 中医药 + 针灸) * (人工智能 + 深度学习 + 机器学习 + 神经网络 + 支持向量机 + 随机森林 + 决策树)”为检索式,英文数据库以“TS = (traditional Chinese medicine OR acupuncture) AND (artificial intelligen* OR machine learning OR deep learning OR neural network)”为检索式,时间段为2000—2022年。中文文献按照CiteSpace所需的Refworks格式予以导出,英文文献按照全记录与引用的参考文献格式予以导出,得到本文相关研究样本数据库,筛重并剔除不相关文献后共获得CNKI数据库文献1 337篇,Web of Science核心合集数据库文献474篇。

Time Slicing设置为From2000.01 TO 2022.12,Years Per Slice(时间切片)选择1,即时间跨度为每年分析1次,确保获取文献中的关键词能够被全部统计,节点类型(Node Types)选择Keywords,c、cc、ccv阈值保持默认状态,Pruning板块中勾选寻径网络法(Pathfinder)与裁剪单个切片的网络(Pruning sliced networks),确保节点数量不变,连线数量减少,具有完备性,有唯一解。最后点击GO,运行结束后弹出窗口,选择visualize,可视化图谱产生。

3.2 文献数量分布

机器学习方法在中医药领域研究文献分布情况如插页Ⅴ图1所示,从线形图可以看出CNKI数据库及Web of Science核心合集数据库从2000—2022年中英文文献总体呈稳步增长势头,中英文数据库线形图对比发现Web of Science核心合集数据库总体发文量较CNKI数据库少。如插页Ⅴ图1所示,从2000—2004年,文献发表数量较少,表明机器学习在中医药领域的研究处于萌芽阶段;从2004—2016年发文量在1~68篇波动,文献发表总体走势低平;2017—2022年发文量呈直线上升趋势,由于2022年文献还未补充完整,所以CNKI数据库2022年发文量比2021年降低。从文献总体走势可以看出机器学习在中医药领域稳步发展,并在近5年发展迅猛,机器学习方法应用于中医药领域研究越来越受到研究人员关注。

3.3 关键词分析及趋势探讨

关键词可以快速、准确地反映论文的主题和重点,它在研究领域文献中出现的频次能够反映某时间段内整个研究领域的热点话题,本文以CNKI数据库为例进行关键词分析,并探讨其发展趋势。关键词共现图谱能够通过共现原理分析纳入文献主题词之间的联系,其中线条代表关键词之间的联系,圆圈大小代表关键词聚集频次。关键词共现图谱如插页Ⅴ图2所示,关键词共涉及520个(N = 520),线段共有1027条(E = 1027),密度为0.007 6(Density = 0.007 6)。其中“人工智能”“机器学习”“数据挖掘”“中医”等关键主题出现的频次较高,除基本关键词,“神经网络”“中药”“支持向量机”“决策树”“中医证候”“舌诊”等关键词出现频次较高,为当前研究热点。关键词时区图能够从整体反映相关研究逐年关注度变化及趋势,如插页Ⅵ图3所示,2000年“机器学习”“人工智能”就有相关研究出现;从2002—2015年机器学习方法在中医药领域研究层出不穷,多采用“神经网络”“决策树”“支持向量机”“随机森林”等传统机器学习方法;到2016年,“深度学习”进入研

究人员视野,随后“大数据”也接踵而来,由此可预测“深度学习”“大数据”仍是未来的研究热点。关键词突现是指在特定时期某研究领域对于某个主题关注程度的变化情况,对其进行研究可以分析出该研究领域的前沿问题。如插页Ⅵ图4所示,CNKI数据库大部分的研究热点爆发阶段都维持在2~3年内,之后逐渐失去热度。近2~3年内明显的研究问题的倾向性为“人工智能”“深度学习”“针灸”“特征融合”“舌象”,可以看到研究热度持续到2022年,之后可以在此方向进行持续跟踪,并预测上述关键词将是机器学习在中医药领域研究的未来趋势。

4 机器学习在中医药领域研究面临的挑战

机器学习方法在中医药研究中的应用,极大促进中医药研究标准化、高效化、智能化进程,但仍处于探索阶段。目前,机器学习在中医药领域的交融发展面临以下挑战:①学习算法选择困难,学习深度不够。随着人工智能研究不断深入,深度学习算法正在逐步取代传统的机器学习方法,其与中医诊疗复杂系统的运作模式具有相似性,因此可以很好模拟系统复杂性特征及功能[58]。但在中医药研究领域,深度学习多集中于个别领域[59],且缺乏解释性已成为限制其在中医药领域进一步应用的巨大障碍。现阶段中医药研究多集中于传统机器学习方法,针对中医数据维数高、信息量大等特点,浅层学习难以找到深层关联,且样本量过小,研究时限较短,易造成对数据过拟合[60]。②中医药临床数据的复杂性。中医药文化绵延千年,中医古籍文献、临床数据存在大量结构化和(或)非结构化数据,特征模糊,对特征提取难度较大。随着时代进步,人们对中医疗法的接受度普遍提高,中医诊疗疗效的预测分析显得尤为重要,但由于中医数据的复杂高维,目前针对中医优势病种疗效预测模型构建较少。中医学作为经验医学,其诊疗多基于医家自身的经验,对名老中医经验精准把握仍任重而道远。③中医药研究人员机器学习实践素养不足[61],多学科创新型人才缺乏。信息技术进步,推动医疗产业高速发展,各类大型数据库的建立为机器学习提供更便捷的平台。机器学习方法不仅为中医药事业发展提供循证医学客观数据支持,也为培养中医研究人员科研思维提供了新思路[62]。但现阶段中医药文化的传承发展多基于传统学习方法,相较于现代医学智能技术的相对成熟,中医药研究人员对智能技术的融合应用仍存在较大困难。

5 小结

综上,本文从古籍文献、名医经验、中医辨证诊断、疾病预测及中药针灸范畴对机器学习方法在中医药领域的研究进展概述,并探寻其发展趋势。机器学习方法运用于中医药研究中,是时代的选择,也是促使中医药传承创新发展的契机,智能化发展能够促进对海量中医数据的应用,最大程度发挥对中医优势病种高效诊疗优势,以期更好地服务于临床。从小样本、单一的传统机器学习方法到大数据、多种算法融合应用、深度学习机制深入探究的驱动转换是中医药智能化发展的下一个重要任务。未来,机器学习将作为重要引擎,推动祖国中医药研究事业高速发展。

参考文献

[1] 郇家铭, 陈晓晴, 杨雯晴, 等. 基于机器学习模型分析镇肝熄风汤治疗高血压病的配伍特征研究[J]. 科技导报, 2024, 42(21): 149-162.
[2] 李紫玥, 丛开源, 吴诗琪, 等. 人工智能技术在先导化合物发现与优化中的应用进展[J]. 药学学报, 2024, 59(9): 2443-2453.
[3] JORDAN M I, MITCHELL T M. Machine learning: trends, perspectives, and prospects[J]. Science, 2015, 349(6245): 255-260.
[4] PATEL V L, SHORTLIFFE E H, STEFANELLI M, et al. The coming of age of artificial intelligence in medicine[J]. Artif Intell Med, 2009, 46(1): 5-17.
[5] JAMES G, WITTEN D, HASTIE T, et al. An introduction to statistical learning[M]. New York: Springer, 2017.
[6] PEDREGOSA F, VAROQUAUX G E L, GRAMFORT A, et al. Scikit-learn: Machine Learning in Python[J]. Journal of Machine Learning Research, 2011, 12(85): 2825-2830.
[7] COX D R. The regression analysis of binary sequences[J]. J R Stat Soc Ser B Stat Methodol, 1958, 20(2): 215-232.
[8] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.
[9] CIOS K J, LIU N. A machine learning method for generation of a neural network architecture: a continuous ID3 algorithm[J]. IEEE Trans Neural Netw, 1992, 3(2): 280-291.
[10] SPEYBROECK N. Classification and regression trees[J]. Int J Public Health, 2012, 57(1): 243-246.
[11] SALZBERG S L. C4.5: programs for machine learning by J. ross quinlan. morgan kaufmann publishers, inc., 1993[J]. Mach Learn, 1994, 16(3): 235-240.
[12] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323: 533-536.
[13] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Comput, 1989, 1(4): 541-551.
[14] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Commun ACM, 2017, 60(6): 84-90.
[15] 罗粤铭. 基于肾系水肿古籍挖掘的茯苓类方药在糖尿病肾病中的应用研究[D]. 广州: 广州中医药大学, 2019.
[16] 王倩. 基于《伤寒论》热利三方医案文献的中药处方用量研究[D]. 北京: 北京中医药大学, 2015.
[17] 孙燕. 基于机器学习技术的《伤寒论》方证分析方法研究[D]. 北京: 北京中医药大学, 2007.
[18] 王国玺, 李兵, 张华敏, 等. 基于知识组织的医案古籍知识库的构建与思考[J]. 西部中医药, 2019, 32(9): 49-52.
[19] 王阶, 熊兴江. 名医经验传承探索之路[J]. 中医杂志, 2011, 52(7): 545-549.
[20] 刘英锋, 查青林, 黄利兴, 等. 对名医传承规律与传承模式研究的设想[J]. 中医药管理杂志, 2010, 18(12): 1070-1072, 1078.
[21] 肖晔. 陈潮祖医案“五经五纬”方证对应数字化模型研究[D]. 成都: 成都中医药大学, 2020.
[22] 庞博. 施今墨学派名老中医诊治糖尿病学术思想与经验传承研

究[D]. 北京：北京中医药大学，2012.
[23] JIANG M，LU C，ZHANG C，et al. Syndrome differentiation in modern research of traditional Chinese medicine[J]. J Ethnopharmacol，2012，140(3)：634－642.
[24] 杨涛，朱学芳. 中医辨证智能化研究现状及发展趋势[J]. 南京中医药大学学报，2021，37(4)：597－601.
[25] HU Q N，YU T，LI J H，et al. End－to－End syndrome differentiation of Yin deficiency and Yang deficiency in traditional Chinese medicine[J]. Comput Methods Programs Biomed，2019，174：9－15.
[26] 刘丽蓉，詹秀菊. 基于BP神经网络和支持向量机的荨麻疹证候分类探讨[J]. 广州中医药大学学报，2020，37(3)：573－577.
[27] LIU G P，YAN J J，WANG Y Q，et al. Deep learning based syndrome diagnosis of chronic gastritis[J]. Comput Math Methods Med，2014，2014：938350.
[28] 刘婧玮. 基于卷积神经网络的中医舌象辨识人工智能方法学研究[D]. 北京：北京中医药大学，2020.
[29] 陈松晔. 基于深度学习和集成学习的脉象信号分析识别研究[D]. 上海：华东理工大学，2020.
[30] 魏戍，谢雁鸣，田峰，等. 病证结合构建慢病风险预测模型的思路与方法[J]. 中国中医基础医学杂志，2017，23(2)：180－183.
[31] 马红丽，徐长英，杨新鸣. 决策树模型在中医药领域的应用现状[J]. 世界中医药，2021，16(17)：2648－2651，2656.
[32] 方锐，杨勇，任建萍，等. “中西医并重”视角下中医“治未病”与健康管理的对比及融合研究[J]. 世界科学技术－中医药现代化，2018，20(11)：1929－1935.
[33] LIN S，LI Z G，FU B W，et al. Feasibility of using deep learning to detect coronary artery disease based on facial photo[J]. Eur Heart J，2020，41(46)：4400－4411.
[34] 佟旭，杨纯，孟庆刚. 基于多标签机器学习的糖尿病肾病中医“同病异证”风险评估模型的构建[J]. 中华全科医学，2022，20(2)：181－185，227.
[35] 李军，胡晓娟，屠立平，等. 基于舌象参数与多指标特征联合的2型糖尿病风险预测模型[J]. 中国中医基础医学杂志，2021，27(3)：451－456，501.
[36] 王梓民. 基于病例大数据的中医“治未病”机器学习方法[J]. 电子技术与软件工程，2019(5)：161－163.
[37] 夏淑洁，杨朝阳，李灿东. 智能化中医“治未病”健康管理模式探析[J]. 中华中医药杂志，2019，34(11)：5007－5010.
[38] 钟赣生. 中药学[M]. 3版. 北京：中国中医药出版社，2012.
[39] 吴冲. 基于深度学习的中药材鉴别方法研究[D]. 成都：四川大学，2021.
[40] 鲁路. 基于机器学习优化分子对接筛选肾衰营养胶囊有效成分[D]. 广州：南方医科大学，2017.
[41] 马磊，王佳辉，郝宣润，等. 基于人工神经网络模型的黄芩提取物性状－成分与药效相关性研究[J]. 中草药，2020，51(8)：2151－2161.
[42] REN X，SHAO X X，LI X X，et al. Identifying potential treatments of COVID－19 from Traditional Chinese Medicine (TCM) by using a data－driven approach[J]. J Ethnopharmacol，2020，258：112932.
[43] YU J S，HO C H，HSU Y C，et al. Traditional Chinese medicine treatments for upper respiratory tract infections/common colds in Taiwan[J]. Eur J Integr Med，2014，6(5)：538－544.
[44] CHU S M，SHIH W T，YANG Y H，et al. Use of traditional Chinese medicine in patients with hyperlipidemia：a population－based study in Taiwan[J]. J Ethnopharmacol，2015，168：129－135.
[45] BU D，XIA Y，ZHANG J，et al. FangNet：Mining herb hidden knowledge from TCM clinical effective formulas using structure network algorithm[J]. Comput Struct Biotechnol J，2021，19：62－71.
[46] 王鹏，张静宇，石文，等. 基于决策树算法的古代方剂分类研究[J]. 信息与电脑(理论版)，2021，33(10)：45－48.
[47] 卢云，张梦月，夏赫，等. 基于LightGBM及SHAP对1055例新型冠状病毒肺炎重型患者中西医结合及西医治疗的多中心回顾性研究[J]. 北京中医药大学学报，2021，44(12)：1098－1107.
[48] 颜彩琴，范睿琦，宁雨坪，等. 深度学习模型在中药毒性预警中的应用和前景[J]. 中国药理学与毒理学杂志，2022，36(3)：231－240.
[49] 叶立，王新洲，朱永亮，等. 利用树模型预测中药成分的肝毒性[J]. 中国药学杂志，2014，49(18)：1583－1588.
[50] 吴淇慧. 中草药肝毒性预测模型的构建及其应用研究[D]. 广州：广州中医药大学，2018.
[51] 王华，杜元灏. 针灸学[M]. 3版. 北京：中国中医药出版社，2012.
[52] 刘涛涛，王志辉，张弓，等. 神经网络在针灸训练系统中的应用研究[J]. 计算技术与自动化，2020，39(3)：102－107.
[53] 谢天宇. 针灸治疗膝骨关节炎智能辨证选穴的算法模型研究[D]. 成都：成都中医药大学，2019.
[54] 温川飙，陈菊，贾伟，等. 基于深度学习的穴位敏化客观化方法研究[J]. 辽宁中医杂志，2017，44(8)：1723－1725.
[55] 韦晓燕，张洪来，魏航，等. BP和Elman神经网络评价针灸治疗颈椎病疗效研究[J]. 现代计算机(专业版)，2016，22(22)：3－6，12.
[56] 尹涛，何昭璇，孙睿睿，等. 机器学习在针灸研究领域的应用现状与展望[J]. 中国针灸，2020，40(12)：1383－1386.
[57] 梁吉，韩名媛，王承斌，等. 机器学习与针灸学领域结合的研究进展[J]. 针刺研究，2021，46(6)：460－463.
[58] 孙肇阳，吕桂娇，郭义，等. 中医诊疗复杂巨系统的人工神经网络建模原理[J]. 中华中医药杂志，2022，37(10)：5841－5844.
[59] 陈曦. 深度学习(DL)算法在中医药领域的应用探索[C]//第一届中国中医药信息大会论文集. 北京，2014：81－84.
[60] 关菀，马志龙，徐春，等. 深度学习技术在中医领域中的应用[J]. 中国卫生信息管理杂志，2022，19(2)：281－285，292.
[61] 刘大路，李静. 培养医学生机器学习的实践素养[J]. 基础医学与临床，2021，41(7)：1071－1075.
[62] 陈靖，许欣竹，龚杰. 新时期中医药创新创业人才培养模式探究[J]. 中国中医药现代远程教育，2022，20(5)：11－13.