

常见机器学习技术在中医证候诊断中的应用进展

柏 杨, 郝平生

成都中医药大学附属医院皮肤科, 四川 成都

收稿日期: 2024年2月20日; 录用日期: 2024年3月21日; 发布日期: 2024年3月29日

摘 要

在中医学中, 辨证诊断占据着核心地位, 它依托于对患者症状、体征的全面评估, 以及舌象、脉象等关键信息的分析, 以准确界定疾病的本质。辨证诊断过程面临的挑战包括证型识别的主观差异、证型变异的复杂度以及中医四诊信息的非标准化问题, 这些都极大地提升了对医生专业知识和经验的需求, 使得对新手医生而言, 实现“精准辨证”显得尤为艰难。随着人工智能技术的不断进步, 其与多学科的交叉融合日益加深, 并被广泛利用于探索中医证候的诊断研究之中。这种融合为中医的辨证诊断研究带来了新颖的方法与工具。在此背景下, 我通过文献回顾, 对中医证候研究中应用的主要机器学习技术进行了分类和总结评价, 分析了各类机器学习技术的优缺点, 并讨论了利用机器学习进行中医研究所遇到的问题及其在未来的应用潜力, 旨在为中医证候诊断领域内机器学习算法的深入研究与应奠定参考基础。

关键词

机器学习, 中医辨证, 证候诊断, 应用进展

The Application Progress of Common Machine Learning Techniques in Traditional Chinese Medicine Syndrome Diagnosis

Yang Bo, Pingsheng Hao

Department of Dermatology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu Sichuan

Received: Feb. 20th, 2024; accepted: Mar. 21st, 2024; published: Mar. 29th, 2024

Abstract

In Traditional Chinese Medicine (TCM), evidence-based diagnosis occupies a central position, re-

lying on a comprehensive assessment of the patient's symptoms and signs, as well as the analysis of key information, such as the tongue and pulse, in order to accurately define the nature of the disease. Challenges to the diagnostic process include the subjective differences in pattern recognition, the complexity of pattern variation, and the non-standardization of the four diagnostic information in TCM, all of which greatly increase the need for doctors' expertise and experience, making accurate pattern recognition particularly difficult for novice doctors. With the continuous progress of AI technology, its cross-fertilization with multiple disciplines is deepening, and has been widely utilized in diagnostic research to explore TCM evidence. This integration has brought novel methods and tools for TCM diagnostic research. In this context, I classify and summarize the main machine learning techniques applied in TCM diagnostic studies through a literature review, analyze the advantages and disadvantages of each type of machine learning technique, and discuss the problems encountered in using machine learning for TCM research and its potential for future application, with the aim of laying a reference foundation for the in-depth research and application of machine learning algorithms in the field of TCM diagnostic diagnosis.

Keywords

Machine Learning, Traditional Chinese Medicine Identification and Typing, Evidence-Based Diagnosis, Advances in Applications

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

“辨证论治”在中医学中占据着显著的地位，它不仅是中医诊断的核心，也是基于中医理论对疾病本质进行判断的关键思维过程。这一过程能够深入揭示疾病在某一特定阶段的病理变化，为药物的选择和治疗方案的制定提供理论依据。中医诊断涉及到从多个源头收集、处理和综合信息的复杂过程，其中证候的特征包括整体性、非线性、模糊性和隐蔽性。医生往往需要依据自己的主观经验来做出诊断判断，这一过程容易受到环境因素如光线和温度的影响，缺少客观的判断标准，导致辨证的准确性和可重复性较低，诊断结果易受到医生个人主观意识的影响。因此，如何将大量的中医信息进行有效整合和分析，揭示证候要素的内在规律，探究证候的本质，从而提升临床诊断效率，实现中医证候诊断的客观化和标准化，成为当前研究的难题和焦点。

机器学习作为人工智能领域的一个关键分支，旨在赋予计算机通过简化编程实现学习的能力，进而让计算机能够基于数据进行决策或进行预测。通过将机器学习技术应用于中医的辨证诊断中，分析海量的中医四诊等临床信息，可以仿真中医的临床诊断推理过程，揭示复杂的“病-证-症”关系，这有助于突破人类认知和思维的限制，充分体现中医的整体观念、模糊性和非线性特征，从而发现大数据背景下隐藏的中医知识和规律，有效地分析和解释中医证候[1]。这种方法不仅能够提升辨证诊断的精确度和效率，也能够为中医诊断过程提供科学、可量化和客观的评价标准。因此，本研究综述了机器学习技术在中医证候诊断领域的应用现状和成果，探讨了采用机器学习算法进行中医研究所遇到的挑战以及未来的应用潜力，目的是为中医证候诊断领域中机器学习算法的深入研究和应用提供理论支持和实践指导，同时为临床医生和科研人员提供一定参考。

2. 机器学习概述

机器学习技术作为一门跨学科的综合领域,旨在使计算机能够在无需明确编程指令的前提下,通过数据训练构建模型,模仿人类学习行为,赋予机器自主学习的能力[2] [3]。这一技术特别适用于那些对传统方法而言过于复杂,或至今未有解决算法的问题,通过分析积累的经验数据,揭示问题的内在本质和规律,从而不断提升性能,实现更优的结果。机器学习,作为人工智能的基础,既包括了传统的机器学习算法也覆盖了深度学习算法。从学习方式来看,传统机器学习算法可以分为监督学习、无监督学习、半监督学习和强化学习,涵盖了 Logistic 回归、主成分分析、多层感知机、支持向量机、随机森林、AdaBoost、k 近邻法、决策树、多核学习、核主成分分析、独立成分分析、贝叶斯算法、聚类算法等多种方法[4] [5] [6]。而深度学习技术则突破了人工特征构造的限制,无需深入理解领域知识,即可自动从原始数据中提取和构建特征,通过层层加工处理实现数据的深度分析。

在中医医疗领域,众多机器学习算法已经得到了广泛的应用。特别是监督学习模型,如随机森林算法、决策树、支持向量机、贝叶斯网络和人工神经网络等,它们在处理分类和归类的问题方面表现出显著的效果,并在辨证模型构建、疾病风险预测以及证素相关危险因素的研究中展现出巨大的应用潜力[7]。另一方面,无监督学习算法,如关联规则、聚类分析、主成分分析和信息熵等,能够在不依靠预定目标的情况下揭示数据内在的规律性,这类模型主要被用于数据处理的研究,以支持临床的诊断和治疗工作。在基于深度学习的研究领域内,卷积神经网络被普遍应用于医疗影像数据的分析,而循环神经网络则更多地用于处理电子病历数据。同时,自动编码器和深度置信网络也在医疗影像和电子病历数据的应用中发挥作用。这些技术的应用不仅提高了中医医疗领域的研究和诊疗效率,也推动了该领域向数据驱动的精准医疗转型。

3. 常见机器学习技术在中医辨证中的应用

从传统的统计学方法到机器学习,使用现代化的技术对疾病临床症状和体征进行辨析并试图诊断证型,逐渐成为新的研究领域。机器学习技术在中医证候研究中的应用,是现代科技与传统中医相结合的一个重要表现。这一领域的发展不仅反映了对中医传统知识的现代化探索,也代表了医学领域内对精准医疗和个性化治疗需求的响应。机器学习的引入,特别是在数据处理能力和模式识别方面的优势,为中医证候研究提供了新的视角和方法。

3.1. 应用的缘起与发展历程

将机器学习应用到中医辨证诊断中,最初的动因来自于将中医诊断过程量化、系统化的需求,旨在提高诊断的标准化和重复性,减少人为误差。在发展初期,人们尝试利用计算机技术进行数据收集和简单处理,如使用统计分析来识别证候与症状之间的相关性。随着机器学习和人工智能技术的发展,研究者开始利用这些技术进行更复杂的数据分析和模式识别,如使用决策树、随机森林、支持向量机、神经网络等算法来识别和预测证候。近年来,深度学习技术的引入使得对证候的识别和预测更为精准,特别是在处理图像(如舌象)和语音(如问诊)数据方面显示出了巨大的潜力。同时,大数据的利用也为中医证候的研究提供了更丰富的数据资源。

3.2. 与其他相关技术的对比

3.2.1. 传统统计方法

相比于传统的统计方法,机器学习技术能够学习和建模中医证候中大量症状和体征的非线性、高度复杂和多维度的关系,而传统统计方法在处理这种复杂性时可能会遇到困难。同时,机器学习模型能自

动地从大量临床数据中识别和提取关键特征, 显著降低了传统方法对手动特征选择和提取的依赖。这一点对中医证候分类至关重要, 因为有效的特征可能是细微和非直观的, 难以通过传统方法识别。此外, 机器学习的迅速发展带来了众多创新的算法和技术, 为预测和分类任务提供了灵活的方案, 开启了解决新挑战和优化现有方案的新途径。特别是在预测未知数据的能力方面, 经过适当训练和调整的机器学习模型能够在不断变化的环境中提供出色的性能, 这对于需要预测未来事件的应用至关重要。

3.2.2. 专家系统

机器学习算法能自动识别数据中的复杂模式, 无需预设规则, 这一特性使得机器学习模型能够揭示人类专家可能尚未发现的新模式和联系, 进而提升诊断准确率。随着医疗数据增加, 机器学习模型有效处理大规模数据, 提高预测准确性和可信度, 而专家系统可能因知识和规则限制难以适应。机器学习通过持续学习自我优化, 适应新数据, 保持高效预测, 对比下, 专家系统更新耗时且可能延迟。在中医证候分类中, 机器学习整合多源数据, 如症状、舌象, 提供精准预测, 考虑个体差异, 符合中医个体化治疗原则, 展现出比专家系统更高的灵活性和个性化。随数据增长, 机器学习模型能自我改进, 适应新任务, 而专家系统扩展则复杂耗时。

3.2.3. 其他人工智能技术

结合自然语言处理等其他人工智能技术, 机器学习在分析和理解中医古籍文献以及电子病历等非结构化数据方面显现出巨大的潜能。这种技术融合使得机器不仅能够识别和解析文本中的关键信息, 还能深入理解其语义内容, 极大地提高了从古籍和电子记录中提取有价值信息的能力。通过这种方式, 机器学习和自然语言处理技术共同为中医领域的数据分析和知识提炼提供了强大的支持, 使得处理和解析这些传统上难以自动化处理的非结构化数据成为可能。

3.3. 具体应用

K-最近邻是最简单的机器学习算法之一, 主要关注线性可分问题, 而随着数据的复杂性增加, 支持向量机通过引入核技术提供了处理高维数据的能力, 可处理非线性问题。随机森林作为集成学习的代表, 通过集成多个决策树提高了预测的稳定性和准确性。K-最近邻和支持向量机依赖于有效的手动特征提取, 而随机森林虽然减少了这一需求, 但在特征选择上仍有限制。神经网络, 尤其是深度学习, 通过自动特征学习完成从手动特征工程到自动特征学习的转变, 标志着机器学习向处理更复杂任务的演进。以下将详细介绍 K-最近邻、支持向量机、决策树、随机森林、人工神经网络这几种常见机器学习技术在中医辨证中的具体应用。

3.3.1. K-最近邻

K-最近邻(K-NN)算法, 作为一种基础而广泛使用的机器学习方法, 主要用于分类与回归任务。该算法的核心原理是识别一个给定样本周围的 k 个最近邻居, 并利用这些邻居的信息来做出预测。K-NN 算法的主要优势在于其理论模型的简洁性, 易于理解和执行。作为一种懒惰学习策略, 它避免了训练阶段的学习过程, 适用于既定的分类及回归问题。此外, 由于该算法的决策过程仅与邻近的局部数据集有关, 它对处理非线性数据集具有较好的适应性。

孙鑫亮[8]等通过应用多标记近邻算法 ML-KNN, 基于收集到的 1315 份冠心病患者的医案资料, 成功构建了冠心病的中医辨证模型。该模型平均汉明损失为 5.47%, 1-错误率为 2.13%, 覆盖率为 12.33%, 排序损失为 1.56%, 平均精度为 93.55%。这一结果表明, 使用 ML-KNN 算法从临床数据中构建的冠心病辨证模型能够有效地反映中医专家的诊断经验。另一方面, 贺鹏飞[9]等采用决策树、随机森林、极限梯度提升、支持向量机和 K 最近邻等五种算法构建慢性肾炎的中医智能辨证模型, 发现在这些模型中, K

最近邻算法的准确率最低, 为 62.2%。

K-NN 算法其简洁和直观的特性在中医证型的预测及疾病的分类上展现出了广阔的应用潜力。在中医证型分类中, K-NN 能够依据患者的临床特性匹配相近的案例, 进而辅助医生做出诊断。尽管这种方法在操作上简便且有效, 但它面临着计算复杂性高和处理大型数据集时效率较低的问题, 且对噪声数据和非相关特征敏感, 需要事先选择适当的 K 值, 而这个选择通常是基于交叉验证。为应对这些实践中遇到的挑战, 可能需要采取融合其他机器学习策略或对算法本身进行优化的措施。

3.3.2. 支持向量机

支持支持向量机(Support Vector Machine, SVM)是由 Vapnik 和 Cortes 于 1995 年提出的一种机器学习模型概念[10], 其主要思想是旨在通过构建一个最优化的超平面来有效区分不同类别的数据点, 实现精确的分类效果。SVM 遵循的是结构风险最小化原则, 即机器判定的结果与实际结果发生错误概率的上界尽可能最小化。该模型一定程度上解决了“过学习”和“维数灾难”等传统机器学习存在的问题, 具有运算速度快、适应性良好, 结构简单等诸多优势。

孙资金等[11]建立基于人工智能的中风中医辨证模型, 结果发现 SVM 模型的准确率最高, 可达 95%, 显示出 SVM 在中医诊断和预测中的高效能力。顾天宇[12]等通过比较 SVM、BP 神经网络和 GBDT 在中风病中医证候分类中的应用, 发现 SVM 模型以 86% 的准确率领先, 表示相较于其他分类器模型, 基于支持向量机建立的模型在中风病中医证候分类上更具优势, 突显了其在复杂病症分类中的优越性。此外, 刘丽蓉等[13]的比较研究也证明了 SVM 在荨麻疹证候分类上以 92.3% 的准确率优于 BP 神经网络, 进一步印证了其优势, 提示利用支持向量机分类器进行荨麻疹证候分类均可取得较好的结果。

在中医证型分类中, SVM 可以处理高维数据, 并且在小样本情况下也能表现良好, 适用于处理诊断信息较为复杂的证型分类问题。SVM 在中医辨证诊断中的应用提高了辨证诊断的准确性和效率, 尤其在处理复杂和高维度的医疗数据时, 显示出其独特的优势。然而, SVM 模型的性能很大程度上依赖于正确选择核函数和调整模型参数, 对参数和核函数的选择敏感, 需要彻底的数据预处理, 如归一化, 这需要丰富的专业知识和经验。同时, 处理大规模数据集时的计算效率仍是 SVM 面临的挑战之一。

3.3.3. 决策树

决策树(Decision Tree, DT)作为一种树状结构的分类和预测工具, 通过分析数据集中的特征来实现决策过程。在该模型中, 每个节点代表一种特征, 每条分支代表这个特征的一种可能值, 而叶节点则表示最终的预测结果或分类。DT 因其具有训练速度快、精确度高、分类模式相对简单等优点, 并且能够处理非线性关系和高维数据, 适用于辨证规则的提取这类大规模数据集。

李人亮[14]等人运用 CART 决策树算法共建立了气滞血瘀证、湿热阻络证和肝肾亏虚证三个中医辨证模型, 结果显示: 气滞血瘀证、湿热阻络证和肝肾亏虚证的预测准确率分别为 73.54%、84.56%、84.98%, 各模型的曲线下面积(ACU)值分别为 0.738、0.795、0.748。贺鹏飞[6]等通过分析 786 个慢性肾炎病例的医案数据, 构建了包括决策树在内的五种中医智能辨证模型, 其中决策树模型的准确率达到 81.5%。此外还有学者用决策树模型构建高血压痰湿壅盛证诊断模型[15], 慢性阻塞性肺病中医诊断模型[16], 慢性乙型肝炎肝胆湿热证和小儿肺炎痰热闭肺证中医证候诊断模型[17]等。这些模型的成功构建, 展示 CART 决策树在中医诊断建模领域的广泛应用性和潜在可行性。

尽管决策树模型在中医证候分类中得到了广泛应用, 但其在处理包含众多证素的情况时容易产生过拟合现象, 同时对数据中的噪声和异常值较为敏感, 在某些噪声较大的分类或回归问题上会过拟合。对于有不同级别的属性的数据, 属性的重要性评估可能偏差。这些问题仍需在未来的研究中得到进一步探索和解决。

3.3.4. 随机森林

随机森林算法(Random Forest, RF)是由 Leo Breiman 于 2001 年提出的一种新的机器学习算法[18], 是基于多个决策树的一种集成学习方法, 主要用于分类、回归和其他任务。随机森林具有高准确性、鲁棒性、可解释性等优点, 能够处理高维数据和大规模数据集, 对于数据中的噪声、异常值和缺失值包容性较好, 可以提供特征选择和变量重要性分析的结果, 帮助我们理解问题的本质, 进而有效地处理复杂的现实问题。

宫文浩[17]等以 918 例多中心小儿肺炎痰热闭肺证病例作为数据源, 利用 RF 和偏相关分析(PCA)的方法构建小儿肺炎痰热闭肺证中医证候诊断模型, RF 模型在超参数调优后显示出 90.58%的准确率、94.52%的灵敏度和 86.15%的特异度, 展现了高预测性能。贺鹏飞[8]等通过分析 786 个慢性肾炎相关的中医医案数据, 构建了包括随机森林在内的五种中医智能辨证模型, 其中随机森林以 85.4%的准确率先, 显示出在慢性肾炎证型判断中的高效性, 证明了人工智能在中医辨证方法学中的可行性。姜超[19]等人采用人工神经网络、贝叶斯网络、随机森林三种机器学习算法对 518 例脑出血急性期患者进行中医辨证分类, 发现随机森林的分类效果最好, 准确率、中经络正确率、中脏腑正确率分别为 81%、80%、81%。许明东[20]等通过随机森林算法, 为 627 例高血压病患者构建了中医辨证分型模型, 预测准确率高达 98.50%, 在测试集上更是达到了 99.50%。表明在中医临床指标被标准化的情况下, 通过结合 Boruta 特征选择, 基于随机森林算法的高血压病中医辨证分型建模可以对高血压病中医证型进行有效区分。

随机森林能有效管理和分析具有大量特征的高维数据, 对数据集中的异常值和噪声具有较高的容忍度, 展现出优秀的并行处理能力和泛化性能, 适用于处理中医证型分类中的复杂数据。通过整合多个决策树的预测结果, 随机森林提供了比单个决策树更准确、更稳定的性能。然而, 随机森林在处理噪声较大的数据集时仍有过拟合的风险, 并且由于其随机性, 模型的解释性较单个决策树更加复杂。

3.3.5. 人工神经网络

人工神经网络(Artificial Neural Networks, ANN)通过模拟人脑神经网络的结构和功能对数据进行有效处理, 进而识别出数据中隐藏的复杂规律[21]。这种网络结构由众多简单的单元互联组成, 展现出并行处理、高容错率、智能化及自我学习的特质。ANN 的特点在于其非线性的高度以及能够表达复杂知识和进行智能自适应学习的能力, 使其能够执行高级的逻辑运算和处理非线性关系。

臧晓彤[22]等通过人工神经网络算法建立小细胞肺癌中医智能辨证模型, 该模型训练集总准确率 91%, 验证集总准确率 92.6%, 模型平均 AUC 为 0.842, 为小细胞肺癌中医临床诊断提供参考依据。丁亮[23]等人开发了一个基于深度神经网络的原发性肝癌证型诊断预测模型, 该模型对原发性肝癌病例的证型预测准确率介于 82.86%~92.76%, 关联规则验证数据集的符合率介于 75%~100%, 证明了其在中医证型分类预测中的高效性和精确度。刘超[21]等基于 BP 神经网络方法针对冠状动脉临界病变患者证候要素及其常见组合, 建立 7 种证候要素的 BP 神经网络中医辨证诊断模型, 训练集平均准确率为 94.51%, 验证集平均准确率为 88.29%, 模型平均 AUC 为 0.953。对 5 种常见证候要素组合形式构建 BP 神经网络中医辨证诊断模型, 训练集平均准确率为 99.47%, 验证集平均准确率为 94.34%, 模型平均 AUC 为 0.996。为冠心病早期中医证候诊断提供了客观依据, 为中医辨证体系规范化、客观化提供了可行策略。曹云[24]应用支持向量机、神经网络和自动编码器分别构建的胃食管反流病中医智能辨证模型显示, 这些算法在诊断两种证候的准确率分别达到 78.3%、79.2%和 79.2%, 表明神经网络及自动编码器降维基础上的神经网络模型具有很好的诊断、预测能力。

ANN 可在大量中医数据中发现规律, 并总结其潜在的逻辑关系, 进而发现各项信息之间的相关规则, 以实现数据的预测, 从而更好地解决中医证候分类问题。然而, ANN 在实际应用中仍然存在一些问题,

比如难以避免陷入局部最优解、收敛速度慢、训练精度低于预期、输出结果与真实数据存在差异等, 训练过程需要大量计算资源。对初学者而言, 调参过程复杂且不直观。因此, 未来需要对其进行改进, 可以考虑增加算法复杂性或与其他技术相结合, 以克服这些缺陷所带来的预测效果不佳的挑战。与其他技术的结合已经逐渐成为研究的热点和未来的发展方向。

3.4. 未来应用前景与发展趋势

未来机器学习的发展将重点在于更高效地融合与分析多来源数据, 例如结合基因组学数据、生化指标和临床表现等多模态信息, 目的是实现更全面的疾病分析和预测。此外, 机器学习预计将促进中医在个性化医疗领域的应用, 通过细致分析个体间的差异, 向每个病人提供定制化的诊断与治疗方​​案。结合机器学习、生物信息学和系统生物学等学科, 有潜力深化我们对中医证候与疾病生物学基础之间联系的理解。同时, 发展智能化的医疗工具和应用程序, 例如利用智能穿戴设备进行持续的健康监控和疾病预防, 以及开发基于移动技术的个性化健康管理解决方案, 也是机器学习未来的重要发展方向。

4. 面临的挑战与展望

在将机器学习技术应用于中医辨证诊断过程中, 可能会遇到多种挑战。例如, 中医辨证诊断依赖于丰富的医学特征, 然而这些特征通常是非结构化且难以量化的, 比如舌苔、脉搏等, 因此特征提取成为一个挑战。目前, 中医四诊数据的收集缺乏一个被广泛认可的范式, 这导致了机器学习所需的大量规范、全面和准确的原始数据难以获取。此外, 中医辨证诊断的过程要求医生能够理解模型的决策基础, 然而传统的机器学习模型往往缺乏解释性, 因此难以解释其预测结果。因此, 如何将复杂深奥的辨证原理转化为计算机可理解的语义符号仍需进一步研究。

随着人工智能技术的发展与变革, 越来越多的新兴技术应用于中医研究, 基于机器学习的中医辨证诊断模型虽然面临挑战, 但也有着巨大的发展潜力。以下是一些具体的应用展望: 1) 高度个性化的中医诊疗: 机器学习技术通过分析和处理包括遗传资料、生活习惯、环境因素及临床表现在内的大量个体健康数据, 能够为每位患者定制高度个性化的中医证候分析及治疗方案。这种基于数据驱动的个性化医疗方法, 预计将显著提升治疗成效, 并减少非必要的医疗介入, 从而优化医疗资源的使用, 提高患者的治疗满意度和生活质量。通过精确匹配患者的具体病情与最适合的治疗方法, 机器学习技术在提高治疗精确度和效率方面展现出巨大潜力。机器学习技术在中医证候辨证诊断模型中的应用正逐渐成为推动中医现代化和个性化治疗发展的重要力量。2) 精准的证候识别和分类: 未来的研究可以整合文本、图像、声音等多种数据模态, 提供更全面、多维度的信息来支持中医辨证诊断。后续研究可以融合舌象和脉象构建多模态模型, 利用深度学习技术提高诊断准确性, 并加强模型的可解释性, 以更好地应用于临床实践并为患者提供更好的医疗服务, 有助于提高患者满意度和治疗的有效性。这种基于数据和算法的方法, 为中医的现代化和标准化提供了新的路径。3) 跨学科研究的深化: 未来通过整合机器学习技术与生物信息学、遗传学、系统生物学等领域的研究, 可以深入探究中医证候与疾病的生物学基础之间的关联。这种跨学科的方法不仅助力于中医证候的科学化和标准化进程, 而且促进了中西医结合研究的深化, 为理解复杂疾病机理和开发新的治疗策略提供了新视角。通过揭示中医证候背后的分子机制和生物标记物, 这种方法有望提高中医诊断和治疗的准确性和效率, 同时为中西医融合提供坚实的科学基础。4) 智能化诊疗辅助系统的发展: 通过机器学习算法, 开发的智能中医诊疗辅助系统, 如智能问诊、疾病分析、治疗方案推荐等, 能够极大提升诊疗流程的效率和精确性。这类系统能够辅助医生快速准确地做出诊断和治疗决策, 通过分析大量的临床数据和患者信息, 提供个性化的治疗建议。此外, 智能系统通过提供便捷的在线问诊服务, 不仅减轻了医生的工作负担, 还为患者带来了更加方便、快捷的医疗体验, 改善了

患者的就医流程和满意度。这种技术的应用, 不仅优化了医疗资源的分配, 还推动了中医诊疗服务向数字化、智能化的转型。5) 实时健康监测和疾病预防: 结合智能穿戴设备与移动健康应用, 机器学习技术能够实时监控个人健康状态, 并依据中医理论进行健康风险预警。此方法预期在疾病预防与健康管理领域具有重要价值。通过持续收集和分析个体的生理数据, 如心率、血压、睡眠质量等, 机器学习模型可以识别健康趋势和异常模式, 及早警示潜在的健康问题。这不仅有助于个人对自身健康状况有更深入的了解, 也使医生能够基于数据作出更准确的诊断和治疗计划, 促进个性化医疗的实施。此外, 该技术的应用还能激励个人积极参与健康管理, 通过改善生活习惯预防疾病, 提高生活质量。6) 大数据驱动的中医知识发现: 应用机器学习技术于海量中医临床数据、古籍文献及现代科研成果的分析, 能显著加快中医知识的挖掘与创新过程。这种技术能够揭示新的证候、疾病机制及治疗手段。通过深度分析和学习复杂数据模式, 机器学习不仅能够识别现有知识中的隐含关联, 还能预测未被发现的治疗效果和药物作用。这一进程不仅加深了对传统中医理论的理解, 还促进了中医与现代医学方法的融合, 为中医的现代化和国际化提供了强有力的技术支持。

综上, 机器学习技术在中医证候辨证诊断模型的应用展望广阔, 有望为中医的现代化、国际化和科学化提供强大的技术支持, 同时为全人医疗和个性化治疗贡献重要力量。

5. 小结

机器学习技术在中医辨证候断领域的应用已经取得了一定的进展, 但仍然存在着挑战和发展空间, 未来需要进一步深入研究, 不断完善模型和方法, 以更好地服务于临床实践和患者需求。中医作为中国传统医学的重要组成部分, 其现代化和智能化发展已成为必然趋势。基于机器学习的中医辨证诊断模型将传统中医知识与现代医学技术相结合, 辅助医生做出更精准、更规范的诊断, 从而实现中医诊疗过程的智能化。同时, 这种模型能够分析患者的个体特征、病史以及临床表现等多方面信息, 为每位患者提供个性化的诊疗方案, 有助于实现精准医疗的目标, 提高患者的治疗效果和生活质量。此外, 该模型具有普适性和可迁移性, 可以在全球范围内推广应用, 这将推动中医现代化进程, 促进中医国际化发展, 使中医在当今医疗领域中发挥更重要的作用。

本文综合评述了机器学习在中医证候诊断中的应用进展, 强调了机器学习技术在提高中医诊断准确性、实现个性化治疗方案、以及促进中医证候科学化和标准化方面的巨大潜力。通过分析不同机器学习技术的优势与局限, 并探讨其在中医证候诊断研究中遇到的挑战和未来的发展方向, 如数据标准化、模型可解释性等问题。展望未来, 提出整合多学科研究、开发智能化诊疗工具和应用程序, 以及利用大数据驱动的中医知识发现, 为机器学习技术在中医诊断领域的深入研究和应用提供了理论支持和实践指导。

参考文献

- [1] 夏淑洁, 杨朝阳, 周常恩, 等. 常见机器学习方法在中医诊断领域的应用述评[J]. 广州中医药大学学报, 2021, 38(4): 826-831. <https://doi.org/10.13359/j.cnki.gzxbtcm.2021.04.032>
- [2] 张晓航, 石清磊, 王斌, 等. 机器学习算法在中医诊疗中的研究综述[J]. 计算机科学, 2018, 45(z2): 32-36.
- [3] 曹云. 基于机器学习的胃食管反流病中医证候分类的应用研究[D]: [硕士学位论文]. 北京: 北京中医药大学, 2019. <https://doi.org/10.26973/d.cnki.gbjzu.2019.000087>
- [4] 赵文, 张佳, 徐佳君, 等. 四诊合参智能化发展现状及实现路径[J]. 中医杂志, 2020, 61(1): 58-62, 67.
- [5] 赵若琳, 常运立. 中医智能诊断的应用优势及其伦理问题[J]. 中国医学伦理学, 2021, 34(6): 746-752.
- [6] Wang, Y.L., Shi, X.M., Li, L., et al. (2021) The Impact of Artificial Intelligence Ontraditional Chinese Medicine. *The American Journal of Chinese Medicine*, 49, 1297-1314. <https://doi.org/10.1142/S0192415X21500622>
- [7] 刘福栋, 姜晓晨, 王桂彬, 等. 机器学习方法在中医学传承研究中的应用及思考[J]. 中医杂志, 2022, 63(8):

- 720-724, 738. <https://doi.org/10.13288/j.11-2166/r.2022.08.005>
- [8] 孙鑫亮, 杨涛, 章颖, 等. 基于 ML-KNN 算法的冠心病辨证模型研究[J]. 山东中医药大学学报, 2019, 43(5): 438-442. <https://doi.org/10.16294/j.cnki.1007-659x.2019.05.004>
- [9] 贺鹏飞, 刘聪, 李红典, 等. 运用机器学习构建慢性肾炎中医辨证智能模型及内部验证[J/OL]. 辽宁中医药大学学报, 1-11. <https://link.cnki.net/urlid/21.1543.R.20240124.1503.010>, 2024-02-18.
- [10] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1007/BF00994018>
- [11] 孙资金, 吉静, 马重阳, 等. 基于机器学习的中风中医辨证模型的构建与应用[J]. 湖南中医药大学学报, 2023, 43(4): 694-699.
- [12] 顾天宇, 严壮志, 蒋皆恢. 基于支持向量机的中风病中医证候分类[J]. 中医药信息, 2021, 38(9): 1-3. <https://doi.org/10.19656/j.cnki.1002-2406.20210901>
- [13] 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. <https://doi.org/10.13359/j.cnki.gzxbtcm.2020.03.036>
- [14] 李人亮, 张平, 胡子毅, 等. 基于 CART 决策树与 BP 神经网络算法探析蒋小敏教授治疗骨痹的辨证规律[J]. 世界科学技术-中医药现代化, 2023, 25(1): 401-412.
- [15] 田艳鹏, 丁学义, 朱羽硕, 等. 基于决策树和神经网络的高血压病痰湿壅盛证诊断模型研究[J]. 中华中医药杂志, 2018, 33(8): 3579-3584.
- [16] 苏翀, 任瞳, 王国品, 等. 利用决策树建立慢性阻塞性肺病中医诊断模型[J]. 计算机工程与应用, 2019, 55(3): 225-230.
- [17] 宫文浩, 兰天莹, 杨燕, 等. 基于随机森林和偏相关分析的小儿肺炎痰热闭肺证中医证候诊断模型研究[J]. 中华中医药杂志, 2023, 38(9): 4497-4501.
- [18] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [19] 姜超, 冯哲, 王均琴, 等. 三种机器学习方法在脑出血中医辨证分类中应用的比较研究[J]. 中国卫生统计, 2023, 40(6): 921-924, 928.
- [20] 许明东, 孙雨琪, 夏玉婷, 等. 基于随机森林的高血压病中医辨证分型模型[J]. 中医临床研究, 2021, 13(28): 30-32.
- [21] 刘超, 高嘉良, 董艳, 黄信生, 林飞, 李易, 张振鹏, 李军, 王阶. 基于 BP 神经网络的冠状动脉临界病变患者证候要素及其常见组合中医辨证诊断模型研究[J]. 中国中医药信息杂志, 2021, 28(3): 104-110.
- [22] 臧晓彤, 张培彤. 基于神经网络的小细胞肺癌智能辨证研究[J]. 中医学报, 2022, 37(5): 1067-1070.
- [23] 丁亮, 章新友, 刘莉萍, 牛晓录, 郭永坤. 基于深度神经网络的原发性肝癌证型诊断分类预测模型[J]. 世界科学技术-中医药现代化, 2020, 22(12): 4185-4192.
- [24] 曹云, 卢毅, 陈建新, 等. 基于机器学习的胃食管反流病中医智能辨证模型的应用[J]. 北京中医药大学学报, 2019, 42(10): 869-874.