

DOI:10.13288/j.11-2166/r.2025.15.014

文献研究

智慧化中医药循证医学证据获取模式探索 ——以冠心病证候研究的系统综述为范例

车前子^{1△}, 曾庆阳^{2△}, 宋填¹, 陈琳¹, 万静², 史楠楠^{1✉}

1. 中国中医科学院中医临床基础医学研究所, 北京市东城区东直门内南小街16号, 100700; 2. 北京化工大学

[摘要] **目的** 以冠心病证候研究的系统综述为范例, 探索智慧化文献筛选、获取和信息提取模式, 以提高中医循证医学研究效率。**方法** 基于中国知网 (CNKI)、维普中文科技期刊数据库 (VIP)、万方数据资源系统, 检索2000年1月1日至2023年12月31日发表的冠心病证候研究相关文献, 采用自动化工具批量获取文献元数据。通过分析标题和摘要, 利用文本相似度算法对文献进行合并、去重和初步筛选。进一步采用目标检测和图像处理技术, 根据全文和统计表格进行复筛。最后运用自然语言处理技术和预训练模型进行文献信息提取。**结果** 初步检索冠心病证候研究相关文献56 255篇, 通过人工智能辅助的自动化初筛和复筛将人工核查范围缩小至1075篇, 经人工全文筛选最终纳入冠心病证候研究文献646篇。经核查, 证候与证素的语义识别与拆分过程的准确性均在90%以上, 实现了文献基本信息及38种证素统计结果的自动化提取和标准化处理。**结论** 通过引入自然语言处理、预训练模型、人工智能图像处理等技术, 实现了对中医研究文献的高效检索、筛选和标准化数据提取。

[关键词] 人工智能; 冠心病; 证候; 文献筛选; 文献提取

数据驱动思维在循证医学领域的影响促成了新兴交叉研究模式^[1]。循证医学将已发表的文献研究结果作为证据生产资源, 可以更加全面且真实地总结相关研究进展, 具有多样性、准确性和时效性的特点^[2]。然而, 近年来中医研究的快速发展积累了大量研究文献, 使科研人员和临床医生在检索、筛选和提取信息时消耗越来越多的时间和人力资源。研究表明, 临床研究者完成并发表一篇系统评价平均耗时为67.3周^[3]。如何在庞大的医学文献资源中快速获取信息以提升医疗集体经验, 已成为亟待解决的问题。

为应对这些挑战, 人工智能 (artificial intelligence, AI) 技术逐渐被引入到系统评价的文献筛选、信息提取及偏倚评估等过程中^[4-5]。机器学习算法, 特别是近年来基于深度学习的大语言模型,

通过学习文献的特征, 可以自动识别与其相关的文献, 从而大大减少手工筛选的工作量^[6]。在提取文献数据的环节, 研究者需要手动从每篇文献中抄录特定研究指标, 如样本量、研究设计和治疗效果等, 耗时长且易出错。AI技术能够批量化地从文献的非结构化表达中自动提取结构化的研究信息, 提高循证研究的效率和准确率^[7]。针对中医证候的系统综述通常需要在长时间跨度下纳入各种研究设计类型的文献。一方面, 早期发表的文献存在研究设计、撰写格式和用词用语不规范、不统一的问题; 另一方面, 相较于干预性研究, 观察性研究的结构更为灵活多样^[8-9]。因此, 现有的自动化文献获取和信息提取工具多难以实现证候研究文献的快速、精确筛选, 尤其是对研究结果的规范化提取。本文以冠心病证候研究的系统综述研究为范例, 采用自动化工具对相关中医研究文献进行批量检索、筛选和获取, 并对文献数据进行自动提取和标准化, 旨在为智慧化中医循证医学研究建立行之有效的工作模式。

基金项目:北京市自然科学基金(7244507); 国家自然科学基金(82405243); 中国中医科学院科技创新工程(CI2022C004-01)

✉ 通讯作者:13811839164@vip.126.com

△同等贡献

1 资料与方法

1.1 文献批量检索

1.1.1 文献检索 本研究基于中文期刊文献库,包括中国知网(CNKI)、维普中文科技期刊数据库(VIP)、万方数据资源系统。通过前期调查,1999年及之前发表的冠心病中医研究文献数量较少,且鲜少统计患者证候或证素分布,因此检索2000年1月1日至2023年12月31日发表的冠心病中医证候研究相关文献。

1.1.2 文献纳入标准 1) 经过同行评审的以中文或英文发表的原始临床研究,具体研究设计类型包括病例系列、横断面研究、病例对照研究、队列研究、注册登记研究、病例队列研究、病例交叉研究、随机对照试验、实用性临床试验;2) 研究对象符合国内外相关诊断标准、指南中冠心病的诊断标准;3) 研究结果需报告患者证候或证素的频数或频率;4) 可获得全文,或有证型统计数据的摘要。

1.1.3 文献排除标准 1) 研究对象仅针对特定证候患者,或按照特定比例纳入不同证候患者;2) 临床研究中总病例数≤30例;3) 重复发表的文献仅保留资料最全的一篇。

1.1.4 预检索制定检索词 为适应不同时期中医研究文献格式和文字表达的差异,开展文献预检索,检索和筛选2000—2001年、2010—2011年、2020—2021年CNKI收录的冠心病中医证候研究文献。预检索结果显示,冠心病中医证候研究文献的题目和关键词多来源于中西医疾病名、症状、体征、干预措施。故本研究汇总预检索文献的关键词,经3位中西医结合心内科专家核对相关病名词汇,制定正式检索式为:篇名=‘心肌缺血’OR‘缺血性心脏病’OR‘冠状动脉疾病’OR‘冠状动脉粥样硬化性心脏病’OR‘冠心病’OR‘心绞

痛’OR‘UA’OR‘急性冠脉综合征’OR‘ACS’OR‘PCI’OR‘冠状动脉介入’OR‘CABG’OR‘心肌梗死’OR‘心肌梗塞’OR‘STEMI’OR‘心力衰竭’OR‘心衰’OR‘心律失常’OR‘心痛’OR‘胸痛’OR‘真心痛’OR‘卒心痛’AND主题=‘中医’OR‘中医学’OR‘中药’OR‘中药学’OR‘中医药’OR‘中成药’OR‘中草药’OR‘中西医结合’OR‘针灸’AND全文=‘辨证’OR‘证候’OR‘证型’OR‘证素’。

1.2 文献批量筛选与获取

1.2.1 基于文献元数据的初步筛选 基于中文期刊文献库和检索词,使用Python语言,利用自动化工具Selenium WebDriver,以多线程的方式获取开放权限的文献元数据,包括摘要、标题、关键词、作者、期刊、出版日期等。基于文献元数据,使用pandas、openpyxl库对来自多数据源的异构文献元数据进行对齐。通过余弦相似度算法计算文献元数据的标题相似度,设定相似度阈值0.85,相似度超过阈值的文献进一步计算元数据(作者、期刊、出版日期)相似度,相似度仍超过阈值则视为重复文献进行排除。

为通过标题快速筛选部分文献,根据文献排除标准,通过预检索总结综述、个人经验阐释、动物实验、文献研究、生信研究、单一证型/证素的研究标题中常见的关键词(表1)。排除标题中含有以上关键词的文献,并进一步将文献摘要与纳入排除标准进行比对筛选。

1.2.2 基于文献正文的精确筛选 本研究采用Python编程语言实现文献全文的自动化批量获取,基于初步筛选获得的文献元数据自动生成跨平台下载链接池,结合requests库与Selenium WebDriver构建双通道下载体系:通过requests直接访问开放

表1 冠心病中医证候研究文献预检索排除文献题目关键词

Table 1 Keywords summary of titles in the excluded articles in preliminary search

分类	排除关键词
综述或述评	综述、概述、概况、现状、进展、近况、浅谈、浅见、浅析、诊疗方案、展望、略论、治疗方略、述略、简述、浅探、简况、析义、探析、简析、述要、探要、述略、探源、教学、体会、理论研究、启示、meta、知识图谱、综合评价、系统评价、评价标准、纂要、循证、源流、可视化分析
个人经验阐释	刍议、雏议、病案、举案、经验、思路、思考、举隅、经验、临证解析、体会、心得、例析、赏析、验案、思路、认识、对策、体悟、研讨
动物实验	兔、鼠、猪、犬、动物
生信研究	网络药理、分子对接、机理研究
其他	诊疗指南、专家共识、标准、释义、解读
单一证型	血瘀、淤血、气虚、阳虚、阴虚、血虚、痰湿、痰浊、水饮、寒湿、热蕴、湿热、气滞、内毒、寒凝、虚寒

API 接口获取标准 PDF 文件, 针对需要交互验证的文献则采用 Selenium 模拟浏览器操作完成身份认证与文件下载。系统集成文件合法性验证模块, 通过二进制特征码检测确保下载文件的完整性及格式合规性, 对仅提供 HTML 阅读模式的文献, 开发基于 XPath 定位的正文区域智能识别算法, 利用 PDFKit 库实现网页内容至 PDF 格式的结构化转换。下载文件采用文献唯一标识符 (如 CNKI 数据库编码) 命名规则, 按出版年份-期刊名称-文献类型的层级化目录体系存储, 确保与后续表格提取及光学字符识别 (OCR) 处理环节的无缝衔接。整个流程支持断点续传与失败重试机制, 保证大批量文献处理过程的稳定性和完整性。

预检索结果显示, 证候或证素统计结果多以表格形式呈现。因此, 首先使用 Tabular 对每个候选文献的 PDF 全文进行处理, 保留具有表格的文献。对于具有表格的文献, 首先使用目标检测技术 YOLOv5 对文献中的表格区域进行定位, YOLOv5 模型的训练采用自定义数据集优化模型精度。然后使用 OCR 技术提取表格内容, TableTransformer 模型进行结构化表格数据的提取与解析。对提取出的内容进行分类和判断, 保留表格中具有“证候”或“证素”等概括性名词, 或包含描述具体证候或证素名词 (血瘀、气虚、阳虚等) 的文献, 排除没有证候或证素统计结果的文献。由具有医学背景的研究人员对精确筛选后的文献进行人工核查, 排除研究对象主要诊断非冠心病、包含古代医案病例、选择性纳入特定证候患者、使用文献数据或医生调查结果的文献。

1.3 文献数据批量提取与标准化

1.3.1 基础信息 采用深度学习技术构建信息抽取模型, 批量提取研究的基础信息。由于研究所需处理的大量文献数据无法完全通过人工标注, 因此引入了低资源情景下的伪标签生成策略, 并结合半监督学习方法进行模型训练。首先对部分文献进行人工标注, 标注内容包括主要诊断、研究地点、样本量等基础信息, 形成初步的标注数据集, 然后使用人工标注数据生成伪标签, 并应用于大规模文献数据, 以进一步优化抽取模型。

在深度学习模型方面, 本研究采用了大规模预训练语言模型 BERT, 并结合特定的抽取网络 W2NER (Word-word Relation Classification Architecture) 和 HGN (Hero-Gang Neural structure) 进行训

练。W2NER 主要用于处理词与词之间的关系分类, 特别适用于抽取涉及实体关系的信息, 如研究时间、地点、样本量等信息; 而 HGN 则通过深度学习神经网络结构, 能够更好地从长文本和复杂语境中抽取研究对象的主要诊断和合并症。通过这两种模型的结合, 能够更准确地从文献中提取基础信息, 增强了模型的精度和鲁棒性。模型还采用了自然语言处理中的文本嵌入技术, 将文献的摘要与正文中的关键词和短语转化为高维向量, 进而增强模型对语义信息的理解。最后生成的基础信息数据通过中医研究者校对, 确保其符合实际研究的要求。

1.3.2 证候和证素统计数据 在预检索阶段, 参照人工检索文献、《冠心病心绞痛证候要素诊断标准》^[10]、《经皮冠状动脉介入治疗围手术期心肌损伤中医诊疗专家共识》^[11]、《慢性心力衰竭中西医结合诊疗专家共识》^[12] 及《中医临床诊疗术语 第 2 部分: 证候》^[13] 制定证候和证素字典, 包括证候-证素映射表, 并由具有临床经验的中医研究者对字典进行校对。

在文献处理阶段, 通过对表格和其中的具体内容进行抽取, 自动化提取证候和证素统计数据。首先采用基于 Tabular 算法的表格识别模型对 PDF 文献中的表格进行自动识别和定位。该算法通过分析文档中的结构特征, 批量提取并汇总包含证候或证素统计的表格页。在此过程中结合图像处理技术和卷积神经网络, 对表格进行精确定位, 有效排除了文档中非表格内容的干扰。表格页的批量处理通过自动化脚本完成, 极大提升了文献处理的效率。然后使用表格自动抽取程序对提取到的表格进行进一步处理。此程序结合了规则引擎和机器学习技术, 根据表格格式的预定义规则判断哪些表格包含证候或证素信息, 应用 OCR 技术提取表格中的文本信息。通过这一过程, 文献中的表格内容被转化为结构化的数据格式 (如 Json、CSV 或 Excel), 为后续分析提供了基础。

数据提取阶段, 基于已纳入文献的表格特点, 总结并设计数据提取模板。模板的构建充分考虑了表格中的数据分布和结构, 通过这些模板可以有效地从表格中识别并提取出证候或证素信息。结合预检索制定的证候和证素字典及余弦相似度算法, 进一步精确定位表格内容中的证候或证素信息。使用 BERT-base-Chinese 词向量模型, 通过计算证候、证素名称与标准模板的语义相似度, 实现自动化的

识别和匹配。对于每一个提取的证候,通过与证候和证素字典进行相似度计算,判断其语义一致性,确保提取数据的准确性。针对证候数据,通过预检索阶段建立的证候-证素映射表,进行证素拆分。根据映射表将证候名称与相应的证素进行匹配,并在此基础上利用语义相似度技术进一步确认拆分的准确性。提取的证候与证素信息汇总到一个标准的 Excel 表格中。最终生成的证候汇总信息表由具有中医研究背景的专家进行校对和验证,以确保数据的准确性、完整性和临床相关性。

2 结果

2.1 文献批量检索、筛选和获取结果

以冠心病证候研究为例,通过预检索制定检索式和排除文献题目关键词,在 CNKI、VIP 及万方检索并批量获得 56 255 篇文献元数据,其中 CNKI 12 363 篇,万方 41 323 篇, VIP 2569 篇。采用相似度算法合并去重,排除重复文献 13 769 篇。根据文献元数据进行初筛,含有排除文献题目关键词的 26 694 篇文献不纳入研究,进一步将文献摘要与纳入排除标准进行比对,排除文献 7476 篇。对剩余文献批量下载全文进行精确筛选,采用目标检测技术定位表格中证候或证素统计结果,排除没有具体统计结果的文献共计 7241 篇。纳入文献均由具有医学背景的研究人员对精确筛选后的文献进行人工核查,排除研究对象主要诊断非冠心病、包含古代医案病例、选择性纳入特定证候患者、使用文献数据或医生调查结果的文献共 429 篇,最终纳入 646 篇文献。人工智能辅助的自动化文献检索、初筛和复筛共用时不足 1 个月,将人工核查的文献数量减少了 55 180 篇 (98.09%),显著提升了研究者工作效率。文献筛选过程见图 1。

2.2 文献数据批量提取和标准化结果

为验证文献数据提取过程的准确性,本研究对模型输出结果进行了人工核查。从最终纳入的文献中随机抽取 50 篇,提取出证候名称共计 187 条,采用 BERT-base-Chinese 词向量模型计算其与标准证候词典中条目的语义相似度 (余弦相似度),输出相似度大于 0.85 的匹配结果,人工确认其语义一致性的准确率为 93.4%。在证素拆分任务中,通过证候-证素映射表对 91 条常见证候进行拆分,平均每个证候拆分出 2.4 个证素,拆分准确率为 91.2%。此外,比较不同模型的效果发现,使用 BERT-base-

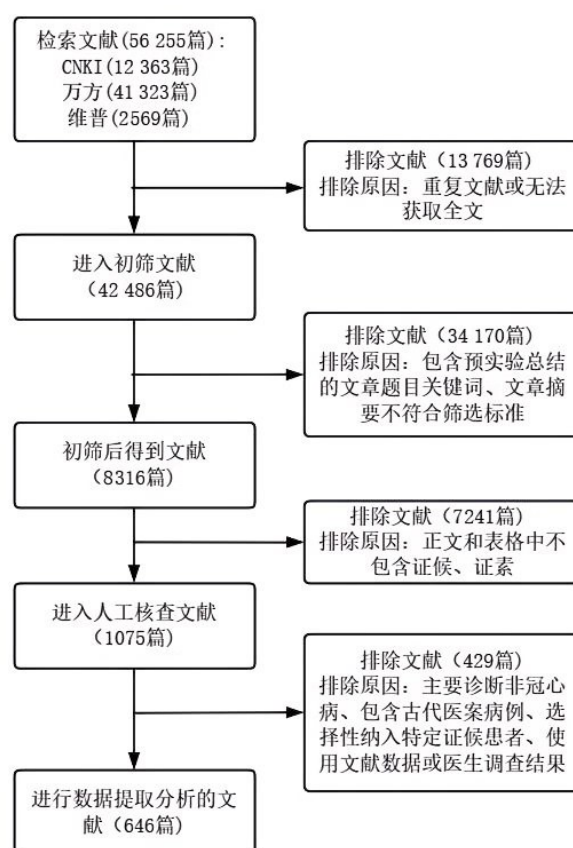


图 1 冠心病证候研究文献筛选流程图

Figure 1 Flowchart for coronary heart disease syndrome-related literature screening

Chinese 词向量模型计算语义相似度的平均准确率 (93.4%) 高于 Word2Vec (82.9%), 且在处理存在语义歧义的证候 (如“气虚血瘀”和“血瘀气滞”) 时表现更为稳定。

文献数据通过深度学习自动化提取和处理后以 Excel 格式保存。其中文献基本信息包括 doi 编码、文献标题、研究时间、研究地点、研究对象主要诊断、合并症、分组、样本量、年龄、性别。证候统一拆分为证素, 最终纳入的文献中共包含 23 个病性证素和 15 个病位证素。病性证素分别为血瘀、气虚、痰湿、阴虚、毒、气滞、火 (热)、阳虚、津液伤、血虚、阳亢、水饮、精亏、亡阳、湿、寒凝、血寒、血热、气陷、气不固、内风、外风、暑; 病位证素分别为: 心、肺、脾、肝、肾、胃、胆、小肠、大肠、膀胱、胸膈 [上焦]、表、经络、神、脑。

3 讨论

应用人工智能技术的自动化文献获取与信息提取工具, 节约处理中医研究文献的时间和人力资

源,有助于推进中医循证医学研究证据的应用和转化。然而,现有的AI辅助文献筛选与数据提取工具在中医研究领域的应用仍存在一定局限。本研究构建了智慧化中医药循证医学证据获取模式,系统整合了人工智能技术与循证医学方法。首先,基于多源文献智能检索策略,通过跨平台数据库集成及中西医术语融合的检索式构建,形成基础文献池。其次,采用双层次自动化筛选机制:在元数据层面建立基于语义相似度的重复文献排除算法,结合关键词过滤规则库实现初步筛选;在全文层面创新性地开发了表格导向的深度学习筛选体系,集成YOLOv5表格定位、OCR文本提取及语义匹配技术,精准识别证候相关数据。再次,构建低资源环境下的信息抽取模型,通过BERT-W2NER-HGN混合架构实现基础信息自动化提取,结合伪标签生成和半监督学习优化模型性能。最后,建立证候数据标准化流程:基于权威共识构建动态证候-证素映射词典,运用词向量模型计算语义相似度,实现证候术语的智能匹配与证素拆分。通过人工智能与循证医学的深度融合,系统解决了传统证据获取中存在的效率瓶颈与标准化难题。文献检索环节,建立跨平台数据库整合、中西医术语融合的检索策略;文献筛选环节,开发以表格特征识别为核心的深度学习筛选模型,突破非结构化数据提取的技术壁垒;数据提取环节,创建证候-证素动态映射的语义标准化流程,攻克中医术语异质性的核心难题。该模式在冠心病证候研究中成功实现了大规模文献数据的智能化处理,为中医药循证研究提供了可推广的方法学范式。

目前较为成熟的自动化文献筛选工具,如Cochrane Crowd, Abstrackr 和 SWIFT-Review, 通常依据用户设定的纳入标准,自动识别文题和摘要中描述患者、干预措施、对照措施、结局指标、研究设计(PICOS原则)的句子,以进行文献的初步筛选,然后基于全文内容进行复筛,排除不符合纳入标准的文献。但是,与西医研究文献相比,中医文献的自动筛选更加困难。首先,中医文献在中文期刊文献库中的分类并不完善,大量文献并没有标记“随机对照试验”“综述”等类型索引。其次,中医文献的行文格式更加多样,部分述评和阐释个人临床经验的文献难以和原始研究简单区分。再次,中医研究对于同一概念有更加丰富的表达形式,例如“证候”可以用“证”“证候”“证素”以及具体的

证候和证素类型等关键词表示。最后,早期文献的摘要内容较为简略,缺乏规范化结构,无法充分反映研究设计和统计结果。因此,仅通过题目和摘要筛选中医文献的准确率较低。本研究在制订纳入和排除标准的基础上,通过对不同时间段的研究文献进行预检索,总结出了不同类型文献的关键词,根据文本相似度计算,进一步精简初筛结果,从而达到了较好的筛选效果。

在复筛阶段,需要进一步判断文献是否包含证候或证素分布的统计结果。由于这类统计结果可能仅在正文中出现,无法通过摘要快速筛选,一般只能通过人工下载并阅读大量文献进行核查。因此,本研究通过全文检索和表格筛选来确定文献是否符合纳入标准。文献库下载的文献全文通常以PDF格式存储,目前自动化全文筛选过程大多需要将PDF转化为xml或txt等可读取的文本格式,转化过程中很容易发生格式混乱^[14]。本研究使用Tabular对候选文献的PDF格式进行处理,可以将文字、表格、数据提取并转化为结构化格式,便于后续分析和处理^[15]。在表格筛选过程中,使用的目标检测技术YOLOv5能够识别图像中不同形状和复杂度的表格区域,其模型在处理文档中的表格识别时可以达到高精度水平,且具有非常快的检测速度,能够实时处理大批量的表格数据^[16]。TableTransformer模型能够准确识别并分割文献中的复杂表格,特别是在处理具有嵌套结构和跨页的表格时表现优异^[15]。本研究与李易真等^[17]开发的中医药文献信息中循证证据的自动选择和推荐系统有相似之处,该系统使用机器视觉方法将中医药文献信息划分为多个循证证据,然后采用BERT预训练模型与模式规则对循证证据进行信息抽取。相比之下,本研究能够处理文献中结构复杂的表格数据,为循证证据的提取和综合提供更高的精度和效率。

本研究仍存在一定的局限性。一方面,由于中医文献中术语的复杂性,文献检索过程仍然需要预检索,以制订检索策略、纳入标准和排除标准,并总结标题筛选关键词,以辅助文献筛选。尽管本研究的自动化模式在一定程度上解决了这一问题,但在后续人工审核中发现仍无法完全避免误判。尤其是针对需要结合中医专业知识背景判断的问题,例如研究对象为冠心病引起的心力衰竭,研究指标为中医体质而非中医证候,文献没有明确为系统综述但病例数据来源于专家既往发表的病例分析研究

等。因此 AI 文献筛选可以大大缩小文献范围,但仍然需要人工判读核查筛选结果。另一方面,由于现有中医文献自动化工具尚未开源,因此无法与本研究建立的模式进行比较分析。此外,由于工作量较大,本研究无法将整个流程与人工方法对比验证,而是在每个环节引入了人工核查保证结果准确性,保证算法准确率在 90% 以上。因此,未来研究需进一步优化人工智能算法,以更好地适应中医文献的复杂性,并解决文献质量不一致带来的问题,从而提高自动化处理的准确性和适用性。

本研究以冠心病证候研究为范例,建立了人工智能辅助的智慧化中医药循证医学证据获取模式,其中包括基于自然语言处理和自动化工具的文献批量筛选与下载,以及结合深度学习技术和图像处理技术的表格自动化识别和数据提取,显著提高了中医研究文献的检索、筛选、数据提取和标准化的效率。尽管如此,现有的人工智能技术在中医文献复杂术语的自动化判断和处理上仍存在局限,未来需要进一步优化算法,提升本模式的准确性和普适性。

参考文献

- [1] 范美玉. 高质量循证医学证据获取与应用研究[J]. 协和医学杂志, 2023, 14(1): 39-43.
- [2] 阎小妍, 董冲亚, 姚晨. 大数据时代的循证医学研究[J]. 中国循证医学杂志, 2017, 17(3): 249-254.
- [3] BORAH R, BROWN AW, CAPERS PL, et al. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry[J]. BMJ Open, 2017, 7(2): e012545.
- [4] O'MARA-EVES A, THOMAS J, MCNAUGHT J, et al. Using text mining for study identification in systematic reviews: a systematic review of current approaches[J]. Syst Rev, 2015, 4(1): 5.
- [5] CHEERKOOT-JALIM S, KHEDO KK. A systematic review of text mining approaches applied to various application areas in the biomedical domain[J]. J Knowl Manag, 2021, 25(3): 642-668.
- [6] O'KEEFE H, RANKIN J, WALLACE SA, et al. Investigation of text-mining methodologies to aid the construction of search strategies in systematic reviews of diagnostic test accuracy: a case study[J]. Res Synth Methods, 2023, 14(1): 79-98.
- [7] MARSHALL IJ, WALLACE BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis[J]. Syst Rev, 2019, 8(1): 163.
- [8] 卫茂玲, 辜蕊, 张鸣明, 等. 11 种地方性中医药期刊 1995~2000 年临床研究文献质量的初步分析[J]. 中国循证医学杂志, 2004, 4(6): 410-413.
- [9] 李敬华, 王家良, 崔蒙. 中医临床文献质量评价研究现状及方法分析[J]. 中国中医药信息杂志, 2008, 15(6): 95-98.
- [10] 王阶, 邢雁伟. 冠心病心绞痛证候要素诊断标准[J]. 中医杂志, 2018, 59(6): 539-540.
- [11] 世界中医药学会联合会介入心脏病专业委员会, 中华中医药学会介入心脏病专业委员会, 中国中西医结合学会心血管病专业委员会介入心脏病学组, 等. 经皮冠状动脉介入治疗围手术期心肌损伤中医诊疗专家共识[J]. 中国中西医结合杂志, 2017, 37(4): 389-393.
- [12] 中国中西医结合学会心血管病专业委员会. 慢性心力衰竭中西医结合诊疗专家共识[J]. 心脑血管病防治, 2016, 16(5): 340-347.
- [13] 国家市场监督管理总局, 国家标准化管理委员会. 中医临床诊疗术语 第 2 部分: 证候[S]. 北京: 中国标准出版社, 2021: 416.
- [14] 郭玉杰, 张雪芹, 孙文字, 等. 自动化文献筛选工具在系统评价中的应用[J]. 协和医学杂志, 2024, 15(4): 921-926.
- [15] LI Y, WEI Q, CHEN X, et al. Improving tabular data extraction in scanned laboratory reports using deep learning models[J]. J Biomed Inform, 2024, 159: 104735. doi: 10.1016/j.jbi.2024.104735.
- [16] XIA Y, CAI J, LI Y, et al. A precision-preferred comprehensive information extraction system for clinical articles in traditional Chinese Medicine[J]. Int J Intell Syst, 2022, 37(8): 4994-5010.
- [17] 北京航空航天大学. 一种中医药文献信息中循证证据的自动选择和推荐系统: CN202410910999.7[P]. 2024-10-22.

Exploration of an Intelligent Evidence Achieve Mode of Evidence-Based Chinese Medicine: Take Systematic Review of Coronary Heart Disease Syndrome Research as an Example

CHE Qianzi¹, ZENG Qingyang², SONG Tian¹, CHEN Lin¹, WAN Jing², SHI Nannan¹

1. Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, 100700; 2. Beijing University of Chemical Technology

ABSTRACT Objective To construct an intelligent model for literature screening, retrieval, and data extraction with a systematic review of coronary heart disease syndromes as an example, so as to improve the efficiency of evidence-based Chinese medicine research. **Methods** Based on China National Knowledge Infrastructure (CNKI), VIP and Wanfang Data Resource System, the articles related to coronary heart disease syndrome research published from January 1, 2000 to December 31, 2023 were retrieved. Automated tools were used to batch retrieve paper metadata. Using text similarity algorithms, papers were merged, deduplicated, and subjected to preliminary screening based on titles and abstracts. Further screening was performed using object detection and image processing technologies on the full texts and statistical tables. Natural language processing (NLP) techniques and pre-trained models were applied to extract information. **Results** The initial search retrieved 56 255 coronary heart disease syndrome-related articles. By artificial intelligence-assisted preliminary and secondary screening, the manual verification scope was narrowed to 1075 articles. Ultimately, 646 coronary heart disease syndrome related studies were included manually. With accuracy verification showing over 90% consistency in semantic recognition and element decomposition processes, we achieved data extraction and standardization processing for both basic literature information and 38 syndrome element statistics. **Conclusion** By incorporating natural language processing, pre-trained models, artificial intelligence image processing and other technologies, this study enabled efficient retrieval, screening and standardized data extraction of Chinese medicine research literature.

Keywords artificial intelligence; coronary heart disease; syndrome; literature screening; data extraction

(收稿日期: 2025-01-10; 修回日期: 2025-06-23)

[编辑: 柴倩云]

广告

《中医杂志(英文版)》欢迎投稿

Journal of Traditional Chinese Medicine [《中医杂志(英文版)》] (ISSN 0255-2922, CN 11-2167/R, 双月刊) 由中华中医药学会与中国中医科学院联合主办, 是我国中医药界具有权威的中医药英文科技期刊, 现发行和交换 90 多个国家和地区。

《中医杂志(英文版)》自 1981 年创刊以来一直坚持正确的办刊宗旨, 国际影响力不断提升。2010 年入选 SCI-E (《科学引文索引》扩刊版), 2011 年获得中国中医科学院自主选题项目资助, 2013 年获得“中国科技期刊国际影响力提升计划”资助, 自 2012 年起连续被评为“中国最具国际影响力学术期刊”, 2023 年再次入选“T1 级中医药英文科技期刊目录”。本刊先后被国际知名文献数据库 MEDLINE、《化学文摘》(CA)、《医学文摘》(EM) 等收录, 是 MEDLINE 收录的第一份中国出版的中医药期刊, 是 SCI 收录的 30 余种全科医学和替代医学期刊中的中国传统医学综合性学术期刊之一。

《中医杂志(英文版)》主要栏目包括循证研究、理论研究、基础研究、临床观察、中药研究、文献研究, 思路与方法、综述等。内容以报道中医药基础、临床研究成果和科研前沿进展为主, 覆盖中医药学科各个领域。欢迎广大专家学者积极投稿。作者可直接登录本刊网站 <http://www.journaltcm.cn> 投稿。

联系邮箱: jtcmen@126.com。联系电话: 010-64089190。

联系地址: 北京市东城区东直门内南小街 16 号, 邮政编码: 100700。