

医疗垂直领域中医药人工智能多模态测评体系构建研究 新进展

马燕 张霞婧 史文璞 闫辉 孙绪德^{通讯作者}

陕西中医药大学第二附属医院

DOI:10.32629/bmtr.v8i3.20475

[摘要] 随着人工智能技术与中医药领域的深度融合，多模态模型已成为推动中医药数字化、智能化转型的核心载体，但其缺乏性能评估和系统测评体系的统一，制约了技术落地与行业规范化发展。本文围绕中医药人工智能多模态测评体系的构建背景、核心构成、关键技术进展、现存挑战及未来方向展开综述，重点分析数据标注标准化、多维度评测框架搭建、临床转化验证等关键环节的最新成果，旨在为该领域的研究优化与实践应用提供参考，助力破解“模型多、测评少；概念多、数据少；论文多、落地少”的行业痛点。

[关键词] 中医药；人工智能；多模态；测评体系

中图分类号：R2-03 文献标识码：A

Research Progress on the Construction of a Multi-modal Evaluation System for Traditional Chinese Medicine Artificial Intelligence in the Medical Vertical Field

Yan Ma, Xiajing Zhang, Wenpu Shi, Hui Yan, Xude Sun

Corresponding Author

The Second Affiliated Hospital of Shaanxi University of Chinese Medicine

[Abstract] With the deep integration of artificial intelligence technology and the field of traditional Chinese medicine, multi-modal models have become the core carrier for promoting the digital and intelligent transformation of traditional Chinese medicine. However, their lack of unified performance evaluation and system assessment systems has restricted the implementation of the technology and the standardized development of the industry. This article conducts a review on the construction background, core components, key technological progress, existing challenges, and future directions of the multi-modal assessment system for traditional Chinese medicine artificial intelligence. It focuses on analyzing the latest achievements in key aspects such as standardized data annotation, the establishment of multi-dimensional evaluation frameworks, and clinical transformation verification, aiming to provide references for research optimization and practical application in this field, and help solve the industry pain points of “many models, few evaluations; many concepts, few data; many papers, few implementations”.

[Key words] Traditional Chinese Medicine; Artificial Intelligence; Multimodal; Assessment System

引言

中医药作为我国独特的医疗健康资源，其整体观念、辨证论治的核心理念与多模态数据的整合分析特性高度契合，为人工智能技术的应用提供了广阔场景。近年来，舌诊、脉诊等中医特色诊疗数据的客观化采集技术日趋成熟^[1]，结合电子病历文本、代谢组学数据、医学影像等多源信息，中医药人工智能多模态模型在证候诊断、方剂推荐、疗效评价等领域实现了诸多突破^[2]。然而，当前领域内存在模型研发碎片化、测评标准不统一、临床实用性验证不足等问题，多数模型仅停留在实验室阶段，难以满足临床实际需求^[3]。构建科学、全面的多模态测评体系，成为衡量模型性能、指导技

术优化、推动临床落地的关键支撑，对促进中医药现代化与智能化发展具有重要意义^[2]。本文系统梳理中医药人工智能多模态测评体系的研究现状与最新进展，为行业规范化发展提供理论与实践参考^[1]。

1 中医药人工智能多模态测评体系构建背景与核心需求

1.1 构建背景

中医药与人工智能的融合是时代发展的必然趋势，也是中医药创新的重要路径，但二者融合过程中面临多重瓶颈制约^[5]。从数据层面看，中医药多模态数据涵盖文本、图像、生理、代谢等多种类型，存在格式异构、标准不一、隐私保

护与共享矛盾突出等问题，导致测评数据质量参差不齐^[6]；从技术层面看，现有模型多聚焦单一任务优化，缺乏全流程能力评估，且中西医理论体系差异较大，模型的中医药专业性难以量化衡量^[7]；从应用层面看，模型临床效果评估难、运行稳定性保障难，缺乏与医师诊疗能力的对标验证，制约了临床推广应用。在此背景下，亟需建立一套适配中医药特点的多模态测评体系，实现对模型性能的客观、全面、精准评估^[8,9]。

1.2 核心需求

中医药人工智能多模态测评体系的构建需满足三大核心需求：一是专业性需求，需贴合中医药理论体系，覆盖四诊合参、辨证论治、方剂配伍等核心诊疗环节，体现中医“整体观念”与“个体化诊疗”特色^[10]；二是全面性需求，需涵盖数据质量、技术性能、临床应用等多维度评估，既关注模型的准确率、召回率等技术指标，也重视其临床实用性、安全性与可解释性^[11]；三是动态性需求，需适配中医药数据的更新迭代与模型技术的快速发展，建立灵活、可扩展的测评框架，支持场景化、个性化测评需求^[10]。同时，测评体系还需兼顾行业标准统一与跨机构协作需求，推动形成从数据到模型进一步测评并且实践应用的良性循环生态^[12]。

2 中医药人工智能多模态测评体系核心构成与技术进展

2.1 测评数据集构建与标准化

高质量数据集是测评体系的基础，当前中医药多模态测评数据集构建已形成从标准化标注、质量控制到迭代更新的完整流程，核心进展集中在以下三方面：

1、特色模态数据标注规范化：针对舌诊、脉诊等中医特有数据，已建立细分维度的标注指南，实现客观化、精细化标注^[13]。舌诊图像标注涵盖舌质（颜色、形态、活动度）、舌苔（颜色、厚度、形态、湿度）、舌下络脉（颜色、形态）等维度，采用结构化格式记录标注信息及对应证型提示；脉诊数据标注则聚焦寸关尺脉位特征、脉象基本属性（脉率、脉律、脉力、脉形）及复合脉象，明确各脉象的量化标注标准，确保数据的专业性与一致性^[13]。文本数据方面，通过中医 NLP 技术构建术语本体映射，将非结构化的四诊记录、医案文献转化为标准化证候要素，提升文本数据的结构化程度与可复用性^[14]。

2、数据集质量控制体系完善：采用三级标注体系保障数据准确性，即初级标注由中医研究生完成，中级审核由主治医师负责，最终由副主任医师及以上专家确认，同时通过3名及以上专家独立标注并计算 Kappa 系数进行一致性检验，Kappa 系数 <0.7 的项目需经专家讨论达成共识^[1]。此外，引入边界案例与挑战性案例构建机制，筛选症状不典型、多证

型交叉、罕见证型及模态信息不一致的案例，全面检验模型的复杂场景适配能力，同时通过10%随机抽检进行质量把关，确保数据集的全面性与严谨性^[15]。

3、数据集迭代更新机制建立：遵循“基础-进阶-专业”的迭代路径，基础版数据集覆盖常见证型与基础诊疗能力，包含500个左右评测项目；进阶版增加复杂证型与疑难病例，扩展至800个项目并纳入中医各家学说内容；专业版则全面覆盖中医各科疾病，项目数量突破1000个，新增中西医结合相关评测内容^[1]。同时建立常态化更新机制，每季度分析模型表现优化评测难点，每半年新增评测项目，每年开展全面评估更新，适配模型技术与临床需求变化。代表性成果包括中国中医科学院研发的“中医辨证思维测评数据集（TCMEval-SDT）”，涵盖300份高质量医案，聚焦信息抽取、病机推理、证候诊断、解释总结四大辨证环节，已在主流大模型验证中展现出良好的测评效能^[2]。

2.2 多维度测评框架搭建

1、技术能力领域：聚焦模型的多模态融合与核心任务执行能力，涵盖单诊能力、多诊合参能力、知识应用能力三个子域。单诊能力评估针对各单一模态数据的处理效果，如舌诊图像的证型匹配准确率、脉诊数据的脉象识别精度、文本数据的证候要素抽取召回率等；多诊合参能力评估模型整合多模态信息的协同分析能力，重点考察不同模态数据不一致场景下的决策稳定性；知识应用能力则通过方剂推荐合理性、病机阐释准确性等指标，衡量模型对中医药理论知识的掌握与应用水平^[16]。

2、应用场景领域：关注模型的场景覆盖广度与适配性，涵盖中医内科、外科、妇科、儿科等专科场景，以及预防保健、慢病管理、急危重症辅助诊疗等功能场景，同时纳入不同人群（儿童、孕妇、老年人）的场景适配性评估，确保模型的临床应用普适性^[1]。针对中西医结合场景，新增中西医协同诊疗方案生成、中药与西药相互作用提示等评测项目，贴合临床实际应用需求^[17]。

3、临床价值领域：聚焦模型的实际应用效果与安全性，包含应用成熟度、临床有效性、安全性三个子域^[1]。应用成熟度评估模型的运行稳定性、响应速度及易用性；临床有效性通过与医师诊疗结果对比、真实临床场景疗效追踪等方式验证，核心指标包括证型诊断符合率、治疗方案有效率、患者症状改善程度等；安全性则重点评估模型的决策风险，如误诊率、误治风险及算法偏见，同时考察模型的可解释性，要求明确决策依据，降低临床应用风险^[18]。中国中医科学院发布的《面向行业的大规模预训练模型技术和应用评估方法第12部分：中医药大模型》首次明确该框架，涵盖13个能力子域、30余个能力项，实现中医药大模型的量化评估^[2]。

2.3 临床转化验证与跨场景适配

1、临床验证方法创新：通过构建模拟临床诊疗场景，采用动态对话式评测方法，模拟医师与患者的诊疗交互过程，评估模型的实时决策能力与问诊引导能力，更贴近真实临床诊疗流程^[1]。在真实临床验证中，依托三甲医院急诊科、慢病管理中心等场景开展试点，将模型嵌入中西医协同诊疗流程，追踪模型推荐方案的临床落地效果，通过患者症状积分、实验室指标变化、随访复发率等多模态数据综合评估模型的临床价值^[2]。例如在中成药临床评价中，通过多模态模型整合临床试验数据、医案文献与经济学证据，自动生成综合评价报告，精准对接医疗机构、药企监管部门的差异化需求，提升临床转化效率^[2]。

2、跨场景适配与对比评测：针对不同地域、不同流派的中医药应用差异，扩展测评体系的跨文化、跨流派适配性，纳入各流派的诊疗特色与地域疾病谱特点，确保模型在不同场景下的稳定性^[19]。同时开展模型与人类专家的对比评测，选取不同职称、不同专科的中医医师组成评测团队，从诊断准确率、治疗方案合理性、诊疗效率等维度进行对标，明确模型与人类专家的差距，为模型优化提供方向^[17]。此外，建立跨机构协作验证网络，推动不同医院、科研机构共享评测数据与验证结果，提升测评结论的权威性与普适性^[20]。

3 中医药人工智能多模态测评体系现存挑战

3.1 数据层面：共享壁垒与异构整合难题

中医药多模态数据分散存储于不同医疗机构，受数据隐私保护与利益壁垒影响，跨机构数据共享难度大，导致测评数据集规模有限、地域代表性不足，难以覆盖多样化的临床场景与人群特点^[21]。同时，文本、图像、生理、代谢等多模态数据格式异构、尺度差异大，数据采集设备与标准不统一，导致多源数据整合难度高，影响测评结果的客观性与可比性，亟需建立统一的数据共享机制与标准化采集规范^[22]。

3.2 技术层面：中医药专业性与可解释性测评不足

当前测评指标多借鉴现代医学 AI 测评体系，对中医药理论的适配性有待提升，如中医“气”“经络”等抽象概念难以量化测评，方剂配伍的君臣佐使逻辑、证候演变的动态规律等核心专业能力缺乏针对性评测指标^[23]。此外，多模态模型存在“算法黑箱”问题，决策过程的可解释性不足，现有测评体系难以全面评估模型决策依据的合理性与中医药理论契合度，不利于临床医师的信任与应用^[2]。

3.3 临床层面：真实世界验证与风险管控滞后

多数测评仍停留在实验室模拟场景，真实世界临床验证不足，模型在复杂临床环境中的稳定性、容错性及与现有诊疗流程的适配性缺乏系统评估^[24]。同时，临床应用中的伦理风险与法律责任界定不清晰，如模型误诊导致的医疗纠纷责

任划分、患者数据泄露风险等，现有测评体系尚未纳入完善的风险管控评估模块，难以满足临床安全应用需求^[2]。

3.4 行业层面：标准不统一与协同生态缺失

不同研究机构的测评方法、指标体系差异较大，缺乏统一的行业标准与互认机制，导致模型性能评估结果难以横向对比，制约了优质模型的筛选与推广^[22]。此外，“中医药+AI”跨学科协同生态尚未形成，中医专家、AI 研发人员、临床医师、监管部门的协作不足，测评体系的构建与更新难以充分贴合各方需求，影响其实际应用价值^[4]。

4 未来发展方向

4.1 构建跨机构数据共享与标准化平台

依托国家层面政策引导，建立符合伦理规范与数据安全要求的中医药多模态数据共享平台，采用联邦学习等技术在保护数据隐私的前提下实现跨机构数据协同，扩大测评数据集规模与多样性^[25]。同时制定统一的多模态数据采集标准与接口规范，规范舌诊、脉诊等特色数据的采集设备参数与操作流程，推动不同模态数据的结构化整合，提升测评数据的质量与一致性^[26]。

4.2 优化中医药特色化测评指标与可解释性评估

深度融合中医药理论，构建贴合辨证论治、方剂配伍、证候演变规律的特色测评指标体系，开发中医抽象概念的量化测评方法，提升测评体系的专业性与适配性^[27]。加强模型可解释性测评，引入决策路径可视化技术，评估模型决策依据与中医药理论的契合度，同时建立模型解释能力的量化评估指标，推动模型从“能决策”向“会解释”转变，增强临床医师的信任度^[23]。

4.3 强化真实世界临床验证与风险管控

建立“实验室-模拟临床-真实世界”全链条验证体系，扩大真实世界验证的样本量与场景覆盖范围，重点评估模型在复杂病例、特殊人群中的应用效果及与现有诊疗流程的融合适配性^[8]。将伦理风险与法律责任评估纳入测评体系，新增算法偏见、数据安全、决策风险等评估模块，明确模型临床应用的安全阈值，建立风险预警与管控机制，保障临床应用安全^[28]。

4.4 推动行业标准统一与协同生态构建

加快制定中医药人工智能多模态测评行业标准，明确测评框架、指标体系、数据规范与验证流程，推动测评结果的行业互认，引导模型研发向临床实用化方向聚焦^[2]。构建跨学科协同创新共同体，整合中医专家、AI 技术团队、临床机构、监管部门等多方资源，建立常态化沟通机制，推动测评体系的动态优化与落地应用，形成“研发-测评-落地-反馈”的良性循环^[29]。同时加强交叉学科人才培养，支持高校开设“中医药+AI”专业，培养兼具中医药理论与 AI 技术的复合

型人才,为测评体系的持续发展提供人才支撑^[30]。

5 结论

中医药人工智能多模态测评体系的构建是推动“中医药+AI”技术规范化、临床化发展的关键支撑,当前已在数据集标准化、多维度评测框架搭建、临床转化验证等方面取得显著进展,形成了贴合中医药特色的测评思路与方法。但仍面临数据共享壁垒、中医药专业性测评不足、真实世界验证滞后、行业标准不统一等挑战。未来需通过构建数据共享平台、优化特色化测评指标、强化临床验证与风险管控、推动标准统一与协同生态构建,进一步完善测评体系,为中医药人工智能技术的研发优化与临床落地提供科学支撑,助力千年中医药智慧在智能时代焕发新生机。

[参考文献]

- [1]Wang Z, Hao M, Peng S, et al. TCMEval-SDT: a benchmark dataset for syndrome differentiation thought of traditional Chinese medicine[J]. Scientific data, 2025,12(1):437.
- [2]张霄.中国中医科学院发布《中医药大模型评测标准》[J].中医药管理杂志,2025,33(10):24.
- [3]刘钰涵,胡芳.人工智能时代创新型中医药数据分析人才的培养模式研究[J].学校党建与思想教育,2020(16):71-73.
- [4]Qin S, Wang Y, Cui T, et al. Intelligent Chinese patent medicine (CPM) recommendation framework: Integrating large language models, retrieval-augmented generation, and the largest CPM dataset[J]. Pharmacological research, 2025,219:107883.
- [5]Deng J, Qin Y. Current Status, Hotspots, and Prospects of Artificial Intelligence in Ophthalmology: A Bibliometric Analysis (2003-2023)[J]. Ophthalmic epidemiology, 2025,32(3):245-258.
- [6]Liu X, Li S, Zhu Q, et al. Interpretable Semi-federated Learning for Multimodal Cardiac Imaging and Risk Stratification: A Privacy-Preserving Framework[J]. Journal of imaging informatics in medicine, 2025:10-1007.
- [7]李莉,户万.AI赋能,中医药传承创新谱新篇——记北京交通大学计算机科学与技术学院教授周雪忠[J].科学中国人,2025(6):64-65.
- [8]鄢灯莹,郑琪光,常凯,等.构建真实世界中医药临床研究人工智能技术体系:重要价值、研究问题与发展方向[J].上海中医药杂志,2025,59(10):1-8.
- [9]Chen Z, Zhang D, Nie P, et al. Developing the Artificial Intelligence Method and System for "Multiple Diseases Holistic Differentiation" in Traditional Chinese Medicine and Its Interpretability to Clinical Decision[J]. Journal of evidence-based medicine, 2025,18(2):e70016.
- [10]李小阳,花睿,周雪忠,等.基于命名实体识别和关系抽取的中药现代药理学知识图谱构建研究[J].中国中医药图书情报杂志,2025,49(5):35-41.
- [11]张艳宏,高伟,张润顺,等.基于辨证论治个性化治疗满意度评估简表的应用评价[J].中华中医药杂志,2025,40(9):4664-4669.
- [12]Liu T, Gu Y, Chen H, et al. A foundational triage system for improving accuracy in moderate acuity level emergency classifications[J]. Communications medicine, 2025,5(1):322.
- [13]陆俊红,徐艺峰,王忆勤,等.基于计算机图像技术的中医舌质客观化研究概述[J].中华中医药杂志,2024,39(1):352-354.
- [14]Trivedi S, Gildersleeve R, Franco S, et al. Evaluation of a Concept Mapping Task Using Named Entity Recognition and Normalization in Unstructured Clinical Text[J]. Journal of healthcare informatics research, 2020,4(4):395-410.
- [15]张鹏飞,曾鹏飞,王德贤,等.智能融合赋能中医四诊合参客观化[J].中国中西医结合杂志,2025,45(10):1165-1172.
- [16]Chakraborty S, Chowdhury R, Shuvo S R, et al. A scalable framework for evaluating multiple language models through cross-domain generation and hallucination detection[J]. Scientific reports, 2025,15(1):29981.
- [17]Wang J, Liu Y, Li J, et al. Artificial Intelligence in Traditional Chinese Medicine: Multimodal Fusion and Machine Learning for Enhanced Diagnosis and Treatment Efficacy[J]. Current Medical Science, 2025,45(5):1013-1022.
- [18]祖雅琪,唐旭东,于琦,等.中医消化领域优势病种循证大数据平台的构建及应用[J].中国数字医学,2025,20(02):56-62.
- [19]方俊泽,高怀婷,邢素霞,等.人工智能胸部疾病辅助诊断系统在基层医疗卫生机构全科医生诊断中的实效性评估[J].中国全科医学,2025,28(31):3948-3953.
- [20]Wang J, Liu Y, Li J, et al. Artificial Intelligence in Traditional Chinese Medicine: Multimodal Fusion and Machine Learning for Enhanced Diagnosis and Treatment Efficacy[J]. Current medical science, 2025,45(5):1013-1022.
- [21]王一萌,李敬华,田野,等.基于数据挖掘和网络药理学探讨中医药治疗原发性肝癌用药规律和作用机制[J].中国中医药图书情报杂志,2025,49(2):72-77.

[22]Ali A, Al-Rimy B A S, Alsubaei F S, et al. HealthLock: Blockchain-Based Privacy Preservation Using Homomorphic Encryption in Internet of Things Healthcare Applications[J].Sensors(Basel, Switzerland),2023,23(15):6762.

[23]Li Y, Liu X, Zhou J, et al. Artificial intelligence in traditional Chinese medicine: advances in multi-metabolite multi-target interaction modeling[J]. Frontiers in pharmacology, 2025,16:1541509.

[24]Evangelou K, Zemperligkos P, Politis A, et al. Diagnostic, Therapeutic, and Prognostic Applications of Artificial Intelligence (AI) in the Clinical Management of Brain Metastases (BMs)[J]. Brain sciences, 2025,15(7):730.

[25]陆承双,楼步青,曾宇平,等.联邦学习框架下中医药数据要素传承创新策略研究[J].中国卫生事业管理,2025,42(1):96-100.

[26]D. R K, T. S M, A. B, et al. Federated Learning for Privacy-Preserving Data Mining: 2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA)[C], 2024.2024

[27]Duan Y, Jia S, Ye S, et al. Artificial Intelligence for Spleen-Stomach Disorders in Traditional Chinese Medicine: Integrating Knowledge Graphs with Intelligent Diagnosis and Treatment[J]. Current medical science, 2025:10-1007.

[28]陈智慧,朱欢,胡婷霞,等.中药真实世界研究电子源数据采集与治理平台研究:构建中医药临床科研一体化新范式[J].中国食品药品监管,2025(9):78-89.

[29]赵根良.长三角科技创新共同体相关问题研究综述[J].内江科技,2025,46(8):129-131.

[30]孙荪,胡孔法.中医药人工智能研究生培养模式构建与实践——基于知识生产模式转型理论[J].时珍国医国药,2023,34(6):1487-1489.

作者简介:

马燕(1993-),女,民族:汉,陕西省咸阳市人,博士研究生,陕西中医药大学第二附属医院,讲师,研究方向:围术期医学,神经毒理。

基金项目:

咸阳市重大科技创新专项:重大技术攻关专项-中医药与人工智能(批准号:L2025-ZDKJ-ZDGG-ZYRGZN-003)。