

人工智能驱动中药质量标志物发现与评价：数据基础、证据链与验证框架

李 欢^{1,2}, 周子欣², 栗悦潇², 刘 洋¹, 薛 蕊^{3,4}, 王 杰⁵, 张双喜¹, 史孟华^{1*}, 张相安^{1*}

1. 河南中医药大学第一附属医院 肛肠科, 河南 郑州 450000

2. 河南中医药大学, 河南 郑州 450046

3. 云南中医药大学暨云南省南药可持续利用研究重点实验室, 云南 昆明 650500

4. 云南省药食同源饮片工程中心, 云南 昆明 650500

5. 上海中医药大学, 上海 201203

摘 要: 中药质量标志物 (quality marker, Q-Marker) 是指存在于中药中, 与中药的质量、药效、安全性密切相关, 能够反映中药品质特征、区分中药真伪优劣、追溯中药来源, 且可被精准检测的化学物质或特征组合, 是中药质量控制与评价的核心指标。传统中药质量控制多依赖经验判断与单一指标成分, 存在筛选效率低、主观性强、难以兼顾“化学特异性-药效关联性-整体协调性”的局限, 难以适配中药“多成分、多靶点、整体调节”的核心特点。随着人工智能 (artificial intelligence, AI) 技术的快速发展, 机器学习、深度学习、知识图谱等技术凭借强大的多源数据融合、高维特征挖掘与非线性关联建模能力, 逐步渗透到中药 Q-Marker 研究的全流程, 实现了从“经验驱动”向“数据驱动”的转变, 显著提升了 Q-Marker 筛选的精准度、效率与科学性。系统梳理 AI 驱动的中药 Q-Marker 研究的多源数据基础、AI 预处理技术、核心算法、研究范式, 分析当前研究面临的挑战并展望未来发展方向, 为中药 Q-Marker 的智能化研究与标准化应用提供参考。

关键词: 人工智能; 中药; 质量标志物; 深度学习; 多组学; 知识图谱

中图分类号: R28; TP18 **文献标志码:** A **文章编号:** 0253-2670(2026)20-8013-15

DOI: 10.7501/j.issn.0253-2670.2026.20.008

Artificial intelligence-driven discovery and evaluation of quality markers of traditional Chinese medicine: Data foundation, evidence chain and verification framework

LI Huan^{1,2}, ZHOU Zixin², LI Yuexiao², LIU Yang¹, XUE Rui^{3,4}, WANG Jie⁵, ZHANG Shuangxi¹, SHI Menghua¹, ZHANG Xiang'an¹

1. Department of Anorectology, the First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou 450000, China

2. Henan University of Chinese Medicine, Zhengzhou 450046, China

3. Yunnan University of Chinese Medicine & Yunnan Key Laboratory of Southern Medicine Utilization, Kunming 650500, China

4. Engineering Research Center for Medicine and Food Homologous Beverage of Yunnan Province, Kunming 650500, China

5. Shanghai University of Chinese Medicine, Shanghai 201203, China

Abstract: Quality marker (Q-Marker) of traditional Chinese medicine (TCM) refers to a chemical substance or characteristic combination that exists in TCM, which is closely related to its quality, efficacy and safety, can reflect its quality characteristics, distinguish genuine and inferior products, trace its origin, and be accurately detected. It serves as the core indicator for TCM quality control and evaluation. Traditional TCM quality control mostly relies on empirical judgment and single-component chemical analysis, which suffers from low efficiency, strong subjectivity, and difficulty in balancing chemical specificity, efficacy correlation and overall coordination. Such limitations make it incompatible with the core characteristics of TCM featuring multi-components, multi-targets

收稿日期: 2026-04-16

基金项目: 国家自然科学基金面上项目 (81973625); 河南省科技研发计划联合基金 (242301420114, 252301420095); 河南省科技攻关项目 (232102311221, 262102311235); 河南省中医药科学研究专项 (2024ZY2011); 河南省医学科技攻关项目 (LHGJ20250412)

作者简介: 李 欢, 男, 硕士研究生, 研究方向为中医药防治肛肠病研究。E-mail: 724359388@qq.com

*通信作者: 史孟华, 女, 医学博士, 从事中医药防治肛肠病研究。Tel: 19937104600 E-mail: 358945774@qq.com

张相安, 男, 医学博士, 博士生导师, 主任医师, 从事中医药防治肛肠病研究。Tel: 13592581889 E-mail: zxa2105@163.com

and holistic regulation. With the rapid development of artificial intelligence (AI), technologies including machine learning, deep learning and knowledge graphs have been gradually applied in the whole process of TCM Q-Marker research due to their powerful capabilities in multi-source data fusion, high-dimensional feature mining and nonlinear correlation modeling, realizing the transformation from “experience-driven” to “data-driven” and significantly improving the accuracy, efficiency and scientificity of Q-Marker screening. This paper systematically reviews the multi-source data basis, AI preprocessing techniques, core algorithms and research paradigms of AI-driven Q-Marker research in TCM, analyzes the current challenges and discusses future development directions, so as to provide references for the intelligent research and standardized application of TCM Q-Marker.

Key words: artificial intelligence; traditional Chinese medicine; quality marker; deep learning; multigenomics knowledge graph

中医药以多成分、多靶点、整体调节为核心特点,但其基原混杂、产地多变、炮制各异、配伍复杂,导致质量控制长期面临指标单一、脱离药效、难以溯源、标准不统一等困境^[1]。2016 年刘昌孝院士^[2]提出中药质量标志物 (quality marker, Q-Marker) 核心概念,明确其应具备特有性、有效性、可测性、传递与溯源性、配伍关联性 5 大属性,构建起“物质基础-作用机制-质量控制”一体化研究框架。

随着色谱、质谱、光谱、多组学技术快速发展,中药质量研究进入数据密集型时代,高维、异构、小样本、强噪声数据使传统统计方法难以胜任深度挖掘^[3]。在人工智能 (artificial intelligence, AI) 技术广泛普及前,化学计量学已将多元统计分析与机器学习方法融入中药指纹图谱研究,形成了成熟的方法体系;随着 AI 技术的演进,经典机器学习与深度学习方法进一步突破了传统方法的局限,在特征提取、模式识别、关联推理上展现出更强优势,已广泛应用于中药真伪鉴别、产地溯源、炮制工艺优化等场景,并逐步渗透至 Q-Marker 全链条研究^[4]。然而该领域仍缺乏统一的方法演进脉络梳理、证据分级体系与可解释模型框架,且存在 Q-Marker 概念边界模糊、方法价值定位不清等问题。基于此,本文结合近年来国内外相关研究成果,系统梳理从传统化学计量学到 AI 的方法演进路径,明确不同技术的增量价值,构建 7 级证据分级体系 (分类鉴别→候选标志物→结构确证→药效关联→传递验证→可测性验证→标准转化),归纳提出 AI 驱动的 Q-Marker 5 类研究路径,完善 AI 辅助的 Q-Marker 验证与评价体系,剖析当前研究面临的核心挑战,展望未来发展方向,为推动 Q-Marker 研究的智能化、标准化、产业化发展提供参考。

1 Q-Marker 研究的数据基础与 AI 预处理体系

AI 模型的训练与优化依赖高质量、多维度的数据支撑,Q-Marker 研究的多源数据涵盖化学、生物、文本、图像等多个维度,不同类型数据的特性差异

较大,需通过 AI 预处理技术消除干扰、统一格式、提取核心特征,为后续模型构建奠定基础^[5]。

1.1 Q-Marker 研究的多源数据类型

中药 Q-Marker 的特有性、有效性、可测性、传递与溯源性、配伍关联性 5 大核心属性需要多维度数据从物质基础、生物效应、理论依据等不同层面共同支撑^[6]。围绕 Q-Marker 发现、验证与评价的全流程需求,当前研究的核心数据可分为化学分析、元生物效应、文本知识、图像特征 4 大类,各类数据从不同维度支撑 Q-Marker 的对应属性,辅以基原、产地、炮制工艺、贮藏条件等工艺辅助数据,形成覆盖“化学物质-生物效应-理论依据-外观表征”的多维度数据集,为 AI 模型挖掘特征关联、筛选高质量 Q-Marker 提供全面的数据基础。4 类核心数据的具体信息见表 1。

1.2 多源数据的 AI 预处理技术

中药 Q-Marker 研究的多源数据存在高维异构、批次差异显著、噪声干扰强、样本异质性大等特有问题,易导致假阳性候选标志物、量效关系偏差、溯源准确率下降等结果,直接影响 AI 模型对 Q-Marker 的识别精度与泛化能力。针对上述问题,需通过数据清洗、标准化、特征提取、数据融合 4 级预处理流程,消除系统干扰、统一量纲尺度、浓缩核心信息,构建适配 Q-Marker 筛选的标准化数据集,为后续 AI 建模提供可靠数据基础^[11]。

1.2.1 数据清洗 核心是消除数据中的噪声、缺失值、异常值,确保数据的准确性与完整性。针对不同类型数据采用差异化清洗策略:①化学分析数据,采用平滑处理 [如 Savitzky-Golay (SG) 滤波] 消除光谱、色谱数据中的噪声,通过插值法 (线性插值、拉格朗日插值) 填补缺失值,采用 Z-score 法、四分位距 (interquartile range, IQR) 法识别并剔除异常值^[12]。其中 SG 滤波通过移动窗口内的多项式拟合实现噪声平滑^[13]。②生物效应数据,采用多重插补法填补缺失的活性数据,通过箱线图法剔除异

表 1 中药 Q-Marker 研究多源数据分类与特征

Table 1 Classification and characteristics of multi-source data in research on traditional Chinese medicine Q-Marker					
数据类型	对应 Q-Marker 核心属性	核心研究价值	主要检测/数据内容	典型 Q-Marker 研究应用	文献
化学分析数据	特有性、可测性、传递溯源性	筛选化学特异性 Q-Marker 的核心依据；补充无机特征维度，发现元素 Q-Marker	色谱数据（HPLC、UPLC、GC 等）、光谱数据（NIRS、拉曼、NMR 等）、质谱数据（LC-MS/MS 等）；无机元素含量与元素指纹（LIBS、ICP-OES、ICP-MS 等）	结合 AI 无监督算法挖掘指纹差异峰，初筛基原/产地专属 Q-Marker；通过定量分析验证标志物可测性；挖掘道地药材元素特征，筛选元素质量标志物；支撑矿物药 Q-Marker 体系建立	7
生物效应数据	有效性、安全性	验证 Q-Marker 药效关联的核心支撑	体外细胞活性数据、体内动物药效数据、毒理学数据、多组学数据	结合 AI 关联算法构建谱效关系，筛选与药效直接相关的机制型 Q-Marker，量化其活性贡献度	8
文本知识数据	配伍关联性、特有性	提供理论先验，引导虚拟筛选	中医药古籍、现代期刊文献、行业标准、中药数据库（TCMSP、TCMID 等）	结合 NLP 与知识图谱挖掘成分-靶点-药效关联，辅助筛选复方配伍环境下的核心 Q-Marker	9
图像特征数据	可测性、传递溯源性	支撑 Q-Marker 快速无损检测	饮片外观图像、显微结构图像、高光谱图像	结合 CV 算法建立图像特征与内在 Q-Marker 含量的关联模型，实现现场快速定量预测	10

常实验数据，确保实验数据的可靠性。③文本数据，采用去重、去停用词、纠错处理，消除文本中的冗余信息与错误内容，例如剔除“的、了、和”等无意义停用词，修正文本中的错别字与术语错误。④图像数据，采用图像去噪（高斯滤波、中值滤波）、图像增强（对比度调整、直方图均衡化）、图像分割等技术，消除图像中的噪声干扰，突出图像的核心特征^[14]。

1.2.2 数据标准化 核心是消除不同维度、不同单位数据之间的量纲差异，使数据处于同一量级，避免因数据量级差异影响 AI 模型的训练结果。常用标准化方法包括 Z-score 标准化、Min-Max 标准化、归一化等^[15]。①Z-score 标准化：适用于服从正态分布的数据，将数据转化为均值为 0、标准差为 1 的标准化数据；②Min-Max 标准化：将数据映射到[0, 1] 区间，适用于非正态分布的数据；③色谱/光谱数据：采用峰对齐、归一化处理，确保不同批次、不同仪器检测的数据具有可比性。

1.2.3 特征提取 核心是从高维数据中提取与 Q-Marker 高度相关的核心特征，降低数据维度，减少冗余信息，提升 AI 模型的训练效率与精度。针对不同类型数据采用差异化特征提取方法，①化学分析数据：采用峰识别、峰匹配、特征峰筛选等方法，提取色谱/光谱中的特征峰（如保留时间、峰面积、峰高）作为核心特征；采用主成分分析（principal component analysis, PCA）、偏最小二乘判别分析（partial least squares discriminant analysis, PLS-DA）等降维算法，从高维特征中提取主成分，实现特征降维^[16]。②生物效应数据：采用特征选择算法（如

互信息法、 χ^2 检验），提取与药效、安全性高度相关的生理指标、活性数据作为核心特征。③文本数据：采用自然语言处理（natural language processing, NLP）技术，通过词袋模型、词频-逆向文件频率（term frequency-inverse document frequency, TF-IDF）、Word2Vec、BERT 等方法，将文本转化为数值型特征向量，提取文本中的关键信息（如中药成分、药效、靶点等）^[17]。④图像数据：采用计算机视觉（computer vision, CV）技术，通过边缘检测、轮廓提取、特征点检测等方法，提取图像的纹理、颜色、形状等核心特征，转化为数值型特征向量^[18]。

1.2.4 数据融合 核心是将不同类型、不同维度的预处理后数据进行整合，构建统一的特征矩阵，充分挖掘多源数据之间的内在关联，提升 AI 模型的筛选精度。常用数据融合策略包括早期融合（数据级融合）、中期融合（特征级融合）、晚期融合（决策级融合）^[19]。①早期融合：将原始数据经过清洗、标准化后直接整合，适用于数据类型相近的场景（如不同色谱数据的融合）；②中期融合：将不同类型数据的提取特征进行整合，构建统一的特征集，是中药多源数据融合的主流策略，适用于多类型数据的融合（如化学数据与生物数据的融合）；③晚期融合：将不同 AI 模型的预测结果进行整合，通过投票、加权平均等方法得到最终结果，适用于多模型协同筛选的场景。例如，将色谱数据提取的化学特征与生物效应数据提取的活性特征进行中期融合，构建统一的特征矩阵，用于 Q-Marker 的筛选，可同时兼顾化学特异性与药效有效性^[20]。

综上，通过上述 AI 预处理技术可有效消除多

源数据中的干扰信息,提取核心特征,构建标准化、高质量的数据集,为后续 AI 算法建模与 Q-Marker 筛选奠定坚实基础。

2 中药 Q-Marker 研究方法演进:从化学计量学到 AI

2.1 方法演进脉络:从化学计量学到 AI

在 AI 技术广泛普及之前,化学计量学已将多元统计分析与机器学习方法系统融入中药指纹图谱与谱效关系研究,形成了以 PCA、PLS-DA、系统聚类分析(hierarchical cluster analysis, HCA)、判别分析(discriminant analysis, DA)为核心的方法体系,解决了中药多变量数据的降维、模式识别与差异成分初筛问题^[21]。需要明确的是,AI、机器学习与深度学习是包含而非并列关系:AI 是最广义的概念,机器学习是 AI 的核心子领域,深度学习则是机器学习中基于多层神经网络的分支。因此,本文所述“AI 驱动”并非对化学计量学方法的简单替代,而是方法链条的自然延伸^[22]。从方法演进角度,可将中药 Q-Marker 研究中的计算方法划分为 3 个相互衔接的阶段:第 1 阶段为化学计量学时期(以 PCA、PLS-DA、HCA 为代表),主要依赖人工特征工程与线性/广义线性模型,解决“差异成分可视化与初筛”问题^[23];第 2 阶段为以支持向量机(support vector machine, SVM)、随机森林(random forest, RF)、逻辑回归(logistic regression, LR)、K 近邻(K-nearest neighbor, KNN)、最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)为代表的经典机器学习时期,引入核方法、集成学习与正则化特征选择,解决“小样本高维数据下的非线性分类与特征重要性排序”问题^[24];第 3 阶段为以卷积神经网络(convolutional neural network, CNN)、循环神经网络(recurrent neural network, RNN)、图神经网络(graph neural network, GNN)、自编码器(autoencoder, AE)、Transformer 为代表的深度学习时期,通过层级化自动特征提取与注意力机制,解决“高维原始信号直接建模、多模态融合与复杂关联网络推理”问题^[25]。从增量价值看,经典机器学习相对于化学计量学的主要进步在于:在相同样本量下提升了外部验证集的分类/回归精度,并能输出特征重要性排序(如 RF 的 Gini 重要性、LASSO 的系数稀疏化),从而将差异峰筛选从“载荷图目视判断”推进到“定量排序”;深度学习相对于经典机器学习的增量价值则主要体现

在 3 方面:①可直接处理光谱、色谱、图像等原始高维信号,减少人工峰识别与特征工程引入的信息损失;②通过多组学融合与图网络建模,能够发现传统线性方法难以捕捉的“成分-靶点-通路”非线性协同关系,部分研究报道了传统方法遗漏的低丰度差异成分或新的协同作用通路;③借助 Transformer 与大模型实现了从海量文献、多模态数据中进行知识推理与虚拟筛选。但必须客观指出,目前多数已发表研究中,深度学习相对于经典机器学习的增量仍主要表现为分类准确率的提升,而非系统性发现传统方法遗漏的新化学成分;且深度学习模型在小样本中药数据集上的外部批次验证与跨实验室验证仍不充分,其增量价值有待更多对照研究证实^[26]。本文按“经典机器学习-深度学习-集成与小样本学习”的顺序,分别讨论各类方法在 Q-Marker 研究中的应用。

2.2 面向特征筛选与分类的传统机器学习算法

经典机器学习以结构化数据为主要处理对象,具备可解释性强、样本门槛低、结果稳健的特点,与化学计量学方法一脉相承。目前应用于 Q-Marker 研究的传统机器学习模型主要包括 SVM、RF、PLS、LR、KNN 及正则化特征筛选算法等,常规应用流程为多源数据清洗与标准化、Q-Marker 相关特征筛选与编码、模型训练与调优、候选标志物优先级排序。

2.2.1 RF RF 属于集成学习框架下的树模型,通过构建多棵决策树并以投票或均值方式输出最终结果,既能降低单模型的过拟合风险,又能输出特征重要性排序,可高效处理指纹图谱这类多峰、多变量的异构数据,常用于药效关联 Q-Marker 筛选、成分效价预测等研究^[27]。RF 通过计算特征重要性(Gini 系数、袋外误差),筛选出对模型预测贡献最大的特征(如化学成分)作为候选 Q-Marker^[28]。富尧等^[29]构建生柴胡与炒柴胡指纹图谱,在《中国药典》2025 年版原有柴胡皂苷 A、柴胡皂苷 D 测定指标基础上新增柴胡皂苷 B₁、柴胡皂苷 B₂、柴胡皂苷 C 及醇浸出物评价,联合 HCA、PCA、正交偏最小二乘判别分析(orthogonal partial least squares discriminant analysis, OPLS-DA)化学计量学,融合变量重要性投影值、人工神经网络与 RF 筛选得到柴胡皂苷 A、柴胡皂苷 B₁、柴胡皂苷 B₂、柴胡皂苷 D 为炒柴胡 Q-Marker,可用于炮制品定性定量及炮制过程动态监测,为柴胡多成分质量控制、炮制工

艺数字化及炮制技术传承提供数据支撑与方法依据。此外, RF 还可用于中药质量等级分类、真伪鉴别, 通过构建分类模型实现不同质量等级、不同基原中药的精准区分^[30]。

2.2.2 SVM SVM 是经典的监督学习分类与回归模型, 通过核函数将低维数据映射至高维特征空间, 以最大间隔原则构建分类超平面, 在小样本、高维特征场景下表现突出, 恰好适配中药光谱、色谱数据维度高、样本量有限的特点, 广泛用于基原与产地专属 Q-Marker 的初筛、标志物特异性判别等场景^[31]。冯佳豪等^[32]采用 SVM 结合近红外光谱 (near infrared, NIR) 数据构建淫羊藿与柔毛淫羊藿的基原鉴别模型, 筛选出 2 类药材的特异性 Q-Marker 特征波段, 模型判别准确率达 100%, 为基原专属标志物的快速筛查提供了高效方法。在黄芪产地溯源研究中, SVM 结合激光诱导击穿光谱 (laser-induced breakdown spectroscopy, LIBS) 和 NIR 数据, 实现不同产地黄芪的精准分类, 同时筛选出与产地相关的特征标志物, 溯源准确率达 96.0%^[33]。

2.2.3 PLS PLS 是多元统计与机器学习结合的经典算法, 通过最大化自变量与因变量的协方差构建关联模型, 包含 PLS-DA 与偏最小二乘回归 (partial least squares regression, PLSR) 2 种形式, 是中药谱效关系研究、差异 Q-Marker 筛选的主流方法^[34]。马盼盼等^[35]基于苍术 UPLC 指纹图谱与抗溃疡性结肠炎活性数据, 采用 PLS-DA 构建谱效关联模型, 筛选出 6 种与药效显著相关的核心 Q-Marker, 明确了标志物含量与药效的量化对应规律。

2.2.4 LR LR 是基于广义线性模型的概率分类算法, 通过 Sigmoid 函数将线性拟合结果转化为类别概率, 可直接量化各特征对分类结果的贡献权重, 兼具建模简洁与结果可解释的优势。在 Q-Marker 研究中, LR 可通过拟合化学成分含量与药材质量等级、真伪属性之间的概率关联, 筛选对分类判别贡献显著的特征成分, 作为候选 Q-Marker^[36]。何家群等^[37]采用 QuEChERS 提取结合超高效液相色谱串联质谱联用技术, 对不同等级酱香型高温大曲中 83 种潜在香气物质进行准确定量分析, 并结合香气活性值评估筛选出 59 种重要香气活性组分, 主要涵盖吡嗪类、酚类、酸类及醛酮类; 在此基础上进一步整合机器学习算法构建高温大曲质量等级评价模型, 比较发现 LR 与 SVM 表现最优, 准确率达 100%; 该研究建立的评价方法为深入解析芳香类中

药 Q-Marker 与质量判定提供了可靠的技术方法。此外, LR 还可用于 Q-Marker 的安全风险评估, 通过标志物含量与不良反应的概率关联, 为中药安全质控阈值的设定提供参考^[38]。

2.2.5 KNN KNN 是基于相似性度量的惰性学习算法, 无需预先训练模型, 通过计算待判别样本与已知样本的特征距离, 以最近邻样本的属性作为判别依据, 原理直观、实现简便, 尤其适合小样本数据集的快速分类判别。在 Q-Marker 研究中, KNN 可通过样本间化学特征的相似性比对, 识别出区分不同基原、产地、等级的特异性成分, 辅助完成候选标志物的初筛^[39]。张云飞等^[40]依托 CM-5 分光测色仪获取薏苡仁色度参数, 构建决策树、KNN、贝叶斯判别机器学习模型, 结合超快速气相电子鼻采集气味信息, 建立判别因子分析与 PLS-DA 模型挖掘产地差异标志物, 并开展颜色特征与气味成分的关联分析; 筛选得到 2-丙烯酸、2-甲基丁醛、甲苯、丙醛为区分不同产地薏苡仁的核心气味标志物, 丁二酮、2,3-乙酰基丙酮、3-己醇等多种气味成分与 L^* 、 a^* 、 b^* 色度参数显著相关, 证实电子眼联合超快速气相电子鼻可实现不同产地薏苡仁快速精准鉴别, 为多产地中药的真伪产地判别与质量控制提供新思路。

2.2.6 正则化特征筛选算法 以 LASSO、弹性网络为代表的正则化算法通过在损失函数中加入惩罚项, 将冗余特征的系数压缩至 0, 可同步实现特征选择与数据降维, 核心解决中药多维成分数据中多重共线性干扰 Q-Marker 优先级判断的问题^[41-43]。吴鑫雨等^[44]基于智能感官多源数据, 采用 LASSO 回归筛选与香附炮制程度高度相关的特征 Q-Marker, 实现了炮制过程的快速辨识, 为炮制过程标志物的动态监测提供了方法支撑。

2.3 面向深层机制解析的深度学习算法

深度学习是机器学习的重要分支, 核心是通过多层神经网络的层级化特征提取机制自动从原始数据中学习复杂的隐性规律, 突破了传统算法依赖人工特征工程的局限, 具备强大的非线性关联建模能力。面向中药“多成分、多靶点、整体调节”的复杂体系, 深度学习可深度挖掘光谱、色谱、多组学、文本等多源数据的内在关联, 实现潜在 Q-Marker 的精准识别与作用机制的系统阐释, 是当前机制型 Q-Marker 发现的核心创新技术^[45]。目前应用于中药 Q-Marker 研究的深度学习模型主要包括 CNN、RNN、

GNN、AE 以及 Transformer 与大模型等,不同模型分别适配不同类型的数据与研究场景。

2.3.1 CNN CNN 是深度学习领域应用最广泛的模型之一,通过卷积层的局部感受野与权值共享机制提取数据的空间结构特征,经池化层压缩维度后由全连接层输出最终结果,无需人工设计特征参数,可自动捕捉数据中的深层模式。在 Q-Marker 研究中,CNN 主要用于光谱、色谱数据的特征挖掘与标志物定量检测,可有效解决重叠峰、低丰度成分难以识别的技术瓶颈,实现 Q-Marker 的快速无损定量^[46]。Gao 等^[47]采用 CNN 融合双向长短期记忆网络(long short-term memory, LSTM)网络,基于 NIR 数据精准预测血府逐瘀口服液 7 种活性 Q-Marker 的含量,同时实现了产品生产过程中的质量动态监测,为 Q-Marker 的传递规律解析提供了成熟的时序建模工具。

2.3.2 RNN RNN 是专门面向时序数据的深度学习模型,通过神经元的循环连接保留序列前后的依赖信息,其改进模型 LSTM 与门控循环单元(gated recurrent unit, GRU)通过门控机制有效解决了长序列训练中的梯度消失问题。在 Q-Marker 研究中,RNN 系列模型主要用于解析标志物在炮制、提取、贮藏、全产业链加工过程中的含量动态变化规律,支撑 Q-Marker 的传递溯源性评价与生产过程质量控制^[48]。Zhang 等^[49]以 354 味抗癌中药及其对应归经、5 471 种成分化合物为数据基础,整合 CancerHSP、ETCM 与 Hit 2.0 公共数据库资源,采用图卷积网络(graph convolutional network, GCN)与 RNN 耦合的深度学习混合模型 GRMC,基于分子图拓扑结构生成归经多标签序列,实现中药多靶点归经的智能分类,可用于 Q-Marker 筛选及预测。

2.3.3 GNN GNN 是针对图结构数据的深度学习模型,通过消息传递机制聚合邻居节点的特征信息,可天然适配中药“成分-靶点-通路-药效”的复杂关联网络,是机制型 Q-Marker 发现的核心创新技术。根据具体架构差异,GNN 可细分为 GCN、图注意力网络(graph attention network, GAT)等变体,分别适用于不同复杂度的网络建模。Zhu 等^[50]基于自建环氧化酶-2(cyclooxygenase-2, COX-2)抑制剂数据集与中药来源化合物库,系统对比随机森林分类器、深度学习及 GAT、GCN、MPNN 等图神经网络模型在 COX-2 抑制剂分类任务中的性能,筛选最优模型挖掘潜在 COX-2 抑制剂,解析调控抑制活性的关键分

子亚结构,并对候选分子开展实验验证,为挖掘潜在中药 Q-Marker 及验证提供了思路。

2.3.4 AE AE 及其变种变分自编码器(variational autoencoder, VAE)属于无监督深度学习模型,通过“编码-解码”的重构结构,将高维原始数据压缩为低维隐性特征,再通过解码重构还原数据,可挖掘传统方法难以识别的隐性特征^[51]。在 Q-Marker 研究中,AE 主要用于无标签数据的特征挖掘,可识别低丰度、弱差异的潜在 Q-Marker,补充标志物筛选的维度与深度^[52]。Wan 等^[53]以对氯苯丙氨酸诱导的失眠大鼠模型为对象,构建融合转录组、代谢组与深度学习的整合分析框架,采用基于 AE 降维与跨模态注意力机制的多组学特征融合策略解决数据异质性问题,共筛选得到 1 847 个差异基因与 286 个差异代谢物,五折交叉验证下深度神经网络分类准确率达 91.2%,曲线下面积(area under curve, AUC)为 0.956,置换检验证实模型性能显著优于随机水平,提示天王补心丹可协同调控神经递质合成与分解代谢通路发挥药效,也为其他中药复方的机制解析提供可复用的计算研究框架。

2.3.5 Transformer Transformer 基于自注意力机制捕捉数据的全局关联,突破了循环模型的序列长度限制,是当前多模态数据融合与知识推理的核心架构,以此为基础的中医药大模型进一步推动了 Q-Marker 研究向端到端、智能化方向升级^[54]。中医药专用大模型可高效挖掘海量文献中的成分-药效关联知识,加速 Q-Marker 的虚拟筛选;多模态大模型可同步整合化学、生物、文本多源数据,实现标志物筛选、机制阐释与检测方法设计的一体化输出。Yuan 等^[55]开发的 HSQCid 工具基于 Transformer 架构实现了天然产物结构的高通量解析,可快速确认候选 Q-Marker 的化学结构,效率较传统人工解析提升数十倍,为 Q-Marker 的结构验证提供了高效工具。多模态大模型是未来 Q-Marker 研究的重要发展方向,后续可进一步融合多组学数据与中医药理论知识,构建一体化的智能筛选体系^[56]。

2.4 面向小样本与泛化提升的集成学习与小样本学习

集成学习与小样本学习是机器学习领域针对模型泛化能力优化的重要技术方向,核心解决中药 Q-Marker 研究中样本异质性强、稀缺药材数据不足、单一算法稳定性弱的现实问题。2 类技术分别从多模型协同决策、样本利用效率优化 2 个维度切

入，有效提升筛选模型的鲁棒性与场景适配性，是推动 AI 辅助 Q-Marker 筛选从实验室研究向实际应用落地的重要支撑^[57]。

2.4.1 集成学习 集成学习通过组合多个结构、原理各异的基学习器，以投票、加权平均等群体决策方式输出最终结果，可显著降低单一模型的过拟合风险与结果偏差，提升预测的稳定性与泛化能力，常用实现策略包括 Bagging、Boosting、Stacking 3 类^[58]。在 Q-Marker 研究中，集成学习可整合不同算法的筛选优势，降低单一样本、单一建模方式带来的假阳性偏差，常用于 Q-Marker 含量的精准预测、多维度质量分级与综合评价。谢文涌等^[59]提出基于叶片子区间多特征提取和 Stacking 集成学习的金线莲品系分类方法，分类正确率达 94.49%，*F* 值为 93.91%，显著优于单一分类模型。周正等^[60]采用 Bagging 策略集成多个偏最小二乘基模型，将黄柏提取过程中 Q-Marker 小檗碱含量预测的均方根误差从 0.961 4 降至 0.413 6，模型稳定性与预测精度均显著优于单一模型，实现了标志物含量的在线精准监测。

2.4.2 小样本学习 小样本学习旨在样本量有限的条件下完成模型的有效训练，核心技术涵盖迁移学习、数据增强、元学习等，可破解珍稀濒危药材、经

典名方、罕见病证相关中药等实验样本稀缺场景下的 Q-Marker 筛选难题^[61]。其中，迁移学习通过将通用数据集上训练得到的预训练模型迁移至目标小样本场景，大幅降低对目标领域样本量的要求^[62]；数据增强则通过对现有样本进行符合专业逻辑的变换扩充训练集提升模型的泛化表现。郭兆华等^[57]采用光谱数据增强结合 CNN，在小样本条件下实现白芷产地溯源准确率 94.7%，显著优于传统 SVM 模型，为道地药材专属 Q-Marker 的筛选提供了可行的技术路径。小样本学习大幅降低了 AI 模型的样本门槛，拓展了 AI 技术在稀缺药材 Q-Marker 研究中的应用范围；但模型迁移效果受源域与目标域数据相似度的影响较大，后续可结合元学习框架进一步提升小样本场景下的模型快速适配能力^[63-64]。

综上，集成学习与小样本学习并非独立于传统机器学习、深度学习的算法，而是通过优化模型组合、样本利用效率，弥补单一算法在中药 Q-Marker 研究中的不足，与传统机器学习、深度学习算法协同作用，构建适配中医药数据特点的 Q-Marker 筛选模型，提升筛选的精准度、稳定性与泛化能力。以上模型的适用场景及优缺点见表 2，需根据不同的基础数据及需求选择合适准确的算法模型。

表 2 常用 AI 模型在中药 Q-Marker 研究中的适用场景及优缺点

Table 2 Applicable scenarios, advantages and disadvantages of common AI models in Q-Marker research of traditional Chinese medicine			
模型	适用场景	优点	缺点
SVM	基原/产地专属 Q-Marker 初筛、标志物特异性判别 ^[27-28]	小样本高维数据表现优异、稳定性好	参数调优敏感、大规模数据处理效率低
RF	药效关联 Q-Marker 筛选、特征重要性排序 ^[24]	抗过拟合能力强、可输出特征权重、对缺失值不敏感	对低丰度弱差异标志物识别灵敏度有限
PLS	谱效关系建模、差异 Q-Marker 筛选与量效分析 ^[30]	小样本下稳定、可同步实现筛选与关联建模	难以捕捉多成分非线性协同作用
LR	质量等级关联 Q-Marker 筛选、标志物安全风险评估 ^[31-33]	可解释性强、结果易转化为质控标准	为线性模型，无法刻画复杂非线性关联
KNN	小样本 Q-Marker 快速初筛、基原属性分类 ^[34-35]	原理直观、无需预训练、小样本可用	高维大数据计算效率低、对特征尺度敏感
CNN	Q-Marker 快速无损定量、低丰度成分识别 ^[42]	自动提取空间特征、检测精度高	需大量标注数据、文本类数据适配性差
RNN	Q-Marker 传递溯源规律解析、全产业链过程质控 ^[44]	擅长处理时序数据、捕捉动态变化规律	长序列易出现梯度消失、训练难度较高
GNN	机制型 Q-Marker 发现、配伍关联网络分析 ^[45]	适配图结构数据、可阐释作用机制	训练复杂度高、可解释性弱
AE	无监督潜在 Q-Marker 挖掘、低丰度隐性特征识别 ^[48]	无需标签数据、可挖掘深层隐性特征	无监督训练可控性弱、结果需实验验证
Transformer	多模态 Q-Marker 筛选、标志物结构解析与虚拟筛选 ^[50]	捕捉全局关联、支持多模态融合、知识推理能力强	训练数据与算力需求高、存在生成幻觉风险
集成学习	Q-Marker 多维度综合评价、标志物含量精准预测 ^[55]	鲁棒性强、泛化能力优于单一模型	模型复杂度高、可解释性低于单模型

3 AI 驱动的 Q-Marker 发现研究路径

基于 AI 的多源数据整合、高维特征挖掘与非线性关联建模能力，结合 Q-Marker “特有性、有效性、可测性、传递溯源性、配伍关联性” 5 大核心属性，笔者构建了中药 Q-Marker 标准化、可重复、高效化的研究流程，打破了传统 “单一成分

筛选、经验判断主导” 的局限，实现了 “数据驱动-模型挖掘-机制验证-标准落地” 的闭环，见图 1。需要说明的是，以下 5 类研究路径系本文基于现有文献归纳提出的分析框架，并非领域已形成共识的固定分类；其覆盖从初筛到机制阐释的全流程，见表 3。

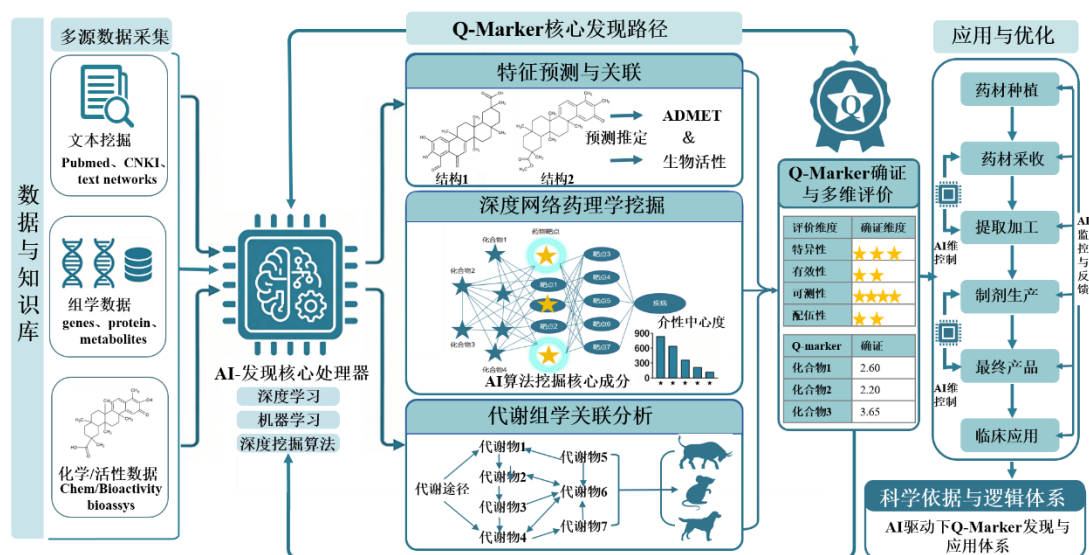


图 1 基于 AI 的中药 Q-Marker 发现与评价框架

Fig. 1 Discovery and evaluation framework for Q-Marker of traditional Chinese medicine based on AI

表 3 5 类 AI 驱动中药 Q-Marker 研究路径对比

Table 3 Comparison of five types of AI-driven research pathways for traditional Chinese medicine Q-Markers

研究路径	核心痛点	关键流程	优势	应用边界	适用条件
化学指纹+AI 无监督	缺乏快速初筛差异标志物的手段	指纹采集→预处理→降维聚类→差异成分筛选→验证确证	快速、简便、无需药效数据	仅反映化学差异，无法关联药效，适用于初筛	仅具备化学/光谱数据，无药效数据
多组学+AI 关联	难以解析多成分-多靶点非线性关联	多组学检测→数据融合→关联建模→筛选→药效验证	机制阐释，一体化关联	成本高、样本量大，需多组学检测支撑	具备多组学与药效数据
文本知识+AI 知识图谱	珍稀药材实验数据匮乏，难以衔接传统理论	文本采集→知识抽取→图谱构建→推理→候选筛选→实验验证	挖掘传统理论，低成本快速筛选	结果为预测性，需实验验证，适用于虚拟筛选	文献丰富、实验样本稀缺
图像特征+AI 计算机视觉	外观评价依赖经验，难以关联内在成分	图像采集→预处理→特征提取→特征-成分关联→筛选→验证	快速、无损、现场检测	受采集条件影响大，间接关联标志物	饮片现场快速检测、质量分级
多模态融合+AI 端到端	单模态数据片面，难以兼顾 5 大属性	多模态采集→预处理→特征融合→端到端建模→多维度验证	深度融合，端到端智能，效率高精准	技术门槛高，可解释性待提升，是前沿方向	具备多源数据的复杂中药体系

3.1 化学指纹+AI 无监督研究路径

该范式是中药 Q-Marker 研究中应用最广、最基础的方法，依托化学指纹图谱（HPLC、UPLC、近红外光谱、拉曼光谱等）的整体特征性，结合 AI 无监督学习算法（如 PCA、自编码器、K-means 等），无需先验标签即可挖掘指纹差异特征，快速筛选与基原、产地、炮制工艺等质量属性相关的候选标志物，适用于道地溯源、真伪鉴别及未知质量差异机制的初筛^[32]。黄冬兰等^[65]采用红外三级鉴定与 K-means 聚类法进行党参快速鉴别，采集纹党参与白条党参共 49 个样本的红外光谱，经一维、二阶导数和二维相关红外分析后，利用 K-means 聚类实现分类，二维相关红外光谱显示纹党参在 1459、1298、1209 cm^{-1} 处有 3 个强自动峰（以 1209 cm^{-1} 最强），

而白条党参仅在 1450 cm^{-1} 处有 1 个强自动峰，其余自动峰极弱，该方法为相似中药的快速鉴别提供了有效手段。

3.2 多组学+AI 关联研究路径

该范式以系统生物学为核心，整合代谢组、蛋白组、转录组等多维数据，结合 AI 关联建模，挖掘“组学特征-中药质量-生物效应”的内在关联，筛选机制型 Q-Marker，实现“成分-靶点-通路-药效”一体化，适用于药效物质基础解析与机制阐释。Gao 等^[66]通过 CNN 学习 LC-MS/MS 谱图特征，并联合候选结构指纹的 Tanimoto 相似度计算，实现了代谢产物的注释，进而为中药体内代谢产物的鉴定与预测提供了有力支持。所鉴定的代谢产物可映射至原型成分及代谢途径，进而筛选出能在体内暴露

并传递至靶器官的 Q-Marker, 为基于体内传递溯源的 Q-Marker 发现提供了有力工具。

3.3 文本知识+AI 知识图谱研究路径

该范式以中医药传统理论与现代文献为核心, 依托 NLP 与知识图谱技术, 构建“中药-成分-药效-靶点-通路”知识网络, 通过知识推理挖掘潜在 Q-Marker, 实现传统理论与现代科技融合, 适用于理论筛选及缺乏实验数据的珍稀中药、经典名方。此研究范式整合了中医药传统知识与人工智能技术, 为中药 Q-Marker 的发现提供了智能化新路径, 如相关中医药知识挖掘平台^[67]采用此路线, 通过图算法从已有知识中推理出与功效强相关的成分节点, 作为候选 Q-Marker, 实现了传统理论与现代智能技术的融合。

3.4 图像特征+AI 计算机视觉研究路径

该范式以中药外观、显微特征为核心, 依托计算机视觉与深度学习, 自动提取图像特征并关联内在质量成分, 筛选反映真伪、等级的 Q-Marker, 实现快速无损检测, 适用于饮片鉴别、质量分级、基原区分^[68]。在中药 Q-Marker 研究中, 采集不同产地与等级的中药饮片及显微图像, 经中值滤波去噪和直方图均衡化增强后, 采用 CNN 提取纹理、颜色及细胞结构特征, 结合 HPLC 测定的活性组分含量构建随机森林关联模型, 筛选出候选标志物, 其中纹理密度可能与差异成分含量显著正相关^[69]。孙佳琪等^[70]进一步利用高光谱成像, 提取图像灰度共生矩阵纹理特征及小波能量, 并与特征波长光谱值融合, 构建 XGBoost 模型, 实现了对紫苏 Q-Marker 紫苏醛含量的高精度预测 (测试集 $R^2=0.989\ 44$, 预测均方根误差 $=0.036\ 40$), 证实图像纹理与特征波长光谱可作为 Q-Marker 含量快速无损预测的间接代理变量; 需注意图像/高光谱特征本身并非 Q-Marker, 其预测目标 (如紫苏醛) 才是经化学确证的 Q-Marker。

3.5 多模态融合+AI 端到端研究路径

该范式是当前前沿方向, 整合化学、生物、文本、图像等多模态数据, 依托多模态深度学习与 Transformer 大模型, 实现“数据输入-特征融合-模型挖掘-标志物筛选-验证评价”的端到端智能化流程, 无需人工分步处理, 充分挖掘多模态数据内在关联, 筛选兼具化学特异性、药效有效性、传递溯源性的核心标志物, 适用于复杂中药制剂、道地药材、濒危药材的高质量研究。梁振忠等^[71]提出的基

于 Transformer 改进的中药-靶点相互作用 (traditional Chinese medicine herb-target interaction, TCMHTI) 模型预测出青附蠲痹汤治疗类风湿关节炎的肿瘤坏死因子- α (tumor necrosis factor α , TNF- α)、白细胞介素- 1β (interleukin- 1β , IL- 1β) 等 9 个核心靶点。TNF- α 、IL- 1β 等属于药效作用靶点 (疾病相关生物标志物), 并非中药 Q-Marker, 该类研究的意义在于以预测得到的核心靶点为反向线索, 引导研究者在复方中富集筛选能够调控上述靶点的化学成分, 进而通过谱效关联与实验确证锁定真正的化学 Q-Marker, 从而为 Q-Marker 筛选提供机制导向的新依据, 模型预测准确性优于经典网络药理学方法。

4 AI 辅助的中药 Q-Marker 验证与评价体系构想

AI 不仅能实现中药 Q-Marker 的高效筛选与机制挖掘, 还能为 Q-Marker 的验证与评价提供标准化、智能化技术支撑, 破解传统验证方法效率低、评价维度单一、标准不统一的难题。基于 AI 技术构建的 Q-Marker 验证与评价体系, 以“特异性、有效性、可测性、传递溯源性、配伍关联性”5 大核心属性为评价准则, 整合多源数据与多算法模型, 实现从实验室验证到行业标准落地的全链条评价, 确保筛选出的 Q-Marker 科学、可靠、可推广, 具体涵盖验证技术、评价维度与标准化流程 3 部分, 见图 2。

4.1 Q-Marker 研究的证据分级框架

在系统梳理现有文献时可以发现, 不同研究所完成的工作深度差异极大, 部分研究仅完成了基于光谱/指纹图谱的分类与鉴别 (即样品能否被正确区分), 部分研究进一步筛选出了差异化学成分, 但未进行结构确证与药效关联, 少数研究则完整覆盖了从候选 Q-Marker 发现到标准转化的全链条。为客观评价各文献在 Q-Marker 发现与验证链条上所处的位置, 本文提出 7 级证据分级框架, 由低到高依次为 I 级 (分类/鉴别证据): 仅证明不同基原、产地、等级或炮制规格的样品可被算法正确区分, 未指向具体化学成分; II 级 (候选化学标志物证据): 通过差异分析、特征重要性排序等筛选出候选化学成分 (或差异峰), 但未完成结构确证; III 级 (结构确证证据): 候选成分经 LC-MS/MS、NMR 等完成结构鉴定与指认; IV 级 (药效关联证据): 候选成分含量与体外/体内药效数据建立显著量效或谱效关联, 或通过受体/酶学实验验证其活性; V 级 (传递验证证据): 候选成分在药材-饮片-中间体-制剂

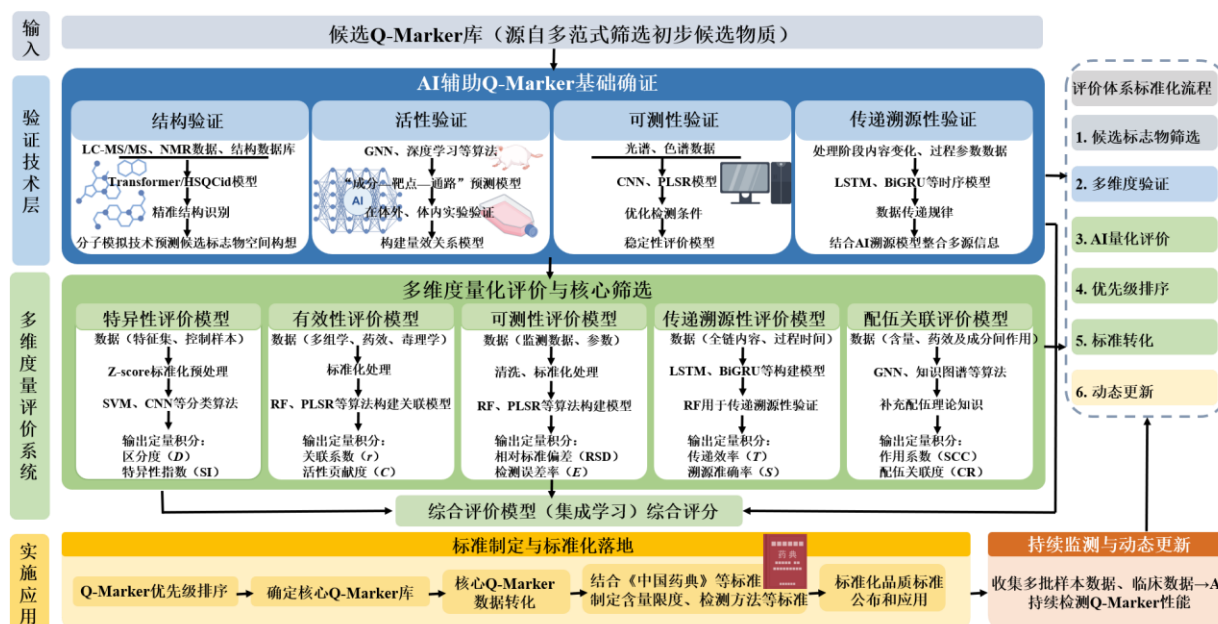


图2 AI辅助中药Q-Marker验证与评价构想技术路线

Fig. 2 Conceptual technical route for AI-assisted verification and evaluation of Q-Markers in traditional Chinese medicine

产业链各环节可被稳定追踪，其传递规律明确；VI级（可测性验证证据）：候选成分可经常规色谱/光谱方法准确定量，方法学验证（专属性、精密度、重复性、稳定性）符合《中国药典》要求；VII级（标准转化证据）：候选成分已纳入或具备纳入药典/行业标准的含量限度与检测方法，并通过外部批次与跨实验室验证。现有文献多数AI相关研究处于I~III级，即完成了分类鉴别与候选成分筛选，但同时完成IV~VII级验证的研究仍属少数；尤其缺乏外部批次验证与跨实验室验证支撑的研究，其结论尚不足以直接支撑Q-Marker的标准转化。本文将围绕IV~VII级验证所需的技术、量化评价与标准化流程展开。

4.2 AI辅助的Q-Marker验证技术

Q-Marker的验证是确认其科学性与实用性的关键环节，核心包括结构验证、活性验证、可测性验证与传递溯源性验证4大模块，AI技术可显著提升验证效率与精准度，实现验证过程的智能化与标准化。

4.2.1 结构验证 核心是确认候选Q-Marker的化学结构，避免成分鉴定误差，AI技术可辅助实现化合物结构的快速解析与确认。常用AI技术包括基于Transformer的化合物结构解析模型（如HSQCid模型）、分子结构预测模型，可结合LC-MS/MS、核磁共振波谱法等检测数据快速匹配化合物结构

数据库，实现结构的精准鉴定，相比传统人工解析方法，效率、准确率均可得到提升。此外，AI分子模拟技术可预测候选标志物的空间构象，辅助确认其与靶点的结合位点，为活性验证提供理论支撑。

4.2.2 活性验证 核心是确认Q-Marker与中药药效、安全性的关联，明确其生物活性与作用机制，AI技术可实现活性的快速预测与高效验证。一方面，通过GNN、深度学习等算法构建“成分-靶点-通路”预测模型，快速预测候选标志物的潜在活性与作用机制，缩小体外、体内实验的范围，降低实验成本；另一方面，AI辅助的高通量筛选技术（如基于CNN的细胞活性检测模型）可实现多候选标志物的同步活性验证，快速筛选出具有明确生物活性的标志物。同时，AI可整合多批次活性实验数据，构建量效关系模型，明确Q-Marker含量与药效的关联规律，为含量限度制定提供支撑。

4.2.3 可测性验证 核心是确认Q-Marker可通过常规检测技术精准定量，满足质量控制的实际需求，AI技术可辅助优化检测方法、提升检测精度。常用AI技术包括CNN、PLSR等，可结合光谱、色谱数据，优化检测条件（如色谱柱选择、流动相比、检测波长），实现标志物的快速、精准定量，同时解决传统检测中重叠峰、微弱峰难以识别的问题^[72]。此外，AI可构建检测方法的稳定性评价模型，自动监测检测过程中的误差，确保检测结果的可靠

性与重复性。

4.2.4 传递溯源性验证 核心是确认 Q-Marker 在中药产业链（药材-饮片-复方制剂）中的传递规律，实现质量溯源与全程质控。AI 技术可实现传递规律的动态监测与精准溯源，通过 LSTM、BiGRU 等时序模型，分析 Q-Marker 在不同加工环节（炮制、提取、制剂）的含量变化数据，捕捉其传递规律，明确加工过程对标志物含量的影响，优化加工工艺；同时，结合 AI 溯源模型（如 SVM、RF），整合产地、加工工艺、检测数据等多源信息，实现中药质

量的全链条溯源，确保 Q-Marker 的可追溯性。

4.3 多维度评价体系与 AI 建模

基于中药 Q-Marker 的 5 大核心属性，或许可尝试构建多维度评价体系，结合 AI 建模技术，实现 Q-Marker 的全面、客观评价，明确其优先级与应用价值，为质量标准制定提供科学依据。评价体系围绕上述 5 大属性展开，均遵循“数据准备-AI 建模-量化评价-优化”的通用框架，但各维度在数据输入、模型选择、核心公式及阈值依据上存在显著差异。各维度的关键量化公式与评价标准汇总见表 4。

表 4 5 大核心属性量化评价公式与指标

Table 4 Five core attribute quantitative evaluation formulas and indicators

评价维度	量化公式	符号说明与推荐模型	合格阈值及依据
特异性	$D=1-\frac{\bar{X}_t-\bar{X}_c}{\max(\bar{X}_t,\bar{X}_c)}, SI=\frac{TP}{TP+FP}$	X_t, X_c 分别为目标与对照样品特征均值；TP 为正确识别目标中药的样本数，FP 为误判样本数；常用 SVM、CNN	区分度 (D) ≥ 0.8 （参考化学计量学分类判据 ^[73] ）；特异性指数 (SI) ≥ 0.85
有效性	$r=\frac{n\sum_{i=1}^n \hat{y}_i y_i - (\sum_{i=1}^n \hat{y}_i)(\sum_{i=1}^n y_i)}{\sqrt{[n\sum_{i=1}^n \hat{y}_i^2 - (\sum_{i=1}^n \hat{y}_i)^2][n\sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2]}}$ $C=\frac{w_i}{\sum_{k=1}^m w_k}$	x_i, y_i 分别为第 i 个样本的标志物含量与药效值， n 为样本数； w_i 为模型特征重要性权重；常用 RF、PLSR，SHAP 值辅助	关联系数 $ r \geq 0.8$ ；活性贡献度 (C) 由 SHAP 值推导； $R^2 \geq 0.8$, RMSEP ≤ 0.1 （参照 QbD 指导原则 ^[74] ）
可测性	$A=\frac{TP+TN}{TP+TN+FP+FN}, RSD=\frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2}}{\bar{x}}, E= \frac{x^2-x_0}{x_0} $	TP/TN 为正确定量/判为未检出样本数，FP/FN 为误判样本数； x_i 为重复测定值， $\bar{x}$ 为均值； $\hat{x}, x_0$ 分别为模型预测值与实际检测值；常用 CNN、PLSR	检测准确率 (A) $\geq 95\%$ ；RSD $\leq 2\%$ （参照《中国药典》分析方法验证指导原则 ^[75] ）；检测误差率 (E) $\leq 5\%$
传递溯源性	$T=\frac{X_k}{X_{k-1}}, S=\frac{\text{正确溯源的样本数}}{\text{总溯源样本数}}$	X_k, X_{k-1} 为第 k 与 $k-1$ 环节的标志物含量均值；常用 LSTM、BiGRU（传递预测）、RF（溯源）	传递效率 (T) $\geq 80\%$ （单环节）、全链均值 $\geq 75\%$ ；溯源准确率 (S) $\geq 95\%$ （参照食品/药品追溯系统阈值 ^[76] ）
配伍关联性	$SCC=\frac{Y_{1+2}}{Y_1+Y_2}, CR=\frac{\sum_{j=1}^k w_{ij} \times r_{ij}}{k}$	Y_{1+2} 为联合药效， Y_1, Y_2 为单独药效； w_{ij}, r_{ij} 为第 j 个成分的交互权重与关联系数， k 为成分数；常用 GNN、知识图谱	协同作用系数 (SCC) > 1 ；配伍关联度 (CR) ≥ 0.7

基于表 4 的 5 项指标得分，可构建集成学习综合评价模型（如 Stacking 集成），通过加权或投票机制对候选 Q-Marker 进行综合评分与优先级排序。上述公式与阈值多源自相关领域已有文献或指导原则的推荐，实际应用中需根据具体品种和检测条件进行校正与优化。AI 模型在此体系中始终作为产生假设、辅助排序的智能化工具，Q-Marker 的最终确立仍须以药效实验、含量测定等湿法验证为根本依据。当然，上述构想还有待在未来进一步验证，同时目前尚无法替代湿法验证，仅能辅助中药 Q-Marker 发现与评价。

4.4 评价体系的标准化流程

为确保 Q-Marker 验证与评价的规范性、可重复性，结合 AI 技术构建标准化流程，明确各环节的技术要求与操作规范，具体流程如下：（1）候选标志物筛选，基于上述 Q-Marker 发现范式初筛候选物，质控要求保证来源样品具代表性且经权威鉴定，纳入阴性对照样品，避免假阳性^[77]；（2）多维度验证，依次完成结构、活性、可测性、传递溯源性及（复方）配伍关联性验证^[5]。每项验证须设定性能确认标准（如结构确证匹配度 $> 95\%$ ），并全程记录仪器型号、参数与标准操作程序（standard

operating procedure, SOP), 确保可重复; (3) AI 量化评价^[78], 调用表 4 中对应模型计算核心指标。质控上, 模型须通过交叉验证并保存超参数, 预测集性能指标需达到预设阈值(如 $R^2 \geq 0.8$)。当前缺乏统一基准时, 可依据该批次数据分布设定内部阈值, 并与经典方法结果对照; (4) 优先级排序, 综合评分排序后, 须进行灵敏度分析, 考察权重变化对排序的影响, 防止极值误判; (5) 标准转化, 依据核心 Q-Marker 含量数据, 参照《中国药典》相关技术规范, 拟订含量限度与检测方法^[79]。可行性方面, 需开展多实验室协同验证, 确保方法可转移; (6) 动态更新, 利用 AI 持续监测上市后多批次样本及临床数据, 设定偏差预警限(如含量波动超过 $\pm 20\%$ 历史均值), 触发再评价。此步骤要求建立长期数据库与数据采集规范。

整个流程通过嵌入质控节点与阈值判据, 力求使 Q-Marker 验证与评价从“理论构想”走向规范化、可操作, 化解不同团队间标准不一的难题。同时, AI 在此体系中的定位始终是高效假设生成与数据处理的辅助工具, 最终确认仍需依赖实验验证。

5 模型规范与质量标准转化条件

AI 辅助筛选的 Q-Marker 要实现从实验室研究向法定质量标准的转化, 需满足数据、模型、验证、合规 4 方面的规范要求, 确保结果科学、可靠、可重复、可监管。在数据规范性方面, 样本需具备充分的代表性, 覆盖不同产地、基原、批次、规格, 充分反映中药的质量波动范围, 避免样本偏倚导致的假阳性结果^[80]。数据采集需遵循统一的标准操作流程, 仪器参数、前处理方法、检测条件需完整记录, 确保数据的可追溯性与可重复性。数据标注需由专业人员完成, 建立标注审核机制, 确保基原鉴定、药效评价、成分定性等标注信息准确无误。同时鼓励研究数据公开共享, 接受同行验证, 推动领域基准数据集的构建, 为模型性能对比提供统一基准。在模型规范性方面, 算法选择应优先选择可解释性强、成熟度高的算法, 使用深度学习等黑箱模型时, 需配套可解释性分析, 如 SHAP^[81]等方法, 说明标志物筛选的科学依据, 解释模型决策逻辑。模型需具备可复现性, 公开代码、超参数设置、训练数据划分方式, 确保其他研究团队可重复验证结果。模型性能评价需采用统一的规范指标, 分类模型采用准确率、精确率、召回率、F1 值等, 回归模型采用决定系数、均方根误差、相对标准偏差等,

同时报告内部交叉验证与外部验证结果。在验证层级要求方面, 需建立 3 级验证体系, 确保模型的泛化能力与实用性^[82]。一级为内部交叉验证, 在同一批次数据集内进行交叉验证, 初步评估模型性能; 二级为外部批次验证, 采用不同批次、不同时间、不同仪器采集的独立数据集进行验证, 评估模型的批次泛化能力; 三级为跨实验室验证, 在多个实验室、多台仪器上进行验证, 评估模型的跨实验室转移能力, 这是纳入标准的必要条件。目前多数研究仅完成了内部交叉验证, 外部验证与跨实验室验证严重不足, 这是制约成果转化的关键瓶颈。在合规性适配方面, 筛选出的 Q-Marker 需符合《中国药典》的整体质控框架, 含量限度设定需结合安全性、有效性与生产实际, 具备可执行性。检测方法需适配常规实验室的仪器条件, 避免依赖特殊、昂贵设备, 确保方法可推广。需充分论证 AI 辅助方法的科学性与可靠性, 符合药品监管部门的技术要求, 逐步推动 AI 方法纳入《中国药典》指导原则。

6 结语与展望

AI 技术凭借强大的多源数据融合、高维特征挖掘与非线性关联建模能力, 已深度渗透到中药 Q-Marker 研究的全流程, 打破了传统筛选方法的局限, 实现了从“经验驱动”向“数据驱动”的转变, 构建了化学指纹+AI 无监督、多组学+AI 关联、文本知识+AI 知识图谱、图像特征+AI 计算机视觉、多模态融合+AI 端到端 5 类研究路径, 在单味药、复方中药、道地药材、濒危药材的 Q-Marker 研究中取得了显著成果^[83]。

但当前研究仍面临多重挑战: (1) 多源数据存在“碎片化、标准化不足”问题, 不同研究团队的检测技术、样本来源、实验条件差异导致化学、生物效应、图像等数据难以整合形成“数据孤岛”, 缺乏统一的数据采集、预处理、标注规范, 稀缺药材与经典名方的样本量不足及多模态数据完整性欠缺, 限制了 AI 模型的泛化能力与筛选精度^[84]。(2) AI 模型可解释性差、适配性不足, 深度学习等模型的“黑箱”特性使其难以阐明筛选 Q-Marker 的内在逻辑, 与中医药“辨证论治、整体调节”的理论体系难以契合, 现有模型多直接迁移自化学、生物学领域, 未能充分融合中药多成分、多靶点、整体调节以及炮制、配伍对成分影响等核心特点, 导致泛化能力有限, 多模型协同筛选策略尚未成熟^[85]。(3) 技术融合深度不足、落地难度大, AI 技术多停留在

数据处理与特征筛选环节,与化学检测、药效实验、质量标准制定等环节尚未形成“数据采集-模型筛选-实验验证-标准落地”的完整闭环。复合型人才匮乏、AI 检测成本较高,多数模型难以应用于生产流通等实际场景。(4) Q-Marker 验证体系不完善、标准不统一, AI 筛选出的候选标志物验证方法单一,缺乏对传递溯源性、配伍关联性的系统考量,尤其复方中药的协同作用复杂,单一标志物难以反映整体质量,不同团队筛选结果差异较大,且缺乏基于 AI 技术的统一评价标准,导致结果难以纳入《中国药典》等行业规范。

针对上述挑战,未来研究应朝着“数据标准化、模型可解释化、技术深度融合、标准体系化”的方向发展:(1) 构建中药多源数据标准化体系,建立统一的数据采集、预处理、标注、共享规范,搭建数据共享平台并补充稀缺样本,采用数据增强技术提升数据利用率;(2) 研发可解释性 AI 模型,结合知识图谱与中医药理论,融入多成分、多靶点等特性,通过集成学习与小样本学习提升泛化能力,并推动模型轻量化;(3) 深化 AI 与化学检测、药效实验、炮制工艺、质量标准制定等环节的融合,形成完整闭环,同时加强复合型人才培养,开发便携低成本检测设备;(4) 完善 Q-Marker 验证体系,加强对传递溯源性、配伍关联性及复方协同作用的验证,建立基于 AI 技术的统一筛选与评价标准,推动成果纳入《中国药典》;(5) 拓展研究场景至道地药材溯源、濒危药材保护等领域,对接国际质量标准,助力中药现代化与国际化发展。

综上所述,通过构建数据标准化体系、研发可解释性模型、深化技术融合、完善验证体系、拓展研究场景, AI 驱动的中药 Q-Marker 研究将向更高质量、更标准化、更规模化的方向发展,实现智能化筛选与标准化应用,为中医药现代化、国际化提供有力支撑。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 李淑贤,朱紫含,邵俊然,等. 芥子化学成分、药理作用研究进展及其质量相关功效物质预测分析 [J]. 中草药, 2026, 57(7): 2768-2786.
- [2] 刘昌孝,陈士林,肖小河,等. 中药质量标志物(Q-Marker): 中药产品质量控制的新概念 [J]. 中草药, 2016, 47(9): 1443-1457.
- [3] 陶益,陈林,朱菲,等. 人工智能视域下拉曼光谱分析技术融合深度学习算法在中药生产过程质量控制中的应用 [J]. 中国现代中药, 2022, 24(1): 176-180.
- [4] 韩星星,朱华旭,唐志书,等. 人工智能推动经典名方研发范式重塑的进展与应用展望 [J]. 中草药, 2026, 57(4): 1209-1220.
- [5] 段成思,刘志勇,史兰云,等. 中药质量标志物多维度评价体系的构建与应用: 化学、生物及整合分析视角 [J]. 中草药, 2025, 56(22): 8395-8402.
- [6] 中华中医药学会《中药质量标志物研究技术指南》编写组. 中药质量标志物研究技术指南 [J]. 中草药, 2025, 56(7): 2249-2252.
- [7] 贺亚丽,梁怡,周琼芝,等. 基于智能感官技术结合化学计量学的蜜化橘红炮制程度快速辨识研究 [J]. 中国现代应用药学, 2025, 42(21): 3681-3690.
- [8] 张瑞,曹庆伟,李科,等. 黄芪药材等级与化学成分和抗疲劳药效的相关性分析 [J]. 药学学报, 2019, 54(6): 1092-1100.
- [9] 杜朋丽,赵丹阳,杜姚,等. 《本草纲目》止吐药物总结及应用规律初探 [J]. 时珍国医国药, 2023, 34(8): 2048-2050.
- [10] 李渊彤,罗裕升,朱珍民. 基于深度学习的舌象特征分析 [J]. 计算机科学, 2020, 47(11): 148-158.
- [11] 汪舒琦,申蕾,宋立亚,等. 深度学习技术在中药质量控制中的应用与探讨 [J]. 中草药, 2025, 56(16): 6048-6057.
- [12] Zhou Y, Zhang H, Sheng X H, *et al.* Evaluation of saffron quality using rapid quantitative inspection technology with near-infrared spectroscopy [J]. *Molecules*, 2024, 29(17): 3983.
- [13] Zhang L, Guan Y, Wang N, *et al.* Identification of growth years for *Puerariae Thomsonii Radix* based on hyperspectral imaging technology and deep learning algorithm [J]. *Sci Rep*, 2023, 13: 14286.
- [14] 解文欣,张紫萱,刘越,等. 人工智能在中药质量中的应用及研究进展 [J]. 中草药, 2025, 56(15): 5616-5631.
- [15] Li Z H, Zhang X H, Liao J, *et al.* An ultra-robust fingerprinting method for quality assessment of traditional Chinese medicine using multiple reaction monitoring mass spectrometry [J]. *J Pharm Anal*, 2021, 11(1): 88-95.
- [16] Long W J, Lei G H, Guan Y T, *et al.* Classification of Chinese traditional cereal vinegars and antioxidant property predication by fluorescence spectroscopy [J]. *Food Chem*, 2023, 424: 136406.
- [17] Wang Y N. NLP-driven integration of electrophysiology and traditional Chinese medicine for enhanced diagnostics and management of postpartum pain [J]. *SLAS Technol*, 2025, 32: 100267.
- [18] 周丽媛,高红梅,赵启军,等. 基于多特征融合结合深度学习模型的药材切片鉴别 [J]. 世界科学技术—中医药现代化, 2024, 26(1): 211-217.
- [19] 杨浣菲,李丝雨,俞辰倩,等. 全流程数据要素融合促进中药调剂创新与应用 [J]. 中国中药杂志, 2025, 50(11): 3189-3196.
- [20] 顾志荣,毛小文,祁梅,等. 数据融合技术在中药研究

- 领域的应用 [J]. 中草药, 2023, 54(18): 6159-6171.
- [21] 张雨晴, 何民友, 粟春枝, 等. 基于指纹图谱和多指标成分含量测定结合化学计量学评价不同产地决明子质量 [J]. 中草药, 2026, 57(15): 6159-6171.
- [22] 桂新景, 李涵, 王艳丽, 等. 基于人工智能嗅觉技术和化学计量学的白及饮片真伪鉴别 [J]. 医药导报, 2024, 43(3): 441-451.
- [23] 孙仁爽, 马承俊, 李东飞, 等. 不同产地南柴胡多指标定量及化学计量学联合 CRITIC-TOPSIS 法的质量差异评价 [J]. 中草药, 2026, 57(17): 6965-6972.
- [24] 冯嘉楠, 商周怡, 李萌, 等. 机器学习在药物质量分析中的应用进展 [J]. 厦门大学学报: 自然科学版, 2026, 65(1): 68-80.
- [25] 石佳, 康帅, 王子骏, 等. 面向小样本的细粒度不同规格黑豆及其混淆品、伪品的人工智能检测方法研究 [J]. 中国新药杂志, 2026, 35(4): 387-395.
- [26] 赵书言, 王书睿, 申镇华, 等. 基于人工智能算法的中药主归经判别研究 [J]. 中草药, 2024, 55(16): 5573-5582.
- [27] 夏伯候, 胡玉珍, 熊苏慧, 等. 随机森林算法在中药指纹图谱中的应用: 以不同品牌夏桑菊颗粒指纹图谱分析为例 [J]. 中国中药杂志, 2017, 42(7): 1324-1330.
- [28] 石岩, 李宁, 魏锋, 等. 机器学习相关技术在以黄酮为特征的黄芪分类中的应用研究 [J]. 药物分析杂志, 2024, 44(5): 866-873.
- [29] 富尧, 郑长辉, 何婷, 等. 基于指纹图谱与多成分动态分析的炒柴胡炮制成分变化规律及质量标志物构建 [J]. 中国中药杂志, 2026, 51(9): 2525-2535.
- [30] 吴强, 赵鹏, 王猛, 等. 基于有效成分含量与高光谱的蒙古黄芪品质分级 [J]. 光谱学与光谱分析, 2025, 45(11): 3066-3071.
- [31] 沈晔, 王继森, 向晴, 等. 基于机器学习的中药产地溯源研究 [J]. 中草药, 2026, 57(5): 1798-1813.
- [32] 冯佳豪, 关志琪, 张智勇, 等. 基于近红外光谱与化学计量学的淫羊藿与柔毛淫羊藿基原鉴别及模型转移研究 [J]. 分析测试学报, 2025, 44(6): 1034-1042.
- [33] 刘宇强, 石胜群, 张猛胜, 等. LIBS-NIR 光谱信息融合的黄芪产地鉴别 [J]. 红外与激光工程, 2025, 54(6): 20250003.
- [34] 王海霞, 所同川, 余河水, 等. 基于近红外光谱技术的甘草提取过程最优建模方法研究 [J]. 中国中药杂志, 2016, 41(19): 3537-3542.
- [35] 马盼盼, 冯佳鑫, 张洪宝, 等. 基于谱效关系联合效应成分指数评价不同产地的苍术质量 [J]. 中药新药与临床药理, 2026, 37(3): 543-551.
- [36] 张传耀, 周晴, 王校, 等. 基于模式识别的唇形科中药寒热药性物质基础研究 [J]. 药学报, 2023, 58(2): 429-438.
- [37] 何家群, 李奕, 韩业慧, 等. 基于 QuEChERS-UPLC-MS/MS 定量与机器学习结合的酱香型高温大曲的质量评价方法研究 [J/OL]. 食品与发酵工业 [2026-10-14]. <https://doi.org/10.13995/j.cnki.11-1802/ts.045913>.
- [38] 位云朋, 胡会强, 毛晓波, 等. 基于高光谱成像技术与异构集成学习的龟甲药材生长年限鉴别 [J]. 光谱学与光谱分析, 2024, 44(9): 2613-2619.
- [39] Chen Z, Cao Y F, He S B, et al. Development of models for classification of action between heat-clearing herbs and blood-activating stasis-resolving herbs based on theory of traditional Chinese medicine [J]. *Chin Med*, 2018, 13(1): 12.
- [40] 张云飞, 张玖捌, 陈鹏, 等. 基于性状电子检测技术结合机器学习的不同产地薏苡仁的快速鉴别 [J]. 中草药, 2024, 55(12): 4185-4193.
- [41] 向科霖, 章晓云, 黎征鹏, 等. 胞葬相关基因 UCP2、EGLN3、IL1B 对骨关节炎的诊断价值及中药治疗预测: 基于生物信息学及机器学习 [J]. 南方医科大学学报, 2026, 46(3): 693-706.
- [42] 廖昊森, 陈翠兰, 马月辉, 等. 基于机器学习与实验验证的缺血性卒中核心机制解析及中药防治策略预测 [J]. 中草药, 2026, 57(7): 2640-2654.
- [43] Nie W N, Guo S B, Li C H, et al. A machine learning-based assessment strategy for *Artemisia Argyi Folium*: Exploring the dose-effect relationship between anti-inflammatory activity and chemical composition [J]. *J Ethnopharmacol*, 2026, 357: 120952.
- [44] 吴鑫雨, 邱丽媛, 王又迪, 等. 基于智能感官与多源信息融合技术的香附炮制程度快速辨识方法研究 [J]. 中草药, 2023, 54(21): 7007-7016.
- [45] 黄婧. 基于合成图像数据集和深度学习的中药饮片识别方法 [J]. 中国现代应用药学, 2025, 42(17): 2990-2996.
- [46] 刘英慧, 丁盈, 李林岚, 等. 基于一维卷积神经网络-长短期记忆网络(LDCNN-LSTM)协同模型驱动的中红外-拉曼融合光谱鉴别石菖蒲/水菖蒲 [J]. 中国无机分析化学, 2026, 16(2): 281-291.
- [47] Gao L L, Zhong L, Feng T T, et al. An AI-driven strategy for active compounds discovery and non-destructive quality control in traditional Chinese medicine: A case of Xuefu Zhuyu Oral Liquid [J]. *Talanta*, 2025, 287: 127627.
- [48] 陶益, 陈林, 江恩赐, 等. 人工智能和工业 4.0 视域下高光谱成像技术融合深度学习方法在中药领域中的应用与展望 [J]. 中国中药杂志, 2020, 45(22): 5438-5442.
- [49] Zhang S, Zhang X W, Du J Y, et al. Multi-target meridians classification based on the topological structure of anti-cancer phytochemicals using deep learning [J]. *J Ethnopharmacol*, 2024, 319: 117244.
- [50] Zhu Z X, Liu B, Xiao Y W, et al. Machine learning-aided drug repurposing for screening COX-2 inhibitors from traditional Chinese medicines [J]. *Pharmaceuticals*, 2026, 19(6): 878.
- [51] 吴继忠, 时艺丹, 黄慧, 等. 基于改进堆叠自编码器结合 LightGBM 的近红外光谱回归算法研究 [J]. 分析测试学报, 2023, 42(9): 1112-1118.
- [52] 郭兆华, 文师召, 李亚薇, 等. 基于近红外光谱技术的冀产黄芩鉴别方法研究 [J]. 中国药学杂志, 2024, 59(20): 1939-1947.

- [53] Wan M C, Cai R, Zhang X S, *et al.* Integrated multi-omics and deep learning analysis reveals neurotransmitter metabolism regulatory mechanisms of Tianwang Buxin Dan [J]. *Front Neurosci*, 2026, 20: 1837233.
- [54] Yang H X, Chen M Y, Wang J Y, *et al.* Rapid identification of trace pharmacodynamic substances in traditional Chinese medicine via SERS and deep learning [J]. *Biosensors*, 2026, 16(3): 139.
- [55] Yuan B, Zhang C, Ji C Q, *et al.* HSQCid: A powerful tool for paving the way to high-throughput structural dereplication of natural products based on fast NMR experiments [J]. *Anal Chem*, 2025, 97(6): 3227-3235.
- [56] 吴泓嘉, 张弛, 张宏鑫, 等. 药症方关联的中医药古籍交互可视分析方法 [J]. 计算机辅助设计与图形学学报, 2025, 37(8): 1439-1452.
- [57] 郭兆华, 文师召, 李思凡, 等. 基于近红外光谱结合数据增强 CNN 算法的白芷产地溯源方法 [J]. 中国药学杂志, 2024, 59(21): 2022-2029.
- [58] 谢文涌, 柴琴琴, 林旒, 等. 基于 Stacking 集成学习的马兜铃酸及其类似物鉴别 [J]. 江苏农业学报, 2021, 37(2): 503-508.
- [59] 谢文涌, 柴琴琴, 甘勇辉, 等. 基于多特征提取和 Stacking 集成学习的金钱莲品系分类 [J]. 农业工程学报, 2020, 36(14): 203-210.
- [60] 周正, 吴志生, 史新元, 等. Bagging-PLS 的黄柏中试提取过程在线近红外质量监测研究 [J]. 世界中医药, 2015, 10(12): 1939-1942.
- [61] 金承亮, 王永军, 黄河, 等. 高维红外光谱数据预处理在中药材产地鉴别中的应用 [J]. 光谱学与光谱分析, 2023, 43(7): 2238-2245.
- [62] 马骞, 杨婉琪, 李福生, 等. 基于迁移学习与 TrAdaBoost-SVM 方法的 XRF 中药重金属超标研究 [J]. 光谱学与光谱分析, 2023, 43(9): 2729-2733.
- [63] 王一丁, 郝晨宇, 李耀利, 等. 基于深度学习的小样本中药材粉末显微图像识别 [J]. 计算机应用, 2020, 40(5): 1301-1308.
- [64] 于旭华, 闫蓉, 苏如玉, 等. 中药临床试验样本量合理优化的探讨 [J]. 中药新药与临床药理, 2013, 24(3): 326-328.
- [65] 黄冬兰, 徐永群, 陈小康, 等. 基于红外三级鉴定与聚类分析法的党参快速鉴别研究 [J]. 光谱学与光谱分析, 2017, 37(10): 1849.
- [66] Gao S, Chau H Y K, Wang K J, *et al.* Convolutional neural network-based compound fingerprint prediction for metabolite annotation [J]. *Metabolites*, 2022, 12(7): 605.
- [67] Xiao W K, Zhang M Q, Zhao D N, *et al.* TCMKD: From ancient wisdom to modern insights-a comprehensive platform for traditional Chinese medicine knowledge discovery [J]. *J Pharm Anal*, 2025, 15(6): 101297.
- [68] 徐雅静, 俞捷, 余远盼, 等. 人工智能在中药材及饮片鉴别领域的应用 [J]. 中华中医药学刊, 2022, 40(8): 47-50.
- [69] 田春芳, 胡倩南, 詹志来, 等. 基于传统品质评价的野生与栽培赤芍比较 [J]. 中国实验方剂学杂志, 2024, 30(14): 165-174.
- [70] 孙佳琪, 殷勇, 于慧春, 等. 紫苏中紫苏醛含量的高光谱预测方法 [J]. 光谱学与光谱分析, 2024, 44(9): 2667-2674.
- [71] Liang Z Z, Ding C S. TCMHTI: A transformer-based herb-target interaction prediction model for Qingfu Juanbi Decoction in rheumatoid arthritis [J]. *Digital Chin Med*, 2025, 8(2): 206-218.
- [72] 白钢, 侯媛媛, 丁国钰, 等. 基于中药质量标志物构建中药材品质的近红外智能评价体系 [J]. 药学报, 2019, 54(2): 197-203.
- [73] 吕斌, 黄现青, 吕文浩. 山茱萸产地 LIBS 结合多模型融合判别方法的建立 [J]. 郑州大学学报: 医学版, 2025, 60(5): 618-622.
- [74] 刘子稀, 孟庆民, 饶小勇, 等. 基于 QbD 理念的山楂配方颗粒制备工艺关键属性的辨析和传递关系研究 [J]. 中草药, 2026, 57(9): 3343-3352.
- [75] 邵方娴, 黄朝瑜, 严菲, 等. AQbD 理念下的《中华人民共和国药典》与 ICH Q2 分析方法验证指导原则解析与实例探讨 [J]. 中国新药杂志, 2026, 35(1): 6-10.
- [76] 刘雯丽, 李灵犀, 聂黎行, 等. 基于 UHPLC-Q Exactive Orbitrap MS/MS 技术分析不同叶色紫苏叶化学成分及其与色数值的相关性 [J]. 药物分析杂志, 2026, 46(1): 173-186.
- [77] 李思雨, 裴志东, 王嘉唯, 等. 基于谱效关系的通关藤质量标志物筛选 [J]. 中成药, 2025, 47(7): 2327-2332.
- [78] 阳长明, 杨平, 刘乐环, 等. 中药质量标志物 (Q-Marker) 研究进展及对中药质量研究的思考 [J]. 中草药, 2021, 52(9): 2519-2526.
- [79] 郭玥, 康显杰, 叶佩, 等. 基于中药炮制特色的饮片质量标准研究现状与问题探析 [J]. 中华中医药杂志, 2024, 39(11): 6163-6167.
- [80] 于姗姗, 刘洪超, 郭东晓, 等. 从质量标准复核和审评视角浅析中药配方颗粒标准制定工作 (I) [J]. 中草药, 2022, 53(14): 4243-4248.
- [81] 张恩景, 韦琼, 江娟, 等. 基于 XGBoost 模型的围孕期用药风险预测与个性化药学服务策略 [J]. 医药导报, 2026, 45(3): 524-529.
- [82] 江振作, 王跃飞. 基于“药材基原-物质基础-质量标志物-质控方法”层级递进的中药质量标准模式研究 [J]. 中草药, 2016, 47(23): 4127-4133.
- [83] 李珊珊, 李玮婕, 王萍. 整合药理学在中医药现代研究的若干进展 [J]. 中国实验方剂学杂志, 2024, 30(14): 270-278.
- [84] 鲁路, 倪世豪, 陈星伶, 等. 人工智能赋能下基于人用经验的中药新药研发展望 [J]. 中国中药杂志, 2025, 50(20): 5605-5612.
- [85] 江启煜, 曾慧妍. 基于人工智能和组学数据驱动的中药潜在机制新型分析预测方法 [J]. 中国组织工程研究, 2025, 29(35): 7552-7561.

[责任编辑 潘明佳]