

• 专题报道 •

多源信息融合技术在中药质量评价中的应用研究进展

李 涵¹, 谢梦迪¹, 桂新景^{1,2}, 侯富国¹, 范雪花¹, 李海洋¹, 王艳丽²,
李学林^{2,3,4,5*}, 刘瑞新^{2,3,4,5*}

(1. 河南中医药大学, 河南 郑州 450046; 2. 河南中医药大学第一附属医院, 河南 郑州 450000; 3. 河南省中药临床应用、评价与转化工程研究中心, 河南 郑州 450000; 4. 河南中医药大学呼吸疾病中医药防治省部共建协同创新中心, 河南 郑州 450046; 5. 河南省中药临床药学中医药重点实验室, 河南 郑州 450000)

摘要: 中药质量评价是关乎中药现代化发展进程的关键问题之一, 中药质量评价技术体系主要包括传统评价(性状、显微和理化)、化学评价和生物评价。由于中药组成复杂, 上述评价技术体系中的单一检测手段通常无法获得足够的信息, 信息融合策略可有机整合来自多个分析检测技术的数据, 获取样本更全面的信息, 提高质量评价效果。目前, 信息融合策略已广泛应用于军事、工业和食品等领域, 在中药质量评价领域尚处于起步阶段。本文介绍了多源信息融合的定义、结构、方法及融合层次, 综述了其在中药产地溯源、品种鉴别和制药分析中的研究进展, 并整理了各篇文献中数据融合的具体方法, 最后对信息融合技术及其在中药质量评价方面的应用、发展及存在问题进行总结、展望和讨论, 以期拓宽该技术在中医药领域的应用范围提供参考。

关键词: 中药; 质量评价; 信息融合

中图分类号: R286.0

文献标识码: A

文章编号: 0513-4870(2023)10-2835-18

Research progress of multi-source information fusion technology in quality evaluation of traditional Chinese medicine

LI Han¹, XIE Meng-di¹, GUI Xin-jing^{1,2}, HOU Fu-guo¹, FAN Xue-hua¹, LI Hai-yang¹,
WANG Yan-li², LI Xue-lin^{2,3,4,5*}, LIU Rui-xin^{2,3,4,5*}

(1. Henan University of Chinese Medicine, Zhengzhou 450046, China; 2. The First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou 450000, China; 3. Henan Province Engineering Research Center for Clinical Application, Evaluation and Transformation of Traditional Chinese Medicine, Zhengzhou 450000, China; 4. Co-construction Collaborative Innovation Center for Chinese Medicine and Respiratory Diseases by Henan & Education Ministry of China, Henan University of Chinese Medicine, Zhengzhou 450000, China; 5. Henan Provincial Key Laboratory for Clinical Pharmacy of Traditional Chinese Medicine, Zhengzhou 450000, China)

Abstract: The quality evaluation of traditional Chinese medicine is one of the key issues related to the modernization of traditional Chinese medicine. The quality evaluation technology system of traditional Chinese medicine mainly includes traditional evaluation (traits, microscopic and physicochemical identification), chemical evaluation and biological evaluation. Due to the complex composition of traditional Chinese medicine, the single detection method in the above evaluation technology system usually cannot obtain sufficient quality information. The multi-source information fusion strategy can organically integrate data from multiple analysis and detection

收稿日期: 2023-02-20; 修回日期: 2023-07-05.

基金项目: 国家自然科学基金资助项目(81773892, 81774452); 河南省中医药拔尖人才培养项目(重点项目)(2019ZYBJ07); 河南省卫生健康中青年学科带头人专项(HNSWJW-2020014); 国家中医药管理局2022年青年岐黄学者培养项目(国中医药人教函[2022]256号).

*通讯作者 Tel: 86-371-66233562, E-mail: liuruixin7@163.com;

Tel: 86-371-66245142, E-mail: xuelinli450000@163.com

DOI: 10.16438/j.0513-4870.2023-0195

technologies to obtain more comprehensive information of samples and improve the quality evaluation effect. At present, multi-source information fusion strategy has been widely used in the fields of military, industrial and food, and it is still in its infancy in the field of quality evaluation of traditional Chinese medicine. This research introduces the definition, structure, method (algorithm) and fusion level of multi-source information fusion, summarizes its research progress in the origin traceability, variety identification and pharmaceutical analysis of traditional Chinese medicine, and sorts out the specific methods of data fusion in each literature. Finally, we summarized, prospected and discussed the application, development and existing problems of information fusion technology and its application in the quality evaluation of traditional Chinese medicine, in order to provide reference for broadening the application of this technology in the field of traditional Chinese medicine.

Key words: traditional Chinese medicine; quality evaluation; multi-source information fusion

自国家药品监督管理局2020年提出促进中医药传承创新发展实施意见和2021年提出“十四五”国家药品安全及促进高质量发展规划以来,在日臻完善的现代分析检测技术和监管体系的推动下,中药质量迈上了一个新台阶^[1]。中药(中药材、中药饮片、中药复方制剂)作为中医药产业链的物质基础,其质量是保证临床有效性、安全性的基础,是促进中药国际化、产业化和中医药现代化的关键^[2]。

中药的质量是中药化学物质的综合生物效应的整体表现,其内涵特点是多成分、多功效和整体性^[3]。影响中药质量的因素有多种:药材品种、栽培种植和产地生态条件、采收加工、炮制工艺、运输储藏、提取纯化、制剂工艺、组方以及药物代谢等。传统中药质量评价技术体系围绕中药“形、色、气、味”等性状表征来判断中药真伪及品质优劣,由此而提出的“辨状论质”思想^[4,5]在一定时期的中药质量评价中发挥着作用,电子鼻、电子眼、电子舌等电子感官技术的发展使该体系能够实现药材或饮片部分性状的客观数据表达^[6]。现代中药质量评价体系以“找成分,测含量”为主要研究模式,包括化学评价和生物评价。高效液相色谱、气相色谱、液相色谱-质谱联用、核磁共振、近/中红外光谱、紫外光谱等色谱光谱技术为主的化学评价^[7-10]以及综合中药成分和生源途径的质量标志物(Q-marker)^[11,12]可以宏观地反映中药中复杂的物质成分,是中药质量的内在表征。基于生物体系(整体动物、离体组织、器官、微生物和细胞及相关生物因子等)和药效实验的生物评价^[13]能够总体评价和表征中药的安全性和有效性,是对中药化学评价模式的补充。

信息融合技术是指对来自不同时空、多个来源的同质或异构信息进行合并和集成,从而促进对事件或有关实体的有效决策。由于中药基原的复杂性和成分、药效的多样性,单一的分析技术测得的数据难以实现对样品质量的准确评价,多指标、多手段的质量评价模式更符合中药多成分、多功效的特点,信息融合技术

的提出较好地适应了这一评价模式。正如中医药“整体观”、“系统论”思想,基于不同检测原理的多种仪器的数据能够提供样本的多维信息,信息融合策略可以将这些数据有机地整合在一起,获得更为全面、互补的样品质量信息,从而实现更有效、更精确的识别及预测结果。信息融合技术最早由美国开始研究,目前已广泛应用于军事领域、工业领域、食品检测、环境监测、航空航天等多个领域^[14-17],但在中药学科中的研究还仍处于起步阶段。本文综述国内外多篇文献,围绕信息融合技术在不同类型的中药质量评价中的应用,梳理各文献中的融合数据源、融合层次、融合算法等具体数据融合过程,从而总结信息融合技术应用于中药质量评价时的研究思路,为拓宽该技术在中医药领域的应用范围提供参考。

1 多源信息融合概述

1.1 定义

多源信息融合是指把从不同空间或时间上获得的多个来源的同质或异质信息进行不同抽象层次的集成,从而得到有关实体、关系或事件的更完整、更精确、更可靠的信息或推论,促进有效决策,其本质是对大脑综合分析信息能力的模拟^[18]。多源信息融合起源于军事领域中的多传感器联合使用,因此又称为多传感器数据融合(或数据融合),随着现代科学技术的发展,“传感器”泛指各种信息来源,除了电子传感器外,还包括数据库、网络系统等。具体在中药质量评价领域,“传感器”指的是能够获得样本质量数据的分析检测技术,如近红外、中红外、紫外-可见光、拉曼光谱等光谱技术;液相、气相、液质联用等色谱质谱技术;电子鼻、色差仪、电子眼等智能感官技术等。

1.2 信息融合结构

信息融合的结构(图1)可以分为集中式、分布式和混合式3大类^[19],集中式融合中,所有传感器的测量信息不经任何预处理,直接输送至融合中心,在融合中心处统一运算,处理后采用融合算法对目标进行预测

和评估;分布式融合结构中,来自每个传感器的数据在输送至融合中心前,都先由其各自的数据处理器进行预处理、跟踪和分类,这样做的优点是信息量较原始数据大幅减少,融合中心的计算量也会减少,此时既能部署复杂的算法又不至于损失太多的时间;混合式融合结构就是集中式和分布式并存的结构,可以依据各传感器的信息,灵活调整每个传感器的融合途径。在中药质量评价中,信息融合的结构多采用混合式,因为不同分析检测仪器所获得的信息是不同质、不同量纲的,所以其预处理方法也不同,需采用分布式处理;但某些分析仪器如电子感官的传感器,其数据获取途径以及模式识别、数据预处理、数据形态方面具有一定的相似性,因此也可以选择集中处理。

1.3 信息融合方法(算法)

信息融合的方法(即融合算法)有多种,由于其所面临的问题多样且复杂,目前仍没有完全统一的算法能够适应所有场景,需要针对不同的问题场景选择合适的算法。文献中常见的方法分类有:①基于物理模型、参数模型和认知识别模型的方法分类^[20];②基于状态模型、统计理论、知识规则和信息理论的方法分

类;③基于信号处理与估计理论、统计推断理论、信息论、决策论、人工智能和几何分析的方法分类;④基于概率统计的传统方法和基于人工智能的现代方法分类;⑤面向数据/信息不同特性的方法分类;⑥面向不同融合功能级别的方法分类。⑤和⑥分类方法能够为信息融合中数据的不确定、不精确、不完整、多义、模糊的特性以及融合系统的设计、传感器资源的管理和支撑平台的建立提供参与指导,具有一定的实用性,较前4种分类方法更加通俗易懂。但不论分类方法如何,常用的信息融合算法主要有加权平均法、Bayes方法、Kalman滤波法、D-S证据合成方法、模糊推理、神经网络和粗糙集方法。具体在中药质量评价领域,常用算法有:偏最小二乘-判别分析(PLS-DA)、支持向量机(SVM)、随机森林(RF)、主成分分析(PCA)、偏最小二乘回归(PLSR)等。

1.4 信息融合层次

在混合式信息融合结构中,融合层次从低到高依次为数据级、特征级和决策级(图2)。针对实际问题与信息源数据特征的差异性,可单独使用不同层次的融合方法或组合某两个层次的递进融合方法,从而得到

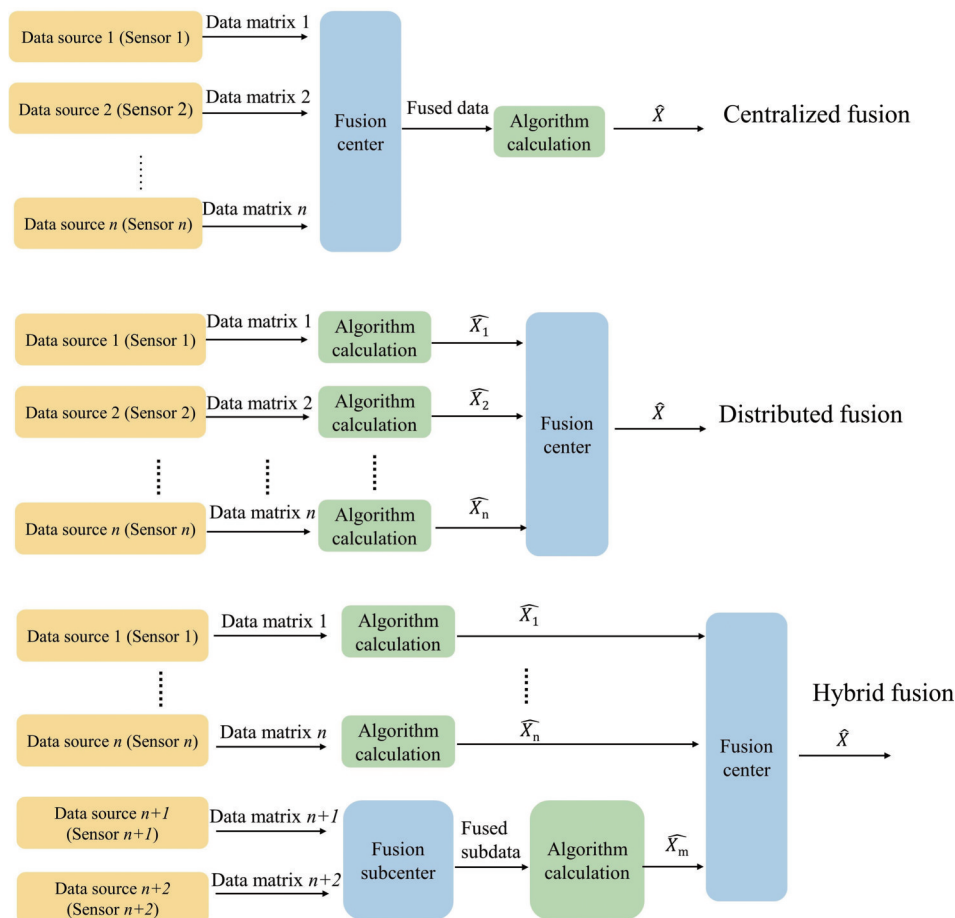


Figure 1 Structure of centralized, distributed and hybrid fusion

使系统性能较优的融合方案^[21]。

在数据级融合(低级融合)中,输入层是多个传感器的原始串联信息,输出层是原始数据经特征识别处理后的信息或局部决策的结果,其优点是不错失细微的样本信息,算法处理简单,缺点是样本信息冗杂,与建模无关的样本会影响模型性能。

在特征级融合(中级融合)中,首先对各传感器原始信息提取特征矢量,然后基于融合的特征矢量对样本识别和评估,由于原始数据被压缩和特征提取,因此部分信息损失,模型准确性降低,但冗余信息的去除能使算法运行效率提高。在使用多源仪器对中药质量评价中,特征变量一般由重要变量或原始数据的变量转化组成,如VIP融合、主成分融合或潜变量融合等。

决策级融合又称高级融合,首先分别提取各传感器原始数据的特征信息,生成特征矢量来识别和预测目标,然后融合各传感器的识别结果,得到被测目标的一致性解释说明。决策级融合具有很好的容错性和时效性,在一个或几个传感器失效时也能正常工作,但预处理成本高。多项研究表明,与未进行数据融合相比,数据经融合后的识别和预测结果会更好^[22-24]。

2 多源信息融合在中药质量评价中的应用

2.1 地理来源鉴定

中药材具有一定的农产品特性,不同产地药材的特殊品质是其基因型、特定生态环境和栽培措施共同作用的结果,其共同塑造了某一药材区别于其他产地同类药材在外观性状、化学组成、药理作用及临床疗效方面特有的品质特征^[25]。因此,产地鉴别对于保证中药材质量可控及用药安全稳定意义重大。以下根据所使用的不同质量评价技术对其分类叙述。

2.1.1 光谱、色谱数据融合 红外光谱可以提供化合物分子官能团结构,从整体上反映植物代谢产物化学特征,液相色谱能得到特定指标成分或成分群化学特征^[26],光谱色谱数据融合可以高效、全面、准确地鉴别

药材地理来源。研究^[27]采用液相色谱(LC₂₄₂和LC₂₁₀)与傅里叶变换红外光谱(FTIR)表征云南省8个产地的茯苓和茯苓皮特征,比较了两种数据融合水平(低级和中级)和两种数据组合方式(色谱、光谱数据组合和两个药用部位数据组合)所构建的茯苓产地鉴别模型性能(图3),结果表明,不同产地的茯苓和茯苓皮中三萜类成分含量不尽相同,FTIR和LC₂₄₂低级融合数据所构建的模型产地鉴别性能最好,正确率为100%。茯苓地理溯源的另一文献^[28]从野生和栽培两个角度出发,融合FTIR和HPLC特征变量数据,相比于单一技术,融合数据提高了对野生样品的溯源准确率。同样是茯苓地理溯源,文献^[29]着重考察了茯苓药材中受地理环境影响较大的矿质元素,采用电感耦合等离子体原子发射分光光度计(ICP-AES)和FTIR对来自12个地理区域茯苓中的9种矿物元素和光谱进行了分析,融合两类数据并建模,多元素分析表明不同采样区域茯苓药材元素含量差异显著,高海拔地区更有利于元素积累,中级融合结果优于单一的FTIR数据。研究^[26]采集了云南、四川、广西、贵州4个产地的滇龙胆药材的光谱并测定其环烯醚萜苷活性成分含量,经光谱色谱低、中级数据融合后模型预测集正确率为96.43%,高于单一技术;滇龙胆产地鉴别的另一文献^[30]分析了该药材在不同纬度地区的化学物质地理变异情况,对龙胆药材根、茎、叶色谱数据分别进行低、中级融合(图4),结果表明根、茎数据低级融合后的OPLS-DA模型分类效果最好,正确率为98.94%。

在产地鉴别中,由于样本质量信息多样,模型对样本数据转换时不只是线性的,也有非线性识别的可能,因此,使用的算法有PLS-DA、LDA等线性判别模型,也有LS-SVM、BP-NN等非线性模型。研究^[31]以芍药(白芍与赤芍)为载体,比较了HPLC和FTIR融合数据与单一技术的鉴别能力,以及线性分类模型(LDA)和非线性分类模型(RBF-ANN和LS-SVM)的识别能力,

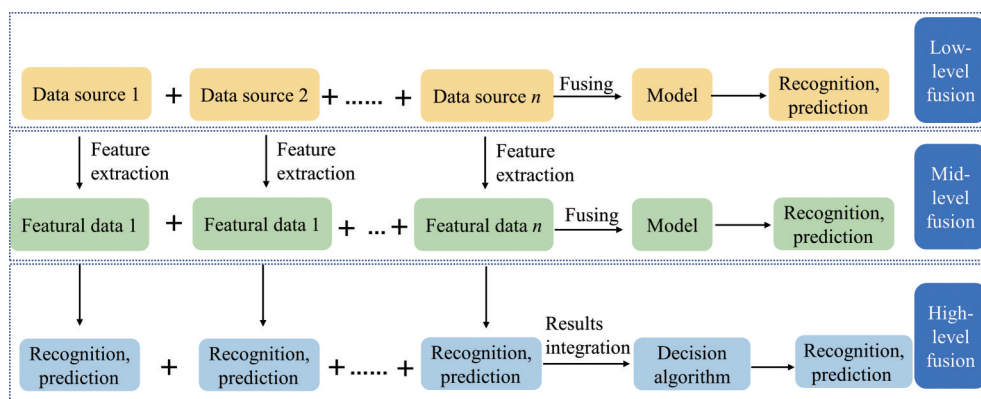


Figure 2 Level of information fusion

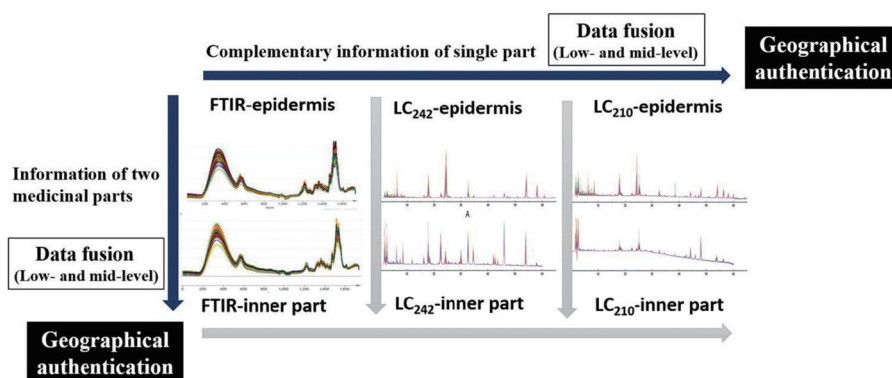


Figure 3 The workflow of geographical authentication of the epidermis and inner part of *Macrohyporia cocos* using data fusion^[27]

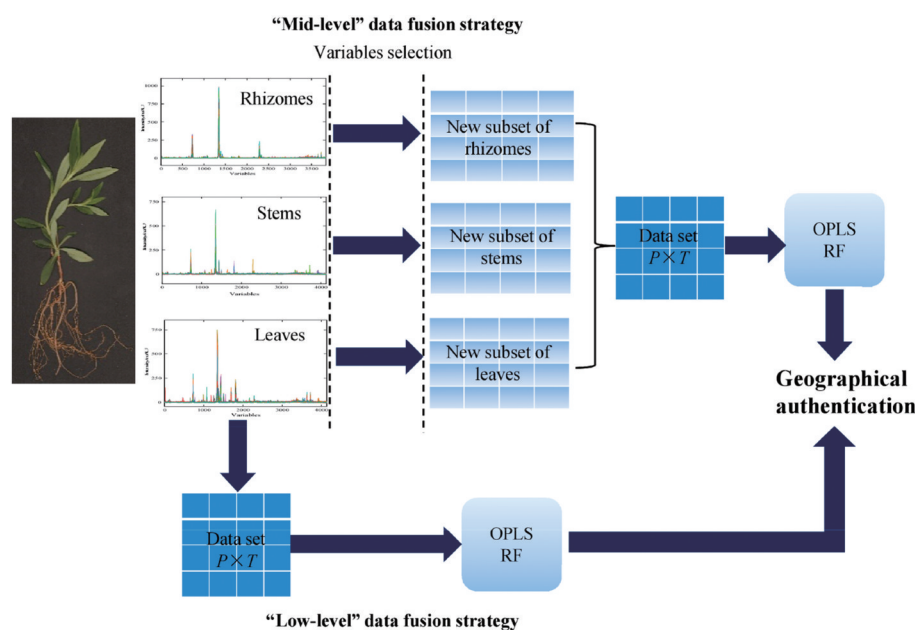


Figure 4 The workflow of geographical authentication of *Gentiana Rigescens* with data fusion strategy^[30]

结果表明, 经遗传算法-偏最小二乘 (GA-PLS) 提取最优变量或经 PCA 提取主成分后的融合数据分类性能优于单一技术, 且以非线性分类模型效果最好, 可准确识别芍药的产地。

2.1.2 光谱数据融合 近红外光谱主要反映含 H 化学结构 (C-H、N-H、O-H) 的合频峰和泛频峰, 包含大多数有机化合物的组成和分子结构信息, 中红外光谱 (MIR) 反映了分子振动的基频信息, 敏感性及特征吸收峰数目高于近红外, 因此近-中红外光谱融合后可以获取更多样品信息, 提高模型预测性能。研究^[32]分别从低、中、高层次来融合 NIR 和 FT-MIR 光谱, 建立了云南 5 个地区三七的稳健地理追溯模型。在低级融合中, 预处理后的光谱经归一化后被拼接成新的数据矩阵; 中级融合使用递归特征消除 (recursive feature elimination, RFE) 函数选择关键变量或主成分分析提取光谱特征; 高级融合中, 首先用特征变量分别建立单

一模型, 然后融合单一模型结果, 根据模糊集合论的最大值、最小值、乘积和平均值进行投票得到最终结果: 近红外和中红外光谱数据经高级融合后的模型分类性能最优, 原因可能是高级融合避免了原始光谱中的无用和干扰信息。在文献^[33]中, 三七地理溯源也通过 FT-MIR 和 NIR 的低、中、高融合数据的 RF 模型来进行, 不同的是, 中级融合是通过变量选择 (RF-Vs) 和 RF-Boruta 来筛选变量, 分类结果也以高级融合最优, 且 RF-Boruta 的重要变量更精简有效, 数据分析更快。

文献^[34]基于衰减全反射-傅里叶变换中红外光谱 (ATR-FTMIR) 和 NIR 光谱, 采用 PCA、PLS-DA、RF 等化学计量学模型结合低、中、高级融合, 对云南 5 个地区的野生滇重楼进行研究, 结果表明 Boruta 筛选变量方法比递归特征消除 (RFE) 筛选变量方法有更高正确率, 高级融合 RF 模型是野生滇重楼的最优产地鉴别模型。文献^[35]采集了不同产地黑果枸杞的近、中红外光

谱并测定其多糖含量,通过PCA降维分析融合光谱数据建立LS-SVM分类模型,通过Si-PLS模型筛选波段建立多糖含量预测PLS模型,结果表明,近中红外光谱融合后分类模型和预测模型性能均提高,能够对黑果枸杞进行产地及品质检测。

红外光谱是分子振动光谱,使用最多的是中红外和近红外,但近红外光谱信息存在很大程度上的重叠,无法精准反映某些特殊官能团上的信息;紫外-可见光(UV-Vis)光谱可反映生色团和助色团的光谱特征;拉曼光谱通过散射方式研究晶格以及分子振转模式信息,可分析出化学物质内部结构,因此多种光谱互补融合可弥补单一光谱反映信息模糊的弊端,并得到更全面的化学信息,从而实现更精准的产地鉴别。研究^[36]采集了9个产地的滇黄精的红外光谱(ATR-FTIR)和紫外-可见光光谱(UV-Vis),结合RF算法建立单一光谱以及低、中、高级融合数据的产地辨识模型,结果表明,两种光谱数据都能反映出药材所含成分,提取潜变量(LVs)后的中级融合分类模型正确率为97.50%,性能最优。文献^[37]融合了云南6个地区的野生重楼的FT-MIR和UV-Vis数据,通过PCA提取光谱特征,建立低、中级融合的SVM-GS模型和高级融合的RF模型,高层次融合模型中训练集和测试集的准确率分别在99%和98%左右,显著高于单一光谱数据和低、中层次融合数据,表明基于高级融合的RF算法具有高效、良好的地理溯源性能。文献^[38]在融合野生重楼的FT-MIR和UV-Vis数据时,在无监督的PCA算法提取PCs特征的基础上,对比了有监督算法PLS-DA提取的LVs特征,然后进行数据融合,结果基于LVs选择的特征变量分类效果较单一光谱和低级融合好。另有文献分别对杜仲叶^[39]、黄芪饮片^[40]的不同光谱数据进行融合,从而进行产地溯源,均取得较好结果。

2.1.3 光谱、色谱、质谱数据融合 稳定同位素分析技术已广泛用于地理溯源,根据样本自身天然存在的碳(¹²C、¹³C)、氢(¹H、²H)、氧(¹⁶O、¹⁸O)和氮(¹⁴N、¹⁵N)等稳定同位素的特征,可以估算多种特定的环境参数,从而鉴别其产地。文献^[41]将UPLC-QTOF-MS、电感耦合等离子体质谱(ICP-MS)、稳定同位素比例质谱(IRMS)及FTIR、NIR、拉曼光谱(Raman)联合使用,对新西兰4个地区的蜂蜜进行花源的产地鉴别,融合数据的模型预测能力与单一UPLC-QTOF-MS数据相同,表明数据融合并未提高鉴别性能。同样是蜂蜜花源溯源,文献^[42]结合¹H-NMR、LC-Orbitrap-MS和LC-TOF-MS三个技术平台的优势,分析了来自4个单花来源和2个不同地区多花来源的蜂蜜,并比较了选择变量和不选择

变量的特征提取方式,结果表明,经变量选择后的NMR-MS中层融合模型识别效果较好,没有误分类现象。

2.1.4 高光谱成像数据融合 高光谱成像(hyperspectral imaging, HSI)融合了传统成像和光谱学原理,可在数百个波长的连续光谱范围内收集光谱和图像数据,从而提供被测物体的物理特征和内部化学成分,来表征和鉴定不同地理区域的药材和饮片,是一种快速无损、无污染、低成本的检测技术。研究^[43]采集了3个产地白术切片的可见光-近红外(Vis-NIR)和短波红外(short-wave infrared, SWIR)光谱和图像数据,通过连续投影算法(successive projection algorithm, SPA)提取光谱特征,通过灰度共生矩阵(gray-level co-occurrence matrix, GLCM)和灰度运行长度矩阵(gray-level run-length matrix, GLRLM)提取图像纹理特征,融合光谱和图像数据,结果表明,Vis-NIR-SWIR光谱和图像数据融合后分类能达到最优精度,明显高于单一光谱(80.9%)和单一图像(79.7%)的分类精度。文献^[44]测定了黄芪饮片的Vis-NIR和NIR高光谱图像,分别基于PCA和CNN提取原始光谱的特征,并进行低、中级融合,融合后数据分类能力均优于单一高光谱数据,表明HSI结合CNN和数据融合策略能够确定黄芪的产地。研究^[45]基于HSI从可见-近红外波段和短波红外波段提取不同产地甘草药材光谱和图像特征,从Vis-NIR和SWIR融合、光谱和图像融合以及全数据融合3个维度来建立产地分类模型,结果表明基于连续投影算法波段对全数据融合可以在保证分类准确率的基础上有效提高分类效率。

2.1.5 人工智能感官数据融合 人工智能感官技术能够模拟人体感官,快速获得药材饮片的形色气味并通过模式识别来表征其质量。研究^[46]分别基于无监督算法(PCA、HCA)和有监督算法(LDA、QDA)来考察单独使用电子鼻、电子舌和超高效液相色谱对2个产地厚朴饮片的鉴别能力,3类数据融合后基于LDA算法的产地分类模型性能最好,正确率为100%,表明电子感官和化学分析相结合能够从多维度来表征厚朴饮片的质量并鉴别其产地来源。

地理来源鉴定中数据融合的具体策略见表1^[26-46]。

2.2 药材及饮片品质鉴别及评价

药材和饮片是中医药发挥疗效的物质基础,从药材到饮片的生产流通过程中,野生与栽培、采收、炮制加工、运输、储藏等因素均会影响其质量。因此,对其进行真伪、品种、炮制品等鉴别也是中药质量评价的重要组成部分。

Table 1 Data fusion strategy of geographical origin identification

Aim of fusion	Data source of fusion	Methods of data pretreatment	Fusion level	Methods of feature extraction or variable selection	Data analysis
Geographical authentication of the epidermis and inner part of <i>Macrohyporia cocos</i> ^[27]	LC ₂₁₀₅ , LC ₂₄₂ ; FTIR; the epidermis and inner part of <i>Macrohyporia cocos</i>	Correlation optimized warping algorithm (chromatogram); second derivative (SD) and Savitzky-Golay filter (spectra)	Low-level; mid-level	PCs by PCA	PLS-DA
Geographical traceability of wild and cultivated <i>Macrohyporia cocos</i> ^[28]	FTIR; HPLC	Correlation optimized warping algorithm (chromatogram); SD and Savitzky-Golay filter (spectra)	Low-level; mid-level; high-level	Feature extraction by Boruta in RF; PCs by PCA	RF; PLS-DA
Geographical traceability and multielement analysis of <i>Wolfiporia cocos</i> ^[29]	ICP-AES; FTIR	MSC; SD (FTIR)	Mid-level	Variables extracted by VIP values; PCs by PCA	SVM; PLS-DA
Geographical origins identification of <i>Gentiana rigescens</i> ^[26]	HPLC; FTIR	SNV+SG-SD	Mid-level	PCs by PCA	PCA; PLS-DA
Geographical origin of <i>Gentiana Rigescens</i> ^[30]	HPLC data of rhizomes, stems and leaves	Correlation optimized warping algorithm	Low-level; mid-level	Relative variables extracted by Boruta in RF; variables selected by VIP values in OPLS-DA	RF; OPLS-DA
Geographical origin discrimination of <i>Radix Paeoniae</i> varieties ^[31]	HPLC; FTIR	Autoscaling	Mid-level	Important variables by GA-PLS; PCs by PCA	LDA; RBF-ANN; LS-SVM
Geographical traceability of <i>Panax notoginseng</i> ^[32]	FT-MIR; NIR	SNV; Savitzky-Golay filter	Low-level; mid-level; high-level	Variable selection by recursive feature elimination (RFE); PCs by PCA	RF; SVM; fuzzy set theory
Origin identification of <i>Panax notoginseng</i> ^[33]	FT-MIR; NIR	SNV; S-G; SD	Low-level; mid-level; high-level	Variable selection in RF; variable extraction by Boruta in RF	RF
Geographical origin identification of <i>Paris polyphylla</i> Smith var. <i>yunnanensis</i> ^[34]	ATR-FTMIR; NIR	SD	Low-level; mid-level; high-level	Variable selection by Boruta; variable selection by RFE; PCs by PCA	PLS-DA; RF+ fuzzy set theory
Geographical origin identification and quality evaluation of <i>Lycium ruthenicum</i> Murr. ^[35]	NIR; MIR	SNV	Mid-level	PCs by PCA	PCA; LS-SVM; Si-PLS
Origin identification of <i>Polygonatum Kingianum</i> ^[36]	ATR-FTIR; UV-Vis	First derivative (FD); SD; SNV; normalization	Low-level; mid-level; high-level	PCs by PCA; LVs by PLS-DA	RF; fuzzy set theory
Traceability of wild <i>Paris polyphylla</i> Smith var. <i>yunnanensis</i> ^[37]	FT-MIR; UV-Vis	SNV+SD+S-G; normalization	Low-level; mid-level; high-level	PCs by PCA	SVM-GS; RF+ fuzzy set theory
Origins discrimination of wild <i>Paris Polyphylla</i> Smith var. <i>yunnanensis</i> ^[38]	FT-MIR; UV-Vis	SNV+SD	Low-level; mid-level	PCs by PCA; LVs by PLS-DA	PLS-DA; SVM-GS
Origins discrimination of <i>Eucommia ulmoides</i> leaves ^[39]	ATR-FTIR; UV-Vis	SD; exponentially weighted moving average (EWMA)+SG; normalization	Low-level; mid-level	PCs by PCA; LVs by PLS-DA	HCA; PLS-DA; GS-SVM
Discrimination of geographical origins of <i>Radix Astragali</i> ^[40]	Raman spectra; UV-Vis	Wavelet denoising; normalization	Low-level; mid-level	PCs by PCA	KPCA; SVM; KNN; SRC (sparse representation classification)
Discrimination of honey of different floral origins ^[41]	UPLC-QTOF-MS; ICP-MS; IRMS; FTIR; NIR; Raman	Peak alignment; normalization; pareto scaling; logarithm transformation; FD; S-G smoothing; SNV	Low-level	-	PCA; OPLS-DA

Continued

Aim of fusion	Data source of fusion	Methods of data pretreatment	Fusion level	Methods of feature extraction or variable selection	Data analysis
Honey botanical origin characterization ^[42]	¹ H-NMR; LC-orbitrap-MS; LC-TOF-MS	Baseline correction and logarithm transformation (NMR); peak detection, alignment and integration (orbitrap-MS); peak alignment by non-linear locally estimated scatterplot smoothing and Pareto scaling (LC-TOF-MS)	Mid-level	Variable selection by VIP values; PCs by PCA	PCA; PLS-DA
Classification of geographical origins of <i>Rhizoma Atractylodis Macrocephalae</i> ^[43]	VNIR; SWIR hyperspectral data	SG (9-point, 13-point, 17-point, 21-point); SNV; MSC; FD; SD	Mid-level	Effective wavelengths by SPA; texture features extracted by GLCM and GLRLM	PLS-DA; SVM
Geographical origin identification of <i>Radix Astragali</i> ^[44]	Vis-NIR; NIR hyperspectral data	Wavelet transformation	Low-level; mid-level	PCs by PCA; deep spectral feature extraction by CNN	PCA; CNN; SVM; LR
Geographical origins identification of <i>Glycyrrhizae Radix et Rhizoma</i> ^[45]	Vis-NIR; SWIR hyperspectral data	S-G smoothing; SD	Mid-level	Featural wavelengths by SPA; texture features extracted by GLCM	PLS-DA; SVC (support vector classification); RF
Discrimination of two botanical origins of <i>Magnolia Officinalis Cortex</i> ^[46]	E-Nose; E-tongue; UPLC	Normalization	Low-level; mid-level	PCs by PCA; variable selection by VIP values (PLS)	PCA; HCA; LDA; QDA

2.2.1 真伪鉴别及评价 药材及饮片真伪鉴定是其质量评价的重要部分, 伪品流通市场, 将不利于保证饮片疗效的稳定一致, 使用不当可能会导致潜在的用药安全事故, 因此有必要对其进行鉴别。研究^[47]结合中红外和近红外光谱, 建立了一种区分半夏、虎掌南星(伪品)、姜半夏和姜虎掌南星(炮制伪品)的数据融合方法, 结果表明, 中级融合对分类结果的解释比低级融合更加直观, 能有效鉴别4类样本, 准确率达100%, 优于单一光谱。研究^[48]通过小波压缩(wavelet compression)和间隔偏最小二乘(iPLS)筛选近红外和中红外特征波段, 结合多种分类算法, 建立了快速、无损、准确辨识正品大黄和伪品大黄的中层次数据融合方法。文献^[49]测定了姜半夏及其伪品姜虎掌南星的NIR和MIR光谱, 以光谱PLS-DA模型潜变量为特征进行数据融合, 能够准确地预测两类样本并改善其在潜变量空间的分类情况。研究^[50]基于气、味信息, 联合电子鼻和电子舌对川贝母粉及其掺假品粉末进行鉴别, 数据融合后正判率优于单一电子舌数据, 与单一电子鼻鉴别能力相等。

2.2.2 品种鉴别及品质评价 药材品种不同, 其质量也会有差异, 如南柴胡与北柴胡存在14个差异成分且保肝降脂、药效结果也具有极显著差异^[51]; 禹白芷和杭白芷药材中香豆素类成分含量不尽相同^[52]。近源种属药材虽有相似功效, 但成分含量与临床疗效却也存在差异, 如金银花中绿原酸含量低于山银花, 木犀草苷含

量高于山银花, 金银花主治外感发热而山银花主治温病热症^[53]; 关黄柏与川黄柏中木兰花碱、巴马汀和小檗碱含量差异显著^[54]等。因此, 为保证临床用药的稳定一致, 有必要对药材或饮片进行品种鉴别。文献^[55]基于湿化学法和HPLC分别测定了4种莲子(湘莲、建莲、赣莲和湖莲)的总糖、总多糖和芦丁、莲心碱含量并鉴别了4个品种, 通过融合特征色谱峰和NIR特征光谱建立了快速预测莲子成分含量的方法, 结果表明, 数据融合后模型预测集 R^2 均在0.95以上, 能快速准确预测成分含量。研究^[56]使用HPLC-DAD和GC-MS获得了莪术3个品种(蓬莪术、广西莪术和温郁金)的二维指纹图谱, 基于PCA解析了各类样本与HPLC或GC-MS中各变量的相关性, 经数据融合后, PCA展示了各类样本更优的分离效果且分类模型性能也均提高。文献^[57]通过HPLC、FT-NIR和FT-MIR, 借助PCA和HCA, 分析了4种黄连属药材(短萼黄连、三角叶黄连、峨眉黄连和云南黄连)的整体代谢特征变异和分类关系, 并基于低、中层次数据融合构建了黄连素含量的预测模型, 结果表明基于VIP值的中级融合能降低模型预测误差, 为最优预测方法。研究^[58]联合NIR和UV-Vis光谱对11种石斛属药材进行鉴别, 数据融合后模型分类错误率降低, 且低级融合更能反映不同种属石斛之间的化学成分差异并有效鉴别, 推测中级融合中主成分提取虽获得了光谱有效信息, 但吸收较弱的重要光谱特征可能被剔除, 导致模型误差变大; 此外,

还采集了采收期为1~12月的石斛药材茎段和叶片的ATR-FTIR数据,融合两个部位数据对采收期进行了鉴别,结果显示,两部位的特征光谱融合可提高模型鉴别准确率。

2.2.3 炮制品鉴别及评价 饮片炮制会导致其外部色、性、味的物理变化和内部化学成分的变化,从而达到减毒增效的目的,对生品和炮制品的鉴别和全面质量评价有助于了解炮制过程中的物化变化,规范临床应用,保证其安全有效。研究^[59]分别用比色仪、电子鼻和近红外评价莪术炮制前后的整体颜色变化、挥发性成分差异及化学成分变化,通过改进的特征提取和数据融合方法(图5)快速鉴定并评价了莪术及其炮制品,预测准确率100%,此外,3个数据源的相关性分析结果表明,炮制过程中产生的芳香族衍生物与样品的色、味等外部变化密切相关,建立的数据融合评价方法可深入探究莪术炮制后的内部成分(近红外)和外部色味(比色仪、电子鼻)差异,全面了解饮片变化。研究^[60]结合NIR和高分辨率质谱的代谢组学数据,通过两种中级融合方法实现了硫熏与未硫熏麦冬的鉴别(图6),结果表明,数据融合能够可视化变量选择,增强样本分类能力,此外,数据特征中前3个主成分的其

他主成分的融合也可能达到与前者相等的分类效果。文献^[50]分别基于电子眼、电子鼻和电子舌分析了生黄连、酒黄连、姜黄连和萸黄连的色、气、味信息,并构建智能感官信息的单源及融合LDA模型,结果以电子鼻和电子舌融合后准确率最高,优于单一智能感官。

2.2.4 气味辨识及评价 四气五味是中药发挥疗效的整体物质基础的反映,不同药材的气味各有规律,融合药材气味信息构建软硬件分类系统是多学科领域发展的方向。文献^[61]融合辛味中药材气-味信息,结合控制变量法求最佳参数,使用流形算法中的局部线性嵌入算法(locally linear embedding, LLE)和LDA算法对3种生产日期的白术和6类辛味中药材进行数据降维、特征提取和分类,可准确辨别药材。研究^[62]利用图像、气体和味道传感器量化药材特征,提取特征向量,经SVM训练学习后建立药材特征数据库,进而开发了基于多传感器信息融合的药材分类和评级系统。文献^[63]基于辛味中药材气-味融合信息,建立了药材种类、产地及采收期分类鉴别的方法。研究^[64]在对中药水煎液的4种味觉(酸、苦、甜、咸)分类辨识时,尝试融合两类电子舌数据,结果发现数据融合后LS-SVM算法正判率较单一电子舌数据提高,提示了多传感器信息融合

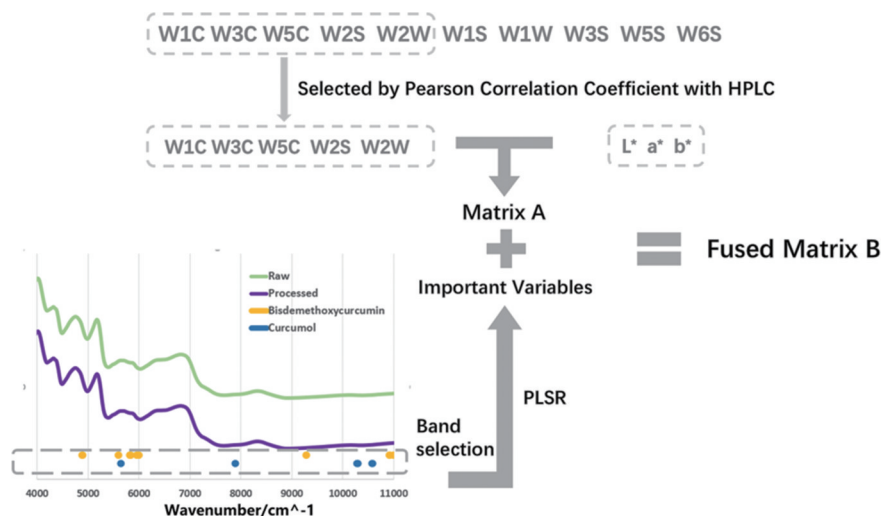


Figure 5 A mid-level data fusion approach for evaluating the internal and external changes in *Curcuma Rhizoma* processing^[59]

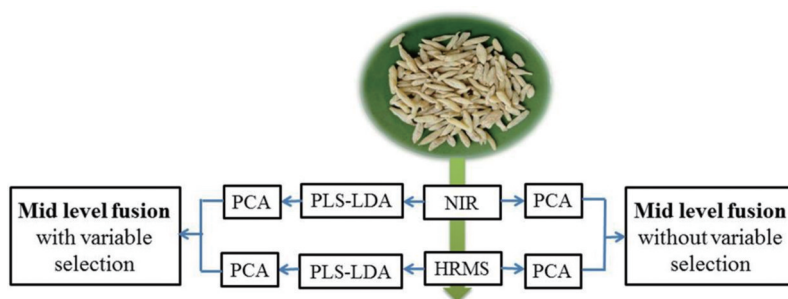


Figure 6 Data fusion strategy of discrimination of sulfur fumigation and non-sulfur fumigation *Ophiopogonis Radix*^[60]

用于中药味觉区分的可行性。

“走油”是药材和饮片储藏过程中常发生的变质现象,药材走油后,药性散失,疗效下降,更甚者会产生毒性,严重影响中医临床用药安全和稳定。研究^[65]将苦杏仁的颜色和气味信息相结合,分别建立了基于苦杏仁粉末颜色-气味融合信息、苦杏仁剖面颜色-气味融合信息对其走油程度的判别模型和苦杏仁苷含量、酸值含量预测模型,从而对其“走油”的变质情况进行全面监测。

2.2.5 辨色论质 中药饮片作为一种特殊商品,辨状论质是其在市场流通过程中的重要评价手段,也是其定档分级的指标。文献^[66]建立了红花“辨色论质”方法,根据指标成分含量、水浸出物含量和指纹图谱相似度对其分级,融合红花色值数据和NIR光谱,建立羟基红花黄色素A和水浸出物含量的综合定量预测模型,结果显示融合数据模型预测性能最优。

药材及饮片品质鉴别及评价中数据融合的具体策略见表2^[47-50,55-66]。

Table 2 Data fusion strategy of quality and variety assessment and identification

Aim of fusion	Data source of fusion	Methods of data pretreatment	Fusion level	Methods of feature extraction or variable selection	Data analysis
Identification of genuine and adulterated processed products of <i>Pinellia ternata</i> ^[47]	MIR; NIR	SG+FD (MIR); SG+SD (NIR); mean centering	Low-level; mid-level	Variable selection by VIP values; LVs by PLS-DA	PLS-DA
Identification of processed products of <i>Coptidis Rhizoma</i> ^[50]	Electronic nose; electronic tongue	-	Low-level	-	LDA
Identification of official rhubarb and unofficial rhubarb ^[48]	MIR; NIR	SG; FD; SD; MSC; normalization	Low-level; mid-level	Wavelet compression; wavelengths selection by iPLS	PLS-DA; SIMCA; SVM; ANN
Identification of ginger-processed <i>Pinelliaternata</i> and its counterfeits ^[49]	MIR; NIR	SG+FD (NIR); SG (MIR)	Mid-level	LVs by PLS-DA	PLS-DA
Identification of adulteration of <i>Fritillariae cirrhosae bulbus powder</i> ^[50]	Electronic nose; electronic tongue	Normalization	Low-level	-	DFA
Differentiation of lotus seed cultivars and quantitative predicting ^[55]	HPLC; NIR	SD	Mid-level	Wavelengths selected by GA-PLS	LDA; KNN; PLSR
Identification of similar, complex <i>Rhizoma Curcumae</i> samples ^[56]	HPLC; GC-MS	Normalization	Mid-level	Important variables selected by PCA loadings	LDA; BP-ANN; LS-SVM
Comprehensive quality assessment for <i>Rhizoma Coptidis</i> ^[57]	NIR; MIR	SNV+SD (NIR); SD (MIR)	Low-level; mid-level	PCs by PCA; variable selection by VIP values	PLSR
Spectrum fingerprint of <i>Dendrobium officinale</i> and similar species ^[58]	NIR; UV-Vis	SD; MSC	Mid-level	PCs by PCA	PLS-DA; SVM; KNN
Identification of harvesting period of <i>Dendrobium officinale</i> ^[58]	ATR-FTIR spectra of stems and leaves	SD; MSC	Mid-level	Important variables selected by RF	RF
Evaluation of the internal and external change in <i>Curcumae Rhizoma</i> processing ^[59]	Colorimeter; E-nose; FT-NIR	MSC; SNV; SG+FD; deviation normalization	Mid-level	GA (genetic algorithm); IRIV (iteratively retains informative variables); CARS (competitive adaptive reweighted sampling); wavelengths selection by VIP values > 1; the most related sensors with compounds selected in electronic nose	PLSR; PCA; PLS-DA
Discrimination of sulfur fumigation and non-sulfur fumigation <i>Ophiopogonis Radix</i> ^[60]	NIR; UHPLC-HRMS	Z-score; normalization	Mid-level	PCs by PCA; variables selection by VIP values	PCA; PLS-DA
Classification of pungent Chinese herbal medicines ^[61]	Electronic nose; electronic tongue	K neighboring points; the best regularization parameter	Low-level	-	LLE+LDA

Continued

Aim of fusion	Data source of fusion	Methods of data pretreatment	Fusion level	Methods of feature extraction or variable selection	Data analysis
Development of medicinal material classification system ^[62]	Image data; electronic nose; electronic tongue	Normalization	Mid-level	Color features by mean value, standard deviation and entropy; texture features by local binary pattern and GLCM	SVM
Differentiation of varieties, growing areas and production dates of pungent Chinese herbal medicines ^[63]	Electronic nose; electronic tongue	-	Low-level	-	PCA+LDA
Classification of four kinds of taste of natural medicines ^[64]	TS-5000Z electronic tongue; α -ASTREE electronic tongue	Autoscaling	Low-level	-	LS-SVM; DA; PCA-DA; PLS-DA
Oil-releasing monitoring of <i>Armeniacae Semen Amarum</i> ^[65]	Color (spectrophotometer); odor (electronic nose)	Z-score	Mid-level	Color features; sensors with high response value	Naive Bayes; SVM; logistic; multiple layer perception; KStar; RF
Color grading of safflower ^[66]	Machine vision; NIR spectra	Baseline; SNV	Mid-level	Variable selection by VIP values; LVs by PLS	PLSR

2.3 制药过程分析与制剂质量控制

2.3.1 制药过程分析 中药原料质量波动和工艺过程参数波动是影响中药产品(或中间体)质量的主要变异来源^[67],过程分析技术(process analytical technology, PAT)和质量源于设计(quality by design, QbD)理念鼓励采用增强的分析工具、工艺建模方法和先进过程控制策略,增加对制药过程质量传递规律、关键质量属性(critical quality attributes, CQA)变化的理解和产品质量一致性的控制。研究^[68]融合了NIR和MIR光谱数据,采用PLS对小儿消积止咳口服液提取过程中辛弗林、槟榔碱、绿原酸、连翘苷A、柚皮苷、橙皮苷和新橙皮苷7种CQA的含量变化进行了研究,结果表明,基于NIR光谱的模型含量预测能力优于MIR;两类光谱特征融合后的模型性能优于单一光谱数据, R_p^2 均大于0.97, RMSEP最小为0.003,所建立的光谱融合方法可为中药生产过程快速监控提供参考(图7)。基于QbD理念,研究^[69]分别提取文献数据和融合文献数据及少量实验室数据建立PLS模型,输入处方材料属性矩阵和工艺参数矩阵的融合数据,对高剪切湿法制粒工艺过程中颗粒CQA(中值粒度G50)进行了预测,结果显示,文献数据由于各实验条件差异,模型稳健性不强(RMSEP为152.6),而融入少量实验室数据可显著提高其预测精度(RMSEP为115.7)。研究^[67]将丹参饮片原料属性、提取过程工艺参数和过程状态变量(提取液NIR光谱)组合为自变量,以其醇提液中的有效成分(丹参酮II A、隐丹参酮和丹参酮I)含量为因变量,建立了丹参提取液质量预测的PLS模型(图8),结果表明融合模型的校正和预测性能较优,NIR光谱包

含了提取过程动态变化的综合信息,可有效填补原料属性、工艺质量控制盲区,信息有效集成,可共同参与产品CQA的预测。

2.3.2 制剂质量控制 溶出度是评价中药固体制剂质量一致性的重要指标,其取决于原料性质、制剂工艺和受控程度。文献^[70]基于NIR和显微拉曼光谱建立了自制盐酸青藤碱缓释片累积溶出率的预测模型,结果显示NIR和处方配比、工艺参数融合后模型预测能力更好,相比于PLSR模型,BP-ANN的外推能力更强,该方法可为其他中药固体制剂溶出行为的模型化及生产监控提供参考。

制药工艺优化有助于提高中药生产效率、降低成本、加强环境保护和劳动保护等,但终产品质量会因工艺变化而改变,因此,制备工艺一致性评价对保证中药制剂质量及中医药疗效安全稳定至关重要。文献^[71]为探究盐酸青藤碱制备工艺变更的可行性,融合NIR和拉曼光谱,建立了苯提取和酸水渗漉-氯仿萃取后所制的两类盐酸青藤碱样品的相似度匹配值(similarity match value, SMV)预测模型,并综合PCA分析中Hotelling T²和DModX两种统计量对其质量进行了表征,中水平融合模式下,检测出有质量差异的样本仅1个,低、高水平融合模式下,分别检测出9、8个,结果以中水平融合的质量一致性评价方法最优,两法制得样品质量一致性良好。

中药制剂质量不仅受原材料的影响,生产厂家、批次也会导致同种中药制剂的疗效差异,因此准确鉴别其质量是保障用药安全和制药行业发展的关键。文献^[72]融合了分别在原子和分子层面上测定样本信息的

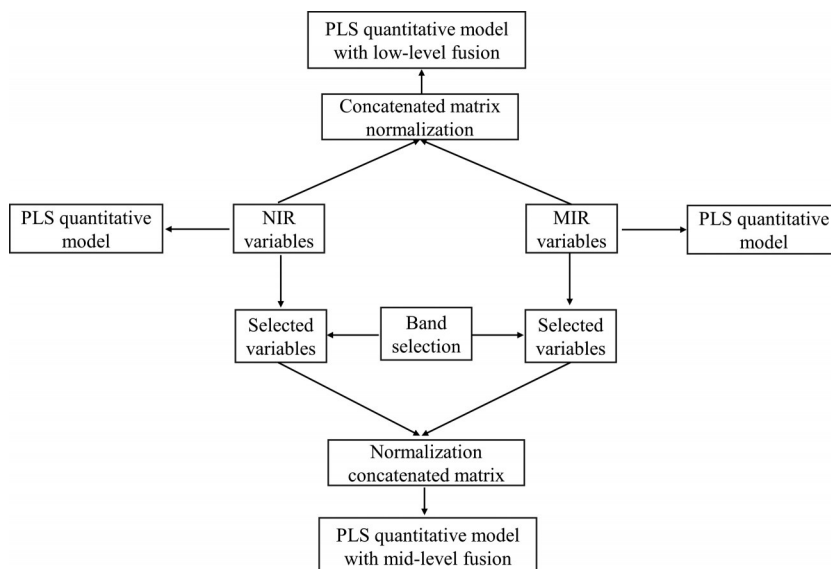


Figure 7 Data fusion strategy of multi critical quality attributes monitoring in Xiao'er Xiaoji Zhike Oral Liquid extraction process^[68]

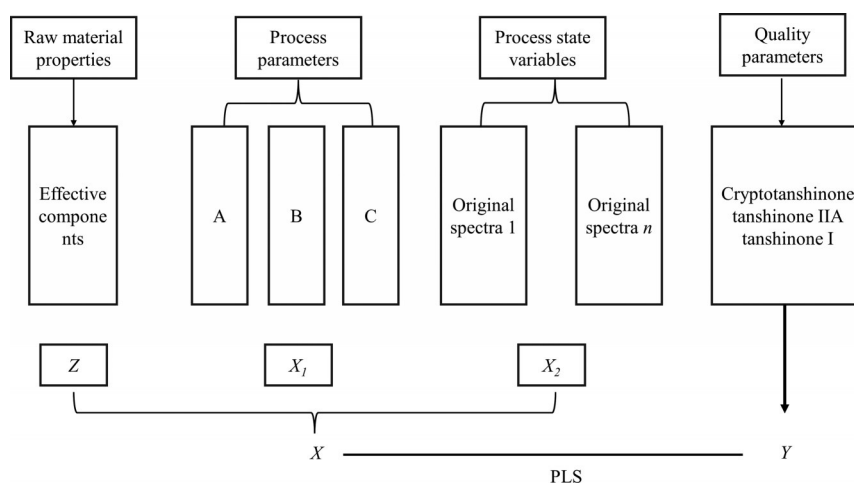


Figure 8 Flowchart of multi-source information fusion modeling in the extraction process of *Salvia miltiorrhiza*^[67]

激光诱导击穿光谱 (LIBS) 和 NIR 光谱, 对来自 3 个厂家的 18 种复方丹参片进行了鉴别, RF 判别模型以经变量重要性测量指标 (VIM) 选择变量后的融合数据性能最好。研究^[73]基于 UPLC、智能感官对 26 个厂家收集到的小柴胡颗粒进行定性和定量评价, 融合电子鼻、电子眼、电子舌数据, 结合化学计量学方法, 建立了小柴胡颗粒中柴胡皂苷 b2 和甘草皂苷 G2 的含量预测模型, 可为中药制剂的含量评价提供新方法 (图 9)。

制药过程分析与制剂质量控制数据融合具体策略见表 3^[67-73]。

3 总结与进展

中药质量受多种因素干扰, 单一检测技术无法全面表征其质量信息, 具有一定局限性。文献表明, 信息融合技术已逐渐应用于中药质量评价领域, 可为该问题的解决提供参考。多种仪器数据的有机整合, 可充

分利用每类检测技术的优点, 结合化学计量学方法构建融合模型, 使中药质量评价更科学、全面、精准。

整理相关文献发现: ① 信息融合技术用于药材或饮片产地鉴别的最多, 表明产地信息更适合通过多种分析技术的融合数据来表征, 是中药质量评价的重要部分; 用于制药过程分析的较少, 还有待开发, 制药过程监测需要来自多个分析技术的支撑, 信息的有效融合对于药品生产者增强工艺过程理解、提高制药效率、保证生产质量有重要参考价值。在大类别上, 该技术主要用于中药产地溯源、品种鉴别和制药过程分析, 具体来看, 涉及品质评价、产地、真伪、品种、炮制品、性味、采收期鉴别及走油监测、中药制剂质量评价、过程工艺指标预测等, 即多源信息融合建模的 Y 值有多种, 可见信息融合技术能够用于多方面中药质量评价, 有广阔的应用前景; ② 采用光谱技术融合的文献数目最

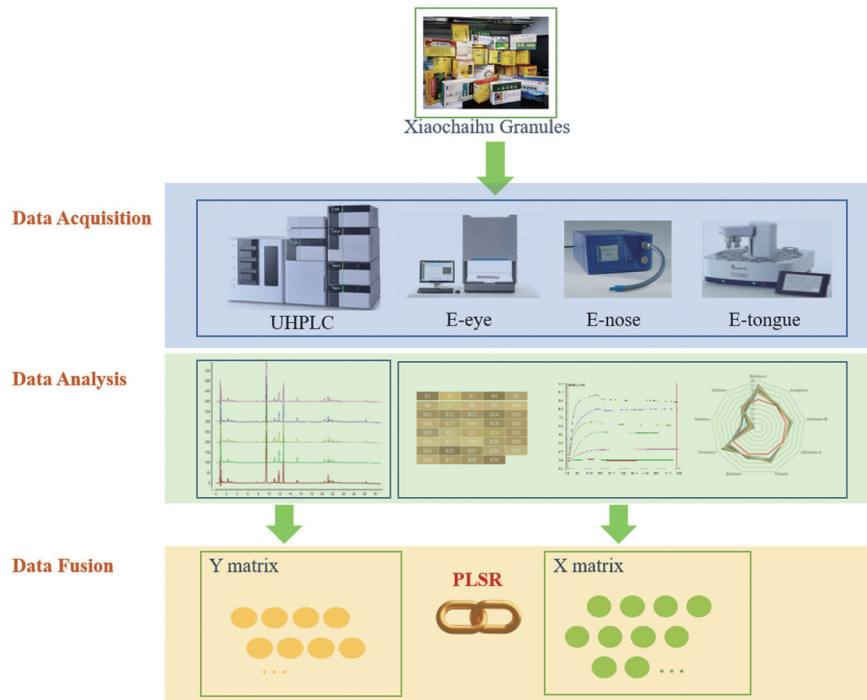


Figure 9 Qualitative and quantitative assessment of Xiaochaihu granules^[73]

Table 3 Data fusion strategy of pharmaceutical process analysis and quality control

Aim of fusion	Data source of fusion	Methods of data pretreatment	Fusion level	Methods of feature extraction or variable selection	Data analysis
Multi critical quality attributes monitoring in the extraction process of Xiao'er Xiaoji Zhike Oral Liquid ^[68]	NIR; MIR	Normalization; SNV; SG; MSC; CWT (continuous wavelet transform); derivatives	Low-level; mid-level	UVE (uninformative variable elimination); CARS; CC (correlation coefficients)	PLS
Development of formulation-process-quality model for high shear wet granulation ^[69]	Literature data; laboratory data; material properties; process parameters	Centering; scaling	Low-level	-	PLSR
Quality prediction in the extraction process of <i>Salvia miltiorrhiza</i> ^[67]	Raw material properties; process parameters; NIR spectra in process state	MSC; SNV	Low-level	-	PLS
Prediction of dissolution rate of sinomenine hydrochloride sustained-release tablets ^[70]	NIR; Roman spectra; prescription ratio; process parameters	SNV+FD (NIR); normalization	Mid-level	PCs by PCA	PLSR; BP-ANN
Consistency evaluation of preparation of sinomenine hydrochloride ^[71]	NIR; Roman spectra	S-G; FD; normalization	Low-level; mid-level; high-level	PCs by PCA	SMV; hotelling T ² ; DModX
Identification of the manufacturers of compound <i>Salvia miltiorrhiza</i> tablets ^[72]	LIBS; NIR	Normalization	Mid-level	VIP; variable importance measure (VIM)	RF
Qualitative and quantitative assessment of Xiaochaihu granules ^[73]	Electronic nose; electronic tongue; electronic eye	Normalization	Low-level	-	PLSR

多,共46个(本文中占比80%),由于光谱技术具有快速无损和操作简单等特点,其在中药质量评价中的应用最为广泛,色谱、质谱、智能感官技术融合的研究虽然相对较少,但仍是中药质量评价的重要手段,未来可针对其他检测技术的信息融合策略开展研究;③在中药质量评价领域,融合对象不仅可以是不同分析仪器

的数据,还可以是药材不同部位的信息,这是中药材的特殊属性,非传统药用部位信息的加入有助于模型识别,提示在其他领域数据融合中,可根据目标问题适当泛化融合对象,从而得到更多被忽略的质量信息;④中级融合被运用于中药质量评价的频次较多(本文中占比80%),特征提取方式集中在PCA提取主成分得

分、VIP 筛选变量和 PLS-DA 提取潜变量 3 种类型,此外,还有光谱选择、图像特征提取的专属算法及基于相关指标筛选原始变量的方法(如文献^[59])等。高级融合使用较少(本文中占比 15%),且大多数是通过 RF 结合模糊集理论被用于产地溯源,同时,并非所有高级融合模型性能均最优,这与数据来源、数据性质和模型算法均有关;⑤ 信息融合技术应用于定性质量评价的较多(本文中占比 80%),应用于定量的较少(本文中占比 24%),且定量评价集中在品质检测和制药过程分析方面。定性评价中,融合算法以 PLS-DA、SVM、RF 使用较多,PLS-DA 有降维、简化样本信息作用,其机制是将数据转换成一组线性潜在变量用于分类变量预测,通过已知类别作为训练集建立分类模型,对未知样本进行预测;SVM 是基于核函数的算法,能够对原始数据进行某种非线性映射,从而使原本难以分类的样本数据经过非线性映射之后在新的高维空间被一个超平面分开聚类;RF 基于 Bagging (套袋算法)思想,利用多棵树对样本进行训练并预测,包含多个决策树,输出的类别由个别树输出的类别的众数而定。定量评价中,融合算法以 PLSR 应用较多,PLSR 是处理自变量和因变量之间关系的一种强有力双线性回归方法,通过交叉验证建立校正模型,提取自变量线性组合的部分正交因子作为回归变量,用于定量预测因变量。

信息融合技术在中药地理来源鉴定、品质鉴别及评价和制药过程分析等研究中已有初步应用,有助于其质量评价和控制,但仍然存在以下问题需进一步关注和研究:① 仪器数据的准确性和重复性。信息融合的效果受检测仪器的原始数据影响,若基础信息不准确,融合后模型性能也将受限,仪器检测方法的重复性也与模型预测结果能否稳定重现相关联;② 数据预处理方法的选择。不同检测设备获得的信息模式和跨度均很大,无法直接关联;此外,还应避免多源数据量级的差异减慢算法迭代收敛速度以及量级较大的属性占主导地位的现象,因此数据预处理方法的选择也很重要,应结合数据集自身特点(偏度是否较大、是否引入哑变量)、模型种类、问题类型(回归或是分类)等^[74],综合分析并选取适宜的预处理方法或多法联合使用;③ 样本量大小及样本均一同质性。样本过少,模型泛化能力不强、过多,则模型运算速度降低;样本同质性对模型的外推性能也有重要影响,在药材及饮片质量评价中,样本多以粉末形式进行检测,虽能保证样本均一同质,但具有破坏性,多示例学习理论能够实现多个量测示例结果的融合,能较好地解决药材及饮片被检测时非均一同质的问题,如林兆洲^[75]以多示例学习理论为基础,以 LS-SVM 为基分类器建立了两种空间数

据融合方法,可以提高分类准确性和表征覆盖度,具有一定参考价值;④ 特征提取方式。从本文来看,PCA 提取主成分的特征提取方式使用频次较多,然而 PCA 是无监督分析,提取的主成分与 Y 值的关系不如监督分析算法提取的特征变量与 Y 值间的关系大,如文献^[28,36,38-40,46,57]中对比了有监督变量 LVs 或 VIP 或 Boruta 特征变量与无监督变量主成分 PCs 的性能,均是前者效果优于后者,表明特征提取方式对融合结果也有影响,未来期望对此有所突破;⑤ 融合层级的选择。并不是融合层次越高模型性能就越好,低、中、高融合各有优缺点,应结合数据性质和目标问题多次尝试,优选融合方法;⑥ 数据共享。云端数据库的建立可实现中药质量信息共享,未来有望基于大数据开发各类药材及饮片产地、品种等鉴别数据库,或中药制剂过程分析数据库,使中药质量评价更科学、高效,从而促进智能制造及中药现代化的发展。

目前多源信息融合技术主要还在研究探索阶段,在日常生产实践中使用度和使用率不足,其原因主要有^[76]:① 多源信息融合技术是一门边缘学科,其应用面非常广泛,各行各业会按照自己的理解给出不同的定义;② 融合信息类型高度相异且内容具有含糊属性,如军事领域中融合信息所处时域上不同步,空域上位于不同坐标系等;③ 信息融合作为对多源信息、多个任务的综合处理过程,具有本质上的、固有的复杂性;④ 目前尚没有数学工具用来统一描述和处理此类复杂的问题,无法实现多领域统一、对等的应用,信息融合还缺乏系统的理论基础,技术成熟度不够,尚未形成一个完整的理论体系和解决方法;⑤ 多学科(计算机、数学、统计学等)复合型人才较少,交叉学科中各类专家专注的问题不同,因此会导致拥有该技术但缺乏可应用的研究对象的现象,或者拥有可应用的研究对象但缺乏技术力量的现象。而且,各学科交流语言不同,容易产生信息语言的歧义问题;⑥ 由于数据融合、传感器技术缺陷及多样性和自然的应用环境而产生的数据缺陷、异常和虚假数据、数据冲突、数据形式、数据关联、数据校准/匹配、数据联合、处理框架、操作时序、静态与动态以及数据维度等需解决的问题。

随着新型传感器的不断涌现,以及现代信号处理技术、计算机技术、网络通信技术、人工智能技术、并行计算的软件和硬件技术等相关技术的飞速发展,多传感器信息融合将成为我国未来大量军用和民用高科技系统的重要技术手段。有望应用于图像融合、工业智能机器人、遥感、刑侦、故障诊断、智能交通系统等多项领域。未来其主要研究方向应在于:① 发展和完善信息融合的基础理论,丰富信息融合理论体系;② 改进

融合算法以提高系统性能。异类信息的建模、融合尚需算法的支持,大量引入现代统计推断方法有助于解决复杂问题;③ 开发并行处理算法及并行计算的软硬件;④ 传感器资源的管理和优化,包括时间、空间、模式管理;⑤ 建立适用于信息融合系统的数据库与知识库,形成优化的存储机制、检索机制、推理机制,提高整个系统的运行效率;⑥ 研究开发对融合算法和系统性能客观评价的仿真和测试平台,可运行大量仿真数据;⑦ 建立性能评估方法和评价机制,对信息融合系统进行综合分析和评价。

中药领域信息资源十分丰富,包含大量的、原始的、需规范的数据和信息等信息原料。信息的规范化和标准化有利于多学科间共通交融。中药来源、成分、药效复杂、多样,信息融合需要多种分析检测技术的支撑和互补,因此需要多个中药学二级学科通力合作,需中药学科工作者扩展思维,挖掘学科共通点,深入数据分析,共同促进中药质量评控。此外,还应增强实验室原始数据管理,提高数据分析质量,加深数据挖掘,提取有价值、有意义信息。

近年来,以 ChatGPT 为代表的生成式人工智能技术以其海量的数据和强大的运算能力为研究者展示了全新的应用前景,其决策过程与基于多源信息融合用于中药质量评价有相似之处,有望应用于中药质量评价领域。但由于该领域专业化程度较高,因此还需为其提供专业化的文本数据库、基于针对性的训练方法来对模型和数据进行调整,使模型被赋予更广泛的领域知识,能够从多个信息源中获取和整合信息;此外,还需注意将海量的仪器量测数据尤其是实时的新数据增加到原有的文本数据库中,对其中不实、偏见或过时的信息进行更正,并以更新的文本数据库为基础,进行进一步的模型训练,以实现技术模型和应用工具的与时俱进。同时生成式人工智能用于医药领域中时还应注意伦理、道德问题,必要时介入法定手段,尽量避免无人决策结果负责的现象。

作者贡献: 李涵为本文第一作者,负责查找文献、数据汇总和论文撰写;谢梦迪负责文献收集;桂新景、侯富国、范雪花、李海洋和王艳丽对论文语言修改提供了一定帮助;刘瑞新和李学林负责论文整体设计及修改指导。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Xiao XH. Towards accurate quality evaluation and control of traditional Chinese medicine [J]. Acta Pharm Sin (药学报), 2019, 54: 2139-2140.
- [2] Sun Y, Xu G, Ma SC. Development of an overall evaluation system for traditional Chinese medicine [J]. Acta Pharm Sin (药学报), 2021, 56: 1749-1756.
- [3] Xiao XH, Jin C, Yan D, et al. Proposition and practice on "integrative quality" in quality control for Chinese materia medica [J]. Chin Tradit Herbal Drugs (中草药), 2010, 41: 505-508.
- [4] Liu TR, Jin Y, Meng HB, et al. Biological research of color and quality evaluation in "quality discrimination by character" of Chinese medicine [J]. Chin J Chin Mater Med (中国中药杂志), 2020, 45: 4545-4554.
- [5] Hu HF, Xi SY, Cao HK, et al. Quality evaluation of Cnidii Fructus in commodity grade based on theory of "quality evaluation through morphological identification" [J]. Chin J Chin Mater Med (中国中药杂志), 2023, 48: 900-907.
- [6] Li ML, Gao WY, Wang HJ, et al. Comprehensive evaluation system for quality of Chinese medicinal decoction pieces based on "experience-ingredients-activity-electronic sensing" [J]. Chin J Chin Mater Med (中国中药杂志), 2022, 47: 1995-2007.
- [7] Fu HT, Zhang QR, Xiong R, et al. Research progress in quality evaluation of traditional Chinese medicine based on fingerprint combined with QAMS [J]. China Pharm (中国药师), 2022, 25: 861-867.
- [8] Zhang L, Liu Y, Liu Z, et al. Comparison of the roots of *Salvia miltiorrhiza* Bunge (Danshen) and its variety *S. miltiorrhiza* Bge f. *Alba* (Baihua Danshen) based on multi-wavelength HPLC-fingerprinting and contents of nine active components [J]. Anal Methods, 2016, 8: 3171-3182.
- [9] Cui S, Wu J, Wang J, et al. Discrimination of American ginseng and Asian ginseng using electronic nose and gas chromatography-mass spectrometry coupled with chemometrics [J]. J Ginseng Res, 2017, 41: 85-95.
- [10] Zuo YM, Deng XH, Wu Q. Discrimination of *Gastrodia elata* from different geographical origin for quality evaluation using newly-build near infrared spectrum coupled with multivariate analysis [J]. Molecules, 2018, 23: 1088-1105.
- [11] Liu CX. Developing the theoretical methods and strategies of quality markers (Q-marker) of traditional Chinese medicine, and improving the scientific and technological level of traditional Chinese medicine. [J]. Acta Pharm Sin (药学报), 2019, 54: 185-186.
- [12] Zhang TJ, Bai G, Liu CX. The concept, core theory and research methods of Chinese medicine quality markers [J]. Acta Pharm Sin (药学报), 2019, 54: 187-196.
- [13] Zhang B, Xin J, Liu YJ, et al. Application of bioassay in quality evaluation of Chinese medicine: a review [J]. Chin J Chin Mater Med (中国中药杂志), 2022, 47: 6264-6270.
- [14] Chen HY, Zhang DW, Yang H, et al. Design and implementation of intelligent knowledge service system of aerospace military experts [J]. Mod Electronics Tech (现代电子技术), 2021, 44: 8-12.
- [15] Xu WB, Wang X, Qin CW, et al. Research on aero-engine

- condition assessment based on D-S evidence theory [J]. Plant Maint Eng (设备管理与维修), 2021, (16): 28-29.
- [16] Zhou G, Qu Y, Zhang SG. Application of ensemble learning in multi-source information fusion and target identification [J]. J Ordnance Equip Eng (兵器装备工程学报), 2021, 42: 166-169, 254.
- [17] Jiang S, Ni C, Chen G, et al. A novel data fusion strategy based on multiple intelligent sensory technologies and its application in the quality evaluation of Jinhua dry-cured hams [J]. Sens Actuator B Chem, 2021, 344: 130324.
- [18] Chen KW, Zhang ZP, Long J. Moultsource information fusion: key issues, reearch progress and new trends [J]. Computer Sci (计算机科学), 2013, 40: 6-13.
- [19] Xu BW, Ma ZY, Li Y. Research progress and application of multi-sensor information fusion technology in environmental perception [J]. Computer Measurement Control (计算机测量与控制), 2022, 30: 1-7, 21.
- [20] Qi YJ, Wang Q. Review of multi-source data fusion algorithm [J]. Aerosp Electronic Warf (航天电子对抗), 2017, 33: 37-41.
- [21] Li Y, Zhao M, Xu MY, et al. A survey of research on multi-source information fusion technology [J]. Intell Computer Appl (智能计算机与应用), 2019, 9: 186-189.
- [22] Alessandra B, Remo B, Antonio LM, et al. Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication [J]. Anal Chim Acta, 2014, 820: 23-31.
- [23] Crockford DJ, Holmes E, Lindon JC, et al. Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies [J]. Anal Chem, 2006, 78: 363-371.
- [24] Buratti S, Malegori C, Benedetti S, et al. E-nose, e-tongue and e-eye for edible olive oil characterization and shelf life assessment: a powerful data fusion approach [J]. Talanta, 2018, 182: 131-141.
- [25] Huang LQ, Chen ML, Xiao PG. Modern biological basis and model hypothesis of genuineness research of Chinese medicinal materials [J]. Chin J Chin Mater Med (中国中药杂志), 2004, 29: 494-496, 610.
- [26] Wang QQ, Shen T, Zuo ZT, et al. Data fusion and multi-components quantitative analysis for identification and quality evaluation of *Gentiana rigescens* from different geographical origins [J]. Chin J Chin Mater Med (中国中药杂志), 2018, 43: 1162-1168.
- [27] Wang QQ, Huang HY, Wang YZ. Geographical authentication of *Macrohyporia cocos* by a data fusion method combining ultra-fast liquid chromatography and fourier transform infrared spectroscopy [J]. Molecules, 2019, 24: 1320.
- [28] Wang L, Wang QQ, Wang YZ, et al. Comparison of geographical traceability of wild and cultivated *Macrohyporia cocos* with different data fusion approaches [J]. J Anal Methods Chem, 2021, 2021: 5818999.
- [29] Yue JQ, Li ZM, Zuo ZT, et al. Geographical traceability and multielement analysis of edible and medicinal fungi: taking *Wolfiporia cocos* (F.A. Wolf) Ryvarden and Gilb. as an example [J]. J Food Sci, 2021, 86: 770-778.
- [30] Shen T, Yu H, Wang YZ. Assessing geographical origin of *Gentiana Rigescens* using untargeted chromatographic fingerprint, data fusion and chemometrics [J]. Molecules, 2019, 24: 2562.
- [31] Ni YN, Li BH, Kokot S, et al. Discrimination of Radix Paeoniae varieties on the basis of their geographical origin by a novel method combining high-performance liquid chromatography and Fourier transform infrared spectroscopy measurements [J]. Anal Methods, 2012, 4: 4326-4333.
- [32] Li Y, Zhang JY, Wang YZ. FT-MIR and NIR spectral data fusion: a synergetic strategy for the geographical traceability of Panax notoginseng [J]. Anal Bioanal Chem, 2018, 410: 91-103.
- [33] Zhou YH, Zuo ZT, Xu FR, et al. Origin identification of *Panax notoginseng* by multi-sensor information fusion strategy of infrared spectra combined with random forest [J]. Spectroc Acta A Mol Biomol Spectrosc, 2020, 226: 117619.
- [34] Pei YF. Identification Research of Paris polyphylla Smith var. yunnanensis Based on Multi-spectral Fingerprint Combined with Chemometrics (滇重楼多光谱指纹图谱结合化学计量学鉴别研究) [D]. Kunming: Yunnan University of Chinese Medicine (云南中医药大学), 2020.
- [35] Li YH, Li YX, Tan WL, et al. Rapid identification and evaluation of *Lycium ruthenicum* Murr. by near-infrared and fourier transform infrared spectroscopy [J]. Spectrosc Spectr Anal (光谱学与光谱分析), 2020, 40: 3878-3883.
- [36] Zhang J, Wang YZ, Yang WZ, et al. Data fusion of ATR-FTIR and UV-Vis spectra to identify the origin of *Polygonatum kingianum* [J]. Spectrosc Spectr Anal (光谱学与光谱分析), 2021, 41: 1410-1416.
- [37] Wu XM, Zhang QZ, Wang YZ. Traceability of wild *Paris polyphylla* Smith var. *yunnanensis* based on data fusion strategy of FT-MIR and UV-Vis combined with SVM and random forest [J]. Spectroc Acta A Mol Biomol Spectrosc, 2018, 205: 479-488.
- [38] Wu XM, Zuo ZT, Zhang QZ, et al. FT-MIR and UV-vis data fusion strategy for origins discrimination of wild *Paris Polyphylla* Smith var. *yunnanensis* [J]. Vib Spectrosc, 2018, 96: 125-136.
- [39] Wang ZY. Authentication and Quality Evaluation of *Eucommia ulmoides* Leaves Based on Multi-information Fusion (基于多源信息融合的杜仲叶产地溯源及品质评价) [D]. Jishou: Jishou University (吉首大学), 2020.
- [40] Wang HY, Song C, Sha M, et al. Discrimination of medicine Radix Astragali from different geographic origins using multiple spectroscopies combined with data fusion methods [J]. J Appl Spectrosc, 2018, 85: 313-319.
- [41] Jandrić Z, Haughey SA, Frew RD, et al. Discrimination of honey of different floral origins by a combination of various chemical parameters [J]. Food Chem, 2015, 189: 52-59.

- [42] Spiteri M, Dubin E, Cotton J, et al. Data fusion between high resolution (1)H-NMR and mass spectrometry: a synergetic approach to honey botanical origin characterization [J]. *Anal Bioanal Chem*, 2016, 408: 4389-4401.
- [43] Ru CL, Li ZH, Tang RZ. A hyperspectral imaging approach for classifying geographical origins of rhizoma *Atractylodis Macrocephalae* using the fusion of spectrum-image in VNIR and SWIR ranges (VNIR-SWIR-FuSI) [J]. *Sensors (Basel)*, 2019, 19: 2045.
- [44] Xiao QL, Bai XL, Gao P, et al. Application of convolutional neural network-based feature extraction and data fusion for geographical origin identification of *Radix Astragali* by visible/short-wave near-infrared and near infrared hyperspectral imaging [J]. *Sensors (Basel)*, 2020, 20: 4940.
- [45] Yin WJ, Ru CL, Zheng J, et al. Fusion of spectrum and image features to identify *Glycyrrhizae Radix et Rhizoma* from different origins based on hyperspectral imaging technology [J]. *Chin J Chin Mater Med (中国中药杂志)*, 2021, 46: 923-930.
- [46] Jing WG, Zhao XL, Li MH, et al. Application of multiple-source data fusion for the discrimination of two botanical origins of *Magnolia officinalis* Cortex based on e-nose measurements, e-tongue measurements, and chemical analysis [J]. *Molecules*, 2022, 27: 3892.
- [47] Sun F, Chen Y, Wang KY, et al. Identification of genuine and adulterated *Pinellia ternata* by mid-infrared (MIR) and near-infrared (NIR) spectroscopy with partial least squares-discriminant analysis (PLS-DA) [J]. *Anal Lett*, 2020, 53: 937-959.
- [48] Sun WJ, Zhang X, Zhang ZY, et al. Data fusion of near-infrared and mid-infrared spectra for identification of rhubarb [J]. *Spectrosc Acta A Mol Biomol Spectrosc*, 2017, 171: 72-79.
- [49] Sun F, Chen Y, Wang KY, et al. Identification method of ginger-processed *Pinelliaternata* based on infrared spectroscopy data fusion [J]. *J Beijing Univ Tradit Chin Med (北京中医药大学学报)*, 2019, 42: 862-868.
- [50] Yang SL. Discrimination of *Bulbus of Fritillaria* and Processed Products of *Coptidis Rhizoma* by Using Intelligent Sensory Technologies (基于智能感官分析技术的贝母及黄连饮片鉴别研究) [D]. Chengdu: Chengdu University of Chinese Medicine (成都中医药大学), 2015.
- [51] She YP, Zhang PP, Hu JY, et al. Precise application of *Beihaihu* and *Nanchaihu* in classical Formulas [J]. *Chin J Chin Mater Med (中国中药杂志)*, 2023, 48: 681-688.
- [52] Chen L, Tang ZS, Song ZX, et al. Quantitative determination of nine furanocoumarins for quality evaluation of *Angelica dahurica* from different habitats [J]. *Chin J Chin Mater Med (中国中药杂志)*, 2019, 44: 3002-3009.
- [53] Li RE. Study on character identification and pharmacological action of *Lonicera japonica* and *Lonicera hypoglauca* [J]. *Clin J Chin Med (中医临床研究)*, 2020, 12: 118-119, 136.
- [54] Feng Y, Guo L, Chang YQ, et al. Research on difference of *Phellodendron amurense* and *Phellodendron chinense* based on fingerprint and multicomponent content determination [J]. *Chin Tradit Herbal Drugs (中草药)*, 2022, 53: 5179-5184.
- [55] Guo Y, Ding XX, Ni YN. The combination of NIR spectroscopy and HPLC chromatography for differentiating lotus seed cultivars and quantitative prediction of four main constituents in lotus with the aid of chemometrics [J]. *Anal Methods*, 2017, 9: 6420-6429.
- [56] Ni YN, Mei MH, Kokot S. One- and two-dimensional gas chromatography-mass spectrometry and high performance liquid chromatography-diode-array detector fingerprints of complex substances: a comparison of classification performance of similar, complex *Rhizoma Curcumae* samples with the aid of chemometrics [J]. *Anal Chim Acta*, 2012, 712: 37-44.
- [57] Qi LM, Ma YT, Zhong FR, et al. Comprehensive quality assessment for *Rhizoma Coptidis* based on quantitative and qualitative metabolic profiles using high performance liquid chromatography, Fourier transform near-infrared and Fourier transform mid-infrared combined with multivariate statistical analysis [J]. *J Pharm Biomed Anal*, 2018, 161: 436-443.
- [58] Wang Y. Study of Spectrum Fingerprint of *Dendrobium officinale* and Similar Species in Yunnan Province (云南铁皮石斛及其相似种光谱指纹图谱研究) [D]. Kunming: Yunnan University of Chinese Medicine (云南中医药大学), 2020.
- [59] Lan ZW, Zhang Y, Sun Y, et al. A mid-level data fusion approach for evaluating the internal and external changes determined by FT-NIR, electronic nose and colorimeter in *Curcumae Rhizoma* processing [J]. *J Pharm Biomed Anal*, 2020, 188: 113387.
- [60] Dai SY, Lin ZZ, Xu B, et al. Metabolomics data fusion between near infrared spectroscopy and high-resolution mass spectrometry: a synergetic approach to boost performance or induce confusion [J]. *Talanta*, 2018, 189: 641-648.
- [61] Fan DJ. Data Fusion Research of Odor and Taste of Pungent Chinese Herbal Medicines Based on Multi-sensors (基于多传感器的辛味中药材气一味信息融合研究) [D]. Guangzhou: Guangdong University of Technology (广东工业大学), 2014.
- [62] Liu CJ, Yang TJ, Ai L, et al. Multi-sensor information fusion based medicinal material classification system [J]. *Adv Eng Sci [四川大学学报 (工程科学版)]*, 2013, 45: 100-105.
- [63] Wu L. Classification and Identification of Chinese Herbal Medicine by Data Fusion based on Machine Olfactory and Gustatory (基于机器嗅觉/味觉中药材气一味信息融合与品鉴方法的研究) [D]. Guangzhou: Guangdong University of Technology (广东工业大学), 2012.
- [64] Wang YL, Chen PJ, Gui XJ, et al. Study on four kinds of taste classification and identification of natural medicines based on electronic tongue [J]. *Chin J Tradit Chin Med Pharm (中华中医药杂志)*, 2021, 36: 423-433.
- [65] Chen HR. Research on Bitter Almond Oil Monitoring System Based on Color Odor Digitization and Information Fusion (基于

- 颜色气味数字化及信息融合的苦杏仁走油监测系统的研究) [D]. Beijing: Beijing University of Chinese Medicine (北京中医药大学), 2019.
- [66] Xu MF. Study on the Methodology of Color Differentiation of Traditional Chinese Medicine Safflower (中药红花辨色论质方法学研究) [D]. Beijing: Beijing University of Chinese Medicine (北京中医药大学), 2016.
- [67] Zhang N, Xu B, Jia SY, et al. Modeling extraction process of *Salvia miltiorrhiza* based on multi-source information fusion technology [J]. Chin Tradit Herbal Drugs (中草药), 2018, 49: 1304-1310.
- [68] Zhang J, Xu XH, Li L, et al. Multi critical quality attributes monitoring of Chinese oral liquid extraction process with a spectral sensor fusion strategy [J]. Spectrosc Acta A Mol Biomol Spectrosc, 2022, 278: 121317.
- [69] Wang Z, Cao JJ, Li WT, et al. Using a material database and data fusion method to accelerate the process model development of high shear wet granulation [J]. Sci Rep, 2021, 11: 16514.
- [70] Zhang Q. A Study on the Dissolution Behavior Modeling Method of TCM Solid Preparations (中药固体制剂溶出行为模型化方法研究) [D]. Tianjin: Tianjin University of Chinese Medicine (天津中医药大学), 2021.
- [71] Wu SJ, Wang L, Wu HG, et al. A study on consistency evaluation of preparation of sinomenine hydrochloride based on multi-source data fusion technique [J]. J Instrum Anal (分析测试学报), 2022, 41: 121-127.
- [72] Liang J. Research on Danshen (*Salvia miltiorrhiza*) Traceability and Multivariate Calibration Method Based on Laser Induced Breakdown Spectroscopy Technology and Machine Learning Strategies (激光诱导击穿光谱技术结合机器学习策略的丹参溯源及多元校正方法研究) [D]. Xi'an: Northwest University (西北大学), 2021.
- [73] Zhang X, Wu HW, Lin LN, et al. The qualitative and quantitative assessment of Xiaochaihu granules based on e-eye, e-nose, e-tongue and chemometrics [J]. J Pharm Biomed Anal, 2021, 205: 114298.
- [74] Li YP, Wu G. Comparative analysis of data preprocessing methods based on typical data set [J]. J Shenyang Univ Technol (沈阳工业大学学报), 2022, 44: 185-192.
- [75] Lin ZZ. Research on High-dimensional Data Fusion Analysis and Evaluation Method of Traditional Chinese Medicine (中药高维数据融合分析与评价方法研究) [D]. Beijing: Beijing University of Chinese Medicine (北京中医药大学), 2015.
- [76] Han CZ, Zhu HY, Duan ZS. Multisource Information Fusion, Third Edition (多源信息融合, 第3版) [M]. Beijing: Tsinghua University Press, 2022: 1-19.