

• 数据挖掘与循证医学 •

基于机器学习的中药产地溯源研究

沈 晔¹, 王继森^{2#}, 向 晴¹, 王名亮¹, 汪 蕾³, 邓 焱¹, 陈 卓¹, 廖 婉^{1*}, 林 薇^{1*}

1. 成都中医药大学智能医学学院, 药学院, 中药种质资源创新与高效利用实验室, 中药材标准化教育部重点实验室, 四川 成都 611137

2. 成都市药品检验研究院, 国家药品监督管理局中药材质量监测评价重点实验室, 四川 成都 610045

3. 四川省骨科医院, 四川 成都 610041

摘要: 目的 针对传统中药产地溯源主观性强、效率低等问题, 探索基于中红外光谱 (mid-infrared spectroscopy, MIRS) 结合机器学习算法模型的中药产地溯源方法, 为中药产地溯源提供新的技术支撑。方法 以中红外光谱数据集为基础, 对中药材的中红外光谱数据进行异常值处理、空缺值填充、一阶差分、高方差特征筛选、距离计算与分组分析及局部线性嵌入算法 (locally linear embedding, LLE) 等预处理, 构建支持向量机 (support vector machine, SVM)、随机森林 (random forest, RF)、轻量级梯度提升机 (light gradient boosting machine, LightGBM)、K-最近邻 (K-nearest neighbor, KNN)、极端梯度提升 (extreme gradient boosting, XGBoost)、人工神经网络 (artificial neural network, ANN) 6 种机器学习模型, 并利用常青藤优化算法 (Ivy algorithm, IVYA) 优化模型参数, 构建受试者工作特征 (receiver operating characteristic, ROC) 曲线下面积 (area under curve, AUC) 与准确率、召回率、精确率、F1 分数 (F1 score, F1) 的多维度评价指标体系, 探讨适合中药产地溯源的最优模型。结果 SVM (宏平均 AUC=0.998、F1=0.949、准确率 0.949、召回率 0.949、精确率 0.956) 与 ANN (宏平均 AUC=0.999、F1=0.940、准确率 0.939、召回率 0.939, 精确率 0.944) 在中药材产地的识别上表现最优, 对未知样本的产地预测结果具有较高一致性, 且二者核心评价指标值显著高于 RF、LightGBM、KNN、XGBoost。SVM 模型可精准捕捉不同产地药材中等化学成分的官能团振动差异, ANN 模型可模拟中药产地形成的多因素非线性耦合效应, 2 种模型不仅实现了 11 个产地的高效区分, 还具有互补性, SVM 的全局最优分类面适合处理光谱特征差异显著的产地, 而 ANN 的分布式表征能力更适用于特征重叠度高的复杂场景。MIRS 结合 SVM 与 ANN 的方法, 在中药产地溯源上具有快速、无损、高通量的优势, 这为地道药材的规模化产地溯源提供了客观化工具, 为中药产业的现代化发展提供了可落地的技术路径。结论 MIRS 结合 SVM、RF、LightGBM、KNN、XGBoost、ANN 用于中药产地溯源具有有效性, 其中 SVM 与 ANN 可作为中药产地溯源的首选模型, 可为中药产地溯源提供跨学科方案, 未来可整合多模态数据及深度学习技术提升模型性能。

关键词: 中药; 产地溯源; 机器学习; 支持向量机; 人工神经网络; 常青藤优化算法; 中红外光谱技术

中图分类号: TP18; R282.23 文献标志码: A 文章编号: 0253-2670(2026)05-1798-16

DOI: 10.7501/j.issn.0253-2670.2026.05.018

Traceability study of traditional Chinese medicine origins based machine learning

SHEN Ye¹, WANG Jisen², XIANG Qing¹, WANG Mingliang¹, WANG Lei³, DENG Yao¹, CHEN Zhuo¹, LIAO Wan¹, LIN Wei¹

1. School of Intelligent Medicine, School of Pharmacy, Lab for Innovation & Effective Uses of Chinese Drug Germplasm Resources, Key Laboratory of Standardization of Chinese Medicine Ministry of Education, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

2. Chengdu Institute for Drug Control, NMPA Key Laboratory for Quality Monitoring and Evaluation of Traditional Chinese

收稿日期: 2025-10-23

基金项目: 四川省自然科学基金 (2024NSFSC2109); 国家自然科学基金面上项目 (82574868); 成都市科学技术局技术创新研发项目 (2024-YF05-01999-SN); 教育部代谢性心血管疾病医药基础研究创新中心开放课题重点项目 (xnykdxcxz-2025-05); 四川省中医药管理局面上项目 (25MSZX079)**作者简介:** 沈 晔, 医学信息工程专业。E-mail: 3209325828@qq.com

#共同第一作者: 王继森, 副主任中药师, 硕士研究生, 从事药品检验研究。E-mail: 93230623@qq.com

***通信作者:** 廖 婉, 教授, 博士, 硕士生导师, 主要从事新制剂、新剂型及中药炮制工艺与机制研究。E-mail: liaowan@cduetcm.edu.cn

林 薇, 副教授, 博士研究生, 主要从事中医药人工智能研究。E-mail: linwei@cduetcm.edu.cn

Medicine, Chengdu 610045, China

3. Sichuan Province Orthopedic Hospital, Chengdu 610041, China

Abstract: Objective To address the issues of strong subjectivity and low efficiency in traditional Chinese medicine origin tracing, we explore the method of Chinese medicine origin tracing based on mid-infrared spectroscopy (MIRS) combined with machine learning algorithm modeling, and provide new technical support for traditional Chinese medicine origin tracing. **Methods** Based on the mid-infrared spectral dataset, this study performed outlier handling, missing value imputation, first-order difference, high-variance feature selection, distance calculation and clustering analysis, and locally linear embedding (LLE) for dimensionality reduction. Subsequently, we constructed six machine learning models including Support Vector Machine (SVM), Random Forest (RF), Light Gradient Boosting Machine (LightGBM), K-nearest neighbor (KNN), extreme gradient boosting (XGBoost), and artificial neural network (ANN) models. We then employed the Ivy Algorithm (IVYA) to optimize model parameters. A multidimensional evaluation framework was established, incorporating the area under the receiver operating characteristic (ROC) curve (AUC) alongside accuracy, recall, precision, and F1 score (F1). This framework was used to identify the optimal model for Chinese herbal medicine origin tracing. **Results** SVM (macro-average AUC = 0.998, F1 = 0.949, accuracy = 0.949, recall = 0.949, precision = 0.956) and ANN (macro-average AUC = 0.999, F1 = 0.940, accuracy = 0.939, recall = 0.939, precision = 0.944) demonstrated optimal performance in identifying the origins of Chinese herbal medicines. They exhibited high consistency in predicting the origins of unknown samples, with core evaluation metrics significantly outperforming RF, LightGBM, KNN, and XGBoost. The SVM model can accurately capture the functional group vibration differences in the medium chemical components of medicinal materials from different origins. ANN models simulate the multifactorial nonlinear coupling effects formed by Chinese herbal medicine origins. Both models not only achieve efficient differentiation among 11 origins but also exhibit complementary strengths: SVM's globally optimal classification hyperplane excels at processing origins with markedly distinct spectral features, while ANN's distributed representation capability better suits complex scenarios with high feature overlap. The MIRS method, integrating SVM and ANN, offers rapid, non-destructive, and high-throughput advantages for Chinese herbal medicine traceability. This provides an objective tool for large-scale origin tracing of authentic medicinal materials and delivers a practical technical pathway for the modernization of the Chinese herbal medicine industry. **Conclusion** MIRS combined with SVM, RF, LightGBM, KNN, XGBoost, and ANN demonstrates effectiveness for Chinese herbal medicine origin tracing. Among these, SVM and ANN emerge as preferred models for this application, offering an interdisciplinary solution. Future integration of multimodal data and deep learning techniques holds promise for enhancing model performance.

Key words: traditional Chinese medicine; origin traceability; machine learning; support vector machine; artificial neural network; Ivy algorithm; mid-infrared spectroscopic technique

道地药材是我国传统优质药材的代表，是经过中医临床长期应用优选出来的、产在特定地域、具有较高知名度的药材，与其他地区所产同种中药材相比，品质和疗效更好、质量稳定^[1]。中药材的道地性以产地为主要指标，因此产地溯源对药材品质把控至关重要。然而，当前中药市场原料混杂，传统的产地溯源方法，如形态学观察、理化分析等存在主观性强、效率低、难以精准区分近缘物种和复杂成分差异等问题，难以满足现代中药产业对精准溯源的需求。

机器学习作为人工智能的核心技术，通过构建算法模型实现从数据中自动提取知识的能力，其核心算法根据数据是否存在人工标记又可分为监督学习和无监督学习^[2]，在模式识别和数据分析方面展现出强大能力。中红外光谱技术（mid-infrared spectroscopy, MIRS）技术凭借快速、无损、环保且能捕捉丰富化学指纹信息的优势^[3]，已初步应用于

中药成分分析领域。将中红外光谱技术与机器学习深度融合，以中红外光谱技术获取的丰富化学指纹信息作为机器学习算法的数据基础，使得算法能够学习到不同产地中药的特征模式，机器学习算法则可以挖掘出这些复杂的光谱数据中隐藏的与产地相关的信息，从而实现对中药产地的精准溯源，为中药产地溯源提供了新的研究方向和思路。

本研究构建了多模型协同与智能优化相结合的技术框架：其一，将常青藤优化算法（Ivy algorithm, IVYA）引入中药产地溯源模型参数优化，通过模拟植物向光性生长的智能搜索机制实现对支持向量机（support vector machine, SVM）、随机森林（random forest, RF）、轻量级梯度提升机（light gradient boosting machine, LightGBM）、K-最近邻分类（K-nearest neighbor, KNN）、极端梯度提升（extreme gradient boosting, XGBoost）、人工神经网络

络 (artificial neural network, ANN) 6 种算法超参数的全局高效寻优, 解决了传统参数调优过程中局部最优陷阱的问题; 其二, 结合局部线性嵌入算法 (locally linear embedding, LLE) 进行非线性降维, 在保留中药材光谱关键几何特征的同时降低维度灾难影响, 为高维光谱数据的有效利用提供新途径; 其三, 建立了受试者工作特征 (receiver operating characteristic, ROC) 曲线下面积 (area under curve, AUC)、准确率、召回率、精准率、F1 分数 (F1 score, F1) 的多维度评价体系, 系统对比不同模型在中药产地溯源中的性能差异。

1 数据来源与方法

1.1 MIRS 数据来源

MIRS 数据源自数学建模大赛官网 (<https://www.mcm.edu.cn>) 所提供的中药材中红外光谱数据集。数学建模大赛由教育部高等教育司指导、中国工业与应用数学学会主办, 是国内影响力最大、参与范围最广的大学生学科竞赛之一。该数据集作为该竞赛的指定数据, 经过多轮专家论证, 体现了科学性与严谨性, 具有较高的权威性和规范性, 为分析中药产地与光谱特征的关联性提供了可靠基础。

该数据集包含 673 条同种类别的 11 个产地的中药材的 MIRS 数据, 光谱采集范围为 $3\ 998\sim 551\text{ cm}^{-1}$, 分辨率为 1 cm^{-1} 。该光谱波数范围包含众多不同化学键和官能团的振动吸收信息, 可捕捉反映中药材中有机分子的官能团特征振动信息, 其中, 在较低波数区域可能包含一些重原子之间的振动

信息, 而较高波数区域则可能与一些含氢官能团的伸缩振动相关。同时, 不同产地的同一种药材可能由于生长环境、土壤条件等因素的不同, 其化学成分的含量和比例会有所差异, 从而导致在中红外光谱上的吸光度表现不同。因此, 机器学习算法可以利用 MIRS 数据所具有的化学指纹特性挖掘出这些复杂光谱数据中隐藏的与产地相关的信息。

1.2 数据预处理

中药材的中红外光谱信号易受生物碱、黄酮类、挥发油等自身复杂成分及仪器噪声干扰, 且原始光谱中可能包含与产地无关的共同成分信息, 或因样本采集过程中炮制工艺差异、仪器波动产生的异常信号, 这会直接影响产地特征的提取精度, 从而影响模型的稳定性和可靠性^[4-5]。因此, 在进行模型建立之前, 需结合中药材的光谱特性对 MIRS 进行针对性预处理, 以消除干扰信息的影响提升数据质量^[6]。预处理流程见图 1。

1.2.1 目标变量处理 样本的 MIRS 原始数据共包含 673 条同种类中药材样本, 涵盖 11 个不同产地, 而目标变量原产地 (origin place, OP) 正是表征这些中药材地理来源的分类变量, 其取值直接与实际生长产地一一对应, 该变量取值范围为 1~11, 代表 11 个不同产地, 是后续模型训练与产地预测的“参照标签”。同时, 目标变量 OP 具有类别唯一性与差异区分性, 唯一性即同一产地的中药材因生长环境、栽培方式等高度一致, 其化学成分与光谱特征呈现显著同质性、对应的 OP 值唯一; 差异区分

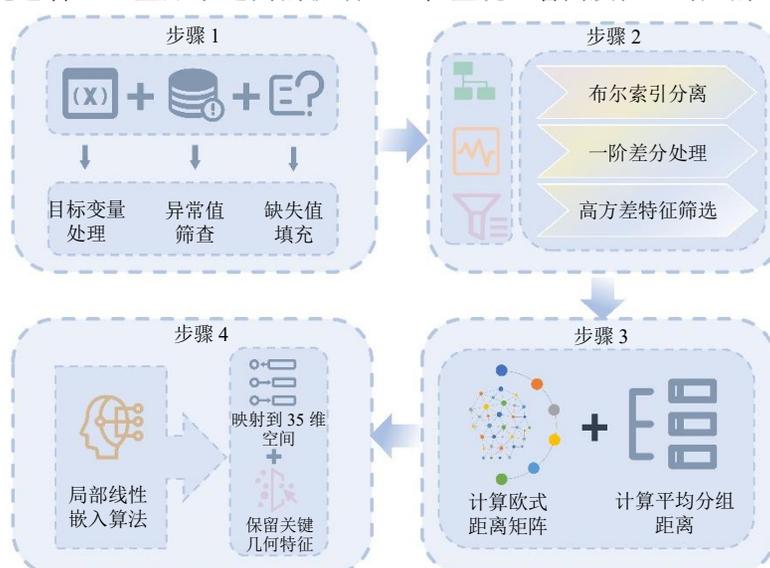


图 1 MIRS 数据预处理流程

Fig. 1 MIRS data preprocessing process

性即不同产地的同一种药材因环境因子、化学成分比例等差异,在MIRS吸光度上表现为特征性信号不同,所对应的OP值也不同。因此,OP与产地具有强关联性,对药材产地划分的错误标记会直接影响产地溯源准确性,如在SVM预训练中,将2条OP=1的样本误标为OP=2,导致SVM出现了OP=12等偏离1~11合理范围的标记,其对OP=1样本的召回率也随之下落,所以需对OP值进行异常值筛查,剔除明显偏离合理范围的异常标记,以避免错误标签对模型训练的干扰。对于目标变量OP存在的少量缺失值采用直接填充法将缺失值统一赋值为0,并将此类样本定义为“未知产地样本”,其余OP值非0的样本为“已知产地样本”,通过分类标记实现2类样本的分离,为后续产地溯源预测提供判别依据。

1.2.2 基于中红外光谱数据的特征工程 原始MIRS数据曲线(图2-A)显示,在500~4000 cm^{-1}

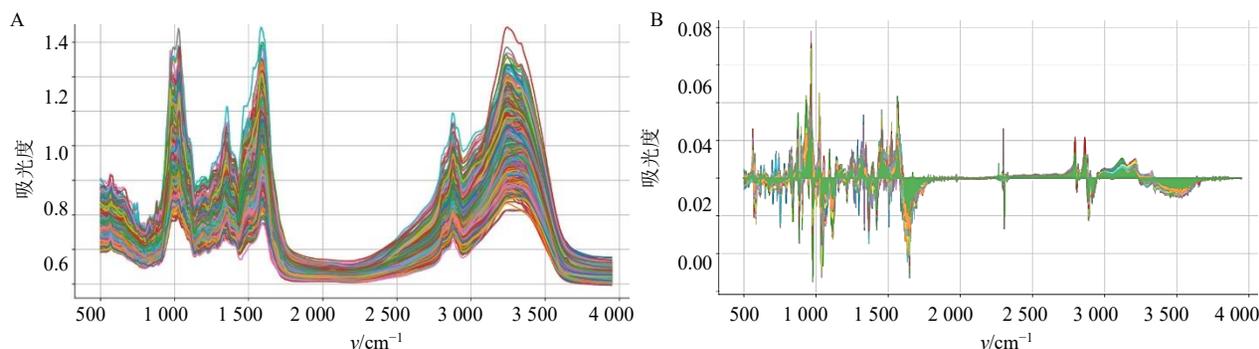


图2 原始MIRS数据曲线(A)和一阶差分处理后的MIRS数据曲线(B)
Fig. 2 Curves of original MIRS data (A) and first-order difference-processed MIRS data (B)

1.2.3 距离计算与分组分析 距离计算与分组分析是为量化样本间的相似性与差异性,从几何度量角度揭示数据结构特征^[7]。因此,采用欧几里得距离构建基于特征工程处理后的MIRS数据中未知样本与已知样本的距离矩阵,其核心作用是客观量化不同产地中药材的光谱特征差异,揭示不同产地中药材的化学组成关联,减少人工判别的主观性误差。该方法相较于曼哈顿距离、余弦距离等所构建的距离矩阵更具独特优势:其一,欧几里得距离对连续型光谱特征的差异捕捉精度更高。中药材MIRS数据的核心信息载体是连续型吸光度数值,不同产地的差异本质是特定波数吸光度的绝对数值差异,欧几里得距离通过计算特征空间中样本点的欧式距离,可直接量化这种绝

对差异。相比之下,曼哈顿距离侧重坐标轴方向的累加差异,对连续型吸光度的细微差异敏感性不足;余弦距离仅衡量样本间的方向相似性,会忽略特征数值的绝对大小,因此均无法满足“精准区分产地细微成分差异”的需求。其二,欧几里得距离与中药产地溯源的“群体相似性”逻辑契合。中药材产地特征具有“群体相似性”,同一产地样本因生长环境一致,光谱特征呈现聚集分布,不同产地样本形成独立簇群。欧几里得距离计算的分组平均距离可直接反映这种群体聚类特性,而曼哈顿距离在群体平均距离计算中,易受个别极端样本影响,导致群体距离偏差,余弦距离则因无法区分“同趋势但不同强度”的群体光谱,导致群体归属误判率升高。

内,各样本曲线整体趋势具有一定相似性,部分波段出现重叠或趋近现象,不同样本特征差异难以清晰辨别。为放大此类差异,增强模型对产地相关的特征判别能力,提高中药产地溯源的准确性,采取一阶差分对原始MIRS进行处理,即将中药材连续波数的吸光度差值转化为数值特征,强化不同产地药材在特征性成分吸收峰处的斜率变化;然后筛选高方差特征,即通过计算一阶差分后各特征的极差,筛选出极差大于均值的高方差特征,从而剔除中药材中普遍存在的共有成分的光谱信号,聚焦于与产地相关的特异性成分信号,减少冗余信息对模型的干扰。

由一阶差分处理后的MIRS数据曲线(图2-B)可见,光谱曲线波动程度显著加剧,相邻波段吸光度变化转化为具体数值特征,有效放大了光谱局部形态差异。图中曲线起伏变化更明显,不同样本曲线在多波段处展现出更清晰差异。

该方法以几何距离度量样本相似性将特征工程处理后的 MIRS 数据通过公式 (1) 量化为特征空间中的几何距离, 其中, d_{ij} 表示未知样本 (u_{ik}) 与已知样本 (k_{jk}) 之间的欧几里得距离, 距离值越小表明样本特征相似度高, 样本在特征成分组成及含量上越接近, 产地归属的可能性越高。通过公式 (2) 计算分组平均距离, 即计算未知样本与某一产地类别的平均距离, 从而形成多维特征向量, 其中, 未知样本与该类别中各样本的距离分别为 $d_{i1}^l, d_{i2}^l, \dots, d_{in_1}^l$, 未知样本与类别 C_l 的平均距离为 \bar{d}_i^l , 未知样本与某类别组的平均距离越小, 其归属于该产地类别的可能性越高。通过公式 (3) 计算组内 (同产地内) 与组间 (不同产地间) 距离矩阵, 从类别结构角度揭示数据分布特征。其中, 类别 C_s 与 C_t 的质心分别为 u_s 和 u_t , 组间距离为 d_{st} 。组内距离矩阵值越小表明类别内样本同质性越高, 同一产地样本的成分稳定性越高; 组间距离矩阵通过类别质心距离度量类别分离度, 值越大说明类别间不

同产地的特征成分差异越显著。该分析尤其适用于“近缘产区”的区分。例如, 当 2 个产区地理邻近但微环境存在差异时, 其药材的光谱特征差异可能细微, 通过距离矩阵可精准捕捉这种差异, 2 类矩阵的结合使用可有效评估数据的类别区分度, 为产地溯源提供量化的特征筛选依据。

$$d_{ij} = \sqrt{\sum_{k=1}^p (u_{ik} - k_{jk})^2} \tag{1}$$

$$\bar{d}_i^l = \frac{1}{n_l} \sum_{j=1}^{n_l} d_{ij}^l \tag{2}$$

$$d_{st} = \sqrt{\sum_{k=1}^p (\mu_{sk} - \mu_{tk})^2} \tag{3}$$

根据特征工程处理后的 MIRS 数据中未知样本与各类别组的平均距离生成的雷达图 (图 3-A) 与柱状图 (图 3-B), 可见未知样本 1~14 与各已知类别组的平均距离均存在显著不同, 这反映了未知样本与已知类别组之间的相似性程度, 为后续模型判断未知样本的潜在归属提供了多维视角的依据。

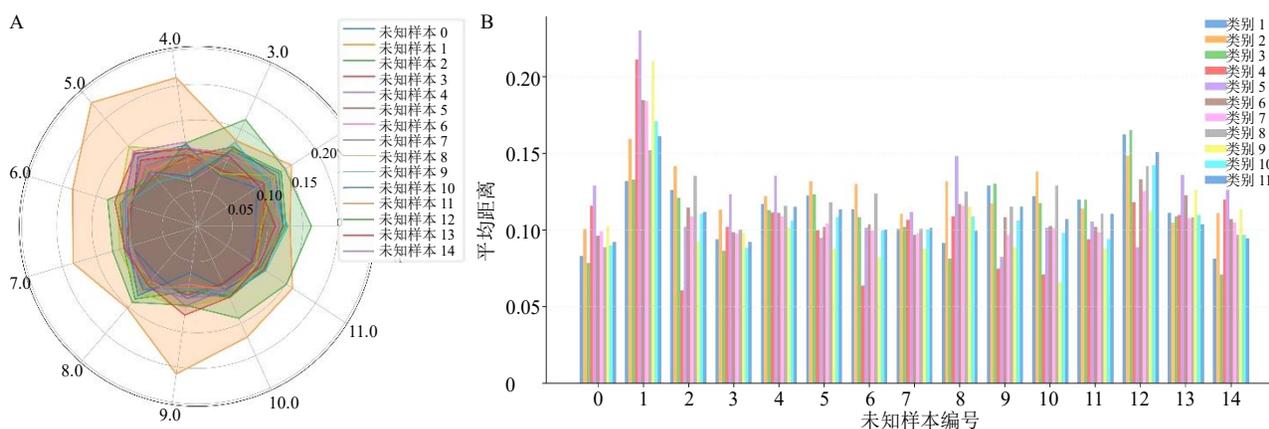


图 3 未知样本与已知类别组的平均距离雷达图 (A) 和柱状图 (B)

Fig. 3 Radar plot (A) and histogram (B) of mean distances between unknown samples and known category groups

1.2.4 基于 LLE 算法的降维处理 LLE 算法是一种非线性降维算法, 其能够使降维后的数据较好地保持原有流行结构^[8]。其核心思想是通过构建样本的 k 近邻局部线性关系, 在低维空间中保持数据的局部几何结构。具体而言, LLE 首先为每个样本点确定 k 个最近邻点, 然后计算该点由这些近邻点线性重构的权重, 使得重构误差最小化; 最后, 利用这些权重在低维空间中重构样本点, 从而在降维过程中保留数据的局部几何特性^[9]。

中药材的 MIRS 包含 $3\ 998 \sim 551\ \text{cm}^{-1}$ 的 3 448 个波数点, 高维数据易导致模型过拟合, 降低产地

溯源准确率。因此, 采用 LLE 算法对 MIRS 进行降维处理, 从而解决高维度灾难问题, 即剔除光谱数据中的冗余特征, 减少模型训练的计算复杂度, 提升训练效率; 同时保留关键特征, 即通过 LLE 算法的局部线性重构特性, 在降维过程中保留不同产地药材光谱的局部几何结构, 确保降维后的数据仍能反映产地间的特征差异。

LLE 算法的关键参数包括降维后目标维度 ($n_components$)、近邻数量 ($n_neighbors$) 及算法实现方式 ($method$)。其中, $method$ 参数通过优化局部权重计算逻辑解决了高维光谱数据中易出现的

矩阵奇异性问题，有效提升了降维过程的稳定性与结果可靠性。对于 $n_components$ 与 $n_neighbors$ 这 2 个核心可调参数，首先结合数据规模，将 $n_components$ 初步限定为 20~50、 $n_neighbors$ 限定为 50~100。经调试发现，当 $n_components < 30$ 时会丢失关键光谱特征，当 $n_components < 40$ 则会引入冗余并增加计算量；当 $n_neighbors < 60$ 易受噪声干扰，当 $n_neighbors > 80$ 会破坏类别边界。最终，确定最优参数为 $n_components = 35$ 、 $n_neighbors = 70$ ，该参数组合在平衡计算复杂度与保留不同产地中药的关键几何特征的同时，能有效避免“维度灾

难”，为后续模型训练提供高效输入。

在对 MIRS 进行 LLE 降维处理后，35 维特征作为 SVM、ANN 等 6 种模型的输入，降维剔除的冗余特征将不再干扰模型学习，后续模型可直接聚焦于产地相关的关键特征，各模型的分类精度与泛化能力将得到提升。根据降维后的 MIRS 中已知样本分布生成的散点图（图 4），已知样本在 35 维空间中呈现显著的聚类趋势，部分类别的样本紧密聚集，产地类别 1、3、4、5、6、7 的样本在低维空间中形成独立簇群，与其他类别界限清晰，表明 LLE 算法成功保留了用于产地区分的关键几何特征。

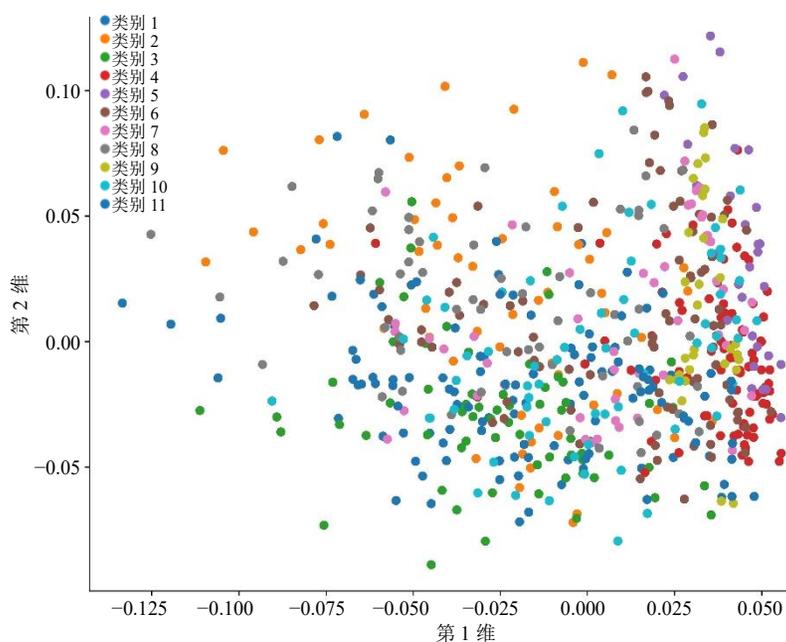


图 4 降维后已知样本分布散点图

Fig. 4 Scatter plot of known sample distribution after dimensionality reduction

1.3 基于机器学习建立中药产地溯源模型

1.3.1 模型选择依据 SVM 是一种基于统计学习理论的监督式分类算法，通过径向基核函数（radial basis function, RBF）将中药材的 MIRS 数据映射到高维空间，构建最大化分类超平面的间隔来实现最优分类边界。SVM 能有效捕捉中药材产地差异对应的非线性光谱特征，解决传统溯源中化学成分细微变化导致的边界模糊问题，兼具高维空间适应性与强抗噪声能力，已成功用于黄连饮片识别及枸杞子产地快速鉴别^[10]。

RF 是一种基于集成学习的监督式分类算法，通过构建多棵决策树并利用“多数投票”机制整合预测结果，以提升模型的泛化能力与分类稳定性^[11]。RF

通过自助采样和随机特征选择，可有效处理中药材光谱中因多成分叠加导致的复杂特征，实现对中药产地复杂特征的稳健识别，已被用于附子产地溯源，准确率达 85%^[12]。

LightGBM 是一种基于梯度提升框架的高效机器学习算法，通过迭代构建弱学习器（决策树）最小化损失函数，特别适用于高维、非线性数据场景^[13]。采用直方图算法离散化连续特征，在压缩维度的同时保留关键特征分布，通过 Leaf-wise 策略优化树结构，兼顾训练速度与微量成分信号保留，已用于金线莲品系识别，为视觉数据驱动的产地溯源提供技术范式^[14]。

KNN 是经典实例驱动型机器学习方法，基于

样本相似性度量与“近邻投票”实现未知样本分类。同一产地的中药材因遗传背景和生长环境一致，其光谱特征呈现聚集分布，而不同产地的样本群体在特征空间中会形成明显聚类，KNN 算法直接通过计算 MIRS 中未知样本与已知样本的欧氏距离确定近邻，适用于特征空间中类别边界清晰、样本分布相对均衡的分类场景^[15-16]，已应用于结合近红外光谱构建杜仲产地识别模型^[17]。

XGBoost 是一种高效的梯度提升框架，通过迭代训练决策树弱学习器，不断优化目标函数，适用于高维、非线性数据的分类与回归任务。该算法通过引入正则化项控制模型复杂度，可避免过度拟合中药材光谱中的噪声，同时利用二阶导数信息加速收敛可高效捕捉特征间的高阶交互，该算法已应用于黄芪饮片干燥过程中水分含量在线检测^[18]。

ANN 是一种受生物神经系统启发的机器学习模型，其核心思想是通过多层神经元的连接与权重调整，实现对复杂非线性数据模式的建模^[19-20]。ANN 的多层感知机 (multi-layer perceptron, MLP) 结构可通过反向传播算法自动学习高阶特征，模拟中药产地形成的多因素交互机制，适合处理近缘产地的溯源问题。该算法具备深度特征提取、非线性映射能力、泛化能力强的优点，已用于基于气味信息和活性成分对三七进行产地溯源研究，其产地的判别准确率达 87% 以上^[21]。

1.3.2 模型建立、优化与评价 MIRS 图谱采用 Pycharm 2023.2.5 的 matplotlib 库绘制。一阶差分、高方差特征筛选，距离计算与分组分析，LLE 降维等数据预处理均采用 Pycharm 2023.2.5 代码实现。

将数据预处理后的 MIRS 数据按 7:3 的划分方式随机分为训练集 (471 个样本) 和测试集 (202 个样本)，利用 SVM、RF、LightGBM、KNN、XGBoost、ANN 6 种机器学习算法进行建模，采用 IVYA 算法对各模型的超参数进行优化，模拟常青藤向光性生长机制，实现对复杂参数空间的高效搜索^[22]。最后，构建 AUC、准确率、召回率、精准率、F1 分数 (F1 score, F1) 的多维度评价指标体系，通过比较不同算法在产地溯源上的 F1 分数与宏平均 AUC，来探讨适合中药产地溯源的最优模型。

2 结果与分析

2.1 基于 SVM 算法的产地溯源

在 SVM 的产地溯源中，引入 IVYA 对惩罚参数 C 和核参数 γ 进行全局寻优，以提升模型对不

同产地中药材光谱特征的泛化能力，并最终确定最优参数组合为 $C=32.81$ 、 $\gamma=9.13$ 。基于训练集与测试集的混淆矩阵 (图 5-A、B)，计算得到训练集的精确率、召回率和 F1 分数均达到 1.000，测试集的整体准确率为 0.949、精确率 0.956、召回率 0.949、F1 分数 0.949。绘制 ROC 曲线 (图 5-C) 可见，宏平均 $AUC=0.998$ ，其中，多数类别 (类别 1、2、4、5、6、7、8、9、11) 的曲线紧密贴近左上角， $AUC=1.00$ ，表明 SVM 对这些产地中药材的特征性成分差异具有近乎完美的区分能力；仅少数 AUC 为 0.99 的类别存在误判，可能与个别样本因生长年限差异导致的成分波动有关。绘制测试集真实值与预测值对比图 (图 5-D) 进一步直观展示了模型分类效果，多数样本点真实值与预测值吻合，仅少数集中在类别 9、10、11 的点偏离，这可能与这些产地中药材的次生代谢产物谱差异较小，导致光谱特征重叠有关。

2.2 基于 RF 算法的产地溯源

RF 算法通过 IVYA 对参数决策树数量 ($n_estimators$) 和最大深度 (max_depth) 进行寻优，最终确定最优参数组合为 $n_estimators=100$ 、 $max_depth=10$ 。结合混淆矩阵 (图 6-A、B) 分析，RF 算法在训练集上准确率、精确率、召回率、F1 分数均达到 1.000，在测试集上 RF 算法整体准确率为 0.934、精确率 0.948、召回率 0.934、F1 分数 0.936。ROC 曲线如图 6-C 所示，宏平均 $AUC=0.996$ ，其中，类别 1、2、3、4、5、7、9 的曲线紧贴左上角 ($AUC=1.00$)，说明这些产地的中药材因特征成分差异显著，模型可完美区分；类别 6、8、10、11 的曲线高度接近左上角 ($AUC=0.99$)，略低于其他类别，可能与这些产地的中药材存在交叉污染或成分趋同现象有关。测试集真实值与预测值对比图如图 6-D 可见，真实值与预测值曲线整体趋势吻合，仅在类别 6、8、10 等相邻区间出现少量集中误判，这可能与决策树分裂时对重叠特征的阈值划分偏差有关。

2.3 基于 LightGBM 算法的产地溯源

LightGBM 利用 IVYA 对 LightGBM 的核心参数进行调优，最终确定最优参数组合为 $n_estimators=50$ 、 $max_depth=-1$ 、 $learning_rate=0.2$ 。结合混淆矩阵 (图 7-A、B) 分析，LightGBM 在训练集上的准确率、精确率、召回率、F1 分数均为 1.000，测试集上的准确率达 0.914、精确率 0.924、召回率

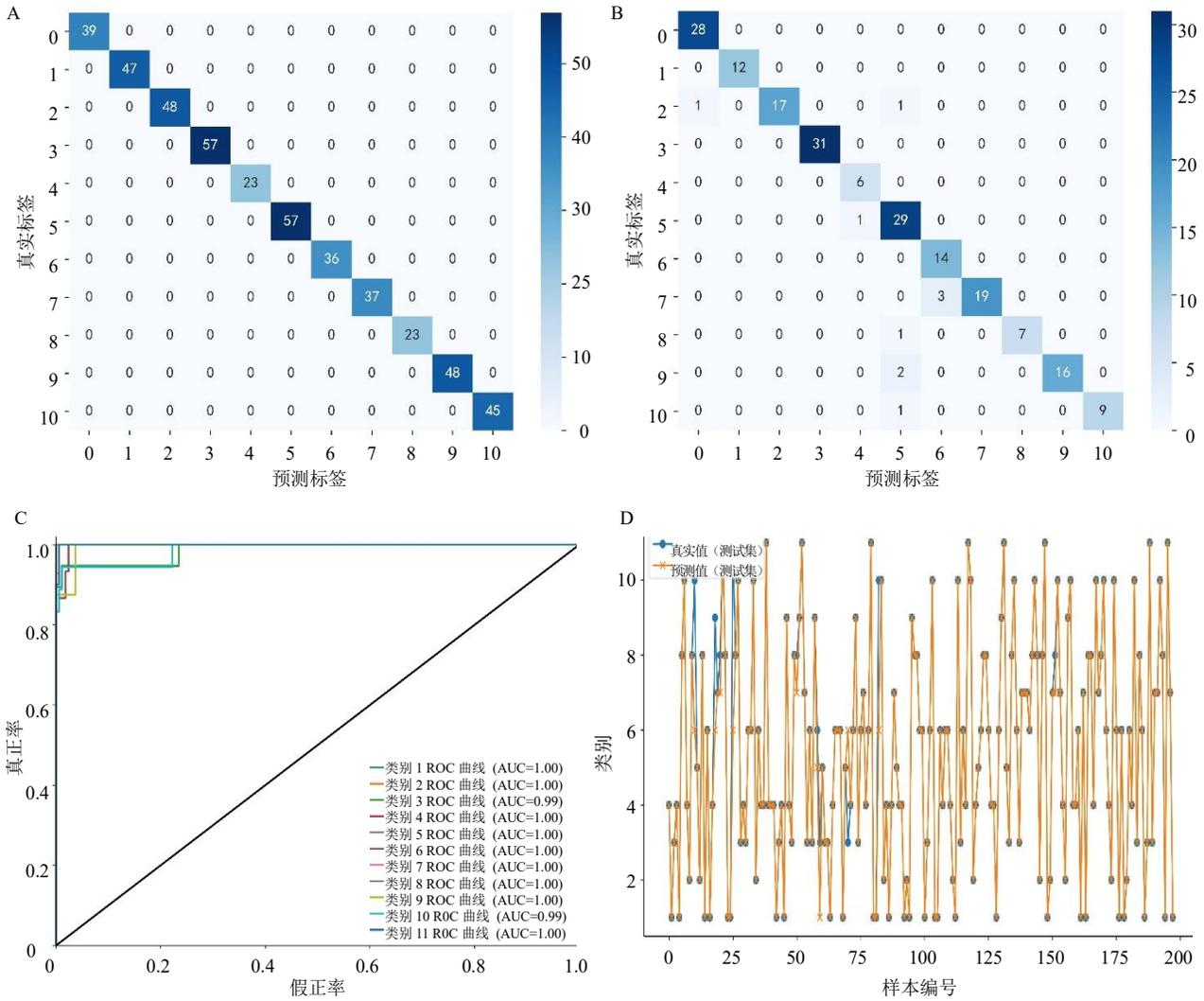


图 5 SVM 的训练集混淆矩阵 (A)、测试集混淆矩阵 (B)、ROC 曲线 (C)、真实值与预测值对比 (D)

Fig. 5 Confusion matrix of training set (A), confusion matrix of test set (B), ROC curve (C), comparison between true values and predicted values (D) for SVM

0.914、F1 分数 0.915。模型在类别 1、3、4、5、6 等光谱特征差异明显的组别中表现优异，召回率均为 1.000，显示出对典型类别样本的强捕捉能力；但在类别 2、7、8 等光谱信号重叠或样本量较少的组别中存在一定误判，主要源于直方图离散化处理导致的特征细节损失及梯度提升对类别不平衡的敏感性。ROC 曲线分析如图 7-C 所示，宏平均 AUC=0.995，类别 1、2、4、5、6、7 的曲线紧贴左上角 (AUC=1.00)，表明对这些类别的完美区分能力；类别 3、8、9、10、11 的曲线接近左上角 (AUC=0.99)，在少数重叠类别中存在轻微性能下降。由测试集真实值与预测值对比图 (图 7-D) 可见，LightGBM 算法真实值与预测值曲线整体趋势吻合，但在样本编号 25、50、115 等位置附近，真实

值和预测值差异明显，说明模型在某些数据特征上的学习还不够充分、预测稳定性不足，这可能源于直方图离散化处理导致的特征细节损失及梯度提升对类别不平衡的敏感性。

2.4 基于 KNN 算法的产地溯源

KNN 直接利用光谱数据的空间分布特性，为未知样本提供基于相似性的产地归属判断，通过 IVYA 确定 KNN 算法的最优参数组合为 n_neighbors=5, weights : distance。训练集混淆矩阵如图 8-A 所示，对角线上元素占据全部样本，KNN 算法展现出极高的拟合精度，准确率、精确率、召回率及 F1 分数均达 1.000。测试集准确率为 0.893、精确率 0.906、召回率 0.894、F1 分数 0.895。由测试集混淆矩阵 (图 8-B) 可见，多数类别的预

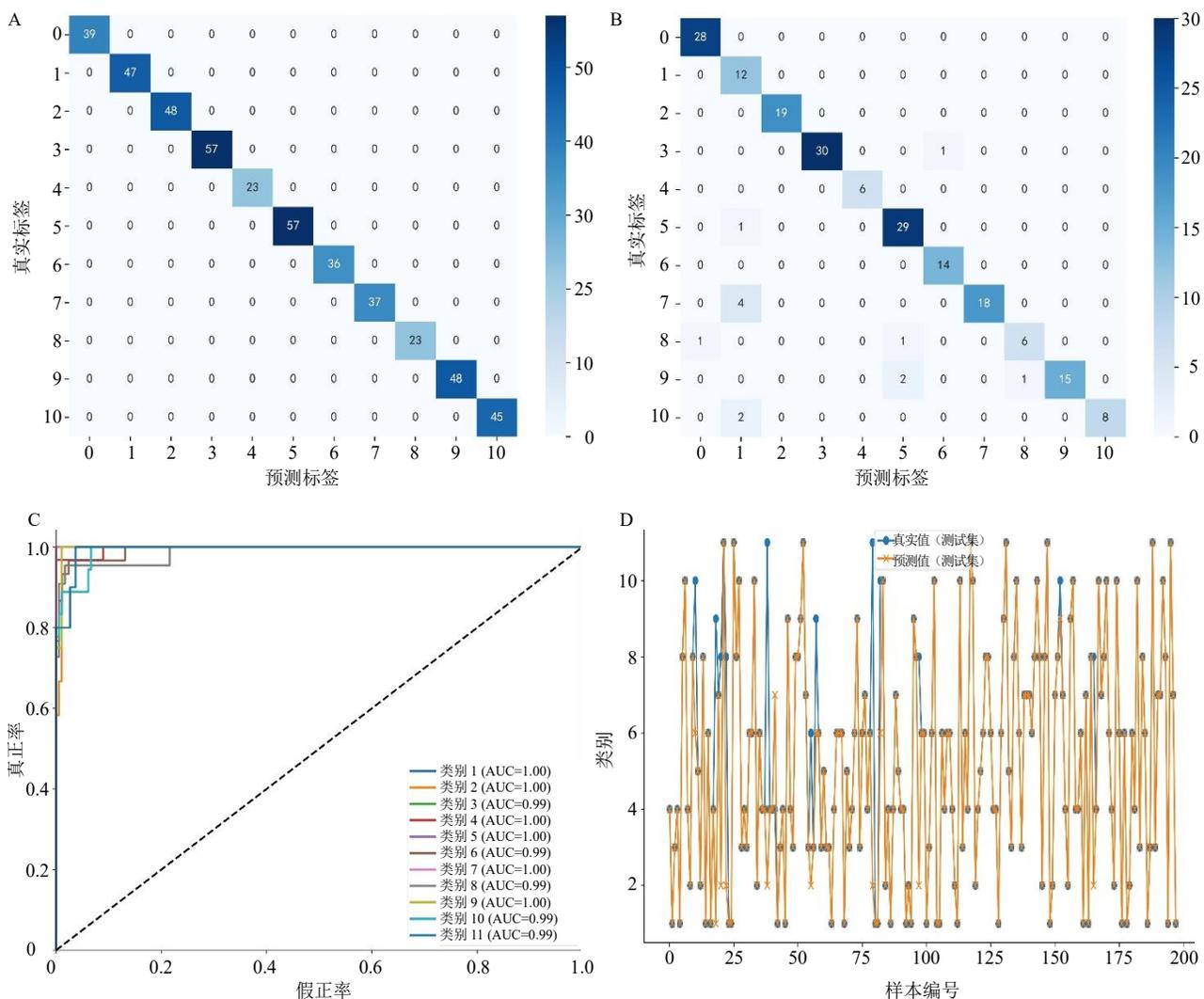


图 6 RF 的训练集混淆矩阵 (A)、测试集混淆矩阵 (B)、ROC 曲线 (C)、真实值与预测值对比 (D)

Fig. 6 Confusion matrix of training set (A), confusion matrix of test set (B), ROC curve (C), comparison between true values and predicted values (D) for RF

测准确率高于 0.900，但部分类别存在少量误判。测试集 ROC 曲线如图 8-C 所示，宏平均 AUC=0.989，KNN 算法在类别 1、2、3、4、5、7、8、11 上的 AUC 均为 1.00，表明这些产地的中药材群体特征鲜明；类别 6 (AUC=0.98)、10 (AUC=0.97) 的曲线接近左上角，仅少数样本的类别边界处出现轻微混淆；而类别 9 的 AUC (0.93) 略低，可能与该产地样本存在明显的个体差异，局部密度不均有关。通过测试集真实值与预测值对比图 (图 8-D) 可见，整体趋势吻合度较高，但在样本编号 25、50 等位置存在显著差异。此类偏差可能源于 2 方面原因：一是 KNN 算法对“k 值”和距离权重的敏感性，当边界样本的近邻类别出现歧义时易导致误判；二是高维特征空间中局部相似性与全局分布的不一致性，

使得少数样本的特征向量虽距离较近，但实际类别归属存在差异。

2.5 基于 XGBoost 算法的产地溯源

引入 IVYA 对 XGBoost 的核心参数进行调优，最终确定最优参数组合为 $n_estimators = 100$ ， $max_depth = 3$ ， $learning_rate = 0.2$ 。基于训练集与测试集的混淆矩阵如图 9-A、B 所示，计算得到 XGBoost 算法在训练集上的准确率、精确率、召回率及 F1 分数均达 1.000，在测试集上的准确率为 0.904、精确率 0.921、召回率 0.904、F1 分数 0.908。ROC 曲线如图 9-C 可知，宏平均 AUC=0.992，类别 1、2、5、6、7 的 AUC 均为 1.00，表明对这些产地的中药材具有完美区分能力；类别 3、4、8、9、10、11 的 AUC 介于 0.98~0.99，曲线接近左上角，

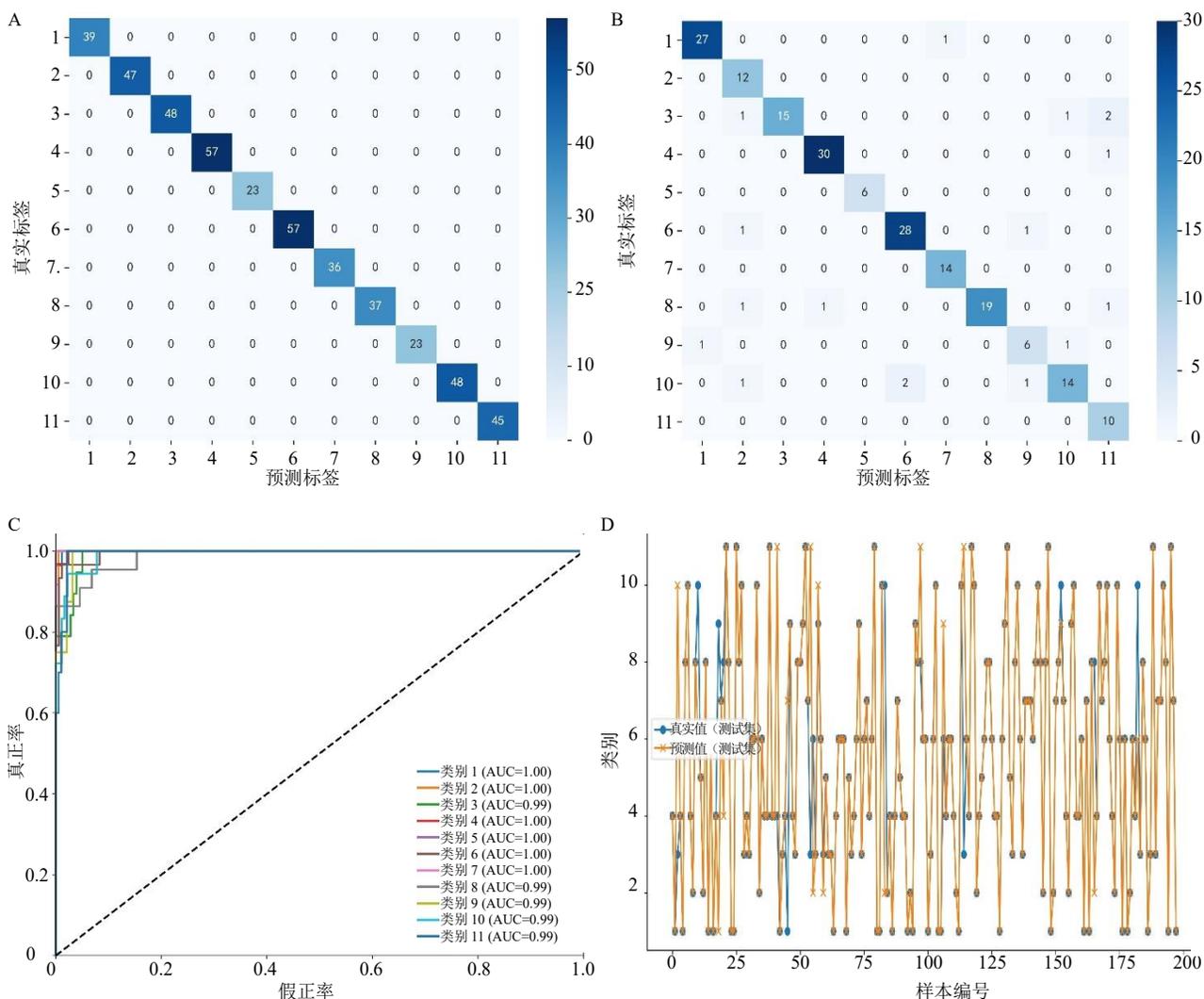


图 7 LightGBM 的训练集混淆矩阵 (A)、测试集混淆矩阵 (B)、ROC 曲线 (C)、真实值与预测值对比 (D)
 Fig. 7 Confusion matrix of training set (A), confusion matrix of test set (B), ROC curve (C), comparison between true values and predicted values (D) for LightGBM

仅在少数样本的类别边界处出现误判。通过测试集真实值与预测值对比图 (图 9-D) 可见, 整体趋势吻合度较高, 但在样本编号 25、50、75、115 附近存在显著差异, 此类偏差可能源于 2 方面原因: 一是 XGBoost 在处理极端不平衡类别时, 正则化项对少数类的关注度不足, 导致其特征未被充分学习; 二是有限的树深度 ($\text{max_depth}=3$) 可能导致复杂特征交互的学习不充分, 使得局部特征组合的判别能力受限。

2.6 基于 ANN 算法的产地溯源

引入 IVYA 对 ANN 的学习率、隐藏层神经元数量及批次大小等关键参数进行全局寻优, 最终确定最优参数组合为学习率=0.005 025、隐藏层神经元数量=191、批次大小=53, 基于混淆矩阵 (图 10-

A、B), 计算得到 ANN 模型在训练集上的准确率、精确率、召回率及 F1 分数均达 1.000, 在测试集上的准确率为 0.939、精确率 0.944、召回率 0.939、F1 分数 0.940。ROC 曲线 (图 10-C) 显示, 宏平均 $\text{AUC}=0.999$, 除类别 10 的 AUC 为 0.99 外, 其余类别的 AUC 均为 1.00, 曲线紧密贴近左上角, 表明模型对绝大多数类别的区分能力近乎完美, 仅在类别 10 的少数样本中出现轻微误判, 这可能与该产地中药材存在罕见的代谢物变异体有关。通过测试集真实值与预测值对比图 (图 10-D) 可见, 整体趋势吻合度较高, 但在类别 8~10 存在显著偏离, 此类偏差可能源于 2 方面原因: 一是 ANN 模型在训练过程中对隐藏层神经元数量和学习率的敏感性, 导致局部最优解的出现; 二是高维特征空间中

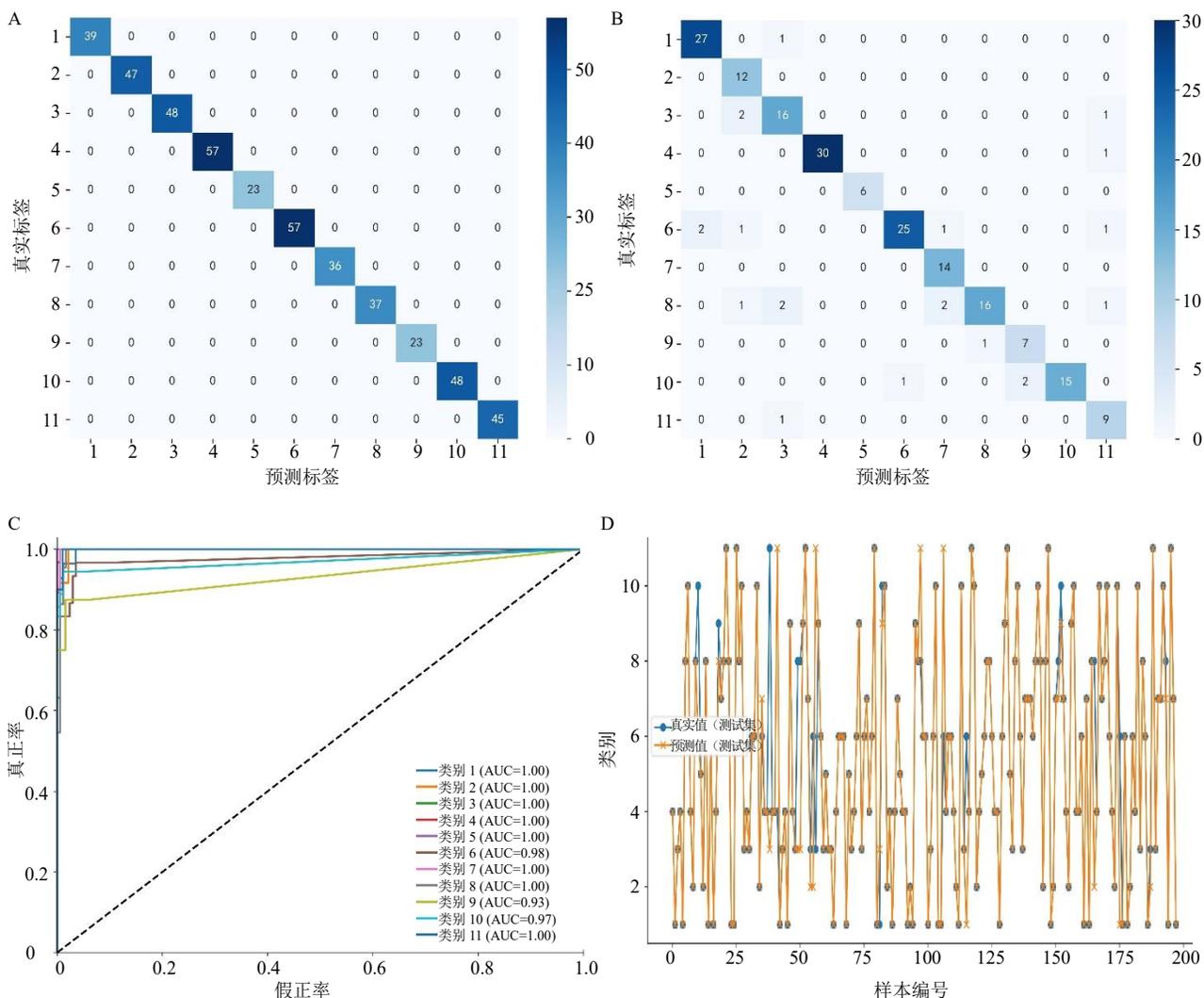


图 8 KNN 的训练集混淆矩阵 (A)、测试集混淆矩阵 (B)、ROC 曲线 (C)、真实值与预测值对比 (D)

Fig. 8 Confusion matrix of training set (A), confusion matrix of test set (B), ROC curve (C), comparison between true values and predicted values (D) for KNN

复杂的非线性交互关系未被完全捕获，使得部分边界样本的预测结果出现偏差。

2.7 结果对比

如表 2 所示，15 个未知样本经 SVM、RF、LightGBM、KNN、XGBoost 及 ANN 模型预测后，多数样本的预测结果具有较高一致性。样本 3、14、38、79、110、134、152、227、618 在 6 种模型中的预测结果完全一致，表明这些样本的光谱特征与特定产地类别的关联性较强，模型对其归属的判别较为稳定；部分样本的预测结果存在差异。样本 48 在 SVM、KNN 预测为类别 7，而 LightGBM、XGBoost 预测为类别 6，这可能与 LLE 降维后样本在低维空间中的局部邻域结构差异有关，导致不同模型对边界样本的判别出现分歧。样本 58 在 SVM、

KNN、ANN 中预测为类别 10，而 RF、LightGBM、XGBoost 预测为类别 6，此类差异或源于模型对高方差特征的敏感度不同。特别地，样本 86 在 SVM、ANN 中预测为类别 11，而其他模型预测为类别 2，反映出不同算法对距离度量和特征权重的处理差异，KNN 等基于距离投票的模型可能受限于局部邻近样本分布，而 SVM 通过全局超平面优化，对类别间的整体分离性更敏感。

3 讨论

3.1 各模型产地溯源评价指标对比分析

结果表明，6 种算法的宏平均 AUC 排序依次为 ANN>SVM>RF>LightGBM>XGBoost>KNN，F1 分数排序依次为 SVM>ANN>RF>LightGBM>XGBoost>KNN。

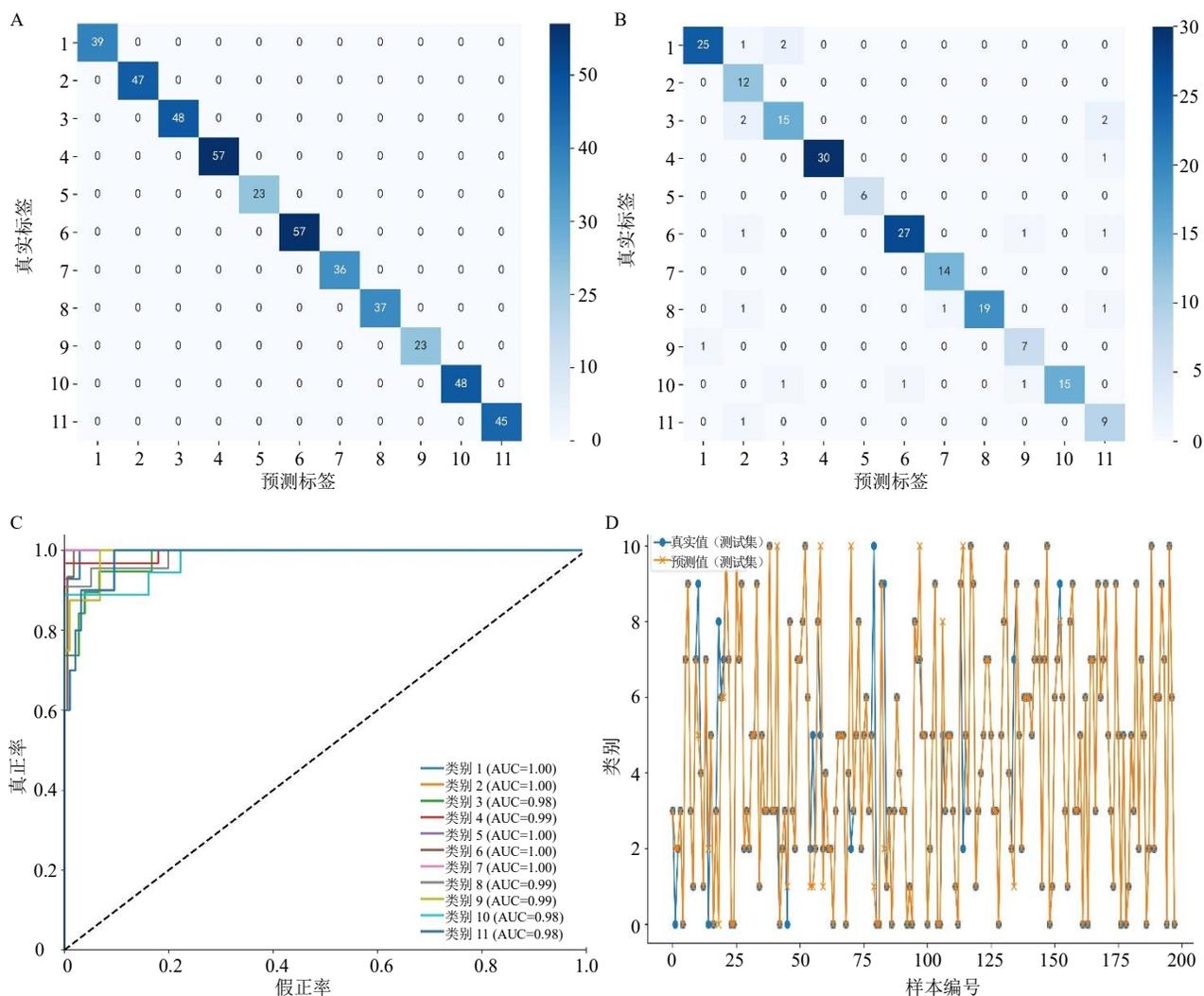


图9 XGBoost 的训练集混淆矩阵 (A)、测试集混淆矩阵 (B)、ROC 曲线 (C)、真实值与预测值对比 (D) for XGBoost

对比结果显示 (表 3), SVM 与 ANN 在产地溯源上的性能指标均显著高于 RF、LightGBM 等其他算法。

3.2 模型泛化性分析

中药产地溯源模型的实际应用价值高度依赖其泛化能力,即模型在不同数据分布、样本规模及验证策略下保持稳定分类性能的能力。为评估 6 种机器学习模型的泛化稳健性,在原实验基础上,将数据集划分比例由 7:3 改为 8:2,随机数种子由 42 调整为 420,交叉验证折数由 5 折提升至 10 折。实验结果显示,在参数调整后,SVM 与 ANN 的各项评价指标仍然显著高于其他算法 (图 11)。

通过对比参数调整前后的模型性能指标变化可知 (图 12),SVM 与 ANN 的各项评价指标均呈

现正向提升,SVM 的测试集准确率、召回率、F1 分数均同步提升 0.013,精确率提升 0.008,宏平均 AUC 虽微降 0.002,但仍保持 0.996 的高水准;ANN 的测试集准确率、精确率、召回率及 F1 分数分别提升 0.016、0.014、0.016 及 0.015,宏平均 AUC 微升 0.001 达 0.998,而其他 4 种算法的各项评价指标却负向降低。这表明 SVM 与 ANN 具备优异的泛化稳健性,在训练样本量增加、验证策略更严格的条件下,仍能稳定捕捉不同产地中药材的光谱特征差异。

3.3 SVM 与 ANN 在中药产地溯源中的独特优势

SVM 通过 RBF 将 MIRS 数据映射至高维空间,以“最大化分类间隔”策略增强对细微光谱差异的区分能力。对于不同产地的中药,其生物碱、

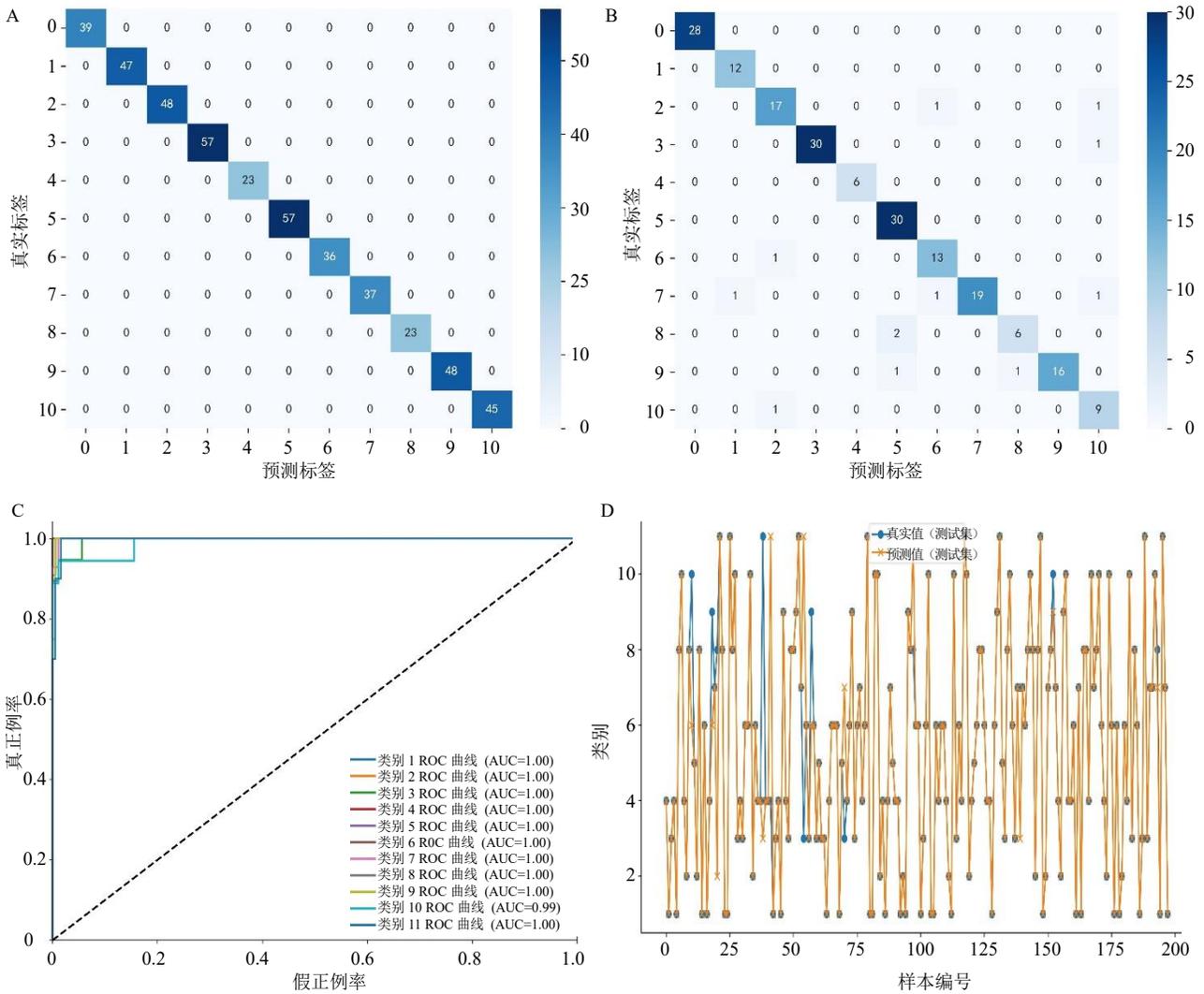


图 10 ANN 的训练集混淆矩阵 (A)、测试集混淆矩阵 (B)、ROC 曲线 (C)、真实值与预测值对比 (D)
 Fig. 10 Confusion matrix of training set (A), confusion matrix of test set (B), ROC curve (C), comparison between true values and predicted values (D) for ANN

表 2 未知样本产地预测结果

Table 2 Predicted results for unknown sample origins

样本编号	产地预测类别					
	SVM	RF	LightGBM	KNN	XGBoost	ANN
3	6	6	6	6	6	6
14	1	1	1	1	1	1
38	4	4	4	4	4	4
48	7	7	6	7	6	7
58	10	6	6	10	6	10
71	11	6	6	11	6	6
79	9	9	9	9	9	9
86	11	2	2	2	2	2
89	3	3	11	3	11	3
110	4	4	4	4	4	4
134	9	9	9	9	9	9
152	2	2	2	2	2	2
227	5	5	5	5	5	5
331	8	2	2	8	8	8
618	3	3	3	3	3	3

表 3 各算法模型在产地溯源评价指标的比较

Table 3 Comparison of various algorithm models in origin traceability evaluation indicators

算法	宏平均 AUC	准确率	精确率	召回率	F1
SVM	0.998	0.949	0.956	0.949	0.949
RF	0.996	0.934	0.948	0.934	0.936
LightGBM	0.995	0.914	0.924	0.914	0.915
KNN	0.989	0.893	0.906	0.894	0.895
XGBoost	0.992	0.904	0.921	0.904	0.908
ANN	0.999	0.939	0.944	0.939	0.940

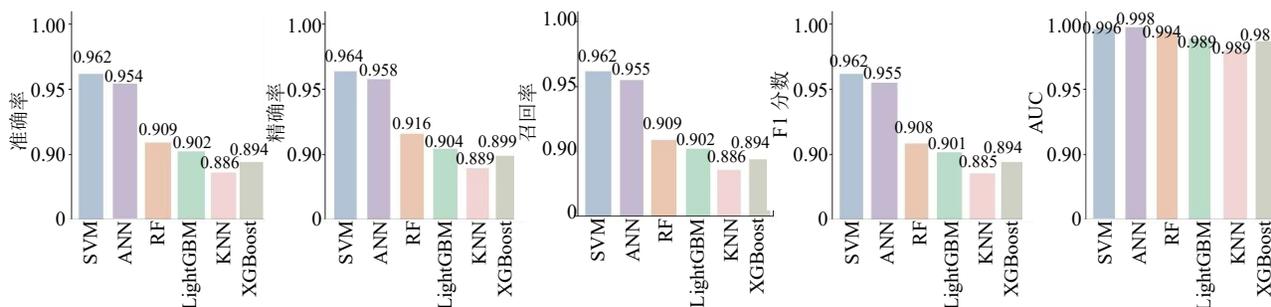


图 11 数据集划分比例 8 : 2、随机种子 420、10 折交叉验证下的各算法测试集的评价指标柱状图

Fig. 11 Bar chart of evaluation metrics for testing sets across algorithms under 8:2 dataset split ratio, random seed 420, and 10-fold cross-validation

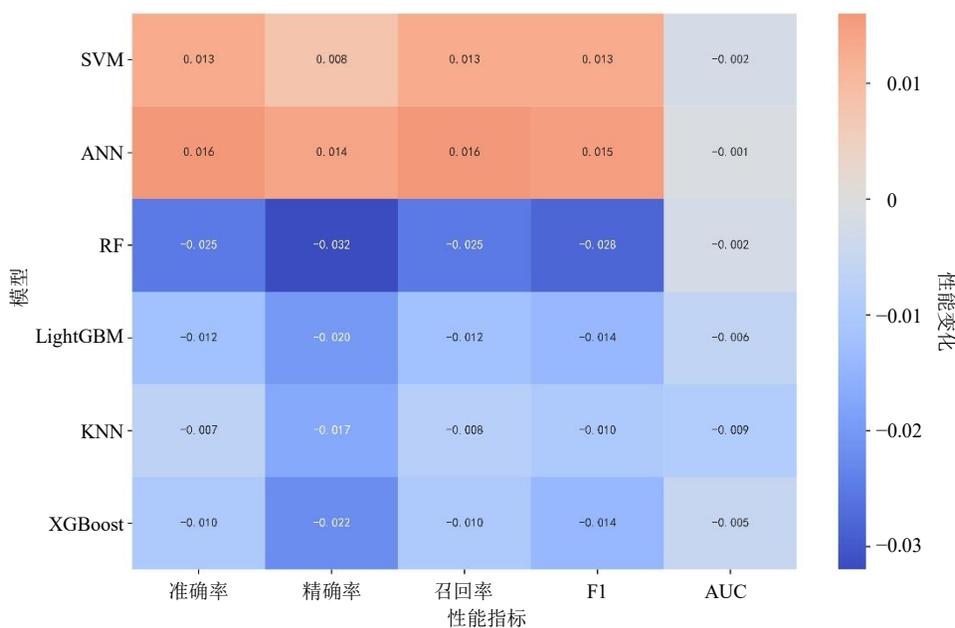


图 12 参数调整前后模型性能指标变化热力图

Fig. 12 Heatmap of model performance metrics before and after parameter adjustment

黄酮类等化学成分的官能团振动差异在光谱中表现为非线性特征，RBF 核可有效将这类特征转化为高维空间中的线性可分问题，尤其适用于边界样本的精准识别。在测试集混淆矩阵中，SVM 对典型产

地特征的捕捉能力突出。ANN 借助多层感知机结构自动学习光谱数据中的高阶特征交互，其深度特征提取能力可挖掘中红外光谱中隐含的官能团振动组合信息，从而模拟中药产地形成的多因素非线性

耦合效应。具备对复杂特征的鲁棒性与对复杂近缘产地的鉴别优势。相比之下, RF、LightGBM 等集成算法虽在特征重要性评估上表现突出, 但对于中药光谱中低丰度成分的特征响应灵敏度稍逊; KNN 依赖局部相似性判断, 在样本分布不均的情况下易受边缘样本干扰, 这也解释了其在部分产地类别中准确率略低的原因。同时, SVM 与 ANN 亦具有互补性, SVM 的全局最优分类面适合处理光谱特征差异显著的产地, 而 ANN 的分布表征能力更适用于特征重叠度高的复杂场景, 这为中药多源产地的精准溯源提供了双重技术路径。

因此, 最终构建出 MIRS 结合 SVM 与 ANN 的方法, 其在中药溯源上具有快速、无损、高通量的优势, 为道地药材的规模化产地溯源提供了客观化工具, 为中药产业的标准化、现代化发展提供了可落地的技术路径。该方法在中药产地溯源中具有 3 重价值: 其一, 相比传统的形态鉴别或单一成分测定, MIRS 可同时反映中药材的整体特征, 结合机器学习实现产地溯源, 可避免因单一成分波动导致的误判; 其二, 算法优化过程针对中药材光谱的高维、高噪声特性进行适配, 提升了模型对微量成分信号的识别能力; 其三, SVM 与 ANN 的互补性为复杂场景提供了灵活方案, 对于道地性明确、成分差异显著的中药材, SVM 可实现快速精准分类; 对于近缘产地、成分差异细微的中药材, ANN 的深度学习能力可挖掘更深层的特征模式。

4 结论及展望

本研究构建了 MIRS 结合机器学习的中药产地溯源方法, 通过数据预处理、多模型对比和参数优化, 证明了该技术路线的有效性。基于 AUC 与准确率、召回率、精准率、F1 的多维度评价指标体系, 得益于 SVM 通过 RBF 最大化分类间隔与 ANN 通过多层感知机捕捉复杂特征交互的优势, SVM (宏 AUC=0.998、F1=0.949、准确率 0.949、召回率 0.949、精确率 0.956) 与 ANN (宏平均 AUC=0.999、F1=0.940、准确率 0.939、召回率 0.939、精确率 0.944) 凭借对非线性特征的高效建模, 在产地溯源上表现最优, 这为中药产地溯源提供了客观、高效的技术方案。同时, MIRS 的化学指纹特性与机器学习的模式识别能力形成互补, 突破了传统鉴别方法的主观性局限, 为中药产地溯源提供了新的技术路径。

需要指出的是, 本研究虽可借助中红外光谱特

征波峰实现中药产地溯源, 但难以全面反映药材产地的多维特征。利用机器学习处理 MIRS 数据时, 存在明显局限性: 其一, 现有 MIRS 数据维度较为单一, 无法覆盖中药产地溯源所需的多维度关键特征, 如药材形态特征、产地生态相关的微生物群落信息、生长周期内的环境因子数据, 这种维度缺失导致模型仅能基于部分化学信息判断产地, 难以全面捕捉产地差异的本质特征; 其二, 模型对数据质量与数量依赖度极高, MIRS 数据易受检测条件影响, 需大量标准化预处理, 且小众样本、不均衡样本会导致模型泛化能力下降; 其三, 特征提取与选择存在短板, MIRS 光谱维度高, 冗余特征与重叠峰易掩盖关键微量成分的弱信号特征, 部分模型还对特征尺度敏感, 易出现强信号主导判断的问题。这些局限使得该应用需结合中药化学成分、炮制原理等领域知识进行人工干预优化, 难以完全脱离专业背景实现“端到端”精准分析。未来可通过数据多模态化路径, 整合近红外光谱 (near-infrared spectroscopy, NIRS)、超高效液相色谱-质谱联用技术 (ultra-high performance liquid chromatography-mass spectrometry, UHPLC-MS) 代谢组学数据及图像信息, 构建多源异构数据融合模型, 以提升对复杂成分差异的捕捉能力, 为道地药材溯源提供新的方法框架。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 赵露颖, 施梦瑶, 张巧艳, 等. 道地药材品质特征及形成机制研究进展 [J]. 中草药, 2022, 53(21): 6931-6947.
- [2] 张瀚文, 李月娥, 喻家伟, 等. 机器学习在中药饮片加工炮制和质量评价中的研究进展 [J]. 中国中药杂志, 2025, 50(13): 3605-3614.
- [3] 王玉琴, 徐芳芳, 张欣, 等. 基于中红外光谱技术桂枝茯苓胶囊浓缩过程快速检测方法研究 [J]. 中草药, 2022, 53(16): 5026-5033.
- [4] Peng Q, Chen J L, Meng K, *et al.* Rapid detection of adulteration of glutinous rice as raw material of Shaoxing Huangjiu (Chinese rice wine) by near infrared spectroscopy combined with chemometrics [J]. *J Food Compos Anal*, 2022, 111: 104563.
- [5] Zhao F Y, Du G R, Huang Y. Exploring the use of Near-infrared spectroscopy as a tool to predict quality attributes in prickly pear (*Rosa roxburghii* Tratt) with chemometrics variable strategy [J]. *J Food Compos Anal*, 2022, 105: 104225.
- [6] 崔萌, 段宝忠, 程蕾, 等. 基于近红外光谱结合化学计

- 量学方法的山里红产地溯源分析 [J]. 中草药, 2024, 55(14): 4897-4906.
- [7] Wang Z W, Liu X W, Li Q N. A euclidean distance matrix model for convex clustering [J]. *J Optim Theory Appl*, 2025, 205(1): 1.
- [8] 徐珍珍, 史星星. 基于电子鼻技术的 LLE+SMA 算法对木香的定性鉴别研究 [J]. 中草药, 2019, 50(24): 6114-6119.
- [9] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. *Science*, 2000, 290(5500): 2323-2326.
- [10] 杨岩, 肖佳妹, 周晋, 等. 支持向量机法及其在中药研究中的应用 [J]. 中草药, 2020, 51(8): 2258-2266.
- [11] 吕红燕, 冯倩. 随机森林算法研究综述 [J]. 河北省科学院学报, 2019, 36(3): 37-41.
- [12] 龚圣, 朱雅宁, 曾陈娟, 等. 近红外光谱结合随机森林算法: 一种快速有效的附子产地溯源策略 [J]. 光谱学与光谱分析, 2022, 42(12): 3823-3829.
- [13] Hajihosseini M, Maghsoudi A, Ghezelbash R. A novel scheme for mapping of MVT-type Pb-Zn prospectivity: LightGBM, a highly efficient gradient boosting decision tree machine learning algorithm [J]. *Nat Resour Res*, 2023, 32(6): 2417-2438.
- [14] 陈舒迪, 柴琴琴, 张勋, 等. 基于多特征融合和 LightGBM 的金线莲品系识别 [J]. 江苏农业学报, 2021, 37(1): 155-162.
- [15] 毋雪雁, 王水花, 张煜东, K. 最近邻算法理论与应用综述 [J]. 计算机工程与应用, 2017, 53(21): 1-7.
- [16] Zhang S. Challenges in KNN classification [J]. *Ieee T Knowl Data En*, 2021, 34(10): 4663-4675.
- [17] 刘庭恺. 多元光谱结合化学计量学对杜仲的产地溯源和真伪鉴别研究 [D]. 武汉: 中南民族大学, 2022.
- [18] 何帅, 周杰, 张福林, 等. 近红外光谱和 XGBoost 的流化床干燥过程水分含量在线检测 [J]. 光谱学与光谱分析, 2024, 44(12): 3347-3352.
- [19] 张文青, 赵珊, 钱文秀, 等. 基于人工神经网络和支持向量机算法的药源性急性间质性肾炎计算模型的建立及在中药域的应用 [J]. 中草药, 2023, 54(2): 416-424.
- [20] 鲍珮瑾, 陈全莉, 赵安迪, 等. 激光诱导击穿光谱和人工神经网络的青白色软玉产地溯源 [J]. 光谱学与光谱分析, 2023, 43(1): 25-30.
- [21] 闫莎莎, 李雪, 洪晶, 等. 基于气味信息和活性成分的三七产地溯源研究 [J]. 中成药, 2024, 46(5): 1740-1745.
- [22] Ghasemi M, Zare M, Trojovský P, et al. Optimization based on the smart behavior of plants with its engineering applications: Ivy algorithm [J]. *Knowl Based Syst*, 2024, 295: 111850.

[责任编辑 潘明佳]