石婷, 许诗咏, 王茹, 等. 紫外-近红外光谱融合结合化学计量学方法鉴别“互助”青稞酒 [J]. 食品工业科技, 2026, 47(3): 313−320. doi: 10.13386/j.issn1002-0306.2025010296
SHI Ting, XU Shiyong, WANG Ru, et al. Identification of Chinese Huzhu Qingke Liquor by UV-NIR Spectral Fusion Combined with Chemometrics Methods[J]. Science and Technology of Food Industry, 2026, 47(3): 313−320. (in Chinese with English abstract). doi: 10.13386/j.issn1002-0306.2025010296

· 分析检测 ·

紫外-近红外光谱融合结合化学计量学方法鉴别“互助”青稞酒

石　婷, 许诗咏, 王　茹, 张明锦*
（青海师范大学化学化工学院, 青海西宁 810000）

摘　要: 本研究旨在开发一种快速、准确的判别方法，以实现对“互助”青稞酒的鉴别，为青稞酒的质量控制与真伪鉴别提供可靠的技术支持。首先利用紫外光谱（Ultraviolet Spectrum，UV）和近红外光谱（Near-Infrared Spectroscopy，NIR）研究了“互助”青稞酒（CHQL）、其它品牌青稞酒（OBQL）和非青稞原料白酒（NQBL）的光谱特性。然后基于 UV、NIR 单光谱和 UV-NIR 融合光谱建立偏最小二乘判别分析（Partial Least Square-Discriminant Analysis，PLS-DA）、支持向量机（Support Vector Machine，SVM）和随机森林（Random Forest，RF）三种分类模型。并通过间隔偏最小二乘法（Interval Partial Least Squares，iPLS）、变量投影重要性（Variable Importance of Projection，VIP）、竞争自适应重加权采样（Competitive Adaptive Reweighted Sampling，CARS）等算法选择特征变量。结果表明：融合光谱能够利用两种光谱的互补信息，提高模型判别能力；此外，特征变量筛选方法进一步优化了模型性能，降低模型复杂度，其中采用 VIP 方法筛选的 78 个最优波长建立的 RF 模型的分类和预测效果最佳，在训练集和测试集上的准确率都达到 100%。综上所述，数据融合策略与化学计量学方法相结合能够有效增强模型性能，实现“互助”青稞酒的快速判别。

关键词: 数据融合, 特征筛选, 青稞酒, 偏最小二乘判别分析, 支持向量机, 随机森林
中图分类号: O657.3　　**文献标识码:** A　　**文章编号:** 1002−0306(2026)03−0313−08
DOI: 10.13386/j.issn1002-0306.2025010296
本文网刊:

Identification of Chinese Huzhu Qingke Liquor by UV-NIR Spectral Fusion Combined with Chemometrics Methods

SHI Ting，XU Shiyong，WANG Ru，ZHANG Mingjin*
（College of Chemistry and Chemical Engineering, Qinghai Normal University, Xining 810000, China）

Abstract: This study aimed to develop a quick and accurate discrimination method to realize the identification of Chinese Huzhu Qingke Liquor (CHQL), and to provide reliable technical support for the quality control and authenticity identification of Qingke liquor. First, ultraviolet spectrum (UV) and near-infrared spectroscopy (NIR) were used to examine the spectral properties of CHQL, Other Brand Qingke Liquor (OBQL), and Non-Qingke Based Liquor (NQBL). Then, partial least square-discriminant analysis (PLS-DA), support vector machine (SVM), and random forest (RF) were established using UV, NIR single spectrum and UV-NIR fusion spectrum. Additionally, interval partial least squares (iPLS), variable importance of projection (VIP), and competitive adaptive reweighted sampling (CARS) were used respectively for extracting feature variables from spectra. The results showed that the fusion spectra could complement each other and improve the performance of the classification model. The feature variable screening method further improved the model performance and reduced the model complexity. The RF model constructed using the 78 optimal wavelengths filtered by the VIP technique had the best classification and prediction results, with 100% accuracy on both the training and

收稿日期: 2025−02−07
基金项目: 国家自然科学基金（22363010）；青海省应用基础研究项目（2022-ZJ-769）。
作者简介: 石婷（2000−）, 女, 硕士研究生, 研究方向: 化学计量学, E-mail: 2572239811@qq.com。
*** 通信作者:** 张明锦（1979−）, 男, 博士, 教授, 研究方向: 化学计量学及应用, E-mail: zhangmingjin@qhnu.edu.cn。

test sets. In summary, the combination of data fusion strategy and chemometrics method can effectively enhance the model performance and realize the rapid discriminant analysis of Huzhu Qingke liquor.

Key words: data fusion; feature selection; high-land barley wine; partial least square-discriminant analysis; support vector machine; random forest

白酒作为中国传统的饮品，在市场上占据着重要的地位。然而，随着白酒市场的逐渐扩大和各品牌竞争的加剧，假冒、稀释充斥市场，严重影响消费者权益和行业健康发展。目前白酒鉴别方法有色谱法、感官评价法、电子鼻和电子舌等。色谱法检测结果准确可靠但费用昂贵且需要费力耗时的分析[1−2]；感官评价法主观性强、准确性受限且需要专业培训[3]；电子鼻和电子舌检测时间短但特异性较低[4]。因此，面对复杂的市场，如何科学地快速、准确鉴别白酒是白酒分析研究的主要问题。青稞酒是以青稞为主要原料酿造的，具有清香醇厚、绵甜爽净的口感[5]。青稞酒富含多种维生素、矿物质以及丰富的 β-葡聚糖[6]。目前对青稞酒的研究主要包括风味物质剖析[7]、健康价值研究[8]、酿造工艺[9] 等，对青稞酒的质量控制研究较少。青海省互助县作为核心产区，“互助”青稞酒被认定为“中国白酒清香型（青稞原料）代表”，是中国国家地理标志产品。对其进行定性研究有助于保障品质和规范市场秩序。

与传统的分析方法相比，光谱技术，包括红外、紫外光谱技术等，通过与化学计量学结合，具有非破坏性、高灵敏度、快速准确等优势，已成为食品质量认证最有前途的方法之一[10−11]。周瑞等[12] 将中红外光谱与化学计量学结合能够快速准确地判别不同酒度以及不同品牌浓香型白酒。其中蚱蜢算法优化支持向量机和误差反向传播人工神经网络模型进行判别时，训练集和测试集的酒样分类准确率均达到 100%。Philippidis 等[13] 利用紫外-可见吸收光谱技术与机器学习建模相结合，对具有“地理标志保护”认证的克里特岛红葡萄酒和白葡萄酒进行鉴定、区分与表征。结果表明，该方法能够区分不同葡萄品种，并首次实现了基于陈酿时间和容器类型的样本分类。然而，当区分极其相似的样本时，单光谱技术的使用可能会导致模型精度较低。因此，需要通过数据融合来提高分类模型的准确性。张维等[14] 采用数据融合策略将近红外光谱（Near-Infrared Spectroscopy, NIR）、GC-MS（Gas Chromatography-Mass Spectrometry, GC-MS）两种数据进行特征级融合建立的白酒等级分类模型准确率达到了 98.28%，分类效果有明显提升。He 等[15] 通过将四向荧光光谱数据与紫外光谱（Ultraviolet Spectrum, UV）、NIR 进行中级数据融合用于识别同一品牌的白酒，模型准确率、灵敏度和特异性分别达到 99.17%、99.15% 和 99.96%。

本研究提出了紫外-近红外（Ultraviolet-Near-Infrared, UV-NIR）光谱数据融合策略，突破了单一光谱的局限性，具有比任何单一数据源更准确和有用的信息[16]。将 NIR、UV 以及 UV-NIR 与化学计量学相结合对“互助”青稞酒（Huzhu Qingke Liquor, CHQL），其它品牌青稞酒（Other Brand Qingke Liquor, OBQL）和非青稞原料白酒（Non-Qingke Based Liquor, NQBL）进行判别分析。UV、NIR 两者的融合能够显著提高分析的全面性和准确性。此外，通过特征变量筛选，能够降低数据维度，提高模型的预测准确率与分析效率。该方法在青稞酒的真伪鉴别和品质控制中展现出广泛的应用潜力，为市场监管和品质保障提供了可靠的技术支持。

1 材料与方法

1.1 材料与仪器

CHQL　青海青稞酒股份有限公司专营店；OBQL、NQBL　西宁市各大商场；甲醇　色谱纯，中国赛默飞世尔科技有限公司；无水乙醇　分析纯，中国医药集团有限公司。

Antaris Ⅱ FT-NIR 光谱仪　Thermo Fisher Scientific 公司；Lambda 系列紫外/可见分光光度计　美国 PerkinElmer 公司。

1.2 实验方法

1.2.1　光谱数据采集　取 3.5 mL 酒样置于石英比色皿中，用近红外光谱仪透射模式在 835~2500 nm 波长范围内扫描样品光谱，分辨率为 0.38 nm，扫描次数为 32 次，三类白酒样品取样量保持一致，重复测定 3 次取平均光谱即得 NIR 光谱。取相同体积样品于石英比色皿中，以无水乙醇为参比，步长为 1 nm，用紫外光谱仪在 200~400 nm 波长范围内扫描样品光谱，三类白酒样品取样量保持一致，重复三次取平均光谱即得 UV 光谱。

1.2.2　光谱预处理　采用多元散射校正（Multiplicative Scatter Correction, MSC）、标准正态变量变换（Standard Normal Variable Transformation, SNV）、一阶导数（First Derivative, FD）、二阶导数（Second Derivative, SD）等方法进行预处理，消除白酒原始光谱中的基线漂移、噪声以及峰重叠等干扰。SNV 和 MSC 在一定程度上能够削弱样品粒度或散射效应的影响[17]。导数算法能够消除基线漂移和叠加效应，提高光谱分辨率[18]。

1.2.3　主成分分析（Principal Component Analysis, PCA）　PCA 是一种探索性分析，通过将原始数据变量线性组合后得到新变量，即主成分，选择保留最大方差的主成分，将高维数据转换为低维数据，从而在更少的维度上表示原始数据的主要变化[19]，因此通常用于数据降维或可视化分析[20]。本研究采用 PCA

对 UV、NIR 原始光谱数据进行处理，以探究三类白酒样品的聚类特征。

1.2.4　模型构建　本研究采用肯纳德-斯顿采样法（Kennard-Stone，K-S）将样本集划分为训练集和测试集，其中 70% 的样本作为训练集，30% 作为测试集。然后建立偏最小二乘判别分析（Partial Least Square-Discriminant Analysis，PLS-DA）、支持向量机（Support Vector Machine，SVM）和随机森林（Random Forest，RF）判别模型。PLS-DA 采用 10 折交叉验证优化潜变量数以提高分类模型性能；SVM 模型采用粒子群算法优化惩罚因子 c 与核参数 g。c 值表示 SVM 算法对离群值的敏感性，影响模型的分类精度。g 值表示映射到高维特征空间的数据分布，影响模型分类训练的速度[21]。RF 通过构建多个决策树并将它们结合起来以提高模型的准确性和泛化能力[22]，本模型为包含 50 棵决策树，每棵树允许最小叶子节点数为 1 的集成训练。

1.2.5　特征变量筛选

1.2.5.1　间隔偏最小二乘法（Interval Partial Least Squares，iPLS）　iPLS 算法的基本原理是将整个频谱划分为若干较小的等宽区域[23]。然后基于每个子区间建立多个 PLS 回归模型，得到每个子区间对应的交叉检验均方根误差（Root Mean Square Error of Cross Validation，RMSECV）。最后选取 RMSECV 最低的区域，并以此为基础构建最终模型[16]。在白酒分析中，不同波段可能对应不同化学成分，iPLS 有助于定位关键化学信息，减少冗余波段的干扰。但区间划分的宽度和数量需要人为设定，可能影响结果的准确性。本研究将 UV-NIR 光谱按等距原则划分为 10、20、30、40、50 和 60 个区间。对每个区间单独建立 PLS 模型，选择 RMSECV 最小的区间作为特征变量。

1.2.5.2　变量投影重要性（Variable Importance of Projection，VIP）　VIP 是一种在多元数据分析中用于评估变量相对重要性的方法。UV-NIR 光谱数据维度高，VIP 能够提供全局变量重要性排序，便于识别关键变量。但是筛选变量时需设定阈值，选择不当可能导致重要变量被忽略或保留无关变量。VIP 值反映了对模型有贡献的自变量重要性和因变量在建模中的表达程度[24]。VIP 值计算公式如下：

$$VIP_j = \sqrt{P \times \sum_{i=1}^{k} \left(\omega_{ij}^2 \times R^2(Y;t_i)\right) \Big/ \sum_{i=1}^{k} R^2(Y;t_i)} \quad \text{式（1）}$$

式中：P 是自变量的个数，k 是偏最小二乘方法从原变量中提取的成分个数，t_i 代表第 i 个潜变量，ω_{ij} 是自变量在计算第 i 个偏最小二乘成分时的权重，$R^2(Y;t_i)$ 是第 i 个偏最小二乘成分 t_i 对因变量 Y 的解释方差。

通常选择 VIP 值大于 1 的变量作为特征变量[25]。本研究基于 UV-NIR 数据建立 PLS 模型，计算各波长的 VIP 值，筛选出 VIP 值高于阈值的波长作为特征变量。

1.2.5.3　竞争自适应重加权采样（Competitive Adaptive Reweighted Sampling，CARS）　CARS 是基于达尔文进化论“适者生存”的原则开发的特征选择方法[26]，它有助于有效地去除冗余变量，同时压缩共线性变量，但是计算较为复杂、耗时。CARS 算法采用蒙特卡罗采样（Monte Carlo Sampling，MCS），选取一定比例的样本组成校准集。然后，根据回归系数，采用指数递减函数（Exponential Decresing Function，EDF）强制波长选择和自适应重加权采样（Adaptive Reweighted Sampling，ARS）竞争波长选择两步方法选择关键波长[27−28]；最后，选择 RMSECV 最小的子集。本研究采用 CARS 算法进行特征变量筛选时，设置蒙特卡洛采样次数为 100 次，通过 10 折交叉验证评估迭代中变量子集的性能，选择 RMSECV 最小的子集作为特征集。

1.2.6　数据融合策略　数据融合策略是合并、集成和分析不同来源或类型的数据以提取更全面、准确和有用信息的过程[29]。数据融合作为一种判断更准确、精度更高的技术被广泛应用于食品[30]、农业[31]和药物[32] 等领域。根据数据处理的形式，数据融合可以分为数据级融合、特征级融合和决策级融合 3 种。数据级融合是指将不同数据源的数据直接合并[33]。特征级数据融合是从单个数据源中提取特征变量，将不同特征的信息进行组合以最小化数据规模并去除冗余信息[34]。决策级融合是将各个数据源独立进行处理和决策，然后将这些决策结果进行融合。本研究基于数据级融合策略来探索模型的分类效果。

1.2.7　模型评价　建模完成后，建立混淆矩阵，分别通过以下评价指标筛选出最佳模型：准确率（Accuracy）、精确率（Precision）和召回率（Recall），三个指标计算公式如下：

准确率：衡量模型预测正确的样本占总样本的比例。

$$\text{准确率} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad \text{式（2）}$$

精确率：衡量模型预测为正类的样本中实际为正类的比例。

$$\text{精确率} = \frac{TP}{TP + FP} \times 100 \quad \text{式（3）}$$

召回率：衡量模型能够正确识别出真实正类样本的能力。

$$\text{召回率} = \frac{TP}{TP + FN} \times 100 \quad \text{式（4）}$$

式中：TP 表示模型正确预测为正类的样本；FP 表示模型错误预测为正类的负类样本；TN 表示模型正确预测为负类的样本；FN 表示模型错误预测为负类的正类样本。

混淆矩阵中的行对应于预测的类，列代表实际类。数字对角线中的数字代表正确分类的样品数，非对角线为错误分类的样本数。

1.3 数据处理

使用 MATLAB R2018b 软件进行光谱数据处理及绘图。

2 结果与分析

2.1 光谱分析

图 1 分别为 113 种白酒的 NIR、UV 原始光谱图。图 1A 显示，NIR 光谱中 1450~1460 nm 和 1930~1940 nm 处的吸收峰主要与水分子和乙醇的羟基(O-H)的伸缩振动一级倍频和组合频有关[35]；在 1690~1780 nm 附近的峰归属于 C-H 键伸缩振动[36]。图 1B 中，在 200~230 nm 区域的吸收主要与白酒中的一些不饱和键和共轭体系有关，如 C=C、C=O 等[37]；在 273 nm 波段主要对应糠醛等物质的吸收峰[38]。

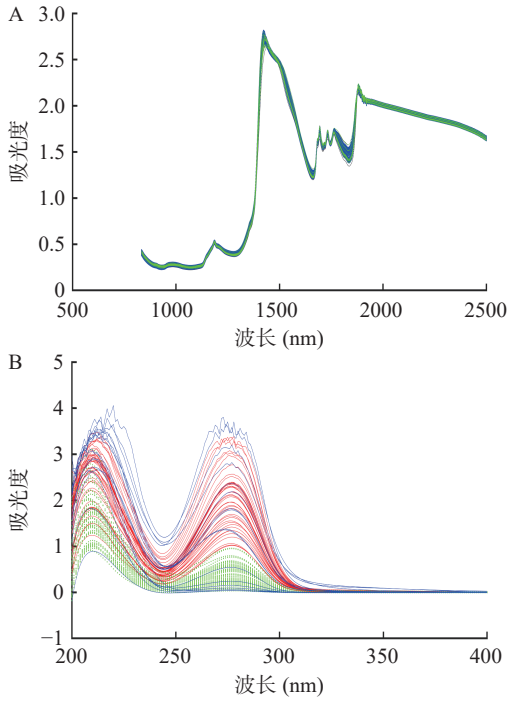


图 1 113 种白酒样品的原始光谱图

Fig.1 Original spectra of 113 liquor samples

注：A: NIR; B: UV。

2.2 主成分分析

将 PCA 应用于 NIR、UV 原始光谱数据集的可视化。在 NIR 获得的数据中，前三个主成分共解释了 98.1% 的数据方差(图 2)。在 UV 获得的数据中，前三个主成分共解释了 97.4% 的数据方差。数据的高解释性证明了前三个主成分代表了大部分原始数据信息。NIR 得分图显示，CHQL 有一定的聚类倾向。UV 得分图显示，沿着 PC2 三类白酒样品聚类较为明显。但是样品重叠都较为严重，可能由于不同来源的白酒在化学成分上具有较高的相似性。因此，在接下来的研究中使用有监督学习方法，结合特征变量筛选，进一步提高分类能力。

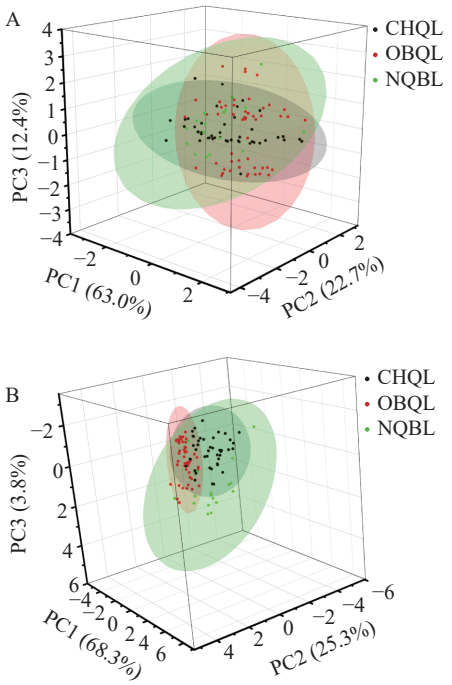


图 2 三类白酒样品的原始光谱得分图

Fig.2 Raw spectral score plots of the three types of liquor samples

注：A: NIR, B: UV。

2.3 基于不同预处理方法的单光谱 PLS-DA 模型

采用 MSC、SNV、FD 和 SD 等 4 种不同的预处理方法对 NIR、UV 进行预处理后建立 PLS-DA 判别模型。表 1 显示了原始光谱(Raw Spectrum, RAW)及不同预处理方法对 PLS-DA 模型的准确率。结果表明，MSC 在两种光谱中表现最优，对于训练集 NIR 和 UV 准确率均提升了 1.26%；对于测试集 UV 准确率提升了 2.94%。

表 1 不同方法预处理后 PLS-DA 建模的准确率

Table 1 Accuracy obtained from PLS-DA modeling after pre-processing with different methods

预处理方法	NIR准确率(%)		UV准确率(%)	
	训练集	测试集	训练集	测试集
RAW	82.28	94.12	86.08	97.06
MSC	83.54	94.12	87.34	100
SNV	82.28	94.12	84.81	94.12
FD	83.54	91.18	81.01	97.06
SD	82.28	91.18	82.28	97.06

2.4 NIR 和 UV 的单光谱分类模型

NIR 和 UV 数据经 MSC 预处理后，分别建立 RF、SVM 和 PLS-DA 分类模型。结果如图 3 所示，对于两种单光谱(NIR、UV)，采用 RF 建立的分类模型虽然训练集准确率达到 100%，但是预测集准确率较低；而采用 PLS-DA 建立的分类模型在测试集上准确率较高，但是训练集准确率一般；对于 SVM 建立的分类模型，UV 数据集在训练集和预测集上准确率较好。结果表明：单光谱的独立建模存在局限性，这可能是由于不同白酒在酿造工艺上存在一定的相

似性，使得白酒的成分和性质差异不明显；同时单一光谱信息量不足，导致分类结果不好。

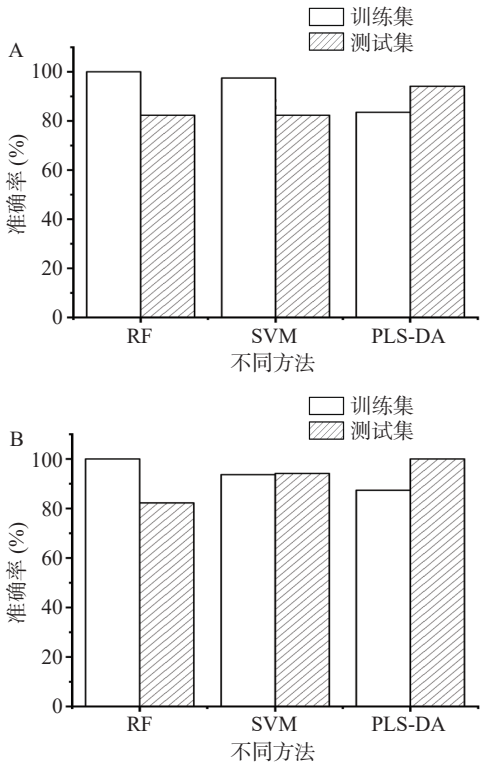


图 3　不同方法在 NIR 和 UV 数据集的建模结果

Fig.3　Modeling results of different methods on NIR and UV datasets

注：A：NIR；B：UV。

2.5　基于数据级融合策略的分类模型

本研究采用数据级融合（Data-Level Fusion，DLF）策略将经 MSC 预处理后的 UV 和 NIR 数据集以端到端的方式组合，并构建 RF、SVM 和 PLS-DA 分类模型。为提高模型的准确率，对 UV 和 NIR 数据进行 LF 融合后筛选特征变量。

2.5.1　基于 DLF 的三种建模方法　表 2 给出了基于 DLF 的三种模型的分类结果。由表 2 可见，基于数据融合的分类模型优于基于单一光谱的模型，其中 RF 方法建立的分类模型准确率最高，训练集和测试集准确率分别为 100% 和 91.18%。结果表明数据级融合策略能够得到更全面、准确和有用的信息以增强模型的判别能力，后续基于融合光谱进行特征变量筛选后分类建模。

表 2　基于 DLF 的三种建模方法准确率

Table 2　Accuracies of three modeling methods based on DLF

数据	建模方法	训练集(%)	测试集(%)
DLF	SVM	100	85.29
	PLS-DA	91.14	91.18
	RF	100	91.18

2.5.2　特征变量筛选

2.5.2.1　间隔偏最小二乘法　表 3 列出了划分为不同区间数时，选择的最优区间、预测误差及区间包含的变量数。结果可见，最佳的划分区间数是 20，最优的子区间是第 5 子区间。图 4 是将光谱划分为 20 个区间时，各区间变量建模的预测误差，将每个子区间获得的 RMSECV 值与全谱 PLS 的 RMSECV 值进行比较。从图中可以看出，第 3、5、7 和 8 区间 RMSECV 值均小于全光谱波段子区间选择的阈值。其中子区间 5 的 RMSECV 值最小。故将数据划分为 20 个区间，选择第 5 区间为特征区间，包含 114 个特征变量。

表 3　划分不同区间数时建模及区间选择结果

Table 3　Modeling and interval selection results when dividing the number of different intervals

区间数	最优子区间	RMSECV	特征变量数
10	3	0.4056	228
20	5	0.3752	114
30	7	0.3789	76
40	10	0.4416	57
50	17	0.4568	46
60	4	0.4633	38

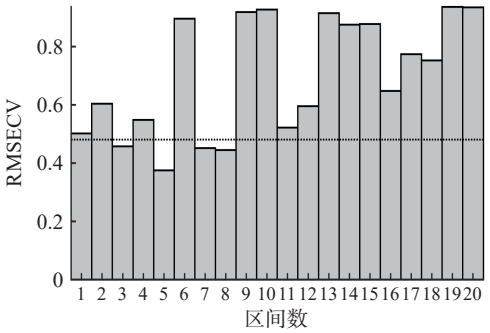


图 4　划分 20 个子区间时 iPLS 建模结果

Fig.4　Modeling results of iPLS when dividing into 20 sub-intervals

2.5.2.2　变量投影重要性　VIP 值反映了变量在解释因变量方差时的贡献程度，VIP 值大于 2 意味着对应的特征在模型中的重要性明显超过平均水平，能够有效提升模型的预测能力和解释力，在模型中起着至关重要的作用。各波长变量对应的 VIP 得分如图 5 所示，共筛选出 VIP 值大于 2 的波长变量 78 个。

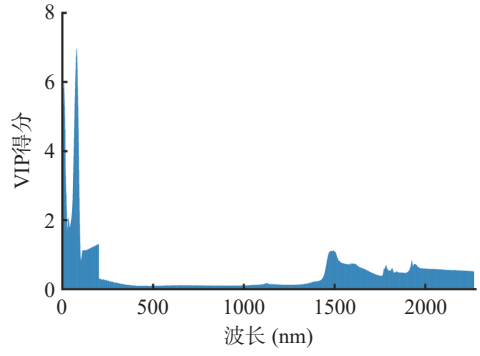


图 5　各波长变量对应的 VIP

Fig.5　VIP corresponding to each wavelength variable

2.5.2.3 竞争自适应重加权采样 采用 CARS 筛选融合光谱的特征变量。图 6 是采样次数为 100、10 折交叉检验时变量数(图 6A)、RMSECV 值(图 6B)和每个变量的回归系数路径(图 6C)随着 CARS 采样运行次数的增加的变化趋势。图 6A 中,所选择的特征变量数量随着采样运行次数的增加而逐渐减少。这个过程可以分为两个阶段。第一阶段为“快速选择阶段”,在这一阶段采样变量的数量快速减少。第二阶段为“精选阶段”,特征变量的数量缓慢减少并趋于稳定,最终保留 13 个特征变量。图 6B 中,当采样运行次数小于 74 次时,RMSECV 值增加缓慢,表明大量与分类无关的信息被去除。当采样运行次数大于 74 次时,RMSECV 增长变得明显,这主要是由于从最优子集中消除一些关键变量导致信息丢失而增加。图 6C 中,当采样运行次数为 74 次时,在星号垂直线处得到变量的最优组合,RMSECV 值最小为 0.4867。

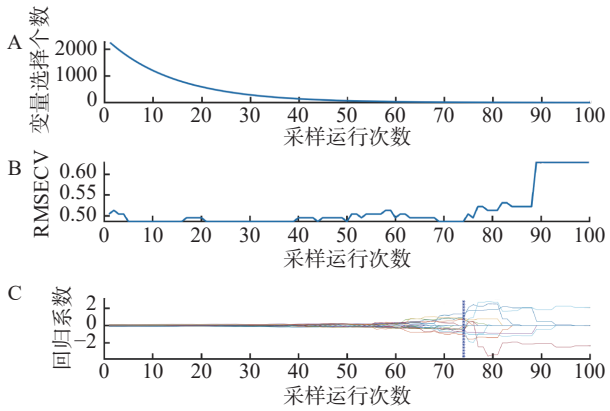


图 6 CARS 选择特征变量的过程

Fig.6 Process of CARS in selecting characteristic variables

2.5.3 基于融合光谱数据的最优建模方法选择 图 7 总结了采用不同变量筛选方法时,不同分类模型的准确率。对于 RF 和 PLS-DA 模型,都是采用 VIP 算法从原始变量中提取 78 个关键变量,筛选后模型训练集和测试集准确率分别达到 100%、100% 和 89.87%、97.06%;对于 SVM 模型,采用 CARS 算法准确率最高(96.2%, 97.06%),关键变量为 13 个。结果表明,特征变量筛选进一步优化了模型的性能,减少了冗余信息对模型的干扰,同时模型的效率显著提升;RF-VIP 和 SVM-CARS 以更高的准确率和更低的运行时间被选为候选模型。

采用混淆矩阵计算精确率、召回率等评价指标以进一步评估两个模型的性能。RF-VIP 分类模型混淆矩阵如图 8 所示:由公式(2)~公式(4)计算得出,模型在训练集和测试集上的准确率达到了 100%,同时其精确率和召回率也均为 100%,表明该模型能够完全正确地区分三类白酒。SVM-CARS 分类模型混淆矩阵如图 9 所示:由公式(3)和公式(4)计算得出,模型对于 CHQL、OBQL、NQBL 在训练集上的精确率分别为 95.45%、100%、95.65%,召回率为 100%、78.57%、100%;在测试集上的精确率分别为 100%、100%、90%,召回率为 100%、66.67%、100%;SVM-CARS 模型在 CHQL 和 NQBL 类别上表现很好,但在 OBQL 类别上的召回率相对较低。综上所述,RF-VIP 模型在本研究中表现最佳,适用于“互助”青稞酒的快速判别分析。

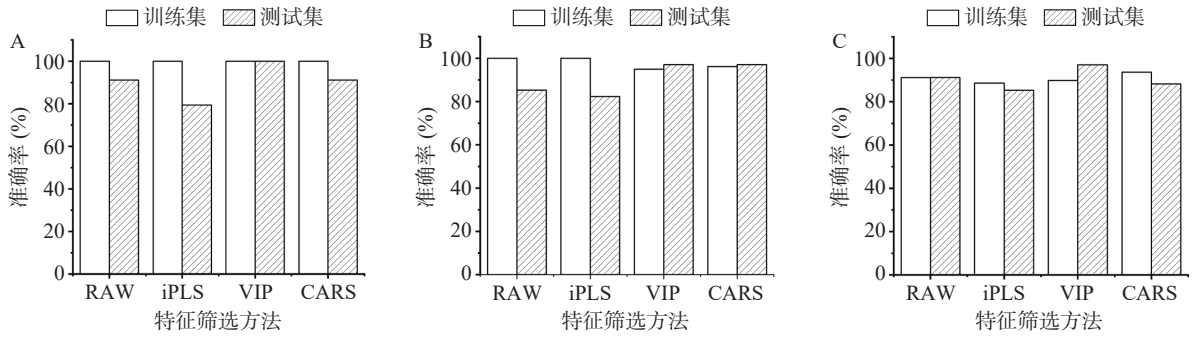


图 7 不同模型在不同变量筛选方法时的建模效果

Fig.7 Modeling effects of different models under different variable screening methods

注:A:RF;B:SVM;C:PLS-DA。

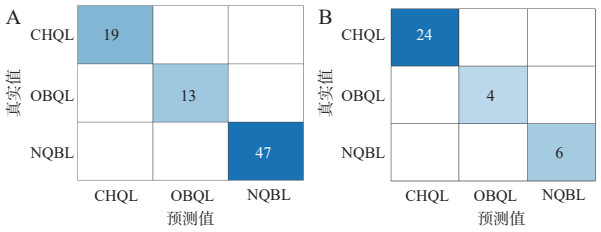


图 8 RF 模型的混淆矩阵

Fig.8 Confusion matrix of the RF model

注:A:训练集;B:测试集。

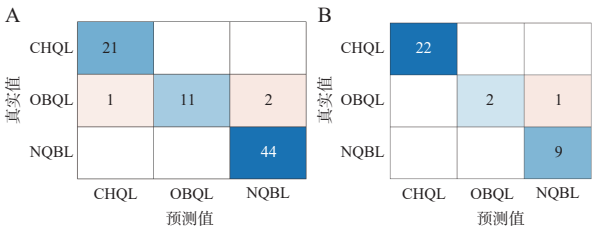


图 9 SVM 模型的混淆矩阵

Fig.9 Confusion matrix of the SVM model

注:A:训练集;B:测试集。

3　结论

本研究基于紫外-近红外融合光谱建立“互助”青稞酒的快速判别模型。首先对原始光谱进行预处理（MSC、SNV、FD、SD），确定 MSC 为最佳预处理方法。为验证数据融合的必要性，对 UV、NIR 单一光谱和 UV-NIR 融合光谱分别建立 PLS-DA、SVM 和 RF 判别模型，结果表明 UV-NIR 的判别模型准确率比单一光谱好，说明融合光谱能够突破单一光谱的局限性，包含更准确、有用的信息。最后基于 UV-NIR 进行特征变量筛选，采用 VIP 方法提取出 78 个特征变量，所建模型的分类准确率提升至 100%，且模型复杂度降低。综上所述，融合光谱技术与化学计量学方法相结合能够对“互助”青稞酒实现快速分类，是一种很有前途的白酒鉴别技术。

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

参考文献

[1] JIA W, FAN Z B, DU A, et al. Recent advances in Baijiu analysis by chromatography based technology-A review[J]. Food Chemistry, 2020, 324: 126899.

[2] FULGENCIO A C C, RESENDE G A P, TEIXEIRA M C F, et al. Screening method for the rapid detection of diethylene glycol in beer based on chemometrics and portable near-infrared spectroscopy[J]. Food Chemistry, 2022, 391: 133258.

[3] LU L, HU Z Q, HU X Q, et al. Electronic tongue and electronic nose for food quality and safety[J]. Food Research International, 2022, 162: 112214.

[4] 周易枚, 陈彬, 丁红梅, 等. 橄榄油鉴伪检测方法的研究进展[J]. 中国粮油学报, 2025, 40(2): 227−234. [ZHOU Y M, CHEN B, DING H M, et al. Research progress on olive oil counterfeiting detection methods[J]. Journal of the Chinese Cereals and Oils Association, 2025, 40(2): 227−234.]

[5] 师成旭. 发酵型藏传青稞酒生产工艺的研究[J]. 酿酒, 2018, 45(5): 59−61. [SHI C X. Research on the production process of fermented Tibetan barley liquor[J]. Liquor-Making Science & Technology, 2018, 45(5): 59−61.]

[6] 游茂兰, 邓婧, 覃小丽, 等. β-葡聚糖对传统青稞酒发酵的影响[J]. 食品与发酵工业, 2019, 45(10): 149−154. [YOU M L, DENG J, QIN X L, et al. Effect of β-glucan on the fermentation of traditional barley wine[J]. Food and Fermentation Industries, 2019, 45(10): 149−154.]

[7] CHEN L H, PENG Q H, LIU B, et al. Study on the dynamic changes of nutrient components and flavor compounds during the fermentation process of high-quality highland barley wine[J]. International Journal of Gastronomy and Food Science, 2024, 35: 100860.

[8] 焦兴弘. 舟曲青稞酒的营养价值及保健功能[J]. 食品安全导刊, 2020(12): 95. [JIAO X H. Zhouqu barley wine nutritional value and health functions[J]. China Food Safety Magazine, 2020(12): 95.]

[9] 杜木英, 伍怡郦, 阚建全, 等. 传统青稞酒发酵过程中化学成分动态变化的研究[J]. 食品工业科技, 2007(9): 94−98. [DU M Y, WU Y L, KAN J Q, et al. Study on the dynamic change of chemical composition during the fermentation process of traditional highland barley wine[J]. Science and Technology of Food Industry, 2007(9): 94−98.]

[10] RUGGIERO L G, AMALFITANO C, DI VAIO C, et al. Use of near-infrared spectroscopy combined with chemometrics for authentication and traceability of intact lemon fruits[J]. Food Chemistry, 2022, 375: 131822.

[11] LIU R, SUN Q, HU T, et al. Multi-parameters monitoring during traditional Chinese medicine concentration process with near infrared spectroscopy and chemometrics[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 192: 75−81.

[12] 周瑞, 陈晓明, 张莉丽, 等. 化学计量学结合中红外光谱的浓香型白酒分类研究[J]. 食品与发酵工业, 2023, 49(5): 88−93. [ZHOU R, CHEN X M, ZHANG L L, et al. Classification of strong aromatic liquor by chemometrics combined with mid-infrared spectroscopy[J]. Food and Fermentation Industry, 2023, 49(5): 88−93.]

[13] PHILIPPIDIS A, POULAKIS E, KONTZEDAKI R, et al. Application of ultraviolet-visible absorption spectroscopy with machine learning techniques for the classification of cretan wines[J]. Foods, 2020, 10(1): 9.

[14] 张维, 张贵宇, 庹先国, 等. 基于 NIR 和 GC-MS 融合技术的浓香型白酒原酒等级鉴别[J]. 食品科学, 2024, 45(21): 288−296. [ZHANG W, ZHANG G Y, TUO X G, et al. Grade identification of strong aroma liquor based on NIR and GC-MS fusion technology[J]. Food Science, 2024, 45(21): 288−296.]

[15] HE M, CHEN X, ZHANG J, et al. Identification of liquors from the same brand based on ultraviolet, near-infrared and fluorescence spectroscopy combined with chemometrics[J]. Food Chemistry, 2022, 400: 134064−134064.

[16] WANG Y Q, YANG Y Z, SUN H J, et al. Application of a data fusion strategy combined with multivariate statistical analysis for quantification of puerarin in *Radix puerariae*[J]. Vibrational Spectroscopy, 2020, 108: 103057.

[17] BADARO A T, E TAVARES J P H, BLASCO J, et al. Near infrared techniques applied to analysis of wheat-based products: Recent advances and future trends[J]. Food Control, 2022, 140: 109115.

[18] QIU X W, DONG Y Q, JIANG L W, et al. Portable near-infrared spectroscopy with variable selection-linear discriminant analysis technology for accurate and nondestructive detection of sulfur-fumigated *Citri Reticulatae Pericarpium*[J]. LWT, 2024, 205: 116518.

[19] LI G Y, LI J Q, LIU H G, et al. Geographic traceability of *Gastrodia elata* Blum based on combination of NIRS and chemometrics[J]. Food Chemistry, 2025, 464: 141529.

[20] BIANCOLILLO A, MARINI F. Chemometric methods for spectroscopy-based pharmaceutical analysis[J]. Frontiers in chemistry, 2018, 6: 576.

[21] XIAO S J, WANG Q H, LI C F, et al. Rapid identification of A1 and A2 milk based on the combination of mid-infrared spectroscopy and chemometrics[J]. Food Control, 2022, 134: 108659.

[22] ZHANG Y, LI Y, ZHOU C, et al. Rapid and nondestructive identification of origin and index component contents of Tiegun yam based on hyperspectral imaging and chemometric method[J]. Journal of Food Quality, 2023, 2023(1): 6104038.

[23] PARISOTTO G, FERRAO M F, MULLER A L H, et al. Total acid number determination in residues of crude oil distillation using ATR-FTIR and variable selection by chemometric methods[J]. Energy & Fuels, 2010, 24(10): 5474−5478.

[24] YANG P P, HOU Y, CHEN A M, et al. Rapid authentication of variants of *Gastrodia elata* Blume using near-infrared spectroscopy combined with chemometric methods[J]. Journal of Pharmaceutical and Biomedical Analysis, 2023, 235: 115592.
[25] MALECHAUX A, LE DREAU Y, ARTAUD J, et al. Control chart and data fusion for varietal origin discrimination: Application to olive oil[J]. Talanta, 2020, 217: 121115.
[26] LI H D, LIANG Y Z, XU Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica Chimica Acta, 2009, 648(1): 77−84.
[27] WU J Z, ZAREEF M, CHEN Q S, et al. Application of visible-near infrared spectroscopy in tandem with multivariate analysis for the rapid evaluation of matcha physicochemical indicators[J]. Food Chemistry, 2023, 421: 136185.
[28] SHE X T, HUANG J, CAO X Q, et al. Rapid Measurement of total saponins, mannitol, and naringenin in *Dendrobium officinale* by near-infrared spectroscopy and chemometrics[J]. Foods, 2024, 13(8): 1199.
[29] SUN W J, ZHANG X, ZHANG Z Y, et al. Data fusion of near-infrared and mid-infrared spectra for identification of rhubarb [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2017, 171: 72−79.
[30] GUO M, WANG K, LIN H, et al. Spectral data fusion in non-destructive detection of food products: Strategies, recent applications, and future perspectives[J]. Comprehensive Reviews in Food Science and Food Safety, 2024, 23(1): e13301.
[31] BARRILE V, SIMONETTI S, CITRONI R, et al. Experimenting agriculture 4.0 with sensors: A data fusion approach between remote sensing, UAVs and self-driving tractors[J]. Sensors, 2022, 22(20): 7910.
[32] STEYAERT S, PIZURICA M, NAGARAJ D, et al. Multimodal data fusion for cancer biomarker discovery with deep learning[J]. Nature Machine Intelligence, 2023, 5(4): 351−362.
[33] ZHANG J, WANG Y Z, YANG M Q, et al. Identification and evaluation of *Polygonatum kingianum* with different growth ages based on data fusion strategy[J]. Microchemical Journal, 2021, 160: 105662.
[34] BORRAS E, FERRE J, BOQUE R, et al. Olive oil sensory defects classification with data fusion of instrumental techniques and multivariate analysis (PLS-DA)[J]. Food chemistry, 2016, 203: 314−322.
[35] 王凡, 张文娟, 李国辉, 等. 近红外光谱技术同时快速定量酱香型白酒基酒醇类物质[J]. 食品与发酵工业, 2025, 51(16): 324−331. [[WANG F, ZHANG W J, LI G H, et al. Simultaneous and rapid quantification of alcohols in soy sauce-type baijiu base wines by near-infrared spectroscopic techniques[J]. Food and Fermentation Industry, 2025, 51(16): 324−331.]
[36] 彭帮柱, 龙明华, 岳田利, 等. 用偏最小二乘法及傅立叶变换近红外光谱快速检测白酒酒精度[J]. 农业工程学报, 2007(4): 233−237. [PENG B Z, LONG M H, YUE T L, et al. Rapid detection of liquor alcohol by partial least squares and fourier transform near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2007(4): 233−237.]
[37] 佘露露, 王蕾, 蒋洪久, 等. 基于紫外光谱技术快速鉴别不同工艺酱香型白酒[J]. 中国酿造, 2023, 42(9): 211−216. [SHE L L, WANG L, JIANG H J, et al. Rapid identification of different processes of soy sauce-type white wine based on ultraviolet spectroscopic technology[J]. China Brewing, 2023, 42(9): 211−216.]
[38] 赵宇. 紫外分光光度法测定十种白酒中的糠醛[J]. 中国林副特产, 2019(6): 29−31. [ZHAO Y. Determination of furfural in ten liquors by ultraviolet spectrophotometry[J]. Forest by-Product and Speciality in China, 2019(6): 29−31.]