作物学报 ACTA AGRONOMICA SINICA 2026, 52(3): 922−935 http://zwxb.chinacrops.org/
ISSN 0496-3490; CN 11-1809/S; CODEN TSHPA9 E-mail: zwxb301@caas.cn

DOI: 10.3724/SP.J.1006.2026.54100

基于高光谱成像的烟叶主要化学成分无损检测与可视化

杨 月[1] 张新新[1,2] 贺增辉[3] 李瑞东[3] 潘昱洁[3] 李嘉康[1] 杜 薇[4] 徐大勇[1] 堵劲松[1,*]

[1] 中国烟草总公司郑州烟草研究院, 河南郑州 450001; [2] 郑州轻工业大学软件学院, 河南郑州 450002; [3] 云南烟叶复烤有限责任公司, 云南昆明 650021; [4] 四川中烟工业有限责任公司, 四川成都 610000

摘 要: 烟叶的化学成分是决定其香气、风味与吸食质量的关键因素, 利用高光谱成像(hyperspectral Imaging, HSI)技术能够实现烟叶主要化学成分的快速、无损检测与可视化。本研究选取云南不同产区、不同等级的 240 份烟叶样本为研究对象, 采集 967.05~2561.33 nm 范围内的高光谱图像, 提出一种结合光谱差值与超像素聚类的感兴趣区域(region of Interest, ROI)提取方法, 有效剔除背景、叶脉及不规则结构干扰。光谱数据经标准正态变量变换(standard normal variate, SNV)与一阶导数(first derivative, FD)联合处理后, 采用主成分分析(principal component analysis, PCA)进行特征降维, 并基于偏最小二乘回归(partial least squares regression, PLSR)对烟碱、总糖、还原糖、总氮、钾和氯 6 种成分进行统一建模。结果表明, SNV+FD 预处理能有效提升模型性能, 其中烟碱、总糖、还原糖和总氮的交叉验证决定系数(coefficient of determination for cross-validation, Q^2)均超过 0.89, 交叉验证均方根误差(root mean square error of cross-validation, RMSECV)最低达 0.09; 在 20 个独立样本测试中, 6 种成分的预测决定系数(coefficient of determination for prediction, R^2)分别为 0.930、0.908、0.854 和 0.915, 平均相对偏差(relative difference, RD)小于 5%, 相对预测偏差(residual prediction deviation, RPD)均高于 2.5, 验证了模型的稳定性与预测能力。基于所建模型实现了烟叶主要化学成分的可视化分布, 揭示了各成分在叶片不同区位的异质性特征。本研究提出的 ROI 提取与建模方法能够有效实现烟叶主要化学成分的无损检测与可视化, 可为烟叶品质评价与精细化加工提供技术支撑。

关键词: 高光谱成像; 无损检测; 化学成分; 感兴趣区域; 偏最小二乘回归; 烟草

Non-destructive prediction and visualization of major chemical components in tobacco leaves using hyperspectral imaging

Yang Yue[1], Zhang Xin-Xin[1,2], He Zeng-Hui[3], Li Rui-Dong[3], Pan Yu-Jie[3], Li Jia-Kang[1], Du Wei[4], Xu Da-Yong[1], and Du Jin-Song[1,*]

[1] Zhengzhou Tobacco Research of China National Tobacco Corporation, Zhengzhou 450001, Henan, China; [2] School of Software, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China; [3] Yunnan Leaf Tobacco Redrying Co., Ltd., Kunming 650021, Yunnan, China; [4] China Tobacco Sichuan Industrial Co., Ltd., Chengdu 610000, Sichuan, China

Abstract: The chemical composition of tobacco leaves plays a crucial role in determining their aroma, flavor, and smoking quality. Hyperspectral imaging (HSI) offers a rapid, non-destructive means of detecting and visualizing key chemical constituents in

本研究由中国烟草总公司重大科技计划项目(110202401033(SZ-07)), 中国烟草总公司云南省公司科技计划重大项目(2024530000241026), 云南烟叶复烤有限责任公司陆良复烤局科技计划项目(2023FK03)和河南省科技攻关项目(252102211069)资助。
This study was supported by the Major Science and Technology Program of China Tobacco Corporation (110202401033(SZ-07)), the Major Science and Technology Program of Yunnan Provincial Tobacco Company of China National Tobacco Corporation (2024530000241026), the Science and Technology Program of Luliang Redrying Plant, Yunnan Tobacco Redrying Co., Ltd. (2023FK03), and the Science and Technology Research and Development Plan of Henan Province (252102211069).

* 通信作者(Corresponding author): 堵劲松, E-mail: djsdxx@126.com
第一作者联系方式: E-mail: y.hean@foxmail.com

Received (收稿日期): 2025-08-07; Accepted (接受日期): 2025-11-18; Published online (网络出版日期): 2025-12-09.
URL: https://link.cnki.net/urlid/11.1809.S.20251208.1658.006

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

tobacco leaves. In this study, 240 flue-cured tobacco samples from various grades and production areas in Yunnan province, China, were analyzed. Hyperspectral images were collected across the 967.05–2561.33 nm spectral range, and an improved region of interest (ROI) extraction method—combining spectral difference analysis with superpixel clustering—was proposed to effectively eliminate interference from background, leaf veins, and irregular structures. Spectral data were preprocessed using a combination of standard normal variate (SNV) transformation and first derivative (FD), followed by principal component analysis (PCA) for dimensionality reduction. Partial least squares regression (PLSR) models were developed under a unified modeling framework to simultaneously predict six chemical components: nicotine, total sugars, reducing sugars, total nitrogen, potassium, and chlorine. The results showed that the SNV+FD preprocessing strategy could effectively enhance model performance. The coefficients of determination for cross-validation (Q^2) for nicotine, total sugars, reducing sugars, and total nitrogen all exceeded 0.89, with the lowest root mean square error of cross-validation (RMSECV) reaching 0.09. In an external test with 20 independent samples, the coefficients of determination for prediction (R^2) were 0.930, 0.908, 0.854, and 0.915, respectively. The average relative deviation (RD) was less than 5%, and the residual prediction deviation (RPD) values were all above 2.5, which verified the stability and predictive capability of the models. The established models enabled pixel-level visualization of chemical constituent distribution, revealing distinct heterogeneity across different leaf regions. The proposed ROI extraction and modeling method provides an efficient, non-destructive, and visual approach for evaluating tobacco quality and supporting precision processing.

Keywords: hyperspectral imaging (HSI); non-destructive testing; chemical component; region of interest (ROI); partial least squares regression (PLSR); tobacco

烟叶的化学成分是决定其香气特征、风味表现与吸食质量的核心指标, 直接影响烟草制品的工业可用性与消费体验[1]。在烟叶分级、工艺配方设计与产品质量控制过程中, 准确掌握烟叶中主要化学成分(如烟碱、总糖、还原糖、总氮、钾、氯等)的含量信息, 对提升产品稳定性和工艺优化具有重要意义[2]。目前, 烟叶主要化学成分的测定普遍依据烟草行业标准, 采用流动分析法[3-7]作为检测手段。该方法具备较强的重复性和较高的检测精度, 是目前工业检测的主流方案。此外, 也有研究和实验室采用气相色谱(gas chromatography, GC)、高效液相色谱(high performance liquid chromatography, HPLC)和质谱(mass spectrometry, MS)等高精度分析技术[8-9]。这些方法虽然在准确性和可靠性方面表现良好, 但普遍存在样品前处理复杂、检测周期长、运行成本高、试剂消耗大等问题, 且均为破坏性检测手段, 难以满足大批量、实时、无损检测的工业需求[10-11]。因此, 需探索更加高效、无损的检测技术, 以满足现代烟草生产对快速检测的实际需要。

光谱技术因其快速、无损的特性, 近年来在农产品成分分析领域得到了广泛应用[12-14]。其中, 近红外光谱(near infrared spectroscopy, NIRS)已在烟叶常规成分检测中取得显著进展。例如, 基于偏最小二乘回归的 NIRS 模型已实现对烟叶主要化学成分的有效预测[15]; 基于 NIRS 的数字化评价方法也在卷烟原料质量评估中展现出应用潜力[16]。然而, 传统 NIRS 技术通常基于点测或区域平均光谱, 无法揭示叶面结构的空间非均质特征, 同时仍依赖一定程度的样品预处理, 限制了其在工业现场的实时检测能力。

高光谱成像(hyperspectral imaging, HSI)技术融合了光谱与空间信息, 兼具光谱分辨率与空间分辨能力, 能够在不破坏样品的前提下实现化学成分的可视化分布, 因而成为农产品无损检测的前沿技术[17-19]。在烟叶检测领域, HSI 已展现出多方面的应用价值。例如, Lu 等[20]结合 HSI 与新型图像处理算法构建了烟叶成熟度判断模型, 有效提升了分类准确性; 李士静等[21]在多元散射校正基础上引入 F-Score 特征选择, 实现了不同等级烟叶的快速分类。郭文孟等[22]利用光谱吸收指数对烟叶泛青特征进行了量化表征; 范鹏飞等[23]结合 SPA-RF 模型成功识别烟叶霉变; 李智慧等[24]通过 SPA-SVM 模型实现了对烟叶中非烟物质的高效分类。这些高光谱成像技术的研究与应用, 有效推动了烟叶质量检测的数字化与智能化进程。然而, 现有研究在化学成分的定量检测方面仍存在不足, 尤其是在如何高效提取烟叶叶面感兴趣区域(region of interest, ROI)以获取纯净光谱数据并构建高精度、可推广的预测模型方面, 仍缺乏系统性的解决方案。

为解决上述问题, 本文提出基于高光谱成像的烟叶化学成分无损检测新方法。该方法融合光谱差值识别与超像素聚类的自动 ROI 提取策略, 结合主成分分析(principal component analysis, PCA)进行特征波段选择与光谱降维, 并采用偏最小二乘回归(partial least squares regression, PLSR)建立统一预测模型, 实现对烟碱、总糖、还原糖、总氮、钾和氯 6

种指标的高效建模与可视化反演。该方法不仅提升了检测的准确性与效率，还为烟叶品质评估、生产过程控制及配方优化提供了新的技术路径，具备良好的工业应用前景和推广价值。

1 材料与方法

1.1 试验材料与设备

试验材料为 2023 年云南省红河、保山、楚雄、昆明 4 个产区的初烤烟叶，品种为云烟 87，涵盖 B2F、C2F、C3F、C4F 四个等级。选取每个产区各等级样品 15 片，共计 240 片样品，确保数据的代表性和广泛性。

为确保光谱采集的一致性，所有烟叶样品在采集前置于恒温恒湿环境(ICH260 型箱体，美墨尔特公司，德国)中，以稳定叶片含水量并减小水分波动对光谱信号的干扰。设定条件为 45℃、相对湿度 80%，平衡 48 h。

采用烟草工艺重点实验室搭建的高光谱成像系统(图 1)采集高光谱数据，该系统核心设备包括 GalaFeld-V10E-AZ4 型及 Image-λ-N25E-HS 型短波红外高光谱相机，覆盖波段范围为 967.05~ 2561.33 nm，共计 288 个波段，光谱分辨率为 5.4 nm，空间分辨率为 384 × 893 像素，信噪比为 900 dB。辅助设备包括 OLES30 型成像镜头(Middleton Spectral Vision，美国)、4 个 50 W 卤素光源以及电动位移平台，并在密闭暗箱内完成图像获取，防止环境光干扰。通过 Spec View 软件设定并控制采集参数，所有样品均在黑白板校正后完成成像。数据采集后，所有处理分析步骤均在 MATLAB R2024a 平台完成。

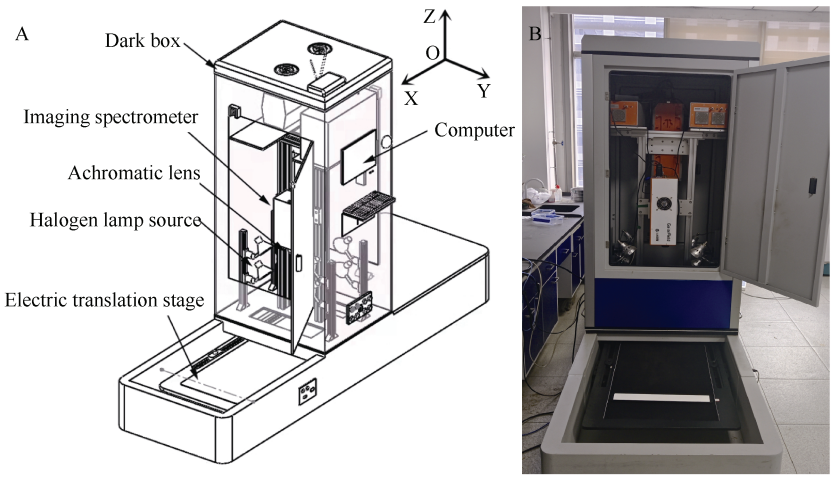


图 1 高光谱成像系统

Fig. 1 Hyperspectral imaging system

A: 高光谱成像系统示意图; B: 高光谱成像系统实物图。

A: schematic diagram of the hyperspectral imaging system; B: physical photograph of the hyperspectral imaging system.

1.2 数据获取

采集开始前将相机提前预热 30 min，将完成平衡后的叶片尽量展平置于暗箱系统内进行高光谱采集，以减少外界光干扰，提高测量的一致性和数据质量[25]。

由 Spec View 软件控制并完成光谱数据的采集。具体参数设置为：相机曝光时间 12 ms，电动移动平台前进速度 1.5 cm s^{-1}、回退速度 2.0 cm s^{-1}，以保证数据采集过程的均匀性和连续性。采集过程中，调整光源强度，避免图像过曝，并减少环境光影响，以获取高质量的光谱数据。为消除光照强度波动及相机暗电流噪声的影响，采集完成后需对原始数据进行黑白板校正，校正公式如下：

$$R = \frac{I - B}{W - B}$$

式中，R 为校正后高光谱数据，I 为原始采集的高光谱数据，B 为无光源条件下采集的暗电流噪声，W 为比反射率 99%的标准白板的高光谱数据。

为建立回归模型所需的真实标签值，本试验委托云南烟叶复烤有限公司陆良复烤厂依据烟草行业标准，对烟叶样品的 6 种主要化学成分(烟碱、总糖、还原糖、总氮、钾、氯)进行检测。

1.3 高光谱数据处理

1.3.1 感兴趣区域(ROI)提取 原始高光谱图像

中包含背景、叶脉、边缘阴影等非目标区域，这些区域的光谱信息与烟叶化学成分无关，若未经有效剔除直接建模，易引入噪声，降低建模的准确性与稳定性[26]。因此，需在建模前精确提取叶面感兴趣区域(ROI)，以提高光谱纯净度。

现有 ROI 提取方法包括固定阈值分割、手动标注及基于算法的自动分割。固定阈值分割通常基于单一波段反射率设定阈值，操作简便，但难以适应样本间颜色、质地及光照差异，易出现误识别；手动标注依赖主观判断，效率低且重复性差，不适合高通量处理。算法方法包括无监督聚类分割(如 K-means)、主成分分割(PC-based segmentation)及深度学习分割等，但均存在局限性。聚类分割在复杂纹理背景中易出现区域混叠，边界精度不足；主成分分割受限于主成分的光谱差异，易受噪声与波段选择影响；深度学习方法精度高，但依赖大规模训练数据且计算成本高，不利于大批量样本的实时处理[27]。针对烟叶主脉明显、边缘不规则、表面褶皱复杂的特点，本文提出一种基于光谱差异分析与结构聚类融合的自动 ROI 提取方法，如图 2 所示，包含“光谱差异掩膜构建”与“超像素聚类优化” 2 个阶段，以确保获得高纯度且一致性强的叶面区域。

首先，通过对典型样本不同区域光谱曲线进行对比，发现背景区域在1000~1800 nm 范围内反射率极低，而叶脉和叶面等区域在1107~1185 nm 和1364~1459 nm 2个近红外波段存在显著差异，具有较强的区分能力。基于此，构建上述波段的差值图谱，并以最大值与最小值差(max-min)作为光谱差异度指标，通过分位数阈值法(70%与65%)提取高变异区域，识别主脉、褶皱与非均质结构。同时，结合低波段(如1006.37 nm)反射率构建背景掩膜，得到初步 ROI 候选区域。

随后，采用简单线性迭代聚类(simple linear iterative clustering, SLIC)算法在光谱-空间联合特征空间内进行超像素分割，细化区域边界。聚类参数设置为超像素数 6，紧致度因子 0.4，迭代 10 次以保证收敛，得到空间-光谱一致性强的像素块。剔除背景与主脉区域后，获得精细化的 ROI 掩膜，并映射至原始高光谱图像以提取真实叶面像素光谱。

该方法无需人工干预，适用于大规模烟叶高光谱图像的自动化处理，显著提升 ROI 光谱纯净度与空间一致性，有效避免主脉、叶缘及背景的干扰，为构建高精度、稳定的光谱预测模型提供了可靠的输入数据。

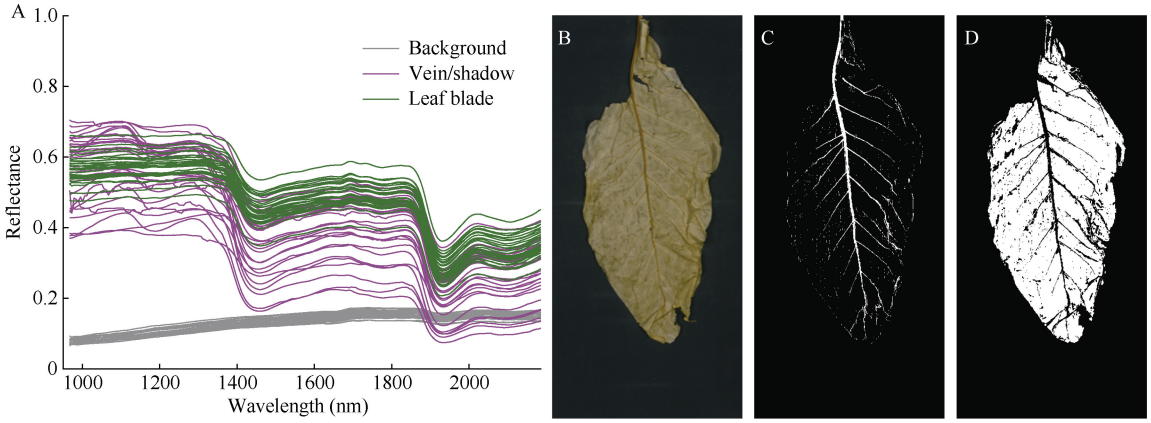


图 2　典型烟叶样本的感兴趣区域(ROI)提取过程

Fig. 2　Extraction process of regions of interest (ROI) in a typical tobacco leaf sample

A: 不同区域光谱曲线对比; B: 高光谱假彩色图; C: 差值识别叶脉图; D: 感兴趣区域掩膜。

A: spectral curve comparison of different regions; B: hyperspectral pseudo-color composite image; C: spectral difference-based midrib identification image; D: region of interest mask.

1.3.2　*光谱数据预处理与降维*　　原始高光谱数据易受散射效应和基线漂移干扰。本研究采用标准正态变量变换(standard normal variate, SNV)和一阶导数(first derivative, FD)对光谱数据进行预处理，并比较了单一方法与组合方法(SNV+FD)的效果。SNV 用于校正散射影响，提高样品间可比性；FD 用于增强光谱特征并消除基线漂移, FD 计算采用 Savitzky-Golay 滤波(窗口宽度 15 点，多项式阶次 2) [28]。

为降低模型复杂度并减少变量相关性，使用主成分分析法对预处理后的光谱数据进行降维。PCA 通过线性变换将高维光谱映射至无关主成分空间，在保留主要信息的同时降低数据维度，减少噪声与

共线性特征, 从而提升模型的泛化能力[29]。

1.4 预测模型构建与评价

在回归建模阶段, 选用偏最小二乘回归(PLSR)作为核心算法。PLSR 适用于处理高维、多共线性数据, 尤其适合“样本数远小于变量数”的高光谱应用场景, 通过最大化预测变量(X)与响应变量(Y)协方差提取潜变量(latent variables, LVs), 建立线性回归模型[30]。

针对多个化学组分的预测任务, 对比 2 种建模策略: 一种为单指标独立建模, 以优化单项预测性能; 另一种为基于统一特征集构建多组分协同预测模型, 以提升模型部署效率与实际应用价值。为保证模型稳健性与泛化能力, 在建模前使用马氏距离(mahalanobis distance)在 99%置信区间下进行异常值检测与剔除[31], 获得 236 份有效样本, 其中 216 份用于建模训练(训练集), 20 份用于外部验证(测试集)。建模过程中使用 K 折交叉验证(K-fold cross validation, K = 10), 将训练集进一步划分为训练数据与验证数据, 以确保模型在不同数据划分下具有良好的泛化能力。

在模型构建过程中, 分别基于原始光谱及不同预处理光谱建立 PLSR 模型, 并以训练集决定系数(coefficient of determination for the training set, R_c^2)、验证集决定系数(coefficient of determination for the validation set, R_v^2)、交叉验证决定系数(coefficient of determination for cross-validation, Q^2)、校正均方根误差(root mean square error of calibration, RMSEC)及交叉验证均方根误差(root mean square error of cross validation, RMSECV)等指标评估模型拟合效果与稳定性。最终选取性能最优的模型, 在独立测试集上进行外部验证, 并计算预测决定系数(coefficient of determination for prediction, R^2)、相对标准偏差(relative standard deviation, RSD)及相对预测偏差(ratio of performance to deviation, RPD), 评价模型的泛化性能及实际应用价值。

1.5 化学成分的可视化预测

在高光谱建模完成后, 回归模型可应用于样品图像中每一个像素点的光谱向量, 逐点计算化学成分的浓度, 实现其叶面分布的可视化预测。具体步骤为: 首先提取样本感兴趣区域(ROI)内各像素的高光谱数据; 随后将预处理和降维后的光谱特征输入到已建立的 PLSR 模型中, 利用模型回归系数计算每个像素的目标化学成分浓度。最后, 将预测结果转换为颜色热图, 不同颜色表示不同浓度水平, 从而直观展现烟叶表面各部位化学组分的含量分布。

2 结果与分析

2.1 化学成分参考值

为分析样品集中 6 种主要化学成分的分布特征, 对保留的 236 份烟叶样本进行统计汇总(表 1)。从统计结果来看, 烟碱、糖类和总氮在样品集中表现出适中的含量及相对均匀的变异性, 说明样品间具有良好的可区分性。糖类组分在含量及波动趋势上较为接近, 提示可能存在一定的协同变化关系。总氮整体稳定, 而钾和氯的分布差异较大, 尤其氯在样品间变化显著, 显示其在建模中可能面临信号响应不足的挑战。通过变异系数分析, 可以初步评估各组分的数值稳定性及潜在建模难度, 可为后续光谱建模提供基础参考。

2.2 感兴趣区域(ROI)提取效果

为验证所提出方法的有效性, 研究对比了常见的固定阈值提取方法、主成分分割方法、K-means 聚类提取方法与本方法在 ROI 提取上的效果(图 3)。结果显示, 不同方法在提取精度和一致性方面存在显著差异。固定阈值提取方法能够较好地区分烟叶

表 1 烟叶化学成分参考值统计结果

Table 1 Statistical results of reference values for chemical components in tobacco leaves (%)

化学成分 Chemical component	平均值 Mean	标准差 Standard deviation	变异系数 Coefficient of variation	最大值 Max.	最小值 Min.
烟碱 Nicotine	2.45	0.65	26.53	4.59	1.33
总糖 Total sugars	35.33	4.64	13.13	43.37	18.22
还原糖 Reducing sugars	31.72	4.27	13.46	39.76	15.97
总氮 Total nitrogen	2.01	0.33	16.42	3.23	1.44
钾 Potassium	2.00	0.30	15.00	3.05	1.29
氯 Chlorine	0.27	0.21	77.78	1.28	0

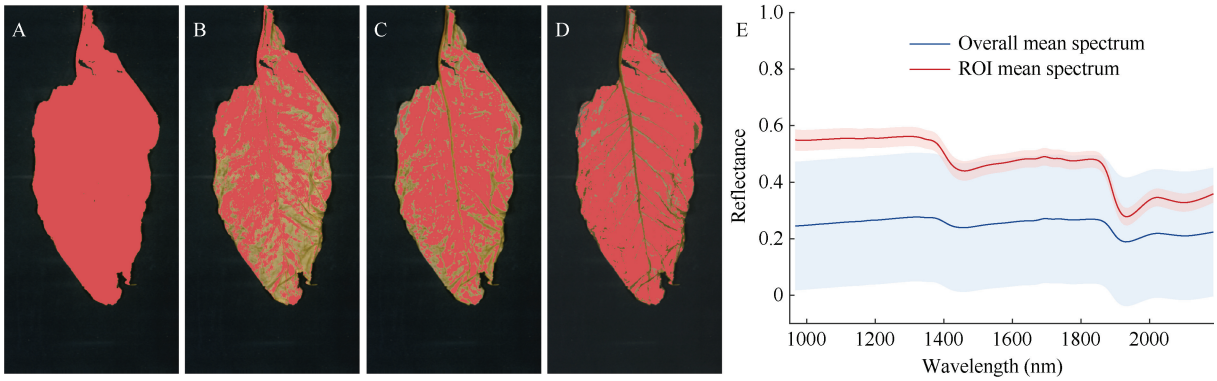


图 3　典型烟叶样本的不同 ROI 提取方法效果对比

Fig. 3　Comparison of the effectiveness of different ROI extraction methods in a typical tobacco leaf sample

A: 固定阈值提取方法; B: 主成分分割提取方法; C: K-means 聚类提取方法; D: 本试验提出的方法; E: 本方法 ROI 提取前后光谱反射率平均值及误差带。

A: fixed threshold extraction method; B: PCA segmentation extraction method; C: K-means clustering extraction method; D: the method proposed in this article; E: mean spectral reflectance and error band before and after ROI extraction by the proposed method.

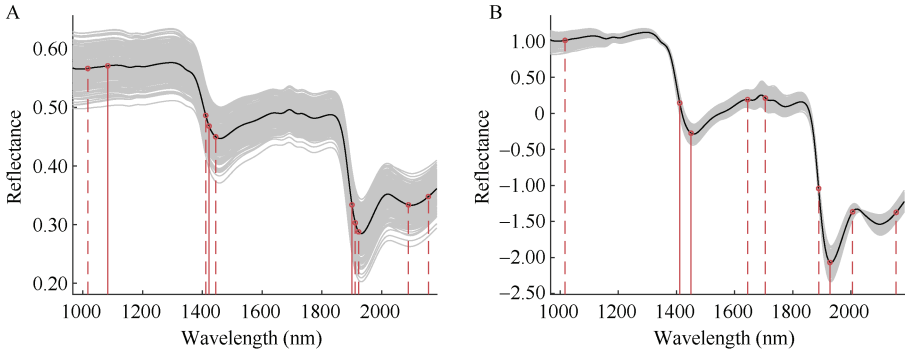
主体与背景，但难以识别叶片内部的细节结构；主成分分割和 K-means 聚类提取方法虽也能区分背景，但并不能有效识别到完整的叶面结果，且难以保证对二级叶脉和复杂褶皱区域的准确提取。相比之下，本文提出的基于光谱差异与超像素聚类融合的 ROI 提取方法，不仅能够完整提取叶面主体，还能有效识别主脉、二级叶脉及褶皱阴影区域，表现出更高的提取精度与稳定性。

图 3-E 展示了所提出方法的 ROI 提取前后样本光谱反射率平均值及误差带的比较结果。可以看到，ROI 提取后光谱曲线更加平滑，误差带范围显著收窄，表明噪声干扰显著减弱。定量分析进一步表明，提取前样本间平均光谱相关性仅为 18.07%，而 ROI 提取后相关性显著提升至 99.61%，非 ROI 部分相关性为 63.28%。以上结果证明，所提出的 ROI 提取方法能够显著提升光谱数据的一致性与纯净度，有效剔除非目标区域信息，为后续基于高光谱成像的定量分析与建模提供了高质量且稳定的输入数据。

2.3　光谱预处理与特征波长选择

图 4 展示了各预处理策略对光谱形态的影响。SNV 预处理使光谱整体趋于均值对齐，有效减少了由样品间物理差异导致的散射影响，增强了样本间可比性；FD 处理突出吸收波段特征，尤其在 1100~1600 nm 区间的微弱特征明显增强，但同时放大了部分高频噪声；SNV+FD 组合兼具两者优势，实现了背景干扰的有效抑制和化学特征的增强，光谱曲线更平滑稳定。

PCA 分析(表 2)进一步揭示了各预处理策略对光谱信息结构的影响。原始数据主成分载荷聚焦于 1422.56 nm (水分第一泛音区)、1083.70 nm (C-H 区)及 1900.29 nm (水分强吸收带)等典型有机官能团吸收区。SNV 处理后，主载荷向 1928.06、1411.45、1450.33 nm 集聚，强化了 C-H 第一泛音(1429~1493 nm)和 O-H 官能团的吸收特征。FD 处理强调 1872.51、1411.45、2122.49 nm 波段，有利于微弱羰



(图 4)

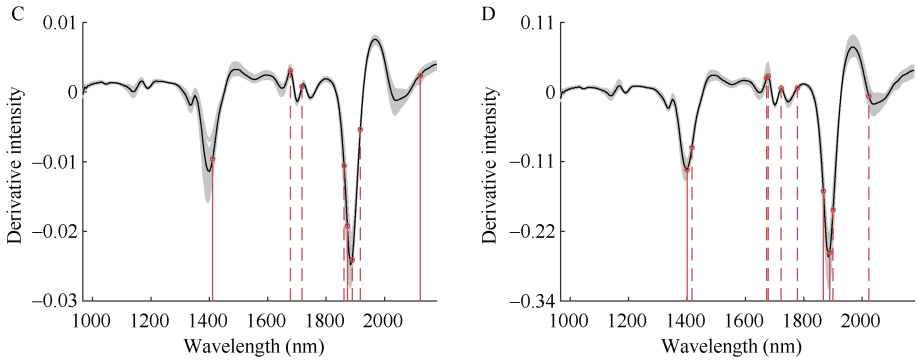


图 4 不同预处理光谱与前 10 个主成分对应波长

Fig. 4 Spectra of samples with different preprocessing methods and wavelengths corresponding to the first ten principal components

A: 原始光谱; B: 标准正态变换(SNV)预处理; C: 一阶导数(FD)预处理; D: 标准正态变换+一阶导数预处理。

A: raw spectra; B: spectra after standard normal variate (SNV) preprocessing; C: spectra after first derivative (FD) preprocessing; D: sepectra after SNV and FD combined preprocessing.

表 2 不同预处理方法主成分分析结果

Table 2 Principal component analysis results of different preprocessing methods

预处理方法 Preprocessing method	主成分 Principal component	波长 Wavelength (nm)	载荷值 Loading value	贡献率 Contribution rate (%)
RAW	PC1	1422.56	0.0814	81.77
	PC2	1083.70	0.0975	17.19
	PC3	1900.29	0.1598	0.62
	PC4	2089.16	0.1434	0.26
	PC5	1444.78	–0.1349	0.11
SNV	PC1	1928.06	0.1404	65.22
	PC2	1411.45	0.1499	25.70
	PC3	1450.33	0.1602	7.54
	PC4	1017.04	0.1113	0.60
	PC5	1644.76	0.1270	0.35
FD	PC1	1872.51	0.2968	3.57
	PC2	1411.45	0.2237	15.76
	PC3	2122.49	0.1721	7.22
	PC4	1889.18	0.3345	1.63
	PC5	1916.95	0.2346	0.66
SNV+FD	PC1	1889.18	0.3541	54.40
	PC2	1866.96	0.2746	34.25
	PC3	1400.34	0.2338	8.74
	PC4	1672.53	0.2128	0.88
	PC5	1678.09	0.3012	0.51

RAW: 原始数据; SNV: 标准正态变换; FD: 一阶导数。RAW: raw data; SNV: standard normal variate; FD: first derivative.

基和酯类特征识别。SNV+FD 联合处理下, 前 3 主成分累计贡献率高达 97.39%, 关键波段集中于 1889.18、1866.96、1400.34 nm, 显著增强了对亲水基团(O-H/N-H)及糖类结构区间的响应能力[32]。

综合来看, SNV+FD 预处理不仅显著压制了系统噪声和散射影响, 还提升了对烟叶中糖类、氮素和水分等重要成分的敏感度与波段聚集度, 充分体现了物理化学特征与波段选择的高度一致性。此方法在多组分建模中表现出优异的鲁棒性和解释性, 为后续回归模型提供了可靠且富有信息量的输入数据。

2.4 化学成分预测模型

2.4.1 分别建模策略结果 在分别建模策略下, 针对烟碱、总糖、还原糖、总氮、钾和氯 6 种主要

化学组分分别建立 PLSR 模型, 不同光谱预处理方式的模型性能指标见表 3。

从整体结果来看, FD 与 SNV+FD 预处理在大多数化学成分的建模中表现最优。以烟碱为例, SNV+FD 预处理下模型的 R_c^2 达到 0.92, R_v^2 为 0.89, RMSEP 为 0.16, 对应的 Q^2 为 0.94, 说明模型具有良好的拟合性与稳健性。糖类成分的建模结果同样突出, 其中总糖和还原糖在 FD 和 SNV+FD 条件下的 R_v^2 分别超过 0.90 和 0.85, RMSECV 均小于 1.50, 表明碳氢键(C-H)与羟基(O-H)等官能团在近红外区域具有明显的光谱吸收特征, 从而实现较高的建模精度。总氮在 FD 与 SNV+FD 条件下同样表现出色, R_v^2 分别为 0.88 和 0.86, Q^2 均为 0.93, RMSECV 均为 0.09, 说明氨基(N-H)等官能团在特征波段具有良好的响应特性, 模型能有效捕捉其光谱变化规律。相比之下, 钾和氯的建模性能明显较弱。钾在 SNV+FD 条件下的 Q^2 仅为 0.62, 氯的 Q^2 为 0.71, 均显著低于有机成分。这主要归因于无机离子在近红外波段的吸收信号较弱, 其特征峰不明显, 加之在样品中的含量较低、分布差异较大, 导致模型对其变化敏感性不足。

综合比较可知, FD 及 SNV+FD 预处理方法能显著提升主要含碳、氢及氮有机组分(如烟碱、糖类、总氮)的定量建模性能, 而对无机离子类组分(如钾、氯)的增强效果有限。

表 3　各化学成分不同预处理方法模型性能对比

Table 3　Comparison of model performance under different preprocessing methods for each chemical component

化学成分 Chemical component	预处理方法 Preprocessing method	主成分数 Number of principal components	R_c^2	RMSEC	R_v^2	RMSEP	Q^2	RMSECV
烟碱 Nicotine	RAW	4	0.65	0.38	0.41	0.32	0.75	0.32
	SNV	10	0.82	0.28	0.66	0.23	0.87	0.23
	FD	15	0.92	0.18	0.88	0.17	0.94	0.17
	SNV+FD	**15**	**0.92**	**0.18**	**0.89**	**0.16**	**0.94**	**0.16**
总糖 Total sugars	RAW	5	0.83	1.90	0.80	1.97	0.83	1.93
	SNV	10	0.85	1.81	0.83	1.82	0.87	1.66
	FD	**15**	**0.90**	**1.46**	**0.91**	**1.34**	**0.91**	**1.43**
	SNV+FD	17	0.91	1.42	0.90	1.38	0.90	1.48
还原糖 Reducing sugars	RAW	5	0.83	1.74	0.73	2.10	0.83	1.75
	SNV	9	0.85	1.64	0.80	1.83	0.89	1.44
	FD	**19**	**0.91**	**1.28**	**0.86**	**1.55**	**0.90**	**1.33**
	SNV+FD	12	0.91	1.31	0.85	1.56	0.90	1.36
总氮 Total nitrogen	RAW	5	0.77	0.16	0.68	0.18	0.78	0.15
	SNV	10	0.88	0.12	0.83	0.13	0.91	0.10
	FD	**12**	**0.92**	**0.09**	**0.88**	**0.11**	**0.93**	**0.09**
	SNV+FD	12	0.92	0.10	0.86	0.12	0.93	0.09
钾 Potassium	RAW	5	0.37	0.24	0.09	0.23	0.35	0.24
	SNV	10	0.49	0.22	0.50	0.17	0.51	0.21
	FD	15	0.52	0.21	0.32	0.20	0.64	0.18
	SNV+FD	**14**	**0.55**	**0.20**	**0.51**	**0.17**	**0.62**	**0.18**
氯 Chlorine	RAW	5	0.15	0.20	0.33	0.12	0.23	0.19
	SNV	10	0.36	0.18	0.24	0.13	0.64	0.13
	FD	13	0.68	0.12	0.81	0.06	0.70	0.12
	SNV+FD	**14**	**0.69**	**0.12**	**0.57**	**0.10**	**0.71**	**0.12**

R_c^2: 训练集决定系数; RMSEC: 训练集均方根误差; R_v^2: 验证集决定系数; RMSEP: 验证集均方根误差; Q^2: 交叉验证决定系数; RMSECV: 交叉验证均方根误差。RAW: 原始数据; SNV: 标准正态变换; FD: 一阶导数。加粗字体表示性能最佳的处理方法及参数。

R_c^2: coefficient of determination for the training set; RMSEC: root mean square error of the training set; R_v^2: coefficient of determination for the validation set; RMSEP: root mean square error of the validation set; Q^2: coefficient of determination for cross-validation; RMSECV: root mean square error of cross-validation. RAW: raw data; SNV: standard normal variate; FD: first derivative. Bold font indicates the processing method and parameters.

2.4.2 多组分统一建模结果 多组分统一建模策略通过使用统一特征集同时预测多个化学成分, 在保持建模效率的同时评估其精度表现。不同光谱预处理方法统一模型的性能指标见表 4。

整体结果表明, SNV+FD 在统一建模条件下表现最优, 兼顾了预测精度与模型稳健性。以烟碱为例, SNV+FD 处理下模型的 R_c^2 和 R_v^2 均达到 0.93, RMSEP 为 0.15, Q^2 值为 0.92, 与分别建模的最优结果相近, 说明统一建模并未显著削弱对关键组分的预测能力。对于总糖、还原糖和总氮, SNV+FD 预处理同样取得优异的结果, 其 R_c^2 与 R_v^2 均超过 0.85, Q^2 值接近 0.90, RMSECV 分别控制在 1.50 (糖类)或 0.10 (氮)以内, 表明该策略在多目标预测中依然能保持较高拟合度与预测稳定性。相比之下, 无机离子类成分钾和氯的模型表现相对较弱。钾在 SNV+FD 条件下的 Q^2 仅为 0.52, 而氯的 Q^2 可达 0.71, 预测性能虽略低于有机组分, 但仍体现出一定的模型稳定性。

在不同预处理方法间比较发现, 未经预处理(RAW)及单独 SNV 处理的模型在多数组分上拟合不足、预测偏差较大, 尤其在钾和氯的预测中误差明显; FD 虽在部分成分(如氯)中表现良好, 但在整体一致性上略逊于 SNV+FD。而 SNV+FD 通过结合散射校正与特征增强, 显著改善了光谱稳定性, 使模型在不同成分间保持较高的统一性与鲁棒性。

表 4 各化学成分不同预处理方法统一模型性能对比

Table 4 Comparison of unified model performance under different preprocessing methods for each chemical component

预处理方法 Preprocessing method	化学成分 Chemical component	主成分数 Number of principal components	R_c^2	RMSEC	R_v^2	RMSEP	Q^2	RMSECV
RAW	烟碱 Nicotine	5	0.75	0.32	0.51	0.40	0.74	0.33
	总糖 Total sugars	5	0.83	1.90	0.80	1.97	0.83	1.93
	还原糖 Reducing sugars	5	0.83	1.74	0.73	2.10	0.83	1.75
	总氮 Total nitrogen	5	0.77	0.16	0.68	0.18	0.77	0.16
	钾 Potassium	4	0.37	0.24	0.09	0.23	0.35	0.24
	氯 Chlorine	3	0.14	0.20	0.22	0.13	0.16	0.20
SNV	烟碱 Nicotine	10	0.82	0.28	0.66	0.34	0.80	0.29
	总糖 Total sugars	8	0.85	1.82	0.83	1.82	0.84	1.88
	还原糖 Reducing sugars	8	0.85	1.64	0.79	1.87	0.84	1.69
	总氮 Total nitrogen	10	0.88	0.12	0.83	0.13	0.87	0.12
	钾 Potassium	10	0.49	0.22	0.50	0.17	0.43	0.23
	氯 Chlorine	8	0.36	0.18	0.27	0.13	0.32	0.18
FD	烟碱 Nicotine	16	0.93	0.17	0.90	0.18	0.82	0.18
	总糖 Total sugars	17	0.91	1.43	0.90	1.39	0.89	1.53
	还原糖 Reducing sugars	15	0.91	1.29	0.86	1.50	0.90	1.36
	总氮 Total nitrogen	18	0.94	0.08	0.91	0.09	0.92	0.09
	钾 Potassium	19	0.59	0.19	0.58	0.16	0.51	0.21
	氯 Chlorine	19	0.77	0.10	0.71	0.08	0.70	0.12
SNV+FD	**烟碱 Nicotine**	**19**	**0.93**	**0.17**	**0.93**	**0.15**	**0.92**	**0.18**
	总糖 Total sugars	**14**	**0.90**	**1.43**	**0.90**	**1.36**	**0.89**	**1.51**
	还原糖 Reducing sugars	**11**	**0.91**	**1.31**	**0.85**	**1.58**	**0.90**	**1.35**
	总氮 Total nitrogen	**19**	**0.94**	**0.08**	**0.91**	**0.09**	**0.93**	**0.09**
	钾 Potassium	**19**	**0.58**	**0.20**	**0.50**	**0.17**	**0.52**	**0.21**
	氯 Chlorine	**19**	**0.76**	**0.11**	**0.70**	**0.08**	**0.71**	**0.12**

缩写同表 3。加粗字体表示性能最佳的处理方法及参数。

Abbreviations are the same as those given in Table 3. Bold font indicates the best-performing method and parameters.

综上所述, SNV+FD 预处理下的多组分统一建模策略能够在保证整体性能的前提下减少建模复杂度, 为高光谱成像在烟叶多成分同步检测中的实际应用提供了可行的技术路径。

2.4.3　模型选择与泛化能力评价　　基于前述建模结果, 最终选取以 SNV+FD 为预处理方式、采用多组分统一建模策略的 PLSR 模型作为最优方案。为进一步验证模型在未知样本上的泛化能力与应用可靠性, 选取 20 份未参与建模的烟叶样品进行外部测试, 并依据表 5 所列指标对模型的外部预测性能进行评估。

由表 5 可知, 模型在烟碱、总糖、还原糖和总氮 4 种有机组分的预测中表现出较强的解释力和稳定性, R^2 分别为 0.930、0.908、0.854 和 0.915, 均高于 0.850, 平均相对误差控制在 5%以内, RSD 维持在 1.35%~2.71%之间, 表明预测结果具有良好的重复性和一致性。RPD 均高于 2.5, 表明模型已具备较高的定量分析能力, 可满足烟叶主要有机成分的快速、可靠检测需求。

相较之下, 对无机离子类成分钾和氯的预测性能明显偏低。钾的 R^2 仅为 0.488, 氯虽略高(0.698), 但 RPD 分别为 1.45 与 1.88, 仍处于定性预测水平。两者的 RD 和 RSD 均显示误差较高且离散性大, 反映出无机离子在近红外波段的光谱响应较弱、信噪比低, 导致模型预测稳定性不足, 这与训练阶段的结论一致。

表 5　外部样本模型性能指标
Table 5　Performance indicators of the model on external samples

化学成分 Chemical component	R^2	RD 均值 RD mean (%)	RSD (%)	RPD (%)
烟碱 Nicotine	0.930	4.91	1.35	3.90
总糖 Total sugars	0.908	3.11	2.50	3.31
还原糖 Reducing sugars	0.854	4.36	2.71	2.63
总氮 Total nitrogen	0.915	3.63	2.59	3.43
钾 Potassium	0.488	7.23	2.72	1.45
氯 Chlorine	0.698	31.91	36.57	1.88

R^2: 预测决定系数; RD: 相对偏差; RSD: 相对标准偏差; RPD: 相对预测偏差。
R^2: coefficient of determination for prediction; RD: relative difference; RSD: relative standard deviation; RPD: residual prediction deviation.

图 5 展示了统一模型在训练集与外部测试集上的拟合与预测表现。可以看到, 烟碱、总糖、还原糖和总氮的拟合曲线斜率接近 1.0, 截距较小, 训练与预测数据均呈良好线性关系, 说明模型拟合充分、泛化性强。虽然钾和氯的拟合偏差较大, 但整体预测趋势与真实值保持一致, 未出现系统性偏移。进一步的残差分布分析显示, 有机组分残差集中且分布均匀, 表明模型具有较好的稳定性; 而无机组分残差范围稍宽、分布离散, 体现出预测的不确定性较高。

综上所述, 采用 SNV+FD 预处理结合统一 PLSR 建模策略的模型在主要有机组分预测中实现了高精度与高稳健性的平衡。该方法不仅保持了较强的泛化能力与可靠性, 同时显著提高了建模与预测效率, 能够满足烟叶主要有机化学成分的快速、同步、无损检测需求。尽管在无机离子类指标上仍存在一定性能瓶颈, 但该研究为高光谱成像技术在烟叶多指标在线表征与质量控制中的工程化应用提供了坚实的建模基础与方法论支撑。

2.5　化学成分分布可视化

在前述建模结果基础上, 将优化后的 PLSR 模型应用于像素级高光谱数据, 实现对烟叶 6 种主要化学成分含量的叶面分布反演与可视化, 结果如图 6 所示。

从图 6 中可以看出, 烟叶不同区域的化学成分呈现出明显的空间异质性。其中, 烟碱含量在主脉附近相对较低, 而在叶缘区域升高; 总糖和还原糖含量整体表现为由叶基向叶尖递增的趋势; 总氮分布则呈相反趋势, 从叶基至叶尖逐渐降低; 钾与氯在叶面上的分布模式较为相似, 二者均表现为由主脉向叶缘逐渐增加的空间梯度。

总体而言, 像素级的高光谱反演结果直观揭示了烟叶主要化学成分在不同区位的含量分布规律, 验证了高光谱成像结合化学计量学建模在烟叶样品化学成分可视化与空间异质性表征中的有效性。该方法不仅实现了烟叶内部主要化学成分的无损、同步定

量映射，也为后续烟叶品质评价、分级控制及加工工艺调控提供了可视化的理论依据与数据支撑。

3 讨论

3.1 模型综合性能分析

本研究建立了基于高光谱成像的烟叶化学成分无损检测方法，通过 ROI 提取、光谱预处理与降维以及 PLSR 建模，实现了快速、准确的多组分预测。ROI 提取显著提升了光谱数据质量。提取后样本间的光谱相关性由 18.07%提升至 99.61%，有效减少了背景与非目标区域的干扰，提升了数据一致性与建模可靠性。相比传统手动选区或其他自动化方法，

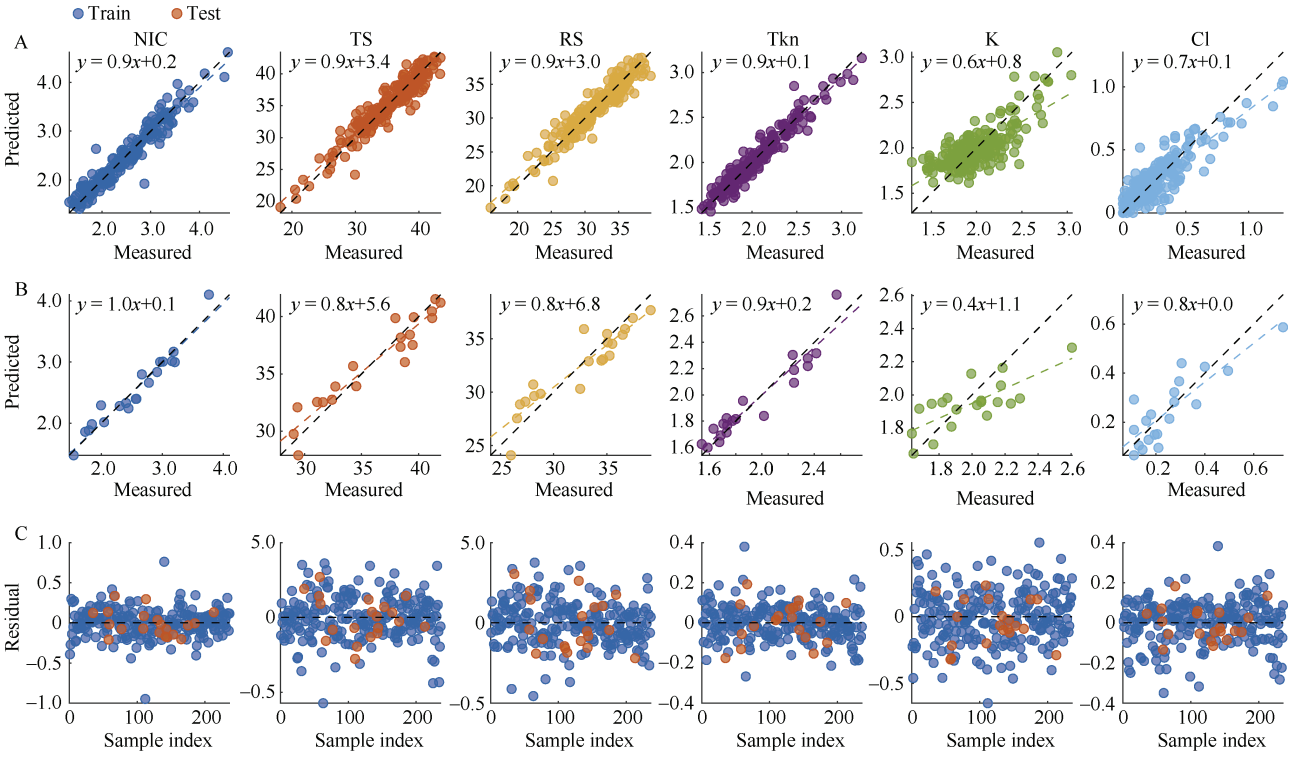
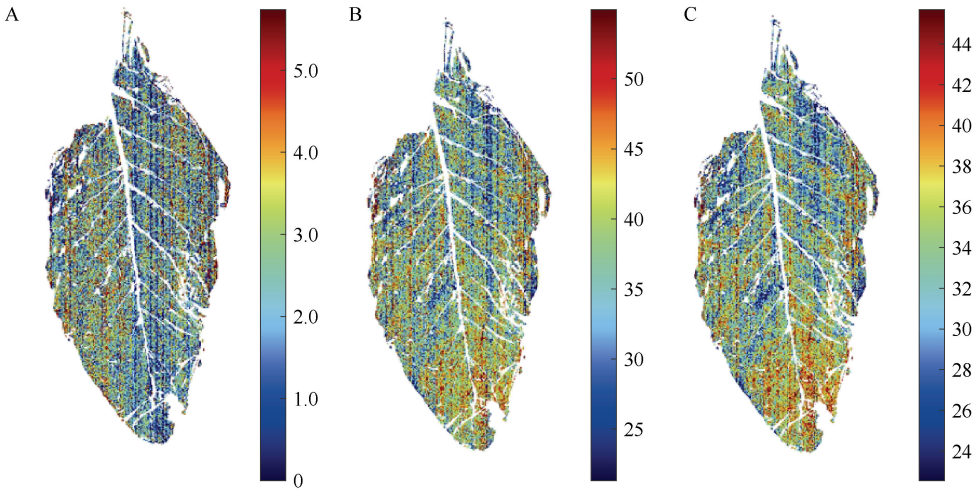


图 5 烟叶主要化学成分预测模型性能

Fig. 5 Performance of prediction model for major chemical components of tobacco leaves

A: 训练集数据分布及拟合; B: 测试集数据分布及拟合; C: 模型残差分布。NIC: 烟碱; TS: 总糖; RS: 还原糖; Tkn: 总氮; K: 钾; Cl: 氯。

A: distribution and fitting curve of training set data; B: distribution and fitting curve of testing set data; C: residual distribution of model. NIC: nicotine; TS: total sugars; RS: reducing sugars; Tkn: total nitrogen; K: potassium; Cl: chlorine.



(图 6)

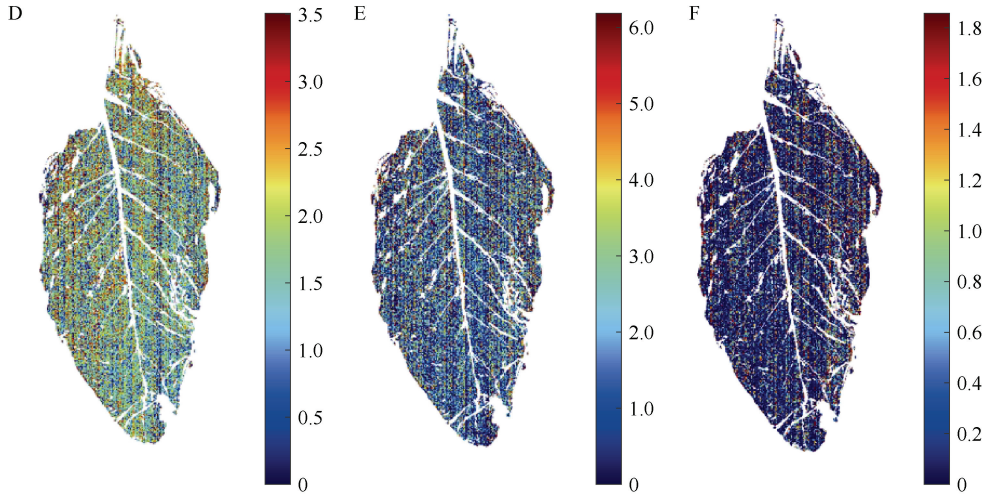


图 6　烟叶表面主要化学分布反演

Fig. 6　Visualization of major chemical component distribution on the surface of tobacco leaves

A: 烟碱分布; B: 总糖分布; C: 还原糖分布; D: 总氮分布; E: 钾分布; F: 氯分布。

A: nicotine distribution; B: total sugars distribution; C: reducing sugars distribution; D: total nitrogen distribution; E: potassium distribution; F: chlorine distribution.

该方法具备高通量、自动化、无需样本标注、分割细致全面等优势。

降维处理显著优化了模型的计算效率, 未降维时全波段建模平均运行时间为 41.01 s, 而采用主成分分析(PCA)降维后, 平均运行时间缩短至不足 0.01 s, 效率提升超过 4000 倍。同时, 降维对预测精度的影响微弱, 例如, 烟碱组分 Q^2 仅由 0.93 下降至 0.92, 表明降维可在保持化学信息完整性的前提下显著提升计算效率并降低过拟合风险。散射校正与导数预处理进一步强化了关键波段特征, 显著提升了主要有机组分的预测性能。NIC、TS 和 Tkn 的 Q^2 分别达到 0.92、0.89 和 0.93, 均优于未经预处理的模型, 验证了该预处理策略在多组分建模中的有效性。

相比之下, 钾和氯的预测性能显著低于主要有机组分, 这主要源于两方面因素: 一是参考值分布偏低且变异性大, 二是在近红外波段缺乏显著吸收特征, 即便经过 ROI 提取与预处理, 光谱信号仍较弱, 导致模型在训练与外部预测中稳定性不足。为提升预测精度, 未来可从两方面进行改进: 其一, 拓展光谱测量至中红外或短波红外, 以增强无机离子及低浓度组分的响应能力; 其二, 优化样本采集策略, 增加低浓度和极端值样本, 以提升模型在不同浓度范围的预测能力。这些改进有望在保持主要有机组分高精度预测的基础上, 同时改善低响应组分的检测可靠性。

综合来看, 本研究提出的基于 ROI 提取、光谱预处理及降维的 PLSR 建模方法, 在主要有机组分预测中具有较高的解释力与稳定性。训练样本交叉验证决定系数(Q^2)均大于 0.89, 交叉验证均方误差(RMSECV)均小于 1.55, 外部测试样本决定系数(R^2)高于 0.85, 平均相对偏差(RD)低于 5%, 相对预测偏差(RPD)高于 2.5, 显示出较高的应用潜力, 为高光谱成像在烟叶品质快速检测领域的实际应用提供了坚实的技术支撑。

3.2　叶面化学成分分布与文献对比

基于像素级高光谱反演结果, 研究揭示了烟叶主要化学成分在不同区位的分布特征。烟碱含量在叶脉附近较低, 而在叶缘区域较高, 这一结果与贺帆等[33]和王小东等[34]的结论一致; 总糖和还原糖含量从叶基到叶尖呈上升趋势, 与殷全玉等[35]的“先升后降”的趋势部分吻合; 总氮含量从叶基到叶尖逐渐降低, 与王小东等[34]的结论相符, 但与贺帆等[33]和殷全玉等[35]的报道存在差异; 钾、氯含量从近主脉向叶缘方向递增, 这一趋势在部分文献中有不同的观察结果。

需要注意的是, 本研究的预测模型是基于整体样本的平均光谱建立的, 与真实像素级参考值存在理论偏差。这种偏差在像素级分布精度上可能导致局部浓度的轻微偏离, 尤其在叶尖和叶缘等面积较小的区域表现更明显。然而, 这种理论偏差对主要趋势和整体分布规律影响有限, 仍能有效反映不同区位间的化学成分分布特征。同时, 样品来源、品

种差异以及叶片区划分操作误差也可能造成局部分布结果与文献略有差异, 但整体趋势与部分前人研究结果保持一致, 验证了高光谱成像在揭示烟叶化学成分分布规律方面的可行性。

为了进一步提升像素级预测与反演精度, 未来可探索基于深度学习的光谱-图像融合建模方法, 充分整合光谱特征与空间信息, 实现多组分同步预测和局部区域高精度反演, 更准确地揭示烟叶化学成分在不同区位的分布规律。

4 结论

本研究选取云南不同产区、不同等级的 240 片初烤烟叶, 采集 967.05~2561.33 nm 范围的高光谱图像, 构建了一种结合 ROI 提取、SNV+FD 预处理以及 PCA 降维的 PLSR 建模方法, 实现烟叶主要化学成分的快速、无损检测与可视化分布。基于光谱差值与 SLIC 聚类的 ROI 提取策略, 有效提升了光谱数据纯度与建模一致性。SNV+FD 预处理结合 PCA 降维, 实现了高效且精准的多组分预测, 尤其在烟碱、糖类及氮素的检测中表现优异(R^2 > 0.85, RD < 5%, RPD > 2.5)。像素级反演揭示了烟叶化学成分的可视化分布特征, 验证了叶片不同区位的含量差异规律。该方法不仅具备快速、无损、自动化的检测优势, 还具有重要的工程应用潜力, 为实现烟叶非破坏性质量评价、区域分级加工及智能化分选提供了理论依据和方法支撑。

References

[1] Soares F L F, Marcelo M C A, Porte L M F, et al. Inline simultaneous quantitation of tobacco chemical composition by infrared hyperspectral image associated with chemometrics. Microchem J, 2019, 151: 104225.

[2] 郝贤伟, 黄文勇, 徐志强, 等. 基于近红外光谱技术的云南片烟综合质量评价. 中国烟草科学, 2022, 43(2): 58–63.
Hao X W, Huang W Y, Xu Z Q, et al. Comprehensive quality evaluation of YunnanTobacco strips based on near infrared spectroscopy. Chin Tob Sci, 2022, 43(2): 58–63 (in Chinese with English abstract).

[3] 国家烟草专卖局. 烟草及烟草制品-总植物碱的测定-连续流动法, YC/T 160-2002. 北京: 中国标准出版社, 2002.
State Tobacco Monopoly Administration. Tobacco and Tobacco Products-Determination of Total Alkaloids-Continuous Glow Method, YC/T 160-2002. Beijing: Standards Press of China, 2002 (in Chinese).

[4] 国家烟草专卖局. 烟草及烟草制品-水溶性糖的测定-连续流动法, YC/T 159-2019. 北京: 中国标准出版社, 2019.
State Tobacco Monopoly Administration. Tobacco and Tobacco Products-Determination of Water Soluble Sugars-Continuous Flow Method, YC/T 159-2019. Beijing: Standards Press of China, 2019 (in Chinese).

[5] 国家烟草专卖局. 烟草及烟草制品-总氮的测定-连续流动法, YC/T 161-2002. 北京: 中国标准出版社, 2002.
State Tobacco Monopoly Administration. Tobacco and Tobacco Products-Determination of Total Nitrogen-Continuous Flow Method, YC/T 161-2002. Beijing: Standards Press of China, 2002 (in Chinese).

[6] 国家烟草专卖局. 烟草及烟草制品-钾的测定连续流动法, YC/T 217-2007. 北京: 中国标准出版社, 2007.
State Tobacco Monopoly Administration. Tobacco and Tobacco Products-Determination of Potassium-Continuous Flow method, YC/T 217-2007. Beijing: Standards Press of China, 2007 (in Chinese).

[7] 国家烟草专卖局. 烟草及烟草制品-氯的测定-连续流动法, YC/T 162-2011. 北京: 中国标准出版社, 2011.
State Tobacco Monopoly Administration. Tobacco and Tobacco Products-Determination of Chloride-Continuous Flow Method, YC/T 162-2011. Beijing: Standards Press of China, 2011 (in Chinese).

[8] 熊珍, 关体青, 周桂园, 等. 黑龙江烟叶化学成分及挥发性成分分析. 现代农业科技, 2023(8): 184–187.
Xiong Z, Guan T Q, Zhou G Y, et al. Analysis of chemical and volatile components of tobacco leaves from Heilongjiang. Mod Agric Sci Technol, 2023(8): 184–187 (in Chinese with English abstract).

[9] 吕纯纯, 姜余婷, 胡永华, 等. 印迹-解析电喷雾/光电离质谱成像技术研究烤烟叶中化学成分的空间分布. 分析化学, 2024, 52: 876–884.
Lyu C C, Jiang Y T, Hu Y H, et al. Study on spatial distribution of chemical components in flue cured tobacco leaves by imprinting analytical electrospray photoionization mass spectrometry. Chin J Anal Chem, 2024, 52: 876–884 (in Chinese with English abstract).

[10] 刘洪剑, 金红岗, 黄晓明, 等. 茄衣烟叶 7 种化学成分近红外预测模型的建立. 广东农业科学, 2023, 50(7): 64–73.
Liu H J, Jin H G, Huang X M, et al. Establishment of near-infrared prediction model for seven chemical components of wrapper tobacco. Guangdong Agric Sci, 2023, 50(7): 64–73 (in Chinese with English abstract).

[11] 廖宇. 基于高光谱图像的茶叶异物在线检测研究. 华东交通大学硕士学位论文, 江西南昌, 2023.
Liao Y. Online Detection of Foreign Matter in Tea Based on Hyperspectral Image. MS Thesis of East China Jiaotong University, Nanchang, Jiangxi, China, 2023 (in Chinese with English abstract).

[12] 毛立宇, 宾斌, 张洪明, 等. 基于近红外光谱的小麦成分检测仪. 光谱学与光谱分析, 2024, 44: 2768–2777.
Mao L Y, Bin B, Zhang H M, et al. Development of wheat component detector based on near infrared spectrum. Spectrosc Spectr Anal, 2024, 44: 2768–2777 (in Chinese with English abstract).

[13] 胡美玲, 郅晨阳, 薛晓梦, 等. 单粒花生蔗糖含量近红外预测模型的建立. 作物学报, 2023, 49: 2498–2504.
Hu M L, Zhi C Y, Xue X M, et al. Establishment of near-infrared reflectance spectroscopy model for predicting sucrose content of single seed in peanut. Acta Agron Sin, 2023, 49: 2498–2504 (in Chinese with English abstract)

[14] 王若楠, 张颖星, 于筱菡, 等. 基于近红外快速检测技术的谷

子淀粉多样性分析及模型构建. 作物学报, 2025, 51: 1757–1768.
Wang R N, Zhang Y X, Yu X H, et al. Near-infrared spectroscopic evaluation of starch diversity and model construction in foxtail millet. Acta Agron Sin, 2025, 51: 1757–1768 (in Chinese with English abstract).

[15] 王发勇, 李一辉, 杨洪明, 等. 基于手持式近红外光谱仪的化学成分预测模型构建及光谱样本采集方法研究. 烟草科技, 2021, 54(增刊 1): 63–68.
Wang F Y, Li Y H, Yang H M, et al. Establishment of prediction model for chemical components based on handheld near-infrared spectrometer and collection method of sample spectra. Tob Sci Technol, 2021, 54(S1): 63–68 (in Chinese with English abstract).

[16] 罗智勇. 基于近红外光谱的卷烟原料数字化评价方法研究. 青岛科技大学硕士学位论文, 山东青岛, 2022.
Luo Z Y. Study on Digital Evaluation Method of Cigarette Raw Materials Based on Near Infrared Spectroscopy. MS Thesis of Qingdao University of Science & Technology, Qingdao, Shandong, China, 2022 (in Chinese with English abstract).

[17] Ram B G, Oduor P, Igathinathane C, et al. A systematic review of hyperspectral imaging in precision agriculture: analysis of its current state and future prospects. Comput Electron Agric, 2024, 222: 109037.

[18] Das M, Yeo W S, Saptoro A. A review of machine learning in hyperspectral imaging for food safety. Vib Spectrosc, 2025, 139: 103828.

[19] Wang Y J, Liu Y, Chen Y Y, et al. Spatial distribution of total polyphenols in multi-type of tea using near-infrared hyperspectral imaging. LWT, 2021, 148: 111737.

[20] Lu X C, Zhao C, Qin Y Q, et al. The application of hyperspectral images in the classification of fresh leaves' maturity for flue-curing tobacco. Processes, 2023, 11: 1249.

[21] 李士静, 潘羲, 陈熙卓, 等. 基于高光谱信息的烟叶分级方法比较. 烟草科技, 2021, 54(10): 82–91.
Li S J, Pan X, Chen X Z, et al. Comparison of tobacco grading methods based on hyperspectral information. Tob Sci Technol, 2021, 54(10): 82–91 (in Chinese with English abstract).

[22] 郭文孟, 薛宇毅, 罗靖, 等. 基于高光谱成像的烟叶泛青特征分析与表征. 烟草科技, 2023, 56(7): 84–91.
Guo W M, Xue Y Y, Luo J, et al. Analysis and characterization of tobacco leaf greenish based on hyperspectral imaging. Tob Sci Technol, 2023, 56(7): 84–91 (in Chinese with English abstract).

[23] 范鹏飞, 马建伟, 姚思愚, 等. 基于高光谱成像和机器学习的烟叶霉变检测方法. 烟草科技, 2024, 57(12): 96–105.
Fan P F, Ma J W, Yao S Y, et al. Method for detecting mildew on tobacco leaves based on hyperspectral imaging and machine learning. Tob Sci Technol, 2024, 57(12): 96–105 (in Chinese with English abstract).

[24] 李智慧, 梅吉帆, 李辉, 等. 高光谱成像的非烟物质分类识别研究. 中国烟草学报, 2022, 28(3): 81–88.
Li Z H, Mei J F, Li H, et al. Research on classification and recognition of non-tobacco related material (NTRM) based on hyperspectral imaging technology. Acta Tab Sin, 2022, 28(3): 81–88 (in Chinese with English abstract).

[25] 梅吉帆. 烟草物料的高光谱检测方法与特征分析. 郑州烟草研究院硕士学位论文, 河南郑州, 2021.
Mei J F. Hyperspectral Detection Method and Characteristic Analysis of Tobacco Materials. MS Thesis of Zhengzhou Tobacco Research, Zhengzhou, Henan, China, 2021 (in Chinese with English abstract).

[26] Delwiche S R, Baek I, Kim M S. Does spatial region of interest (ROI) matter in multispectral and hyperspectral imaging of segmented wheat kernels? Biosyst Eng, 2021, 212: 106–114.

[27] Kartakoullis A, Caporaso N, Whitworth M B, et al. Gaussian mixture model clustering allows accurate semantic image segmentation of wheat kernels from near-infrared hyperspectral images. Chemom Intell Lab Syst, 2025, 259: 105341.

[28] Li L Q, Wang Y J, Cui Q Q, et al. Qualitative and quantitative quality evaluation of black tea fermentation through noncontact chemical imaging. J Food Compos Anal, 2022, 106: 104300.

[29] Im S H, Faqeerzada M A, Cho B K, et al. Optimized feature selection and machine learning for non-destructive estimation of soil volumetric water content in Chinese cabbage using hyperspectral imaging. Vib Spectrosc, 2025, 139: 103816.

[30] Ouyang Q, Wang L, Park B, et al. Simultaneous quantification of chemical constituents in matcha with visible-near infrared hyperspectral imaging technology. Food Chem, 2021, 350: 129141.

[31] 付博, 杨永锋, 刘向真, 等. 数据集划分及预处理方法对烟叶化学成分近红外定量模型的影响. 河南农业大学学报, 2025, 59: 516–527.
Fu B, Yang Y F, Liu X Z, et al. Influence of dataset partitioning and spectral pre-processing methods on the near infrared quantitative model of chemical ingredients in tobacco leaves. J Henan Agric Univ, 2025, 59: 516–527 (in Chinese with English abstract).

[32] Chen Q S, Chen M, Liu Y, et al. Application of FT-NIR spectroscopy for simultaneous estimation of taste quality and taste-related compounds content of black tea. J Food Sci Technol, 2018, 55: 4363–4368.

[33] 贺帆, 王涛, 孙建锋, 等. 烟叶不同区位主要化学成分差异分析. 江西农业学报, 2013, 25(12): 49–52.
He F, Wang T, Sun J F, et al. Analysis of difference in main chemical compositions of tobacco leaves among different positions. Acta Agric Jiangxi, 2013, 25(12): 49–52 (in Chinese with English abstract).

[34] 王小东, 顾会战, 郭东锋, 等. 烤烟叶片不同区位化学成分含量及协调性分析. 烟草科技, 2021, 54(6): 22–29.
Wang X D, Gu H Z, Guo D F, et al. Contents and harmony of chemical components in different parts of flue-cured tobacco leaves. Tob Sci Technol, 2021, 54(6): 22–29 (in Chinese with English abstract).

[35] 殷全玉, 许希希, 张玉兰, 等. 烟叶不同区位常规化学成分差异分析. 湖南文理学院学报(自然科学版), 2018, 30(1): 21–29.
Yin Q Y, Xu X X, Zhang Y L, et al. Analysis in main chemical composition of tobacco leaves among different positions. J Hunan Univ Arts Sci (Sci Technol), 2018, 30(1): 21–29 (in Chinese with English abstract).