

基于近红外光谱的武夷水仙焙火色泽品质评价模型构建

王宇翔^{1,2}, 黄艳², 孙辉^{2*}, 庞杰^{1*}

(1. 福建农林大学食品科学学院, 福建福州 350002) (2. 武夷学院茶与食品学院, 福建武夷山 354300)

摘要: 为了实现武夷水仙焙火品质的高效、无损检测, 优化焙火工艺参数, 提升岩茶标准化生产水平, 本研究构建了基于近红外光谱的焙火色泽品质分析模型。在武夷岩茶产区采集 90 份不同焙火强度(轻火、中火、重火)和时间序列(30~300 min, 间隔 30 min)的武夷水仙毛茶样品, 采用赛默飞 Nicolet IS 5 光谱仪(400~4 000 cm⁻¹)采集全波段近红外光谱, 并结合 YS6010 色差仪测定茶汤 *L*、*a*、*b* 色度参数。通过 SG 卷积平滑(Savitzky-Golay, SG)等方法进行光谱预处理, 利用遗传算法(Genetic Algorithm, GA)结合偏最小二乘法筛选关键特征波长, 并耦合支持向量机(Support Vector Machines, SVM)、BP 神经网络(Back Propagation Neural Network, BP)和随机森林(Random Forest, RF)建立预测模型。结果表明, 焙火过程中 *L* 值随焙火增强呈指数下降, *a* 值与焙火时间正相关, *b* 值持续降低; GA-PLS-SVM 模型表现最优, 定标集 *L*、*a*、*b* 的相关系数分别为 0.963、0.958、0.902, 均方根误差均小于 1.25, 验证集相关系数均高于 0.85, 具备良好泛化能力, 可有效反演焙火参数并识别工艺偏差。该模型能精准预测武夷水仙焙火过程中的色泽变化, 为岩茶焙火工艺的智能化调控与品质标准化提供理论依据和技术支持。

关键词: 色差系统; 近红外光谱; 武夷水仙茶; 品质评价模型

DOI: 10.13982/j.mfst.1673-9078.2027.4.0185

Construction of a Prediction Model for Roasting Color Quality of Wuyi Shuixian Based on Near-Infrared Spectroscopy

WANG Yuxiang^{1,2}, HUANG Yan², SUN Hui^{2*}, PANG Jie^{1*}

(1.Fujian Agriculture and Forestry University, College of Food Science, Fuzhou 350002, China)

(2.Wuyi University, College of Tea and Food, Wuyishan 354300, China)

Abstract: To achieve efficient and non-destructive detection of baking quality in Wuyi Shuixian tea, optimization of baking process parameters, and improvement of standardization in rock tea production, a near-infrared spectroscopy-based model for predicting baking-related color quality was developed. A total of 90 Wuyi Shuixian raw tea samples were collected from the Wuyi tea-producing region, covering different baking intensities (light, medium, and heavy fire) and time intervals (30~300 min, at 30 min intervals). Full-range near-infrared spectra (400~4 000 cm⁻¹) were acquired using a Thermo Scientific Nicolet IS 5 spectrometer, while the *L*, *a*, and *b* color parameters of tea infusions were measured by a YS6010 colorimeter. The spectral data were preprocessed using methods such as Savitzky-Golay (SG) convolution smoothing. Key characteristic wavelengths were selected using genetic algorithm combined with partial least squares (GA-PLS), and predictive models were established by coupling support vector machine (SVM), back propagation neural network (BP), and random forest (RF). It was found that the *L* value decreased exponentially with increasing baking intensity, the *a* value showed a positive correlation with baking time, and the *b* value continuously decreased. The GA-PLS-SVM model was determined to be optimal, with calibration set correlation coefficients of 0.963, 0.958, and 0.902 for *L*, *a*, and *b*, respectively, and Root Mean Square Error below 1.25. Validation set correlation coefficients were all above 0.85, indicating good generalization ability. The model is capable of accurately inverting baking parameters and identifying process deviations. It is concluded that the developed model can precisely predict color changes during the baking process of Wuyi Shuixian tea,

收稿日期: 2026-02-02; 修回日期: 2026-04-05; 接受日期: 2026-04-10

基金项目: 国家重点研发计划项目(2022YFD2101102); 福建省自然科学基金项目(025J011068); 南平市科技局资源产业化学联合创新项目(N2021Z005); 国家级大学生创新创业训练计划项目(202510397015); 武夷学院引进人才科研启动经费项目(YJ202508)

作者简介: 王宇翔(2000-), 男, 硕士研究生, 研究方向: 食品加工与安全, E-mail: m19157718298@163.com

通讯作者: 孙辉(1986-), 博士, 副教授, 研究方向: 食品化学与营养学相关研究, E-mail: sun_hui_edu@163.com; 共同通讯作者: 庞杰(1965-), 博士, 教授, 研究方向: 多糖结构化学相关研究, E-mail: pang3721941@163.com

providing theoretical and technical support for intelligent control and quality standardization in rock tea processing.

Key words: color difference system; near-infrared spectroscopy; wuyi shuixian tea; quality evaluation mode

武夷水仙 (*Camellia sinensis* cv *Shuixian*) 作为武夷岩茶的典型代表, 以其独特的“岩骨花香”品质特征享誉中外。福建省作为我国乌龙茶核心产区, 2025 年茶叶总产量达 57.5 万 t, 其中武夷岩茶年产值突破 170 亿元。在加工工艺中, 焙火作为形成武夷水仙茶“醇厚甘活”品质风格的核心工序, 直接影响茶汤色泽、香气组分及滋味协调性^[1]。传统焙火工艺依赖制茶师经验调控, 存在主观性强、稳定性不足、专家数量有限等问题, 导致品质批次差异显著^[2]。近年来, 随着消费者对茶叶品质标准化需求的提升, 构建基于数字化技术的焙火品质监控体系已成为茶产业升级的迫切需求。

茶叶色泽作为感官审评的核心指标, 其量化评价长期受限于人工判读的主观性。然而, 传统色差检测需破坏性取样, 难以满足在线工艺调控需求^[3,4]。近红外光谱 (Near-Infrared Spectroscopy, NIRS) 技术凭借其快速、无损、高通量分析优势, 已在茶叶水分^[5]、儿茶素^[6]及化学成分^[7]检测中取得突破性进展, 但针对焙火品质动态监测的研究尚未见系统报道。

当前研究多聚焦单一品质指标的定量分析。例如, 杜绍龙等^[8]基于 GA-BP 构建红茶色差预测模型, 建模集相关系数达 0.91; 林怿箐等^[9]利用 PLS (Partial Least Squares, PLS) 算法实现红茶拼配参数数字化, 预测误差降低 23.6%。然而, 乌龙茶焙火过程中的美拉德反应与焦糖化作用导致茶汤色泽呈现非线性变化特征, 传统线性模型难以准确解析其复杂动力学过程^[10]。Wu 等^[11]采用可见近红外高光谱技术耦合区间组合总体分析 (Interval Combination Population Analysis, ICPA), 竞争自适应重加权抽样 (Competitive Adaptive Reweighted Sampling, CARS), PLS 算法, 成功预测抹茶粒径与酚氨比 (验证集相关系数大于 0.93), 为复杂基质分析提供了新思路。

本研究以武夷水仙茶为对象, 创新性集成多光谱预处理与机器学习算法, 突破传统焙火工艺经验依赖瓶颈。通过采集不同焙火梯度 (轻火/中火/重火) 与时间序列 (30~300 min) 样本光谱数据, 结合色差参数动态响应规律, 构建 GA+SVM 智能预测模型, 旨在实现茶汤色泽的快速检测, 为武夷岩茶数字化加工提供技术支撑。

1 材料与方法

1.1 材料与仪器

本研究的武夷水仙样品由 2 部分构成。第 1 部分为近红外光谱模型建模材料, 为采自武夷学院茶山同一批次的“水仙”茶树鲜叶, 按照 GB/T 18745—2006《地理标志产品 武夷岩茶》标准加工制备的 90 份毛茶样品, 涵盖不同焙火强度 (轻火/中火/重火) 和时间序列 (30~300 min, 间隔 30 min)。从中随机挑选 60 份作为定标集样本, 剩余的 30 份作为验证集样本。同时挑选 10 份不同时间采集的样本作为外部验证集^[12]。第 2 部分为焙火工艺模拟材料, 以干燥拣剔后的水仙毛茶为原料, 按每笼摊叶 1.5 kg、厚度 5 cm 规格进行焙火处理, 设置轻火 70 °C、中火 90 °C、重火 110 °C 温度组, 烘焙 5 h, 每 30 min 上下翻动取样, 重复 3 次。

60 型茶叶温控电焙篓, 厦门城众帮机械设备有限公司; NICOLET IS 5 傅里叶近红外光谱分析仪, Thermo Fisher Scientific 公司; YS6010 台式分光测色仪, 深圳市三恩时科技有限公司。

1.2 实验方法

1.2.1 色泽指标测定

参照 GB/T 23776-2018《茶叶感官审评方法》。取 1.1 项下焙火茶样, 按标准冲泡后, 通过滤纸过滤冷却至 40 °C, 倒入 YS6010 色差仪配备的石英比色皿, 以蒸馏水为对照, 用色差仪测定色度, 重复 3 次, 取平均值。色度值采用国际照明委员会 1976 年推荐的均匀色彩空间, 测定三个分量 L (明度, 0~100, 0 为黑、100 为白)、 a (红绿轴, 正值偏红、负值偏绿) 和 b (黄蓝轴, 正值偏黄、负值偏蓝)^[13]。

1.2.2 近红外光谱的采集

采用 Thermo Fisher NICOLET IS 5 FTIR 光谱仪, 将焙火茶样经干燥处理后与溴化钾按 1:75 质量比研磨压片。设置扫描波数 4 000~400 cm^{-1} , 分辨率 4 cm^{-1} , 扫描 32 次累加信号, 实验室环境恒温 (25±1) °C。每组样品独立测定 3 次, 取特征峰均值分析。

1.2.3 数据预处理和建立模型方法



图1 实验流程示意图

Fig.1 Experimental Flowchart

茶汤色泽的高光谱智能检测基于近红外光谱快速分析技术,涵盖光谱预处理、特征筛选与模型构建三大步骤(图1)。首先,对原始光谱进行多维预处理以强化色泽关联信号:采用SG联合基线校正(Baseline Correction, BC)消除基线漂移与随机噪声,结合标准正态变换(Standard Normal Variate Transformation, SNV)、多元散射校正(Multiplicative Scatter Correction, MSC)、一阶导数(First Derivative, FD)、二阶导数(Second derivative, SD)抑制光散射效应;随后,引入GA对全波段进行全局寻优,以PLS模型的交叉验证均方根误差(Root Mean Square Error, RMSE)为适应度函数,筛选与 L 、 a 、 b 值显著相关的特征波段;最后,基于优选波段构建多元预测模型体系,系统比较BP神经网络、RF与SVM三种模型的性能,通过决定系数(Correlation Coefficient, R^2)与RMSE双指标量化精度,并验证模型泛化能力。

1.2.3.1 光谱预处理

由于样品颗粒大小和形态等物理特性的差异,光散射效应对光谱信号影响显著^[14,15]。为有效去除干扰信号、提高模型的稳健性,本文采用多种光谱预处理方法的组合,包括SG与FD、SD、BC、MSC及SNV。通过试验确定最优预处理组合,以最大化目标变量与光谱特征的相关性。

1.2.3.2 特征波段筛选

为有效降低光谱维度、剔除冗余信息并提升模型运算效率与可解释性,本研究采用GA对全波段($400\sim4000\text{ cm}^{-1}$)光谱进行特征筛选。GA是一种模拟自然选择与遗传机制的启发式优化算法,通过选择、交叉和变异等操作迭代进化,能够从高维光谱数据中高效搜索出与目标变量(L 、 a 、 b)关联性最强的特征波数子集^[16]。本研究设置初始种群规模为50,最大进化代数为100,交叉概率为0.8,变异概率为0.01。以验证PLS模型的RMSE作为适应度函数,通过迭代优化,最终筛选出用于各色泽参数预测的最优特征波段组合。

1.2.3.3 预测模型构建与验证

为系统评估基于GA筛选特征波段所建模型的预测性能与普适性,本研究采用PLS评估预测性能并使用BP、RF、SVM三种机器学习算法构建茶汤色泽(L 、 a 、 b)预测模型。

PLS是一种经典的线性建模方法,能有效处理光谱变量间存在的多重共线性问题,在高维小样本数据建模中应用广泛^[17]。RF是一种基于Bagging集成学习的非线性算法,通过构建大量决策树并集成其输出,对异常值和过拟合具有较好的稳健性,且能评估变量的重要性^[18]。SVM基于统计学习理论的结构风险最小化原则,通过核函数将数据映射到高维特征空间以解决非线性回归问题,在小样本情况下表现出良好的泛化能力^[19]。

具体建模流程如下:首先,将全部样本按2:1的比例随机划分为校正集(用于模型训练与超参数调优)和独立验证集(用于最终模型性能的无偏评估)。以GA筛选出的特征光谱数据为自变量(X),色泽参数实测值为因变量(Y),分别构建上述四种预测模型。各模型关键超参数通过网格搜索结合五折交叉验证进行优化确定:PLS模型的主成分数通过最小化RMSE确定;RF模型中决策树数量($n_{\text{estimators}}$)设为100,最大深度(max_depth)为10;SVM采用径向基核函数,正则化参数 C 和核参数 γ 通过搜索优化;BP模型采用单隐层结构,隐层神经元数设为10,学习率为0.01,训练目标误差为 1×10^{-5} 。

采用 R^2 和 $RMSE$ 作为模型性能的核心评价指标。 R^2 越接近 1, 表明模型对数据变异的解释能力越强; $RMSE$ 越小, 代表模型的预测精度越高。计算公式如下:

$$R^2 = 1 - \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

式中:

y ——第 i 个样本的色泽参数实测值、模型预测值和所有样本实测值的平均值;

M ——样本数。

1.2.3.4 预测模型的外部验证

为验证模型的预测性能及泛化能力, 选取独立于建模数据集的武夷水仙样本 10 份, 采样间隔与原方法 1.1 一致, 覆盖不同焙火时长以充分反映武夷水仙色泽在焙火过程中的动态演变特征。利用已建立的模型对验证集样本的色泽指标进行预测, 并采用色差仪测定相应的色泽参数。通过对比模型预测值与测定值并计算 R^2 与 $RMSE$, 以评估模型的准确性。

1.3 数据分析

采用 ORIGIN 2024 和 EXCEL 2021 软件对试验数据进行分析处理。采用 MATLAB 2020b 软件进行特征波长筛选和模型构建, 随机取 90 份茶样中的 60 份测试结果用于评价模型的建立, 余下 30 份用于模型的验证。

2 结果与讨论

2.1 茶叶样品光谱特征

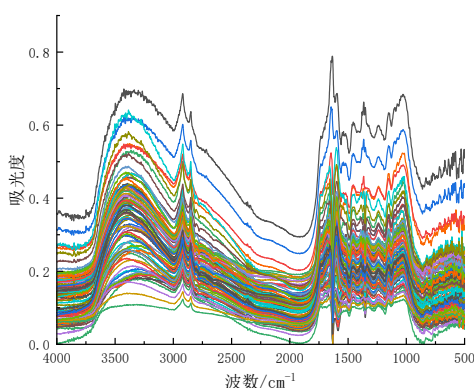


图 2 武夷水仙焙火样品近红外光谱图

Fig.2 Shuixian Tea Roasting Sample Near-Infrared Spectrum

如图 2 所示, 不同焙火程度水仙茶的红外光谱曲线整体高度相似, 表明焙火虽诱导热变性, 但茶叶中多糖、蛋白质和茶多酚等主要化学成分保持较高结构一致性^[20,21]。这些相似性源于基质多糖的 C-O 伸缩 ($1\ 000\sim1\ 200\text{ cm}^{-1}$)、蛋白质 amide I/II 带 ($1\ 500\sim1\ 700\text{ cm}^{-1}$) 以及茶多酚的芳香 C=C 振动 ($1\ 400\sim1\ 600\text{ cm}^{-1}$), 焙火差异主要表现为峰强度微变: 轻度焙火增强多酚氧化聚合, 重度焙火促进 Maillard 反应生成类黑精, 但茶叶样本的主要化学结构特征仍保持一致。在 $2\ 800\sim3\ 500\text{ cm}^{-1}$ 区, 所有样品显示宽阔 O-H/N-H 吸收峰 ($3\ 300\text{ cm}^{-1}$), 焙火增强 $1\ 450\text{ cm}^{-1}$ C-H 变形峰和 $1\ 500\sim1\ 600\text{ cm}^{-1}$ 双峰, 这些峰的位移与积分面积可量化焙火对色泽、香气和口

感的调控。上述光谱特征与焙火化学变化紧密相关：茶多酚降解聚合影响苦涩与持久香，蛋白/多糖变性调控黏度和色泽。类似于 Chen 等^[22]和 Cai 等^[23]在 NIRS 研究中的发现，IR 振动模式（C-H/O-H/C=O）对应 NIRS 倍频区，可作为品质指标。因此，红外光谱提供分子信息，支持水仙茶焙火色泽品质的非破坏性评估，并为 NIRS 模型构建和加工优化奠定数据库基础。

2.2 茶叶样品色泽指标测定值

焙火过程中色差指标变化如图 3 所示，随着焙火时间延长（0~300 min）和温度升高，茶叶叶绿素降解，茶多酚与氨基酸经 Maillard 和焦糖化反应生成茶褐素、茶红素等色素^[24]三种温度下，*L* 值呈线性下降：低焙火从 81.3 降至 71.71，中焙火 78.7 至 65.33，高焙火 76.57 至 53.11，茶汤由浅变深；*a* 值指数增长：低焙火从 4.32 增至 14.71，中焙火 5.83 至 18.8，高焙火 6.72 至 23.16，红色调显著增强；*b* 值指数下降：低焙火峰值 66.1 后降至 49.31，中焙火 60.57 降至 41.76，高焙火 58.66 降至 35.55，初始黄色调因氧化聚合减弱，红褐色主导。整体色泽从鲜亮黄橙演变为深沉红橙，体现焙火对明度降低和暖色调强化的视觉调控。这些规律显示，本研究定标集色泽指标分布广泛且差异显著，满足 NIRS 模型构建的谱质多样性要求^[25,26]。承接前述 IR/NIRS 振动特征（如 O-H/C=O 在 1 400~2 500 cm^{-1} 倍频区），焙火色泽动态可通过 PLS 回归量化，为武夷水仙品质预测模型优化提供关键指标，并桥接后续建模验证。

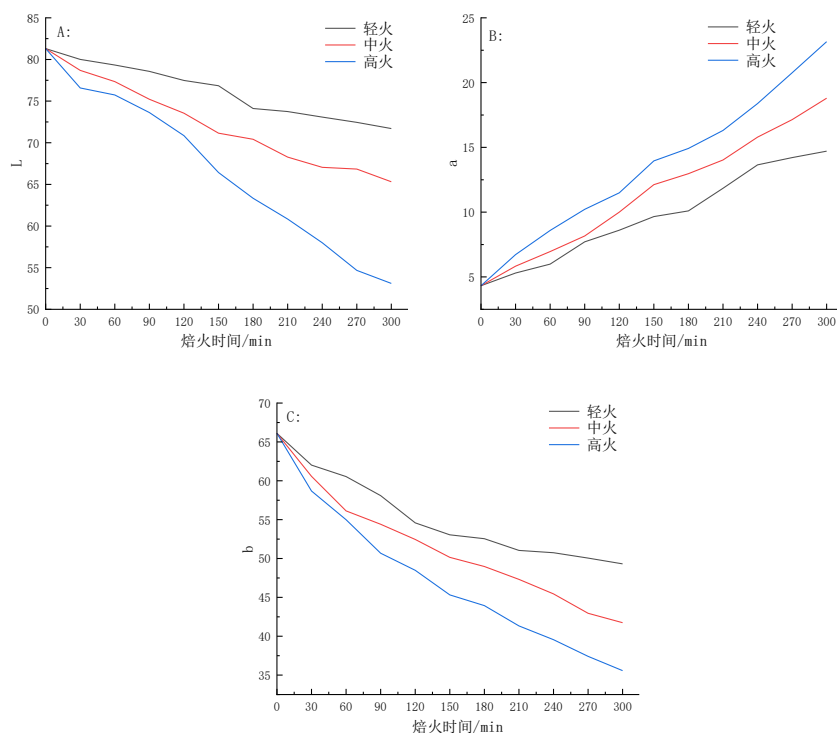


图3 武夷水仙焙火茶汤色泽变化图

Fig.3 Color Change Chart of Shuixian Tea Liquor During Roasting

2.3 茶叶样品色泽指标近红外光谱分析模型构建

2.3.1 光谱预处理方案选择

近红外光谱易受环境噪声影响，需预处理以提升信噪比和模型稳定性^[27]。表 1 列示最佳预处理下各色差指标定标集的 *RMSE* 和 R^2 值，本研究采用 SG+BC 为基础组合，辅以 SNV、MSC、FD 或 SD 优化建立 PLS 模型并采用交叉验证，选取最佳主成分因子数进行茶汤色泽指标预测。在 *L* 值模型中，三种焙火温度最佳预处理方法均为 SG+BC+SNV，*RMSE* 分别为 0.926（低焙火）、0.786（中焙火）和 2.882（高焙火）， $R^2>0.81$ ，表明该方案有效抑制散射干扰，捕捉明度衰减的 O-H 振动特征。高焙火 *RMSE* 较高，源于高温聚合诱导的非线性谱变^[28]。*a* 值模型中，高焙火优选 SG+BC+MSC（*RMSE*=1.296， $R^2=0.834$ ），增强红色调的 C-H/C=O 倍频信号；低/中焙火则为 SG+BC+SNV（*RMSE*=0.894/0.726， $R^2=0.934/0.975$ ），突出 Maillard 反应谱峰。*b* 值模型差异最大：低焙火 SG+BC+FD

($RMSE=1.621, R^2=0.907$)突出黄色调导数峰;中焙火 SG+BC+MSC($RMSE=1.144, R^2=0.960$);高焙火 SG+BC+SNV ($RMSE=1.247, R^2=0.871$), 适应氧化聚合的谱偏移。这些方案优化了焙火色泽的谱质关联 (整体 $R^2>0.83$, $RMSE<3.0$), 可为后续模型构建提供稳健输入, 支持武夷水仙茶汤色泽品质 NIRS 预测的温度特异性校准。

表 1 焙火水仙茶汤色泽指标最佳光谱预处理方案

Table 1 Optimal Spectral Pretreatment Scheme for the Color Indicators of Roasted Shuixian Tea Soup				
焙火程度	色差指标	预处理方法	均方根误差	决定系数
轻火	L	SG+BC+SNV	0.926	0.916
	a	SG+BC+SNV	0.894	0.934
	b	SG+BC+FD	1.621	0.907
中火	L	SG+BC++SNV	0.786	0.976
	a	SG+BC+SNV	0.726	0.975
	b	SG+BC+MSC	1.144	0.960
高火	L	SG+BC+SNV	2.882	0.817
	a	SG+BC+MSC	1.296	0.834
	b	SG+BC+SNV	1.247	0.871

2.3.2 GA 遗传算法优化茶汤色泽近红外光谱分析模型

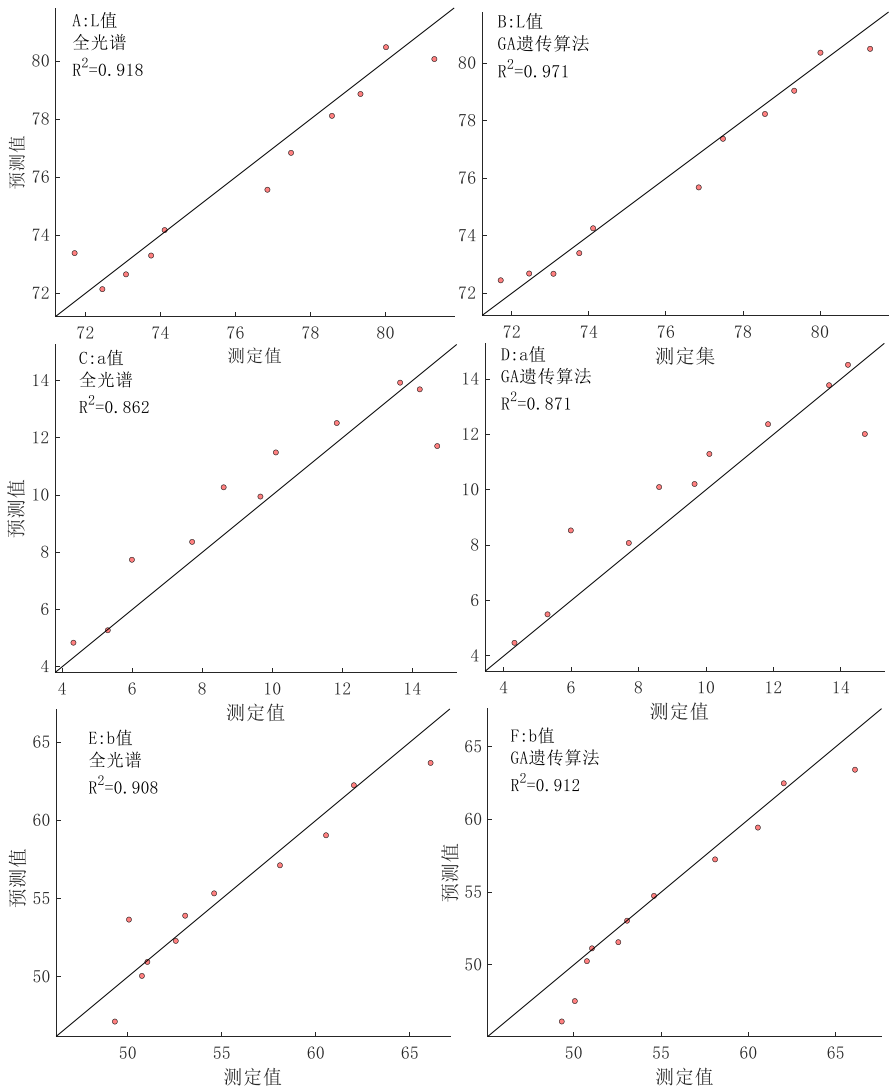


图 4 验证集近红外光谱 PLS 模型散点对比图

Fig.4 Scatter plot comparison of PLS model for near-infrared spectra of the validation set

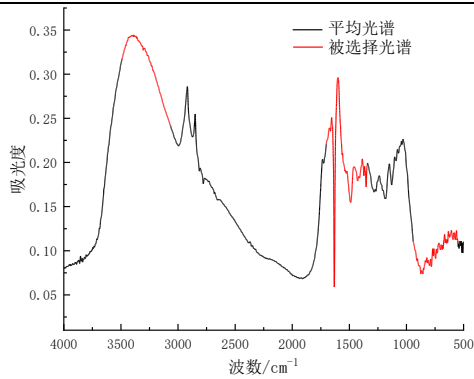


图 5 GA 遗传算法特征波段选取示意图

Fig.5 GA Genetic Algorithm Feature Band Selection

为提升武夷水仙焙火色泽近红外光谱预测模型的效率与稳健性，本研究采用 GA 对预处理后光谱进行特征波长筛选，构建 GA-PLS 模型。如图 4 所示，GA-PLS 模型在验证集上表现优异，尤其对 L 值的预测精度显著提升 ($R^2=0.971$)，优于全光谱 PLS 模型 ($R^2=0.918$)，表明 GA 有效提取了与明度相关的关键光谱信息。对 a 值与 b 值的预测中，GA-PLS 模型 (R^2 分别为 0.871 和 0.912) 亦较全光谱模型 (R^2 分别为 0.862 和 0.908) 有所改善或保持相当性能，说明特征筛选在降低数据冗余的同时，增强了模型对复杂色泽指标的解析能力^[29,30]。结果表明，基于 GA 的特征优选能够显著优化近红外光谱模型，为实现武夷水仙焙火色泽的快速、准确评价提供了可靠方法基础。

GA 筛选出的特征波长如图 5 所示，三种温度下共同筛选的波段大多集中在 550~900，1 342~1 701，3 066~3 490 cm^{-1} 左右。综合分析来看，GA 筛选的平均特征波段数为 362，占原波段数的 10.05%，GA-PLS 具有良好的全局搜索和信息提取能力^[31]，可以较大程度消除无用波段，但不能明显确定对茶汤色泽变化敏感性强的波段。

2.4 基于特征优选的武夷水仙焙火色泽近红外光谱预测模型构建与验证

焙火工艺是武夷水仙茶品质形成的关键环节，其中茶汤色泽 (L 、 a 、 b 值) 是消费者感知品质的重要指标。为实现焙火过程中色泽品质的快速、无损评价与精准调控，本研究系统构建了基于近红外 (NIR) 光谱技术的武夷水仙焙火茶汤色泽预测模型。

2.4.1 全光谱模型性能分析及其局限性

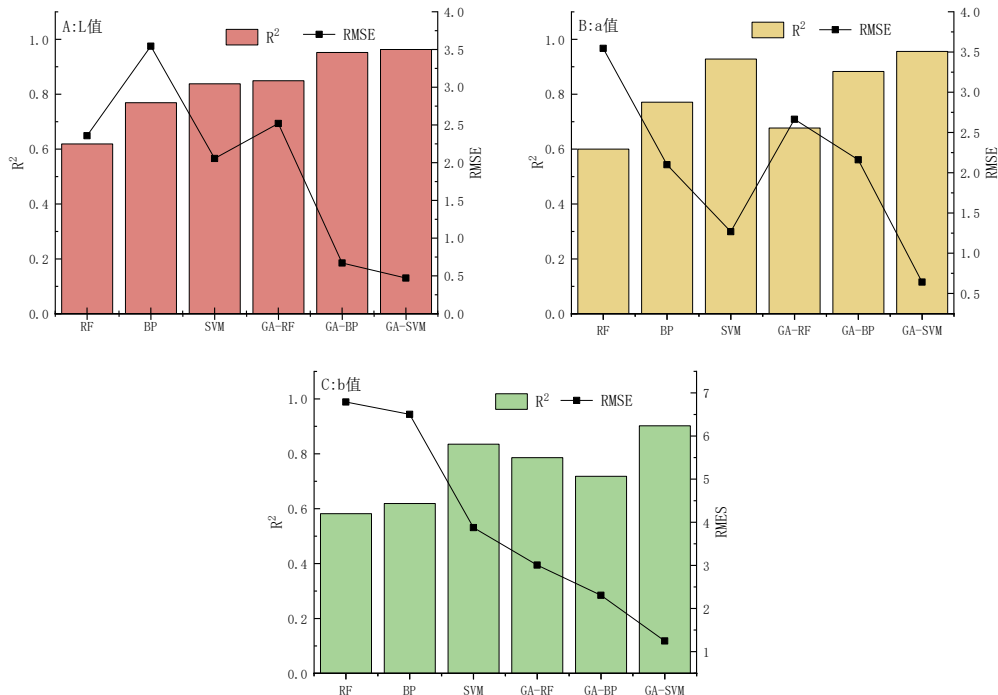


图 6 不同特征处理方法下的模型预测焙火茶汤色泽指标结果对比

Fig.6 Comparison of Model-Predicted Color Indicators of Roasted Tea Infusion under Different Feature Processing Methods

建模之前, 将经过 SG 等预处理的茶样全光谱数据和经 GA-PLS 筛选后的特征波段数据以及色泽指标光学测定值, 分别导入到 RF, BP 和 SVM 三种经典机器学习算法中进行建模。不同数据集建模预测性能的对比结果(图 6)可见, 不同模型对茶汤色泽 L 、 a 、 b 值的预测能力差异显著, 全光谱建模的结果揭示了模型的初步预测能力及潜在问题。在未经特征优选的情况下, SVM 模型表现相对最优, 对 L 、 a 、 b 值的验证集 R^2 分别达到了 0.838、0.928 和 0.835。然而, RF 与 BP 模型的预测精度则普遍较低 (R^2 范围在 0.582~0.771 之间), 且即使是 SVM 模型的 R^2 也未达到卓越水平。这强烈表明, 全光谱数据中可能存在大量的冗余信息和共线性波段, 这些非诊断性特征不仅增加了模型的计算负担, 更可能导致模型过拟合, 影响其泛化能力和预测稳定性^[32], 从而限制了其在实际应用中的效能。

2.4.2 GA-PLS 特征优选显著提升模型性能

GA-PLS 通过结合遗传算法的全局搜索能力和偏最小二乘法的降维去噪优势, 能够高效地从庞大的全光谱数据中筛选出与茶汤色泽 L 、 a 、 b 值高度相关的关键特征波段, 从而剔除冗余信息, 聚焦核心光谱变量。经 GA-PLS 特征优选后, 各模型预测精度均显著提升。优化后的 GA-SVM 模型表现最为突出, 该模型的验证集 R^2 则分别达到 L : 0.963、 a : 0.958、 b : 0.902, 同时验证集均方根误差 $RMSE$ 显著降低, 分别为 L : 0.471、 a : 0.642、 b : 1.246。与全光谱建模结果相比, GA-PLS-SVM 模型的 R^2 提升幅度最大, 尤其在 L 值和 b 值预测上展现出较好的拟合效果。GA-BP 与 GA-RF 模型性能亦明显改善, R^2 普遍升至 0.71 以上。结果表明, GA-PLS 特征优选法通过剔除冗余波段、聚焦关键光谱变量, 有效增强了模型解析能力与预测稳健性。

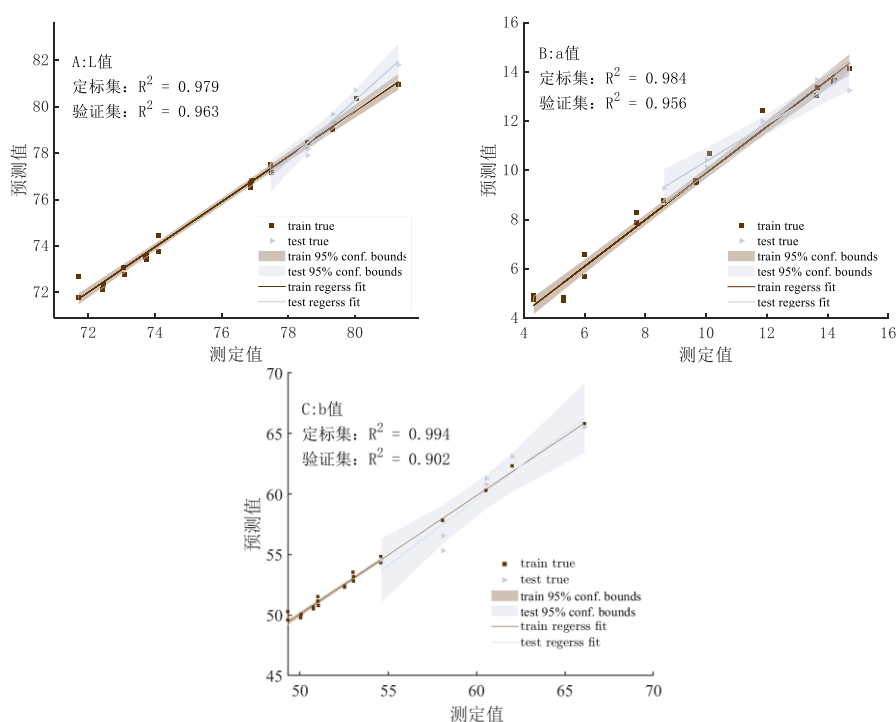


图 7 焙火水仙茶汤色泽指标预测最佳模型散点图

Fig.7 Scatter Plot of the Best Model for Predicting the Color Indicators of Roasted Narcissus Tea Infusion

2.4.3 外部验证模型泛化能力与稳定性

为了全面评估并证实 GA-PLS-SVM 模型的实际应用潜力和泛化能力, 我们采用 10 份独立的、在模型训练过程中从未出现过的外部验证集样本进行预测分析。图 8 直观展示了外部验证集样本的预测值与实测值之间的散点图。结果显示, L 、 a 、 b 值的外部验证集 R^2 分别达到了 0.957、0.954 和 0.914。这些外部验证指标与内部验证集(图 7)的结果相比, 并未出现显著的衰减。预测点密集分布在回归线附近, 且置信区间狭窄, 这有力地表明了所构建的 GA-PLS-SVM 模型不仅在训练集和内部验证集上表现优异, 更具备出色的跨样本泛化能力和预测稳定性。它能够准确捕捉并反映焙火武夷水仙茶汤色泽的真实变化趋势, 避免了过拟合现象, 确保了模型在未知样本上的可靠性和准确性。

综上, 将 GA-PLS 作为一种高效且稳定的特征波段筛选方法, 与 SVM 算法优异的非线性拟合与预测能力相

结合,最终构建的 GA-PLS-SVM 模型在所有茶汤色泽评价指标上均表现出最佳性能。其在 L 、 a 、 b 值预测验证上的高 R^2 ,充分证明该模型对武夷水仙焙火色泽参数的快速、准确、稳定预测能力。这一研究结果不仅为 NIR 光谱技术在茶叶品质检测中的应用提供了新方法,更为武夷水仙茶的智能化加工提供了坚实的技术支持。

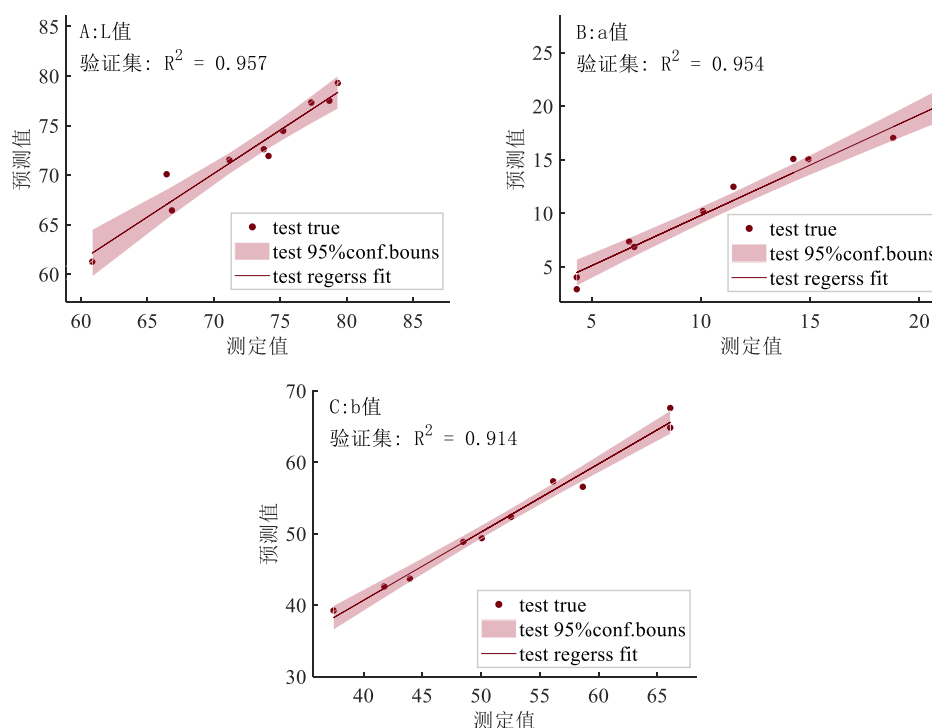


图 8 焙火水仙茶汤色泽指标最佳预测模型外部验证散点图

Fig.8 External Validation Scatter Plot of the Best Predictive Model for the Color Index of Roasted Narcissus Tea Soup

3 结论

茶汤色泽受焙火时间和温度等因素影响,呈现复杂的非线性关系,而传统研究多采用线性回归,难以捕捉其动态变化。本研究针对武夷山水仙茶焙火过程,开发基于 GA-PLS 优化的 SVM 茶汤色泽指标智能预测模型,采用 SG 等多种预处理方法处理光谱数据,有效减少原始光谱中的冗余变量以及无关信息。在特征波段选择上采用 GA-PLS 进行筛选,结果显示,与色泽指标相关的信息主要集中于 550~900、1 342~1 701 和 3 066~3 490 cm^{-1} 波段区间。该策略有效去除了无用信息,显著简化了模型结构,提高了模型的简洁性与解释能力。同时相比于 GA-PLS-RF 和 GA-PLS-BP 模型,GA-PLS-SVM 模型在茶汤焙火色泽的预测中具有一定优势:最佳模型训练集 $R^2=0.979$ 、 $RMSE=0.425$;验证集 $R^2=0.963$ 、 $RMSE=0.471$,显著优于 GA-PLS-RF ($R^2=0.619\sim0.849$) 和 GA-PLS-BP ($R^2=0.769\sim0.952$),后者易过拟合焙火噪声。GA-PLS-SVM 的非线性拟合优势桥接谱质映射,高效捕捉焙火色泽演变,证实其为武夷水仙焙火色泽品质估测的优选方法。

本研究构建的 GA-PLS-SVM 模型实现了焙火过程中武夷水仙茶汤色泽的快速准确预测,为茶叶加工过程中茶汤色泽品质的预测模型研究提供了理论依据和思路,具有推广至岩茶加工在线监测的潜力,未来的研究将致力于构建一个“多产地、多年度、多品种”的动态光谱数据库,进一步丰富训练数据的多样性,扩大样本覆盖范围以进一步提升模型跨场景适用性。

参考文献

- [1] TRUONG V L, JEONG W S. Antioxidant and anti-inflammatory roles of tea polyphenols in inflammatory bowel diseases [J]. Food Science and Human Wellness, 2022, 11(3): 502-511.
- [2] LANG Y, LU J L, ZHANG L Y, et al. Estimation of black tea quality by analysis of chemical composition and colour difference of tea infusions [J]. Food Chemistry, 2003, 80: 283-290.
- [3] 苏静,李长文,梁慧珍,等.色差技术在普洱茶品质控制中的应用研究[J].食品研究与开发,2017,389(20):148-151.

- [4] 王家勤,姚月凤,袁海波,等.基于色差系统的工夫红茶茶汤亮度的量化评价方法研究[J].茶叶科学,2020,40(2):259-268.
- [5] 陈琳,董春旺,高明珠,等.基于近红外光谱的红茶干燥中含水率无损检测方法[J].茶叶科学,2016,36(2):184-190.
- [6] 吴瑞梅,赵杰文,陈全胜,等.近红外光谱技术结合特征变量筛选快速检测绿茶滋味品质[J].光谱学与光谱分析,2011,32(7):1782-1785.
- [7] SONG H J, KIM Y D, JEONG M J, et al. Rapid selection of the anine-rich green tea trees and metabolites profiling by fourier transform near-infrared spectroscopy [J]. Plant biotechnology reports, 2015, 9: 55-65.
- [8] 杜邵龙,周志梅,雷亚兰,等.基于色差系统的宝庆桂丁红茶品质量化评价模型构建[J].食品工业科技,2022,43(14):329-335.
- [9] 林怵箐,周巧仪,李臻锋,等.基于近红外光谱的红茶滋味评价与拼配策略[J].食品与生物技术学报,2025,43(14):329-335.
- [10] 杨平,武夷岩茶焙火过程中关键 Maillard 反应香气成分及形成途径研究[D].北京:北京工商大学:2022.
- [11] WU J Z, MUHAMMAD Z, CHEN Q S, et al. Application of visible-near infrared spectroscopy in tandem with multivariate analysis for the rapid evaluation of match aphysicochemical indicators [J]. Food Chemistry, 2023, 421: 136185.
- [12] 冯丹,罗西,臧利艳,等.唐古特大黄多指标关键质量属性近红外光谱评价研究[J].分析测试学报,2024,43(11):1697-1708.
- [13] 陈丽霞,高枫,许玉金,等.基于色彩色差计和电子鼻的不同清炒程度白芍饮片的辨识[J].中国现代医药,2025,27(11):2135-2143.
- [14] 高瑞贞.基于现代质评技术对不同焙火程度肉桂的品质评价[D].福州:福建农林大学,2024.
- [15] 赵珠蒙.龙井茶外形品质特征的量化提取及数字化评价[D].杭州:浙江大学,2023.
- [16] ZHANG Y B, CHEN X W, CHEN D D. Machine learning-based classification and prediction of typical chinese green tea taste profiles [J]. Food Research International, 2025, 203: 115796.
- [17] ZENG X Q, CAO R, XI Y, et al. Food flavor analysis 4.0:a cross-domain application of machine learning [J]. Trends in Food Science & Technology, 2023, 138: 116-125.
- [18] CHEN Y, AN H M, HUANG Y W, et al. Analysis of non-volatile compounds in Jasmine tea and Jasmine based on metabolomics and sensory Evaluation [J]. Foods, 2023, 12: 3708.
- [19] BAI X, ZHANG L, KANG C Y, et al. Near-infrared spectroscopy and machine learning-based technique to predict quality-related parameters in instant tea [J]. Scientific Reports, 2022, 12(1): 3833.
- [20] GALVIN-KING P, HAUGHEY S A, ELLIOTT C T. Garlic adulteration detection using NIR and FTIR and chemometrics [J]. Journal of Food Composition and Analysis, 2021, 96: 103757.
- [21] SEHER G, METE S, ERIK G, et al. Differentiation of anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate analysis [J]. Food Chemistry, 2015, 170: 234-240.
- [22] CHEN Q C, 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]. Journal of Food Science & Technology, 2018, 55(10): 4363-4368.
- [23] CAI J X, WANG Y F, XI X G, et al. Using FTIR spectra and pattern recognition for discrimination of tea varieties [J]. International Journal of Biological Macromolecules, 2015, 78: 439-446.
- [24] CHEN Q C, FU Y, HENG W T. Re-rolling treatment in the fermentation process improves the taste and liquor color qualities of black tea [J]. Food Chemistry: X, 2024, 21: 101143.
- [25] YANG J R, WANG J, LU G D, et al. Teanet: deep learning on near-infrared spectroscopy data for the assurance of tea quality [J]. Computers and Electronics in Agriculture, 2021, 190: 106431.
- [26] 蔡海荣,宋文静,林建奔,等.花生品质性状近红外光谱分析模型构建及应用[J].华南农业大学学报,2025,46(4):450-458.
- [27] JIA J M, ZHOU X F, LI Y, et al. Establishment of a rapid detection mode for the sensory quality and components of yuezhou longjing tea using near-infrared spectroscopy [J]. LWT-Food Science and Technology, 2022, 164: 113625.
- [28] OUYANG Q, WANG L, BOSOOON P, et al. Simultaneous quantification of chemical constituents in matcha with visible-near infrared hyperspectral imaging technology [J]. Food Chemistry, 2021, 350: 129141.
- [29] YANG J, YANG S X, ZHANG Y Y, et al. Improving characteristic band selection in leaf biochemical property estimation considering interrelations among biochemical parameters based on the PROSPECT-D model [J]. Optics Express, 2021, 29(1): 400-414.
- [30] 李博,朱莉,姚庆宇,等.基于近红外光谱的草莓多品质参数通用预测模型研究[J].现代食品科技,2025,41(8):227-236.
- [31] 王冰玉,孙威江,黄艳,等.基于遗传算法的安溪铁观音品质快速评价研究[J].光谱学与光谱分析,2017,37(4):1100-1104.
- [32] LIU J B, DONG Z Y, XIA J S, et al. Estimation of soil organic matter content based on CARS algorithm coupled with random forest [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 258: 119823.