中国农业大学学报 2026,31(11):205-218
Journal of China Agricultural University http://zgnydxxb.ijournals.cn

叶春,赵爽,吴罗发,舒时富,谢枫. 基于光谱预处理与特征筛选的茶鲜叶SPAD值高光谱定量监测研究[J]. 中国农业大学学报,2026,31(11):205-218.
YE Chun,ZHAO Shuang,WU Luofa,SHU Shifu,XIE Feng. Hyperspectral monitoring of SPAD values in fresh tea leaves based on spectral preprocessing and feature selection[J]. *Journal of China Agricultural University*,2026,31(11):205-218.
DOI: 10.11841/j.issn.1007-4333.2026.11.15

# 基于光谱预处理与特征筛选的茶鲜叶SPAD值高光谱定量监测研究

叶 春[1] 赵 爽[1] 吴罗发[2*] 舒时富[1] 谢 枫[2]
(1.江西省农业科学院 农业工程研究所,南昌 330200;
2.江西省农业科学院 园艺研究所,南昌 330200)

**摘 要** 茶鲜叶SPAD值的快速无损监测是实现茶园精准氮肥管理的关键技术环节。针对茶鲜叶含水量高、品种间异质性大导致现有光谱监测模型泛化能力不足的问题,本研究旨在筛选适用于茶鲜叶特性的"光谱预处理—特征筛选—建模"最优技术路线。于2025—2026年春茶季采集2 844份茶鲜叶样本,其中有效配对数据2 842组(2份样本因SPAD读数变异系数超限予以剔除),利用ASD FieldSpec 4光谱仪采集光谱数据,同时使用SPAD-502 Plus叶绿素仪测定SPAD值。系统比较5种光谱预处理方法、3种特征筛选策略(竞争性自适应重加权采样、连续投影算法、无信息变量消除)及其组合对4种模型(偏最小二乘回归、随机森林、支持向量回归、反向传播神经网络)预测效能的影响,并通过网格搜索交叉验证优化模型超参数。一阶导数预处理结合无信息变量消除算法的特征选择方案表现较优,从全波段中筛选出225个关键光谱波段;基于该特征子集构建的随机森林模型训练集精度较高($R^2$=0.955 0,RMSE=2.366 5),测试集性能明显下降($R^2$=0.655 4,RMSE=6.414 7),表明模型在跨品种、跨采摘期条件下的泛化能力仍有局限;经超参数优化后,随机森林模型的独立验证性能得到提升($R^2$=0.703 2,RMSE=6.022 4),有效缓解了过拟合现象;跨年度验证进一步表明,模型年际泛化能力仍有改进空间。利用一阶导数结合无信息变量消除方法筛选出的225个特征波段建立的优化随机森林模型,可有效降低数据维度并保留叶绿素敏感信息,为不同品种、不同采摘期茶鲜叶SPAD值的快速无损预测及茶园精准氮肥管理提供了方法学支撑。

**关键词** 茶鲜叶;叶绿素;高光谱;SPAD;监测
**中图分类号** S571.1;S127 **文章编号** 1007-4333(2026)11-0205-14 **文献标志码** A

# Hyperspectral monitoring of SPAD values in fresh tea leaves based on spectral preprocessing and feature selection

YE Chun[1], ZHAO Shuang[1], WU Luofa[2*], SHU Shifu[1], XIE Feng[2]
(1. Institute of Agricultural Engineering, Jiangxi Academy of Agricultural Sciences, Nanchang 330200, China;
2. Institute of Horticulture, Jiangxi Academy of Agricultural Sciences, Nanchang 330200, China)

**Abstract** Rapid and non-destructive monitoring of SPAD values in fresh tea leaves is a key technical approach for achieving precision nitrogen management in tea plantations. Given that the high moisture content and substantial inter-varietal heterogeneity of fresh tea leaves lead to insufficient generalization capacity of existing spectral monitoring models, this study aimed to identify an optimal technical pipeline of "spectral preprocessing-

收稿日期:2026-06-08;接受日期:2026-08-12
基金项目:江西省农业科学院基础人才计划项目(JXSNKYJCRC202450、JXSNKYJCRC202525);江西省自然科学基金青年基金项目(20244BCE52264、20252BAC200085);江西现代农业科研协同创新项目(JXXTCX202603)
第一作者:叶春(ORCID:0000-0001-7407-4203),副研究员,主要从事农业工程研究,E-mail:lychunzi@126.com
通讯作者:吴罗发(ORCID:0000-0001-8464-9971),研究员,主要从事信息农业研究,E-mail:297449566@qq.com

feature selection-modeling" tailored to the characteristics of fresh tea leaves. A total of 2 844 fresh tea leaf samples were collected during the spring tea harvest seasons of 2025 and 2026. After excluding two samples whose SPAD readings exceeded the coefficient of variation threshold, 2 842 pairs of valid data were retained. Spectral data were acquired using an ASD FieldSpec 4 spectroradiometer, and SPAD values were simultaneously measured with a SPAD-502 Plus chlorophyll meter. The effects of the combinations of five spectral preprocessing methods with three feature selection strategies [competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA), and uninformative variable elimination (UVE)] on the predictive performance of four modeling algorithms-partial least squares regression (PLSR), random forest (RF), support vector regression (SVR), and back propagation neural network (BPNN)-were systematically compared. Model hyperparameters were optimized via grid search combined with cross-validation. The results showed that the combination of first-derivative preprocessing and the UVE algorithm yielded superior performance, selecting 225 key spectral bands from the full spectrum. The RF model built on this feature subset achieved high accuracy on the training set ($R^2$=0.955 0, RMSE=2.366 5) but exhibited a marked decline in performance on the test set ($R^2$=0.655 4, RMSE=6.414 7), indicating limited generalization capacity under cross-variety and cross-harvest-period conditions. After hyperparameter optimization, the independent validation performance of the RF model was improved ($R^2$=0.703 2, RMSE=6.022 4), effectively alleviating the overfitting problem. Cross-year validation further indicated that the inter-annual generalization capacity of the model still has room for improvement. In conclusion, the optimized RF model established using the 225 feature bands selected by first-derivative preprocessing combined with UVE effectively reduces data dimensionality while preserving chlorophyll-sensitive information, providing methodological support for the rapid and non-destructive prediction of SPAD values in fresh tea leaves across different varieties and harvest periods, as well as for precision nitrogen management in tea plantations.

**Keywords** fresh tea leaves; chlorophyll; hyperspectral; SPAD; monitoring

茶(*Camellia sinensis* (*L.*) *O. Kuntze*)是世界上最受欢迎的非酒精饮料之一,其产量与品质直接关系茶农收益和区域经济发展[1]。叶绿素测量仪(Soil plant analysis development,SPAD)获得的数值(SPAD值)是衡量茶树叶片光合生理活性与氮素盈亏状况的关键指示因子[2-3],研究表明,茶鲜叶SPAD值与鲜叶产量、持嫩性及氨基酸等主要品质成分的代谢积累存在显著关联[4],进而影响成品茶的经济价值。实现对茶鲜叶SPAD值的快速、无损监测,可为茶园精准追肥提供决策依据[5],同时可辅助名优茶适采期判定[6]。当前茶园氮肥管理存在过量施用的突出问题,导致氮肥利用率不足30%[7]。实现茶园精准施氮,前提是具备能够快速、无损获取茶树氮素状态的技术手段。

在茶树光谱监测领域,近红外光谱技术较早应用于成品茶中茶多酚、咖啡碱、游离氨基酸等生化组分的定量分析[8]。随后逐步从成品茶检测向鲜叶监测延伸,实现了茶树叶片叶绿素和氮素含量的估算[9]。原始光谱数据中包含大量与目标组分无关的散射及背景噪声,需进行预处理[10]。针对不同样本集优选适配的光谱预处理方法已被证实是提升模型预测精度的关键步骤[11],常用的方法包括多元散射校正(Multiplicative scatter correction,MSC)、标准正态变量变换(Standard normal variate,SNV)、平滑(Savitzky-golay,SG)及导数变换等。预处理后的全波段光谱仍包含数百个波长变量,高维共线性易导致模型过拟合。为降低光谱维度并增强有效信号,常用的策略有变量筛选,如竞争性自适应重加权采样(Competitive adaptive reweighted sampling,CARS)基于偏最小二乘回归(Partial least squares regression,PLSR)系数权重迭代筛选关键波长[12]。连续投影算法(Successive projections algorithm,SPA)通过最小化共线性迭代选取信息量最大的光谱波段[13]。无信息变量消除(Uninformative variable elimination,UVE)基于PLSR的稳定性系数剔除对模型贡献较小的变量。在建模算法方面,偏最小二乘回归、支持向量回归(Support vector regression,SVR)等回归模型被广泛应用于茶树光谱与生化组分间的定量映射[14-15],随机森林(Random Forest,RF)及反向传播神经网络(Back propagation neural network,BPNN)、卷积神经网络等深度学习方法也逐步引入茶鲜叶内含物的光谱快速检测,显示出较强的非线性拟合能力[16-17]。然而,茶鲜叶

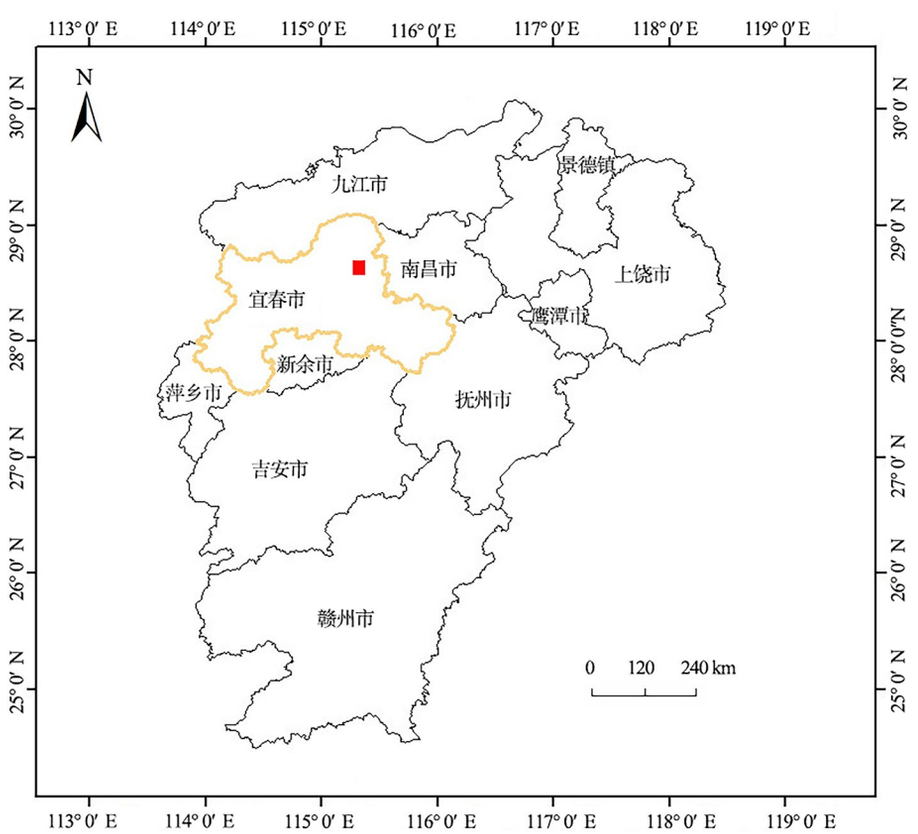
SPAD值的快速无损监测面临特殊挑战：其一，茶鲜叶含水量范围通常在75%～82%（因品种和采摘期而异）[18]，水分子在近红外区域存在强烈的特征吸收峰，导致光谱动态范围被压缩，降低了叶绿素特征波段的信噪比[19]；其二，不同品种茶树在叶片厚度、叶绿素基础含量及内部组织结构等方面存在显著差异[20]，使光谱-叶绿素响应关系呈现品种间异质性，增加了建模难度。因此，筛选相对适用于茶鲜叶特性的SPAD值监测方法成为当前亟需解决的核心技术问题。

目前鲜见针对茶鲜叶的跨年度、多品种、全流程（预处理-特征筛选-建模）高光谱监测方法系统性比较研究[21]。现有基于大田作物建立的光谱监测模型能否直接适用于茶鲜叶，尚缺乏系统性的验证与评估。鉴于此，有必要针对茶鲜叶的光谱特性，开展系统性的全流程方法比较研究。拟以南方茶区6个主栽品种为对象，建立跨品种的SPAD值统一估测模型，避免大多数同类研究仅以1～2个品种建模导致品种迁移性不足的局限。本研究旨在系统比较不同光谱预处理、特征选择和建模算法组合对6种主要茶树品种在2个生长季鲜叶中SPAD值估算精度的影响。确定的最优技术有望为茶园非破坏性叶绿素监测和精准氮素管理提供方法学基础。

## 1　材料与方法

### 1.1　试验地点

试验在江西省农业科学院高安试验基地茶园进行（图1），该试验站位于江西省宜春市高安市相城镇华阳村（28°25′27″ N，115°12′15″ E）。茶园面积500 hm²，海拔41 m，属山地丘陵地貌。土壤类型为红壤，pH约4.5，有机质含量约20 g/kg。茶园按常规管理，年施氮肥量30 kg/hm²，采用机采方式管理。




基于审图号GS（2024）0650，底图无修改。

Based on the drawing approval GS（2024）0650 and not modified.

**图1　试验点所在位置**

Fig. 1　Location of the test site

供试茶树品种有‘鸠坑早’‘白金芽’、‘中茶102’‘新安2号’‘鸠16’‘福鼎大白茶’，分别于2025年（3月28日、4月2日、4月10日）和2026年（4月2日、4月14日）采集了6个茶树品种的高光谱反射率数据，共获得2 844个叶片样本。2026年气温回升较快、茶芽萌发集中，春茶采摘窗口期缩短，故采样次数相应调整为2次。为确保样品一致性，所有叶片均采集自新梢自上而下第3片完全展开叶。

### 1.2 测定项目与方法

#### 1.2.1 光谱反射率测定

叶片的光谱反射率使用ASD Fieldspec 4野外便携式高光谱仪（Analytical Spectral Devices，简称ASD，美国），测定波长范围为350～2 500 nm，采样间隔1 nm。测定均在晴天10:00～14:00进行，以保证太阳辐射角度一致。测定前，使用标准白板进行校正，以此得出的光谱数据为无量纲的相对反射率（标准白板相对反射率为1）。测定时，将叶片置于野外便携式高光谱仪的叶片夹持器中，选取叶片中脉两侧平展区域进行光谱采集，避开叶脉及叶缘破损区域。每个品种取标准叶在相同样点采集光谱数据，取5个光谱反射率的平均值作为该点叶片的光谱反射率。

#### 1.2.2 叶绿素相对值测定

使用SPAD-502 Plus叶绿素计（Konica Minolta，日本）来估算茶鲜叶的叶绿素相对值，该仪器在叶绿素值10～50范围内测量误差为±1.0单位，50～80范围内误差为±2.0（Konica Minolta技术手册）。所有叶片均采集自当年生新梢自上而下第3片完全展开叶，每片叶片在叶中脉两侧等间距取3个测定点，以3点读数的算术平均值作为该叶片的SPAD值。最终获得2 842组有效配对数据（2份样本因连续3次SPAD读数变异系数＞5%而剔除）。所有茶鲜叶的SPAD读数均记录于Excel文件中，以便进一步处理。

#### 1.2.3 统计与分析

本研究采用单因素方差分析（One-way ANOVA）检验不同茶树品种及采样时期对SPAD值的影响，显著性水平设为$\alpha=0.05$，多重比较采用LSD法。建模过程中，偏最小二乘回归（PLSR）、随机森林（RF）、支持向量回归（SVR）及BP神经网络（BPNN）分别基于scikit-learn库的PLSRegression、RandomForestRegressor、SVR及MLPRegressor模块实现。所有模型均在Python 3.9环境中运行。

### 1.3 光谱预处理与特征选择

所采集的茶鲜叶光谱共2 151个波长。针对茶鲜叶高光谱数据中普遍存在的光程差异、散射效应及基线漂移等问题，引入多元散射校正（MSC）、标准正态变量变换（SNV）、一阶导数（First derivative，FD）、去趋势（Detrending，DT）及卷积平滑（SG）共5种预处理方法进行光谱预处理。采用竞争性自适应重加权采样（CARS）、无信息变量消除（UVE）与连续投影算法（SPA）及其组合进行特征波长优选。CARS算法设置迭代次数为50次，采用5折交叉验证，并利用指数衰减函数动态调整采样比例；UVE算法通过引入随机噪声矩阵，基于50次蒙特卡洛采样计算变量稳定性；SPA算法设定最大投影向量数为30个，并在投影前对光谱数据进行标准化处理。上述算法均基于Python 3.9（scikit-learn 1.2.1）实现，关键随机过程设置为固定随机种子（random_state=42），以保证实验结果的可复现性。

### 1.4 模型评价标准

经过特征选择后，利用筛选后的特征变量构建茶鲜叶SPAD值的估测模型。模型评价选择决定系数$R^2$和均方根误差（Root mean square error，RMSE）为评估指标。$R^2$取值范围为0～1，越接近1代表回归模型预测精度越高。RMSE与SPAD值单位相同，其值越小，表明回归模型预测误差越小。各指标表达式分别如下：

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

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

式中：$y_i$为数据集中的真实值，$\hat{y}_i$为预测值，$\bar{y}$为观测值的平均值，$n$为样本数，$m$为测试集样本数。

## 2 结果与分析

### 2.1 叶绿素相对值变化情况

茶鲜叶SPAD值的变化范围为2.1～66.8，平均值为22.6，标准差为13.8，变异系数达61.1%，表明供试样本叶绿素含量分布离散，品种间差异显著。对数据进行单因素方差分析，不同茶树品种间的

SPAD值存在极显著差异（$F=336.12$，$p<0.01$）。从品种特性上看（图2），‘福鼎大白茶’整体维持较高水平，均值最高（约29.5），表现出较强的叶绿素积累能力；‘中茶 102’和‘新安 2 号’SPAD数值相对较低，多在20～40。相比之下‘白金芽’的SPAD值在整个观测期内持续处于低位，大部分时间低于20，叶绿素合成能力显著受限。‘鸠坑早’和‘鸠 16’受极端值影响明显，均值虽居中游但离散程度较大。从时间维度看，所有品种的SPAD值在3月28日至4月2日期间波动较大，未呈现统一趋势。4月2日至4月14日期间，‘福鼎大白茶’、‘鸠坑早’和‘中茶 102’的SPAD值总体呈上升趋势，而‘白金芽’在整个观测期内始终维持于低位，未呈现明显的时间性上升规律。2026年4月2日和4月14日的采集数据与2025年对应时期的品种排序总体一致，表明品种间SPAD差异具有跨年度的稳定性。

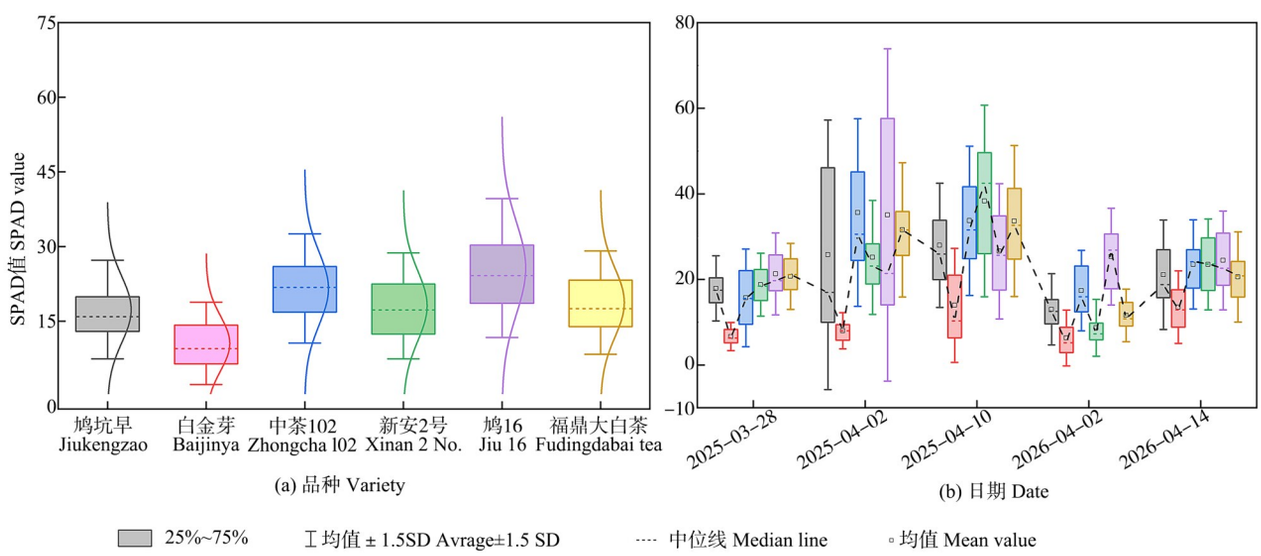



**图2　不同品种和时期茶鲜叶叶绿素变化情况**

Fig. 2　Chlorophyll variation in fresh tea leaves of different varieties and periods

## 2.2　光谱数据预处理结果

经多元散射校正（MSC）、标准正态变量变换（SNV）、一阶导数（FD）、去趋势（DT）及卷积平滑（SG）5种方法预处理后，茶鲜叶高光谱反射率曲线如图3所示。原始光谱曲线（图3（a））整体形态平滑，呈现典型绿色植被反射光谱特征：蓝光区（约450 nm）与红光区（约670 nm）各出现1个明显的叶绿素吸收谷，绿光区（约550 nm）形成反射峰，红边至近红外波段（700～750 nm）反射率急剧上升，并在近红外平台区保持较高反射水平。同时，原始光谱样本间基线差异显著，在400～500 nm与900～1 000 nm等波段存在明显的交叉重叠现象，提示光程差异与散射效应对原始光谱产生了较强干扰，不利于后续特征波段的稳定提取。

MSC（图3（b））与SNV（图3（c））处理后，样本光谱的基线漂移得到明显校正，曲线间平行度提高、离散程度减小。2种方法均通过消除叶片表面微观结构差异所导致的光程与散射效应，使光谱数据更趋一致，为后续建模提供了可比性较好的输入变量。FD处理（图3（d））显著增强了光谱的细微结构信息，绿光区与红光区的吸收峰位置更加突出，峰形锐度提高，同时原始光谱中弱信号区域的肩峰与吸收谷辨识度明显增强；此外，FD处理可在一定程度上消除加性基线偏移，有助于凸显与叶绿素含量相关的特征波段。DT处理（图3（e））有效压制了长波段区域的非线性基线漂移，使光谱曲线围绕零基线分布，提升了模型在长波区域（700～1 000 nm）的稳定性，并使光谱对叶绿素含量的微小变化更为敏感。SG平滑处理（图3（f））在保留原始光谱整体轮廓特征的基础上，显著抑制了高频噪声，曲线平滑度提高、信噪比改善，但过度平滑亦可能在一定程度上弱化光谱的细微结构信息。

## 2.3　特征变量选择结果

为降低高光谱数据的冗余度并有效筛选对茶鲜叶SPAD值敏感的特征波长，在光谱预处理的基础上，分别采用CARS、SPA、UVE及其组合方式对

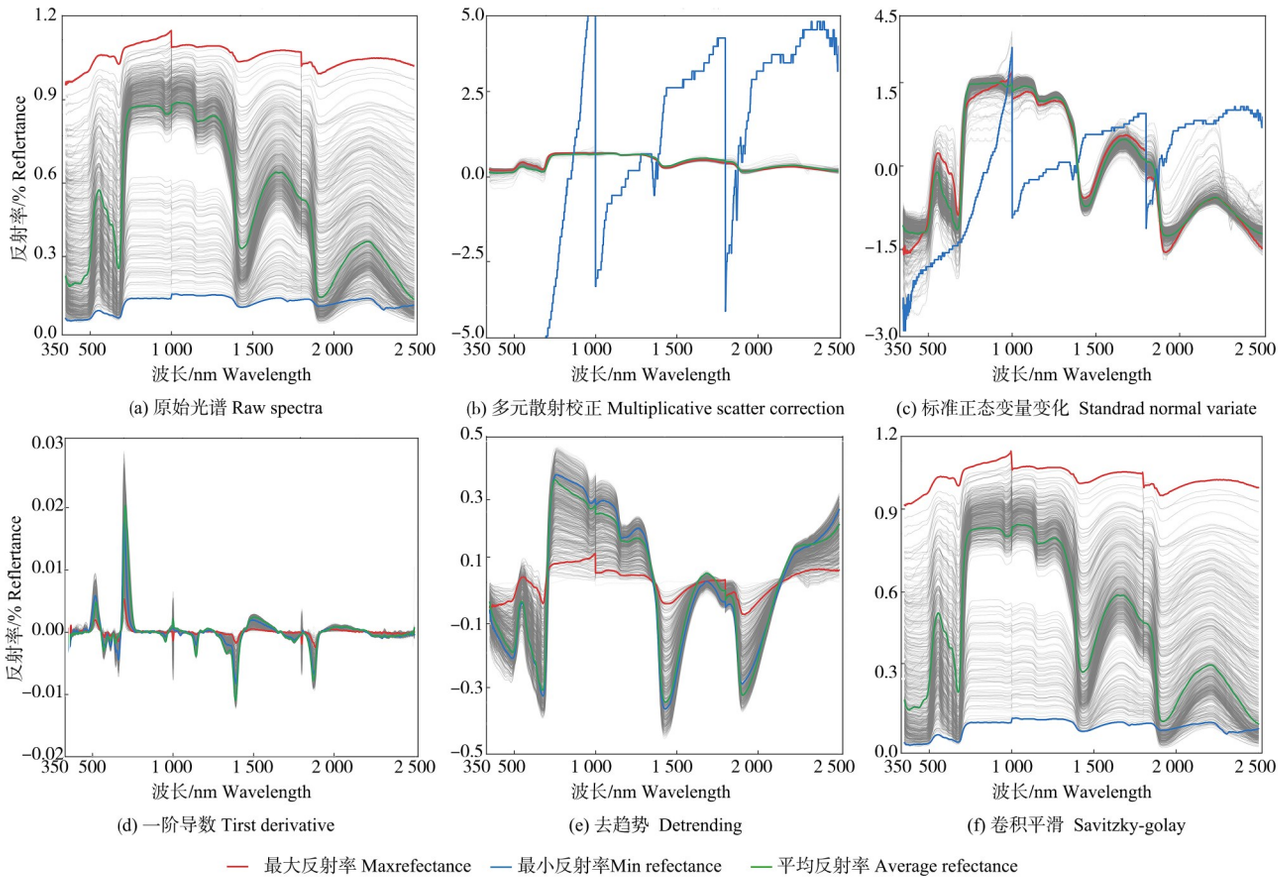



**图 3 不同光谱预处理方法对比**

Fig. 3 Comparison of different spectral preprocessing methods

全波段光谱数据进行特征筛选，各组合的特征波段数量、相关系数（$R^2$）及交叉验证均方根误差（RMSECV）结果如表 1 所示。在所有预处理方法下，CARS、SPA 和 UVE 3 种单一算法的表现存在明显差异。UVE 和 CARS 筛选出的波段数量较多（超过 200 个），保留了较为完整的光谱结构信息，但在多数预处理条件下 $R^2$ 仍处于 0.326 3～0.702 6，RMSECV 为 7.673 3～11.050 1；SPA 的筛选结果均稳定压缩至 30 个波段，体现了其投影寻优的波段简化能力，但部分预处理组合下 $R^2$ 有所下降（如 MSC＋SPA，$R^2$ ＝0.446 4）。值得注意的是，MSC 预处理下 3 种单一算法的 $R^2$ 均不超过 0.446 4，RMSECV 超过 9.081 0，表明 MSC 校正不适用于本数据集的预处理。

与单一算法相比，两两联合（CARS＋SPA、CARS＋UVE、UVE＋SPA）和三者联合（UVE＋CARS＋SPA）并未在所有预处理条件下带来一致的性能提升。在多数情形下，联合算法的最终波段数压缩至 30 个以下（UVE＋CARS＋SPA 最少仅 15～23 个），但 $R^2$ 反而低于对应预处理下的最优单一算法，如 SNV 处理下 UVE＋CARS＋SPA（$R^2$＝0.384 7）远低于 SNV＋CARS（$R^2$＝0.549 8）。这说明多步骤级联筛选存在信息过度剔除的风险，且不同算法对同一光谱数据的“有效波段”判定存在冲突。综合所有 42 组结果，一阶导数 FD＋UVE 组合取得最高 $R^2$（0.702 6）和最低 RMSECV（7.673 3），筛选后保留了 225 个特征波段（图 4）。其中，225 个特征波段在紫外-蓝紫边缘区（350～400 nm）分布 17 个（占 7.6%），在可见光区（400～700 nm）分布 32 个（占 14.2%），主要密集分布于叶绿素 a 的蓝紫光吸收峰（约 430 nm）和叶绿素 b 的红光吸收峰（约 660 nm）附近；在红边区（680～780 nm）分布 9 个（占 4.0%），在近红外区（780～1 300 nm）分布 102 个（占 45.3%），在短波红外区（1 300～2 500 nm）分布 65 个（占 28.9%），部分波段邻近水分强吸收带（1 450 nm、1 940 nm）及蛋白质/

**表 1　茶鲜叶SPAD值特征波长筛选结果**

Table 1　Selected feature wavelengths for fresh tea SPAD value

| 预处理方法 Preprocessing methods | 特征提取方法 Feature extraction method | 特征波段数量/个 Number of characteristic bands | 相关系数 Correlation coefficient $R^2$ | 交叉验证均方根误差 Cross-validation root mean square error RMSECV |
|---|---|---|---|---|
| 原始光谱 Raw spectrum | CARS | 720 | 0.583 4 | 8.078 2 |
| | SPA | 30 | 0.574 6 | 8.148 0 |
| | UVE | 753 | 0.581 0 | 8.096 8 |
| | CARS＋SPA | 30 | 0.583 9 | 8.074 2 |
| | CARS＋UVE | 429 | 0.588 7 | 8.115 5 |
| | UVE＋SPA | 30 | 0.574 6 | 8.148 1 |
| | UVE＋CARS＋SPA | 23 | 0.580 9 | 8.098 4 |
| 多元散射校正 Multiplicative scatter correction | CARS | 548 | 0.326 3 | 10.332 4 |
| | SPA | 30 | 0.446 4 | 9.081 0 |
| | UVE | 455 | 0.406 5 | 9.389 8 |
| | CARS＋SPA | 30 | 0.369 1 | 10.203 2 |
| | CARS＋UVE | 228 | 0.342 2 | 10.091 3 |
| | UVE＋SPA | 30 | 0.385 6 | 9.701 4 |
| | UVE＋CARS＋SPA | 17 | 0.288 7 | 11.050 1 |
| 标准正态变量变换 Standard normal variate | CARS | 663 | 0.549 8 | 8.240 7 |
| | SPA | 30 | 0.493 3 | 8.708 9 |
| | UVE | 231 | 0.538 4 | 8.328 3 |
| | CARS＋SPA | 30 | 0.492 1 | 8.712 4 |
| | CARS＋UVE | 430 | 0.537 8 | 8.332 9 |
| | UVE＋SPA | 30 | 0.508 5 | 8.571 8 |
| | UVE＋CARS＋SPA | 15 | 0.384 7 | 9.570 5 |
| 一阶导数变换 First derivative | CARS | 720 | 0.621 5 | 7.708 8 |
| | SPA | 30 | 0.353 5 | 9.586 1 |
| | UVE | 225 | 0.702 6 | 7.673 3 |
| | CARS＋SPA | 30 | 0.494 5 | 8.601 6 |
| | CARS＋UVE | 320 | 0.612 7 | 7.699 0 |
| | UVE＋SPA | 30 | 0.426 9 | 9.079 5 |
| | UVE＋CARS＋SPA | 19 | 0.393 5 | 9.302 6 |
| 去趋势 Detrending | CARS | 663 | 0.586 6 | 8.057 4 |
| | SPA | 30 | 0.589 9 | 8.030 8 |
| | UVE | 705 | 0.597 6 | 8.049 3 |
| | CARS＋SPA | 30 | 0.584 5 | 8.074 5 |
| | CARS＋UVE | 400 | 0.580 3 | 8.107 5 |
| | UVE＋SPA | 30 | 0.592 9 | 8.007 5 |
| | UVE＋CARS＋SPA | 23 | 0.590 9 | 8.022 3 |
| 卷积平滑 Savitzky-golay | CARS | 720 | 0.592 3 | 8.004 0 |
| | SPA | 30 | 0.595 9 | 7.975 5 |
| | UVE | 784 | 0.586 3 | 8.052 2 |
| | CARS＋SPA | 30 | 0.579 3 | 8.187 1 |
| | CARS＋UVE | 418 | 0.588 2 | 8.037 0 |
| | UVE＋SPA | 30 | 0.591 9 | 8.007 9 |
| | UVE＋CARS＋SPA | 20 | 0.577 6 | 8.121 2 |

注：CARS，竞争性自适应重加权采样（Competitive adaptive reweighted sampling）；SPA，连续投影算法（Successive projections algorithm）；UVE，无信息变量消除（Uninformative variable elimination）。“＋”表示方法的组合使用。原始光谱指未经预处理的全波段数据。

Note: CARS, competitive adaptive reweighted sampling; SPA, successive projections algorithm; UVE, uninformative variable elimination. “＋” indicates the combined use of methods. Raw spectrum refers to the full-band data that has not been preprocessed.

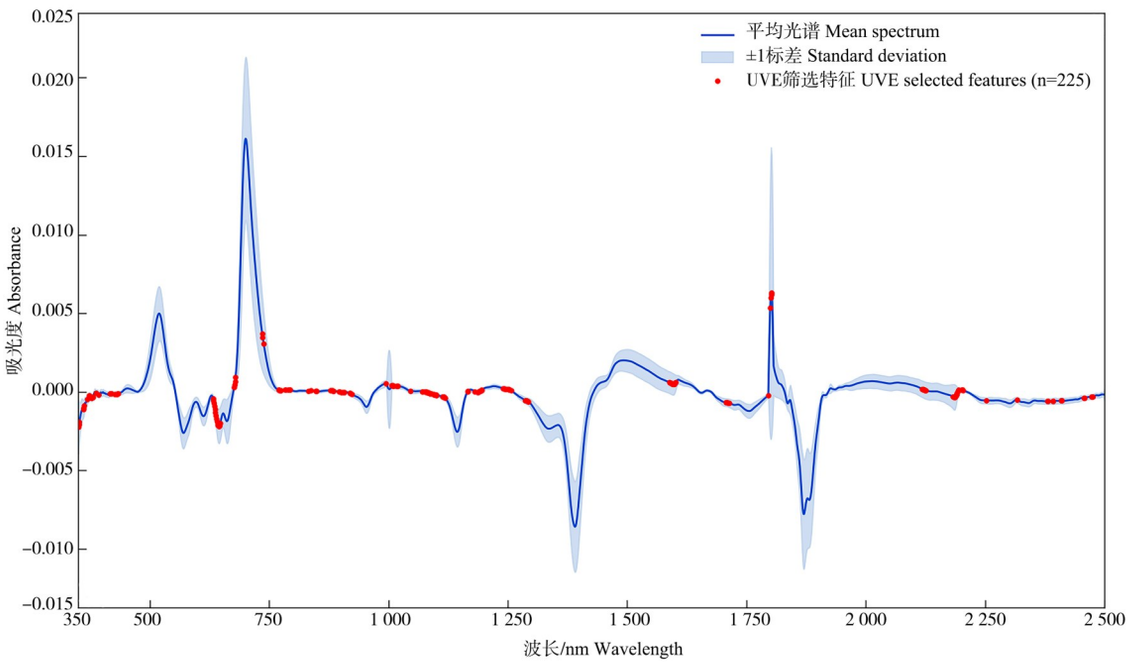



**图 4 特征波长选择可视化图**

Fig. 4 Visualization of feature wavelength selection

氮素吸收区(约 2 050 nm)。

### 2.4 SPAD 值监测模型构建

为验证模型的泛化能力,选取一阶导数 FD+UVE 特征筛选后的数据,分别构建偏最小二乘回归(PLSR)、随机森林(RF)、支持向量回归(SVR)及 BP 神经网络(BPNN)回归模型。将 2025 年采集的 1 705 组数据用作训练集,2026 年采集的 1 137 组数据用作测试集,预测结果如表 2 所示。从决定系数($R^2$)与均方根误差(RMSE)来看,非线性模型(SVR、BPNN 和 RF)在训练阶段均能较好地捕捉样本特征,其中 SVR 展现了最强的拟合潜力。绘制训练集和测试集样本的实测值与预测值散点图,如图 5 所示,SVR 模型在训练集上拟合程度极高,散点紧密分布于 1∶1 参考线附近,决定系数($R^2$)达 0.999 9,均方根误差(RMSE)仅为 0.099 6。BPNN 模型次之,$R^2$ 为 0.992 5,RMSE 为 0.965 0,RF 模型亦表现良好($R^2$=0.955 0,RMSE=2.366 5),三者均显著优于线性 PLSR 模型($R^2$=0.587 1,RMSE=7.169 5)。在测试集上各模型表现出现显著分化(图 6),虽然 SVR 模型在训练集上效果极佳,但在测试集上出现了明显的过拟合现象,散点严重偏离 1∶1 线,预测性能大幅下降($R^2$=0.127 7,RMSE=10.205 6)。相

**表 2 不同回归模型结果对比**

Table 2 Comparison of results from different regression models

| 回归模型 Regression model | 训练集 $R^2$ $R^2$ of training set | 训练集 RMSE RMSE of training set | 测试集 $R^2$ $R^2$ of test set | 测试集 RMSE RMSE of test set |
|---|---|---|---|---|
| 偏最小二乘回归 Partial least squares regression | 0.587 1 | 7.169 5 | 0.498 3 | 7.739 4 |
| 随机森林 Random forest | 0.955 0 | 2.366 5 | 0.655 4 | 6.414 7 |
| 支持向量回归 Support vector regression | 0.999 9 | 0.099 6 | 0.127 7 | 10.205 6 |
| BP 神经网络 Back propagation neural network | 0.992 5 | 0.965 0 | 0.510 4 | 7.646 0 |

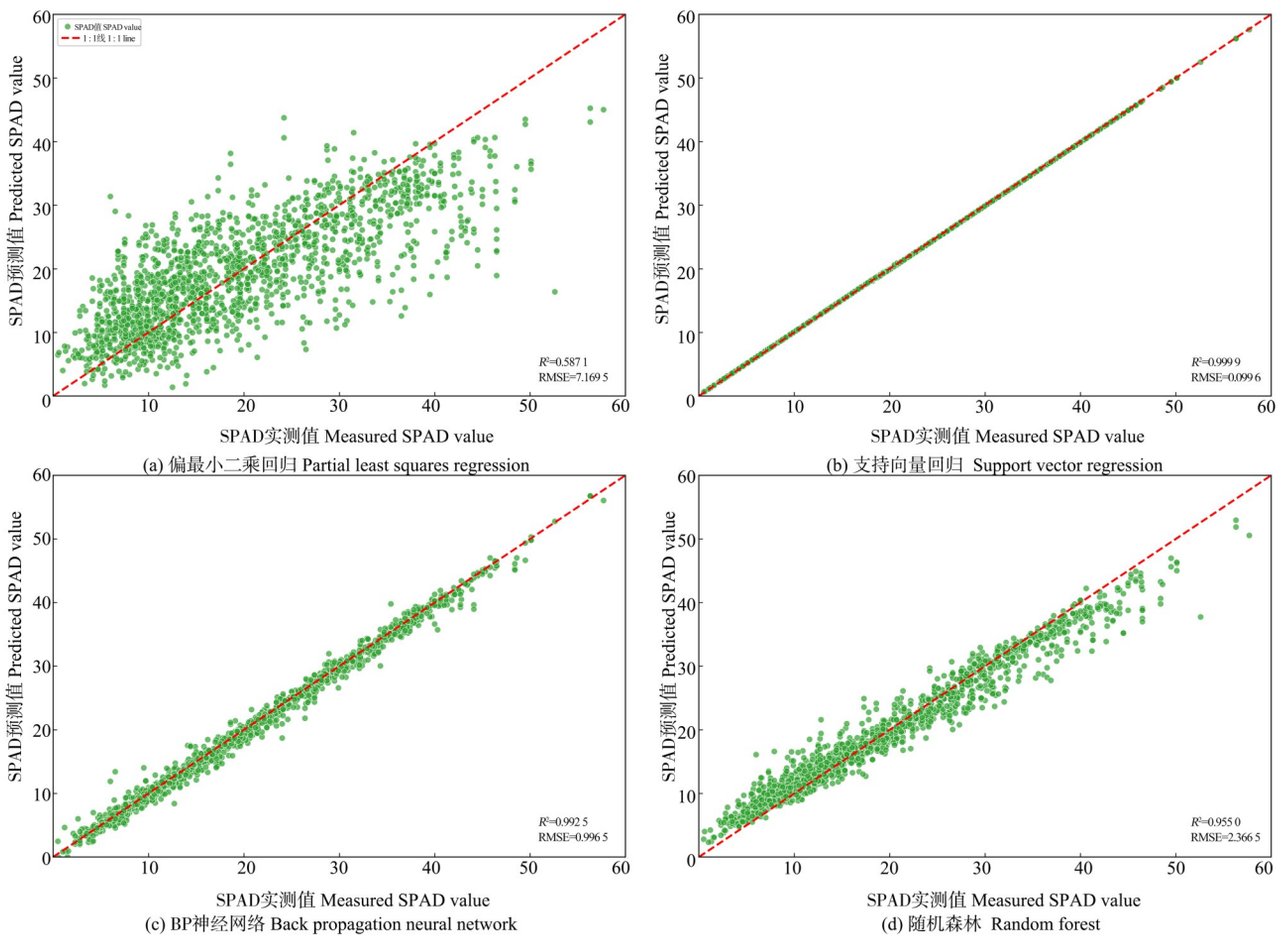



**图 5　4种模型在训练集上预测SPAD值的散点图对比**

Fig. 5　Comparison of scatter plots showing four models predicting chlorophyll content (SPAD value) in tea leaves on the training set

比之下，RF 模型表现出较好的泛化能力，其测试集 $R^2$达到 0.655 4，RMSE 为 6.414 7，显著优于 BPNN（$R^2$ =0.510 4，RMSE=7.646 0）和 PLSR（$R^2$ =0.498 3，RMSE=7.739 4），RF 模型在处理未知数据时表现出较好的稳定性和预测能力。

## 2.5　模型优化

为进一步提升模型的预测精度，采用网格搜索交叉验证（Grid Search CV）对随机森林（RF）模型的关键参数进行寻优，结果如表 3 所示。考察的参数包括决策树数量（n_estimators）、最大深度（max_depth）、最小分裂样本数（min_samples_split）及最大特征数（max_features）。超参数搜索空间设置为：n_estimators∈[200，500]，max_depth∈[10，50]，min_samples_split∈[3，5]，max_features∈

**表 3　模型优化结果**

Table 3　Model Optimization Results

| 数据集 Dataset | 相关系数 Correlation coefficient（$R^2$） | 均方根误差 Root mean square error（RMSE） |
|---|---|---|
| 训练集 Training set | 0.953 0 | 2.400 4 |
| 测试集 Test set | 0.703 2 | 6.022 4 |

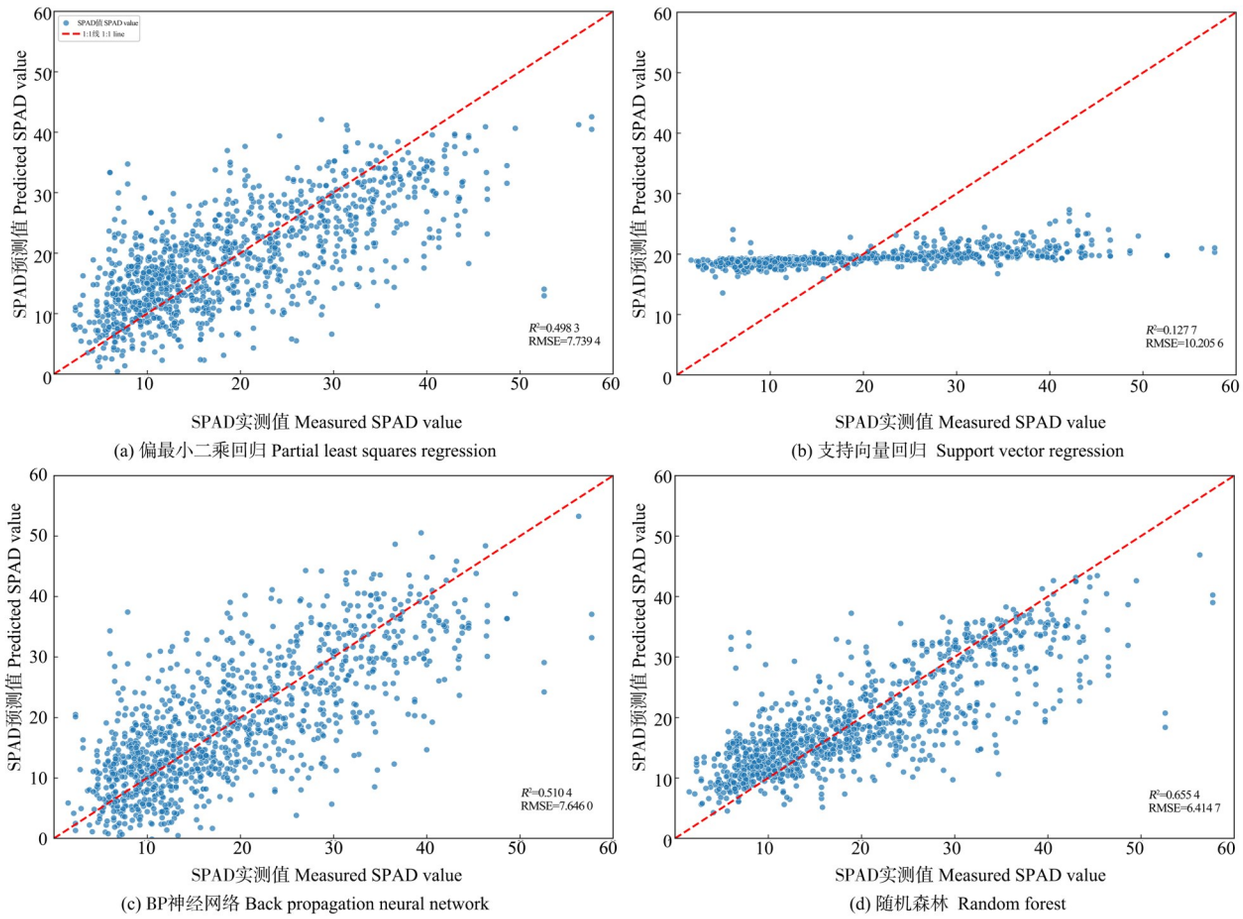



**图 6 4种模型在测试集上预测 SPAD 值的散点图对比**

Fig. 6 Comparison of scatter plots showing the predicted chlorophyll content SPAD value in tea leaves by four models on the test set

['sqrt', 0.3]。模型评估采用5折交叉验证(5-fold CV)。优化后的RF模型在训练集与测试集上的预测性能如图7所示。训练集$R^2$略有下降(0.955 0→0.953 0),但测试集$R^2$显著提升(0.655 4→0.703 2),表明超参数优化有效抑制了过拟合,模型具有较好的泛化能力,能够用于实际的SPAD值预测。经网格搜索得到的最优参数组合为:n_estimators=500, max_depth=20, min_samples_split=3,max_features=0.3。

## 3 讨 论

### 3.1 茶树品种与时期对叶绿素含量的显著影响

单因素方差分析结果表明,不同茶树品种间SPAD值差异极显著($F$=336.12,$p$<0.01),结合总体标准差(13.8)与变异系数(61.1%),表明不同品种间叶绿素积累能力存在显著的遗传分化[22]。白化品种'白金芽'在整个观测期内SPAD值大部分时间低于20,显著低于其他供试品种,这与白化品种因叶绿体发育受阻的已有研究结论一致[23]。相比之下,'福鼎大白茶'SPAD值较高且后期增长显著,表明传统品种在维持较高叶绿素水平方面具有遗传优势[24]。清明前后各品种SPAD值波动明显,可归因于气温剧烈波动及阶段性低温胁迫;除白化品种外,各品种在4月2日后整体变化趋势一致。2025和2026年度2个春茶季品种间SPAD值排序基本稳定,表明品种间叶绿素含量的相对差距具有年际稳定性。

### 3.2 光谱预处理与特征选择方法

通过系统比较5种不同光谱预处理方法与3种特征筛选方法的组合效应,发现不同策略对SPAD监测精度的影响差异显著。在预处理方面,MSC和SNV能有效校正散射效应[25],然而模型$R^2$均未超

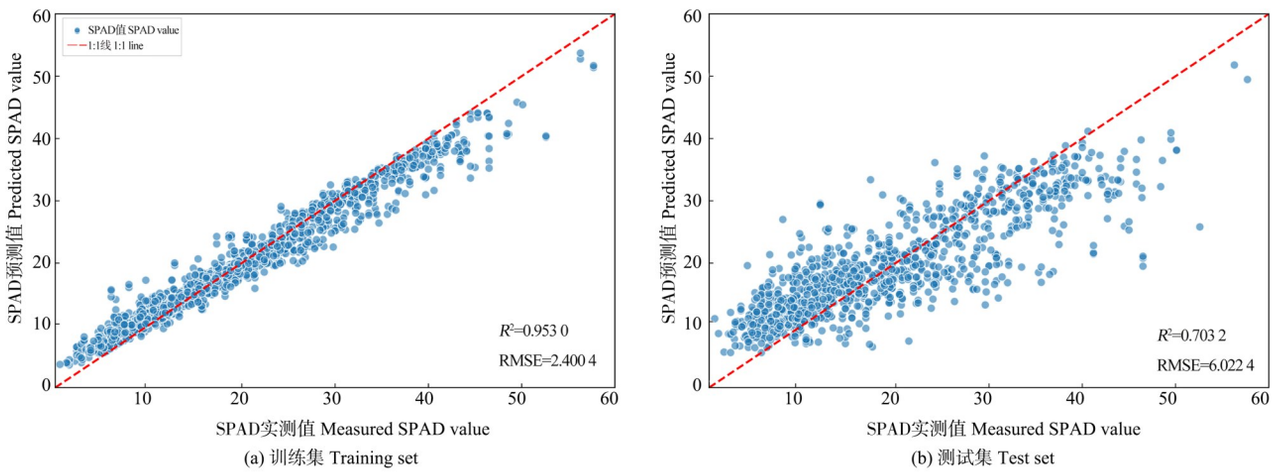



**图 7　优化后的随机森林模型预测性能**

Fig. 7　Predictive performance of the optimized random forest model

过 0.75。一阶导数(FD)在锐化叶绿素吸收峰的同时会放大高频噪声[26]。去趋势处理(DT)和 SG 平滑因过度滤除有效宽带信息[27-28]，$R^2$ 仅为 0.58 左右。在特征选择上，SPA 及 UVE+CARS+SPA 联合算法因波段压缩过激或逻辑互斥造成信息流失，性能均不及单一算法[29-30]。最终，一阶导数+UVE 组合取得最优预测性能($R^2$=0.702 6，RMSECV=7.673 3)，保留 225 个特征波段，所选 225 个波段并非随机分布，而是高度集中于叶绿素、水分及细胞结构等关键生化组分的光谱响应敏感区。其中近红外区(780～1 300 nm)贡献了高达 45.3% 的特征波段，该区主要反映叶片内部细胞排列结构(如栅栏组织密度)，引入这些波段有效校正了由叶片厚度或含水量差异引起的系统误差。

### 3.3　回归模型的监测能力差异

在 4 种回归建模方法中，非线性模型(SVR、BPNN 和 RF)在叶绿素 SPAD 值监测中表现整体优于线性模型，表明叶绿素与光谱反射率间呈非线性关系。其中 SVR 模型呈现出典型的“过拟合”特征[31]，其在训练集上达到了近乎完美的拟合，但在测试集上性能显著下降($R^2$ 由 0.999 9 降至 0.127 7)。其原因可能是模型将品种与年际间差异误作有效信息加以拟合，导致决策边界过度复杂化。相比之下，随机森林(RF)模型展现了较好的综合性能。不仅在训练阶段保持了高拟合度，在测试阶段其散点依然围绕 1∶1 线分布($R^2$=0.655 4，RMSE=6.414 7)。尽管 RF 模型在 4 类模型中表现最优，但跨年度独立验证结果表明，测试集 $R^2$(0.655 4)与训练集 $R^2$(0.955 0)之间仍存在约 0.30 的差距，表明模型在不同品种和年份的泛化方面确实存在挑战。分析原因可能有：一是品种因素，不同品种的叶绿素光谱响应模式可能存在本质差异(尤其是白化品种)，单一全局模型难以同时精确捕捉所有品种的特征；二是年份因素，年际间气象条件的差异可能导致相同 SPAD 值对应不同的光谱表达。未来研究可考虑引入领域自适应或迁移学习策略，将不同品种和年份作为独立的“域”，通过对抗训练或特征对齐来增强模型的跨域泛化能力。

### 3.4　参数寻优对模型性能的提升机制

本研究中，适度限制决策树最大深度并提高估计器数量，有效降低了 RF 模型对训练集中品种特异性噪声的过度拟合，使测试集 $R^2$ 提升 4.78 个百分点(0.655 4→0.703 2)、RMSE 降低 6.1%(6.414 7→6.022 4)，表明针对农业光谱数据的高维强相关特性[32]，通过控制模型复杂度与集成规模之间的平衡，可在保留叶绿素敏感特征的同时抑制噪声拟合，是提升跨品种、跨年度泛化能力的有效途径[33]。最终优化后的 RF 模型验证了“一阶导数 FD+UVE 特征筛选+RF 优化建模”技术路线的有效性。值得指出的是，网格搜索受限于搜索范围与步长设置[34]，未来可引入贝叶斯优化或分品种分层建模策略以进一步改善性能。

## 4　结　论

本研究基于高光谱技术系统探究了光谱预处理、特征波长筛选及机器学习算法对茶鲜叶 SPAD

值监测精度的影响，主要结论如下：

1）不同茶树品种间的SPAD值存在显著差异。6个供试品种间SPAD值变异系数达61.1%。2个春茶季的品种间SPAD值排序基本一致，表明品种间叶绿素含量的相对差异具有跨年度稳定性，为建立具有年际泛化能力的光谱监测模型提供了生物学基础。

2）在对比的42组预处理-特征筛选组合中，一阶导数FD+UVE组合取得了最优的平衡性能（$R^2$=0.702 6，RMSECV=7.673 3），最终筛选得到225个特征波段。UVE+CARS+SPA组合未带来性能提升，表明在茶鲜叶高光谱分析中，单一方法的适度降维优于多方法的激进压缩。

3）在监测性能上，非线性模型（RF、SVR、BPNN）整体优于线性PLSR，其中RF模型表现最优：训练集$R^2$为0.955 0、RMSE为2.366 5，测试集$R^2$为0.655 4，RMSE为6.414 7，在四类模型中测试集预测精度最高，其集成学习机制有效降低了模型方差，对噪声和异常值具有较好的鲁棒性。

4）经网格搜索超参数优化后，RF模型的测试集$R^2$提升4.78个百分点（0.655 4→0.703 2），训练集与测试集性能差距显著缩小，表明超参数调优在调节模型复杂度、改善偏差—方差权衡方面发挥了重要作用。

综上，本研究通过42组预处理-特征筛选组合的系统比较，明确了“一阶导数FD+UVE+RF优化”方法在茶鲜叶SPAD值高光谱监测中的适用性。该方法在6个品种跨年度数据集上取得了较好的预测精度，且筛选波段与叶绿素、水分及氮素吸收特征区间具有物理对应关系。但跨年度验证表明模型年际泛化能力仍有局限，实际茶园原位应用还面临光照变化等环境干扰。后续研究将引入迁移学习与滚动更新策略，并结合主动光源光谱仪开展室外验证，以验证该技术在田间原位条件下的适用性。

## 参考文献 References


[1] Batool D, Shahbaz M, Shahzad Asif H, Shaukat K, Alam T M, Hameed I A, Ramzan Z, Waheed A, Aljuaid H, Luo S H. A hybrid approach to tea crop yield prediction using simulation models and machine learning[J]. *Plants*, 2022, 11(15): 1925

[2] Wu W D, Pei G Q, Lu Z Q, Zhou B, Qian X Y, Wang B J, Yang L N. Lightweight multi-view fusion network for non-destructive chlorophyll and nitrogen content estimation in tea leaves using front and back RGB images[J]. *Agronomy*, 2025, 15(10): 2355

[3] 涂淑萍，黄航，杜曲，陶凌剑. 不同品种茶树叶片光合特性与叶绿素荧光参数的比较[J]. 江西农业大学学报，2021，43(5)：1098-1106
Tu S P, Huang H, Du Q, Tao L J. Comparison of photosynthetic characteristics and chlorophy ll fluorescence parameters of different *Camellia sinensis* cultivars[J]. *Acta Agriculturae Universitatis Jiangxiensis*, 2021, 43(5): 1098-1106 (in Chinese)

[4] Chen J M, Wu S H, Dong F, Li J L, Zeng L T, Tang J C, Gu D C. Mechanism underlying the shading-induced chlorophyll accumulation in tea leaves[J]. *Frontiers in Plant Science*, 2021, 12: 779819

[5] Raza A, Li L, Luo L, Asghar M A, Shoaib N, Yin C Y. Rational application of combined fertilizers improved tea growth and quality components[J]. *Journal of the Science of Food and Agriculture*, 2025, 105(7): 3923-3930

[6] 刘燕德，林晓东，高海根，王舜，高雪. 基于叶绿素荧光光谱技术的茶叶藻斑病模型研究[J]. 光谱学与光谱分析，2021，41(7)：2129-2134
Liu Y D, Lin X D, Gao H G, Wang S, Gao X. Research on tea *Cephaleuros* virescens kunze model based on chlorophyll fluorescence spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(7): 2129-2134 (in Chinese)

[7] 邹振浩，沈晨，李鑫，张丽平，张兰，韩文炎，颜鹏. 我国茶园氮肥利用和损失现状分析[J]. 植物营养与肥料学报，2021，27(1)：153-160
Zou Z H, Shen C, Li X, Zhang L P, Zhang L, Han W Y, Yan P. Status quo of nitrogen fertilizer application and loss in tea garden of China[J]. *Journal of Plant Nutrition and Fertilizers*, 2021, 27(1): 153-160 (in Chinese)

[8] 余涛，胡波，孙睿，金志凤，王岳飞，张蕾，徐伟燕，刘刚. 茶叶光谱与叶绿素、茶氨酸、茶多酚含量关系分析[J]. 遥感技术与应用，2016，31(5)：872-878
Yu T, Hu B, Sun R, Jin Z F, Wang Y F, Zhang L, Xu W Y, Liu G. Relationship between tea spectra and contents of chlorophyll, theanine and polyphenols[J]. *Remote Sensing Technology and Application*, 2016, 31(5): 872-878 (in Chinese)

[9] 潘俊虹，程仁贵，王平，余文森. 导数变换结合机器学习的茶树冠层高光谱氮素反演[J]. 江苏农业科学，2026，54(4)：98-105
Pan J H, Cheng R G, Wang P, Yu W S. Hyperspectral nitrogen retrieval of tea plant canopy by derivative transformation combined with machine learning [J]. *Jiangsu Agricultural Sciences*, 2026, 54(4): 98-105 (in Chinese)

[10] Vaddi R, Phaneendra Kumar B L N, Manoharan P, Agilandeeswari L, Sangeetha V. Strategies for dimensionality reduction in hyperspectral remote sensing: A comprehensive overview[J]. *The Egyptian Journal of Remote Sensing and Space Sciences*, 2024, 27(1): 82-92

[11] Wang S P, Altaner C, Feng L, Liu P P, Song Z Y, Li L Q, Gui A H, Wang X P, Ning J M, Zheng P C. A review: Integration of NIRS and chemometric methods for tea quality control-principles, spectral preprocessing methods, machine learning algorithms, research progress, and future directions[J]. *Food Research International*, 2025, 205: 115870

[12] 赵星辉，冯雪琦，聂小军，张岩，郭二辉. 基于CARS-CNN和高光谱遥感的矿区土壤磷含量估算[J]. 生态学杂志，2024，43(12)：3553-3562
Zhao X H, Feng X Q, Nie X J, Zhang Y, Guo E H. Hyperspectral estimation of soil phosphorus content in mining area based on CARS-CNN[J]. *Chinese Journal of Ecology*, 2024, 43(12): 3553-3562 (in Chinese)

[13] 孟珊，李新国，江远东，麦麦提吐尔逊·艾则孜. 基于CWT-RF模型估算博斯腾湖湖滨绿洲土壤有机碳含量[J]. 中国农业大学学报，2023，28(3)：198-207

Meng S, Li X G, Jiang Y D, EZIZ Mamattursun. Estimation of soil organic carbon content in lakeside oases based on CWT-RF model[J]. *Journal of China Agricultural University*, 2023, 28(3): 198-207 (in Chinese)

[14] 江景, 赵紫薇, 蔡唱, 张劲松, 程志庆. 滞尘影响下的茶树叶片水分高光谱估算[J]. 光谱学与光谱分析, 2021, 41(11): 3532-3537
Jiang J, Zhao Z W, Cai C, Zhang J S, Cheng Z Q. Hyperspectral estimation of tea leaves water content under the influence of dust retention [J]. *Spectroscopy and Spectral Analysis*, 2021, 41(11): 3532-3537 (in Chinese)

[15] 徐阳, 毛艺霖, 李赫, 王玉, 王双双, 钱文俊, 丁兆堂, 范凯. 基于多光谱和高光谱的茶树越冬期REC、SPAD和MDA预测模型[J]. 光谱学与光谱分析, 2025, 45(1): 256-263
Xu Y, Mao Y L, Li H, Wang Y, Wang S S, Qian W J, Ding Z T, Fan K. Multispectral and hyperspectral prediction models of REC, SPAD and MDA in overwintered tea plant[J]. *Spectroscopy and Spectral Analysis*, 2025, 45(1): 256-263 (in Chinese)

[16] 李晓丽, 张东毅, 董雨伦, 金娟娟, 何勇. 基于卷积神经网络的茶鲜叶主要内含物的光谱快速检测方法[J]. 中国农业大学学报, 2021, 26(11): 113-122
Li X L, Zhang D Y, Dong Y L, Jin J J, He Y. Spectral rapid detection of phytochemicals in tea (*Camellia sinensis*) based on convolutional neural network[J]. *Journal of China Agricultural University*, 2021, 26(11): 113-122 (in Chinese)

[17] Jiang J, Ji H T, Yan Y, Zhao L Y, Pan R Y, Liu X J, Yin J, Duan Y, Ma Y C, Zhu X J, Fang W P. Mining sensitive hyperspectral feature to non-destructively monitor biomass and nitrogen accumulation status of tea plant throughout the whole year[J]. *Computers and Electronics in Agriculture*, 2024, 225: 109358

[18] 廖珺, 方洪生, 苏有健, 王烨军, 张永利, 孙宇龙, 方雅各. 不同摊放环境下茶鲜叶失水的变化规律[J]. 中国农学通报, 2023, 39(4): 160-164
Liao J, Fang H S, Su Y J, Wang Y J, Zhang Y L, Sun Y L, Fang Y G. Change law of water loss in fresh tea leaves under different spreading environments[J]. *Chinese Agricultural Science Bulletin*, 2023, 39(4): 160-164 (in Chinese)

[19] Wei Y Z, Wu F Y, Xu J, Sha J J, Zhao Z F, He Y, Li X L. Visual detection of the moisture content of tea leaves with hyperspectral imaging technology [J]. *Journal of Food Engineering*, 2019, 248: 89-96

[20] 朱栗琼, 覃冬梅, 招礼军, 邓斌胜, 刘晟源, 蒋昌杰. 6种野生金花茶叶表皮微形态特征及其系统学意义[J]. 植物研究, 2021, 41(6): 841-850
Zhu L Q, Qin D M, Zhao L J, Deng B S, Liu S Y, Jiang C J. Leaf epidermal micromorphological features and their systematic significance of six wild species of Camellia chrysantha[J]. *Bulletin of Botanical Research*, 2021, 41(6): 841-850 (in Chinese)

[21] 覃爱颖, 陈恩海, 劳福增, 陈达辉, 钟勇, 莫干辉. 基于叶片解剖结构特征的50份茶树品种(系)综合评价与性状关联分析[J]. 热带作物学报, 2025, 46(10): 2376-2389
Qin A Y, Chen E H, Lao F Z, Chen D H, Zhong Y, Mo G H. Leaf anatomical determinants of stress resistance and productivity: A comprehensive evaluation of 50 tea accessions in Guangxi[J]. *Chinese Journal of Tropical Crops*, 2025, 46(10): 2376-2389 (in Chinese)

[22] 刘东娜, 龚雪蛟, 李兰英, 黄藩, 尧渝, 胥亚琼, 高远, 罗凡. 黄化茶树叶片光合及荧光特性分析[J]. 茶叶科学, 2023, 43(6): 757-768
Liu D N, Gong X J, Li L Y, Huang F, Yao Y, Xu Y Q, Gao Y, Luo F. Analysis of photosynthetic and fluorescence characteristics of albino tea plants [J]. *Journal of Tea Science*, 2023, 43(6): 757-768 (in Chinese)

[23] Chen Y J, Han Y X, He S Q, Cheng Q H, Tong H R. Differential metabolic profiles of pigment constituents affecting leaf color in different albino tea resources[J]. *Food Chemistry*, 2025, 467: 142290

[24] 刘晓璐, 朱亚兰, 于敏, 盖新月, 范延艮, 孙平, 黄晓琴. 低温胁迫下茶树叶片细胞壁结构变化及光合特性[J]. 茶叶科学, 2024, 44(6): 917-927
Liu X L, Zhu Y L, Yu M, Gai X Y, Fan Y G, Sun P, Huang X Q. Changes in cell wall structure and photosynthetic characteristics of tea leaves under low temperature stress[J]. *Journal of Tea Science*, 2024, 44(6): 917-927 (in Chinese)

[25] 王鑫磊, 韩鲁佳, 杨增玲, 肖卫华, 王婕妤, 于涞源, 戴嘉伟, 朱礼强. 预处理方法对异源近红外秸秆营养成分速测模型的影响[J]. 农业工程学报, 2025, 41(9): 216-226
Wang X L, Han L J, Yang Z L, Xiao W H, Wang J Y, Yu L Y, Dai J W, Zhu L Q. Influence of preprocessing methods on the rapid quantitative models for crop straw nutrients using different near-infrared spectroscopy devices data[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2025, 41(9): 216-226 (in Chinese)

[26] 第五鹏瑶, 卞希慧, 王姿方, 刘巍. 光谱预处理方法选择研究[J]. 光谱学与光谱分析, 2019, 39(9): 2800-2806
Diwu P Y, Bian X H, Wang Z F, Liu W. Study on the selection of spectral preprocessing methods [J]. *Spectroscopy and Spectral Analysis*, 2019, 39(9): 2800-2806 (in Chinese)

[27] Yuan R R, Liu G S, He J G, Wan G L, Fan N Y, Li Y, Sun Y R. Classification of Lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis[J]. *Computers and Electronics in Agriculture*, 2021, 182: 106043

[28] Zhou L, Wu H B, Jing T T, Li T H, Li J S, Kong L J, Zhou L N. Estimation of relative chlorophyll content in lettuce (*Lactuca sativa L.*) leaves under cadmium stress using visible: Near-infrared reflectance and machine-learning models[J]. *Agronomy*, 2024, 14(3): 427

[29] Du L, Jin Z L, Chen B W, Chen B W, Gao W, Yang J, Shi S, Song S L, Wang M M, Gong W, Wang W. Application of hyperspectral LiDAR on 3-D chlorophyll-nitrogen mapping of rohdea japonica in laboratory[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 9667-9679

[30] Gao L, Darvishzadeh R, Somers B, Johnson B A, Wang Y, Verrelst J, Wang X F, Atzberger C. Hyperspectral response of agronomic variables to background optical variability: Results of a numerical experiment [J]. *Agricultural and Forest Meteorology*, 2022, 326: 109178

[31] Jarro-Espinal I, Huanuqueño-Murillo J, Quille-Mamani J, Quispe-Tito D, Ramos-Fernández L, Pino-Vargas E, Torres-Rua A. Field-scale rice yield prediction in northern coastal region of Peru using sentinel-2 vegetation indices and machine learning models[J]. *Agriculture*, 2025, 15(19): 2054

[32] 周婧, 雷刚, 卢鑫, 邓萌, 阚飞, 黄扬. 中高分辨率影像种植结构遥感提取方法比选研究[J]. 节水灌溉, 2026: 1-15
Zhou J, Lei G, Lu X, Deng M, Kan F, Huang Y. Research on the comparison of remote sensing extraction methods for planting structures in medium and high resolution images[J]. *Water Saving Irrigation*, 2026: 1-

15 (in Chinese)

[33] Sun Z G, Wang G T, Li P F, Wang H, Zhang M, Liang X W. An improved random forest based on the classification accuracy and correlation measurement of decision trees[J]. *Expert Systems with Applications*, 2024, 237: 121549

[34] 熊向阳, 杨小周, 赵银超, 李伟坡. 基于超参数优化随机森林算法的森林生物量遥感反演[J]. 中南林业科技大学学报, 2024, 44(5): 102-111
Xiong X Y, Yang X Z, Zhao Y C, Li W P. Remote sensing inversion of forest biomass based on hyperparametric optimized random forests algorithm[J]. *Journal of Central South University of Forestry & Technology*, 2024, 44(5): 102-111 (in Chinese)

责任编辑: 梁芷健



**第一作者简介:** 叶春,博士,副研究员。主要从事农业工程研究,发表期刊论文10余篇,授权发明专利8项,参与获得江西省科技进步奖二等奖1项。



**通讯作者简介:** 吴罗发,研究员,硕士生导师。主要从事农业工程研究,江西省农业机械化协会会长,江西省农业工程学会副会长,发表期刊论文100余篇,授权发明专利32项,主持获得江西省科技进步二等奖等奖项10余项。