中国粮油学报
Journal of the Chinese Cereals and Oils Association
ISSN 1003-0174,CN 11-2864/TS

《中国粮油学报》网络首发论文

题目：基于 XGBoost 的掺伪茶油光谱鉴别模型
作者：龚中良，刘强，李大鹏，文韬，管金伟，易宗霈，申飘
DOI：10.20048/j.cnki.issn.1003-0174.000047
收稿日期：2022-07-16
网络首发日期：2022-10-27
引用格式：龚中良，刘强，李大鹏，文韬，管金伟，易宗霈，申飘．基于 XGBoost 的掺伪茶油光谱鉴别模型[J/OL]．中国粮油学报．
https://doi.org/10.20048/j.cnki.issn.1003-0174.000047

DOI：10.20048/j.cnki.issn.1003-0174.000047 网络首发时间：2022-10-27 14:58:58
网络首发地址：https://kns.cnki.net/kcms/detail/11.2864.ts.20221026.1733.002.html

基于 XGBoost 的掺伪茶油光谱鉴别模型

龚中良，刘强，李大鹏，文韬，管金伟，易宗霈，申飘

（中南林业科技大学机电工程学院，长沙 410004）

摘 要：为实现采用紫外−可见−近红外光谱技术鉴别掺伪茶油的目的，本研究首先通过向茶油中掺入不同比例的葵花籽油、玉米胚芽油和花生油制备掺伪茶油，然后采用自制的透射光谱采集试验平台获得光谱数据，对原始光谱进行预处理后，分别以竞争性自适应重加权算法（CARS）、连续投影算法（SPA）、Boruta 算法进行特征波长筛选，最后建立了基于 XGBoost 的掺伪茶油鉴别模型。研究结果表明，原始光谱经过 SG-连续小波变换（CWT（分解尺度2^5，L5））预处理和 Boruta 特征波长筛选后，所建立的 XGBoost 模型鉴别性能最佳，测试集的准确率、灵敏度和特异性分别达到了 98.18%、100%和 97.62%。通过与常用的支持向量机（SVM）和极限学习机（ELM）模型对比后得到，XGBoost 模型的准确率分别提高了 3.63%和 1.82%，特异性分别提高了 4.76%和 2.38%。研究成果为开发基于紫外-可见-近红外光谱的掺伪茶油检测装置奠定了基础。

关键词：茶油；紫外-可见-近红外光谱；XGBoost；鉴伪

中图分类号：TS225.1 TS227 文献标识码：A

Spectral Authentication Model of Adulterated Camellia Oil Based on XGBoost

Gong Zhongliang，Liu Qiang，Li Dapeng，Wen Tao，Guan Jinwei，Yi Zongpei，Shen Piao

（School of Mechanical and Electrical Engineering, Central South University of Forestry and Technology，Changsha 410004）

Abstract：To realize adulterated camellia oil (CAO) authentication using UV-Vis-NIR spectroscopy, this study firstly prepared the samples of adulterated CAO by mixing different proportions of sunflower oil, corn germ oil and peanut oil, respectively, into CAO. Then, a homemade transmission spectrum acquisition rig was used to obtain their raw spectral data, which was preprocessed and screened to get the characteristic wavelengths by competitive adaptive re-weighting algorithm (CARS), continuous projection algorithm (SPA), and Boruta algorithm, respectively. Finally, an XGBoost-based authentication model of adulterated CAO was established. The results show that the XGBoost model has the best performance after the preprocessing of SG- continuous wavelet transform (CWT (decomposition scale 2^5, L5)) and Boruta feature wavelength screening. This best model leads to the accuracy, sensitivity and specificity of 98.18%, 100% and 97.62%, respectively. Compared with the commonly used support vector machine (SVM) and extreme learning machine (ELM) models, the accuracy of the XGBoost model is improved by 3.63% and 1.82%,respectively, and the specificity by 4.76% and 2.38%.The study laid the foundation for the development of an adulterated CAO detection device

基金项目：湖南省科技计划重点研发项目（2022NK2048）；湖南省教育厅科学项目（18B192，20A515）；湖南省自然科学基金（2020JJ4142）；湖南省林业杰青培养科研项目（XLK202108-7）

收稿日期：2022-07-16

第一作者：龚中良，男，1965 年出生，教授，主要从事农业装备自动化研究

通讯作者：李大鹏，男，1983 年出生，博士后，主要从事农林产品无损品质检测技术及装备研究

based on UV-vis-NIR spectroscopy.

Keywords：Camellia oil; UV-Vis-NIR spectroscopy;XGBoost;authentication

茶油是世界四大木本油脂之一，在食品保健、医疗、美妆、化工等领域极具发展潜力[1,2]。由于茶油所具备的优越理化指标导致其售卖价格为普通植物油的 5~10 倍[3]，面对市场诱惑，不良商贩通过在高价茶油中掺入低价油非法牟利，从而损害消费者利益。

目前，用于食用油鉴伪的方法主要有色谱法[4]、核磁共振法[5]、电子鼻[6]、光谱法[7]等，但色谱法、核磁共振法、电子鼻检测分别存在试剂污染油样、设备操作复杂、仪器数据稳定性低等问题，因此无法满足无损、便捷、稳定的茶油快速鉴伪需求。近年来，各类光谱检测技术依靠其直接、无损的特点被广泛应用于食用油鉴伪中，Zhang 等[8]通过 12000~4000cm^{-1} (833~2500nm)光谱波段对五种油掺伪玉米油制作多元样本，通过偏最小二乘（PLS）模型的建立得到预测集决定系数（R^2）均在 0.93 以上。Wu 等[9]利用紫外可见光谱（350~800nm）和加权多尺度支持向量机（EMD-SVR）建立大豆油、花生油、芝麻油组成的二元、三元定量模型，其相关系数（R）分别为 0.9533 和 0.9866，证明低频波段也可以作为食用油有效鉴伪的依据。郭文川等[3]以 833~2500nm 的近红外光谱完成对茶油掺伪 4 种低价油的光谱采集，并通过对比多种方法得出在连续投影法（SPA）波长选择后的随机森林（RF）模型的效果最佳，其识别准确率为 99.34%，但其光谱采集是使用傅里叶光谱仪的近红外波段,鉴别成本较高。韩建勋等[10]通过 4000~650cm^{-1}(2500~15385nm)的光谱波段结合主成分分析法（PCA）实现山茶油掺伪大豆油、菜籽油、玉米油的定性判别，同时以偏最小二乘回归算法（PLSR）建立山茶油掺伪大豆油的定量模型，其校正集和验证集的决定系数(R^2)均能达到 0.99。荣菡等[11]利用 10000～4200cm^{-1}(1000~2381nm)的光谱波段以马氏距离聚类分析法和反向传播神经网络分别建立茶油掺伪菜籽油、棕榈油定性模型（掺伪浓度比例 10%~40%），其准确率均为 100%，但未能实现掺伪浓度比例 10%以下的快速鉴伪。可见，目前基于光谱法的茶油鉴伪研究通常采用近红外光谱（1000~2500nm）,尽管可以较准确的鉴别出掺伪茶油，但是其采用的光谱设备成本显著高于紫外-可见-近红外光谱（200~1100nm）[12],因此限制了掺伪茶油光谱鉴别仪器的开发和推广。

本研究拟采用紫外-可见-近红外光谱（200~1100nm）进行掺伪茶油的鉴别，通过对比多种预处理方法、特征波长选择方法和建模算法，建立了准确率、灵敏度和特异性较高的掺伪茶油鉴别模型，从而为开发低成本的掺伪茶油光谱检测装置奠定了基础。

1 材料与方法

1.1 材料

从长沙大型超市购买不同品牌植物油，包括 3 种成品茶油、2 种成品花生油、2 种成品葵花籽油、2 种成品玉米胚芽油。上述试验用植物油均为压榨工艺生产，上述成品茶油均符合 GB/T 11765-2018，试验期间所用油品均在保质期内。

在制备掺伪茶油样品时，以 10ml 液体容量为标准，将 2 种花生油、葵花籽油、玉米胚芽油分别混入 3 种茶油中，并按照掺伪比例分别为 1%、3%、5%、7%、9%、12%、15%、20%、35%制备 162 个掺伪茶油样品。制备时，将油样放入磁力搅拌机中，在 35°C 下搅拌 1h，随后静置 24h。另外，为增加纯茶油的区分性，按照茶油品牌各制备 20 个样品，共得到 60 个纯茶油样品。

1.2 试验装置

搭建了用于油样透射光谱采集的试验平台（图 1），该平台主要由暗箱、比色皿固定支架、两个探头固定支架、两根光纤、OceanView Maya2000 pro 光谱仪、HL1000 卤钨灯光源、OceanView 光谱采集软件、石英比色皿组成。

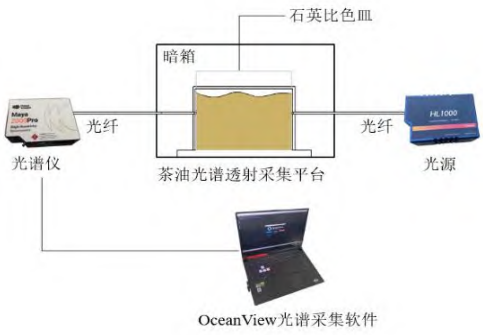

图 1 透射光谱采集试验平台

光谱采样范围为 200~1100nm，设置光谱积分时间为 32ms，扫描次数为 100 次。试验前，将光谱仪与灯源设备提前开启 40min，以达到预热效果。

同一样品采集三次光谱后计算平均值，并通过公式（1）得到样品吸光度数据。

$$A_\lambda = -\mathrm{Log}_{10}\left(\frac{S_\lambda - D_\lambda}{R_\lambda - D_\lambda}\right) \quad (1)$$

式中，A_λ为吸光度，S_λ为采集样本光谱强度，D_λ为暗光谱背景强度，R_λ为空比色皿参考光谱强度。

1.3 分析方法

1.3.1 光谱预处理与样本划分

在采集油样光谱过程中会出现杂散光、基线偏移、电噪声等问题，从而使光谱有效信息混乱、丢失。为提高模型准确性，选用标准正态变量变换（SNV)、多元散射矫正(MSC)、Savitzky-Golay（S-G）平滑处理、移动平均平滑法（MA）、一阶导数（1^{st}Deriv）、二阶导数（2^{nd}Derive）、SG-1^{st}Deriv、SG-2^{nd}Derive、SG-连续小波变换（CWT）等方法进行光谱预处理。

样品训练集与测试集的划分合理性影响模型的预测能力，本研究以 Kennard-Stone(K-S)方法对数据集进行有效划分。

1.3.2 特征波长选择

为实现茶油快速鉴伪，本研究以竞争性自适应重加权算法（CARS）、连续投影算法（SPA）、Boruta 算法对全光谱进行特征波长筛选。

CARS 以权重较大波长点建立 PLS 模型，经过多次循环筛选出特征波长[13]。CARS 运行时设置蒙特卡洛运行次数为 1000 次，每次抽取 80%样品作为校正集，通过 10 折交叉验证循环筛选。

SPA 利用向量投影来优选出冗余度低、共线性小、反应样品光谱关键信息的有效特征波长[14]。设置 SPA 降维后的波长数量范围为 1~30。

Boruta 算法是基于随机森林（RF）[15]构建出的特征筛选方法，它通过加入与真实光谱变量相同数目的乱序影子变量构建新特征集，并基于 RF 计算影子变量和真实光谱变量之间的重要性得分（Z-scores），将得分大于影子变量的光谱变量认定为特征变量[16]。Boruta 以全光谱 2068 个波长变量和对应产生的 2068 个乱序影子变量组成 4136 个全新子集，运行过程中将影子变量中重要性得分最大值标记为 Max_Shadow，得分大于 Max_Shadow 的波长变量被认定为特征变量。

1.3.3 模型的建立与评价

本研究拟采用 XGBoost 算法建立掺伪茶油鉴别模型。XGBoost[17]是基于梯度提升决策树的改进，利用不断新增树的形式来学习新函数，通过新函数去拟合前次产生的残差，从而不断降低误差。XGBoost 通过权重缩减参数（η）调节每棵树的影响，为后续迭代余留更

大的学习空间[18]。由于 XGBoost 算法具有高效可扩展、鲁棒性强等优势，本研究以 gbtree 作为弱学习器类型建立 XGBoost 茶油鉴伪模型。

另外，还将 XGBoost 算法与常用的支持向量机（SVM）和极限学习机（ELM）的建模效果进行对比。SVM 常被用来处理非线性、高维模式识别方面的问题[19]，本文选取径向基函数作为核函数建立 SVM 鉴别模型。 ELM 属于单隐层神经网络，其具备计算快、泛用性强等优势[20]，本文将 Sigmoidal 函数作为激活函数建立 ELM 鉴伪模型。

本研究以准确率（ACC）、灵敏度(TPR)、特异性(FPR)作为各模型分类能力的评价指标。其中，ACC 代表所有正确分类样本数与总样本数的比例。TPR 代表被正确分类为纯茶油的样本数与总纯茶油样本数的比例，检验了模型对纯茶油的鉴别能力；FPR 代表被正确分类为掺伪茶油的样本数与总掺伪茶油样本数的比例，检验了模型对掺伪茶油的鉴别能力。具体公式如（2）（3）（4）。

$$ACC = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

$$TPR = \frac{TP}{(TP+FN)} \quad (3)$$

$$FPR = \frac{TN}{(FP+TN)} \quad (4)$$

式中，TP-正确分类为纯茶油的样本数量；TN-正确分类为掺伪茶油的样本数量；FP-错误分类为纯茶油的样本数量；FN-错误分类为掺伪茶油的样本数量。

2 结果与讨论

2.1 光谱分析

样品吸光度曲线如图 2 所示。在 200~1100nm 范围内出现了 5 个吸收峰，其中紫外光部分在 250nm 左右处出现吸收峰，可见光部分在 430nm 左右处、660nm 左右处出现吸收峰，近红外光部分在 930nm 左右处、1050nm 左右处出现吸收峰。250nm 左右处吸收峰由二元共轭结构产物和三元共轭结构产物产生，430nm 左右处为索雷特特征峰，660nm 左右处吸收峰为-C-H 伸缩振动的五级倍频，940nm 左右的吸收峰为-C-H 三级倍频，1050nm 左右的吸收峰为-O-H 伸缩振动的二级倍频[21–23]。样品光谱曲线之间重叠严重难以直接进行区分，因此本研究借助化学计量学和机器学习方法对全光谱进行进一步分析。

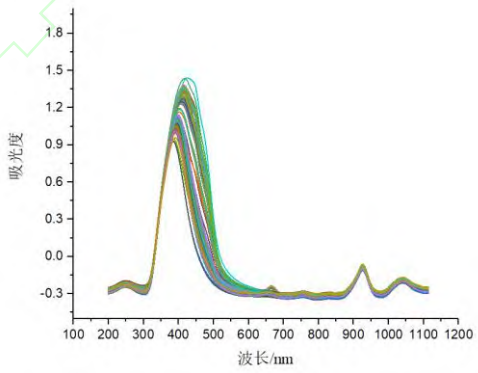

图 2 样品原始光谱

2.2 光谱预处理及样本划分

针对全光谱数据分别以 MSC、SNV、MA、SG、1^{st}Deriv、2^{nd}Derive、SG-1^{st}Deriv、SG-2^{nd}Derive、SG-CWT 等方法进行预处理，其中 SG-CWT 预处理中分解尺度 (n)按照2^n应小于全光谱波长数目（2068）的原则，将其分为 10 个尺度。光谱预处理后将光谱矩阵带

入 SVM 茶油鉴伪模型中，通过对比各光谱矩阵交叉验证率与验证集准确率来选取最优预处理方式。

通过图 3，综合分析多种预处理后的 SVM 茶油鉴伪模型的交叉验证率和验证集准确率，得出 SG-CWT(L5)在各指标的综合性能上优于其他预处理方法，原因在于 SG-CWT 预处理中起初将光谱曲线中不显著的特征峰谷逐步放大（L1-L5），但后续随着分解尺度的增大光谱曲线变得更加平滑（L6-L10），从而使一些不明显的特征峰谷被逐步去除，增大了光谱特征信息的捕捉难度。因此后续模型建立中只针对 SG-CWT(L5)预处理方法进行分析。

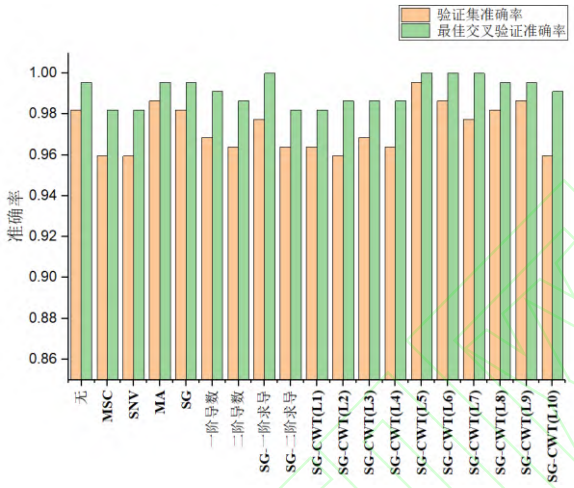

图 3 不同光谱预处理方法对应的验证集准确率与交叉验证准确率

通过 K-S 划分法按照 3：1 的比例将预处理后的 222 个样品光谱数据分为训练集与测试集。训练集共 167 个数据，其中纯茶油样品 47 个掺假油样品 120 个；测试集共 55 个数据，其中纯茶油样品 13 个掺假油样品 42 个。

2.3 特征波长选择和分布特性

2.3.1 特征波长选择

CARS 算法在 1~60 次筛选过程中 RMSECV 值在不断减少，在 60 次后 RMSECV 值不断增大。由于第 60 次筛选时 RMSECV 值为最小值为 0.2084，因此 60 次筛选后剩余的 33 个波长变量即为所选择的最优特征波长。

SPA 算法在选择 21 个波长数目时 RMSE 值最小为 0.25968，之后虽然选择波长数目增加但 RMSE 降幅很小，因此 RMSE 最低处所选择的 21 个波长即为所筛选的特征波长。

Boruta 特征选择算法中通过网格搜索法，得到 RF 模型优化后的最佳决策树数量为 61 棵。Boruta 通过 100 次迭代对比各光谱变量与 Max_Shadow 之间的重要性得分，将 47 个波长认定为重要性特征波长，1954 个波长认定为不重要性波长，67 个波长认定为可能重要的波长，其中 67 个可能重要的波长经后续判断被全部认定为不重要的波长。

2.3.2 特征波长分布特性研究

通过 CARS、SPA、Boruta 方法对光谱数据进行特征波长筛选，分别将波长数目降至全光谱的 1.59%、1.01 %、2.27%。对比紫外、可见、近红外光光谱波段占比特征波长数量（表 1 ），发现 CARS 提取的特征波长主要集中在近红外波段，SPA 提取的特征波长在三个波段的比重较为均匀，Boruta 提取的特征波长主要集中在紫外波段，同时对 3 种方法所筛选的波长分布特性进行研究（图 4），可见 CARS 相较于 SPA、Boruta 其所筛选波长在可见光波段主要分布于一端处，忽略了 400~900nm 主要波段内峰谷周围的重要信息，SPA 相较于 CARS、Boruta 其所筛选波长分布相对疏散，Boruta 相较于 CARS、SPA 其所筛选波长分布集中且更加趋向于陡峭位置。

表 1 各光谱波段占比特征波长数量

波段	波段范围	各特征筛选方法选择的波长数目		
		CARS	SPA	Boruta
紫外光	200~360nm	8	7	24
可见光	360~780nm	5	6	11
近红外光	780~1100nm	20	8	12

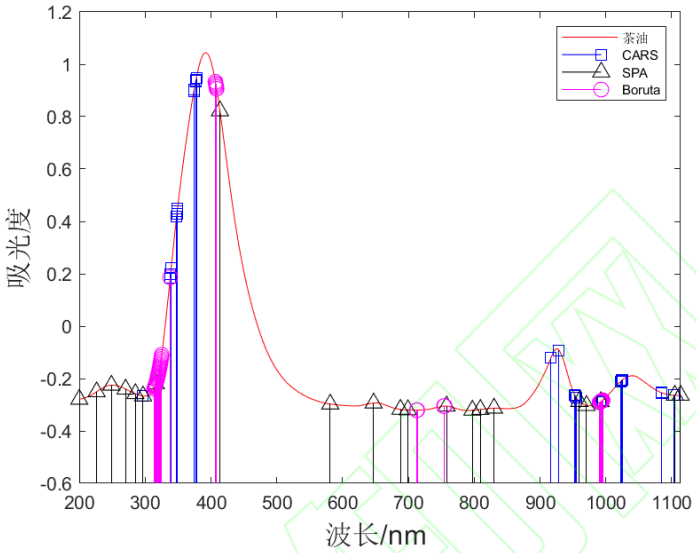

图 4 基于 CARS、SPA、Boruta 波长分布特性

2.4 模型预测与评估

2.4.1 XGBoost 模型的建模结果

XGBoost 模型利用参数 η 调节学习过程中的权重缩减进而提高模型的鲁棒性，本研究通过十折交叉验证以 0.05 为步长对参数 η 进行循环优选，将最高准确率下的 η 作为 XGBoost 模型建立的依据。表 2 给出了不同特征波长选择算法对应的最优 η 值。在此基础上，采用最优模型参数建立的 XGBoost 模型评价指标如表 3 所示。

Boruta-XGBoost 模型的预测性能最佳，其鉴伪准确率、灵敏度和特异性均高于 CARS-XGBoost 和 SPA-XGBoost 模型，而采用全光谱建立的 XGBoost 模型对应的各项评价指标均为最低（表 3）。下面以测试集评价结果为例进行详细说明。首先，Boruta-XGBoost 模型的鉴伪准确率达到了 98.18%，而 CARS-XGBoost 和 SPA-XGBoost 模型均为 96.36%，全光谱 XGBoost 模型的准确率（89.09%）最低。然后 Boruta-XGBoost 可以将鉴伪灵敏度从 CARS-XGBoost 和 SPA-XGBoost 的 92.31%提升到了 100%，并显著高于全光谱 XGBoost 模型（84.62%）。最后，在鉴伪特异性上 Boruta-XGBoost、CARS-XGBoost 和 SPA-XGBoost 模型表现相当，三者的特异性均为 97.62%，但与全光谱 XGBoost 模型（90.48%）相比得到了明显提升。上述结果表明，Boruta 算法与 CARS 和 SPA 相比可以有效提升 XGBoost 模型对掺伪茶油的鉴别能力，体现了 Boruta 使用特征波长与随机影子变量在多次迭代下相互竞争的优势。

此外进一步探讨了不同掺伪比例下各个 XGBoost 模型的鉴别性能，如表 4 所示。由于表 4 仅考虑了掺伪茶油的样品，因此以特异性指标（式 5）说明鉴别精度。可见当掺伪比例在 3%及以上时，CARS-XGBoost、SPA-XGBoost、Boruta-XGBoost 的特异性均达到了 100%；但当掺伪比例 1%时上述三种模型的特异性下降至 83.33%，而全光谱 XGBoost 模型的特异性仅为 50%。

表 2 SVM、ELM、XGBoost 模型参数选择

特征波长选择方法	建模方法			
	XGBoost	SVM		ELM
	η	c	γ	n
FS(全光谱)	0.2	147.0334	0.0039	82
CARS	0.15	256	0.1895	38
SPA	0.1	27.8576	1.7411	34
Boruta	0.2	16	0.1895	50

表 3 XGBoost 茶油快速鉴伪模型评价指标

建模方法	特征筛选方法	训练集			测试集		
		灵敏度	特异性	准确率	灵敏度	特异性	准确率
XGBoost	FS	91.49%	94.17%	93.41%	84.62%	90.48%	89.09%
	CARS	93.62%	98.33%	97.01%	92.31%	97.62%	96.36%
	SPA	95.74%	97.50%	97.01%	92.31%	97.62%	96.36%
	Boruta	95.74%	99.17%	98.20%	100%	97.62%	98.18%

表 4 不同掺伪比例下 XGBoost 茶油快速鉴伪模型测试集特异性

掺伪比例	特征选择方法			
	FS	CARS	SPA	Boruta
1%	50%	83.33%	83.33%	83.33%
3%	80%	100%	100%	100%
5%	100%	100%	100%	100%
7%	100%	100%	100%	100%
9%	100%	100%	100%	100%
12%	100%	100%	100%	100%
15%	100%	100%	100%	100%
20%	100%	100%	100%	100%
35%	100%	100%	100%	100%

2.4.2 模型对比

为进一步阐明 XGBoost 模型的鉴别能力，对比了两种传统的模型，即 SVM 和 ELM。SVM 模型通过十折交叉验证和网格搜索法确定最优的惩罚因子 c 与核函数γ，将最高准确率下 c 和γ作为 SVM 模型建立的依据。ELM 模型中由于隐含层神经元个数要小于训练集样本数[24]，因此在小于训练集样本数范围内以 2 为步长寻优隐含层神经元个数 n，将最高准确率下的 n 作为 ELM 模型建立的依据。SVM 和 ELM 在不同特征波长选择算法下对应的最优模型参数如表 2 所示。在此基础上，以最优模型参数分别建立的 SVM 和 ELM 模型评价指标如表 5 所示，可见通过 Boruta 算法筛选特征波长建立的 SVM 和 ELM 模型的各评价指标均优于 CARS、SPA 以及全波长建立的模型，这也再次显示了 Boruta 算法的优势。因此下面将仅以 Boruta 算法为例对比 XGBoost、SVM 和 ELM 模型的鉴别能力。

对比表 3 和表 5 得到，XGBoost 模型的准确率和特异性最高、ELM 模型次之、SVM 模型最差，而三种模型的灵敏度均达到了 100%。具体而言，XGBoost 模型的准确率与 ELM 和 SVM 模型相比分别提高了 1.82%和 3.63%，特异性分别提高了 2.38%和 4.76%。上述结果表明，相比于 SVM 和 ELM 模型，基于梯度提升原理的 XGBoost 模型通过不断拟合前一棵树的残差来弥补真值与预测值的误差范围，从而有效提升了掺伪茶油的鉴别精度。

表 5 SVM、ELM 茶油快速鉴伪模型评价指标

建模方法	特征筛选方法	测试集		
		灵敏度	特异性	准确率
SVM	FS	92.31%	90.48%	90.91%
	CARS	100%	90.48%	92.73%
	SPA	100%	90.48%	92.73%
	Boruta	100%	92.86%	94.55%
ELM	FS	100%	90.48%	92.73%
	CARS	100%	90.48%	92.73%
	SPA	100%	92.86%	94.55%
	Boruta	100%	95.24%	96.36%

3 结论

本文研究了采用紫外-可见-近红外光谱建立掺伪茶油鉴别模型的方法。首先，对比了 MSC、SNV、MA、SG、 1^{st} Deriv、 2^{nd} Derive、SG- 1^{st} Deriv、SG- 2^{nd} Derive、SG-CWT（L1-L10）等方法对全光谱的预处理结果，得到 SG-CWT(L5)算法的预处理效果最佳。然后，通过 CARS、SPA、Boruta 算法对预处理后的全光谱进行特征波长筛选，得到了不同波段中特征波长的分布特性；进一步对比得到，Boruta-XGBoost 模型表现出最佳的鉴别能力，鉴伪准确率、特异性和灵敏度分别达到了 98.18%、97.62%和 100%。最后，通过将 XGBoost 模型与常用的 SVM 和 ELM 模型进行对比，进一步验证了 XGBoost 可以有效提高掺伪茶油的鉴别能力；XGBoost 模型的准确率与 SVM 和 ELM 模型相比分别提高了 3.63%和 1.82%，而特异性分别提高了 4.76%和 2.38%。本研究为采用紫外-可见-近红外光谱进行茶油鉴伪提供了可行性，为开发低成本的掺伪茶油光谱鉴别装置奠定了基础。

参考文献

[1] 钟丹, 蒋孟良, 王霆. 茶油的化学成分、药理作用及临床应用研究进展[J]. 中南药学, 2012, 10(4): 299-303

ZHONG D, JIANG M L, WANG T. Research progress on chemical constituents、pharmacological effects and clinical application of camellia oil[J]. Central South Pharmacy, 2012, 10(4): 299-303

[2] 李丽, 吴雪辉, 寇巧花. 茶油的研究现状及应用前景[J]. 中国油脂, 2010, 35(3): 10-14

LI L, WU X H, KOU Q H. Research advance and application prospect of camellia seed oil[J]. China Oils and Fats, 2010, 35(3): 10-14

[3] 郭文川, 朱德宽, 张乾, 等. 基于近红外光谱的掺伪油茶籽油检测[J]. 农业机械学报, 2020, 51(9): 350-357

GUO W C, ZHU D K, ZHANG Q, et al. Detection on adulterated oil-tea camellia seed oil based on near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 350-357

[4] 肖义坡, 邓丹雯, 罗家星, 等. 茶叶籽油中角鲨烯的定性与定量分析[J]. 中国粮油学报, 2016, 31(4): 108-112

XIAO Y P, DENG D W, LUO J X, et al. Qualitative and quantitative analysis of squalene in tea seed oil[J]. Journal of the Chinese Cereals and Oils Association, 2016, 31(4): 108-112

[5] EMILIO S M, ALBERTO A, M P J, et al. Solvent-based strategy improves the direct determination of key parameters in edible fats and oils by 1H NMR[J]. Journal of the science of

food and agriculture, 2020,100(4): 1726-1734
[6] GILA M, M D, SANMARTIN, et al. Classification of olive fruits and oils based on their fatty acid ethyl esters content using electronic nose technology[J]. Journal of Food Measurement and Characterization, 2021,15(6):5427-5438
[7] 莫欣欣, 周莹, 孙通, 等. 可见/近红外光谱的油茶籽油三元体系掺假检测模型优化[J]. 光谱学与光谱分析, 2016, 36(12): 3881-3884
MO X X, ZHOU Y, SUN T, et al. Model optimization of ternary system adulteration detection in camellia oil based on visible/near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2016, 36(12): 3881-3884
[8] HUAN Z, XIAOYUN H, LIMEI L, et al. Near infrared spectroscopy combined with chemometrics for quantitative analysis of corn oil in edible blend oil[J]. Spectrochimicaacta. Part A, Molecular and biomolecular spectroscopy, 2022, 270:120841
[9] XINYAN W, XIHUI B, EN L, et al. Weighted multiscale support vector regression for fast quantification of vegetable oils in edible blend oil by ultraviolet-visible spectroscopy[J]. Food Chemistry, 2020, 342:12825
[10] 韩建勋, 孙瑞雪, 陈颖, 等. 傅里叶变换红外光谱结合化学计量学用于山茶油中掺杂大豆油的鉴别[J]. 食品与发酵工业, 2019, 45(18): 222-227
HAN J X, SUN R X, CHENG Y, et al. Discrimination of soya bean oil in adulterated camellia oil by FTIR spectroscopy combined with chemometrics[J]. Food and Fermentation Industries, 2019, 45(18): 222-227
[11] 荣菡, 罗懿, 黄镘淳. 近红外光谱技术法快速鉴别茶油掺伪[J]. 安徽农业科学, 2019, 47(19): 204-206
RONG H, LUO Y, HUANG M C. Study on fast identification of camellia oil adulteration based on near infrared spectroscopy[J]. Journal of Anhui Agricultural Sciences, 2019, 47(19): 204-206
[12] YALI W, YANKUN P, XIN Q, et al. Discriminant analysis and comparison of corn seed vigor based on multiband spectrum[J]. Computers and Electronics in Agriculture, 2021, 190:106444
[13] LI H, LIANG Y, XU Q, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica Chimica Acta, 2009, 648(1):77-84
[14] ARAÚJO M C U, SALDANHA T C B, GALVÃO R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2):65-73
[15] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38
FANG K N, WU J B, ZHU J P, et al. A review of technologies on random forests[J]. Journal of Statistics and Information, 2011, 26(3): 32-38
[16] KURSA M B, RUDNICKI W R. Feature selection with the Boruta package[J]. Journal of Statistical Software, 2010, 36(11): 1-13
[17] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 785-794
[18] 王琦琪, 戴家佳, 崔熊卫. 基于集成学习模型的糖尿病患病风险预测研究[J]. 软件导刊, 2022, 21(4): 62-66

WANG Q Q, DAI J J, CUI X W. Research on prediction of diabetes risk based on ensemble learning model[J]. Software Guide, 2022, 21(4): 62-66

[19] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300

[20] 关婷予, 黄咏梅, 林敏, 等. 大米蛋白粉多组分含量近红外光谱快速检测[J]. 中国粮油学报, 2021, 36(6): 136-142

GUAN T Y, HUANG Y M, LIN M, et al. Determination of multi-component constituents of rice protein powder rapidly by near infrared spectroscopy[J]. Journal of the Chinese Cereals and Oils Association, 2021, 36(6): 136-142

[21] 杰尔·沃克曼, 罗伊斯·文依. 近红外光谱解析实用指南[J]. 分析化学, 2011, 39(4): 551

WORKMAN J, WEYER L. Practical guide to interpretive near-infrared spectroscopy [J]. Chinese Journal of Analytical Chemistry, 2011, 39(4): 551

[22] ZHANG Y, GUO W. Moisture content detection of maize seed based on visible/near-infrared and near-infrared hyperspectral imaging technology[J]. International Journal of Food Science & Technology, 2020, 55(2): 631-640

[23] 丁俭, 齐宝坤, 王立敏, 等. 5 种不同植物油脂氧化程度与脂肪酸比例变化的相关性研究[J]. 中国粮油学报, 2017, 32(8): 84-91

DING J, QI B K, WANG L M, et al. Correlation of the degree of five kinds of different vegetable oil oxidation to proportions change of fatty acid[J]. Journal of the Chinese Cereals and Oils Association, 2017, 32(8): 84-91

[24] 程介虹, 陈争光. 基于高光谱数据的乳香产地快速鉴别[J]. 黑龙江八一农垦大学学报, 2021, 33(4): 93-98.

CHENG J H, CHEN Z G. Rapid identification of frankincense origin based on hyperspectral data[J]. Journal of Heilongjiang Bayi Agricultural University, 2021, 33(4): 93-98.