

# 红外光谱快速测定油茶籽油脂肪酸组成的模型建立

吴雪辉<sup>1,2</sup>, 何俊华<sup>1</sup>, 王泽富<sup>1</sup>

(1. 华南农业大学 食品学院, 广州 510642; 2. 广东省油茶工程技术研究中心, 广州 510642)

**摘要:**通过气相色谱和傅里叶红外光谱仪测定 86 个油茶籽油样本的脂肪酸组成和红外光谱图, 采用支持向量机(SVM)和 BP 人工神经网络(ANN)的非线性建模方法, 构建油茶籽油中主要脂肪酸的定量回归模型。结果表明:ANN 建立的油酸和棕榈酸定量回归模型精确度比 SVM 高, 校正集的相关系数( $R$ )分别为 0.998 7 和 0.945 1, 预测集的相关系数分别为 0.955 7 和 0.926 2, 相对标准偏差分别小于 1% 和 5%; SVM 和 ANN 建立的亚油酸定量分析模型精确度都非常高, 相对标准偏差均小于 1%。说明红外光谱用于油茶籽油中主要脂肪酸的快速检测是完全可行的。

**关键词:**油茶籽油; 脂肪酸组成; 红外光谱; 支持向量机; 人工神经网络; 模型

中图分类号: O657.33; TS227 文献标识码: A 文章编号: 1003-7969(2022)02-0124-05

## Establishment of rapid detection model of fatty acid composition of oil – tea camellia seed oil by infrared spectroscopy

WU Xuehui<sup>1,2</sup>, HE Junhua<sup>1</sup>, WANG Zefu<sup>1</sup>

(1. College of Food Science, South China Agricultural University, Guangzhou 510642, China;

2. Guangdong Engineering Research Center for Oil – Tea Camellia, Guangzhou 510642, China)

**Abstract:** The fatty acid composition and infrared spectra of 86 kinds of oil – tea camellia seed oil samples were determined by gas chromatography and Fourier transform infrared spectrometer, and the nonlinear modeling methods of support vector machine (SVM) and BP artificial neural network (ANN) were used to construct the quantitative regression model of main fatty acids in oil – tea camellia seed oil. The results showed that the quantitative regression models of oleic acid and palmitic acid established by ANN were more accurate than those by SVM, the correlation coefficients( $R$ ) of the correction set and the prediction set were 0.998 7, 0.945 1 and 0.955 7, 0.926 2, respectively, and the relative standard deviations were less than 1% and 5%, respectively. The accuracies of linoleic acid quantitative analysis models established by SVM and ANN were both very high, and the relative standard deviation was both less than 1%. It showed that infrared spectroscopy was feasible for the rapid detection of main fatty acids in oil – tea camellia seed oil.

**Key words:** oil – tea camellia seed oil; fatty acid composition; infrared spectroscopy; support vector machine; artificial neural network; model

油茶籽油又称为茶油、山茶油、茶籽油、野山茶油等, 是从山茶科植物油茶(*Camellia oleifera* Abel.) 种子中提取的木本食用油。油茶籽油的脂肪酸组成

与橄榄油相似, 其中的单不饱和脂肪酸油酸含量高达 80% 以上<sup>[1-3]</sup>。油茶籽油中还含有多种功能活性成分, 如角鲨烯、植物甾醇、多酚、脂溶性维生素等, 具有很高的营养价值和保健功效<sup>[4-8]</sup>。

脂肪酸作为油茶籽油的主要成分, 其组成、含量是评价油茶籽油品质的重要指标, 也是决定油茶籽油高营养价值的主要因素。因此, 快速、准确地测定油茶籽油的脂肪酸组成十分重要。目前, 检测油脂

收稿日期: 2021-03-04; 修回日期: 2021-08-23

基金项目: 广东省林业科技计划项目(2019-02); 广东省林业科技创新项目(2017KJCX005)

作者简介: 吴雪辉(1965), 女, 教授, 博士, 研究方向为粮油食品加工(E-mail)xuehu@scau.edu.cn。

中脂肪酸组成主要采用气相色谱法,样品前处理过程烦琐,时间长,成本高,不能满足快速分析的需要<sup>[9]</sup>。红外光谱分析技术是近年来迅速发展起来的无损检测技术,被广泛应用于油脂品质及掺伪检测<sup>[10-13]</sup>。吴雪辉等<sup>[14]</sup>利用偏最小二乘法(PLS)建立了油茶籽油中甾醇、维生素E和类胡萝卜素含量的预测模型,应用于油茶籽油中功能活性成分含量的快速检测。He等<sup>[15]</sup>基于分离富集得到的植物油不皂化物红外光谱图,采用偏最小二乘判别分析法(PLS-DA)和正交偏最小二乘判别分析法(OPLS-DA)构建了芝麻油鉴定模型。陈洪亮等<sup>[16]</sup>应用近红外光谱(NIR)分析技术,结合联合间隔偏最小二乘法(SiPLS)和带极值扰动的简化粒子群优化算法(tsPSO)建立了芝麻油中大豆油掺伪含量预测模型。Du等<sup>[17]</sup>采用近红外光谱结合偏最小二乘法,通过优化建模的预处理方法,建立了山茶油、菜籽油和葵花籽油多元掺假模型,模型的预测效果良好,其决定系数大于0.995,校正标准偏差(RMSEC)和预测标准偏差(RMSEP)分别小于6.79和4.98。何小三等<sup>[18]</sup>采用NIRFlexN-500型近红外光谱仪采集油茶籽油的红外光谱图谱,通过PLS建立了脂肪酸组成模型,除十七烷酸和 $\alpha$ -亚麻酸有个别值相对偏差大于10%,其他脂肪酸相对偏差均小于10%。但是这些研究主要集中在采用近红外光谱技术结合线性偏最小二乘法建立食用油相关指标的快速检测模型,而应用红外光谱建立油脂中主要成分脂肪酸组成检测的模型研究较少,且大多采用线性的PLS建模方法,效果不理想。因此,本研究采用傅里叶红外光谱结合非线性的支持向量机(SVM)和BP神经网络(ANN)建立油茶籽油中主要脂肪酸的定量回归预测模型,旨在寻找一种快速、有效、无损检测油茶籽油中脂肪酸组成的方法。

## 1 材料与方法

### 1.1 试验材料

86个油茶籽油样本采集于广东省多家油茶籽油生产企业,样本包含有不同原料品种、不同提油方法、不同精炼工艺条件得到的油茶籽油。

Nicolet iS 10 傅里叶变换红外光谱仪,赛默飞世尔科技有限公司;Agilent 7890B 气相色谱仪,安捷伦科技(中国)有限公司。

### 1.2 试验方法

#### 1.2.1 脂肪酸组成的测定

脂肪酸甲酯化的方法参照GB 5009.168—2016稍作修改。称取50 mg 油茶籽油样本于20 mL 试管中,分别依次加入正己烷1 mL、浓度为0.5 mol/L 氢

氧化钾-甲醇溶液1 mL,振摇1 min 之后再加入水5 mL,继续振摇1 min 左右,等待溶液分层,取上清液用于色谱分析。

脂肪酸组成测定采用气相色谱法。检测条件:色谱柱采用DB-23 毛细管柱(60 m × 0.25 mm × 0.25  $\mu$ m);进样口温度250  $^{\circ}$ C;进样量1.0  $\mu$ L;载气( $N_2$ )流速2 mL/min;燃气( $H_2$ )流速30 mL/min;分流比100:1;检测器温度250  $^{\circ}$ C;升温程序为色谱柱起始温度130  $^{\circ}$ C,然后以10  $^{\circ}$ C/min 速度上升至180  $^{\circ}$ C,保持10 min,然后以15  $^{\circ}$ C/min 速度上升至215  $^{\circ}$ C,保持5 min,最后以5  $^{\circ}$ C/min 速度上升到230  $^{\circ}$ C,保持5 min。

#### 1.2.2 红外光谱采集及光谱数据预处理

参照何小三等<sup>[18]</sup>的方法稍作修改,红外光谱采集范围400~4 000  $cm^{-1}$ ,分辨率4  $cm^{-1}$ ,扫描次数64次,每个样品平行采集3次,以平均光谱作为样品光谱。

由于采集的红外光谱原始数据不仅包含了样本的化学信息,还包含了许多外界干扰信息,因此有必要采用合理的预处理方法消除干扰因素,以提高模型的准确性。经前期研究优化出油酸、棕榈酸和亚油酸的最佳预处理方法分别为Savitzky-Golay平滑(SG)、标准正态变换(SNV)和二阶导数(SD)<sup>[19]</sup>。

#### 1.2.3 模型的建立与评价

分别采用支持向量机(SVM)和BP神经网络(ANN)建模,内部验证采用留一法交叉验证。

#### 1.2.4 数据处理

采用MATLAB 2016b、Unscrambler X10.1 和Microsoft Excel 2016 等分析软件进行数据处理。

## 2 结果与分析

### 2.1 油茶籽油脂肪酸组成及样本分配

采用气相色谱法测定了86个油茶籽油样本的脂肪酸组成,由测定结果可知,油茶籽油中主要脂肪酸为棕榈酸(C16:0)6.92%~11.33%、硬脂酸(C18:0)1.51%~3.34%、油酸(C18:1)71.72%~83.42%、亚油酸(C18:2)0.83%~13.20%、亚麻酸(C18:3)0%~0.67%和花生一烯酸(C20:1)0.25%~0.64%等6种,其中棕榈酸、油酸和亚油酸含量较高,平均值分别为9.16%、79.42%、7.88%,合计占油茶籽油脂肪酸总量的97%左右,因此后面的模型建立选择这3种脂肪酸作为研究对象。同时将86个油茶籽油样本按照3:1左右的比例,随机划分为校正集和预测集,校正集样本66个,预测集样本20个。油茶籽油划分样本主要的脂肪酸含量见表1。

表1 油茶籽油划分样本主要的脂肪酸含量 %

脂肪酸	参数	总样本	校正集样本	预测集样本
油酸	最小值	71.72	71.72	75.49
	最大值	83.42	83.42	81.58
	均值	79.42	79.56	78.96
	标准差	1.82	1.89	1.51
亚油酸	最小值	0.83	0.83	4.27
	最大值	13.20	13.20	11.45
	均值	7.88	7.73	8.47
	标准差	1.96	1.99	1.72
棕榈酸	最小值	6.92	6.92	8.56
	最大值	11.33	11.33	10.17
	均值	9.16	9.17	9.15
	标准差	0.67	0.70	0.51

## 2.2 油茶籽油红外光谱分析

86个油茶籽油样本的红外光谱图如图1所示。

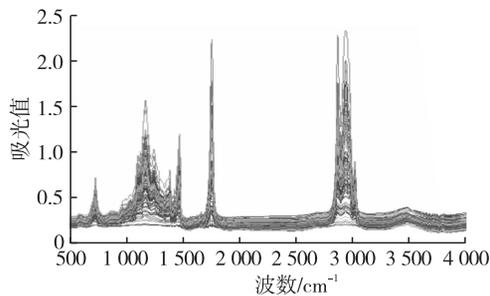


图1 86个油茶籽油样本的红外光谱图

由图1可知,油茶籽油在 $650 \sim 1\ 010\ \text{cm}^{-1}$ 、 $1\ 290 \sim 1\ 430\ \text{cm}^{-1}$ 、 $1\ 630 \sim 1\ 695\ \text{cm}^{-1}$ 、 $2\ 600 \sim 3\ 100\ \text{cm}^{-1}$ 有稳定的吸收带。其中: $650 \sim 1\ 010\ \text{cm}^{-1}$ 和 $1\ 290 \sim 1\ 430\ \text{cm}^{-1}$ 分别为C—H的面外和面内变曲振动; $1\ 630 \sim 1\ 695\ \text{cm}^{-1}$ 为不饱和脂肪酸中C=C的伸缩振动,双键数目对谱带的频率有一定影响;脂肪酸中O=C—OH的特征吸收在 $2\ 500 \sim 3\ 400\ \text{cm}^{-1}$ 。因此,红外光谱图的差异与油茶籽油中主要脂肪酸的含量有一定的关联。

## 2.3 油茶籽油中主要脂肪酸SVM模型的建立与预测

由于油茶籽油中物质含量与红外光谱吸收是非线性的,导致采用线性的最小二乘法回归建模方法存在一定的局限性,其相对标准偏差虽然小于10%,但是没有达到小于5%的水平<sup>[19]</sup>。因此,本研究选用常用的非线性建模方法——支持向量机进行建模。为了提高模型的预测效果,很好地提取3种脂肪酸的光谱信息,首先通过前期研究筛选出油酸、棕榈酸和亚油酸各自的最优子区间,其光谱波段分别为:油酸, $763 \sim 1\ 125\ \text{cm}^{-1}$ 、 $2\ 575 \sim 2\ 934\ \text{cm}^{-1}$ 和 $2\ 995 \sim 3\ 644\ \text{cm}^{-1}$ ;棕榈酸, $400 \sim 763\ \text{cm}^{-1}$ 、 $763 \sim 1\ 125\ \text{cm}^{-1}$ 和 $1\ 847 \sim 2\ 205\ \text{cm}^{-1}$ ;亚油

酸, $763 \sim 1\ 125\ \text{cm}^{-1}$ 、 $1\ 125 \sim 1\ 488\ \text{cm}^{-1}$ 和 $1\ 850 \sim 2\ 209\ \text{cm}^{-1}$ 。

支持向量机的参数优化主要包括惩罚因子( $C$ )、径向基核函数和松弛系数( $g$ )。本研究基于网格全局法寻找最优 $C$ 和 $g$ ,网格全局寻优算法是将参数的寻优范围划分为网格形式并遍历网格内的所有参数点去搜寻最优值,采用网格全局寻优算法得到油酸、棕榈酸、亚油酸的惩罚因子( $C$ )分别为0.25、0.1、0.25,松弛系数( $g$ )分别为0.6、0.35、1,在此条件下建立SVM模型,得到油酸、棕榈酸、亚油酸的实测值与预测值的校正集、预测集样本散点分布情况,结果分别如图2、图3、图4所示。

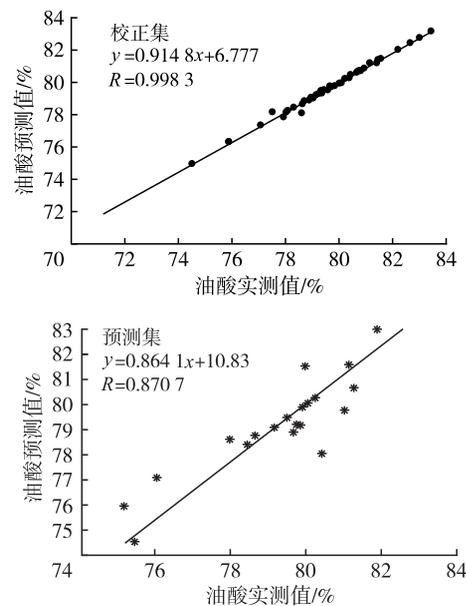


图2 油酸实测值与预测值的校正集、预测集样本散点分布图

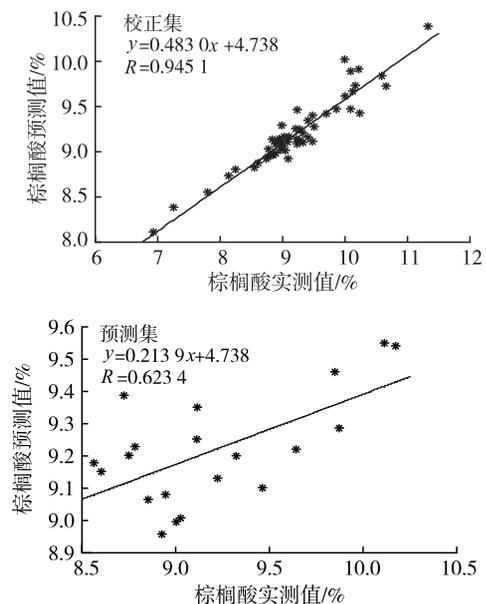


图3 棕榈酸实测值与预测值的校正集、预测集样本散点分布图

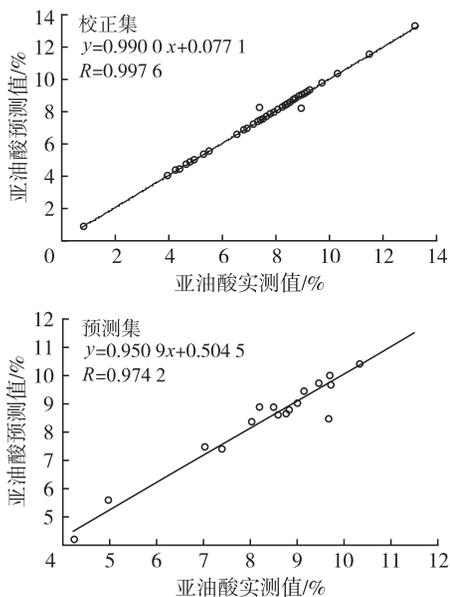


图4 亚油酸实测值与预测值的校正集、预测集样本散点分布图

图2~图4显示:油酸、棕榈酸、亚油酸校正集中所得到预测值与实测值的相关系数( $R$ )分别为0.9983、0.9451、0.9976,预测集中所得到预测值与实测值的相关系数分别为0.8707、0.6234、0.9742,模型的相对标准偏差均小于5%(其中亚油酸的小于1%),预测效果达到应用检测水平;但是棕榈酸的预测集相关系数只有0.6234,表明建立的棕榈酸定量回归模型泛化能力差,出现过拟合情况,可能是选择几个子区间联合作为建模波段,波段点数较多,造成数据量过大,建模过程中支持向量机个数偏多,导致每一个样本出现一个临界域,这样建模集精确度很高,但泛化能力一般会很低。

#### 2.4 油茶籽油中主要脂肪酸 ANN 模型的建立与预测

虽然非线性 SVM 建模效果较好,但没有交叉验证,且3个子区间输入的变量数为282,而建模样本数只有66个,远小于建模输入的变量,很容易导致过度拟合,造成模型的泛化能力差。因此,进一步采用具有内部交叉验证的非线性方法 BP 神经网络建模。

构建 BP 神经网络模型,参数选择:网络结构选择一个3层的人工神经网络,输入分别为优化后的联合子区间,输出分别为油酸、棕榈酸、亚油酸含量,同时需对输入、输出层参数进行归一化处理。对隐含层神经元数、传递函数、训练函数、学习函数、网络性能函数、仿真函数、训练校验次数、学习速率等参数进行优化,将数据输入设定好的网络模型中,对模型进行训练。应用建立 BP 的人工神经网络模

型对校正集和预测集样本进行仿真应用,得到油酸、棕榈酸、亚油酸的实测值与预测值的校正集、预测集样本散点分布情况,结果分别如图5、图6、图7所示。

从图5~图7可以看出:3种脂肪酸的校正集和预测集样本基本均匀地分布在回归线上,油酸、棕榈酸、亚油酸的校正集中所得到预测值与实测值的相关系数( $R$ )分别为0.9987、0.9451、0.9957,预测集中所得到预测值与实测值的相关系数分别为0.9557、0.9262、0.9816,表明模型预测值与实测值非常接近,模型的预测效果较好;3种脂肪酸模型的相对标准偏差分别小于1%、5%和1%。

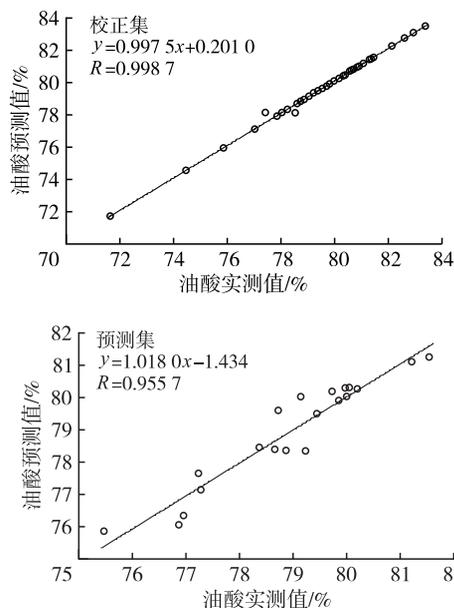


图5 油酸实测值与预测值的校正集、预测集样本散点分布图

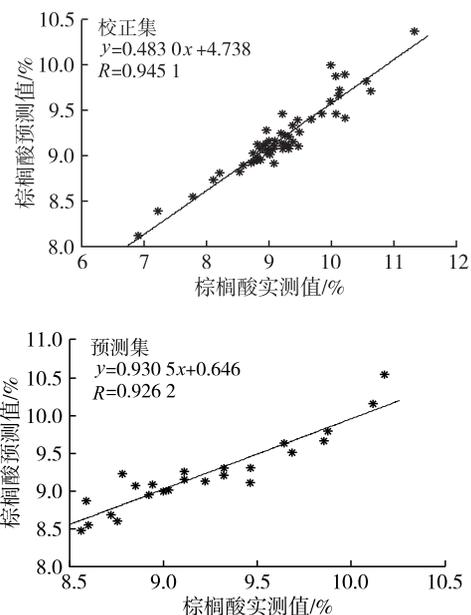


图6 棕榈酸实测值与预测值的校正集、预测集样本散点分布图

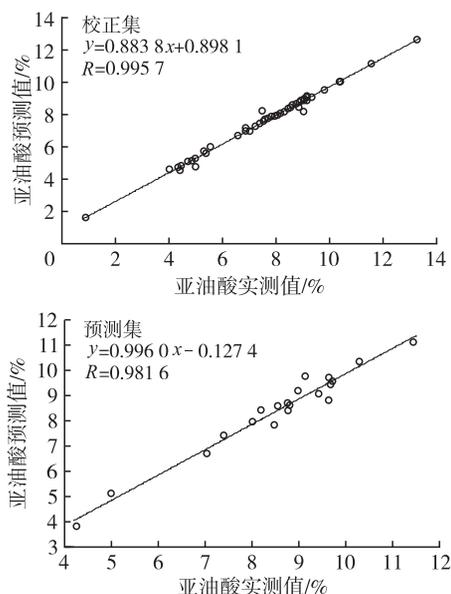


图7 亚油酸实测值与预测值的校正集、预测集样本散点分布图

### 3 结论

通过对从企业收集的86个油茶籽油样本进行脂肪酸组成测定和红外光谱扫描,利用非线性回归方法SVM、ANN分别建立油茶籽油中油酸、棕榈酸和亚油酸的定量回归模型。对油酸和棕榈酸而言,ANN建立的定量回归模型精确度比SVM高,油酸的ANN模型校正集和预测集相关系数分别为0.9987和0.9557,相对标准偏差小于1%,棕榈酸的校正集和预测集相关系数分别为0.9451和0.9262,相对标准偏差小于5%;亚油酸的SVM和ANN定量分析模型相对标准偏差均小于1%,校正集和预测集相关系数分别为0.9976、0.9957和0.9742、0.9816。说明红外光谱替代传统的气相色谱法快速测定油茶籽油中脂肪酸组成是完全可行的。

### 参考文献:

[1] SHI T, WU G, JIN Q, et al. Camellia oil authentication: a comparative analysis and recent analytical techniques developed for its assessment. A review[J]. Trends Food Sci Tech, 2020,97:88-99.

[2] LIANG H, HAO B, CHEN G, et al. Camellia as an oilseed crop[J]. Hortscience, 2017,52(4):488-497.

[3] ZHU G, LIU H, XIE Y, et al. Postharvest processing and storage methods for *Camellia oleifera* seeds[J]. Food Rev Int, 2020,36(4):319-339.

[4] CHENG Y T, WU S L, HO C Y, et al. Beneficial effects of camellia oil (*Camellia oleifera* Abel.) on ketoprofen-induced gastrointestinal mucosal damage through upregulation of HO-1 and VEGF[J]. J Agric Food Chem, 2014,62(3):642-650.

[5] TU P S, TUNG Y T, LEE W T, et al. Protective effect of

camellia oil (*Camellia oleifera* Abel.) against ethanol-induced acute oxidative injury of the gastric mucosa in mice[J]. J Agric Food Chem, 2017,65(24):4932-4941.

[6] 柯城. 茶油主要营养物质组成和保健功能[J]. 现代食品, 2019(13):105-108.

[7] CHOU T, LU Y, INBARAJ B S, et al. Camellia oil and soybean-camellia oil blend enhance antioxidant activity and cardiovascular protection in hamsters[J]. Nutrition, 2018,51/52:86-94.

[8] 曹春顺. 茶油的保健功能及对短跑运动员免疫功能的影响[J]. 长春大学学报, 2020,30(6):45-48.

[9] 荣菡, 甘露菁, 王磊. 基于近红外光谱的茶油掺伪快速检测方法的研究[J]. 中国调味品, 2019,44(12):144-147.

[10] LI X, ZHANG L, ZHANG Y, et al. Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils[J]. Trends Food Sci Tech, 2020,101:172-181.

[11] GOH K M, MAULIDIANI M, RUDIYANTO R, et al. Rapid assessment of total MCPD esters in palm-based cooking oil using ATR-FTIR application and chemometric analysis[J]. Talanta, 2019,198:215-223.

[12] KARUNATHILAKA S R, FARRIS S, MOSSOBA M M. Rapid prediction of low (<1%) trans fat content in edible oils and fast food lipid extracts by infrared spectroscopy and partial least squares regression[J]. J Food Sci, 2018,83(8):2101-2108.

[13] 姚婉清, 彭梦侠, 陈梓云, 等. 山茶油多元掺假近红外模型的建立与研究[J]. 食品安全质量检测学报, 2020,11(2):493-499.

[14] 吴雪辉, 王泽富. 红外光谱快速测定油茶籽油中主要功能活性成分含量的模型建立与评价[J]. 中国油脂, 2018,43(10):140-143.

[15] HE W X, LIN Q. Efficient saponification and separation of unsaponifiables for authentication of sesame oil using FT-IR spectroscopy and chemometrics[J]. Spectrosc Spect Anal, 2021,41(1):319-326.

[16] 陈洪亮, 曾山, 王斌. 优化基于近红外光谱的联合间隔偏最小二乘法建模检测芝麻油掺伪含量[J]. 中国油脂, 2020,45(2):86-90.

[17] DU Q W, ZHU M T, SHI T, et al. Adulteration detection of corn oil, rapeseed oil and sunflower oil in camellia oil by in situ diffuse reflectance near-infrared spectroscopy and chemometrics[J/OL]. Food Control, 2021, 121: 107577[2021-03-04]. <https://doi.org/10.1016/j.foodcont.2020.107577>.

[18] 何小三, 李博, 符树根, 等. 茶油脂肪酸组分近红外模型构建研究[J]. 南方林业科学, 2018,46(5):13-19.

[19] 王泽富. 红外光谱快速检测油茶籽油品质指标研究[D]. 广州:华南农业大学, 2018.