

# 基于近红外光谱技术快速检测广式酱油发酵过程中主要理化指标的含量

黄华丹<sup>1</sup>, 邓健豪<sup>1,2</sup>, 曹庸<sup>1</sup>, 刘占<sup>2</sup>, 符姜燕<sup>2</sup>, 阚启鑫<sup>1</sup>, 温林凤<sup>1</sup>, 徐婷<sup>2</sup>, 扈圆舒<sup>2</sup>, 贺丽苹<sup>1,3\*</sup>

(1. 华南农业大学食品学院, 广东省功能食品活性物重点实验室, 广东广州 510642)(2. 广东美味鲜调味食品有限公司, 广东中山 529800)(3. 华南农业大学测试中心, 广东广州 510642)

**摘要:** 为了实现对广式酱油中理化指标含量的快速检测, 该研究应用近红外光谱技术(Near Infrared Spectroscopy, NIRS)结合5种不同光谱预处理方法以及2种特征波段筛选方法, 运用偏最小二乘回归法与支持向量回归法建立了广式酱油中氨基酸态氮、总氮、可溶性无盐固形物、总酸、铵盐、总糖、还原糖以及盐分8种理化指标的定量预测模型, 并比较建模效果, 以筛选最佳定量预测模型。结果表明, 支持向量回归法所建模型的决定系数均高于PLSR模型, 均方根误差均低于PLSR模型, 说明采用支持向量回归建模效果更优。结果表明, 经异常样品剔除、预处理、特征波段筛选后, 针对广式酱油8种理化指标的建立的支持向量回归定量预测模型效果较好, 其中, 各理化指标定量预测模型的训练集决定系数 $R^2$ 为0.991 1~0.962 1, 测试集决定系数 $R^2$ 为0.977 9~0.857 9。同时对各理化指标筛选出的最佳定量预测模型进行外部验证, 结果显示, 广式酱油各指标的预测值与化学值之间的绝对误差 $\leq 1.31$ ,  $t$ 检验结果显示两组数据的无显著性差异, 所建立的定量预测模型能够快速准确检测广式酱油中各理化指标的含量。该研究可以快速无损测定广式酱油发酵过程的8个重要理化指标, 为酱油实际工业生产过程中的质量控制奠定基础。

**关键词:** 广式酱油; 近红外光谱; 理化指标; 快速检测

文章编号: 1673-9078(2024)12-386-401

DOI: 10.13982/j.mfst.1673-9078.2024.12.0716

## Rapid Determination of Main Physicochemical Indexes of Cantonese Soy Sauce in the Fermentation Process Based on Near-infrared Spectroscopy

HUANG Huadan<sup>1</sup>, DENG Jianhao<sup>1,2</sup>, CAO Yong<sup>1</sup>, LIU Zhan<sup>2</sup>, FU Jiangyan<sup>2</sup>, KAN Qixin<sup>1</sup>,  
WEN Linfeng<sup>1</sup>, XU Ting<sup>2</sup>, HU Yuanshu<sup>2</sup>, HE Liping<sup>1,3\*</sup>

(1. College of Food Science, South China Agricultural University, Guangdong Provincial Key Laboratory of Nutraceuticals and Functional Foods, Guangzhou 510642, China)

(2. Guangdong Meiweixian Flavoring Foods Co. Ltd., Zhongshan 529800, China)

(3. Instrumental Analysis & Research Center, South China Agricultural University, Guangzhou 510642, China)

引文格式:

黄华丹, 邓健豪, 曹庸, 等. 基于近红外光谱技术快速检测广式酱油发酵过程中主要理化指标的含量[J]. 现代食品科技, 2024, 40(12): 386-401.

HUANG Huadan, DENG Jianhao, CAO Yong, et al. Rapid determination of main physicochemical indexes of Cantonese soy sauce in the fermentation process based on near-infrared spectroscopy [J]. Modern Food Science and Technology, 2024, 40(12): 386-401.

收稿日期: 2023-06-13

基金项目: 中山市科学技术局项目(CXTD2020006)

作者简介: 黄华丹(1999-), 女, 硕士研究生, 研究方向: 食品检测分析, E-mail: 2021ddd@stu.scau.edu.cn

通讯作者: 贺丽苹(1965-), 女, 博士, 正高级实验师, 研究方向: 食品质量与安全, 环境与生物分析化学、分析检测技术, E-mail: heliping@scau.edu.cn

**Abstract:** The production of soy sauce is crucial in the food industry. This study aimed to achieve the rapid detection of eight physicochemical indexes of Cantonese soy sauce, namely the content of amino acid nitrogen, total nitrogen, soluble salt-free solids, total acid, ammonium salt, total sugar, reducing sugar, and salt. To achieve this goal, first, quantitative prediction models were established for these physicochemical indexes using near-infrared spectroscopy (NIRS). Specifically, the NIRS spectra of Cantonese soy sauce were processed with five spectral preprocessing methods and two feature band screening methods, and the processed spectra were used to construct quantitative prediction models through partial least-squares regression (PLSR) and support vector regression (SVR). Next, the performances of the models were compared to screen the optimal quantitative prediction models. The results showed that compared with the PLSR-based quantitative prediction models, the SVR-based quantitative prediction models had higher coefficients of determination ( $R^2$ ) and lower root-mean-square errors, indicating that the SVR-based quantitative prediction models were superior after abnormal sample removal, preprocessing, and feature band screening. The  $R^2$  values of the training and test sets for each physicochemical index were between 0.991 1~0.962 1 and 0.977 9~0.857 9, respectively. Furthermore, the optimal quantitative prediction models screened for each physicochemical index was externally validated. The absolute error between the predicted and chemical values of each index of the Cantonese soy sauce was  $\leq 1.31$ , and the  $t$ -test results showed that there was no significant difference between the two groups of data, indicating that the quantitative prediction models can quickly and accurately detect each physicochemical index of the Cantonese soy sauce. The proposed models can quickly and nondestructively determine eight important physicochemical indexes of Cantonese soy sauce in the fermentation process, laying a foundation for quality control in the actual industrial production process of soy sauce.

**Key words:** cantonese soy sauce; near-infrared spectroscopy; physicochemical index; rapid determination

酱油是我国传统调味品，主要以大豆或面粉为原料，经复杂的生化反应形成了具有独特风味和口感的调味液<sup>[1,2]</sup>。目前酿造酱油根据工艺不同，可分为高盐稀态发酵酱油和低盐固态发酵酱油，广式酱油主要以高盐稀态发酵为主，其酱油成品酯香浓郁，色泽红润，口感鲜美，深受广大消费者青睐<sup>[3]</sup>。高盐稀态发酵酱油是在日晒夜露的环境下自然发酵，所需周期一般为3~6个月，开放的发酵方式及较长的发酵周期使得环境因素成为影响其品质的主要因素，而广东地区的气候特征是夏热冬温，季节的变化使不同酱油产品批次间的品质出现波动<sup>[4,5]</sup>。为了保证酱油产品的质量和稳定，需要对酱油生产过程中重要的理化指标进行监测。目前的理化指标检测均采用国家标准规定的传统检测方法，检测结果客观可信，但存在操作繁琐、耗时长，对技术人员要求高并且难以对酱油生产过程进行即时监控等缺点<sup>[6]</sup>。因此，快速无损检测酱油发酵过程中的理化指标对酱油发酵过程控制和成品品质监控具有重要作用。

近红外光谱分析技术(Near Infrared Spectroscopy, NIRS)作为近代迅猛发展的高新分析技术，具有操作简便、分析效率高、环保、不破坏样品等多个优点，在食品、轻工、制药、环境等各领域得到了广

泛应用<sup>[7-11]</sup>。近红外光谱产生于分子振动能级的跃迁，振动信息来自于样品中的含氢基团(如-OH、-CH、-NH等)的倍频和合频吸收，由于这些基团在食品中普遍存在，因此近红外光谱被广泛用于分析食品成分<sup>[12,13]</sup>。但样品中含氢基团种类复杂，不同基团的吸收峰之间存在重叠性，难以直接分析吸收带是由何种基团产生。因此，近红外光谱分析技术作为一种间接分析技术，通常需要结合化学计量学提取近红外光谱中的有效信息<sup>[14]</sup>。

近年来，近红外光谱分析技术在酱油中的应用也有所报道，Hu等<sup>[15]</sup>利用Vis-NIR技术结合改进PSO优化算法成功建立酱油样品中氨基酸态氮、盐、总酸和色比的定量预测模型。Qin等<sup>[16]</sup>通过结合近红外光谱技术与协同区间极限学习机技术，建立了酱油中氨基酸态氮含量无损检测的近红外模型。Chen等<sup>[17]</sup>采用可见光和近红外光谱结合k-近邻算法，建立了不同酱油品牌判别模型，可判别真伪和掺假。但目前大部分近红外研究仅针对市面上的酱油成品进行品质检测和掺假识别，对于品质检测，目前的报道多以2~4个指标为主，未对酱油中其他理化指标进行系统全面的建模研究，不能完全体现酱油品质，而且大部分相关研究并没有进一步验证所建模型的预测精度，使其在实际应用上

的效果难以保证。同时以不同发酵阶段酱油为研究对象的理化指标含量检测目前尚未有相关研究,此方面研究对酱油发酵过程的监测及控制有着重要意义。

本研究以不同发酵阶段的广式酱油样品为研究对象,测定酱油中氨基酸态氮、总氮、可溶性无盐固形物、总酸、铵盐、总糖、还原糖以及盐分 8 种理化指标的含量,同时采集其近红外光谱,结合偏最小二乘回归法(Partial Least Squares Regression, PLSR)和支持向量回归法(Support Vector Regression, SVR)分别建立各理化指标的定量预测模型,并探究不同光谱预处理方式、不同特征波长选择方法和不同建模方法对模型精度的影响,以期筛选出一种快速无损检测酱油发酵过程主要理化指标的方法,为开发便携式酱油品质近红外光谱仪提供参考依据,也为酱油实际工业生产过程中的质量监测奠定基础。

## 1 材料与方法

### 1.1 试验样本

试验样本来自广东美味鲜有限公司工业生产线中不同发酵阶段的广式酱油,酱油样本取自 2020 年 09 月~2021 年 08 月每月投料发酵共 30 组酱油发酵罐,并对其进行周期性取样,间隔 15 d 采集酱油样本,分别于 0、15、30、45、60、75、90、105 d 进行取样,一共采集了 242 组样本,并于 -20 °C 保存。近红外光谱扫描前需对样品进行简单处理,先在 3 000 r/min 转速对样品离心 10 min,取上清液,以避免酱油中颗粒物对光谱产生影响。

### 1.2 仪器设备

S450 近红外光谱仪,配有石英比色皿及 Prolab S450 数据采集软件,上海棱光技术有限公司;L530 台式低速自动平衡离心机,湖南湘仪实验室仪器开发有限公司。

### 1.3 实验方法

#### 1.3.1 样品理化指标的测定

不同发酵阶段的酱油样品中总酸、总氮、氨基酸态氮、可溶性无盐固形物、盐分含量测定参照《GB 18186-2000 酿造酱油》中相关方法<sup>[18]</sup>,铵盐含量测定参照《GB 5009.234-2016 食品中铵盐的测定》中方法<sup>[19]</sup>,还原糖含量测定参照《GB 5009.7-2016 食品中还原糖的测定》中方法<sup>[20]</sup>,总糖含量测定参

照《GB/T 15672-2009 食用菌中总糖含量的测定》中方法<sup>[21]</sup>。

#### 1.3.2 近红外光谱数据采集

将不同发酵阶段酱油样品经离心取上清液后,取 5 mL 样品于样品杯中,放置在仪器旋转台上,加盖反射附件,以漫透反射的方式获取样品光谱信息。在 900~2 500 nm 光谱范围内进行扫描,以仪器内置背景为参比,扫描次数 6 次,仪器分辨率 12 cm<sup>-1</sup>,采样间隔 4 nm。为避免样品间交叉污染,每次测量前先用清水将液体杯洗净样品杯,再装入待测样品。采用仪器自带的 Prolab S450 软件对广式酱油样品的近红外光谱进行采集,每个样品平行测 2 次,取其平均光谱作为最终分析光谱。

### 1.4 建立模型

#### 1.4.1 异常样本剔除

如果建模样品中包含异常样品,模型的准确性可能会受到影响,因此在建模前,需要剔除异常样品光谱。本实验使用主成分分析(Principal Component Analysis, PCA)结合马氏距离法剔除异常光谱<sup>[22]</sup>。

#### 1.4.2 光谱预处理方法及特征波长筛选

在光谱采集过程中,仪器状态、环境温度等会引起噪声,对模型的稳定性和准确性造成影响,所以在建模前需要对光谱进行预处理<sup>[23,24]</sup>。本研究考察了 S-G 平滑(Savitzky-Golay, S-G)、多元散射校正(Multiplicative Scatter Correction, MSC)、标准正态变量变换(Standard Normal Variate Transform, SNV)、MSC+SG、SNV+SG 5 种预处理方法对不同发酵阶段酱油的近红外光谱预处理的效果,确定最佳光谱预处理方法,再进行后续的特征波段筛选及模型建立。

由于近红外光谱数据呈现数量多,维度高和复杂性高等特点,其中不仅包含大量噪声,还使得冗余信息增加,这样建立的模型计算复杂度高,模型精度反而下降<sup>[25]</sup>。因此,需要对光谱数据进行特征波段筛选,本研究主要采用竞争性自适应重加权法(Competitive Adaptive Reweighted Sampling, CARS)和随机蛙跳算法(Random Frog, RF)两种方法进行筛选。竞争性自适应重加权法(CARS)是通过自适应重加权采样技术选择出 PLS 模型中回归系数绝对值大的波长点,去掉权重小的波长点,利用交叉验证选出 RMSECV 最低的子集,可有效寻出最优

变量组合；随机蛙跳算法（RF）是一种通过观察和模仿青蛙群体行为的优化算法，以  $N$  次迭代中变量出现概率为依据筛选特征波段，概率值越高，变量的重要性越大，将具有最高概率的变量选为特征变量<sup>[26-28]</sup>。通过比较经此两种方法特征提取后的建模效果来确定最佳的波段筛选方法，从而提高模型的精确度并降低模型的复杂度。

#### 1.4.3 样本集划分

在建立定量分析模型前，还需将样本集划分为训练集和测试集，先通过训练集的样品结合化学计量学建立模型，再通过检验测试集样品来反映定量分析模型的预测能力。本研究采用了基于  $x$ - $y$  联合距离的样本集划分方法（Sample Set Partitioning Based on Joint X-Y Distance, SPXY），即将  $X$  变量和  $Y$  变量同时考虑在内的样本划分方法，以训练集与测试集样品数量 7:3 比例对样本进行划分<sup>[29]</sup>。

#### 1.4.4 模型建立及评价

近红外光谱定量分析技术的关键是筛选出预测性能优秀的模型，进而实现对目标化学值的预测，因此，选择适合的化学计量学对近红外定量预测模型的建立十分重要。目前常用建模方法分为线性和非线性回归两类，其中线性回归方法包括多元线性回归（Multiple Linear Regression, MLR）、主成分回归（Principal Component Regression, PCR）、偏最小二乘回归（Partial Least Squares Regression, PLSR），非线性回归方法主要包括人工神经网络（Artificial Neural Networks, ANN）和支持向量回归（Support Vector Regression, SVR）等<sup>[30]</sup>。其中，偏最小二乘回归法（PLSR）集合主成分分析、典型相关分析和多元线性回归分析 3 种分析方法的优点，是一种基于因子分析的多元校准方法，可以同时考虑理化值和光谱之间的基本关系，在近红外光谱分析技术中应用十分广泛<sup>[31,32]</sup>。支持向量回归（SVR）属于支持向量机（Support Vector Machine, SVM）中的一个重要应用分支，在出现非线性问题时，SVR 通过引入一个损失函数和一组约束条件来寻找最优的分界线或曲线，从而使得预测值与实际值之间的误差最小化，其优点是可以在高维空间中进行非线性映射，并能够处理具有复杂关系的数据，同时具有较好的鲁棒性和泛化能力，但是 SVR 的计算复杂度较高，本研究采用遗传算法来对 SVR 的惩罚参数  $C$ 、损失函数  $\epsilon$ 、核系数  $\gamma$  进行调参<sup>[33,34]</sup>。由此，本研究采用偏最小二乘回归法（PLSR）构建广式

酱油中各理化指标的定量预测模型，通过 10 折交叉验证确定模型的最佳主因子数，确定最佳光谱预处理方法和特征波长筛选方法，以及采用支持向量回归法（SVR）构建广式酱油中各理化指标的定量预测模型，通过遗传算法对 SVR 的参数进行调整，然后根据训练集和测试集的决定系数（ $R^2_c$ 、 $R^2_p$ ）、训练集均方根误差和测试集均方根误差（ $RMSEC$ 、 $RMSEP$ ）对所建模型进行评价。一般来说，好的定量预测模型应具有较高  $R^2$  值，及较低  $RMSEC$  值和  $RMSEP$  值<sup>[35,36]</sup>。

#### 1.4.5 近红外模型的外部验证

获得最佳定量模型后，还需采用外部验证法检验模型的预测效果及稳定性，通过将 14 组非建模样本的光谱数据导入到筛选出的最佳定量模型，得出样品理化指标的化学预测值，与实际国标方法测定值对比，对所建立模型的准确度进行评价<sup>[37]</sup>。

### 1.5 数据处理

表格采用 Excel 2019 绘制，绘图采用 Origin 2021 软件，用 MATLAB 2021a 软件对光谱进行预处理、波段筛选及模型建立。

## 2 结果与讨论

### 2.1 广式酱油样品的近红外光谱

不同发酵阶段的广式酱油样品的近红外原始光谱图如图 1 所示，共采集了 242 组不同发酵阶段的广式酱油样品数据。由图 1 可见，在波长 1 380~1 550 nm 和 1 850~2 100 nm 范围内酱油样品有明显的特征吸收。其中，950、1 450、1 850、1 950 nm 附近为 O-H 的倍频和合频吸收，1 050、1 500、1 920 附近为 N-H 的倍频和合频吸收，1 300、1 750、2 300 附近为 C-H 的倍频和合频吸收，2 200~2 300 nm 处与碳水化合物的 C-H 和 C=O 组合波段的拉伸振动有关，这也是近红外光谱技术能够检测酱油中各理化指标含量的基础<sup>[15,38,39]</sup>。图 2 为不同发酵阶段的广式酱油样品近红外平均光谱图，可以看出 0 d 的酱油样品吸光度值与其他天数有一定区别，但在总体上不同发酵阶段的广式酱油样品的吸光度趋势一致，这与近红外光谱的特性有关，其近红外特征谱带严重重叠，通常一个特征峰并不能代表单一组分。因此，近红外光谱分析技术作为一种间接分析技术，通常需要结合化学计量学提取近红外光谱中的有效信息。

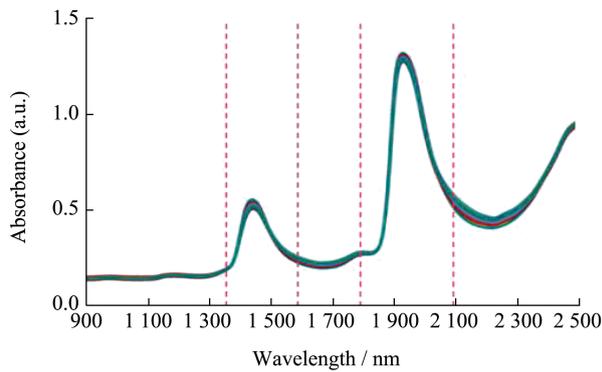


图 1 242 份不同发酵阶段的广式酱油样品原始近红外光谱图

Fig.1 NIR spectra of 242 Cantonese soy sauce samples at different fermentation stages

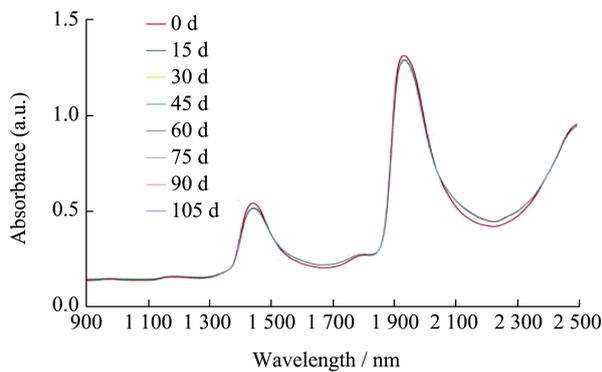


图 2 不同发酵阶段的广式酱油样品近红外平均光谱图

Fig.2 NIR average spectra of Cantonese soy sauce samples at different fermentation stages

## 2.2 样本集划分

所有样本的马氏距离结果如图 3 所示。通过主成分分析结合马氏距离法剔除异常光谱，以马氏距离阈值大于 3 的光谱为异常光谱，将红色直线以上部分是马氏距离大于 3 的样本进行剔除，一共剔除了 22 组异常数据。其中剔除的样本数据中有 15 组酱油样品为发酵 0 d，2 组酱油样品为发酵 60 d，5 组酱油样品为发酵 105 d，造成光谱数据异常的原因很多，如采集时的环境条件干扰、光谱仪本身的误差等，从图 2 可以看出发酵 0 d 的酱油样品平均光谱与其他样品有一定差异，这可能是多组 0 d 酱油样本判定为异常样本的原因。剩余 220 组样品中随机取 2 批发酵罐的样本数据，其中一批为冬季发酵的酱油样本，另一批为夏季发酵的酱油样本，共 14 组样品数据集作为外部验证集，取剩下的 206 组样品按 7:3 划分为训练集和测试集，划分结果及数据统计如表 1 所示，样本训练集的浓度分布范围均大于测试集的浓度分布范围，符合建模样本划分的要求。

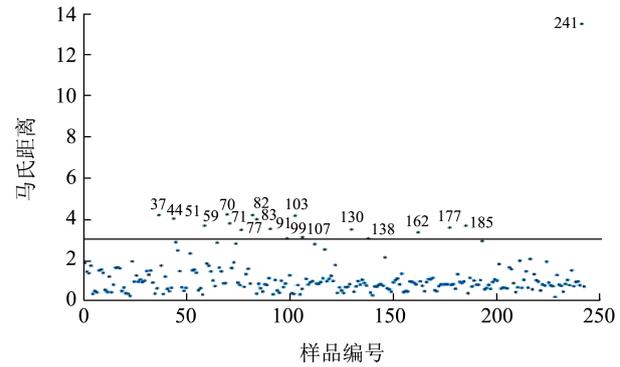


图 3 所有样本的马氏距离

Fig.3 Mahalanobis distance of all samples

表 1 样本划分统计表

Table 1 Sample division statistical table

指标		含量范围 (g/100 mL)	平均值	标准差
氨基酸态氮	训练集	0.16~1.12	0.90	0.19
	测试集	0.31~1.12	0.95	0.13
全氮	训练集	0.35~1.81	1.42	0.28
	测试集	1.28~1.66	1.49	0.10
铵盐	训练集	0.01~0.27	0.15	0.06
	测试集	0.11~0.26	0.17	0.03
总糖	训练集	2.20~8.80	6.24	1.57
	测试集	3.00~8.30	6.31	1.48
还原糖	训练集	2.00~8.30	6.04	1.52
	测试集	3.00~7.80	6.11	1.40
可溶性无盐固形物	训练集	5.60~22.08	18.29	3.17
	测试集	10.65~21.82	18.96	1.75
含盐量	训练集	15.84~20.72	17.23	0.74
	测试集	15.89~18.45	17.12	0.57
总酸	训练集	0.27~2.41	1.68	0.43
	测试集	0.66~2.41	1.86	0.27

## 2.3 光谱数据预处理

为了建立酱油各理化指标最佳定量预测模型，本研究采用 SG 平滑、MSC、SNV、MSC+SG 平滑、SNV+SG 平滑 5 种方法对酱油样品的近红外光谱数据进行预处理，以减少散射和噪声，提高模型的预测性能。然后采用偏最小二乘回归 (PLSR) 和支持向量回归 (SVR) 对不同方法处理后的光谱进行建模分析，确定最优的光谱预处理方法，表 2 为各光谱预处理方法的 PLSR 建模结果，表 3 为各光谱预处理方法的 SVR 建模结果。通常所建立模型的决定系数  $R^2$  越接近于 1，RMSEC 值越小，模

型的预测性能越好。结果显示，在采用 PLSR 建模中，氨基酸态氮、含盐量经 MSC 处理后，得到的  $R^2c$ 、 $R^2p$  为最高， $RMSEC$ 、 $RMSEP$  为最低，因此选择 MSC 为酱油氨基酸态氮、含盐量指标 PLSR 建模的预处理方法；可溶性无盐固形物经 MSC+SG 处理，所得的  $R^2c$ 、 $R^2p$  为最高， $RMSEC$ 、 $RMSEP$  为最低，因此选择 MSC+SG 为酱油可溶性无盐固形物指标 PLSR 建模的预处理方法；总酸、全氮、铵盐、总糖、还原糖无需光谱预处理也能得到较好的预测效果，无需过度处理，因此原始光谱数据可作为其 PLSR 建模的最佳预处理方法。在采用 SVR 建模中，氨基酸态氮、总酸经 SNV+SG 处理后，得到的  $R^2c$ 、 $R^2p$  为最高， $RMSEC$ 、 $RMSEP$  为最低，因此选

择 SNV+SG 为酱油氨基酸态氮、总酸指标 SVR 建模的预处理方法；全氮、可溶性无盐固形物经 MSC 处理后，得到的  $R^2c$ 、 $R^2p$  为最高， $RMSEC$ 、 $RMSEP$  为最低，因此选择 MSC 为酱油全氮、可溶性无盐固形物指标 SVR 建模的预处理方法；铵盐、总糖经 SNV 处理后，得到的  $R^2c$ 、 $R^2p$  为最高， $RMSEC$ 、 $RMSEP$  为最低，因此选择 SNV 为酱油铵盐、总糖指标 SVR 建模的预处理方法；还原糖经 MSC+SG 处理后，得到的  $R^2c$ 、 $R^2p$  为最高， $RMSEC$ 、 $RMSEP$  为最低，因此选择 MSC+SG 为酱油还原糖指标 SVR 建模的预处理方法；含盐量经 SG 处理后，得到的  $R^2c$ 、 $R^2p$  为最高， $RMSEC$ 、 $RMSEP$  为最低，因此选择 SG 为酱油含盐量指标 SVR 建模的预处理方法。

表 2 不同光谱预处理方法酱油样品的理化指标PLSR定量模型的主要评价参数

Table 2 The main evaluation parameters of PLSR quantitative model of soy sauce samples with different spectral pretreatment methods

	预处理方法	PCs	$R^2c$	$RMSEC$	$R^2p$	$RMSEP$
氨基酸态氮	NONE	11	0.989 7	0.019 9	0.914 7	0.016 8
	SG	11	0.989 5	0.020 2	0.918 4	0.016 4
	MSC	10	0.987 5	0.021 8	0.960 5	0.011 0
	SNV	10	0.988 0	0.021 4	0.946 5	0.012 2
	MSC+SG	10	0.987 4	0.021 9	0.960 4	0.011 1
	SNV+SG	10	0.987 9	0.021 5	0.946 5	0.012 2
	全氮	NONE	11	0.982 2	0.036 3	0.958 0
SG		11	0.981 9	0.036 6	0.958 7	0.020 8
MSC		10	0.981 2	0.037 1	0.946 6	0.025 0
SNV		10	0.982 3	0.036 0	0.953 4	0.022 0
MSC+SG		10	0.980 9	0.037 3	0.946 8	0.025 0
SNV+SG		10	0.982 1	0.036 2	0.953 8	0.021 9
铵盐		NONE	12	0.980 3	0.007 6	0.961 5
	SG	12	0.979 7	0.007 7	0.961 8	0.007 7
	MSC	12	0.983 7	0.007 0	0.938 5	0.009 2
	SNV	12	0.983 7	0.007 0	0.938 8	0.009 2
	MSC+SG	12	0.983 6	0.007 0	0.934 1	0.009 6
	SNV+SG	12	0.983 5	0.007 0	0.935 6	0.009 5
	总糖	NONE	13	0.973 1	0.250 8	0.963 6
SG		13	0.970 8	0.261	0.964 9	0.278 4
MSC		9	0.963 2	0.291 9	0.962 7	0.267 6
SNV		9	0.963 2	0.292 2	0.963 8	0.264 4
MSC+SG		9	0.962 5	0.294 7	0.964 1	0.263 0
SNV+SG		9	0.962 6	0.294 4	0.964 2	0.263 0

续表 2

	预处理方法	PCs	$R^2_c$	RMSEC	$R^2_p$	RMSEP
还原糖	NONE	12	0.966 8	0.272 1	0.941 4	0.358 6
	SG	12	0.965 1	0.278 9	0.942 1	0.356 3
	MSC	9	0.952 2	0.322 5	0.954 7	0.296 0
	SNV	9	0.953 0	0.319 6	0.951 0	0.307 1
	MSC+SG	9	0.951 4	0.324 7	0.954 8	0.259 3
	SNV+SG	9	0.952 3	0.321 6	0.951 2	0.306 4
可溶性无盐 固形物	NONE	9	0.970 4	0.550 2	0.892 3	0.399 3
	SG	9	0.970 2	0.551 7	0.892 3	0.399 0
	MSC	7	0.967 4	0.571 2	0.953 1	0.300 4
	SNV	8	0.968 3	0.563 5	0.943 1	0.317 3
	MSC+SG	8	0.967 1	0.573 6	0.953 7	0.298 4
	SNV+SG	8	0.968 0	0.565 9	0.943 9	0.315 1
含盐量	NONE	10	0.926 9	0.193 4	0.884 3	0.196 6
	SG	10	0.926 3	0.194 2	0.882 8	0.195 9
	MSC	9	0.927 3	0.195 5	0.892 7	0.169 7
	SNV	9	0.927 1	0.195 7	0.889 8	0.170 3
	MSC+SG	9	0.926 1	0.196 9	0.893 1	0.169 3
	SNV+SG	9	0.925 9	0.197 1	0.890 2	0.169 9
总酸	NONE	13	0.951 3	0.093 4	0.856 1	0.106 5
	SG	13	0.949 0	0.095 4	0.857 0	0.106 2
	MSC	12	0.951 2	0.093 8	0.808 3	0.103 9
	SNV	12	0.952 9	0.092 2	0.799 0	0.107 2
	MSC+SG	12	0.947 8	0.096 8	0.813 6	0.102 5
	SNV+SG	12	0.950 5	0.094 4	0.800 6	0.107 0

表 3 不同光谱预处理方法酱油样品的理化指标SVR定量模型的主要评价参数

Table 3 The main evaluation parameters of SVR quantitative model of soy sauce samples with different spectral pretreatment methods

	预处理方法	$R^2_c$	RMSEC	$R^2_p$	RMSEP
氨基酸态氮	NONE	0.986 6	0.022 6	0.904 2	0.016 4
	SG	0.989 0	0.020 7	0.932 2	0.014 1
	MSC	0.939 6	0.019 9	0.931 8	0.013 3
	SNV	0.989 5	0.020 1	0.933 2	0.013 2
	MSC+SG	0.990 0	0.019 6	0.936 3	0.012 9
	SNV+SG	0.990 1	0.019 5	0.937 4	0.012 8
全氮	NONE	0.985 0	0.033 3	0.946 8	0.022 7
	SG	0.985 6	0.032 5	0.942 8	0.023 2
	MSC	0.981 1	0.037 2	0.958 0	0.020 4
	SNV	0.981 5	0.036 6	0.954 2	0.020 8
	MSC+SG	0.981 1	0.036 9	0.953 7	0.021 0
	SNV+SG	0.981 8	0.036 4	0.955 1	0.020 8

续表 3

	预处理方法	$R^2_c$	RMSEC	$R^2_p$	RMSEP
铵盐	NONE	0.984 5	0.006 7	0.951 0	0.009 4
	SG	0.983 8	0.006 8	0.950 9	0.009 4
	MSC	0.983 4	0.007 1	0.950 5	0.008 5
	SNV	0.985 0	0.006 7	0.951 0	0.008 4
	MSC+SG	0.985 4	0.006 7	0.947 6	0.008 9
	SNV+SG	0.985 0	0.006 8	0.948 7	0.008 8
总糖	NONE	0.968 4	0.269 8	0.966 8	0.274 4
	SG	0.967 4	0.274 1	0.966 8	0.274 6
	MSC	0.964 7	0.285 5	0.966 1	0.264 4
	SNV	0.975 0	0.244 1	0.968 1	0.261 5
	MSC+SG	0.974 2	0.247 9	0.967 7	0.262 6
	SNV+SG	0.974 9	0.244 5	0.967 9	0.261 7
还原糖	NONE	0.989 1	0.156 1	0.942 5	0.342 6
	SG	0.988 8	0.157 9	0.942 4	0.342 8
	MSC	0.975 5	0.233 5	0.963 6	0.268 8
	SNV	0.968 2	0.265 2	0.956 4	0.298 8
	MSC+SG	0.975 7	0.232 9	0.964 2	0.267 4
	SNV+SG	0.969 2	0.261 4	0.956 6	0.299 0
可溶性无盐固形物	NONE	0.976 1	0.501 8	0.944 7	0.299 0
	SG	0.976 3	0.502 0	0.948 2	0.291 8
	MSC	0.979 1	0.463 2	0.962 7	0.261 9
	SNV	0.979 3	0.460 8	0.961 8	0.264 5
	MSC+SG	0.972 1	0.538 6	0.961 4	0.273 7
	SNV+SG	0.978 7	0.467 5	0.961 8	0.264 5
含盐量	NONE	0.951 5	0.158 7	0.914 5	0.167 8
	SG	0.951 8	0.158 3	0.915 4	0.167 8
	MSC	0.926 0	0.198 5	0.913 3	0.150 8
	SNV	0.925 5	0.199 0	0.913 3	0.150 7
	MSC+SG	0.925 6	0.202 0	0.913 5	0.150 6
	SNV+SG	0.925 9	0.199 2	0.914 5	0.150 2
总酸	NONE	0.943 7	0.099 7	0.784 2	0.122 9
	SG	0.945 2	0.098 2	0.775 3	0.123 7
	MSC	0.963	0.081	0.834 4	0.098 9
	SNV	0.961	0.083	0.833 4	0.099 5
	MSC+SG	0.964 6	0.079 5	0.834 6	0.099 3
	SNV+SG	0.966 1	0.077 9	0.835 2	0.099 8

表 4 不同波段筛选方法及不同建模方法下酱油样品的理化指标的主要评价参数

Table 4 The main evaluation parameters of physicochemical indexes of soy sauce samples under different band screening methods and different modeling methods

	波段筛选方法	建模方法	变量数	$R^2c$	$RMSEC$	$R^2p$	$RMSEP$
氨基酸态氮	MSC-CARS	PLSR	35	0.988 7	0.020 8	0.970 0	0.010 3
	MSC-RF	PLSR	50	0.940 5	0.103 3	0.866 1	0.106 5
	SNVSG-CARS	SVR	22	0.988 7	0.020 8	0.957 3	0.011
	SNVSG-RF	SVR	20	0.991 1	0.018 0	0.977 2	0.014 2
全氮	NONE-CARS	PLSR	32	0.981 1	0.037 4	0.963 7	0.020 5
	NONE-RF	PLSR	50	0.980 4	0.038 0	0.964 4	0.018 4
	MSC-CARS	SVR	14	0.982 8	0.035 5	0.976 4	0.015 2
	MSC-RF	SVR	30	0.982 1	0.036 5	0.960 4	0.018 0
铵盐	NONE-CARS	PLSR	42	0.976 4	0.008 5	0.966 1	0.006 3
	NONE-RF	PLSR	30	0.973 2	0.008 8	0.966 2	0.007 0
	SNV-CARS	SVR	24	0.984 6	0.006 9	0.954 4	0.008 2
	SNV-RF	SVR	30	0.979 9	0.007 9	0.953 4	0.007 9
总糖	NONE-CARS	PLSR	39	0.972 6	0.255 7	0.973 3	0.238 9
	NONE-RF	PLSR	50	0.963 3	0.295 6	0.972 7	0.237 0
	SNV-CARS	SVR	24	0.982 0	0.213 4	0.973 0	0.227 8
	SNV-RF	SVR	35	0.975 1	0.248 8	0.972 5	0.238 3
还原糖	NONE-CARS	PLSR	46	0.961 3	0.286 2	0.952 3	0.325 3
	NONE-RF	PLSR	60	0.960 1	0.298 5	0.955 0	0.301 3
	MSCSG-CARS	SVR	27	0.958 5	0.308 6	0.958 8	0.265 5
	MSCSG-RF	SVR	20	0.980 0	0.213 2	0.961 2	0.285 5
可溶性无盐固形物	MSCSG-CARS	PLSR	26	0.976 0	0.492 2	0.936 5	0.354 5
	MSCSG-RF	PLSR	40	0.966 4	0.566 2	0.964 1	0.329
	MSC-CARS	SVR	41	0.979 3	0.4462	0.977 9	0.269 4
	MSC-RF	SVR	30	0.971	0.530 3	0.971 8	0.297
含盐量	MSC-CARS	PLSR	35	0.937 2	0.181 2	0.916 7	0.162 1
	MSC-RF	PLSR	30	0.930 7	0.186 4	0.921 4	0.164 9
	SG-CARS	SVR	79	0.963 8	0.136 4	0.921 6	0.158 8
	SG-RF	SVR	50	0.951 7	0.154 8	0.923 4	0.168 5
总酸	NONE-CARS	PLSR	32	0.943 8	0.101 4	0.790 9	0.096 0
	NONE-RF	PLSR	50	0.940 5	0.103 3	0.866 1	0.106 5
	SNVSG-CARS	SVR	30	0.969 1	0.075 1	0.849 1	0.095 4
	SNVSG-RF	SVR	30	0.962 1	0.084 7	0.857 9	0.085 8

### 2.4 特征波段筛选

为了进一步提高模型的预测性能，降低模型复杂度，需要对模型进行特征波段筛选。本研究采用了CARS和RF算法进行波段筛选，以各指标最佳预处理方法结合偏最小二乘回归(PLSR)和支持向量回归(SVR)对不同特征波段筛选方法后的光谱数据

进行建模分析，再以 $R^2c$ 、 $RMSEC$ 、 $R^2p$ 、 $RMSEP$ 作为波段筛选方法评价指标。表4为不同发酵阶段广式酱油的各理化指标在不同波段筛选方法下所对应的模型参数结果。结果显示，CARS和RF这两种波段筛选方法均能极大减少近红外光谱的波长变量，降低建模运算时间并提高模型的预测性能，其中，在采用

PLSR 建模中, 广式酱油样品的氨基酸态氮、还原糖、可溶性无盐固形物、含盐量、总酸经 RF 波段筛选方法后所建模型的参数最佳, 全氮、铵盐和总糖经 CARS 波段筛选方法后所建模型的参数最佳; 在采用 SVR 建模中, 广式酱油样品的氨基酸态氮、还原糖、总酸经 RF 波段筛选方法后所建模型的参数最佳, 可溶性无盐固形物、含盐量、全氮、铵盐和总糖经 CARS 波段筛选方法后所建模型的参数最佳。

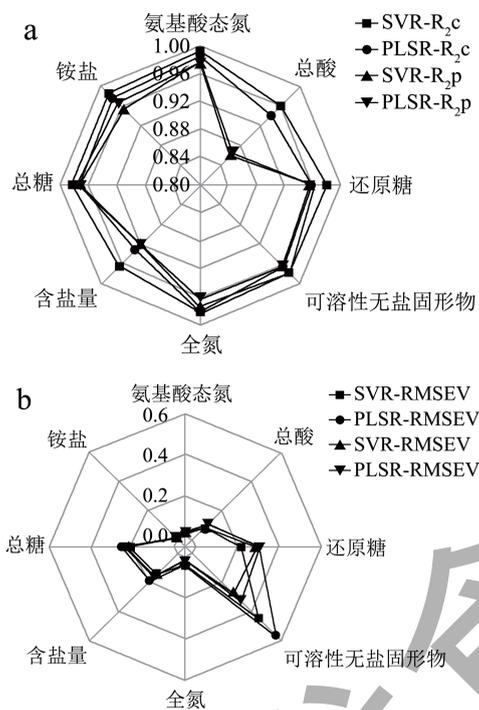


图4 不同建模方法下广式酱油样品的理化指标定量模型评价参数指标对比图

Fig.4 Comparison of parameters of quantitative model evaluation parameters of Cantonese soy sauce samples under different modeling methods

图4为不同建模方法下广式酱油样品的理化指标定量模型评价参数指标对比图。由图可以明显看出, 经波段筛选后 SVR 所建模型的训练集决定系数均高于 PLSR 模型, 训练集均方根误差均低于 PLSR 模型, 说明采用支持向量回归建模效果更优, SVR 模型的预测性能总体上比 PLSR 模型高, 可能是因为酱油发酵是一个不均衡的过程, 品质指标的变化是非线性的。

## 2.5 模型建立及评价

通过以上的条件优化, 以各成分的最优建模参数和各指标成分化学值导入 MATLAB 软件中, 采

用 SVR 建立不同发酵阶段广式酱油中氨基酸态氮、总氮、可溶性无盐固形物、总酸、铵盐、总糖、还原糖以及盐分的近红外光谱定量预测模型, 图5为广式酱油各理化指标化学值与近红外模型预测值之间的相关关系, 图6为广式酱油样品各理化指标的最佳波段分布结果图, 表5为广式酱油各理化指标的 SVR 最佳定量预测模型统计结果。

结果显示, 选用 SVNSG-RF-SVR 定量预测模型对广式酱油中氨基酸态氮、总酸预测效果最佳。其中, 氨基酸态氮的定量模型中训练集决定系数  $R^2$  为 0.991 1, 训练集均方根误差  $RMSEC$  为 0.018 0, 测试集决定系数  $R^2$  为 0.977 2, 测试集均方根误差  $RMSEP$  为 0.014 2; 总酸的定量模型中训练集决定系数  $R^2$  为 0.962 1, 训练集均方根误差  $RMSEC$  为 0.084 7, 测试集决定系数  $R^2$  为 0.857 9, 测试集均方根误差  $RMSEP$  为 0.085 8。

选用 SNV-CARS-SVR 定量预测模型对广式酱油中铵盐、总糖预测效果最佳。其中, 铵盐的定量模型中训练集决定系数  $R^2$  为 0.984 6, 训练集均方根误差  $RMSEC$  为 0.006 9, 测试集决定系数  $R^2$  为 0.954 4, 测试集均方根误差  $RMSEP$  为 0.008 2; 总糖的定量模型中训练集决定系数  $R^2$  为 0.982 0, 训练集均方根误差  $RMSEC$  为 0.213 4, 测试集决定系数  $R^2$  为 0.973 0, 测试集均方根误差  $RMSEP$  为 0.227 8。

选用 MSC-CARS-SVR 定量预测模型对广式酱油中全氮、可溶性无盐固形物预测效果最佳。其中, 全氮的定量模型中训练集决定系数  $R^2$  为 0.982 8, 训练集均方根误差  $RMSEC$  为 0.035 5, 测试集决定系数  $R^2$  为 0.976 4, 测试集均方根误差  $RMSEP$  为 0.015 2; 可溶性无盐固形物的定量模型中训练集决定系数  $R^2$  为 0.979 3, 训练集均方根误差  $RMSEC$  为 0.446 2, 测试集决定系数  $R^2$  为 0.977 9, 测试集均方根误差  $RMSEP$  为 0.269 4。

选用 MSCSG-RF-SVR 定量预测模型对广式酱油中还原糖预测效果最佳。其中, 还原糖的定量模型中训练集决定系数  $R^2$  为 0.980 0, 训练集均方根误差  $RMSEC$  为 0.213 2, 测试集决定系数  $R^2$  为 0.961 2, 测试集均方根误差  $RMSEP$  为 0.285 5。

选用 SG-CARS-SVR 定量预测模型对广式酱油中含盐量预测效果最佳。其中, 含盐量的定量模型中训练集决定系数  $R^2$  为 0.963 8, 训练集均方根误差  $RMSEC$  为 0.136 4, 测试集决定系数  $R^2$  为 0.921 6, 测试集均方根误差  $RMSEP$  为 0.158 8。

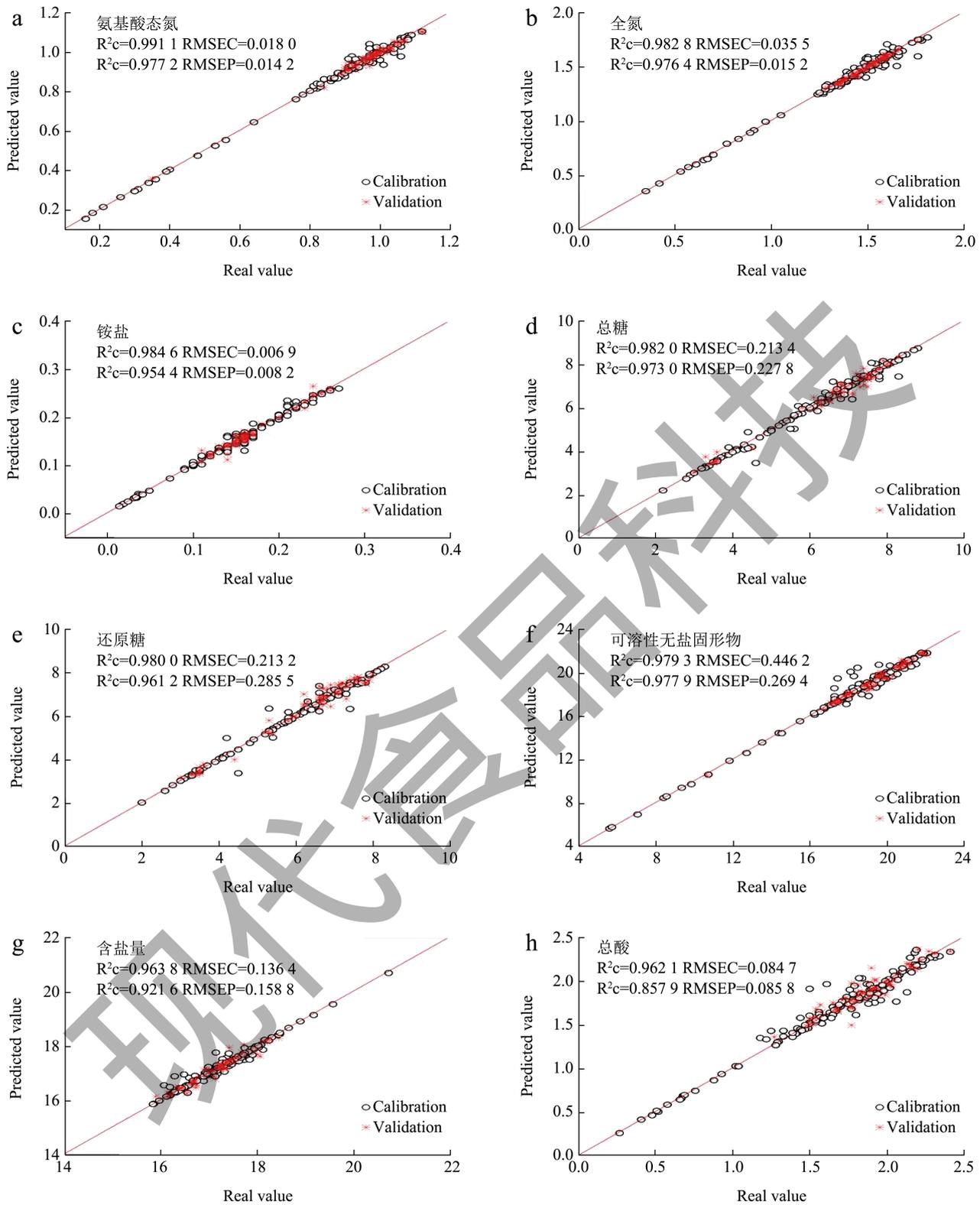


图 5 广式酱油中各理化指标的真实值与近红外预测值的相关图

Fig.5 The correlation graph between the real value and the predicted value of near infrared in Cantonese soy sauce

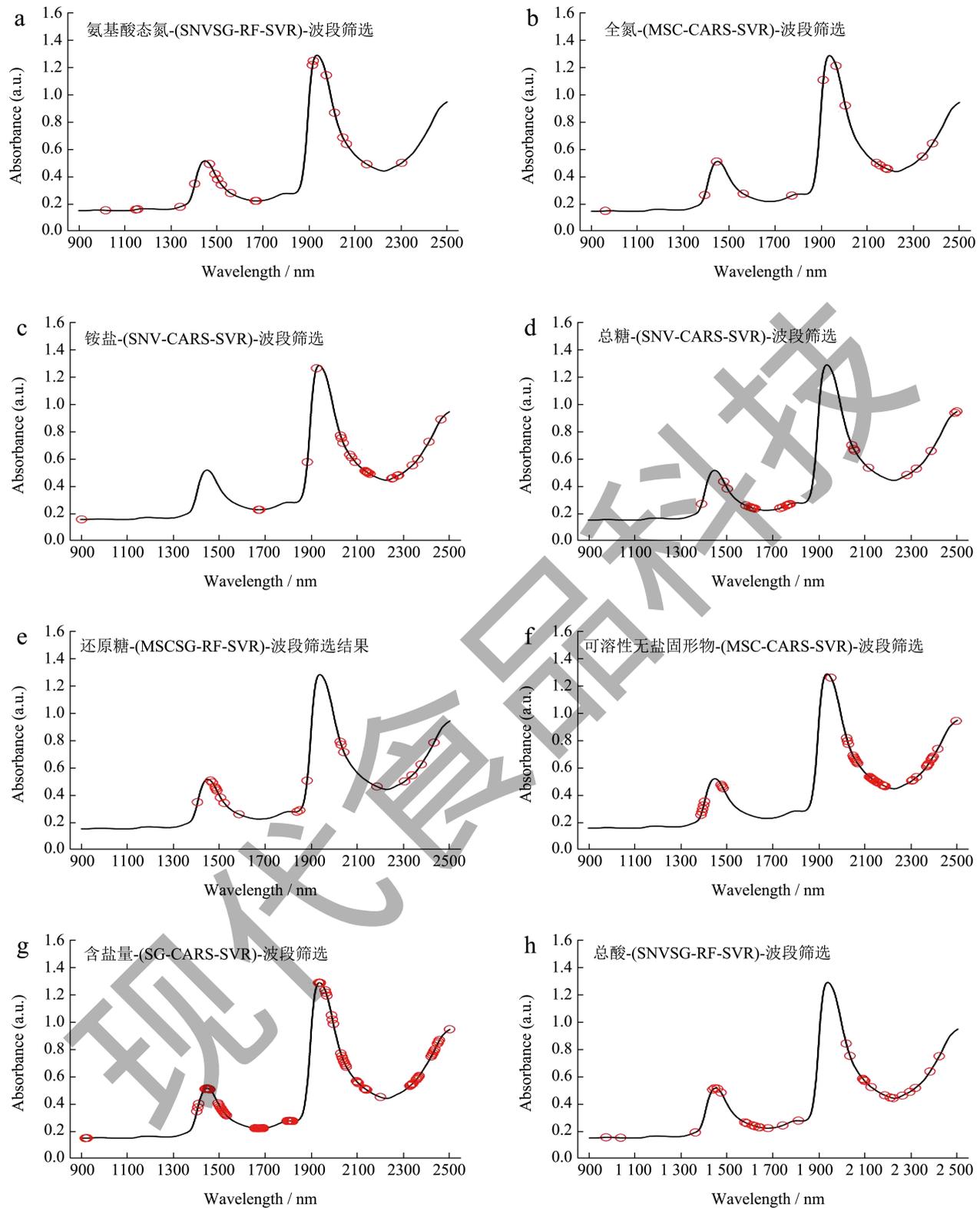


图 6 广式酱油样品各理化指标的最佳波段分布结果图

Fig.6 The optimal band distribution results of various physicochemical indexes of Cantonese soy sauce samples

从图 5 及表 5 可以看出, 酱油样品中各理化指标定量模型均匀分布在回归线两侧, 经特征波段提取后, 模型的变量数大大减少, 酱油中 8 种理化指标的训练集决定系数  $R^2$  均大于 0.90, 说明 8 种理化指标的化学实测值与其近红外预测值高度吻合。

目前, 近红外光谱分析技术对酱油品质检测方面已有不少的研究。付振浩<sup>[38]</sup>以 102 份工业生产中的毛油与成品酱油为研究对象, 利用 CARS 和 SPA 筛选特征波长, 对比 PLSR 和 SVR 的模型效果, 最终筛选出 SG-CARS-SVMR 为毛油与成品酱油氨基酸态氮最佳预测模型, 其训练集决定系数  $R^2$  为 0.979 0, 测试集决定系数  $R^2$  为 0.986 0; Hu 等<sup>[15]</sup>以 60 份不

同品牌的酱油样品为研究对象, 利用 Vis-NIR 技术结合改进 PSO 优化算法成功建立酱油样品中氨基酸态氮、盐、总酸和色比的定量预测模型, 其测试集决定系数  $R^2$  分别达到 0.999 7、0.946 2、0.999 6 和 0.999 8; Qin 等<sup>[16]</sup>以 120 份不同品牌草菇老抽为研究对象, 通过 Si-PLS 变量选择方法结合极限学习机 (ELM) 模型, 建立草菇老抽中氨基酸态氮含量无损检测的近红外预测模型, 其训练集决定系数  $R^2$  为 0.999 3, 测试集决定系数  $R^2$  为 0.965 7。但以上研究均是对成品酱油的 2~4 个品质参数进行研究, 未对酱油中其他理化指标进行系统全面的建模研究, 也没有对所建模型进行外部验证, 难以判断在实际应用上预测效果。

表 5 广式酱油中各理化指标的 SVR 最佳定量预测模型统计表

Table 5 Statistical table of SVR optimal quantitative prediction model for various physicochemical indexes in Cantonese soy sauce

指标	建模方法	变量数	$R^2_c$	RMSEC	$R^2_p$	RMSEP
氨基酸态氮	SNVSG-RF-SVR	20	0.991 1	0.018 0	0.977 2	0.014 2
全氮	MSC-CARS-SVR	14	0.982 8	0.035 5	0.976 4	0.015 2
铵盐	SNV-CARS-SVR	24	0.984 6	0.006 9	0.954 4	0.008 2
总糖	SNV-CARS-SVR	24	0.982 0	0.213 4	0.973 0	0.227 8
还原糖	MSCSG-RF-SVR	20	0.980 0	0.213 2	0.961 2	0.285 5
可溶性无盐固形物	MSC-CARS-SVR	41	0.979 3	0.446 2	0.977 9	0.269 4
含盐量	SG-CARS-SVR	79	0.963 8	0.136 4	0.921 6	0.158 8
总酸	SNVSG-RF-SVR	30	0.962 1	0.084 7	0.857 9	0.085 8

表 6 模型外部验证结果

Table 6 Model external validation results

指标	样品编号	实测值	预测值	绝对误差	指标	样品编号	实测值	预测值	绝对误差
氨基酸态氮	1	0.77	0.78	0.01	全氮	1	1.33	1.32	0.01
	2	0.90	0.94	0.04		2	1.48	1.52	0.04
	3	0.92	0.94	0.02		3	1.50	1.52	0.02
	4	0.95	0.90	0.05		4	1.51	1.49	0.02
	5	0.96	0.95	0.01		5	1.53	1.52	0.01
	6	0.97	0.97	0.00		6	1.51	1.53	0.02
	7	0.96	0.97	0.01		7	1.51	1.54	0.03
	8	0.82	0.81	0.01		8	1.26	1.25	0.01
	9	0.89	0.86	0.03		9	1.28	1.32	0.04
	10	0.97	0.99	0.02		10	1.35	1.33	0.02
	11	0.99	1.00	0.01		11	1.34	1.34	0.00
	12	1.01	1.01	0.00		12	1.37	1.36	0.01
	13	1.02	1.00	0.02		13	1.37	1.36	0.01
	14	1.01	1.01	0.00		14	1.38	1.37	0.01

续表 6

指标	样品编号	实测值	预测值	绝对误差	指标	样品编号	实测值	预测值	绝对误差
铵盐	1	0.08	0.09	0.01	可溶性无盐 固形物	1	17.68	17.56	0.12
	2	0.11	0.13	0.02		2	19.21	20.52	1.31
	3	0.12	0.13	0.01		3	19.98	20.62	0.64
	4	0.13	0.12	0.01		4	20.23	19.77	0.46
	5	0.14	0.14	0.00		5	20.13	20.38	0.25
	6	0.15	0.15	0.00		6	20.27	20.57	0.30
	7	0.15	0.15	0.00		7	20.19	20.56	0.37
	8	0.14	0.13	0.01		8	16.99	16.67	0.32
	9	0.17	0.17	0.00		9	17.06	17.13	0.07
	10	0.23	0.23	0.00		10	17.14	17.41	0.27
	11	0.25	0.26	0.01		11	17.47	17.17	0.30
	12	0.25	0.25	0.00		12	17.31	17.04	0.27
	13	0.26	0.25	0.01		13	17.48	17.55	0.07
	14	0.25	0.26	0.01		14	17.85	17.06	0.79
总糖	1	6.50	6.75	0.25	含盐量	1	17.88	17.86	0.02
	2	7.80	8.15	0.35		2	17.53	17.34	0.19
	3	7.90	7.71	0.19		3	17.16	17.29	0.13
	4	8.30	7.65	0.65		4	17.21	17.34	0.13
	5	8.20	8.03	0.17		5	17.23	17.23	0.00
	6	8.10	7.91	0.19		6	17.87	18.40	0.53
	7	7.90	8.19	0.29		7	17.35	17.26	0.09
	8	5.50	5.50	0.00		8	16.61	16.85	0.24
	9	5.20	5.20	0.00		9	16.74	16.40	0.34
	10	3.60	3.54	0.06		10	16.72	16.82	0.10
	11	3.00	2.98	0.02		11	16.81	16.75	0.06
	12	3.10	3.05	0.05		12	16.65	16.79	0.14
	13	3.30	3.28	0.02		13	16.66	16.56	0.10
	14	2.80	2.81	0.01		14	16.45	16.55	0.10
还原糖	1	6.80	6.41	0.39	总酸	1	1.20	1.39	0.19
	2	7.20	7.50	0.30		2	1.35	1.46	0.11
	3	7.90	7.67	0.23		3	1.40	1.42	0.02
	4	8.00	7.73	0.27		4	1.46	1.33	0.13
	5	7.80	7.78	0.02		5	1.61	1.49	0.12
	6	7.60	7.78	0.18		6	1.68	1.50	0.18
	7	8.00	7.58	0.42		7	1.78	1.71	0.07
	8	5.40	5.03	0.37		8	1.44	1.50	0.06
	9	5.00	4.90	0.10		9	1.58	1.62	0.04
	10	3.50	3.28	0.22		10	1.93	1.95	0.02
	11	3.00	3.10	0.10		11	2.09	2.28	0.19
	12	3.10	3.12	0.02		12	2.19	2.25	0.06
	13	3.00	3.22	0.22		13	2.23	2.22	0.01
	14	2.70	3.11	0.41		14	2.00	2.39	0.39

表 7 酱油样品实测值和近红外预测值  $t$  检验结果Table 7 The  $t$ -test results of the measured values and the near-infrared predicted values of soy sauce samples

配对样本	配对差值					$t$	自由度	Sig. (双尾)
	平均值	标准偏差	标准误差 平均值	差值 95% 置信区间				
				上限	下限			
实测值 - 预测值	-0.008 84	0.238 58	0.022 54	-0.053 51	0.035 83	-0.392	111	0.696

为更好评价所建模型对不同发酵阶段广式酱油样品中各理化指标的预测能力,选取 14 个非建模的不同发酵阶段广式酱油样本进行模型的外部独立样品验证,利用最佳模型对这 14 个酱油样品的理化指标进行预测,表 6 为模型验证结果。结果显示,酱油各指标的预测值与化学值之间的绝对误差均  $\leq 1.31$ 。为了进一步验证所建模型的可靠性,运用 SPSS 软件对样品实测值和近红外预测值进行  $t$  检验,以判定这两组检测结果是否有显著性差异,结果见表 7,由此可知,两组数据的  $P$  值为  $0.696 > 0.05$ ,表明样品的实测值与近红外模型预测值在  $\alpha=0.05$  的水平上无显著性差异<sup>[40]</sup>,通过对近红外光谱进行预处理和特征波段筛选所得定量预测模型的预测精度好,可用来对广式酱油的各理化指标进行检测,同时也说明采用近红外光谱分析技术对酱油品质进行预测具有可行性,为未来在线近红外检测仪的研制提供了有效的数据支撑。

### 3 结论

本研究采用近红外光谱技术构建了不同发酵阶段广式酱油中氨基酸态氮、总氮、可溶性无盐固形物、总酸、铵盐、总糖、还原糖以及盐分 8 种理化指标定量预测模型。酱油近红外光谱经 PCA 结合马氏距离剔除异常值、光谱预处理 (SG 平滑、MSC、SNV、MSC+SG、SNV+SG)、SPXY 划分数据集、特征波段筛选 (CARS、RF) 后,采用偏最小二乘回归 (PLSR) 和支持向量回归 (SVR) 建模,并比较不同建模方法的模型参数以及对筛选出最佳的定量预测模型进行外部检验。结果显示,经光谱预处理和特征波段筛选后,定量预测模型的预测能力和稳定性逐步上升,同时大大降低模型复杂度。SVR 所建模型的确定系数均高于 PLSR 模型,均方根误差均低于 PLSR 模型,说明采用支持向量回归建模效果更优。其中,广式酱油中这 8 种理化指标的训练集决定系数  $R^2c$  分别为 0.991 1、0.982 8、0.984 6、0.982 0、0.980 0、0.979 3、0.963 8、0.962 1,测试

集决定系数  $R^2p$  分别为 0.977 2、0.954 4、0.973 0、0.961 2、0.977 9、0.921 6、0.857 9。外部验证结果显示酱油各指标的预测值与化学值之间的绝对误差  $\leq 1.31$ ,  $t$  检验结果显示两组数据的无显著性差异,表明酱油样品的各理化指标的实际含量与近红外预测值高度吻合,所建定量模型对酱油发酵过程中理化指标的含量具有良好预测能力。本研究表明近红外光谱技术在广式酱油实时监测和快速分析方面具有可行性和优越性,为广式酱油中其他成分的快速检测提供了借鉴,同时为未来在线近红外检测仪的研制提供了有效的数据支撑。

### 参考文献

- [1] 杨晓璇,李阳,马宁,等.酱油中挥发性风味物质的气相色谱-离子迁移谱分析[J].中国酿造,2021,40(3):149-154.
- [2] 张丽.高盐稀态酱油发酵过程中添加酵母的研究[D].贵阳:贵州大学,2019.
- [3] TAN G L, WANG Y, HU M, et al. Comparative evaluation of the microbial diversity and metabolite profiles of Japanese-style and Cantonese-style soy sauce fermentation [J]. *Frontiers in Microbiology*, 2022, 13: 976206.
- [4] QI Q, HUANG J, ZHOU R Q, et al. Characterizing microbial community and metabolites of Cantonese soy sauce [J]. *Food Bioscience*, 2021, 40: 100872.
- [5] 于茜雅,鲁骞,吴昌正,等.温度对广式高盐稀态酱油原油品质的影响[J].食品科学,2023,44(22):55-63.
- [6] ZHANG Y J, FENG Y X, SHI H Q, et al. Impact of steam explosion pretreatment of defatted soybean meal on the flavor of soy sauce [J]. *LWT*, 2022, 156: 113034.
- [7] 桑力青,徐冰洁,杨晨曦,等.近红外光谱技术在食品和农产品领域中的应用[J].食品安全导刊,2023,368(3):127-129.
- [8] COZZOLINO D. The ability of near infrared (NIR) spectroscopy to predict functional properties in foods: challenges and opportunities [J]. *Molecules*, 2021, 26(22): 6981.
- [9] LIM D H, HA J W, CHOI H, et al. Recent progress of ultra-narrow-bandgap polymer donors for NIR-absorbing organic solar cells [J]. *Nanoscale Advances*, 2021, 3(15): 4306-4320.
- [10] 郭洁,解亚娟,李惠,等.近红外光谱法对脑心通胶囊药粉

- 混合均匀度在线检测的应用[J].临床医药文献电子杂志,2020,7(43):125,127.
- [11] 李登珊,李丽娜,张认成.基于SPA和PSO-LSSVM的可见-近红外光谱水质pH值检测[J].激光与光电子学进展,2023,60(1):390-395.
- [12] 袁壮.对比学习的近红外光谱及其在牛奶掺假检测中的应用研究[D].南宁:广西大学,2022.
- [13] PASQUINI C. Near infrared spectroscopy: A mature analytical technique with new perspectives-A review [J]. *Analytica Chimica Acta*, 2018, 1026: 8-36.
- [14] MISHRA P, PASSOS D. Realizing transfer learning for updating deep learning models of spectral data to be used in new scenarios [J]. *Chemometrics and Intelligent Laboratory Systems*, 2021, 212: 104283.
- [15] HU L Q, YIN C L, MA S, et al. Vis-NIR spectroscopy combined with wavelengths selection by PSO optimization algorithm for simultaneous determination of four quality parameters and classification of soy sauce [J]. *Food Analytical Methods*, 2019, 12(3): 633-643.
- [16] OUYANG Q, CHEN Q S, ZHAO J W, et al. Determination of amino acid nitrogen in soy sauce using near infrared spectroscopy combined with characteristic variables selection and extreme learning machine [J]. *Food and Bioprocess Technology*, 2013, 6(9): 2486-2493.
- [17] CHEN J M, FU C L, PAN T. Modeling method and miniaturized wavelength strategy for near-infrared spectroscopic discriminant analysis of soy sauce brand identification [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2022, 277: 121291.
- [18] GB/T 18186-2000,酿造酱油[S].
- [19] GB 5009.234-2016,食品中铵盐的测定[S].
- [20] GB/T 5009.7-2016,食品中还原糖的测定[S].
- [21] GB/T 15672-2009,食用菌中总糖含量的测定[S].
- [22] 石鲁珍,张景川,王彦群,等.马氏距离与浓度残差剔除近红外异常样品研究[J].中国农机化学报,2016,37(6):99-103.
- [23] 褚小立,陆婉珍.近红外分析中光谱预处理及波长选择方法进展与应用[J].化学进展,2004,4:528-542.
- [24] ENGEL J, GERRETZEN J, SZYMAŃSKA E, et al. Breaking with trends in pre-processing? [J]. *TrAC Trends in Analytical Chemistry*, 2013, 50: 96-106.
- [25] RONG Y, ZAREEF M, LIU L, et al. Application of portable Vis-NIR spectroscopy for rapid detection of myoglobin in frozen pork [J]. *Meat Science*, 2023, 201: 109170.
- [26] UPADHYAY R, GUPTA A, MISHRA H N, et al. At-line quality assurance of deep-fried instant noodles using pilot scale visible-NIR spectroscopy combined with deep-learning algorithms [J]. *Food Control*, 2022, 133: 108580.
- [27] 陈蓓,郑恩让,郭拓.多种光谱变量筛选算法在红参提取近红外建模中的应用[J].光谱学与光谱分析,2021,41(8):2443-2449.
- [28] LI H D, LIANG Y Z, XU Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. *Analytica Chimica Acta*, 2009, 648(1): 77-84.
- [29] DU C S, SUN L J, BAI H Y, et al. Quantitative detection of talcum powder in wheat flour based on near-infrared spectroscopy and hybrid feature selection [J]. *Infrared Physics Technology*, 2022, 123: 104185.
- [30] 张美鸣.近红外光谱法快速评定小麦营养价值的研究[D].泰安:山东农业大学,2023.
- [31] WOLD S, SJOSTROM M, ERIKSSON L. PLS-regression: a basic tool of chemometrics [J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, 58(2): 109-130.
- [32] 姜妍,蒙何,赵熠荣,等.青饲大豆主要品质参数的近红外光谱快速分析[J].光谱学与光谱分析,2022,42(3):919-923.
- [33] CONG S L, SUN J, MAO H P, et al. Non-destructive detection for mold colonies in rice based on hyperspectra and GWO - SVR [J]. *Journal of the Science of Food and Agriculture*, 2018, 98(4): 1453-1459.
- [34] FU D D, LI Q Y, CHEN Y, et al. Assessment of integrated freshness index of different varieties of eggs using the visible and near-infrared spectroscopy [J]. *International Journal of Food Properties*, 2023, 26(1): 155-166.
- [35] DI EGIDIO V, SINELLI N, GIOVANELLI G, et al. NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation [J]. *European Food Research and Technology*, 2010, 230(6): 947-955.
- [36] 罗一甲.基于近红外光谱技术的赤霞珠葡萄品质检测及采收期判别研究[D].石河子:石河子大学,2021.
- [37] 刘蕊,李和帅,吴翼.基于近红外光谱的椰肉含油率模型构建[J].食品科技,2021,46(12):276-280.
- [38] 付振浩.基于近红外光谱对高盐稀态酱油工业生产的快速检测研究[D].无锡:江南大学,2022.
- [39] 胡晓云,卞希慧,项洋,等.近红外光谱结合化学计量学对五元调和油中单组分油的定量分析[J].光谱学与光谱分析,2023,43(1):78-84.
- [40] 庞超明,黄弘.试验方案优化设计与数据分析[M].南京:东南大学出版社,2018.