

杨凯,何昱廷,李沃霖,等.基于近红外光谱技术检测全蛋粉掺假[J].华中农业大学学报,2024,43(2):264-272.
DOI:10.13300/j.cnki.hnlkxb.2024.02.029

基于近红外光谱技术检测全蛋粉掺假

杨凯¹,何昱廷¹,李沃霖¹,祝志慧^{1,2},王巧华^{1,2}

1. 华中农业大学工学院,武汉 430070; 2. 农业农村部长江中下游农业装备重点实验室,武汉 430070

摘要 为提高全蛋粉掺假检测的准确度和检测效果,应用近红外光谱技术对全蛋粉掺假进行定性判别并对掺假含量进行定量分析。分别采用标准正态变换、多元散射校正、卷积平滑、归一化、一阶导数、二阶导数等6种不同预处理方法,对原始光谱数据进行预处理,采用竞争性自适应重加权算法(CARS)、连续投影算法(SPA)及CARS-SPA结合算法对光谱数据进行特征波长筛选,建立集成学习(ensemble learning, EL)模型对掺假蛋粉进行定性判别,建立偏最小二乘模型(PLSR)对掺假含量进行定量分析。结果显示:在对掺假蛋粉进行定性判别方面,一阶导数为最佳预处理方法,CARS算法特征筛选效果最佳,EL模型对掺假蛋粉样本总体判别准确率达到98.18%,对各类掺假蛋粉样品的判别准确率在97.78%以上。在对蛋粉掺假含量进行定量分析方面,多元散射校正为最佳预处理方法,CARS算法特征筛选效果更佳,对一组分掺假、二组分掺假、三组分掺假和所有掺假样本的PLSR浓度预测模型的预测集相关系数(R_p)分别为0.958 5、0.931 2、0.945 6和0.955 8,均方根误差(RMSEP)分别为4.689 1、5.813 4、4.604 1和3.802 9。研究结果表明,近红外光谱技术可用于蛋粉掺假检测,为监管机构检测蛋粉掺假提供参考。

关键词 全蛋粉; 近红外光谱; 掺假检测; 集成学习; 精准检测

中图分类号 TS253.7 **文献标识码** A **文章编号** 1000-2421(2024)02-0264-09

蛋粉的蛋白质含量高,矿物质和维生素种类丰富,具有显著的功能性质(如凝胶性、乳化性、保脂性、保水性等),是鸡蛋的理想替代品^[1-2]。相比于新鲜鸡蛋,蛋粉储存期长、使用方便、便于包装运输,已广泛应用于食品工业、畜禽生产、化妆品及医疗等领域^[3-5]。目前,国内蛋粉标准体系尚不完善,监管领域存在一定的空白。少数蛋粉生产企业为获得更高的利润,在蛋粉中掺入淀粉、大豆蛋白、植脂末和麦芽糊精等廉价物质,不仅扰乱了蛋粉市场的秩序,也给食品安全带来了极大隐患。因此,研究蛋粉掺假检测技术,对保障蛋粉质量安全具有重要意义。目前关于蛋粉掺假检测的研究报道较少,现在已有的蛋粉掺假检测主要依据GB 2749—2015《食品安全国家标准》,对蛋粉的主要组分进行分析,方法多为常规的化学分析手段。蛋粉掺假检测方法主要有2种:一是感官识别法,通过闻气味、辨颜色、试手感、尝口感等方法初步判断蛋粉的真假,该方法主观因素太强,一般不单独作为检测标准;二是基于DNA分子、

蛋白质分子进行检测的生物分子技术,主要包括色谱法、质谱法以及DNA条形码等技术^[6],这类方法特异性强、灵敏度高,但操作繁琐,耗时长,成本高,不适合大规模推广使用。此外,刘平等^[7]探索了基于高光谱的全蛋粉掺假的无损检测模型,对掺假含量为1%~50%的蛋粉样品进行掺假识别,总体分类准确率在90%以上,但模型稳健性和检测精度有待提高。

近红外光谱技术是一种便捷的绿色分析检测技术,已经广泛应用于奶粉、莲子粉等粉末状食品掺假检测中。Capuano等^[8]应用近红外光谱技术,建立偏最小二乘(PLSR)模型对奶粉中的掺假物进行预测,所建立的PLSR模型显示出良好的预测效果,可将掺假量为2%的淀粉和麦芽糊精分辨出来。Karunathilaka等^[9]使用近红外光谱技术结合SIMCA软件研制了手持式近红外检测设备,可以快速检测奶粉真实性。胡仁伟等^[10]使用近红外光谱技术结合深度信念网络鉴别莲子粉真伪,建立的判别模型能够有效识别掺杂各类其他作物粉的莲子粉,平均识别

收稿日期:2022-12-09

基金项目:国家自然科学基金项目(32072302)

杨凯, E-mail: 3153271420@qq.com

通信作者,祝志慧, E-mail: zzh@mail.hzau.edu.cn

率达到96%。Muller-Maatsch等^[11]将紫外-可见光、荧光和近红外光谱相结合,鉴别添加植物蛋白粉、乳清粉、淀粉、乳糖、葡萄糖、果糖的脱脂奶粉,检测效果良好。

综上所述,传统的蛋粉生物化学方法检测效率低、仪器价格昂贵,无法满足大规模、快速、无损的检测需求。近红外光谱技术简单方便,适用于食品工业质量控制。本研究基于近红外光谱技术对全蛋粉掺假物进行定性检测,并对掺假物的含量进行预测,旨在为提高全蛋粉掺假检测准确度和检测效果提供参考。

1 材料与amp;方法

1.1 样品制备

购买湖北神地科贸有限公司、安徽荣达食品有限公司、江苏康德蛋业有限公司3种全蛋粉以及实验室自制的全蛋粉共4种全蛋粉。为了确保纯蛋粉之间的差异性,选择不同生产批次的纯蛋粉,最终称取纯蛋粉样品60个。为确保纯蛋粉的真实性,参照GB 5009.5—2016《食品中蛋白质的测定》中的凯氏定氮法测定蛋粉中的蛋白质含量,每次测定取纯蛋粉样品0.5 g,氮换算为蛋白质的系数 F 为6.25,参照GB 5009.6—2016《食品中脂肪的测定》中的索氏抽提法测定蛋粉中的脂肪含量,每次测定取纯蛋粉样品2~5 g,最终测得纯蛋粉样品的蛋白质含量 $\geq 45.0\%$,脂肪含量 $\geq 40.0\%$,在合理范围内^[12]。

选择市面上常见的3种全蛋粉掺假物:大豆分离蛋白(优嘉宝食品有限公司)、麦芽糊精(优嘉宝食品有限公司)、小麦淀粉(宝鼎天鱼有限公司)。蛋粉的掺假分为一组分掺假:单独掺入大豆分离蛋白、麦芽糊精、小麦淀粉的1种,共180个样品;二组分掺假:掺入3种掺假物质中的2类,每类掺假物含量占掺假总含量的50%,共180个样品;三组分掺假:掺入全部3种掺假物质,每类掺假物含量占掺假总含量的1/3,共240个样本。3类掺假物的掺假含量均分别为0.1%、0.5%、1%、2%、5%、10%、15%、20%、30%、40%、50%、60%,最终制备掺假样品600个。对全部纯蛋粉和掺杂物过孔径1 mm标准筛进行处理,确保试验样品的颗粒大小均匀。将掺假后的蛋粉置于搅拌机中搅拌超过1 min,以确保掺假物颗粒均匀分布在蛋粉样品中。试验中把所有样本均放入样品袋(约装有样品10 g)并置于4℃冰箱中贮藏。制备的试验样品如图1所示。

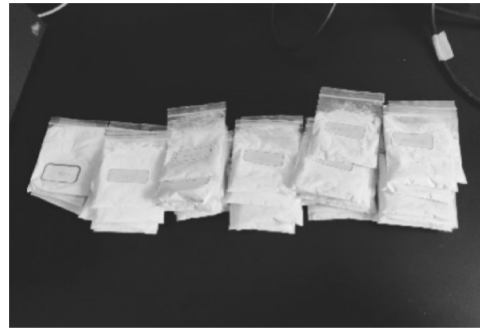


图1 试验样本

Fig.1 Samples of the experiment

1.2 光谱数据采集

光谱仪为美国 Thermo Scientific 公司的 Antaris II 型傅里叶近红外光谱分析仪。波数 10 000~4 000 cm^{-1} (1 000~2 500 nm),分辨率为 3.86 cm^{-1} ,扫描次数 32,测试方式为漫反射。采集光谱时,将样本倒入测试器皿中,以空气作为参比,环境温度为 25℃,每个样品重复扫描 3 次,取其平均光谱作为原始光谱,光谱数据以 *.csv 格式存储。使用 Matlab2016b 和 Python3.9.5 软件进行数据分析处理。

1.3 样本划分与均衡处理

样本类别不平衡是指分类任务中不同类别的样本数目差别很大。如果类别比例超过 4:1,则视为样本不平衡。样本不平衡会导致实际预测结果对多数类别有侧重。为避免样本不平衡问题,在构建分类模型之前,需要对分类不平衡性问题进行处理^[13-14]。合成少数类过采样技术(synthetic minority oversampling technique, SMOTE)是基于随机过采样算法的一种改进方案^[15-16],其算法的流程为:对于少数类中每一个样本 X ,以欧氏距离为标准计算它到少数类样本集中所有样本的距离,得到其 K 个近邻(本研究中 $K=5$);从 K 个近邻中随机选择 1 个样本,假设选择的近邻为 X_i ;生成 1 个 0~1 之间的随机数,用式(1)合成 1 个新样本 X_{new} 。

$$X_{\text{new}} = X + \text{rand}(0, 1) \times (X - X_i) \quad (1)$$

KS(Kennard-Stone, KS)算法可以有效提取光谱数据差异较大的样品作为训练集,剩余样本划分为测试集,本研究使用 KS 算法将样本集按照 3:1 划分为训练集和测试集。KS 划分数据集后,训练集样本 495 个(真样本 45 个、掺假样本 450 个),测试集样本 165 个(真样本 15 个、掺假样本 150 个)。

训练集样本存在类别不平衡的情况,采用 SMOTE 算法,对训练集样本以掺假样本和纯样本 1:1 的比例生成纯蛋粉样本。最终得到训练集数据

900个(真样本450个、掺假样本450个),测试集数据165个(真样本15个、掺假样本150个)。

1.4 光谱预处理

为了减少全蛋粉样本光谱原始数据中噪声和冗余信息的影响,并使有用信息更突出,本研究比较标准正态变换(standard normal variate, SNV)、多元散射校正(multiplicative scatter correction, MSC)、卷积平滑(savitzky golay, SG)、归一化(normalize)、一阶导数(first derivative, FD)、二阶导数(second derivative, SD)共6种不同的预处理方法的建模效果。

1.5 特征波长筛选方法

采集到的原始光谱变量数有1557个,这些光谱变量之间存在大量的冗余变量甚至是干扰变量,如果采用全波段信息建立分析判别模型,不仅会加重计算负荷,还会影响模型精度,降低模型稳健性^[17-18]。本研究采用SPA、CARS以及SPA-CARS算法进行特征波长提取。

1.6 建模方法

采用集成学习模型对掺假蛋粉进行定性分析,集成学习组合了多个弱监督模型以得到一个更全面、更稳定的模型,弥补了单一模型受敏感样本扰动大、稳健性不高的缺陷。集成学习算法主要有2种:基于Bagging的算法和基于Boosting的算法。Bagging方法在训练过程中,各分类器之间并行训练。Boosting方法训练各分类器采用串行的方式^[19-20]。本研究选用Boosting方法(图2),如图2所示,首先通过Bootstrapping产生*k*个训练集样本,然后基于初始带的权重训练出弱学习器 f_1 ,对前一层分类器分错的样本,给予更高的权重 w_k ,如此循环训练。最后通过一定的结合策略形成强学习器 $f(x)$ 。 $f(x)$ 计算公式如下:

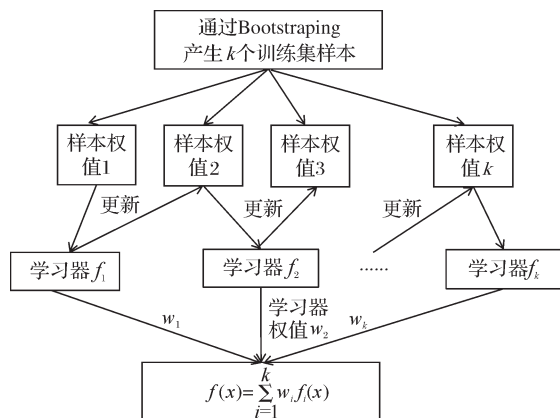


图2 Boosting算法原理图

Fig.2 Schematic diagram of Boosting algorithm

$$f(x) = \sum_{i=1}^k w_i f_i(x) \tag{2}$$

采用PLSR模型对掺假蛋粉的掺假物含量进行定量分析,PLSR建模过程结合了多元线性回归分析、典型相关性分析、主成分回归分析的优点,可以在建模分析过程中兼顾光谱数据和掺假含量对模型的影响,在复杂样品的光谱数据回归分析中应用广泛^[21]。

1.7 模型评价指标

定性分析模型的性能评价指标由准确率(accuracy, Acc,公式中用 A_{cc} 表示)、掺假识别率(adulteration recognition rate, Arr,公式中用 A_{rr} 表示)进行评价。

$$A_{cc} = \frac{T_p + T_N}{T_p + F_N + F_p + T_N} \tag{3}$$

$$A_{rr} = \frac{T_p}{T_p + F_N} \tag{4}$$

式(3)~(4)中: T_p 为掺假蛋粉被正确识别出来的个数; F_N 为掺假蛋粉被误判的个数; T_N 为纯蛋粉被正确识别的个数; F_p 为纯蛋粉被误判的个数。定量分析模型的性能由训练集相关系数 R_c 、训练集均方根误差RMSEC(公式中用 R_{MSEC} 表示)、测试集相关系数 R_p 、测试集均方根误差RMSEP(公式中用 R_{MSEP} 表示)决定,相关系数越接近1,均方根误差越小,模型的效果越好。

$$R_c = \frac{\sqrt{\sum_{i=1}^{n_c} (y_{pi} - y_{mi})^2}}{\sqrt{\sum_{i=1}^{n_c} (y_{pi} - y_{mean})^2}} \tag{5}$$

$$R_p = \frac{\sqrt{\sum_{i=1}^{n_p} (y_{pi} - y_{mi})^2}}{\sqrt{\sum_{i=1}^{n_p} (y_{pi} - y_{mean})^2}} \tag{6}$$

$$R_{MSEC} = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (y_{pi} - y_{mi})^2} \tag{7}$$

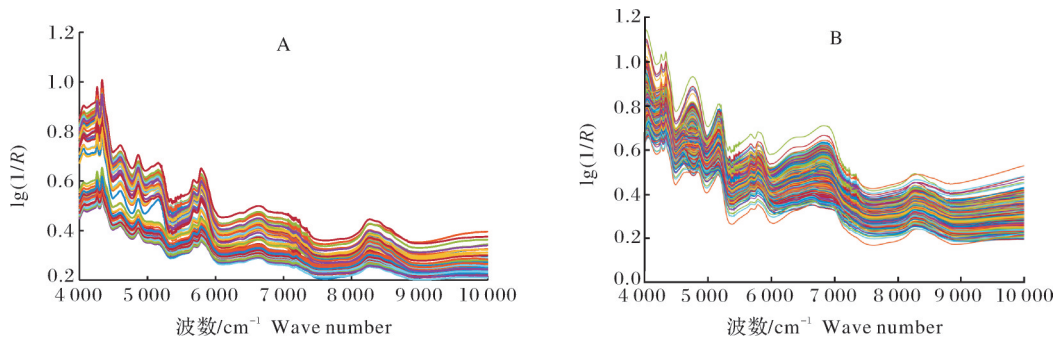
$$R_{MSEP} = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (y_{pi} - y_{mi})^2} \tag{8}$$

式(5)~(8)中, y_{pi} 为预测集中第*i*个样本预测值; y_{mi} 为第*i*个样本的实际值; y_{mean} 为样本的平均值。

2 结果与分析

2.1 原始光谱数据分析

1)纯蛋粉与掺假蛋白粉光谱比较。图3为样品原始光谱图,对比图3A,B可知,纯蛋粉和掺假蛋白粉的原始光谱图不管是在波峰、波谷位置还是变化趋势上都十分相似,随着波数的增加吸光值随之减小,掺假样品的变化幅度比真样品的变化幅度稍大。掺假蛋粉和纯蛋白粉的原始光谱重叠严重,通过肉眼不能进行有效识别,需要通过数学建模的方法来提取光谱的有效信息。



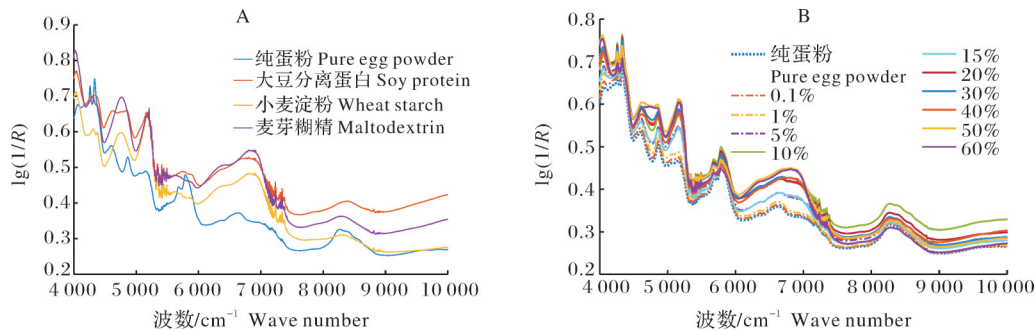
A: 真样品 Real samples; B: 掺假样品 Adulterated samples.

图3 样品原始光谱图

Fig.3 Original spectra of samples

2)不同掺假物蛋粉光谱比较。图4A是3种掺假物和纯蛋粉的光谱图,由图4A可知,掺假物和纯蛋粉在4 600~4 850 cm^{-1} 、6 600~7 000 cm^{-1} 以及7 400~8 150 cm^{-1} 区分比较明显,在4 600~4 850 cm^{-1} 和6 600~7 000 cm^{-1} 范围内吸光值大小为:麦芽糊精>大豆分离蛋白>小麦淀粉>纯蛋粉,在7 400~8 150 cm^{-1} 范围内吸光值大小为:大豆分离蛋白>麦芽糊精>小麦淀粉>纯蛋粉,说明这几个波段存在O-H(多来自水分子)、C-H(多来自有机分子)和N-H(多来自蛋白质)倍频和合频

的吸收峰,这些官能团与蛋粉的主要组分(蛋白质、脂肪、水分等物质)有着密不可分的联系,4 800 cm^{-1} 附近为C-H的一级倍频组合吸收谱带,8 000 cm^{-1} 附近为C-H的一级倍频组合吸收谱带^[22-23]。图4B是不同掺假含量蛋粉的原始光谱图,从图4B可知,不同掺假含量蛋粉和纯蛋粉光谱图的总体变化趋势相同,大体上随着掺假物含量的增加使得光谱吸光值增加,掺假蛋粉和纯蛋粉在4 800~5 300 cm^{-1} 和5 900~7 000 cm^{-1} 有较好的区分性。



A: 不同掺假物 Different adulterants; B: 不同掺假物含量蛋粉 Different adulterated egg powder concentrations.

图4 不同掺假含量蛋粉光谱图

Fig.4 Spectra of egg powder with different adulteration concentrations

2.2 掺假蛋粉定性分析模型建立

1)预处理结果。基于原始光谱和预处理后的光谱数据,建立EL分类模型,比较不同预处理方法对模型分类效果的影响,结果如表1所示。基于原始光谱数据建立的EL模型测试集准确率为93.94%,掺假识别率为96.00%。EL分类模型在一阶导数预处理后建模效果较好,测试集准确率达到98.18%,掺假识别率达到99.33%,总体判别准确率和掺假识别率均有提高。后续本研究定性检测所使用的光谱数据均为经过一阶导数处理的数据。

2)特征波长筛选结果。①CARS筛选结果。使

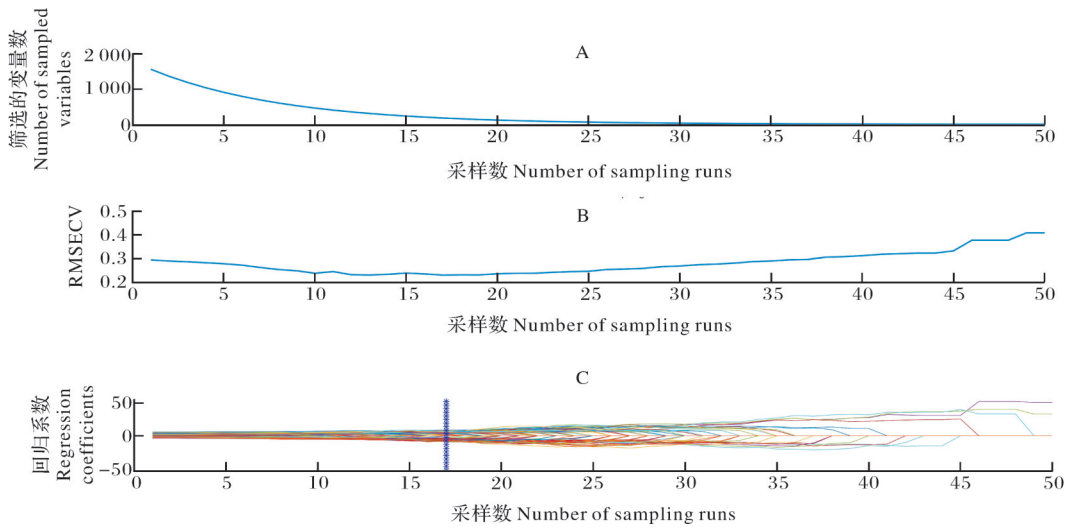
表1 基于不同预处理方法的EL建模结果

Table 1 Modeling results with different preprocessing methods

预处理 Pretreatment	训练集 Training set		测试集 Test set	
	Acc	Arr	Acc	Arr
无 None	98.11	98.00	93.94	96.00
MSC	99.78	99.56	95.15	96.88
SNV	97.44	97.11	97.58	98.67
FD	99.89	99.78	98.18	99.33
SD	99.78	99.78	96.97	98.67
SG	97.67	96.44	95.15	97.33
归一化 Normalize	96.00	95.78	92.12	96.00

用CARS算法对经过一阶导数预处理的蛋粉光谱数据进行特征波长筛选(图5)。初始设定蒙特卡罗采样50次,交叉验证设定50%,重采样率为0.8。如图5A所示,随着迭代次数的增加,筛选的变量会呈指数衰减的趋势减少;图5B中RMSECV值表明了CARS采样选择的特征波长建立模型的效果;图5C中的每条线代表每个特征变量的回归系数变化趋势,*号为CARS特征筛选取得建模效果最好的位置。CARS算法在采样第17次取得最小的均方根误差,最终从1557个波长中筛选出203个特征波长,筛选出的特征波长占波长总数的13.05%。

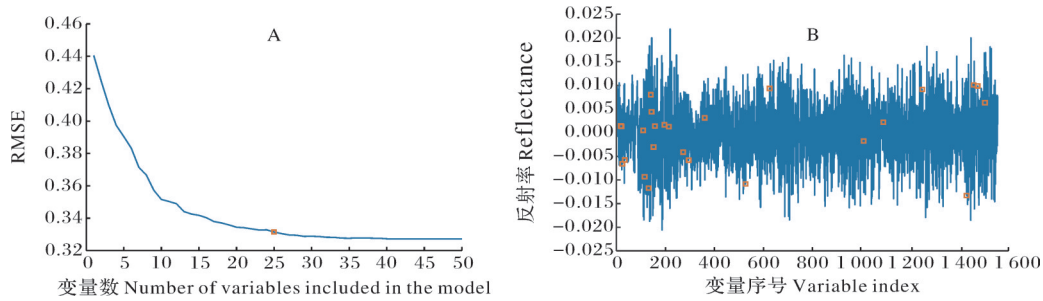
②SPA筛选结果。图6为SPA算法选择蛋粉光谱特征波长过程。SPA算法依据均方根误差RMSE的大小判断建立模型的好坏,如图6A所示,随着被选特征波长数量的增加,RMSE先快速下降,然后趋于平缓,最终筛选出25个特征波长,筛选出的特征波长占波长总数的1.61%,RMSE为0.33,选择此点而没有选择RMSE最低点的原因是在这个位置RMSE已经降到了合适范围,被选的特征变量包含了原始光谱的大部分信息。图6B为被选择的特征变量的分布情况。相对于CARS算法来说,SPA算法所筛选的特征波长较少。



A: 采样变量数 Number of sampled variables; B: RMSECV 趋势 RMSECV trend; C: 回归系数 Regression coefficient path.

图5 CARS变量筛选结果

Fig.5 Results of CARS variable screening



A: RMSE 趋势 RMSE trends; B: 选定的波长点序列 Selected wavelength point sequences.

图6 SPA变量筛选结果

Fig.6 Results of SPA variable screening

③CARS-SPA 筛选结果。为了尽可能在保证准确率的情况下减少特征波长,提高模型运行速度,将CARS和SPA算法结合,在CARS算法筛选一部分变量后使用SPA算法进行二次筛选。CARS-SPA算法共筛选出9个特征波长,筛选出的特征波长仅占波长总数的0.58%,对应波数分别为4 490、4 821、5 234、5 276、5 932、6 021、

6 900、7 617、8 134 cm^{-1} 。所筛选波长中有6个位于4 800~5 300 cm^{-1} 和5 900~7 000 cm^{-1} 范围内,这和图4中掺假蛋粉和纯蛋粉有明显区别的波段基本吻合。说明特征筛选算法对特征波长进行了有效筛选。

3) 蛋粉掺假定性分析结果。分别使用CARS、SPA和CRAS-SPA算法进行特征波长筛选,以筛选

后的波长为自变量,建立 EL 模型,结果如表 2 所示。3 种降维算法均有降低光谱变量的维数和,CARS、SPA 和 CRAS-SPA 筛选后的特征变量数从 1 557 个分别减少到 203、25 和 9 个。3 种特征筛选方法中 CARS 特征筛选后建模结果较佳,CARS-EL 建模的测试集准确率达到 98.18%,说明 CARS 算法在剔除掉干扰的变量的同时保留了对模型判别准确率贡献度高的变量,提高了模型的准确率和检测速度。SPA 和 CARS-SPA 筛选后建模的测试集准确率均

低于原始光谱数据建模的准确率,说明在特征筛选过程中去掉了部分有效的变量导致建模效果变差。模型的单个样本平均检测用时为 0.48 s,符合工业生产检测的要求。

测试集中不同种类的掺假物的检测结果如表 3 所示,一组分掺假、二组分掺假、三组分掺假和所有掺假样本的准确率均在 97.33% 以上,掺假识别率均在 97.78% 以上。说明 EL 模型对单一组分掺假和混合掺假都有较好的判别性能。

表 2 基于不同特征波长筛选方法的 EL 建模结果

Table 2 Modeling results of EL with different characteristic wavelength screening methods

预处理 Pretreatment	筛选算法 Filtering algorithm	特征数 Feature number	训练集 Training set		测试集 Test set		单个检测用时/s Time for individual tests
			Acc/%	Arr/%	Acc/%	Arr/%	
FD	CARS	203	99.67	99.33	98.18	98.67	0.48
FD	SPA	25	96.78	98.22	96.97	97.33	0.24
FD	CARS-SPA	9	99.78	99.55	97.58	98.00	0.17

表 3 不同种类掺假物的检测结果

Table 3 Results of the testing of different adulterants

掺杂种类 Types of adulteration	掺假样本 Adulterationsamples		真样本 Real samples		Acc/%	Arr/%
	总数 Total	误判数 Misjudgment	总数 Total	误判数 Misjudgment		
一组分掺假 One-component	45	0	15	1	98.33	100.00
二组分掺假 Two-component	45	1	15	1	96.67	97.78
三组分掺假 Three-component	60	1	15	1	97.33	98.33
所有掺假 All adulterated	150	2	15	1	98.18	98.67

2.3 掺假蛋粉定量分析模型建立

定量分析不存在不同类样本不均衡的问题,因此,不采用 SMOTE 算法对定量分析的数据集进行扩充。将掺假样本(共 600 个,一组分掺假 180 个,二组分掺假 180 个,三组分掺假 240)按照 2:1 划分为训练集和测试集,建立 PLSR 定量分析模型。

1) 预处理结果。不同预处理方法 PLSR 建模结果如表 4 所示。当主成分数为 14 时,MSC 预处理后建立的 PLSR 模型预测效果较好,模型的 R_c 和 R_p 分别为 0.981 2 和 0.940 9, RMSEC 和 RMSEP 分别为 3.914 2 和 4.053 5。

2) 蛋粉掺假定量分析结果。分别使用 CARS、SPA 和 CRAS-SPA 算法筛选特征波长,以筛选后的波长为自变量,建立 PLSR 模型(表 5)。结果显示,经过 CARS 筛选后建模效果最好,共筛选出 118 个特征波长,剔除了 92.42% 的波长变量,CARS 筛选后的 PLSR 模型预测集 R_p 为 0.955 8, RMSEP 为

表 4 基于不同预处理方法的 PLSR 建模结果

Table 4 Results of PLSR modeling with different pre-processing methods

预处理 Pretreatment	PCs	R_c	RMSEC	R_p	RMSEP
无 None	10	0.955 4	5.987 7	0.883 3	5.306 7
MSC	14	0.981 2	3.914 2	0.940 9	4.053 5
SNV	13	0.973 4	4.639 5	0.927 6	4.598 9
FD	15	0.989 9	3.206 4	0.875 5	5.470 4
SD	8	0.994 2	2.186 2	0.723 3	7.968 2
SG	16	0.989 4	2.942 1	0.931 7	4.422 7
归一化 Normalize	15	0.979 6	4.068 7	0.903 3	5.069 2

3.802 9, 预测效果较全波段建模有所提升。

使用 MSC-CARS-PLSR 模型分别对一组分掺假、二组分掺假、三组分掺假和预测集所有掺假样本的掺假含量进行的预测,结果如图 7 所示,预测集回归系数 R_p 分别为 0.958 5、0.931 2、0.945 6 和 0.955 8,

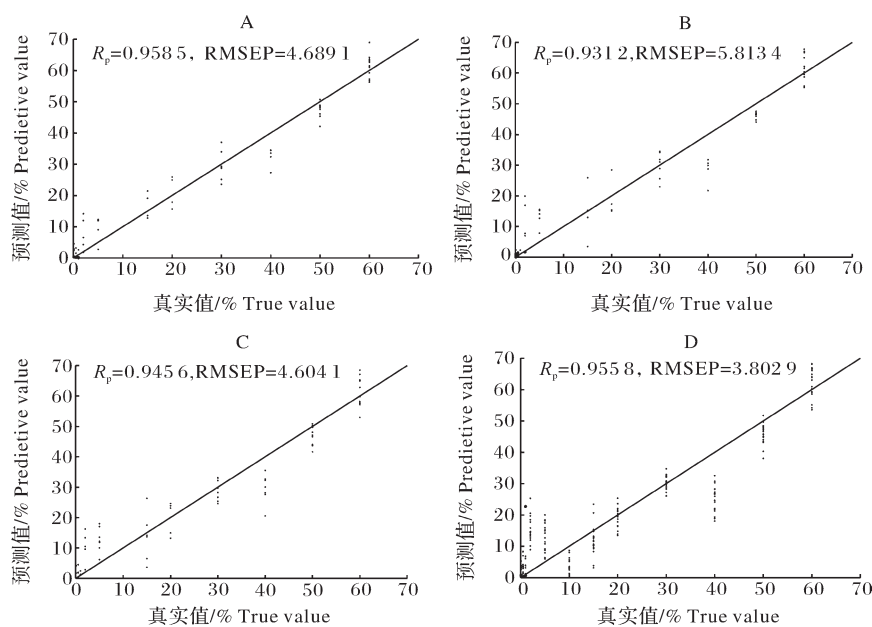
表5 基于不同特征波长筛选方法的PLSR建模结果

Table 5 Modeling results of PLSR with different characteristic wavelength screening methods

预处理 Pretreatment	筛选算法 Filtering algorithm	特征数 Feature number	R_c	RMSEC	R_p	RMSEP
MSC	CARS	118	0.996 2	2.759 3	0.955 8	3.802 9
MSC	SPA	30	0.968 3	5.066 5	0.890 9	7.792 4
MSC	CARS-SPA	27	0.983 9	3.623 5	0.927 6	5.424 3

RMSEP分别为4.689 1、5.813 4、4.604 1和3.802 9。模型对二组分掺假和三组分掺假的掺假含量预测结果略差于所有掺假样本的预测结果,其主要原因有2个:一是掺假含量梯度设置较广,每个含量的数据量偏小,近红外光谱差异较小,在数据量不充足的情况

下定量预测模型的适应性不够高;二是配置掺假样本时存在试验误差,特别是低含量掺假样品的实际掺假含量比样品标签偏高或偏低。此外,为了追求更高的利润,蛋粉市场掺假浓度多是偏高,定量分析模型的预测性能符合蛋粉市场的需求。



A: 一组分掺假样本 One-component adulterated samples; B: 二组分掺假样本 Two-component adulterated samples; C: 三组分掺假样本 Three-component adulterated samples; D: 所有掺假样本 All adulterated samples.

图7 掺假含量预测结果

Fig.7 Results of adulteration concentration prediction

3 讨论

本研究以大豆分离蛋白、麦芽糊精、小麦淀粉等3种市面上常见的掺杂物为掺假物,配置掺假含量为0.1%~60%的掺假蛋粉,采集掺假蛋粉和纯蛋粉的近红外光谱数据,经过比较不同预处理方法和特征波长筛选方法,建立全蛋粉掺假定性判别模型和掺假含量定量分析模型。结果表明,EL定性判别模型的最优预处理方法为FD,最优特征筛选方法为CARS,最优模型对掺假蛋粉的总体判别准确率达到98.18%,掺假识别率达到98.67%,单个样品平均检测用时0.48 s。一组分掺假、二组分掺假、三组分掺假的蛋粉判别准确率分别为98.33%、96.67%、

97.33%,掺假识别率分别为100.00%、97.78%、98.33%,所建立的定性判别模型可以实现全蛋粉掺假的定性判别,满足在线检测需求。PLSR定量分析模型的最优预处理方法为MSC,最优特征筛选方法为CARS,最优模型检测下的一组分掺假、二组分掺假、三组分掺假和所有掺假样本的 R_p 分别为0.958 5、0.931 2、0.945 6和0.955 8, RMSEP分别为4.689 1、5.813 4、4.604 1和3.802 9,可以实现对全蛋粉掺假含量的定量分析。

目前全蛋粉品质检测的研究还较少,相比于生物化学检测技术,本研究使用近红外光谱技术检测全蛋粉的品质,无需复杂的前处理操作且检测速度快,更适合现场检测;与文献[12]的基于高光谱技术

检测全蛋粉掺假的研究相比,本试验的掺假含量梯度扩大到0.1%~60%,增加了样本数据量,模型的准确率和适应性有所提高。此外,相较于高光谱技术,近红外光谱仪器价格更低,更适合研制便携式检测仪器,本研究为开发低成本、高效率的便携式近红外光谱检测仪提供了理论基础。

为了使模型具有更好的稳健性和适用范围,后续研究可以选择更多品牌且不同生产批次的蛋粉,选择更多种类的掺假物,扩充样本的数量,提高模型的适应性;在保证准确率的前提下还可以改进光谱特征的筛选方法,进一步减少光谱特征变量个数,简化模型,提高模型的准确率和检测效率。由于全蛋粉掺假成分多元复杂,针对某一种或几种的掺杂物建立的检测方法不可能有效地覆盖所有可能的掺杂物,如何在不影响掺假判别准确率的前提下利用数据处理技术从复杂、变动的全蛋粉数据中准确提取其真实性本征信息,筛选出有效反映全蛋粉真实性和掺假物质之间差异的特征变量,建立非定向掺假鉴别模型,是之后的一个重要探索方向。

参考文献 References

- [1] 王家鏊,林军,曲可欣,等.喷雾干燥与冷冻干燥鸡蛋粉特性研究[J].食品科技,2020,45(3):80-86.WANG J B, LIN J, QU K X, et al. Comparison of microstructure and physicochemical properties of spray drying and freeze-dried egg powder[J]. Food science and technology, 2020, 45(3): 80-86 (in Chinese with English abstract).
- [2] 马爽.蛋粉冲调特性及复合型功能蛋粉的研究与开发[D].长春:吉林大学,2012.MA S. Research and development of blending characteristics of egg powder and compound functional egg powder[D]. Changchun: Jilin University, 2012 (in Chinese with English abstract).
- [3] 迟玉杰,赵英,鲍志杰,等.高附加值专用型蛋粉的开发现状与展望[J].中国家禽,2014,36(19):2-4. CHI Y J, ZHAO Y, BAO Z J, et al. Development status and prospect of special egg powder with high added value[J]. China poultry, 2014, 36(19): 2-4 (in Chinese).
- [4] 韩帅娟,林辉松,杨丰帆,等.鸡蛋粉在猪生产中的应用研究进展[J].中国畜牧杂志,2019,55(5):20-24. HAN S J, LIN H S, YANG F F, et al. Research progress on egg powder in swine production[J]. Chinese journal of animal science, 2019, 55(5): 20-24 (in Chinese with English abstract).
- [5] 朱冠楠.市场拓展与政府缺位:全球贸易体系中的中国近代蛋粉业[J].清华大学学报(哲学社会科学版),2020,35(1):189-200. ZHU G N. Market expansion and government absence: China's modern egg powder industry in the global trade system[J]. Journal of Tsinghua University (philosophy and social sciences), 2020, 35(1): 189-200 (in Chinese with English abstract).
- [6] 陈佳,王爽,周巍,等. DNA 条形码技术在淀粉掺假鉴别中的应用[J].食品科学,2019,40(16):281-285. CHEN J, WANG S, ZHOU W, et al. Application of DNA barcoding in detection of starch adulteration[J]. Food science, 2019, 40(16): 281-285 (in Chinese with English abstract).
- [7] 刘平,马美湖.基于高光谱技术检测全蛋粉掺假的研究[J].光谱学与光谱分析,2018,38(1):246-252. LIU P, MA M H. Application of hyperspectral technology for detecting adulterated whole egg powder[J]. Spectroscopy and spectral analysis, 2018, 38(1): 246-252 (in Chinese with English abstract).
- [8] CAPUANO E, BOERRIGTER-EENLING R, KOOT A, et al. Targeted and untargeted detection of skim milk powder adulteration by near-infrared spectroscopy[J]. Food analytical methods, 2015, 8(8): 2125-2134.
- [9] KARUNATHILAKA S R, YAKES B J, HE K Q, et al. Non-targeted NIR spectroscopy and SIMCA classification for commercial milk powder authentication: a study using eleven potential adulterants[J/OL]. Heliyon, 2018, 4(9): e00806 [2022-12-09]. <https://doi.org/10.1016/j.heliyon.2018.e00806>.
- [10] 胡仁伟,俞玥,倪明龙,等.基于深度信念网络的近红外光谱鉴别莲子粉掺假[J].食品科学,2020,41(6):298-303. HU R W, YU Y, NI M L, et al. Identification of lotus seed flour adulteration based on near-infrared spectroscopy combined with deep belief network[J]. Food science, 2020, 41(6): 298-303 (in Chinese with English abstract).
- [11] MÜLLER-MAATSCH J, ALEWIJN M, WIJTEN M, et al. Detecting fraudulent additions in skimmed milk powder using a portable, hyphenated, optical multi-sensor approach in combination with one-class classification[J/OL]. Food control, 2021, 121: 107744 [2022-12-09]. <https://doi.org/10.1016/j.foodcont.2020.107744>.
- [12] 马爽,刘静波,王二雷.蛋粉加工及应用的研究现状分析[J].食品工业科技,2011,32(2):393-397. MA S, LIU J B, WANG E L. Existing condition analysis of processing and application of egg powder[J]. Science and technology of food industry, 2011, 32(2): 393-397 (in Chinese with English abstract).
- [13] 黄海松,魏建安,康佩栋.基于不平衡数据样本特性的新型过采样SVM分类算法[J].控制与决策,2018,33(9):1549-1558. HUANG H S, WEI J A, KANG P D. New over-sampling SVM classification algorithm based on unbalanced data sample characteristics[J]. Control and decision, 2018, 33(9): 1549-1558 (in Chinese with English abstract).
- [14] 袁培森,翟肇裕,任守纲,等.基于SVC和过采样的类别非均衡农业高光谱数据分类[J].农业机械学报,2019,50(6):257-264. YUAN P S, ZHAI Z Y, REN S G, et al. Classification of unbalanced agricultural hyperspectral data based on SVC and oversampling[J]. Transactions of the CSAM, 2019, 50(6): 257-264 (in Chinese with English abstract).
- [15] CHEN J Y, HUANG H W, COHN A G, et al. Machine learning-based classification of rock discontinuity trace: smote oversampling integrated with GBT ensemble learning[J]. International journal of mining science and technology, 2022, 32(2): 309-322.
- [16] XU X L, CHEN W, SUN Y F. Over-sampling algorithm for imbalanced data classification[J]. Journal of systems engineering and electronics, 2019, 30(6): 1182-1191.
- [17] NIU C, YUAN Y H, GUO H, et al. Recognition of osmotolerant yeast spoilage in kiwi juices by near-infrared spectroscopy coupled with chemometrics and wavelength selection[J]. RSC advances, 2018, 8(1): 222-229.
- [18] 刘燕德,黎丽莎,李斌,等.多品种苹果可溶性固形物近红外无

- 损检测通用模型研究[J]. 华中农业大学学报, 2022, 41(2): 237-244. LIU Y D, LI L S, LI B, et al. General near-infrared model of soluble solids content in multi-variety apples[J]. Journal of Huazhong Agricultural University, 2022, 41(2): 237-244 (in Chinese with English abstract).
- [19] 郝勇, 王起明, 张书敏. 可见-近红外光谱的鸭梨黑心缺陷在线检测 AdaBoost 集成模型研究[J]. 光谱学与光谱分析, 2021, 41(9): 2764-2769. HAO Y, WANG Q M, ZHANG S M. Study on online detection method of "Yali" pear black heart disease based on vis-near infrared spectroscopy and AdaBoost integrated model [J]. Spectroscopy and spectral analysis, 2021, 41(9): 2764-2769 (in Chinese with English abstract).
- [20] 蒋薇薇, 鲁昌华, 张玉钧, 等. 集成学习算法的红外光谱定量回归模型[J]. 光谱学与光谱分析, 2021, 41(4): 1119-1124. JIANG W W, LU C H, ZHANG Y J, et al. Research on a quantitative regression model of the infrared spectrum based on the integrated learning algorithm[J]. Spectroscopy and spectral analysis, 2021, 41(4): 1119-1124 (in Chinese with English abstract).
- [21] MENDOZA F, LU R F, ARIANA D, et al. Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content[J]. Postharvest biology and technology, 2011, 62(2): 149-160.
- [22] 李跑, 申汝佳, 李尚科, 等. 一种基于近红外光谱与化学计量学的绿茶快速无损鉴别方法[J]. 光谱学与光谱分析, 2019, 39(8): 2584-2589. LI P, SHEN R J, LI S K, et al. Nondestructive identification of green tea based on near infrared spectroscopy and chemometric methods [J]. Spectroscopy and spectral analysis, 2019, 39(8): 2584-2589 (in Chinese with English abstract).
- [23] CHEN J B, SUN S Q, YU J, et al. Tracking the curing process of automotive paint by moving-window two-dimensional infrared correlation spectroscopy and principal component analysis [J]. Journal of molecular structure, 2014, 1069: 112-117.

Detection of adulteration in whole egg powder based on near-infrared spectroscopy

YANG Kai¹, HE Yuting¹, LI Wolin¹, ZHU Zhihui^{1,2}, WANG Qiaohua^{1,2}

1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China;

2. Ministry of Agriculture and Rural Affairs Key Laboratory of Agricultural Equipment in Mid-Lower Yangtze River, Wuhan 430070, China

Abstract Near-infrared spectroscopy was used to qualitatively discriminate the adulteration in whole egg powder and to analyze the concentration of adulteration quantitatively. Six different pre-processing methods were used to pre-process the raw spectral data. The competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA) and combined CARS-SPA algorithms were used to filter the spectral data for characteristic wavelengths. Ensemble learning (EL) model was established to qualitatively discriminate the egg powder adulterated. The partial least squares regression (PLSR) model was established to quantitatively analyze the concentration of adulteration. The results showed that first order derivative was the best pre-processing method in qualitative discrimination. The CARS algorithm had the best effect of feature filtering. The EL model had an overall discrimination accuracy of 98.18% for adulterated samples, with the discrimination accuracy of differently adulterated samples of more than 97.78%. MSC was the best pretreatment method in quantitative analysis. CARS algorithm had the best effect of feature filtering. The prediction set correlation coefficient of PLSR adulteration concentration prediction models for one-component adulteration, two-component adulteration, three-component adulteration, and all adulterated samples adulteration was 0.958 5, 0.931 2, 0.945 6, and 0.955 8, respectively, with the root mean square error RMSEP of 4.689 1, 5.813 4, 4.604 1, and 3.802 9, respectively. It is indicated that the near-infrared spectroscopy can be used for detecting the adulteration in whole egg powder. It will provide a reference for regulatory agencies to detect adulteration in whole egg powder.

Keywords whole egg powder; near-infrared spectroscopy; adulteration detection; ensemble learning; accurate detection

(责任编辑:陆文昌)