

DOI: 10.13652/j.issn.1003-5788.2020.10.011

# 基于近红外光谱的杏仁蛋白软糖理化指标快速无损检测

Rapid and nondestructive detection of physicochemical properties of marzipan based on near infrared spectroscopy

陈素彬

CHEN Su-bin

(南充职业技术学院, 四川 南充 637131)

(Nanchong Vocational and Technical College, Nanchong, Sichuan 637131, China)

**摘要:** 为了实现杏仁蛋白软糖理化指标的快速检测,建立了水分和总糖测定的近红外光谱定标模型,并对其进行了全面优化。以偏最小二乘法分别建立杏仁蛋白软糖水分、总糖测定的近红外光谱定标模型,然后依次运用异常样本剔除、光谱变换、样本集划分和特征波长选择等技术方法优化其性能。结果表明:水分测定模型的建模波点数减至全光谱的0.8%, $R_c^2$ 和 $R_p^2$ 均大于0.99, RMSEC和RMSEP均在0.2左右,RPD大于15.3;总糖测定模型的建模波点数减至全光谱的1.5%, $R_c^2$ 和 $R_p^2$ 均大于0.99, RMSEC和RMSEP均小于0.81, RPD大于17.3。两个优化模型的预测结果皆与相应参考值的差异不显著( $P > 0.05$ ),可用于实际检测工作。

**关键词:** 近红外光谱;快速测定;理化指标;偏最小二乘法;模型优化

**Abstract:** A calibration model of near-infrared spectroscopy for the determination of water and total sugar is established and optimized in order to realize the rapid detection of physical and chemical indexes of marzipan. The partial least square method is used to establish NIR calibration models for the determination of water and total sugar in marzipan. Then, the abnormal sample removal, spectral transformation, sample set partition and characteristic wavelength selection are used to optimize its performance. The results show that the number of modeling wave points of the moisture determination model is reduced to 0.8% of the total spectrum,  $R_c^2$  and  $R_p^2$  are greater than 0.99, with RMSEC and

RMSEP about 0.2 and RPD greater than 15.3. The number of modeling wave points of total sugar determination model is reduced to 1.5% of the total spectrum, and  $R_c^2$  and  $R_p^2$  are greater than 0.99, with RMSEC and RMSEP less than 0.81, and RPD greater than 17.3. There is no significant difference between the prediction results of the two optimized models and the corresponding reference values ( $P > 0.05$ ), which can be used in the actual detection work.

**Keywords:** near infrared spectroscopy; rapid determination; physicochemical properties; partial least square; model optimization

杏仁蛋白软糖是重要的食品 and 食品生产原料,可用作蛋糕内馅、糕面及富有色彩的糕点装饰,制作小杏仁蛋糕、杏仁饼及饼干。杏仁蛋白软糖富含蛋白质、单不饱和脂肪酸、维生素E、纤维和镁等物质,不仅能改善食品的风味、口感和外观,而且还具有很高的医用价值。

根据 GB/T 20977—2007《糕点通则》之规定,杏仁蛋白软糖的理化指标应包括干燥失重(水分)、蛋白质、粗脂肪和总糖,其中总糖含量检验方法为其附录A给出的斐林氏容量法;水分含量按 GB 5009.3—2016《食品安全国家标准 食品中水分的测定》的第一法测定。这些方法的优势为准确度高、重复性好、设备成本较低,适用于仲裁检验,但需用仪器设备和化学试剂较多,具有操作复杂、耗时长、污染环境等缺点<sup>[1]</sup>,致其测定结果易受人因素影响,不利于食品生产中的大样本快速检测。近年来,随着化学计量学方法的发展和仪器研制水平的提高,近红外光谱(Near infrared spectroscopy, NIRS)分析方法检测高效、操作简便、无化学污染、支持多成分同时测定等优势日益凸显,作为一种非破坏性的“瞬间分析”技术<sup>[2]</sup>,被越来越多地应用于食品行业的各种定性和定量分析,但目前中国

**基金项目:**南充市2019年应用技术与开发资金项目(编号:19YFZJ0028);南充职业技术学院2019年重点课题(编号:ZRA1905)

**作者简介:**陈素彬(1967—),女,南充职业技术学院副教授。

E-mail: 442060379@qq.com

**收稿日期:**2020-04-18

尚未见以 NIRS 法检测杏仁蛋白软糖组分的报道。

NIRS 法是一种间接定量分析技术<sup>[3]</sup>,在确保样品光谱和待测成分参考值准确性的情况下,其预测结果取决于所建定标模型的质量。在实际应用和相关研究中,通常采用随近红外光谱仪器附送的 OPUS、TQ Analyst、QUANT 等分析软件<sup>[4-5]</sup>,以偏最小二乘法(Partial least squares, PLS)进行建模和预测。对于食品营养成分、品质指标检测,李路等<sup>[6]</sup>用 PLS 法和 BP 神经网络法建立了大米总糖、水分等成分的近红外检测模型,其决定系数( $R^2$ )均大于 0.9、相对标准差均小于 2.6%。陈冲等<sup>[7]</sup>建立了冬枣水分、糖分无损检测的 PLS 和主成分回归(Principal component regression, PCR)分析模型,其水分预测 PLS 模型的相关系数( $R$ )为 0.997 45、校正均方差(Root mean square error of calibration, RMSEC)为 0.044 5、预测均方差(Root mean square error of prediction, RMSEP)为 0.367,糖分预测 PLS 模型的相关系数( $R$ )、RMSEC 和 RMSEP 分别为 0.960 78, 0.853, 1.64。何云啸等<sup>[8]</sup>建立了诺丽果粉在真空冷冻干燥过程中水分含量与红外光谱的定量模型,当水分含量在 5.00%~9.00%时模型的  $R^2$  为 0.98、交互验证均方根误差(Root mean square error of cross validation, RMSECV)为 0.26。贾柳君等<sup>[9]</sup>利用近红外光谱分析技术对葡萄酒的主要成分进行定量分析,其总糖预测模型的  $R^2$ 、RMSEP 和相对分析误差(Residual predictive deviation, RPD)分别达到 0.943 5, 0.263 6, 4.21。这些研究为应用 NIRS 分析技术进行食品营养指标检测提供了有益参考,但未充分运用化学计量学方法和软件技术优化模型性能,主要体现在:① 忽略了异常样本和不同的样本集划分结果对模型预测能力和稳健性的影响,未进行异常样本判断和样本集划分方法选择。② 根据谱图外观或待测成分对应的化学基团选择变量区间,建模波长较多,模型运算量大而运行效率较低;或用组合方法优选特征波长,增加了建模过程的复杂性。

鉴于目前对各类样品蛋白质和粗脂肪测定的研究较多,相应的 NIRS 法定量分析模型并不鲜见,试验拟以杏仁蛋白软糖为研究对象,以水分和总糖含量测定为代表,采用 PLS 法建立其理化指标检测的 NIRS 快速定量模型,并运用化学计量学方法和软件技术全面优化其性能,包括蒙特卡洛交互验证(Monte Carlo cross validation, MCCV)法剔除异常样本、各种算法光谱预处理、KS(Kenard-Stone)法和 X-Y 共生距离(Sample set partitioning based on joint x-y distance, SPXY)法划分样本集和竞争性自适应重加权采样(Competitive adaptive reweighted sampling, CARS)法选取特征波长等,以提高模型的预测准确性和运行速度,由此形成完整、高效的优化建模方案,为杏仁蛋白软糖理化指标的快速检测提供新的思路。

## 1 材料与方 法

### 1.1 材料与仪器

杏仁蛋白软糖:南充多家糕点企业;

近红外光谱仪:NIRSystems 6500 型,丹麦 FOSS 公司。其光源为钨卤灯,波长范围 400~2 500 nm,检测器为硅(400~1 100 nm)和硫化铅(1 100~2 500 nm);取点间隔 2 nm,扫描速度 1.8 次/s,扫描次数 32,工作温度 15~32 ℃。配有往复移动式样品池,以标准陶瓷片为参比。

### 1.2 样品采集与制备

按 GB/T 20977—2007 规定之方法和数量,在南充糕点企业采集杏仁蛋白软糖样品 32 个,用四分法缩减、混匀后分成 4 份,取其中 2 份为试验样品,分别用于水分、总糖含量参考值测定和光谱数据采集。试样装入广口瓶,保存在冰箱中。

### 1.3 待测成分含量参考值测定

1.3.1 水分 称取试样 8 g(精确至 0.000 1 g),按 GB 5009.3—2016 的第一法测定。

1.3.2 总糖 准确称取样品 2.0 g,按 GB/T 20977—2007 附录 A 的斐林氏容量法测定。

### 1.4 NIRS 定标模型的建立

1.4.1 样品近红外光谱采集 依次将试样装入样品杯,以近红外光谱仪扫描。设置波长范围 450~2 448 nm、间隔 2 nm,每个试样重复装样、扫描 10 次,取其平均值为样品光谱。

1.4.2 NIRS 定标模型建立及其性能评价 以 KS 法将样品按 2:1 划分为定标集和验证集。用定标集光谱数据和水分含量参考值建立 PLS 定标模型,并将验证集光谱数据代入该模型预测相应的样品水分含量;用同样方法建模、预测验证集样品的总糖含量。

以校正决定系数( $Rc^2$ )、预测决定系数( $Rp^2$ )、RMSEC、RMSEP 和 RPD 为评价指标,分析所得水分和总糖测定的 NIRS 定标模型性能。

### 1.5 NIRS 定标模型的优化

1.5.1 异常样本剔除 以蒙特卡洛采样法(Monte Carlo sampling, MCS)从全部样本中抽取 80%为校正集、其余为验证集,建立 PLS 模型进行预测,计算各样本的预测残差。重复该过程使全部样本被预测,计算各样本预测残差的均值(MEAN)和标准差(Standard deviation, STD),绘制全部样本预测残差的 MEAN-STD 分布图,图中位于均值和标准差较高区域的样本即为异常样本<sup>[10]</sup>。

1.5.2 光谱变换处理与样本集划分 以多种算法分别对定标集和验证集光谱数据进行变换处理,包括均值中心化(Mean centralization, MC)、萨维茨基—戈莱平滑(Savitzky-Golay smoothing, SG)、标准正态变量变换

(Standard normal variable, SNV)、去趋势(De-trending, DT)和多元散射校正(Multiple scattering correction, MSC)以及基于MC的组合算法;同时分别以KS法和SPXY法按2:1划分样本集,用变换处理后的光谱数据与相应待测成分参考值分别建立PLS模型预测水分和总糖含量,比较其性能指标以确定最优光谱预处理和样本集划分方案。

1.5.3 特征波长选取 用MCS法抽取80%样本构成校正集、其余为验证集,建立PLS模型进行预测,计算波长 $j$ 对目标的贡献 $|b_j|$ 和权重 $w_j$ ,以指数递减函数计算波长点的保留率 $r_i = ae^{-ki}$ ( $a, k$ 为常数, $i$ 为采样次数);去掉 $|b_j|$ 较小的波长点,用自适应重加权采样(Adaptive re-weighted sampling, ARS)法由 $m \times r_i$ 个波长点中得到样本优选子集( $m$ 为波长点数),取相应光谱数据建立PLS模型,计算其RMSECV。重复此过程200次,取最小RMSECV值对应的优选子集为特征波长。

1.5.4 优化模型建立与预测 从样本集中剔除异常样本,用选定方案进行数据预处理、划分样本集、选取特征波长光谱,得到优化的定标集和验证集,以PLS方法分别建模、预测水分和总糖含量。将所得预测结果与相应成分的参考值对比,计算其平均回收率,并通过配对样本 $t$ 检验判断其差异性。

## 1.6 NIRS定标模型优化性能对比

以 $R_c^2$ 、RMSEC、 $R_p^2$ 、RMSEP和RPD为指标,评估和对比NIRS定标模型优化前后的性能。 $R_c^2$ 越接近1、RMSEC越接近0,则建模质量越好; $R_p^2$ 越接近1、RMSEP越接近0,则模型预测能力越强; $R_c^2$ 和 $R_p^2$ 值大且相差少、RMSEC和RMSEP值小且相差少,则模型精度高、稳健性好。根据国际谷物科技协会标准,NIRS定标模型的 $RPD \geq 3$ 方能用于实际检测。

## 1.7 软件工具

1.7.1 待测成分参考值测定 数据记录、结果计算和数据统计用Microsoft Excel 2010完成。

1.7.2 NIRS定标模型建立和优化 初始模型的建立、预测用The Unscrambler X 10.4完成,模型优化用MATLAB R2019a编程、作图实现。

1.7.3 样品光谱图形绘制 用OriginPro 2019b完成。

## 2 结果与分析

### 2.1 NIRS定标模型建立和预测

2.1.1 样品近红外光谱采集结果 32个样品的近红外光谱如图1所示。

从图1可知,所有杏仁蛋白软糖样品的光谱特征较为相似,谱线较分散、谱图较宽,表明各样品组成基本一致、其成分含量差异较明显;样本光谱在1100,1288,1652,1848 nm左右存在明显波谷,而在1010,1208,1468,1724,1938 nm左右有较明显的波峰,这些波段区

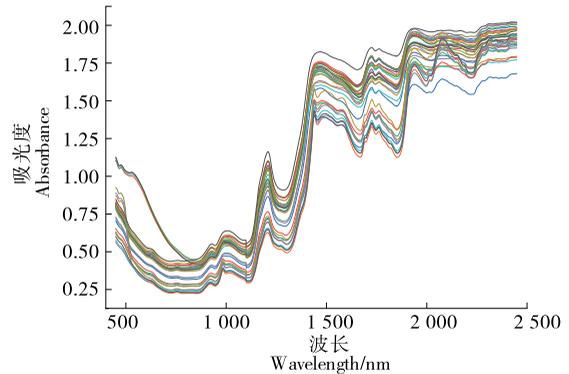


图1 样品近红外光谱

Figure 1 Near infrared spectrum of samples

域含有较多特征光谱信息。同时还可看到,样品光谱谱带较宽,基线漂移和谱带重叠严重,尤其2150~2448 nm波段的光谱信噪比较低,故建模时应应对原始光谱进行变换处理;谱图中有2条谱线的走势明显离群,疑其为异常样本,宜在建模时剔除。

2.1.2 NIRS定标模型预测结果分析 用全部样本、全光谱、无预处理分别所建水分、总糖含量测定的NIRS定标模型预测结果如表1所示。

由表1可知,由于杏仁蛋白软糖样品的原始光谱中存在大量冗余无用信息、共线性变量和异常样本数据,在未采用化学计量学方法和软件技术进行优化时,其水分和总糖含量测定的NIRS定标模型性能都不高。水分测定模型的 $R_c^2$ 和 $R_p^2$ 均小于0.91、RMSEC和RMSEP均大于1.14,且这两对指标值均相差较多,RPD则小于3;总糖测定模型的 $R_c^2$ 和 $R_p^2$ 均小于0.7, RMSEC和RMSEP均大于6.97,且RPD小于2。因此,两个定标模型的测定准确度都较低,预测值与相应参考值的相关性不高;虽然水分测定模型的性能指标优于总糖测定模型,但二者皆未达到可用于实际检测的标准,尚需进行优化。

### 2.2 NIRS定标模型的优化

2.2.1 异常样本剔除结果分析 马氏距离法、主成分得分法、光谱残差检验法等传统方法判断异常样本,皆需根据经验设定阈值,且难以判断多个异常点。试验采用MCCV法,可根据光谱数据和待测成分之间的关系及异常值对模型稳健性的影响,同时筛选出两个方向的异常点,并以全部样本的预测残差MEAN-STD分布图直观地反映其整体分布情况,其结果如图2所示。

表1 NIRS定标模型预测结果

Table 1 Prediction results of NIRS calibration model						
定标模型	PCs	$R_c^2$	RMSEC	$R_p^2$	RMSEP	RPD
水分测定	2	0.903 9	1.144 5	0.855 9	1.347 7	2.85
总糖测定	2	0.693 7	6.975 6	0.693 0	7.465 8	1.82

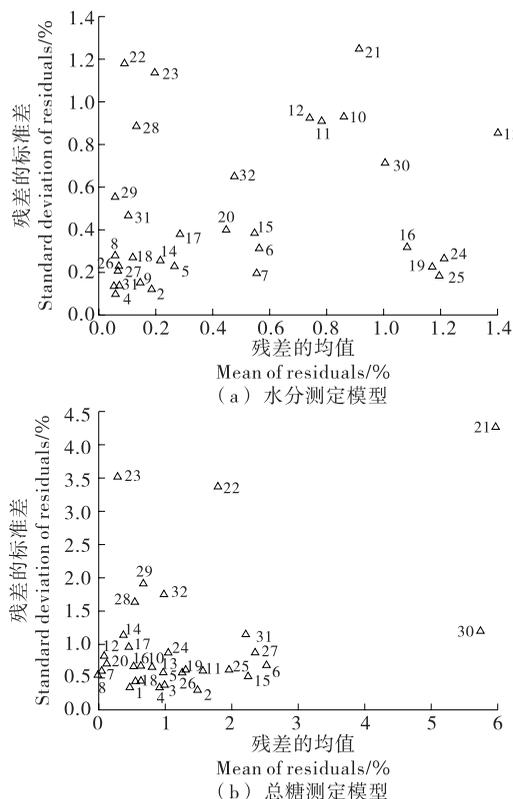


图 2 全部样本预测残差 MEAN-STD 分布图

Figure 2 Mean-STD distribution of predicted residuals of all samples

从图 2(a)可看出,全部样本预测残差的均值和标准差分布较分散,其中位于高均值区域的样本依次为 13、24、25 和 19 号,位于高标准差区域的样本依次为 21、22 和 23 号,反映出在模型的多次运行中,这些样本的水分含量预测准确度或(和)稳定性与总体相差较大,剔除之将使所建模型的准确性和稳健性得到改善。因样本总数较少,实际建模时仅剔除预测残差均值和标准差都较大的 13 和 21 号样本即可;在图 2(b)中,样本预测残差的均值和标准差分布较集中,位于高均值区域的样本为 21 和 30 号,位于高标准差区域的样本依次为 21、23 和 22 号,因此建模时宜剔除离群最远的 21 和 30 号样本。

2.2.2 光谱变换处理与样本集划分

(1) 样本集划分结果比较:在浓度梯度(Concentration gradient, CG)、随机(Random sampling, RS)、KS 和 SPXY 等几种常用样本集划分方法中,KS 法在实际检测和相关研究中采用最多;SPXY 法则在计算样品间距时同时考虑了光谱和参考值变量,能有效覆盖多维向量空间,从而改善所建模型的预测能力。试验将 30 个正常样本分别用 KS 法和 SPXY 法按 2 : 1 划分样本集,得定标集样本 20 个、验证集样本 10 个,其结果如表 2 所示。

表 2 两种方法划分样本集的结果

Table 2 The result of two methods to divide sample set

模型	方法	样本集	最小值	最大值	平均值	标准差
水分	KS	定标集	6.8	18.6	13.52	3.62
		验证集	7.0	18.1	13.92	3.64
	SPXY	定标集	6.8	18.6	14.07	3.45
		验证集	7.0	16.4	12.82	3.84
总糖	KS	定标集	33.2	68.7	45.49	12.15
		验证集	32.7	68.2	44.20	13.47
	SPXY	定标集	32.7	68.7	46.43	11.63
		验证集	33.2	68.2	42.33	14.00

由表 2 可知,用 KS 法划分水分测定模型的样本集较好,所得定标集与验证集样本的水分含量区间分别为 [6.8, 18.6]、[7.0, 18.1],前者完全涵盖了后者;其水分含量平均值(13.52, 13.92)和标准差(3.62, 3.64)都相差很小,表明定标集和验证集样本数据分布相似度高,适合于建立水分含量预测模型。

总糖测定模型则以 SPXY 法划分样本集较好,相应定标集和验证集样本的总糖含量区间分别为 [32.7, 68.7]、[33.2, 68.2],前者亦完全涵盖后者;其验证集样本的总糖含量平均值小于定标集、标准差大于定标集,表明验证集样本数据分布于定标集数据区域内,且其较分散、多态性好,有利于建立一个预测能力强、泛化性能好的总糖测定模型。与之相比,KS 法划分的定标集中未包含总糖含量值最小的样本,将对所建 NIRS 定标模型性能产生不利影响。

(2) 光谱变换处理结果比较:近红外光谱吸收较弱、重叠严重,且存在样品物理性状不均、光的散射及仪器噪声等不利因素影响,致使扫描所得原始光谱信息复杂、信噪比低,以之直接建模难以获得预期的结果。试验分别尝试了多种光谱变换处理算法,其相应 PLS 定标模型的预测结果如表 3 所示。

由表 3 可知,样本光谱经过各种变换处理后,所建水分和总糖定标模型的性能皆有所提高。在多种预处理方法中,组合方法的效果优于单一方法;两个模型都以“MSC+MC”变换处理所得结果最佳。相应水分预测模型的  $R^2$  和  $R_p^2$  均大于 0.99, RMSEC 和 RMSEP 分别为 0.218 9、0.326 4, RPD 为 11.16;总糖预测模型的  $R^2$  和  $R_p^2$  亦大于 0.99, RMSEC 和 RMSEP 分别为 0.762 6, 1.203 2, RPD 为 11.64。表明样品光谱用“MSC+MC”组合方法预处理后,所建水分和总糖定标模型预测准确、运行稳定,预测值与相应参考值的相关性高,其性能已经达到可用于实际检测的标准。

从方法原理的角度讲, MSC 算法能有效消除样品颗

粒大小及其分布不均产生的散射影响,增强光谱与数据之间的相关性;MC变换可增加样品光谱之间的差异,从而提高模型的稳健性和预测能力。就实际效果看,试验将二者组合应用于半固体状的杏仁蛋白软糖样品光谱预处理,取得了优于其他各种方法的效果。

2.2.3 建模波长选取结果分析 试验样品的原始NIRS光谱包括1000个波长,其中存在大量冗余信息和共线性变量,势必增加建模的复杂性和计算量,影响模型的预测精度和运行速度,因此建模时需选取能充分反映待测成分信息的特征波长,以简化模型、提高其性能。在无信息变量消除法(UVE)、连续投影算法(SPA)及遗传算法(GA)等众多方法中,试验采用CARS法进行建模特征波长优选<sup>[11]</sup>,所得结果如表4所示。

由表4可知,用CARS法结合PLS模型筛选后,水分定标模型保留了8个波长变量,仅占全波长的0.8%。其中4个集中在1590~1692 nm波段,正好位于以1652 nm为中心的波谷内;另外4个中有2个位于1938 nm处的波峰区域、1个位于1468 nm处的波峰附

近、1个位于1848 nm处的波谷区域。显然,选出的8个波长与样品谱图的特征基本相符,相应的光谱数据能够表达杏仁蛋白软糖样品中水分的主要信息。

总糖定标模型选取15个波长变量,仅为全波长的1.5%。其中7个集中在1550~1566 nm波段,另外8个中有3个位于1848 nm处的波谷区域、2个位于1208 nm处的波峰区域、2个位于2094~2098 nm波段、1个位于1010 nm处的波峰区域。对照图1可知,所选波长全部位于信噪比较高、特征变化显著的谱区,即第一个波峰(1010 nm)和最后一个波峰(1938 nm)范围,相应的样品光谱数据所含总糖信息足以满足建模要求。

2.2.4 优化模型建立与预测结果分析 将前述各项优化成果综合应用,以PLS法所建样品水分和总糖含量的NIRS定标模型预测结果如表5所示。

由表5可知,得益于多项优化技术的综合应用,样品水分含量定标模型的预测残差为-0.4%~0.5%,其绝对值均小于0.5%;总糖含量定标模型的预测残差为-0.9%~1.0%,其绝对值均不大于1.0%。将预测值与

表3 各种预处理方法的建模预测结果

Table 3 Modeling and prediction results of various preprocessing methods

模型	预处理	PCs	$R^2$	RMSEC	$R^2_p$	RMSEP	RPD
水分	MC	5	0.967 9	0.632 1	0.961 8	0.692 5	5.26
	SG(2,7)	2	0.847 2	1.381 2	0.904 6	1.102 7	3.30
	SNV	2	0.823 7	1.482 6	0.864 0	1.283 2	2.84
	DT	2	0.874 6	1.254 6	0.914 4	1.127 6	3.23
	MSC	2	0.822 8	1.486 1	0.867 4	1.271 2	2.87
	SG+MC	5	0.967 8	0.633 1	0.961 6	0.693 5	5.25
	SNV+MC	7	0.996 4	0.211 7	0.991 7	0.332 3	10.96
	DT+MC	4	0.983 0	0.460 4	0.986 9	0.408 4	8.92
	MSC+MC	7	0.996 2	0.218 9	0.991 8	0.326 4	11.16
总糖	MC	5	0.985 7	1.355 2	0.976 2	2.135 1	6.56
	SG(2,7)	2	0.607 8	7.174 6	0.750 4	9.140 0	1.53
	SNV	2	0.833 8	4.622 5	0.976 4	4.592 1	3.05
	DT	2	0.831 3	4.709 9	0.930 1	5.162 4	2.71
	MSC	2	0.834 2	4.617 5	0.978 7	4.558 7	3.07
	SG+MC	5	0.985 7	1.355 3	0.976 2	2.136 0	6.56
	SNV+MC	4	0.993 9	0.882 0	0.988 3	1.520 7	9.21
	DT+MC	4	0.991 1	1.070 2	0.986 1	1.656 2	8.46
	MSC+MC	4	0.995 5	0.762 6	0.992 6	1.203 2	11.64

表4 NIRS定标模型特征波长选择结果

Table 4 Results of characteristic wavelength selection of NIRS calibration model

模型	波点数	特征波长/nm
水分	8	1 408,1 590,1 686,1 688,1 692,1 868,1 906,1 908
总糖	15	986,1 210,1 212,1 550,1 552,1 554,1 556,1 558,1 562,1 566,1 868,1 874,1 880,2 094,2 098

表 5 优化的 NIRS 定标模型预测结果  
Table 5 Prediction results of optimized NIRS calibration model

样号	水分含量/%			总糖含量/%		
	参考值	预测值	差值	参考值	预测值	差值
1	15.0	15.0	0.0	34.9	34.2	-0.7
2	14.7	14.8	0.1	33.9	33.3	-0.6
3	14.9	15.1	0.2	33.2	33.8	0.6
4	14.6	14.5	-0.1	33.5	33.5	0.0
5	18.1	17.7	-0.4	33.2	33.1	-0.1
6	15.6	15.4	-0.2	44.3	44.8	0.5
7	16.4	16.6	0.2	40.2	39.8	-0.4
8	7.0	6.7	-0.3	67.8	67.8	0.0
9	15.3	15.8	0.5	34.1	35.1	1.0
10	7.6	7.4	-0.2	68.2	67.3	-0.9

表 6 NIRS 定标模型优化性能对比

Table 6 Comparison of the optimized performance of NIRS calibration model

模型	样本数	样本集划分	预处理	波点数	$R_c^2$	RMSEC	$R_p^2$	RMSEP	RPD
水分	32	KS	None	1 000	0.903 9	1.144 5	0.855 9	1.347 7	2.85
	30	KS	MSC+MC	1 000	0.996 2	0.218 9	0.991 8	0.326 4	11.16
	30	KS	MSC+MC	8	0.997 1	0.190 2	0.995 3	0.237 0	15.37
总糖	32	KS	None	1 000	0.693 7	6.975 6	0.693 0	7.465 8	1.82
	30	SPXY	MSC+MC	1 000	0.995 5	0.762 6	0.992 6	1.203 2	11.64
	30	SPXY	MSC+MC	15	0.996 6	0.661 3	0.996 3	0.805 9	17.38

0.903 9, 0.855 9 增至 0.996 2, 0.991 8, RMSEC 和 RMSEP 分别由 1.144 5, 1.347 7 降为 0.218 9, 0.326 4, RPD 由 2.85 提高到 11.16; 在此基础上选取 8 个特征波长后, 相应模型的  $R_c^2$  和  $R_p^2$  随之升到 0.997 1, 0.995 3, RMSEC 和 RMSEP 减为 0.190 2, 0.237 0, RPD 增至 15.37, 表明该模型的预测能力和稳健性得到了较大提高。

总糖定标模型在剔除 2 个异常样本、执行“MSC+MC”预处理、改用 SPXY 方法划分样本集后,  $R_c^2$  和  $R_p^2$  分别由 0.693 7, 0.693 0 增至 0.995 5, 0.992 6, RMSEC 和 RMSEP 分别由 6.975 6, 7.465 8 降为 0.762 6, 1.203 2, RPD 由 1.82 提高到 11.64; 进一步选取 15 个特征波长后, 相应模型的  $R_c^2$  和  $R_p^2$  随之升到 0.996 6, 0.996 3, RMSEC 和 RMSEP 减为 0.661 3, 0.805 9, RPD 增至 17.38, 显然该模型的性能得到了显著改善。

同时, 由于建模波长变量大幅度缩减, 大量非待测成分信息得以消除, 使得优化模型的计算量显著降低, 其运行速度也随之加快。在采用相同方法进行同类检测时, 参照试验所选特征波长采集样品光谱, 可极大节省其工作量。

### 3 结论

利用近红外光谱技术快速检测杏仁蛋白软糖理化指

参考值之比作为预测回收率, 算得两个模型的平均回收率分别为 99.5%, 99.9%, 表明其可用于预测杏仁蛋白软糖的相应成分含量。

分别将水分、总糖定标模型的预测值与参考值进行配对样本  $t$  检验, 在设置置信水平 95% 的情况下, 结果分别为  $P=0.83>0.05$ ,  $P=0.84>0.05$ , 说明 95% 概率下两组数据无显著性差异, 两个模型可分别满足杏仁蛋白软糖水分、总糖含量快速准确测定的要求。

### 2.3 NIRS 定标模型优化性能对比分析

从初始模型开始, 到运用多项技术方法完成模型优化, 各阶段分别所建杏仁蛋白软糖水分、总糖含量预测的 NIRS 定标模型性能对比如表 6 所示。

由表 6 可知, 两个模型的性能皆随优化技术的逐步引入而不断提高。当剔除 2 个异常样本、执行“MSC+MC”预处理后, 所建水分定标模型的  $R_c^2$  和  $R_p^2$  分别由

标, 以水分和总糖含量测定为例, 利用 MCCV 法剔除异常样本, 尝试多种算法进行光谱预处理, 选用 KS 法和 SPXY 法划分样本集, 采用 CARS 法选取特征波长变量, 并以 PLS 法建立水分和总糖含量预测模型。结果表明: 剔除异常样本后, 以“MSC+MC”组合方法预处理后所建模型最优; 水分定标模型以 KS 法划分样本集较好, 总糖定标模型则 SPXY 法划分样本集为佳; 两个模型用 CARS 法所建建模波长变量数仅分别为全波长的 0.8%, 1.5%。优化后的水分定标模型  $R_p^2$ 、RMSEP 和 RPD 分别为 0.995 3, 0.237 0, 15.37, 总糖定标模型  $R_p^2$ 、RMSEP 和 RPD 分别为 0.996 3, 0.805 9, 17.38; 两个优化模型的平均回收率分别为 99.5%, 99.9%, 预测值与相应成分参考值的配对样本  $t$  检验结果均为  $P>0.05$ , 故试验所建模型的预测结果与对应标准<sup>[12-13]</sup>方法测定值无显著性差异。

由此可见, 试验所建杏仁蛋白软糖水分、总糖含量的 NIRS 定标模型均有较高的精度和运算速度, 其性能指标优于同类试验模型, 并达到国际谷物科技协会用于实际检测的标准, 适用于杏仁蛋白软糖理化指标的快速检测。但因时间所限, 试验采集样品数相对较少, 以及未尝试非线性建模和更多波长选择方法, 下一步将对此进行研究。

(下转第 64 页)

罪相关罪名明确规定适用从业禁止,并结合案件性质加强从业禁止的惩处力度。二是要适当延伸从业禁止的期限。在上述 11 起判决书中,判处从业禁止的期限均为 3 年,适用期限普遍较短,无法充分发挥特殊预防的作用。因此,可以考虑设置更长的从业禁止期限,对于社会危害性严重的犯罪人终身剥夺其从事食品行业的资格。通过提高定罪后续成本,可以有效地防止再犯,不仅减少了犯罪人的预期收益,更可以实现精确打击,封堵危害食品安全犯罪源头。

### 参考文献

- [1] 范雪珂. 危害食品安全罪: 法益与立法完善[J]. 政治与法律, 2019(6): 141-149.
- [2] 宋浩波. 犯罪经济学[M]. 北京: 中国人民公安大学出版社, 2002: 1.
- [3] 贝克尔. 人类行为的经济分析[M]. 王业宇, 陈琪, 译. 上海: 上海人民出版社, 2015: 19.
- [4] 张维迎. 博弈论与信息经济学[M]. 上海: 上海人民出版社, 2012: 17.
- [5] 于杨曜. 非法添加类食品安全犯罪刑事规制体系及完善: 以“两高”《关于办理危害食品安全刑事案件适用法律若干问题的解释》为切入点[J]. 政治与法律, 2013(10): 54-63.
- [6] 斯密. 国富论[M]. 唐日松, 译. 北京: 华夏出版社, 2006: 25.
- [7] 王红雨生产、销售不符合安全标准的食品罪一案一审刑事判决书[EB/OL]. (2019-10-24)[2020-06-27]. [https://www.itslaw.com/detail?initialization=%7B%22category%22%3A%22CASE%22%2C%22id%22%3A%22758420ac-f3b7-4f0c-8272-08479b10c2a2%22%2C%22anchor%22%3Anull%2C%22detailKeyWords%22%3A%5B%22%22%5D%7D#content\\_null](https://www.itslaw.com/detail?initialization=%7B%22category%22%3A%22CASE%22%2C%22id%22%3A%22758420ac-f3b7-4f0c-8272-08479b10c2a2%22%2C%22anchor%22%3Anull%2C%22detailKeyWords%22%3A%5B%22%22%5D%7D#content_null).

- [8] 曝商贩做卤牛肉加“保水王”过量使用会影响健康[EB/OL]. (2020-05-19) [2020-06-28]. [https://m.sohu.com/n/413309573/?pvid=000115\\_3w](https://m.sohu.com/n/413309573/?pvid=000115_3w).
- [9] 沈海平. 寻求有效率的惩罚[D]. 北京: 北京大学, 2008: 89.
- [10] 李春雷. 食药安全行政执法与刑事司法[M]. 北京: 研究出版社, 2019: 15.
- [11] 陈涛, 潘宇. 食品安全犯罪现状与治理[J]. 中国人民公安大学学报(社会科学版), 2015, 31(4): 21-27.
- [12] 郝某某生产、销售不符合安全标准的食品罪一审刑事判决书[EB/OL]. (2019-06-21)[2020-06-27]. [https://www.itslaw.com/detail?initialization=%7B%22category%22%3A%22CASE%22%2C%22id%22%3A%22d5e5c7dc-27f3-479b-b408-6e242d868e80%22%2C%22anchor%22%3Anull%2C%22detailKeyWords%22%3A%5B%22%22%5D%7D#content\\_null](https://www.itslaw.com/detail?initialization=%7B%22category%22%3A%22CASE%22%2C%22id%22%3A%22d5e5c7dc-27f3-479b-b408-6e242d868e80%22%2C%22anchor%22%3Anull%2C%22detailKeyWords%22%3A%5B%22%22%5D%7D#content_null).
- [13] 边沁. 道德与立法原理导论[M]. 时殷弘, 译. 北京: 商务印书馆, 2000: 59.
- [14] 梁亮. 基于经济分析的食品安全监管模式的优化[J]. 食品与机械, 2019, 35(2): 103-106.
- [15] 李春雷, 任韧. 我国食品药品犯罪防治回顾与前瞻[J]. 中国人民公安大学学报(社会科学版), 2015, 31(4): 1-14.
- [16] 贝卡里亚. 论犯罪与刑罚[M]. 黄风, 译. 北京: 中国大百科全书出版社, 1993.
- [17] 白建军. 裸刑均值的意义[J]. 法学研究, 2010, 32(6): 135-142.
- [18] 吉丽颖. 法经济学视角下中国食品安全事件的规制维度[J]. 食品与机械, 2018, 34(12): 63-66.

(上接第 58 页)

### 参考文献

- [1] 吕程序, 姜训鹏, 张银桥, 等. 基于变量选择的小麦粗蛋白含量近红外光谱检测[J]. 农业机械学报, 2016, 47(10): 340-346.
- [2] 雷加容, 余敖, 欧俊梅, 等. 小麦子粒粗蛋白 FT-NIRS 分析模型建立的初步研究[J]. 安徽农学通报, 2015, 21(3/4): 22-23.
- [3] 胡振, 陈素彬, 张晓琪, 等. 蝙蝠算法优化近红外光谱校正模型测定柴油黏度[J]. 当代化工, 2019, 48(3): 647-651.
- [4] 沈乐丞, 刘书航, 邓海玲, 等. 近红外光谱结合偏最小二乘法快速测定糖果中水分含量[J]. 食品工业科技, 2018, 39(7): 255-258.
- [5] 郑瑞娜, 谢定, 杨倩圆, 等. 基于近红外光谱的海藻糖浆组分快速定量检测[J]. 食品与机械, 2017, 33(10): 60-63.
- [6] 李路, 黄汉英, 赵思明, 等. 大米蛋白质、脂肪、总糖、水分近红外检测模型研究[J]. 中国粮油学报, 2017, 32(7): 121-126.

- [7] 陈冲, 罗华平, 索玉婷, 等. 基于近红外光谱对冬枣水分糖分的无损检测研究[J]. 农业与技术, 2019, 39(5): 1-3.
- [8] 何云啸, 张晓青, 张阳, 等. 基于 PLSR 及红外光谱技术的冻干过程诺丽果粉水分含量预测[J]. 食品与机械, 2017, 33(10): 56-59.
- [9] 贾柳君, 王健, 张海红, 等. 近红外光谱技术定量分析葡萄酒的主要成分[J]. 食品科技, 2017, 42(5): 273-278.
- [10] 陈素彬, 胡振, 张晓琪, 等. 软件技术在近红外光谱定量分析中的应用[J]. 化学教育(中英文), 2018, 39(16): 62-67.
- [11] 江水泉, 孙通. 基于可见/近红外光谱和变量选择的脐橙可溶性固形物含量在线检测[J]. 食品与机械, 2020, 36(2): 89-93.
- [12] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 糕点通则: GB/T 20977—2007[S]. 北京: 中国标准出版社, 2007.
- [13] 中华人民共和国国家卫生和计划生育委员会. 食品安全国家标准 食品中水分的测定: GB 5009.3—2016[S]. 北京: 中国标准出版社, 2016.