

贮藏期内灵武长枣果糖含量的高光谱预测

万国玲, 刘贵珊, 何建国*, 杨晓玉, 程丽娟, 张 翀

宁夏大学农学院, 宁夏 银川 750021

摘 要 高光谱成像可将图像和光谱相结合, 同时获得目标对象的图像和光谱信息, 已在农产品定性和定量分析检测方面得到广泛利用。利用可见-近红外高光谱成像结合化学计量学方法对贮藏期内灵武长枣果糖含量进行无损检测。采用高效液相色谱测量长枣果糖含量的化学值, 可见-近红外高光谱系统采集长枣的高光谱图像, 提取每个样本感兴趣区域的平均光谱; 建立长枣贮藏期的径向基核函数支持向量机(radial basis kernel function support vector machine, RBF-SVM)模型; 分别选用正交信号校正法(orthogonal signal correction, OSC)、多元散射校正(multiplicative scatter correction, MSC)、中值滤波(median-filter, MF)、卷积平滑(savitzky-golay, SG)、归一化(normalization, Nor)、高斯滤波(gaussian-filter, GF)和标准正态变换(standard normalized variate, SNV)等方法对原始光谱进行预处理; 为减少数据量, 降低维度, 提高运算速度, 采用反向区间偏最小二乘法(backward interval partial least squares, BiPLS)、间隔随机蛙跳算法(interval random frog, IRF)和竞争性自适应加权算法(competitive adaptive reweighted sampling, CARS)对光谱数据提取特征变量; 建立全波段和特征波段的偏最小二乘回归(partial least squares regression, PLSR)和主成分回归(principle component regression, PCR)长枣果糖含量预测模型。结果表明: RBF-SVM 判别模型校正集准确率为 98.04%, 预测集准确率为 97.14%, 能很好地预测长枣的贮藏期; 利用 BiPLS, IRF 及 CARS 进行降维处理, 提取特征波长个数为 100, 63 和 23, 占原光谱数据的 80%, 50.4% 和 18.4%; 为简化模型运算过程并提高模型精度, 采用 CARS 算法对 BiPLS 及 IRF 算法所选取的特征波长进行二次筛选, 分别优选出 18 和 15 个特征波长, 占原光谱数据的 14.4% 和 12%, 显著减少特征波长数; 将全波段光谱与提取出的特征波长分别建立长枣果糖含量的 PLSR 及 PCR 预测模型, 优选出 CARS 提取特征波长建立的 PLSR 模型效果最优, 其中校正集的相关系数 $R_c=0.8544$, 均方根误差 $RMSEC=0.0053$, 预测集的相关系数 $R_p=0.8303$, 均方根误差 $RMSEP=0.0057$, 说明 CARS 有效地对光谱进行降维, 简化了数据处理过程。研究表明, 利用可见-近红外高光谱成像结合化学计量学方法及计算机编程, 可以有效的实现灵武长枣果糖含量的快速无损分析, 为灵武长枣内部品质的检测提供理论依据。

关键词 高光谱成像技术; 果糖; 贮藏期; 高效液相色谱法; 偏最小二乘回归

中图分类号: TS255 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)10-3261-06

引 言

果糖是红枣中单糖的重要组成成分, 果糖的口感和甜度优于其他糖且升糖指数低, 被称为“健康糖”, 受到国内外学者的高度关注^[1]。传统检测糖含量的方法耗时、过程复杂且破坏样本的完整性^[2], 无法快速无损地检测出灵武长枣中果糖的含量; 因此, 灵武长枣中果糖含量的快速无损检测是目前研究的重点。

近年来, 高光谱成像作为一种快速无损的检测方法, 广泛应用于农产品内外部品质的检测^[3-4]。Oliveira 等^[5]、Hu 等^[6]、Guo^[7]等利用光谱检测了百香果、猕猴桃、枣中糖、酸、类胡萝卜素含量、总酚含量和总抗氧化剂活性; 冯迪等^[8]、冯愈钦等^[9]对水果内部成分糖度、硬度等进行检测, 均取得了较好的研究成果。因此, 高光谱成像在农产品品质检测中具有良好的发展前景。

以灵武长枣为研究对象, 以长枣中关键成分果糖为检测指标, 分别采用不同算法对光谱数据进行降维处理, 建立灵

收稿日期: 2018-09-10, 修订日期: 2019-01-22

基金项目: 国家自然科学基金项目(31560481), 国家自然科学基金项目(75002108A1651)资助

作者简介: 万国玲, 女, 1994年生, 宁夏大学农学院硕士研究生 e-mail: glingwan@163.com

* 通讯联系人 e-mail: hejg@nxu.edu.cn

武长枣中果糖含量的预测模型,为实现高光谱成像对灵武长枣内部品质的检测提供理论依据。

1 实验部分

1.1 材料

样品采自灵武果业开发有限责任公司,选取 147 个无损伤的长枣,4 °C 贮藏。

果糖标品,四川省维克奇生物科技有限公司;乙醇(分析纯),超纯水,一次性有机滤膜(0.45 μm×50 mm),水系膜(0.45 μm×50 mm),滤头(0.45 μm)。

1.2 仪器

可见-近红外高光谱成像系统(美国 Headwall Photonics 公司)和高效液相色谱仪 AGILENT1100(美国安捷伦科技公司)。可见-近红外高光谱成像系统由成像光谱仪(V10E-QE 型)、CCD 摄像机(C8484-05G 型)、光纤卤素灯(DCR III 型)、输送装置、计算机和数据采集软件等构成。

1.3 高光谱图像的采集与黑白标定

为了避免图像失真,经多次预试验确定最佳光谱数据采集参数为:曝光时间为 10 ms,曝光物距为 385 mm,电控位移台扫描速度为 0.42 mm·s⁻¹。由于光源强度分布不均匀及暗电流的存在均会导致图像中含有较大的噪声,因此,在采集高光谱图像前需要进行黑白校正来消除噪声的影响^[10]。如式(1)所示。

$$R = \frac{R_0 - D}{W - D} \times 100 \quad (1)$$

式(1)中:R 是校正后图像,R₀ 是原始图像,D 是暗图像,W 是白板图像。

1.4 高效液相色谱法(HPLC)测量灵武长枣果糖含量

标准溶液的配备:称取果糖标品 100 mg,超纯水溶解定容于 5 mL 的容量瓶中,采用梯度稀释为质量浓度为 0.5, 1, 2, 4 和 8 mg·mL⁻¹ 的标准溶液。

样品待测液的制备:取长枣果肉 1 g,加入 10 mL 80% 的乙醇后均质,超声提取 30 min 后,11 000 r·min⁻¹ 离心 15 min,收集上清液,残渣加入 5 mL 80% 乙醇再提取,将 2 次上清液合并,55 °C 真空旋干,再用超纯水定容至 25 mL,经

0.45 μm 针式过滤器过滤。液相条件:流动相为超纯水,进样量为 10 μL,流速为 0.4 mL·min⁻¹,柱温为 30 °C,示差折光检测器温度为 35 °C。

2 结果与讨论

2.1 HPLC 测定结果

2.1.1 果糖标准曲线的绘制

图 1(a)为果糖标准品高效液相色谱图,保留时间为 14.90 min;图 1(b)为长枣提取液的高效液相色谱出峰图,长枣提取液中果糖的保留时间与标准品对应。果糖对应的峰面积与进样量的线性方程为 $y = 4.189 0 \times 10^{-6} X - 4.32 \times 10^{-2}$,相关系数 $r = 0.999 9$,说明两者具有良好的线性相关性。

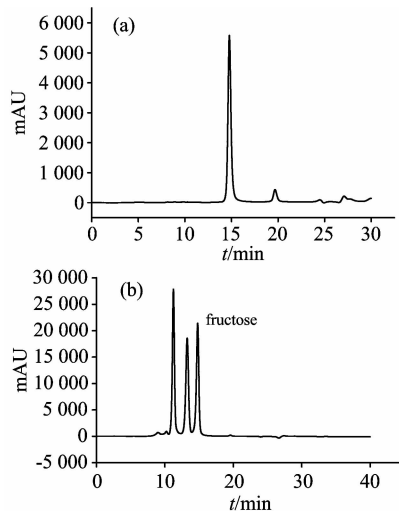


图 1 果糖标品(a)和长枣提取液(b)的高效液相色谱图

Fig. 1 High-performance liquid chromatography analyses of fructose (a) and jujube (b)

2.1.2 加标回收率测定结果

由表 1 可知,加标回收率在 92.96%~99.67%,平均回收率为 96.55%,能准确测量长枣中果糖含量。

表 1 回收率实验

Table 1 Recovery determination test

Component	content/(mg·mL ⁻¹)			Rate of recovery/%	Average recovery rate/%
	Background value	Amount of addition	Measurement		
Content	1.690 7	1	2.620 3	92.96	96.55
	2.524 6	2	4.464 9	97.02	
	3.249 9	4	7.236 6	99.67	

2.2 高光谱数据的提取

采集长枣图像信息,利用 ENVI 4.6 提取长枣感兴趣区域,分别从枣赤道部位提取椭圆大小为 25 mm×15 mm 的 ROI,计算出平均反射光谱作为原始光谱,如图 2(a)所示。将平均光谱再均值化,得到图 2(b)所示的长枣贮藏期的原始平均光谱,由图 2(b)可看出,在 675 nm 附近,光谱反射值达到峰谷,是由于长枣样本的 C—H 伸缩振动引起的^[11],

900~1 000 nm 波段间的吸收峰是由于长枣内部水分中 O—H 基团的二倍频特征吸收^[12]所引起的。

2.3 校正集预测集划分

蒙特卡洛法具有同时检测光谱异常值和理化参照值异常值的优点,采用蒙特卡洛法剔除异常值后将红枣样本按照 3:1 的比例划分如表 2 所示。

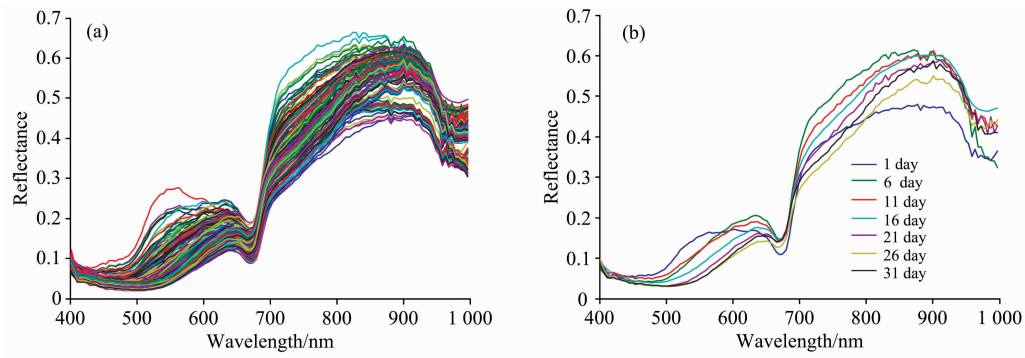


图 2 原始光谱曲线 (a)和平均原始光谱曲线 (b)
 Fig. 2 Raw reflectance spectra (a) and average raw spectra (b)

表 2 长枣果糖数据统计

Table 2 Statistics of fructose content in jujube

Data set	Samples	Fructose content/(mg · g ⁻¹)			
		Max	Min	Mean	Standard deviation
Calibration set	102	0.080 2	0.030 5	0.054 6	0.010 2
Prediction set	35	0.073 1	0.030 8	0.056 0	0.009 7

2.4 长枣贮藏时间 RBF-SVM 判别模型建立与校正

径向基核函数支持向量机 (RBF-SVM) 在低样本、非线性性和多维空间等模式识别方面有较强的优势。在全波段范围内,对原始光谱建立长枣贮藏时间的模型判别模型如图 3 所示。图 3(a)是校正样本训练模型,得到 RBF-SVM 判别模型

对长枣贮藏时间的判别率达 98.04%,所建模型可有效鉴别贮藏时间。利用已建好的 RBF-SVM 校正模型对预测集样本的贮藏时间进行判别,由图 3(b)可以得出,模型对预测集判别准确率高达 97.14%,模型可靠性和预测能力较好。

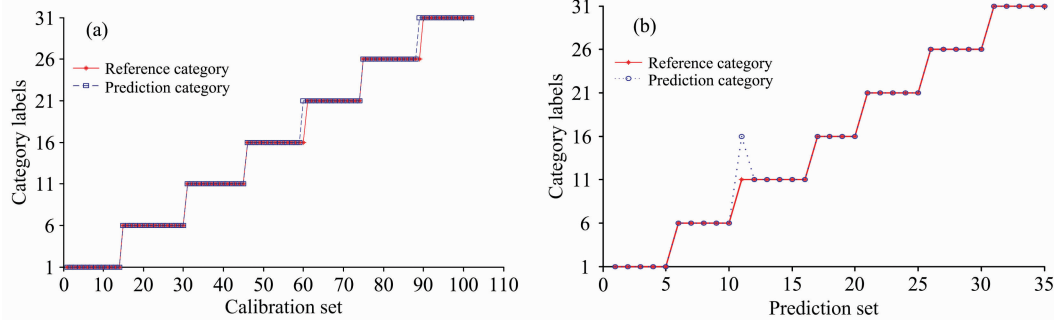


图 3 长枣贮藏时间的 RBF-SVM 判别模型
 Fig. 3 RBF-SVM discriminant model of storage time in long Jujube

2.5 光谱预处理的比较与分析

分别对预处理后的光谱曲线建立长枣果糖含量的 PLSR 预测模型,建模效果如表 3 所示。由表 3 可见,原始光谱的 PLSR 的系数 R_c 和 R_p 值最大且 RMSEC 和 RMSEP 值最小,建模效果最好。预处理后建模效果差可能是消除了与红枣果糖相关的有用信息,导致 PLSR 预测模型效果较差。

2.6 特征波长的提取

2.6.1 应用 CARS 算法选取特征波长

图 4 为采用 CARS 提取特征波长的图。(a)曲线反映出算法对变量进行“粗选”和“精选”的过程;(b)曲线为 RM-

SECV 的变化趋势图,从图中可以看出当采样次数为 12 次时, RMSECV 值最小为 0.006 4;(c)为回归系数的变化曲线图,“*”为残差最低点,与(a),(b)曲线相对应。共优选出 23 个特征波长,占全光谱的 18.4%。

2.6.2 应用 BiPLS 算法选取特征波长

将整条光谱分为 20 个子区间,采用留一法交互验证计算各个区间的 RMSECV 值可知,当组合第 2~5, 7~8 和 10~19 等 16 个区间时, RMSECV 值最小,所以入选区间为 [2 3 4 5 7 8 10 11 12 13 14 15 16 17 18 19],所对应的波长主要分布在 430.2~550.2, 583.8~641.4, 675.0~967.9 nm。

表 3 不同预处理方法的长枣果糖含量的 PLSR 模型

Table 3 Results of fructose content PLSR models using different spectral pretreatments

Pretreatment method	factor	Calibration set		Prediction set	
		RMSEC	R_c	RMSEP	R_p
Raw data	9	0.006 7	0.758 6	0.006 9	0.737 5
OSC	3	0.006 9	0.735 4	0.015 1	0.705 4
SNV	13	0.006 0	0.813 2	0.007 3	0.715 7
MF(5)	9	0.006 6	0.765 4	0.007 3	0.715 7
GF	8	0.006 8	0.745 3	0.007 1	0.702 5
MSC+Nor(range)	11	0.006 2	0.787 9	0.006 9	0.734 3
SNV+SG(9)+Nor(range)	11	0.006 3	0.788 3	0.006 9	0.735 0

2.6.3 应用 IRF 算法优选特征波长

IRF 是基于随机蛙跳的一种波长间隔选择方法。应用 IRF 提取特征波长得到 121 个间隔中排名前 10 的间隔见表 4；计算按照排名组合间隔的 RMSECV 值，间隔波长排名前 10 的 RMSECV 不是最小，RMSECV 最低为第 28 个间隔，因此，选择前 28 名的间隔波长，最终波长为 439.8~473.4, 492.6~516.6, 555.0~607.8, 660.6~679.8, 713.4~751.9, 833.5~862.3, 919.9~967.9 和 977.5~996.7 nm, 共 63 个波长。

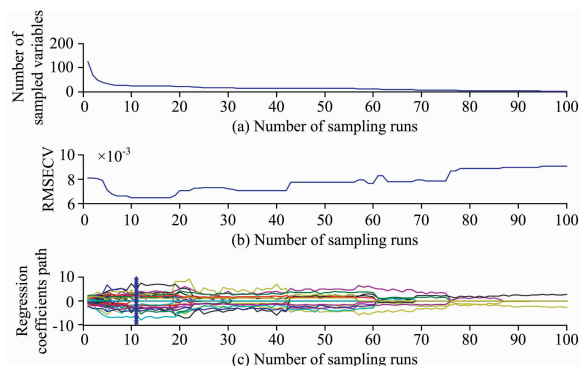


图 4 CARS 优选变量分布图

Fig. 4 Distribution of selected CARS variables

表 4 排名前 10 的波长间隔

Table 4 Top ranked 10 wavelength intervals

排名	间隔	排名	间隔
1	37~41	6	109~113
2	21~25	7	69~73
3	92~96	8	36~40
4	40~44	9	10~14
5	110~114	10	20~24

2.6.4 应用 BiPLS-CARS, IRF-CARS 算法优选特征波长

BiPLS 及 IRF 算法提取特征波长时选取的特征波长数较多，为简化模型运算过程，提高模型精度，采用 CARS 算法对 BiPLS 及 IRF 算法选取的特征波长进行二次筛选，CARS 优选 BiPLS 算法后 RMSECV 变化如图 5(a)所示，RMSECV 逐渐下降，达到最低点 0.006 9，CARS 优选 IRF 算法 RMSECV 变化如图 5(b)所示，RMSECV 先下降后上升最低

点为 0.006 9。

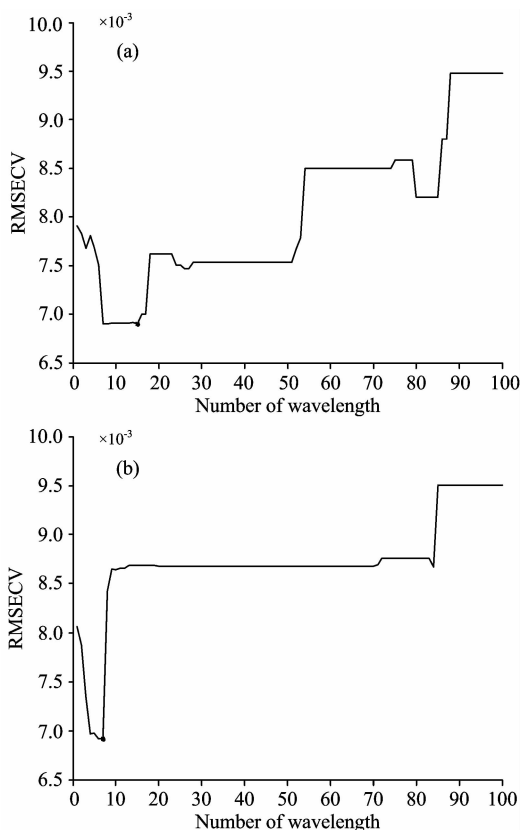


图 5 基于 BiPLS-CARS(a)和 IRF-CARS(b)法的特征波长的选择

Fig. 5 Selection of characteristic wavelengths based on BiPLS-CARS (a) and IRF-CARS (b) methods

2.7 模型效果比较

建立全波段及特征波段的预测模型，结果见表 5。由表 5 可知，CARS-PLSR 模型其 R_c 值、 R_p 值均大于原始光谱模型，RMSEC 值、RMSEP 值均小于原始光谱模型；选用 CARS 算法对 BiPLS 及 IRF 算法筛选出的特征波长进行优化，发现模型的 R_c 显著增大、RMSEC 值及 RMSEP 值显著减小，说明 CARS 算法对这两种算法提取的特征波长进行有效的筛选。

表 5 不同波长提取方法建立的 PLSR 及 PCR 模型结果

Table 5 The results of PLSR and PCR models based on different wavelength extraction methods

Modeling method	Extraction method	Variable number	Calibration set		Validation set		$R_c + R_p$
			RMSEC	R_c	RMSEP	R_p	
PLSR	FS	125	0.006 7	0.758 6	0.006 9	0.737 5	1.496 1
	CARS	23	0.005 3	0.854 4	0.005 7	0.830 3	1.684 7
	BiPLS	100	0.006 8	0.750 5	0.007 3	0.694 9	1.445 4
	IRF	63	0.005 5	0.840 2	0.006 8	0.762 4	1.602 6
	BiPLS+CARS	18	0.006 1	0.800 0	0.005 8	0.815 4	1.615 4
	IRF+CARS	15	0.005 8	0.826 2	0.006 8	0.758 2	1.584 4
PCR	CARS	23	0.005 6	0.837 7	0.006 2	0.802 1	1.639 8
	BiPLS	100	0.006 9	0.741 0	0.007 4	0.692 5	1.433 5
	IRF	63	0.006 9	0.735 9	0.007 1	0.715 9	1.451 8
	BiPLS+CARS	18	0.006 1	0.800 0	0.005 8	0.815 3	1.615 3
	IRF+CARS	15	0.005 8	0.826 2	0.006 8	0.758 2	1.584 4

3 结 论

利用 VIS/NIR 高光谱技术对灵武长枣果糖含量进行预测,同时对贮藏时间灵武长枣的贮藏时间进行预测模型的建立。研究结果如下:

(1)建立灵武长枣反射光谱与贮藏时间的径向基核函数支持向量机(RBF-SVM)模型,校正集判别准确率为 98.04%,预测集判别准确率为 97.14%。

(2)利用 CARS, BiPLS, IRF, BiPLS-CARS 和 IRF-CARS 进行降维处理,提取特征波长个数为 23, 100, 63, 18 和 15, 占原光谱数据的 18.4%, 80%, 50.4%, 14.4% 和 12%。

(3)建立基于原始光谱及特征波段的 PLSR 及 PCR 模型,优选出 FS-CARS-PLSR 模型具有较高的运行速率和预测能力($R_c = 0.854 4$, $RMSEC = 0.005 3$, $R_p = 0.830 3$, $RMSEP = 0.005 7$)。

References

- [1] Simeone M L F, Parrella R A C, Schaffert R E, et al. Microchemical Journal, 2017, 134: 125.
- [2] Das B, Sahoo R N, Pargal S, et al. Spectrochim Acta Part A: Molecular and Biomolecular Spectroscopy, 2017, 192: 41.
- [3] Mo C, Kim M S, Kim G, et al. Biosystems Engineering, 2017, 159: 10.
- [4] SUN Jing-tao, MA Ben-xue, DONG Juan, et al(孙静涛, 马本学, 董娟, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(7): 2184.
- [5] Oliveira G A D, Castilhos F D, Bureau S. Food Research International, 2014, 60(6): 154.
- [6] Hu W, Sun D W, Blasco J. Postharvest Biology & Technology, 2017, 125: 168.
- [7] Guo Y, Ni Y, Kokot S. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2016, 153: 79.
- [8] FENG Di, JI Jian-wei, ZHANG Li, et al(冯迪, 纪建伟, 张莉, 等). Chinese Journal of Luminescence(发光学报), 2017, 38(6): 799.
- [9] FENG Yu-qin, WU Long-guo, HE Jian-guo, et al(冯愈钦, 吴龙国, 何建国, 等). Chinese Journal of Luminescence(发光学报), 2016, 37(8): 1014.
- [10] Zou Xiaobo, Xu Yiwei, Chen Wu, et al. Mod. Food Sci. Technol., 2015, 32(9): 263.
- [11] Siedliska A, Baranowski P, Zubik M, et al. Journal of Food Engineering, 2017, 215: 61.
- [12] Siedliska A, Baranowski P, Zubik M, et al. Postharvest Biology & Technology, 2018, 139: 115.

Prediction of Fructose Content of Lingwu Long Jujube During Storage Using Hyperspectral Imaging Technique

WAN Guo-ling, LIU Gui-shan, HE Jian-guo*, YANG Xiao-yu, CHENG Li-juan, ZHANG Chong

School of Agriculture, Ningxia University, Yinchuan 750021, China

Abstract Hyperspectral imaging technique which is a non-destructive method combines image and spectral techniques to obtain image and spectral information of target objects' and qualitative and quantitative analysis using spectral data has been widely used in the field of agricultural product testing. This paper uses visible/near-infrared spectroscopic imaging technique combined with chemometrics methods to achieve the non-destructive detection of fructose content of Lingwu long jujube during storage. The chemical value of jujube fructose was determined by High performance liquid chromatography (HPLC), and the hyperspectral images of long jujubes were collected using near-infrared hyperspectral system, and the average spectral data for each sample area of interest were extracted. Support Vector Machine With RBF Nucleus (RBF-SVM) Model for establishing storage time of long jujube. Orthogonal Signal Correction (OSC), Multiple Scatter Correction (MSC), Median Filter (MF), Savitzky-Golay (SG), Normalize (Nor), Gaussian filter (GF) and Standard Normalized Variate (SNV) were used to preprocess the original spectral data. To reduce the amount and dimension of data, the characteristic wavelengths were extracted by Backward interval Partial Least Squares (BiPLS), Interval Random Frog (IRF) and Competitive Adaptive Reweighted Sampling (CARS); the partial least squares regression (PLSR) model and principle component regression (PCR) were established based on full spectra and characteristic wavelengths for predicting fructose of Lingwu long jujube. The results indicated that the accuracy of the RBF-SVM model calibration set was 98.04%, and the accuracy of the prediction set was 97.14%, which could well predict the storage time of the jujube; The BiPLS, IRF and CARS methods were used to select characteristic wavelengths with 100, 63 and 23 from 125 wavelengths, respectively. In order to simplify the model and improve the accuracy of prediction of the model, the CARS algorithm was used to perform secondary extracted characteristic wavelengths of BiPLS and IRF and select characteristic wavelengths with 18 and 15, respectively, which significantly reduced the number of characteristic wavelengths. Comparing models of the full band spectrum with the models of extracted characteristic wavelengths of PLSR and PCR, PLSR model based on the characteristic variables selected by CARS was the best, and correlation coefficient of Calibration set (R_c) and root-mean-square error of Calibration set (RMSEC) of the model were 0.854 4 and 0.005 3, and correlation coefficient of prediction (R_p) and root-mean-square error of prediction set (RMSEP) of the model were 0.830 3 and 0.005 7, respectively, which indicated that CARS effectively reduced the dimension of the spectrum and simplified the data processing. The results showed that visible/near-infrared hyperspectral imaging technique combined with chemometrics methods and computer programming can effectively detect fructose content of Lingwu long jujube rapidly and non-destructively, providing a theoretical basis for the detection of internal quality of Lingwu long jujube.

Keywords Hyperspectral imaging technique; Fructose; Storage; High performance liquid chromatography; Partial least squares regression

(Received Sep. 10, 2018; accepted Jan. 22, 2019)

* Corresponding author