

基于BP-ANN和PLS的近红外光谱无损检测李果实品质的研究

赵志磊^{1, 2, 3, 4}, 王雪妹^{1, 2, 3}, 刘冬冬^{1, 2, 3}, 王艳伟^{1, 2, 3}, 顾玉红⁵, 滕佳鑫¹, 牛晓颖^{1, 2, 3, 4*}

1. 河北大学质量技术监督学院, 河北 保定 071002
2. 计量仪器与系统国家地方联合工程研究中心, 河北 保定 071002
3. 河北省能源计量与安全检测技术重点实验室, 河北 保定 071002
4. 河北大学地理标志研究院, 河北 保定 071002
5. 河北农业大学生命科学学院, 河北 保定 071002

摘要 可溶性固形物(SSC)和可滴定总酸(TA)含量是影响李果实品质的重要指标,经典的破坏性检测方法不适用于果实按品质分级,近红外光谱(NIRS)检测方法具有速度快、操作简便、可无损检测果实品质。为实现NIRS无损快速检测安哥诺李果实可溶性固形物和可滴定总酸含量,利用NIRS采集李果实的漫反射光谱,同时采用糖度计测定安哥诺李果实的SSC,采用滴定法测定了李果实TA含量,使用杠杆值和F概率值剔除异常样品,采用软件优化结合人工筛选光谱波段,使用了消除常数偏移量、减去一条直线、矢量归一化(SNV)、最大-最小归一化、多元散射校正(MSC)、一阶和二阶导数结合平滑处理、一阶导数结合减去一条直线和平滑处理、以及一阶导数结合SNV或MSC校正等光谱预处理方法,分别采用偏最小二乘法(PLS)和主成分分析结合反向传播人工神经网络(BP-ANN)建立李果实SSC、TA的定量分析模型。结果表明,李果实SSC和TA的最佳PLS建模效果波段范围分别为4 000~8 852和4 605~6 523 cm^{-1} 。SSC的PLS模型的最佳光谱预处理方法为MSC校正,最佳模型校正相关系数(R_c)为0.914 4,预测相关系数(R_p)为0.878 5,校正均方根误差(RMSEC)为0.91,预测均方根误差(RMSEP)为1.00。经一阶微分结合SNV和9点平滑的方法预处理后,TA的PLS模型效果最佳, R_c , R_p , RMSEC, RMSEP分别为0.860 3, 0.819 6, 0.80和0.86。提取了李果实SSC和TA光谱数据的主成分,并基于前10个主成分得分建立了李果实SSC和TA最佳BP-ANN定量分析模型,其 R_c , R_p , RMSEC和RMSEP分别为0.976 7, 0.889 7, 0.75和0.99;TA的BP-ANN模型的相应参数值依次为0.974 3, 0.897 7, 0.62和0.83,与采用PLS算法建立的定量模型相比较,BP-ANN模型具有较高的 R_c , R_p 和较低的RMSEC, RMSEP,因此BP-ANN模型对SSC和TA指标的定量分析结果更佳。

关键词 李果实; 偏最小二乘法; 反向传播人工神经网络; 近红外光谱

中图分类号: O657.33 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)09-2836-07

引言

“安哥诺”李是李中极晚熟品种,色香味美,品质极佳。随着李果实产量和消费量的增加,消费者在选购李果实时除了关注其大小、形状、颜色等外观品质,对于内在品质也越来越重视,其中糖、酸含量是决定果实口感的主要因素,直接影响消费者的购买意愿。因此实现李果实糖度和酸度的快速无损检测以品质分级实现优质优价显得尤为重要。糖、酸

含量的传统检测方法均属有损检测,无法满足对大批量果实进行快速分级筛选的需求,急需建立一种快速无损检测李果实糖、酸含量的方法。

近红外光谱(near infrared spectroscopy, NIRS)分析技术作为一种无损、快速及同时测定多种组分的检测手段已在多种水果的无损检测方面得到广泛应用。近些年来,国内外众多学者采用近红外光谱分析技术相继开展了对苹果^[1]、梨^[2]、芒果^[3]、桃^[4]、灵武长枣^[5]、猕猴桃^[6]、橙子^[7]、山楂^[8]、柿子^[9]、西瓜、樱桃^[10]等水果的可溶性固形物(solu-

收稿日期: 2021-08-22, 修订日期: 2022-03-17

基金项目: 国家自然科学基金项目(31872907), 河北省自然科学基金项目(C2021201011)资助

作者简介: 赵志磊, 1977年生, 河北大学质量技术监督学院教授 e-mail: 15930721279@163.com

* 通讯作者 e-mail: 408643620@qq.com

ble solids content, SSC)含量、可滴定总酸(titratable acidity, TA)和硬度等品质的无损检测研究。与其他果实相比,李果实表面光洁且果皮较薄,尤其适用于NIRS漫反射光谱信息采集果实品质信息。Paz等获取了9个品种李果实样品的近红外光谱,并建立了评估李果实SSC含量和硬度的校正模型,用以对不同品种的李果实进行分类,预测结果得出SSC的交叉验证的标准误差和决定系数分别为0.77, 0.83^[11]。Louw等建立了波长为800~2700 nm范围内的傅里叶变换近红外反射模型对李果实SSC和TA及糖酸比指标进行预测,也取得了较好的结果^[12]。白凤华等使用偏最小二乘法结合近红外光谱定量检测了李果实的坚实度,其模型校正相关系数、校正和预测均方根误差分别为0.878 1, 1.22和1.51 kg·cm⁻²^[13]。上述研究为近红外光谱无损预测“安哥诺”李果实内部品质的提供了重要参考。

本研究提出一种使用杠杆值和F概率值进行异常样品剔除、人工筛结合软件自动优化比较不同波段和不同光谱预处理方法优化分析模型,采用偏最小二乘法、主成分分析结合反向传播神经网络两种算法建模,通过实验对比寻求最佳建模方案,建立了适用于“安哥诺”李果实SSC和TA的定量分析模型,为实现快速无损检测“安哥诺”李品质提供技术支持。

1 实验部分

1.1 材料

“安哥诺”李采摘自河北省保定市易县独乐村商品李果园。对采摘以后的李果实清洗晾干,挑选果形正常、色泽一致且无机械损伤的李果实样品,用于测定SSC的样品数目为583个、TA的样品数目为402个并对其标记。

1.2 近红外光谱采集

使用MPA近红外光谱仪及漫反射附件(Bruker, 德国)采集李果实的漫反射光谱,光谱采集范围为4000~12500 cm⁻¹,分辨率设定为8 cm⁻¹,扫描次数32,仪器光源为20 W钨卤灯,检测器为PbS,所使用的光谱采集软件为布鲁克公司的OPUS 6.5。

以仪器内置的金背景作为背景光谱,对李果实样品赤道部位每隔120°进行1次光谱扫描,共得到3条扫描光谱,而后取其平均光谱作为该样品的原始光谱用于分析。

1.3 可溶性固形物和总酸测定方法

采集李果实样品的近红外光谱后,取光谱采集点处的果肉挤压果汁,使用PAL-1型数显糖量计(Atago, 日本)测量SSC含量。

总酸含量的分析则依据GB/T 12456—2008进行。精确称取5.00 g李果实样品置于研钵中,充分研磨至果浆状后完全转移至10 mL离心管中,经超声振荡30 min及离心10 min后,取上清液于锥形瓶中作为待测溶液。以酚酞作为指示剂,使用0.10 mol·L⁻¹的氢氧化钠溶液滴定至待测液呈微红色且30 s内不褪色为终点。计算出可滴定酸的数值。

1.4 数据处理

使用杠杆值和F概率值进行异常样品剔除及偏最小二

乘法(partial least squares, PLS)建模和优化均使用OPUS 6.5软件,使用消除常数偏移量、减去一条直线、矢量归一化(standard normal variate, SNV)、最大-最小归一化、多元散射校正(multiplicative scatter correction, MSC)、一阶和二阶导数结合平滑处理、一阶导数结合减去一条直线和平滑处理、以及一阶导数结合SNV或MSC校正等光谱预处理方法^[14]。用MATLAB 7.0软件提取样品全光谱数据的主成分,并进行神经网络的建模和优化。

2 结果与讨论

2.1 安哥诺李果实近红外光谱分析

安哥诺李果实的近红外光谱的图如图1所示。在9000~12500 cm⁻¹波段范围内,李果实光谱包含较多的噪声,而4000~9000 cm⁻¹光谱波段中所包含的光谱噪声则相对较少,且在波数7134, 5264和4463 cm⁻¹附近有较为明显的吸收峰,可能是由于李果实中的SSC和TA等成分所包含的基团(C—H, O—H, N—H)对近红外光谱吸收不同所导致,可为光谱与SSC和TA建立关系提供理论基础。其中,7134 cm⁻¹附近的吸收峰可能与O—H键伸缩振动的一级倍频吸收有关,5264 cm⁻¹附近的吸收峰与水中O—H键的伸缩和变形振动的吸收相关,而4463 cm⁻¹附近的吸收峰则可能与NH₃基团的变形振动和N—H键伸缩振动吸收有关^[14]。

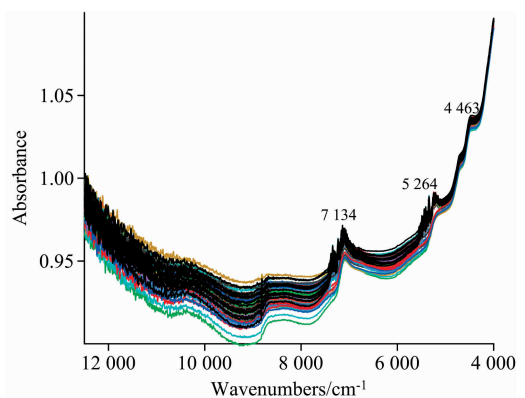


图1 李果实的近红外光谱

Fig. 1 Raw spectra of plum fruit

2.2 可溶性固形物和可滴定酸含量分析

在对李果实品质进行定量分析过程中,建模样品的各指标浓度的检测精度、分布情况和覆盖范围对所建模型的性能优劣具有一定的影响。图2(a,b)分别为李果实样品的SSC、TA两个指标含量的频率分布图。由图2可知,各指标浓度数据的分布情况均大体呈现正态分布的趋势,说明本实验所获得的李果实样品的SSC、TA两个指标的浓度值具有一定的代表性,可用于对李果实SSC、TA进行分析。

2.3 近红外定量分析SSC和TA

2.3.1 有效波段筛选及样品集分配

(1)有效波段筛选

由图1可知,李果实的近红外光谱在9000~12500 cm⁻¹范围内没有较为明显的吸收峰且包含的噪音也相对较

多。因此,将该范围的光谱除去,对剩余波段中包含明显吸收峰的光谱范围进行手动选取和组合,结合软件自动优化,

分别进行 SSC、TA 的定量建模,并比较不同波段的建模效果,如表 1。

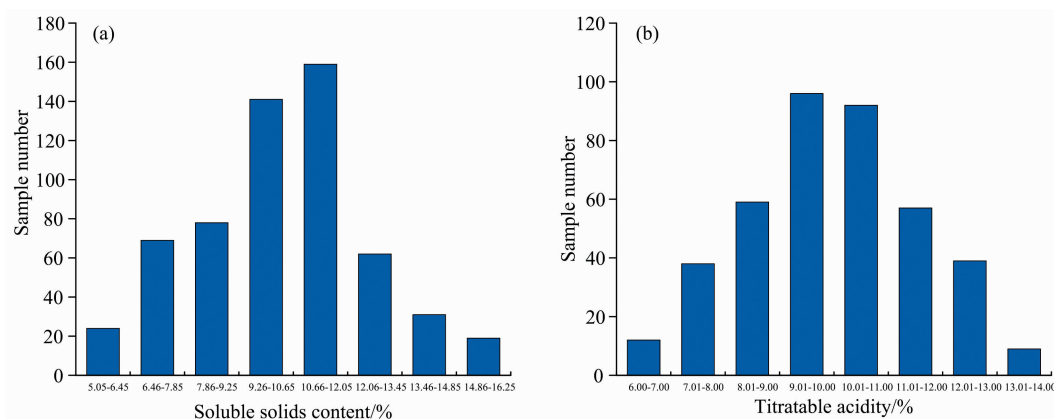


图 2 李果实样品 SSC (a)和 TA (b)含量频率分布图

Fig. 2 The probability distribution of SSC (a) and TA (b) content for plum samples

表 1 不同波段建模效果比较

Table 1 Comparison on performance of different model bands

指标名称	波段/ cm^{-1}	校正集		交叉验证		因子数	剔除样品
		r	RMSEC	r	RMSECV		
SSC/%	4 000~9 000	0.890 1	1.01	0.876 3	1.06	8	剔除 7 个
		0.891 1	1.01	0.877 5	1.06	8	剔除 8 个
		0.901 6	0.96	0.888 7	1.01	8	剔除 7 个
	4 000~8 852	0.901 3	0.96	0.887 3	1.02	8	剔除 8 个
		0.904 8	0.94	0.893 2	0.99	8	剔除 10 个
	4 000~8 000	0.886 2	1.03	0.871 8	1.08	8	剔除 6 个
		0.883 9	1.04	0.872 9	1.07	7	剔除 7 个
4 000~6 000	0.884 5	1.03	0.873 0	1.07	7	剔除 8 个	
6 000~9 000	0.888 5	1.02	0.872 8	1.08	7	剔除 6 个	
TA/%	4 000~9 000	0.821 2	0.89	0.811 4	0.90	6	剔除 5 个
	4 000~6 523	0.827 5	0.87	0.818 7	0.88	5	剔除 5 个
	4 605~9 000	0.822 3	0.88	0.812 2	0.90	6	剔除 5 个
	4 605~6 523	0.833 0	0.86	0.824 5	0.87	5	剔除 5 个
	5 605~9 000	0.824 8	0.88	0.812 3	0.90	6	剔除 5 个
	5 605~6 523	0.821 7	0.88	0.799 6	0.92	5	剔除 6 个
	6 410~7 600	0.818 5	0.89	0.810 2	0.90	5	剔除 6 个

由表 1 可以看出,对于 SSC 的最佳波段范围为 4 000~8 852 cm^{-1} ,在剔除完 10 个异常样品后,该模型的相关系数较高,且校正集和交叉验证的标准误差都有所下降,模型性能趋于稳定。TA 指标的最优波段覆盖范围为 4 605~6 523 cm^{-1} ,因而最终选取该波段用于定量建模分析。

(2) 样品集数据统计

对有效波段进行筛选后,SSC、TA 指标的剩余样品数分别为 573 个、397 个。根据 SSC、TA 的数值大小分别对各指标所对应的全部样品进行排序,并将最大值和最小值样品归入校正集,其余样品则按照大约 2:1 的比例随机分为校正集和预测集。李果实各指标的校正集和预测集样品的数据统计数据如表 2 所示。

表 2 李果实 SSC、TA 校正集和预测集样品统计数据

Table 2 Statistics of SSC, TA in plum for calibration set and prediction set

指标名称	样品集	样品数	数值范围	平均值	标准偏差
SSC /%	校正集	382	5.20~16.00	10.16	2.23
	预测集	191	5.40~14.90	10.19	2.14
TA/%	校正集	267	6.15~13.45	9.88	1.55
	预测集	130	6.44~12.68	9.84	1.52

2.3.2 定量模型的建立和优化

(1) 偏最小二乘法

在各指标的最佳光谱范围内,比较了不同的光谱预处理方法对各预测模型的影响,表 3—表 5 分别列出了结合不同

表 3 不同光谱预处理方法下 SSC 建模性能比较

Table 3 Performance comparison with the SSC model of different spectral pretreatment methods

光谱预处理	校正集		预测集		因子数
	R_c	RMSEC	R_p	RMSEP	
原始光谱	0.903 9	0.96	0.874 0	1.04	8
消除常数偏移量	0.913 5	0.92	0.875 9	1.02	9
减去一条直线	0.909 0	0.94	0.874 8	1.03	8
SNV	0.911 4	0.93	0.874 0	1.04	8
最大-最小归一化	0.903 4	0.97	0.863 3	1.08	8
MSC 校正	0.914 4	0.91	0.878 5	1.00	8
一阶微分+5 点平滑	0.876 9	1.08	0.876 6	1.01	4
一阶微分+17 点平滑	0.844 0	1.20	0.833 0	1.18	4
一阶微分+25 点平滑	0.862 2	1.14	0.874 7	1.03	5
二阶微分+17 点平滑	0.869 8	1.11	0.795 5	1.29	4
二阶微分+25 点平滑	0.842 6	1.21	0.812 2	1.24	4
一阶微分+减去一条直线+5 点平滑	0.904 8	0.95	0.874 1	1.04	4
一阶微分+SNV+25 点平滑	0.905 2	0.95	0.876 0	1.02	8
一阶微分+MSC+25 点平滑	0.905 6	0.95	0.875 4	1.02	7

表 4 不同光谱预处理方法下 TA 建模性能比较

Table 4 Performance comparison with the TA model of different spectral pretreatment methods

光谱预处理	校正集		预测集		因子数
	R_c	RMSEC	R_p	RMSEP	
原始光谱	0.840 3	0.85	0.807 2	0.89	5
消除常数偏移量	0.795 2	0.95	0.794 8	0.92	4
减去一条直线	0.830 6	0.87	0.796 3	0.92	5
SNV	0.843 5	0.85	0.819 3	0.87	4
最大-最小归一化	0.826 3	0.88	0.807 7	0.89	5
MSC 校正	0.841 7	0.84	0.814 0	0.88	4
一阶微分+17 点平滑	0.841 0	0.84	0.813 9	0.88	4
一阶微分+25 点平滑	0.815 5	0.91	0.812 0	0.88	4
二阶微分+17 点平滑	0.841 2	0.84	0.760 8	0.98	4
二阶微分+25 点平滑	0.825 3	0.88	0.790 6	0.93	4
一阶微分+减去一条直线+5 点平滑	0.815 1	0.91	0.774 2	0.96	4
一阶微分+SNV+9 点平滑	0.860 3	0.80	0.819 6	0.86	6
一阶微分+SNV+25 点平滑	0.840 7	0.85	0.816 8	0.87	5
一阶微分+MSC+25 点平滑	0.809 2	0.92	0.775 3	0.95	4

光谱预处理方法的 SSC、TA 含量的建模效果。

由表 3 可知，SSC 含量的定量模型在经过最大-最小归一化、一阶微分(平滑点数：17)和二阶微分(平滑点数：17 和 25)等光谱预处理方法处理后，原始光谱的模型性能无论是校正集还是预测集均有所下降；尽管平滑点数为 5 和 25 的一阶微分处理方法能够略微提高模型预测集的性能，但其校正集性能却变差，从而影响了模型的整体性能；而消除常数偏移量、减去一条直线、SNV、MSC 校正、一阶微分结合减去一条直线(平滑点数：5)、一阶微分分别结合 SNV 和 MSC 校正(平滑点数：25)等方法则使原始光谱的模型性能有所改善，尤其是经过 MSC 校正方法预处理后，原始光谱

的模型性能改善情况相对较为明显。因此，基于 $4\ 000\sim 8\ 852\ \text{cm}^{-1}$ 波段范围结合 MSC 校正的预处理方法建立了李果实 SSC 近红外定量检测模型，最佳模型的定量结果如图 3 (a) 所示。

由表 4 可知，SNV、MSC 校正、一阶微分(平滑点数：17)以及一阶微分结合 SNV(平滑点数：9 和 25)等预处理方法均能够改善李果实 TA 原始光谱模型的相关性和预测精度，且由表 4 数据可知，经一阶微分结合 SNV 和 9 点平滑方法处理后的 TA 模型效果最佳，其校正集和预测集的相关系数都得到提升且误差均有所下降；最大-最小归一化、一阶微分结合 25 点平滑以及二阶微分结合 17 点平滑等方法达不到

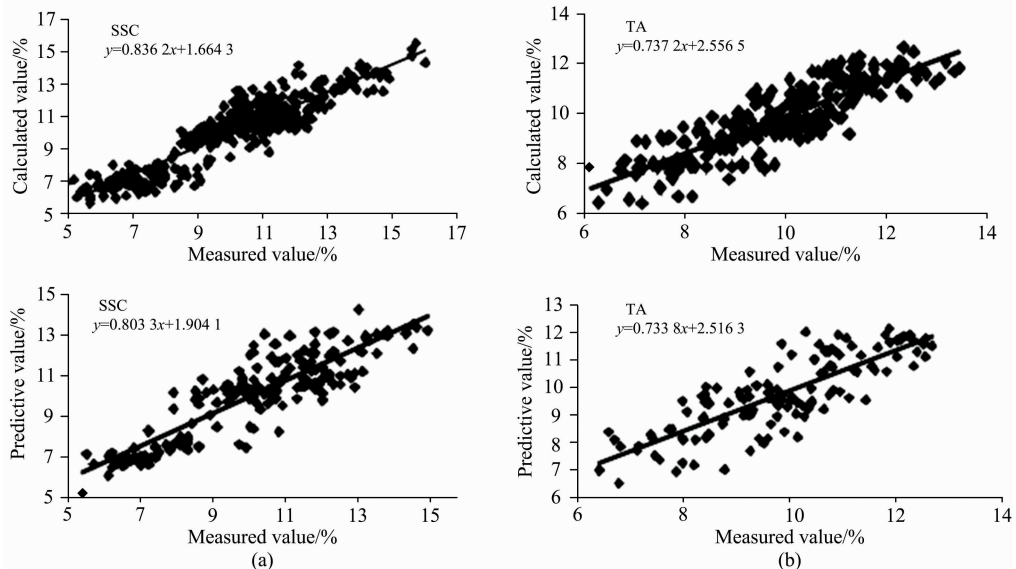


图 3 PLS 模型对李果实 (a)SSC、(b)TA 的定量结果

Fig. 3 Quantitative results of PLS model for (a) SSC and (b) TA in plum samples

从总体上提高模型性能的效果；而消除常数偏移量、减去一条直线、二阶微分结合 25 点平滑、一阶微分结合减去一条直线(平滑点数: 5)以及一阶微分结合 MSC 校正(平滑点数: 25)等预处理方法则使 TA 原始光谱模型的性能变差。即 TA 的最佳 PLS 模型是经一阶微分结合 SNV 和 9 点平滑处理后得到的, 见图 3(b)。

(2) 神经网络

采用反向传播人工神经网络(back propagating artificial neural net, BP-ANN)建模算法对李果实各品质指标进行定量分析。首先使用 MATLAB 软件提取李果实所有样品全波段光谱数据的主成分, SSC、TA 前 10 主成分得分的累积百分比分别达到 99.686% 和 98.614%。因而, 采用这两个指标的前 10 主成分代替原始光谱建模是可行的。将各指标的前 10 主成分得分作为 BP-ANN 模型的输入变量, 建立各自的近红外定量分析模型。

构建一个三层的 BP 神经网络, 用以建立李果实 SSC、TA 的定量模型。由于将前 10 主成分得分作为 BP-ANN 模型的输入变量, 因而各模型的输入层节点数均设为 10, 输出层节点数为 1, 其他各项参数见表 5。

表 5 李果实 SSC、TA 的 BP-ANN 定量模型的参数
Table 5 Parameters of the BP-ANN models for SSC, TA in plums

参数	SSC	TA
输入层函数	tansig	tansig
隐含层函数	logsig	logsig
输出层函数	purelin	tansig
训练函数	trainlm	traingdx
显示间隔	50	50
学习速率	0.05	0.05
训练目标	1×10^{-3}	1×10^{-3}

表 6 不同隐层节点数对李果实 SSC 定量模型的影响

Table 6 Influence of nodes in hidden layer on quantitative results of SSC in plum

隐含层节点数	校正集		预测集	
	R_c	RMSEC	R_p	RMSEP
10	0.935 1	0.84	0.719 3	1.15
11	0.945 1	0.82	0.671 2	1.25
12	0.925 4	0.85	0.660 2	1.29
13	0.950 8	0.81	0.755 3	1.15
14	0.955 7	0.81	0.782 0	1.04
15	0.960 4	0.80	0.797 9	1.04
16	0.964 1	0.79	0.854 0	1.02
17	0.964 3	0.79	0.871 9	1.01
18	0.976 7	0.75	0.889 7	0.99
19	0.967 2	0.78	0.713 2	1.16
20	0.965 9	0.78	0.675 9	1.22

表 7 不同隐层节点数对李果实 TA 定量模型的影响

Table 7 Influence of nodes in hidden layer on quantitative results of TA in plum

隐含层节点数	校正集		预测集	
	R_c	RMSEC	R_p	RMSEP
10	0.941 0	0.66	0.806 4	0.88
11	0.947 9	0.65	0.787 7	0.91
12	0.959 1	0.64	0.817 7	0.86
13	0.958 4	0.64	0.800 6	0.91
14	0.962 1	0.63	0.806 0	0.87
15	0.956 7	0.64	0.830 7	0.85
16	0.955 2	0.64	0.795 4	0.91
17	0.956 7	0.64	0.806 2	0.87
18	0.960 8	0.63	0.877 3	0.84
19	0.974 3	0.62	0.897 7	0.83
20	0.968 4	0.63	0.867 6	0.84

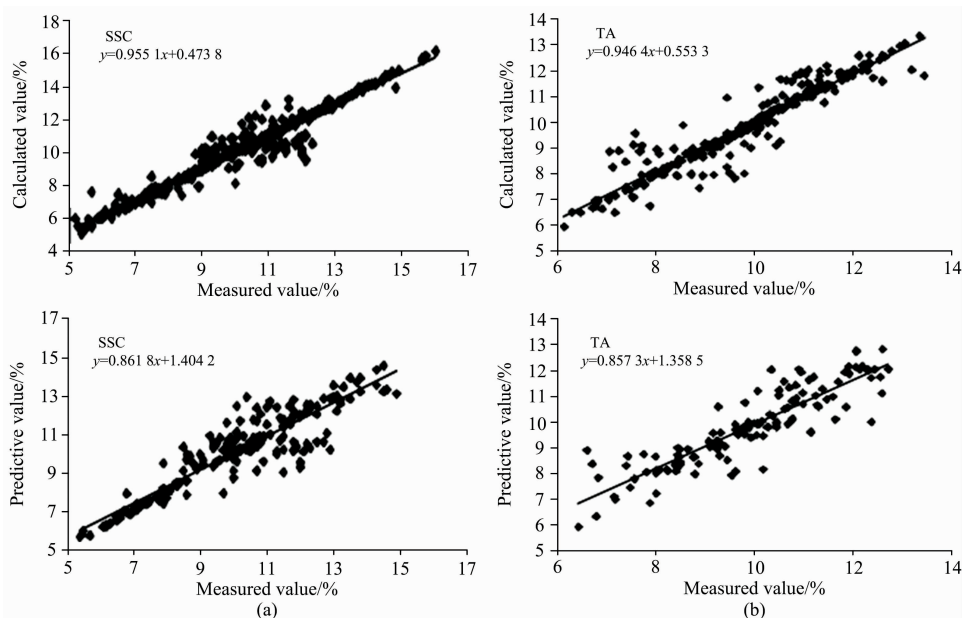


图 4 BP-ANN 模型对李果实 (a)SSC 和 (b)TA 的定量结果

Fig. 4 Quantitative results of BP-ANN model for (a) SSC and (b) TA in plum samples

表6和表7分别比较了不同隐含层节点数所对应的李果实SSC、TA的BP-ANN定量模型的性能。由表7和表8可以看出,当SSC和TA指标的隐含层节点数分别设为18和19时,其BP-ANN模型的性能达到最优。其中,SSC和TA最佳BP-ANN模型校正集和预测集的 R 分别为0.9767, 0.8897和0.9743, 0.8977;二者的校正和预测均方根误差分别是0.75, 0.99和0.62, 0.83。其最佳定量模型分别如图4(a,b)所示。李果实SSC、TA的PLS和BP-ANN定量分析模型的结果比较如表8所示。与PLS定量模型结果相比,BP-ANN模型性能均有较大的改善,即BP-ANN模型对SSC和TA指标的定量分析结果更精准可靠。

表8 李果实SSC、TA的PLS和BP-ANN模型的定量结果
Table 8 Quantitative results of the established models using PLS and BP-ANN methods for SSC, TA in plum

检测指标	建模方法	校正集		预测集	
		R_c	RMSEC	R_p	RMSEP
SSC/%	PLS	0.9144	0.91	0.8785	1.00
	BP-ANN	0.9767	0.75	0.8897	0.99
TA/%	PLS	0.8603	0.80	0.8196	0.86
	BP-ANN	0.9743	0.62	0.8977	0.83

3 结论

建立并优化了基于PLS和BP-ANN算法的李果实的SSC、TA的定量分析模型。

References

- [1] Lan Weijie, Jaillais B, Leca A, et al. Food Chemistry, 2020, 310: 125944.
- [2] Wang J, Wang J, Chen Z, et al. Postharvest Biology and Technology, 2017, 129: 143.
- [3] Shah S, Zeb A, Qureshi W S, et al. Infrared Physics & Technology, 2021, 103639.
- [4] Paloma Andrade Martins Nascimento, et al. Postharvest Biology and Technology, 2016, 111: 345.
- [5] QIANG Feng, WANG Qin-zhi, HE Jian-guo, et al(强 锋, 王芹志, 何建国, 等). Modern Food Science & Technology(现代食品科技), 2017, 212(4): 283.
- [6] Cicooritti R, Paliotta M, Tiziana A, et al. Scientia Horticulturae, 2019, 257.
- [7] JIANG Shui-quan, SUN Tong(江水泉, 孙 通). Food & Machinery(食品与机械), 2020, 36(2): 89.
- [8] Maniwaru P, Nakano K, Ohashi S, et al. Scientia Horticulturae, 2019, 257.
- [9] Ar N H, Purwanto Y A, Budiastra I W et al. IOP Conference Series: Materials Science and Engineering, 2019, 557(1): 9.
- [10] WANG Dong, SUN Jun-peng, YU Shi-feng, et al(王 冬, 孙俊鹏, 于世锋, 等). Journal of Food Safety & Quality(食品安全质量检测学报), 2021, 12(18): 7222.
- [11] Paz P, Sanchez Maria-Teresa, Pérez-Marín D, et al. Journal of Agricultural and Food Chemistry, 2008, 56(8): 2565.
- [12] Louw E D, Theron K I. Postharvest Biology and Technology, 2010, 58: 176.
- [13] BAI Feng-hua, ZHANG Xiao-yu, WANG Yan-wei, et al(白凤华, 张晓瑜, 王艳伟, 等). Food Industry(食品工业), 2018, 39(6): 175.
- [14] Workman J, Weyer J L. Practical Guide to Interpretive Near-Infrared Spectroscopy(近红外光谱解析实用指南). Translated by CHU Xiao-li, XU Yu-peng, TIAN Gao-you(褚小立, 许育鹏, 田高友, 译). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2009. 19.

(1)比较了李果实SSC、TA指标的不同波段的PLS建模效果,最终确定各指标的最佳波段范围分别为4000~8852和4605~6523 cm^{-1} 。采用PLS法分别建立这两个指标的定量分析模型,通过比较不同的光谱预处理方法得出,SSC的PLS模型的最佳光谱预处理方法均为MSC校正,经该方法处理后,SSC最佳PLS模型校正集与预测集的 R 和均方根误差分别为0.9144, 0.8785和0.91, 1.00。采用一阶微分结合SNV和9点平滑的方法预处理后,TA指标的PLS模型效果更佳,校正集的 R 和均方根误差分别为0.8603和0.80,预测集的分别为0.8196和0.86。

(2)提取了李果实SSC、TA样品光谱数据主成分,并基于前10主成分得分建立这两个指标的BP-ANN定量分析模型。经对不同隐含层节点数进行比较后确定各指标的最佳BP-ANN模型。其中,SSC最优BP-ANN模型校正集和预测集的 R 分别为0.9767和0.8897, RMSEC和RMSEP依次为0.75和0.99;TA的BP-ANN模型的相应参数值依次为0.9743, 0.8977, 0.62, 0.83。与采用PLS算法建立的定量模型相比较,BP-ANN模型具有较高的 R_c , R_p 和较低的RMSEC, RMSEP,因此BP-ANN模型对SSC和TA指标的定量分析结果更佳。

本研究采用PLS和BP-ANN两种算法分别建立了安哥诺李果实主要品质指标的定量分析模型,并取得了良好的结果,证实了近红外光谱分析技术无损快速检测李果实品质的可行性,同时也为近红外技术在其他果实检测分析的应用提供参考。

Quantitative Analysis of Soluble Solids and Titratable Acidity Content in Angeleno Plum by Near-Infrared Spectroscopy With BP-ANN and PLS

ZHAO Zhi-lei^{1, 2, 3, 4}, WANG Xue-mei^{1, 2, 3}, LIU Dong-dong^{1, 2, 3}, WANG Yan-wei^{1, 2, 3}, GU Yu-hong⁵, TENG Jia-xin¹, NIU Xiao-ying^{1, 2, 3, 4*}

1. College of Quality and Technical Supervision, Hebei University, Baoding 071002, China
2. National & Local Joint Engineering Research Center of Metrology Instrument and System, Hebei University, Baoding 071002, China
3. Hebei Key Laboratory of Energy Metering and Safety Testing Technology, Hebei University, Baoding 071002, China
4. Institute of Geographical Indications, Hebei University, Baoding 071002, China
5. College of Life Science, Hebei Agricultural University, Baoding 071002, China

Abstract Soluble solid content (SSC) and titratable acidity (TA) are important indexes affecting the fruit quality and the fruit quality grading. Classical destructive detection methods are not suitable for fruit classification by quality. NIRS detection method is fast, easy to operate and can detect fruit quality without damage. In order to achieve non-destructive and rapid determination of SSC and TA in Angeleno plum fruits by near-infrared spectroscopy (NIR), diffuse reflectance spectra of plum fruits were collected by NIR spectrometer, SSC was measured by saccharometer, and TA content was determined by titration. Using leverage and F probability value to eliminate abnormal samples and software optimization combined with a manual screening of spectral bands, eliminating constant offset, subtracting a straight line, standard normal variate (SNV), max-minimum normalization, and multiplicative scatter correction (MSC), first and the second derivative combined smoothing, the first derivative combined minus a straight line and smoothing, and the first derivative combined with SNV or MSC correction. Partial least squares (PLS) and back propagation artificial neural network (BP-ANN) were used to establish the quantitative models of SSC and TA of plum fruit. Results indicated that the best Band ranges of plum fruit SSC and TA are 4 000~8 852 and 4 605~6 523 cm^{-1} respectively. The best spectral preprocessing method of the PLS model of SSC was MSC correction. The best model correction correlation coefficient (R_c) was 0.914 4, the prediction correlation coefficient (R_p) was 0.878 5, the correction root means square error (RMSEC) was 0.91, and the prediction root means square error (RMSEP) was 1.00. After the first order differential combined with SNV and 9-point smoothing, the PLS model of TA was the best, and the R_c , R_p , RMSEC and RMSEP were 0.860 3, 0.819 6, 0.80 and 0.86. The principal components of SSC and TA spectral data of plum fruits were extracted, and the optimal BP-ANN quantitative analysis model of SSC and TA were established based on the first 10 principal component scores. The SSC BP-ANN model's R_c , R_p , RMSEC and RMSEP were 0.976 7, 0.889 7, 0.75 and 0.99. The corresponding parameter values of the BP-ANN model of TA were 0.974 3, 0.897 7, 0.62 and 0.83, respectively. Compared with the quantitative model established by the PLS algorithm, the BP-ANN model has higher R_c and R_p and lower RMSEC and RMSEP than that of the PLS algorithm, so the quantitative analysis results from the BP-ANN model were better than that of the PLS algorithm for SSC and TA indicators.

Keywords Plum fruit; PLS; BP - ANN; Near-infrared spectrum

(Received Aug. 22, 2021; accepted Mar. 17, 2022)

* Corresponding author