

# Research on the Performance of Juicy Peach Sugar Content Detection Model Based on Near Infrared Spectroscopy

Tiantian Guo<sup>1</sup>, Miaomiao Liu<sup>1</sup>, Yue Luo<sup>1</sup>, Wenli Zhang<sup>1</sup>, Shuning Lou<sup>1</sup>, Yiye Sun<sup>2\*</sup>

<sup>1</sup>College of Electrical & Electronic Engineering, Wenzhou University, Wenzhou Zhejiang

<sup>2</sup>Department of Planning Finance, Wenzhou University, Wenzhou Zhejiang

Email: \*16588875@qq.com

Received: Jun. 7<sup>th</sup>, 2020; accepted: Jun. 21<sup>st</sup>, 2020; published: Jun. 28<sup>th</sup>, 2020

## Abstract

The data fusion technology makes full use of the information between different near-infrared spectroscopy instruments, improves the reliability of the peach quantitative detection model, and realizes the rapid non-destructive detection of the peach sugar content. In the experiment, the near-infrared spectrum data of the same peach samples were collected by flame-NIR and USB2000+ spectrum collection system. After that, the true sugar value of all samples was measured by Abel refractometer, and a quantitative detection model of peach sugar content was constructed respectively. On the basis of it, a fusion model of two kinds of spectral data is constructed, and the fusion data is subjected to the elimination of information-less variables and the selection of genetic algorithm variables to construct UVE-PLS and GA-PLS. Results showed that the fusion model developed on two kinds of spectral data performed better than the quantitative model on the basis of the univocal spectral data, and the root mean square errors of cross validation of the optimal GA-PLS fusion model reduced 22.19% and the 33.79% compared with the univocal model, as well as that the root mean squared error of prediction reduced 14.24% and 49.67%. The results show that the data fusion models are superior to the pre-fusion models and have better detection capabilities. They can make full use of the sugar content information of flame-NIR and USB2000+ spectroscopic instruments to achieve quantitative detection of peach sugar content.

## Keywords

Peach, Near Infrared Spectroscopy, Data Fusion, Soluble Solids Content

# 基于近红外光谱技术的水蜜桃糖度检测模型性能研究

郭甜甜<sup>1</sup>, 刘苗苗<sup>1</sup>, 罗越<sup>1</sup>, 张文莉<sup>1</sup>, 娄淑聆<sup>1</sup>, 孙一叶<sup>2\*</sup>

\*通讯作者。

<sup>1</sup>温州大学电气与电子工程学院, 浙江 温州

<sup>2</sup>温州大学财务计划处, 浙江 温州

Email: 16588875@qq.com

收稿日期: 2020年6月7日; 录用日期: 2020年6月21日; 发布日期: 2020年6月28日

## 摘要

通过数据融合技术充分利用不同近红外光谱仪器间的信息, 提高水蜜桃定量检测模型的可靠性。实验通过flame-NIR和USB2000+光谱采集系统分别采集相同水蜜桃样本的近红外光谱数据。之后, 利用阿贝尔折光仪测出所有样本糖度真实值, 分别构建糖度定量模型。在其基础上, 构建两种光谱数据的融合模型, 融合数据经过无信息变量消除(UVE)和遗传算法(GA)变量选择后构建UVE-PLS、GA-PLS。结果显示, 相比于未进行数据融合构建的糖度定量预测模型, 最优融合模型GA-PLS交叉验证均方根误差减少了22.19%、33.79%, 均方根误差分别减少了14.24%、49.67%。结果表明, 数据融合模型能充分利用flame-NIR和USB2000+光谱仪器的糖度信息, 具有更好的检测能力。

## 关键词

水蜜桃, 近红外光谱, 数据融合, 可溶性固形物

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

我国桃子已有 3000 多年的孕育历史, 种类繁多。因此, 在采摘和销售过程中极易混合, 从而影响水果质量, 严重降低了产品的竞争力。糖度作为评估水果风味的一个重要指标, 其在一定程度决定了水果质量的好坏。因此, 依据糖度对采摘后的桃子进行等级分类, 能提高水果的竞争力。

当前多通过人的经验或感官鉴别水果糖度, 这些方法破坏了水果的完整性, 检测后无法再次销售。因此, 探究无损检测对实现水果的大规模检测具有重要意义。目前, 常用的水果无损检测方法有高光谱成像[1]、拉曼光谱[2]、近红外光谱[3]等。其中, 近红外具有检测速度快, 无损、成本低等特点[4], 在水果检测方面有广泛应用, 例如梨[5]、橙子[6]、苹果[7]等。

近红外光谱技术是通过分析物质对光的吸收、反射等获取信息, 光谱仪器的差异会导致结果大相径庭。在以往的研究中大多使用单一仪器对水蜜桃糖度进行研究, 鲜少结合多种仪器进行分析。单一仪器检测虽实现水蜜桃糖度含量检测, 但获取的光谱信息存在差异, 建立的水蜜桃糖度定量检测模型的预测效果并不理想。本研究探究两台近红外光谱仪器采集的相同水蜜桃光谱数据的检测模型性能, 另外通过数据融合技术和变量筛选方法进一步提高模型检测性能, 以实现水蜜桃的等级分类。

## 2. 实验步骤

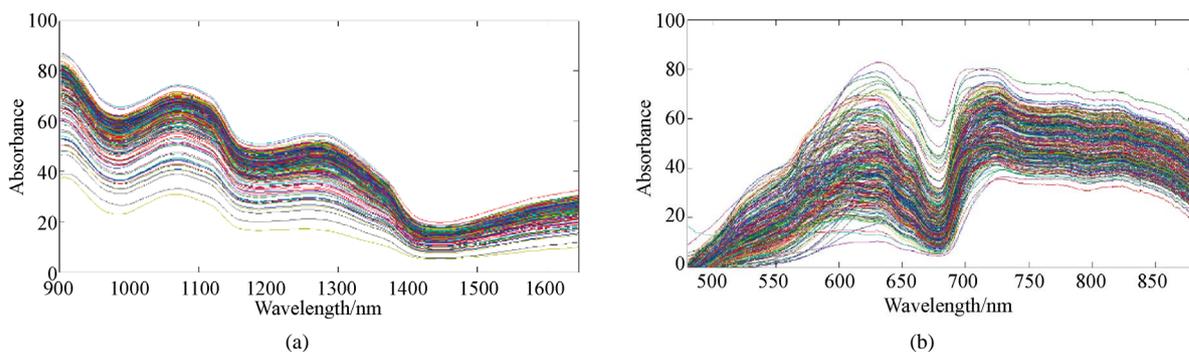
### 2.1. 样本准备

实验以温州某果园的采摘的水蜜桃为实验对象, 挑选出 270 个外形无损坏, 大小相似的水蜜桃作为

实验样本。实验前对全部样本清洁、编号，之后将全部样本均放置在空调室内 6 h，使所有样本和室内温度保持一致。

## 2.2. 光谱采集

本研究采用海洋光学近红外光谱检测系统采集水蜜桃的光谱数据，系统主要由光谱仪(型号为 flame-NIR、USB2000+)、光源、光纤、电源(5 v 直流电源)、电脑等组成。光源照射水蜜桃样品周围，部分光经样品的反射，通过光纤传输到光谱仪，光谱仪将光信号转换后发送给电脑。flame-NIR 光谱仪的扫描波段范围在 900~1700 nm，分辨率为 2 nm。USB2000+光谱仪检测器的控制范围在 200~1100 nm，积分时间的范围为 1 ms~65 s。其中检测器的波段范围、分辨率等均不相同，这种系统差别导致采集的光谱形成不同位置的波峰。按照编号大小依次利用两种光谱采集系统获取相同水蜜桃样本的近红外光谱。根据峰的位置分布，去除掉比较平坦的光谱波段。最终，保留 flame-NIR 光谱仪波段范围在 902.59~1648.61 nm 光谱，保留 USB2000+光谱仪波段范围在 472.02~888.12 nm 光谱，采集光谱图如图 1 所示。



**Figure 1.** Spectra collected by two spectrometers. (a) Flame-NIR spectrometer collected spectrum; (b) USB2000+ spectrometer collected spectrum

**图 1.** 两台光谱仪采集光谱图。(a) flame-NIR 光谱仪采集光谱；(b) USB2000+光谱仪采集光谱

## 2.3. 糖度测量

阿贝尔折光仪广泛地用于液体的定量检测，可以准确的测量出水果中糖度含量的真实值，理论基础是光的折射。近红外光谱数据采集后，分别取水蜜桃不同光谱获取位置的果肉，取出汁液，通过阿贝尔折光仪读出水蜜桃糖度的精确值，水蜜桃样品均检测三次，取平均值作为最终结果。根据编号将光谱数据和真实糖度值一一对应，结合偏最小二乘，建立水蜜桃糖度的定量检测模型，预测未知水蜜桃样本的糖度含量。糖度统计结果如表 1。

**Table 1.** Distribution of sugar content in peach samples

**表 1.** 水蜜桃样品中糖度含量的分布

光谱仪		样本数(个)	糖度分布/°Brix	平均值/°Brix
flame-NIR 光谱仪	训练集	178	6.4~15	10.6
	预测集	88	6.4~15.5	10.9
USB2000+光谱仪	建模集	176	6.4~14.2	10.7
	预测集	87	6.4~15.5	10.7

## 2.4. 数据融合

为了可以提供更多有关样本的信息，获得更高的准确率，将不同来源之间的数据进行融合。通过使

用不同仪器之间的独特信息来增强技术之间的协同作用，以便最终能够获得更好的预测结果。但同时数据融合后易增加光谱维度，对融合后的数据进行变量选择可以减少模型的复杂度。研究中将 flame-NIR 近红外光谱仪和 USB2000+光谱仪分别采集的光谱数据做了数据合并。使两个光谱仪的数据简单地按样本连接到一个矩阵中，形成一个 259 行  $\times$  1444 列的矩阵。该矩阵的行与全部水蜜桃样本个数相等，列与 flame-NIR 和 USB2000+两台光谱仪所测量信号的变量和相等。数据融合示意图如图 2 所示。

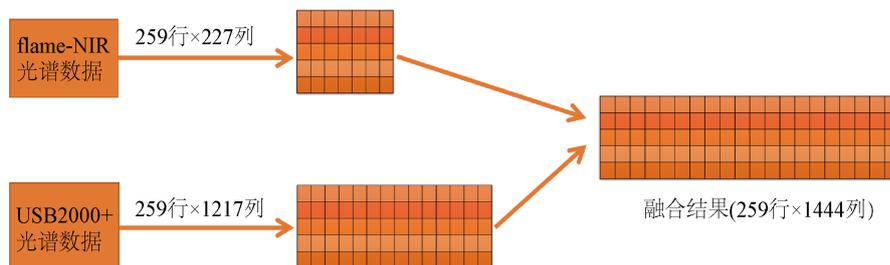


Figure 2. Schematic diagram of the fusion of spectral data collected by two spectrometers  
图 2. 两台光谱仪采集光谱数据融合示意图

## 2.5. 模型评价标准

常用的模型评估指标包括交叉验证均方根误差(Root-Mean-Square Error of cross Validation, RMSECV)、预测均方根误差(Root-Mean-Square Error of Prediction, RMSEP)、训练集相关系数( $R_{cv}$ )、预测集相关系数( $R_p$ )。其中，RMSECV 和 RMSEP 越小， $R$  越大，说明模型的预测能力和相关性强。

## 3. 结果与分析

### 3.1. 原始光谱数据模型分析

分别对 flame-NIR 光谱数据、USB2000+光谱数据、融合光谱数据建立水蜜桃糖度定量检测模型，即  $f_1$ 、 $f_2$ 、 $f_3$ ，数据结果如表 2 所示。从表中可以看出，基于 flame-NIR 和 USB2000+光谱仪建立的模型  $f_1$  和  $f_2$  均在一定程度上实现了水蜜桃糖度定量检测，训练集的相关系数分别为 0.8216、0.7258，预测集均方根误差分别为 0.9946、1.3489，两种光谱检测系统具有不同的预测能力。为了使模型具有更好的可靠性，将两种光谱检测系统测量的光谱数据做了融合，基于融合后的数据建立水蜜桃糖度的定量检测模型，融合后建立的定量检测模型的交叉验证均方根误差相比于  $f_1$ 、 $f_2$  分别减少了 3.24%、14.84%，均方根误差相比于  $f_1$  增加了 9.1%，相比于  $f_2$  减少了 26.33%。融合模型的预测误差低于  $f_1$ ，但仍高于  $f_2$ ，融合数据预测效果并未优于 flame-NIR 光谱数据的建模效果。这可能是由于数据融合后，虽然样本总数未改变，但增加了光谱变量长度，导致模型数据量较大，所以出现单一模型的数据源优于融合模型数据源。

Table 2. PLS prediction model based on single spectrum and fusion spectrum data  
表 2. 基于单一光谱及其融合光谱数据建立的 PLS 预测模型

模型	主成分数	训练集		预测集	
		RMSECV	$R_{cv}$	RMSEP	$R_p$
$f_1$	10	0.9443	0.8216	0.9946	0.8343
$f_2$	10	1.0603	0.7258	1.3489	0.7269
$f_3$	14	0.9119	0.8397	1.0856	0.7897

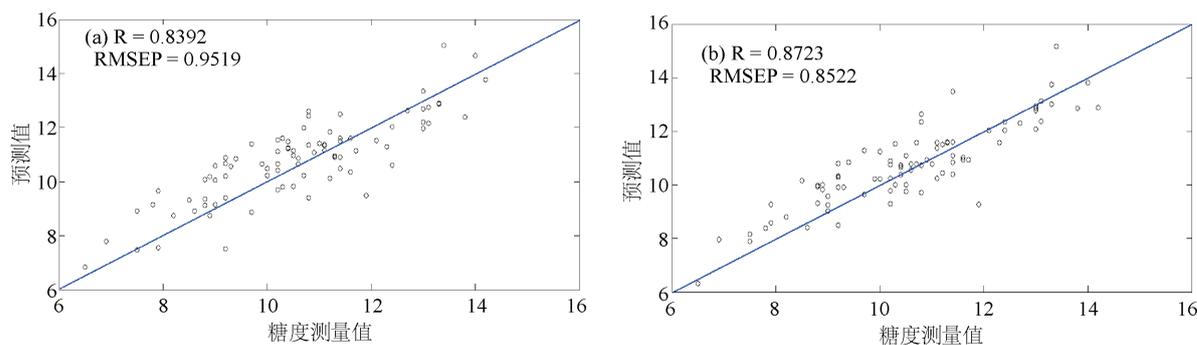
### 3.2. 变量选择模型分析

由于考虑到融合后的数据形成的矩阵较大, 具有很大的维度, 并包含来自不同仪器的冗余信息, 相比于其他未合并的矩阵, 可能会降低数据的性能。因此, 对融合后的数据进行变量筛选, 以便可以找出最优的建模效果, 将其用于水蜜桃糖度定量检测。融合光谱数据经过无信息变量消除法和遗传算法变量选择后构建模型 UVE-PLS、GA-PLS, 数据结果如表 3 所示。从表中可以看出, UVE-PLS 模型用了全部光谱变量的 8.03%, 相关系数  $R_{cv}$  达到 0.8734, 交叉验证误差相比于  $f_1$ 、 $f_2$ 、 $f_3$ , 分别减少了 12.65%、24.25%、9.41%, 均方根误差分别减少了 4.27%、39.7%、13.37%。经过 GA 变量选择后构建的 GA-PLS 模型用了全部光谱变量的 6.44%, 相关系数  $R_{cv}$  达到 0.9028, 交叉验证误差相比于  $f_1$ 、 $f_2$ 、 $f_3$ , 分别减少了 22.19%、33.79%、18.95%, 均方根误差减少了 14.24%、49.67%、23.34%。模型 UVE-PLS 和 GA-PLS 具有比单一模型更好的相关性和预测能力, 且其均优于未进行变量选择的融合模型  $f_3$ 。图 3 为经过变量筛选后构建 UVE-PLS 和 GA-PLS 模型的水蜜桃糖度的测量值和预测值的关系图。

**Table 3.** The results of fusion data variables after selection of UVE and GA variables

**表 3.** 融合数据变量经过 UVE 和 GA 变量选择后结果

模型	主成分数	变量数	训练集		预测集	
			RMSECV	$R_{cv}$	RMSEP	$R_p$
UVE-PLS	13	116	0.8178	0.8734	0.9519	0.8392
GA-PLS	19	93	0.7224	0.9028	0.8522	0.8723



**Figure 3.** The relationship between the measured value and the predicted value of the model after variable selection. (a) UVE-PLS model; (b) GA-PLS model

**图 3.** 变量筛选后模型测量值与预测值关系图。(a) UVE-PLS 模型; (b) GA-PLS 模型

## 4. 结论

实验利用不同近红外光谱仪器实现了水蜜桃定量无损检测, 之后结合不同仪器光谱数据的融合技术用于提高水蜜桃的糖度定量检测模型性能, 与单一技术相比, 不同仪器输出融合之后, 虽优于 USB2000+ 光谱数据建模效果, 但融合后的数据量较大, 出现了 flame-NIR 光谱数据优于融合后的数据。因此, 又对融合后的数据做了变量选择, 建立的融合模型 UVE-PLS 和 GA-PLS, 改善了单个结果, 具有更可靠的预测能力。数据融合为水蜜桃等级划分提供了技术支持, 具有重要的研究价值和经济意义。

## 基金项目

温州大学开放实验室一般项目(JW19SK35); 温州大学大学生创新创业训练计划项目(JWSC2019112)。

## 参考文献

- [1] 马本学, 应义斌, 饶秀勤, 等. 高光谱成像在水果内部品质无损检测中的研究进展[J]. 光谱学与光谱分析, 2009, 29(6): 1611-1615.
- [2] 王涛, 裘正军, 张卫正, 等. 基于拉曼光谱技术的枇杷果实  $\beta$ -胡萝卜素含量无损测定研究[J]. 光谱学与光谱分析, 2016, 36(11): 3572-3577.
- [3] 马毅, 汪西原. 基于近红外光谱无损检测的水果品质定量分析与预测[J]. 农业科学研究, 2010, 31(3): 16-20.
- [4] 孟凡坤, 熊刚, 宋世远. 近红外光谱在油料低温性能检测中的应用[J]. 当代化工, 2018(7): 1529-1532.
- [5] Yuan, L.M., Mao, F., Chen, X.J., *et al.* (2020) Non-Invasive Measurements of “Yunhe” Pears by Vis-NIRS Technology Coupled with Deviation Fusion Modeling Approach. *Postharvest Biology and Technology*, **160**, Article No. 111067. <https://doi.org/10.1016/j.postharvbio.2019.111067>
- [6] Yuan, L.M., Sun, L.I., Cai, J.R., *et al.* (2015) A Preliminary Study on Whether the Soluble Solid Content and Acidity of Oranges Predicted by Near Infrared Spectroscopy Meet the Sensory Degustation. *Journal of Food Process Engineering*, **38**, 309-319. <https://doi.org/10.1111/jfpe.12104>
- [7] Bai, Y.H., Xiong, Y.J., Huang, J.C., *et al.* (2019) Accurate Prediction of Soluble Solid Content of Apples from Multiple Geographical Regions by Combining Deep Learning with Spectral Fingerprint Features. *Postharvest Biology and Technology*, 156, Article No. 110943. <https://doi.org/10.1016/j.postharvbio.2019.110943>