

文章编号: 1005-5630(2020)04-0007-07

DOI: 10.3969/j.issn.1005-5630.2020.04.002

# 基于近红外光谱技术的马铃薯叶片 含水率高效预测

于旭峰, 李红梅, 卓伟, 冯洁

(云南师范大学物理与电子信息学院, 云南昆明 650000)

**摘要:** 提出了运用近红外光谱技术检测新鲜马铃薯叶片中含水量的方法, 并通过预测结果和运算量的对比得出一种高效率的预测方法。采集了 900~2100 nm 波段范围内 110 个新鲜马铃薯叶片的光谱反射率信息, 经 SG(Savitzky-Golay)平滑、多元散射校正(MSC)和标准正态变量变换(SNV)3种预处理后, 分别建立偏最小二乘回归(PLSR)模型和 BP神经网络模型, 再运用回归系数(regression coefficients, RC)法在全波段光谱中提取特征波长, 同样经3种预处理后分别建立预测模型。结果表明: 在运用光谱全波段信息构建的模型中, 经多元散射校正(MSC)预处理建立的 BP神经网络模型预测效果最好, 预测集决定系数  $R^2$  为 0.9791, 均方根误差 RMSE 为 0.3723; 在基于特征波长构建的模型中, 经 SG平滑预处理建立的神经网络模型预测效果最优, 预测集决定系数  $R^2$  为 0.9658, 均方根误差 RMSE 为 0.4759; 验证了特征波段结合 BP神经网络建立的模型与全波段建立的模型预测结果相差不大, 因而能够极大地减少运算量, 提高预测效率。

**关键词:** 马铃薯叶片; 含水率; 光谱; 偏最小二乘回归(PLSR); BP神经网络

**中图分类号:** O 433.4 **文献标志码:** A

## Efficient determination of water content in potato leaves based on spectroscopy technology

YU Xufeng, LI Hongmei, ZHUO Wei, FENG Jie

(College of Physics and Electronic Information, Yunnan Normal University, Kunming 650000, China)

**Abstract:** The determination of moisture content in potato leaves using spectral technique was studied in this paper. Spectral signatures of one hundred and ten fresh potato leaves in the wavelengths of 900-2100 nm were acquired by the spectral device. Then, the moisture content was measured by the drying method. The near-infrared reflection spectrum information was corrected by the Savitzky-Golay (SG) smoothing, multiplicative scatter correction (MSC) and standard normal variable (SNV) correction. The quantitative relationship between spectral information and moisture

收稿日期: 2020-03-30

基金项目: 国家大学生创新创业训练计划(201910681027); 校级研究生核心课程建设项目(YH2018-C04)

作者简介: 于旭峰(1996—), 男, 硕士研究生, 研究方向为多光谱成像技术。E-mail: 944468580@qq.com

通信作者: 冯洁(1975—), 女, 教授, 研究方向为多光谱成像技术、颜色科学与技术。E-mail:

fengjie\_yunnan@126.com

was built by partial least squares regression (PLSR) and BP neural network respectively. The effective wavelength was identified by regression coefficients (RC) and corrected by three pretreatment methods. Then the PLSR and BP neural network models were built respectively. The results showed that for full wavelengths-based models, MSC-BP model performed the best with the coefficient of determination ( $R^2$ ) of 0.979 1 and the root mean square error (RMSE) of 0.372 3 in the prediction. For selected wavelengths-based models, it was the SG-BP model that obtained the optimal result. The  $R^2$  value was 0.965 8 and the RMSE value was 0.475 9 in the prediction. This experiment verified that the prediction results of the model established by combining the characteristic band with BP neural network were not different from those of the model established by the whole band, so it could greatly reduce the computation and improve the efficiency.

**Keywords:** potato leaf; moisture content; spectrum; partial least squares regression(PLSR); BP neural network

## 引 言

马铃薯作为全球第四大粮食作物，其产量仅次于小麦、水稻、玉米，在维持全球粮食安全方面发挥着重要作用。在马铃薯植株的生长发育过程中，水分起着至关重要的作用，水分的缺失会影响马铃薯的生长、产量和品质。常用的作物水分测量方法有蒸馏法、烘干法等，虽测量准确、稳定性高，但时间久、能耗大，且要对叶片造成损害<sup>[1-3]</sup>，不能满足快速检测的要求。

高光谱检测技术具有快速、准确、简洁、无损、无接触等优势，近年来被广泛用于农产品检测领域。朱洁等<sup>[4]</sup>利用高光谱信息构建偏最小二乘回归模型对单粒小麦籽水分分布进行预测，预测集相关系数为 0.90，均方根误差为 1.36%。孙红等<sup>[5]</sup>运用高光谱成像实现了对马铃薯叶片含水率的预测及分布可视化，用建立的最优模型进行水分的预测，得到的相关系数为 0.983 2，均方根误差为 2.32%，为检测马铃薯生长状况以及叶片含水率分析提供了新的参考。张筱蕾等<sup>[6]</sup>利用可见/近红外光谱结合特征波长实现了新鲜茶叶叶片含水率的快速检测。Zhou 等<sup>[7]</sup>利用小波分解提取特征波段，建立莴苣叶片含水率偏最小二乘回归预测模型，实现了莴苣叶片含水率定量检测及分布可视化。Mutlu 等<sup>[8]</sup>利用近红外光谱和人工神经网络成功预测了小麦的多个品质参数，得到水分的预测相关系数为 0.92。在这些研究基础上本文运用近红外光谱技术，通过提取特

征波长和建立优化模型对马铃薯叶片含水率进行高效预测。

## 1 实验部分

### 1.1 马铃薯叶片样品采集及含水率测定

2019 年 11 月，在云南师范大学马铃薯育种试验基地，连续 5 天在 13:00–14:00 之间采摘叶片共 110 片。每次采摘完毕，立即对每个样本进行称重，质量记为  $M_1$ ，精度确定到 0.001 g。随即对样本进行高光谱扫描得到样本的光谱数据，然后将叶片放入电热鼓风恒温干燥箱，温度调至 110 ℃，烘干 24 h 至恒重后，连续称取 3 次取平均值，得到样品质量  $M_2$ 。含水率的具体表达式<sup>[9]</sup>如下：

$$M = \frac{M_1 - M_2}{M_1} \times 100\% \quad (1)$$

式中： $M$  为马铃薯样本叶片的含水率(%)； $M_1$  为样本叶片干燥前的质量(g)； $M_2$  为样本叶片干燥后的质量(g)。

### 1.2 马铃薯叶片高光谱数据采集

本文采用 Gaiasorter-Dual 高光谱分选仪，只使用其近红外相机部分，结构如图 1 所示。采集的光谱范围为 947~2 102 nm，为消除基线漂移误差，测量前预热仪器 30 min，曝光时间设为 6 ms，载物台移动速度设为 0.65 cm/s。在采集

光谱信息时关闭室内光源,使样本处在黑暗的环境中,以消除环境光对实验的影响,并且只使用高光谱分选仪自带的光源。

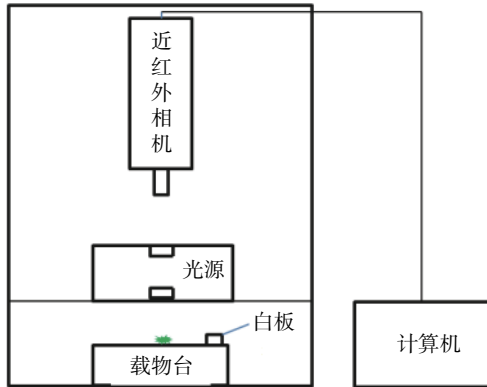


图1 高光谱成像系统

Fig. 1 Schematic of hyperspectral imaging system

在每个高光谱波段下,光谱强度分布不均匀,需要对原始光谱数据进行黑板校正,校正式为

$$R = \frac{I - B}{W - B} \quad (2)$$

式中:  $R$  为马铃薯叶片校正数据;  $I$  为马铃薯叶片的原始光谱数据;  $B$  为黑板;  $W$  为白板。采集完高光谱图像后,提取叶片的高光谱反射率数据,用 Origin 作图,结果如图 2 所示。

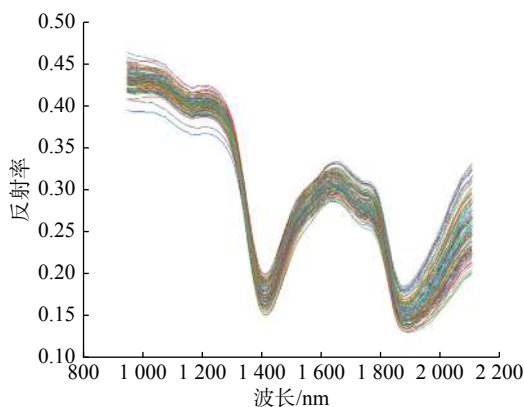


图2 叶片反射光谱曲线

Fig. 2 Reflectance spectra of potato leaves

### 1.3 光谱数据的预处理

对原始数据信息进行预处理的方法有 3 种,即 SG(Savitzky-Golay)平滑、多元散射校正(MSC)和标准正态变量变换(SNV),每种预处理方法都有不同的含义。

平滑是一种有效的去噪方法,本文采用 SG 卷积平滑法提高光谱数据的平滑性,该方法也称为多项式平滑<sup>[10]</sup>,它是运用多项式对移动窗口中的数据点进行多项式最小二乘拟合而得。

MSC 是目前多波长校准建模中常用的一种数据处理方法,经多元散射校正后的光谱数据可以极大地减弱散射效应,加强和成分含量有关的光谱信息<sup>[11]</sup>。

SNV 为标准正态变量变换,是将原始光谱数据与平均光谱数据的差值除以原始光谱的标准偏差,主要是校正由散射造成的样品间的误差。

### 1.4 建模方法

本文采用两种建模方法,分别为偏最小二乘回归、BP 神经网络。

偏最小二乘回归(PLSR)作为运用最普遍的建模方法之一,当变量数多于样本数时,预测效果极为显著<sup>[12-13]</sup>。PLSR 以主成分提取为基础,解决了变量间的自相关和多重共线性,同时对叶片光谱反射率矩阵与含水率矩阵进行主成分分解,然后建立光谱矩阵与叶片含水率矩阵的相关性,构建两者之间的线性回归模型来检测马铃薯叶片的含水率<sup>[5]</sup>。

目前普遍运用的 BP 神经网络是根据误差反向传播算法训练的多层前馈神经网络<sup>[14]</sup>,BP 神经网络可以在不需要揭示描述的输入-输出模式映射关系的前提下,大量地储存和学习这种映射关系。BP 神经网络通常包含输入层、隐含层和输出层,层与层之间常采用全互连接方式,同一层神经元之间无关联。BP 神经网络的实现步骤包括网络构建、训练和预测 3 步,本文将样本按 2:1 的比例随机分为训练集和预测集,有助于提高模型的稳定性。

### 1.5 数据分配及模型评价标准

建模之前首先要把数据分为建模集、验证集、预测集。本文将 110 个样本按照 2:1 的比例随机分为建模集和预测集,然后根据留一法对建模集中的样本进行交互验证,将建模集中的样本同时作为验证集样本,这样做的目的是为了使样本分类更加精确,提高预测模型的稳定性。具体步骤为:利用建模集数据分别建立 PLSR 模型

和 BP 神经网络模型；然后将建模集、验证集、预测集的光谱数据分别代入两个预测模型中，得出预测的含水率结果，再与之相对应的实际含水率相比较；通过计算模型性能的评价参数，比较模型预测效果的质量好坏。

模型性能的评价参数分为建模集决定系数  $R_c^2$  (coefficient of determination in calibration)，验证集决定系数  $R_{cv}^2$  (coefficient of determination in cross-validation)，预测集决定系数  $R_p^2$  (coefficient of determination in prediction)，建模集均方根误差 RMSEC (root mean square error in calibration)，验证集均方根误差 RMSECV (root mean square error in cross-validation)，预测集均方根误差 RMSEP (root mean square error in prediction)。判定一个预测模型性能的优越，需要有较高的  $R_c^2$ 、 $R_{cv}^2$ 、 $R_p^2$  和较小的 RMSEC、RMSECV、RMSEP 值，并且建模集，验证集和预测集的决定系数相差越小，预测效果越好<sup>[6]</sup>。

## 2 结果与分析

### 2.1 含水率统计

实验采用 110 个样本，含水率在 80.00%~92.35% 之间，平均值为 86.82%，标准偏差为 2.579%。将 110 个样本按照 2:1 的比例随机分为建模集和预测集后，每个集合中的最大值、最小值、平均值、标准偏差如表 1 所示。

表 1 马铃薯叶片水分含量统计  
Tab. 1 Descriptive statistics of moisture content in the potato leaves

数据集	样本数	最大值/%	最小值/%	平均值/%	平均偏差/%
总样本	110	92.35	80.00	86.82	2.579
建模集	75	92.35	80.00	86.92	2.575
预测集	35	91.33	80.09	86.59	2.573

### 2.2 基于全波段数据预测结果

为了优化预测效果，需减少噪音、降低基线偏移等因素对预测模型的影响。本文采用 3 种预处理方法对原始数据进行处理，并且根据数据分别建立 PLSR 预测模型和 BP 神经网络预测模型。

#### (1) PLSR 模型建立

将预处理后的数据对应的光谱反射率作为  $X$  变量，叶片的含水率作为  $Y$  变量，构建偏最小二乘回归预测模型，用以预测叶片含水率。结果显示，以光谱全波段数据构建的 PLSR 模型中，采用 SG 平滑预处理后的数据建立的模型，预测效果最好， $R_c^2$  为 0.8907， $R_{cv}^2$  为 0.8656， $R_p^2$  为 0.8493，RMSEC 为 0.8515，RMSECV 为 0.6916，RMSEP 为 0.9988。预测集预测结果对比如图 3 所示，预处理和原始数据模型的预测结果如表 2 所示。

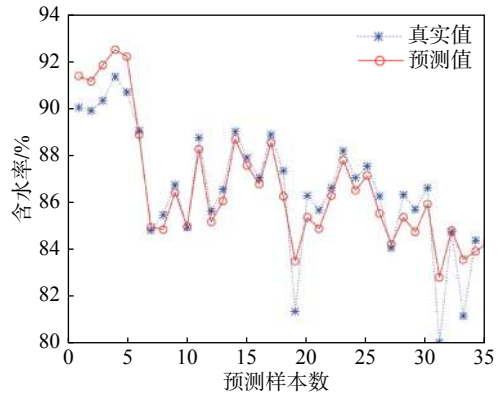


图 3 全波段 PLSR 模型预测集预测结果对比  
Fig. 3 Comparison of prediction results by PLSR models with whole spectra

#### (2) BP 神经网络模型建立

将训练样本和预测样本以 mapminmax 函数为基础作归一化处理。设创建的 BP 神经网络隐含层神经元个数为 9 个，具体网络参数设置为训练次数 1 000 次，训练目标 0.001，学习速率 0.01。将 4 组建模集数据作为训练集分别代入神经网络，经多次训练，分别得出 4 组最优预测集数据。结果显示，以光谱全波段数据建立的 BP 神经网络预测模型中，采用 MSC 预处理后的数据建立的模型，预测效果最好， $R_p^2$  为 0.979 1，RMSEP 为 0.372 3。预测集预测结果对比如图 4 所示，预处理和原始数据模型的预测结果如表 3 所示。

不难发现，基于全波段数据建立的预测模型中，BP 神经网络模型的预测效果要整体优于偏最小二乘回归模型，并且经 MSC 预处理后的数据建立的神经网络模型预测效果最好，并且  $R_p^2$  为 0.979 1，RMSEP 为 0.372 3。

表 2 不同方法的全波段数据 PLSR 模型预测结果  
Tab. 2 All-band data PLSR model prediction results with different methods

处理方法	建模集		验证集		预测集	
	$R_c^2$	RMSEC	$R_{cv}^2$	RMSECV	$R_p^2$	RMSEP
无预处理	0.8907	0.8513	0.8658	0.6911	0.8493	0.9989
SG平滑	0.8907	0.8515	0.8656	0.6916	0.8493	0.9988
MSC	0.8906	0.8513	0.8657	0.6912	0.8492	0.9988
SNV	0.9068	0.7862	0.8725	0.6735	0.8363	1.0412

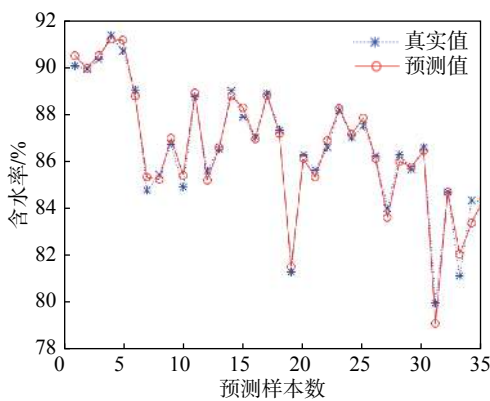


图 4 全波段 BP 神经网络预测集预测结果对比

Fig. 4 Comparison of prediction results by BP neural network models with whole spectra

表 3 不同方法的全波段数据 BP 神经网络模型预测结果  
Tab. 3 All-band data BP neural network mode prediction results with different methods

处理方法	预测集	
	$R_p^2$	RMSEP
无预处理	0.9655	0.4776
SG平滑	0.9747	0.4095
MSC	0.9791	0.3723
SNV	0.8519	0.9902

### 2.3 基于回归系数法提取特征波长

采用回归系数法 (regression coefficients, RC) 提取特征波长可以减少输入变量和简化模型。提取特征波长的目的是将所有原始变量简化为少数关键变量, 而这些新的特征变量包含原始光谱数据的有效信息, 并产生与全波段变量相似的预测结果<sup>[15]</sup>。RC 图中, 波长点的绝对值越大, 此波长对模型的影响越大, 所以这些波长都处在波峰

或者波谷位置<sup>[16]</sup>。

本文通过 RC 法从全波段光谱数据中筛选出 9 个相关性最高的特征波长, 分别为 1 163.5 nm、1 251.2 nm、1 453.8 nm、1 575.3 nm、1 609.0 nm、1 663.0 nm、1 750.8 nm、1 865.5 nm、2 020.8 nm。与全波段数据相比, 特征波长对应的光谱数据量减去了 94.8%, 极大地提高了建模效率。

### 2.4 基于特征波长数据预测结果

将提取出的特征波长对应的光谱反射率分别经 3 种预处理方法处理后作为新的  $X$  变量再次与含水率建立新的预测模型。结果显示, 基于特征波长数据建立的偏最小二乘回归模型中, 采用 SG 平滑预处理后的数据建立的模型, 预测效果最好, 且  $R_c^2$  为 0.877 2,  $R_{cv}^2$  为 0.852 8,  $R_p^2$  为 0.849 2, RMSEC 为 0.902 4, RMSECV 为 0.723 8, RMSEP 为 0.999 1。基于特征波长数据建立的 BP 神经网络预测模型中, 同样为经 SG 平滑预处理后的数据, 建模预测效果最好, 且  $R_p^2$  为 0.965 8, RMSEP 为 0.475 9, 预测结果对比如图 5、图 6 所示。预处理和原始数据模型的预测结果如表 4、表 5 所示。

由此看出, 基于特征波段数据建立的预测模型中, BP 神经网络模型的预测效果同样要整体优于偏最小二乘回归模型, 并且经 SG 平滑预处理后的数据建立的神经网络模型预测结果最好, 且  $R_p^2$  为 0.965 8, RMSEP 为 0.475 9。

### 2.5 结果对比

综合上述试验数据, 可以得出:

(1) 以全波段和特征波段建立的模型预测效果作为比较, 两种处理方法的差别不大, 但是基

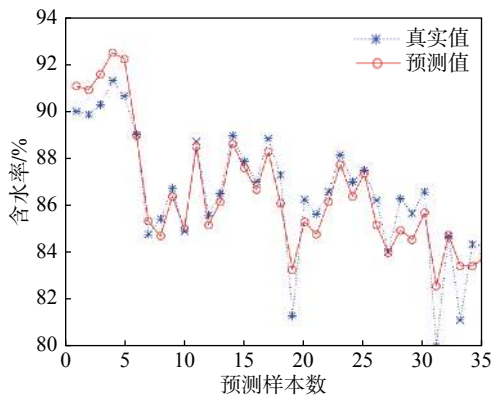


图 5 特征波段 PLSR 模型预测集预测结果对比

Fig. 5 Comparison of prediction results by PLSR models with extracted spectra

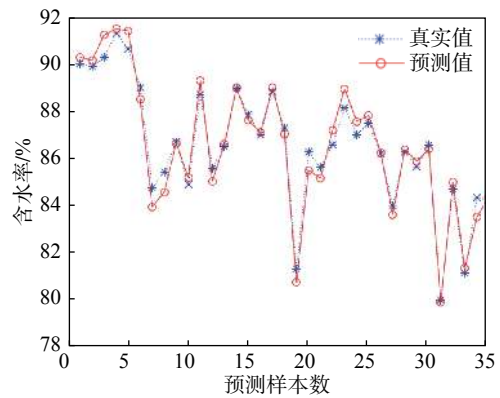


图 6 特征波段 BP 神经网络预测集预测结果

Fig. 6 Prediction results by BP neural network models with extracted spectra

表 4 不同方法特征波段数据 PLSR 模型预测结果  
Tab. 4 Prediction results of different methods feature band PLSR model

处理方法	建模集		验证集		预测集	
	$R_c^2$	RMSEC	$R_{cv}^2$	RMSECV	$R_p^2$	RMSEP
无预处理	0.8783	0.8982	0.8407	0.7529	0.8468	1.0070
SG平滑	0.8772	0.9024	0.8528	0.7238	0.8492	0.9991
MSC	0.8212	1.1788	0.8036	0.8575	0.7745	0.9899
SNV	0.8164	1.1035	0.7624	0.9196	0.7520	1.2813

表 5 不同方法特征波段数据 BP 神经网络模型预测结果  
Tab. 5 Prediction results of BP neural network models with different methods

处理方法	预测集	
	$R_p^2$	RMSEP
无预处理	0.9170	0.7411
SG平滑	0.9658	0.4759
MSC	0.8351	1.0498
SNV	0.7660	1.2446

于全波段数据建立的模型预测效果要整体优于特征波段模型；

(2)以 PLSR 模型和 BP 神经网络预测模型预测效果作为比较,相同数据,基于 BP 神经网络建立的模型预测效果要优于偏最小二乘回归模型；

(3)以预处理方法的处理效果作为比较,全波段 PLSR 模型中,基于 SG 平滑预处理后的数

据构建的预测模型预测效果最好,全波段 BP 神经网络预测模型中,基于 MSC 预处理后的数据构建的预测模型预测效果最好。特征波段 PLSR 模型和 BP 神经网络中,二者都以经 SG 平滑预处理后的数据建立的模型预测效果最优。

由此可见,虽是基于 MSC 预处理后的全波段数据构建的 BP 神经网络模型预测效果最优,但其运算量过大,而基于 SG 处理的特征光谱建立的神经网络模型,特征波段数量极少,可以极大减少运算量,缩短运算时间,提高效率,并且预测效果与之相差不大,因此是一种高效的预测方式。

### 3 结 论

为了实现通过光谱数据预测马铃薯叶片含水率,本文采集了 900~2100 nm 波段范围内 110 个新鲜马铃薯叶片的光谱反射率信息。首先基于 3 种不同预处理方法处理后,分别对叶片的光谱数据建立 PLSR 模型和 BP 神经网络预测模

型;然后运用RC回归系数法提取9个特征波长,同样经过3种预处理方法后,分别建立PLSR模型和BP神经网络模型。

与孙红<sup>[5]</sup>等人利用240个数据、15个特征波段建模、最优预测系数高达0.9832的结果相比,本文虽然特征波段建模预测中的最优预测系数只达到了0.9658,但只采用了75个数据、9个特征波段建模,在取得不错预测效果的同时,大大减少了运算量,达到了高效预测的目的。

#### 参考文献:

- [1] 侯贤清,李荣,何文寿,等.保水剂施用量对土壤水分利用及马铃薯生长的影响[J].浙江大学学报(农业与生命科学版),2015,41(5):558-566.
- [2] 薛俊武,任稳江,严昌荣.覆膜和垄作对黄土高原马铃薯产量及水分利用效率的影响[J].中国农业气象,2014,35(1):74-79.
- [3] 秦军红,陈有君,周长艳,等.膜下滴灌灌溉频率对马铃薯生长、产量及水分利用率的影响[J].中国生态农业学报,2013,21(7):824-830.
- [4] 朱洁,黄华,祝诗平,等.基于高光谱的单粒小麦籽水分分布预测[J].中国粮油学报,2019,34(8):133-138.
- [5] 孙红,刘宁,吴莉,等.高光谱成像的马铃薯叶片含水率分布可视化[J].光谱学与光谱分析,2019,39(3):910-916.
- [6] 张筱蕾,夏威,唐善虎.基于可见/近红外光谱技术的新鲜茶叶叶片含水率快速测定[J].西南民族大学学报(自然科学版),2018,44(4):352-357.
- [7] ZHOU X, SUN J, MAO H P, et al. Visualization research of moisture content in leaf lettuce leaves based on WT-PLSR and hyperspectral imaging technology[J]. *Journal of Food Process Engineering*, 2018, 41(2): e12647.
- [8] MUTLU A C, BOYACI I H, GENIS H E, et al. Prediction of wheat quality parameters using near-infrared spectroscopy and artificial neural networks[J]. *European Food Research and Technology*, 2011, 233(2): 267-274.
- [9] 戴春霞,刘芳,葛晓峰.基于高光谱技术的茶鲜叶含水率检测与分析[J].茶叶科学,2018,38(3):281-286.
- [10] 谢传奇.基于高光谱成像技术的茄子叶片色差值检测和早疫病识别方法研究[D].杭州:浙江大学,2015.
- [11] WILLIAMS P, NORRIS K. Near-infrared technology in the agricultural and food industries[M]. St. Paul, MN: the American Association of Cereal Chemists, 1987.
- [12] ZHANG X L, HE Y. Rapid estimation of seed yield using hyperspectral images of oilseed rape leaves[J]. *Industrial Crops and Products*, 2013, 42: 416-420.
- [13] XIE C Q, LI X L, NIE P C, et al. Application of time series hyperspectral imaging (TS-HSI) for determining water content within tea leaves during drying[J]. *Transactions of the ASABE*, 2013, 56(6): 1431-1440.
- [14] 朱文学,孙淑红,陈鹏涛,等.基于BP神经网络的牡丹花热风干燥含水率预测模型[J].农业机械学报,2011,42(8):128-130,137.
- [15] KAMRUZZAMAN M, ELMASRY G, SUN D W, et al. Application of NIR hyperspectral imaging for discrimination of lamb muscles[J]. *Journal of Food Engineering*, 2011, 104(3): 332-340.
- [16] ELMASRY G, SUN D W, ALLEN P. Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef[J]. *Journal of Food Engineering*, 2012, 110(1): 127-140.

(编辑:刘铁英)