

## 基于近红外光谱技术的空苞山核桃快速识别

俞储泽, 翁定康, 曹烁森, 孙通\*

(浙江农林大学光机电工程学院 杭州 311300)

**摘要** 空苞山核桃是指果实没有种仁或者发育受阻的山核桃,严重影响山核桃产品品质。为实现空苞山核桃的快速无损识别,利用 2 种近红外检测装置在 200~1 160 nm 波长范围采集带壳山核桃样本的光谱,采用 8 种预处理方法进行光谱预处理,利用竞争自适应重加权采样(CARS)方法筛选空苞山核桃的特征波长变量,最后应用线性判别分析(LDA)、二次判别分析(QDA)和马氏距离判别分析(MDA)建立空苞和正常山核桃的分类模型。结果表明,使用检测装置 1 所建立的空苞山核桃分类模型性能优于检测装置 2 的分类模型,经多元散射校正(MSC)预处理后建立的分类模型的识别结果最好,LDA、QDA 及 MDA 模型的特异性、敏感性和正确率均为 1,优于其它预处理方法建立的分类模型。经 CARS 变量筛选后,建模所用的光谱变量数目大大减少,有效简化了分类模型,而模型性能仍与全波长模型性能持平。本文为空苞山核桃的快速、无损识别提供了一种可行的方法。

**关键词** 山核桃; 近红外光谱; 空苞; 竞争自适应重加权采样

**文章编号** 1009-7848(2024)02-0292-11 **DOI:** 10.16429/j.1009-7848.2024.02.028

山核桃(*Carya cathayensis* Sarg)是胡桃科山核桃属乔木植物,有 18 个种、3 个变种,广泛分布在亚洲、欧洲和美洲,其果实是我国著名干果<sup>[1-2]</sup>。在我国,浙江临安和安徽宁国是山核桃主产区,年产量接近全国的 90%<sup>[2-3]</sup>。山核桃品质优良,果仁中含有丰富的蛋白质、脂肪和多种微量元素,营养价值高,深受人们的喜爱<sup>[4]</sup>。在山核桃生长过程中,有部分果实没有果仁或者果仁发育受阻,这种现象被称为“空苞现象”,引起“空苞”的原因包括环境不适、授粉不良、营养缺乏和结果过多等<sup>[5]</sup>。根据 GB/T 24307-2009《山核桃产品质量等级》,瘪籽率和半籽粒率 $\leq 1\%$ 且无空籽为特级山核桃,瘪籽率和半籽粒率 $\leq 2\%$ 且无空籽为一级山核桃,瘪籽率和半籽粒率 $\leq 3\%$ 且无空籽为二级山核桃。“空苞”现象的发生严重影响山核桃产品的品质。有效剔除空苞山核桃,将有利于提升山核桃的整体品质和提高其市场售价,保障种植户的收益,最终促进山核桃产业的健康发展。因此,非常有必要对空苞山核桃进行快速、无损、准确识别。

近红外光是波长在 780~2 526 nm 范围的电

磁波,当近红外光照射有机物样品时,会使样品分子从基态跃迁到激发态,从而产生近红外光谱<sup>[6]</sup>。近红外光谱主要反映有机物含氢基团(C-H、N-H、O-H 等)的倍频与合频吸收信息<sup>[7]</sup>。在不同的分子结构中,同一基团或不同基团的近红外吸收峰位置与强度均有所差别,因此近红外光谱可以反映有机物分子结构与含量等信息<sup>[8]</sup>。因快速、无损、稳定等优点,近红外光谱广泛应用于蔬菜、水果、肉类等农产品的品质检测与分级<sup>[9-12]</sup>。近年来,近红外光谱分析技术也被应用在坚果内部缺陷检测方面。蒋大鹏等<sup>[13]</sup>采用近红外光谱技术对正常和霉变松子进行检测,所建立的分类模型的准确率为 100%。Hu 等<sup>[14]</sup>利用近红外光谱技术对新鲜板栗的品质进行研究,发现线性判别分析(Linear discriminant analysis, LDA)模型对正常板栗和霉变板栗的识别效果最佳,分类模型的验证集准确率为 96.37%。郑剑等<sup>[15]</sup>对褐变和正常板栗进行近红外光谱检测,采用变量选择方法筛选特征波长变量,并建立板栗褐变的分类模型,其模型正确率可达 0.88。Rogel-Castillo 等<sup>[16]</sup>应用 PLS-DA 模型对隐性损伤和正常杏仁进行分类,结果表明其识别错误率在 8.2%~9.2%之间。Pannico 等<sup>[17]</sup>利用近红外光谱技术对缺陷和正常榛子进行识别,分类模型的决定系数为 0.89,交互验证均方根误差为 0.88。Moscetti 等<sup>[18]</sup>通过近红外光谱技术对虫害板

收稿日期: 2023-02-03

基金项目: 浙江省重点研发项目(2020C02019);浙江省属高校基本科研业务费专项资金项目(2021TD002)

第一作者: 俞储泽,男,硕士生

通信作者: 孙通 E-mail: suntong980@163.com

栗和正常板栗进行分类,结果表明分类错误率为 8.41%,AUC 值为 0.952。由上可知,国内外研究者大多将近红外光谱技术应用于杏仁、板栗、松子等坚果的缺陷和病害检测,而对山核桃内部缺陷的检测尚未见研究报道。

本研究以完整带壳山核桃为对象,利用两种近红外光谱检测装置采集山核桃的光谱,采用 8 种预处理方法进行光谱预处理,利用 CARS 方法筛选特征波长变量,采用 3 种判别分析方法空苞山核桃识别模型,并比较不同预处理、近红外检测方式及判别方法等所建立的识别模型性能,以确定较优的空苞山核桃识别模型,为空苞山核桃的快速、无损识别提供方法与技术支持。

## 1 材料与方法

### 1.1 试验材料

山核桃样本采购于杭州临安,随机选择大小均匀、外观正常的 469 个山核桃作为试验样本,依次标号并保存于 4~6 °C 冷库中直到试验使用。光谱采集前,将山核桃样本从冷库中取出并在室温环境中放置 12 h,以保证山核桃温度与室温一致。光谱采集完成后,将全部山核桃样本进行手工破壳,破壳后的山核桃壳和山核桃仁分别置于 A4 打印纸上观察是否存在空苞,将空苞山核桃样本记为类别 A,正常山核桃样本记为类别 B。

样本划分的方法有随机抽样法、kennard-stone 法、光谱-理化值共生距离法等<sup>[9]</sup>。本试验采用 kennard-stone 分类法对 469 个山核桃样本按照 2:1 划分为校正集和预测集,校正集有 30 个空苞山核桃样本和 277 个正常山核桃样本,用于建立分类模型;预测集有 18 个空苞山核桃样本和 144 个正常山核桃样本,用于验证分类模型。

### 1.2 仪器与设备

AvaSpec-HS1024×122TEC 型微型光纤光谱仪,荷兰 Avantes 公司;FC-UV600-2-ME 型光纤,荷兰 Avantes 公司;FA2004 电子分析天平,上海舜宇恒平科学仪器有限公司;游标卡尺,德国美耐特公司。

### 1.3 光谱采集

试验采用两种近红外光谱检测装置采集山核桃样本的光谱,两种检测装置的原理示意图如图

1 所示。检测装置 1 中,2 盏卤素灯分别位于山核桃样本的两侧,三者中心位于同一水平线,光源总功率为 300 W。检测装置 2 中,4 盏卤素灯间隔 90 °C 环绕于山核桃样本四周,五者中心处于同一水平面,光源总功率为 600 W。

光谱采集前,将山核桃样本从冷库中取出放置在通风干燥处 12 h,使样品的温湿度与实验室环境保持一致,以减少试验误差。此外,为保证所测光谱的稳定性,近红外光谱仪需要预热 30 min。本研究采用 2 种山核桃近红外光谱检测装置以漫透射方式采集山核桃样本的光谱,并以直径 30 mm 的聚四氟乙烯球作为参比。在检测装置 1 中,将山核桃顶端和底部与光源垂直放置,采集 1 次光谱记为位置 1,顺时针旋转 90°再采集 1 次光谱记为位置 2,然后将山核桃顶端和底部与光源平行放置,采集 1 次光谱记为位置 3,取 3 次光谱平均值作为山核桃样本的近红外光谱。在检测装置 2 中,将山核桃顶端和底部与光源呈 45°放置,采集 1 次光谱记为位置 1',顺时针旋转 90°再采集 1 次光谱记为位置 2',取 2 次光谱平均值作为山核桃样本的近红外光谱。对整个带壳山核桃样本的近红外光谱采集,检测装置 1 积分时间为 80 ms,检测装置 2 积分时间为 50 ms,扫描次数和平滑点数均为 1 次和 1。

### 1.4 光谱预处理及特征波长选择

光谱预处理可以消除或降低由光谱仪器、采集环境等因素造成的光谱噪声、基线漂移等影响。预处理方法中,SG 平滑 (Savitzky-golay smoothing)是消除光谱噪声的一种最常用的方法<sup>[20]</sup>;导数 (Derivative) 可以用来校正光谱的基线偏移和漂移,消除基线和其它背景的干扰<sup>[21]</sup>;多元散射校正 (Multiplicative scatter correction, MSC)可以有效消除样品光谱中由光散射引起的干扰<sup>[22]</sup>;标准正态变量变换 (Standard normal variate transformation, SNV)可以用来消除光程变化、颗粒大小及其散射产生的干扰<sup>[23]</sup>;归一化 (Normalize)是将数据按比例缩放以减少分布差异、特征、规模等的影响<sup>[24]</sup>;去趋势 (Detrend)可减少光谱数据采集过程中产生的偏移影响<sup>[25]</sup>;基线校正是一种用来减少偏差的方法<sup>[26]</sup>。本研究采用上述 8 种方法对光谱进行预处理,以尽可能消除光谱噪声。

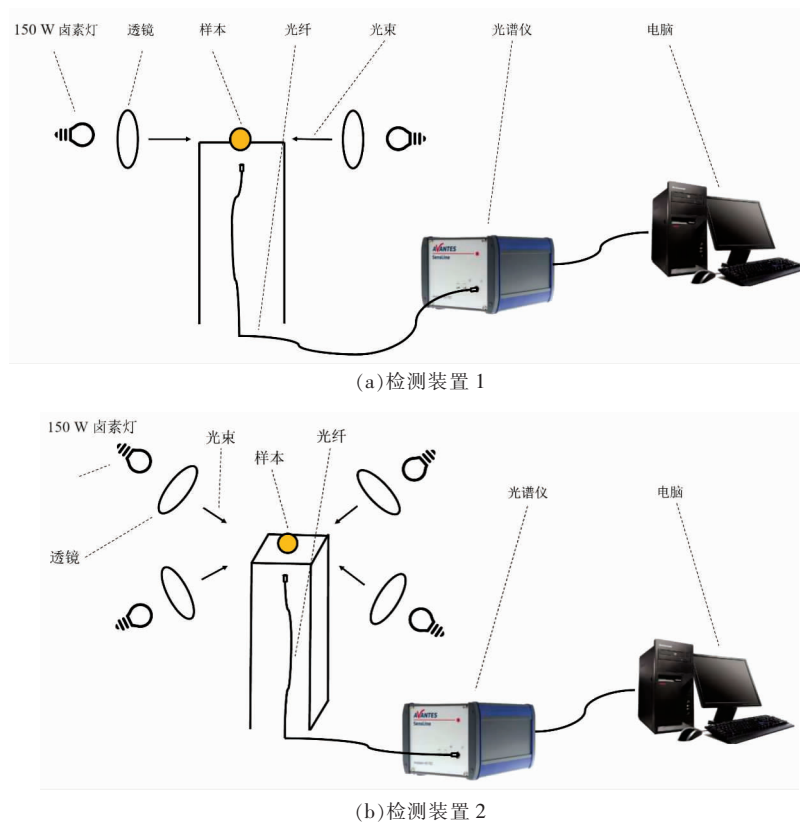


图1 山核桃近红外光谱检测装置示意图

Fig.1 Schematic diagram of *Carya cathayensis* near-infrared spectroscopy detection device

由于近红外光谱数据的变量数目繁多,会存在较多的冗余和无用信息,需要进行特征波长选择,以获取有用信息变量,减少建模时间,提升模型的稳定性。本研究采用CARS方法<sup>[27]</sup>筛选空苞山核桃的特征波长,以剔除与空苞山核桃无关的信息变量,简化山核桃空苞的识别模型,并提高模型的稳定性和精度。在CARS变量选择过程中,蒙特卡罗采样次数为50,采用PLS交叉验证建模,并根据最低错分率确定最优变量子集。

### 1.5 识别模型建立与评价

LDA是一种常用的模式识别方法,其基本原理是将高维光谱数据投射到低维的向量空间,以实现光谱数据压缩,并使同类样本投影点接近<sup>[28]</sup>。二次判别分析(Quadratic discriminant analysis, QDA)和线性判别分析均基于贝叶斯规则,区别在于线性判别分析假设每一种分类的协方差矩阵相同,而二次判别分析中每一种分类的协方差矩阵不同。马氏距离判别分析(Mahalanobis distance discriminant analysis, MDA)使用马氏距离作为相

似性度量。本研究采用上述3种判别分析方法建立空苞与正常山核桃的分类模型,分类模型性能由敏感性、特异性以及总正确率指标进行评价。模型性能评价指标定义如式(1)~(3)。

$$\text{敏感性} = \frac{TP}{TP+FN} \quad (1)$$

$$\text{特异性} = \frac{TN}{TN+FP} \quad (2)$$

$$\text{总正确率} = \frac{TP+TN}{TP+FN+TN+FP} \quad (3)$$

式中,TP——真阳性,表示空苞山核桃被正确判别为空苞山核桃;FN——假阴性,表示空苞山核桃被错误判别为正常山核桃;TN——真阴性,表示正常山核桃被正确判别为正常山核桃;FP——假阳性,表示正常山核桃被错误判别为空苞山核桃。

## 2 结果与分析

### 2.1 山核桃的物理参数及近红外光谱分析

469个山核桃样本的物理参数如表1所示。

由表 1 可知,根据 GB/T 24307-2009《山核桃产品质量等级》,试验所用的山核桃样本的质量、直径等参数均在合理范围内,具有一定的代表性。

从本质来说,可以采用近红外光谱分析技术对正常和空苞山核桃进行分类是由于两类山核桃内部组成成分存在差异,其差异体现在近红外光谱信息中,通过对近红外光谱信息的处理从而实现两者分类。图 2 是 200~1 160 nm 波长范围 2 种检测装置采集的山核桃原始光谱。由图 2 可知,山核桃近红外光谱在 200~520 nm 波长范围的噪声比较大,因此选用 530~1 100 nm 波长范围的山核桃近红外光谱用来分析。山核桃样本的近红外光谱在 800~1 000 nm 波长范围存在明显的波峰和波谷,其中光谱波峰分别位于 930 nm 和 1 000 nm 处,光谱波谷分别位于 880 nm 和 959 nm 左右

表 1 山核桃物理参数

Table 1 *Carya cathayensis* physical parameters

属性	质量/g	直径/mm	高度/mm
范围	1.69~4.92	15.7~21.4	17.5~25.8
均值	2.87	17.6	22.3

处;在 530~800 nm 波长范围存在较为微弱的波峰,分别位于 590 nm 及 780 nm 左右处。从图 2 中还能看出,正常山核桃的吸光度虽普遍高于空苞山核桃,但两者光谱相互重叠,很难直接从形态上加予以区分。因此,需要采用化学计量学方法进行分析处理,以提取有用的光谱差异信息,建立山核桃空苞的分类模型。此外,检测装置 1 与检测装置 2 的山核桃近红外光谱相比,采用检测装置 2 采集的近红外光谱吸光度值比检测装置 1 的略高。

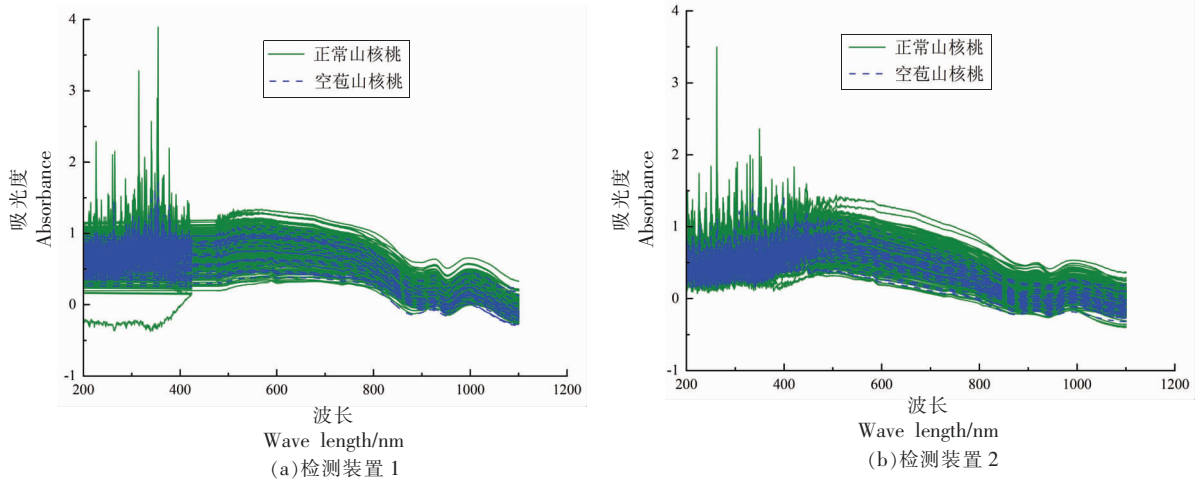


图 2 200~1 100 nm 波长范围的山核桃近红外光谱

Fig.2 Near-infrared spectra of *Carya cathayensis* in the wave length range of 200~1 100 nm

### 2.2 主成分分析

对山核桃原始光谱信息进行主成分分析,其结果如表 2 所示。山核桃光谱的第 1,2,3,4 主成分方差贡献率分别为 84.08%,12.21%,2.65%,0.87%,累计贡献率可达 99.82%,可以代表山核桃的原有光谱信息。

第 1 主成分和第 2 主成分的方差贡献率达到 96.30%,远大于其它主成分,主成分分析不仅可以进行数据降维,同样也能对两类样本间的差距进行初步分析。因此,以第 1 主成分和第 2 主成分为变量绘制散点分布图,结果如图 3 所示。

表 2 山核桃光谱的主成分贡献率

Table 2 Principal component contribution rates of *Carya cathayensis* spectra

主成分	方差贡献率/%	累计贡献率/%
1	84.08	84.08
2	12.21	96.30
3	2.65	98.94
4	0.87	99.82

由图 3 可以看出,正常山核桃和空苞山核桃的主成分散点分布聚集在一起,没有明显的簇拥现象,说明两者之间的主成分散点分布没有明显

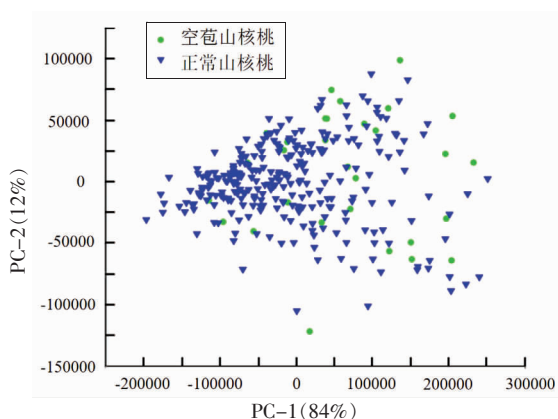


图3 第1和第2主成分散点分布图

Fig.3 1<sup>st</sup> and 2<sup>nd</sup> principal component scatter plots

的差异。用前2个主成分不能对正常和空苞山核桃进行分类,因此需要采用模式识别方法对山核桃光谱作进一步分析。

### 2.3 分类模型建立

**2.3.1 全波长分类模型** 对山核桃样本光谱,采用归一化等8种方法进行光谱预处理,然后应用LDA、QDA和MDA3种方法建立空苞和正常山核桃的分类模型,并利用预测集样本对分类模型精度进行验证,其结果如表3和表4所示。经MSC预处理后的山核桃近红外光谱如图4所示。

从表3和表4可以看出,相较于原始光谱建立的模型,经归一化、一阶导数、标准正态变量和多元散射校正预处理后的光谱建立的模型能提高

对空苞山核桃的识别率,而对正常山核桃的识别率略微降低;经SG平滑和二阶导数预处理后的光谱建立的模型对空苞和正常山核桃的识别率都降低;经去趋势和基线校正预处理后的光谱建立的模型对空苞山核桃和正常山核桃的识别结果影响不明显;经过MSC处理后的光谱建立的模型对空苞和正常山核桃的识别率都大幅上升,3种建模方法所建立的分模型性能均为最优,其模型的敏感性、特异性和总正确率均为1。3种建模方法相比而言,对空苞山核桃识别结果最佳的是MDA模型,QDA模型次之;对正常山核桃识别结果最佳的是LDA模型,QDA模型次之。对两种光谱检测装置而言,使用检测装置1的光谱建立的3种模型精度都高于检测装置2。

为了探究MSC处理对建立的分模型精度的影响,将经MSC处理后的光谱进行可视化,两种检测装置经MSC处理后的光谱如图4所示。从图4中可以看出,相较于原始光谱,经MSC处理后的两种山核桃的光谱曲线更加集中于同种类别,空苞山核桃吸光度值在900~1000 nm波长范围内低于正常山核桃,可以基本区分两者。由此可见,MSC预处理能够有效消除散射的影响,使光谱中与空苞变化相关的有用信息凸显出来。因此,经MSC方法预处理后所建立的山核桃空苞分类模型的性能较优。

表3 不同预处理和建模方法下山核桃空苞的分类模型结果(检测装置1)

Table 3 Classification model results of *Carya cathayensis* with empty bud under different preprocessing and modeling methods (Detection device 1)

建模方法	变量数目	光谱预处理	校正集			预测集		
			敏感性	特异性	总正确率	敏感性	特异性	总正确率
LDA	645	无	0.8333	0.9567	0.9446	0.6667	0.9653	0.9321
	645	归一化	0.8667	0.9314	0.9251	0.7222	0.9583	0.9321
	645	去趋势	0.8333	0.9531	0.9414	0.6667	0.9653	0.9321
	645	基线校正	0.8333	0.9567	0.9446	0.6667	0.9653	0.9321
	645	SG平滑	0.8000	0.9567	0.9414	0.6667	0.9653	0.9321
	645	一阶导数	0.8333	0.9603	0.9479	0.6667	0.9653	0.9321
	645	二阶导数	0.8333	0.9567	0.9446	0.6667	0.9722	0.9383
	645	标准正态变量变换	0.8333	0.9386	0.9283	0.7222	0.9583	0.9321
QDA	645	多元散射校正	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	645	无	0.8667	0.9278	0.9218	0.7222	0.9514	0.9259
	645	归一化	0.9333	0.8809	0.8860	0.7778	0.9236	0.9012



(续表 4)

建模方法	变量数目	光谱预处理	校正集			预测集		
			敏感性	特异性	总正确率	敏感性	特异性	总正确率
MDA	645	无	0.9000	0.8700	0.8730	0.6667	0.9167	0.8889
	645	归一化	0.9000	0.8303	0.8371	0.7778	0.8958	0.8827
	645	去趋势	0.9000	0.8665	0.8697	0.6667	0.9028	0.8765
	645	基线校正	0.8333	0.8592	0.8567	0.6667	0.8889	0.8642
	645	SG平滑	0.8667	0.8700	0.8697	0.6667	0.9167	0.8889
	645	一阶导数	0.9000	0.8664	0.8697	0.7778	0.8681	0.8580
	645	二阶导数	0.8667	0.9458	0.9381	0.7222	0.8958	0.8765
	645	标准正态变量变换	0.9000	0.8484	0.8534	0.7778	0.9028	0.8889
	645	多元散射校正	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

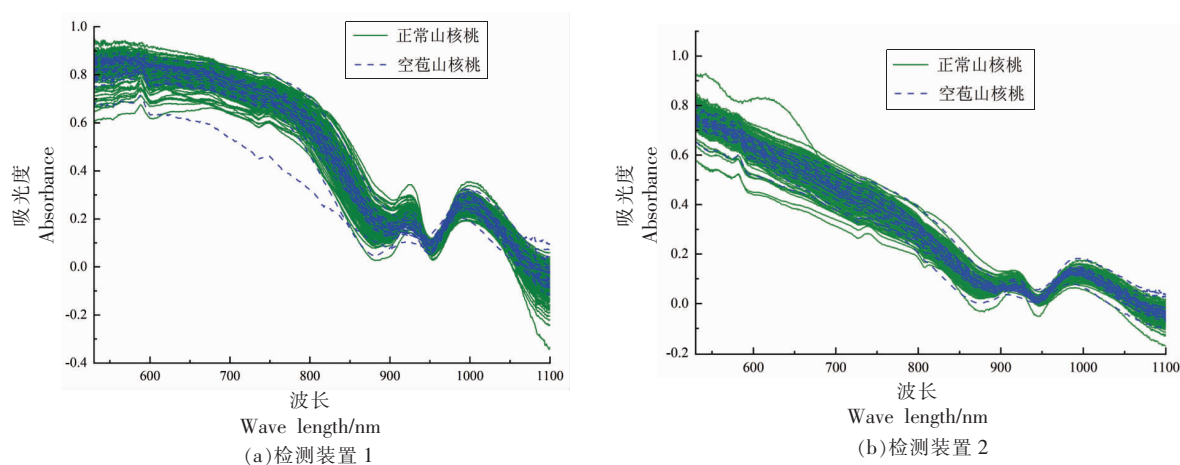
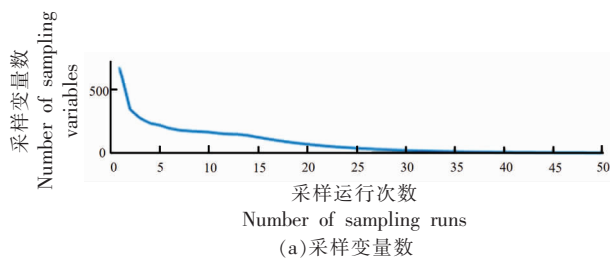


图 4 MSC 预处理后 530~1100 nm 波长范围的山核桃近红外光谱

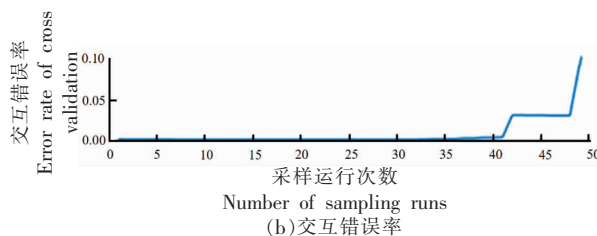
Fig.4 Near-infrared spectra of *Carya cathayensis* after MSC pretreatment in wave length range of 530–1100 nm

2.3.2 基于特征波长的分类模型 特征波长的选取可以有效减少无关变量的数目,从而起到减少变量数、提高模型精度的作用。对于全波长建模结果,由于检测装置 1 中建立的山核桃分类模型优于检测装置 2,因此采用 CARS 方法仅对检测装置 1 中的光谱数据进行特征波长选择。图 5 为经 MSC 预处理后的山核桃光谱的 CARS 变量筛选结果。由图 5a 可知,在 1~5 次 MC 采样过程中,被选择的山核桃空苞的波长变量数快速下降;在 5~25

次 MC 采样过程中,被选择的波长变量数缓慢下降,对应山核桃空苞特征波长变量的粗选与精选两个过程。图 5b 为山核桃空苞分类模型的错分率随采样次数的变化情况。在 26 次 MC 采样时,分类模型的错误率最小,此时选取的特征波长建模效果最佳。图 5c 为 MC 采样过程中山核桃光谱各波长变量的回归系数变化情况,“\*”对应的位置为分类模型错分率最低的 MC 采样次数。最终,共获得 34 个山核桃空苞的特征波长变量。对于原始光



(a) 采样变量数



(b) 交互错误率

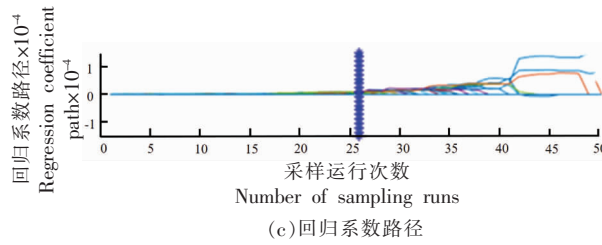


图 5 MSC 预处理后光谱的 CARS 变量筛选过程

Fig.5 CARS variable screening process of spectra after MSC preprocessing

谱及其它预处理光谱, 利用 CARS 方法在相同参数下进行变量筛选, 分别获得 27,13,27,27,48,54,43 和 24 个特征波长变量。

经 CARS 变量筛选后, 采用 LDA、QDA 和 MDA 方法分别建立空苞和正常山核桃的分类模型, 其结果见表 5。相比于全波长分类模型, 采用 CARS 筛选的特征波长建立的分类模型的变量数和计算时间大大减少, 各种分类模型的敏感性基

本不变, 特异性略有降低, 总正确率基本持平。由此可见, CARS 方法可以较为有效地筛选出空苞山核桃的特征波长变量。与全波长分类模型结果类似, MSC-CARS-LDA、MSC-CARS-MDA 和 MSC-CARS-QDA 模型的识别结果最优, 其特异性、敏感性和正确率均为 1, 说明可以很好地识别两类山核桃。

表 5 基于 CARS 特征波长的山核桃空苞的分类模型结果 (检测装置 1)

Table 5 Classification model results of *Carya cathayensis* with empty bud based on CARS sensitive wavelengths (Detection device 1)

建模方法	变量数	光谱预处理	校正集			预测集		
			敏感性	特异性	总正确率	敏感性	特异性	总正确率
CARS-LDA	27	无	0.8000	0.9458	0.9316	0.6667	0.9583	0.9259
	13	归一化	0.8000	0.9278	0.9153	0.7222	0.9375	0.9136
	27	去趋势	0.8000	0.9531	0.9381	0.6667	0.9583	0.9259
	27	基线校正	0.8333	0.9567	0.9446	0.6667	0.9722	0.9383
	48	SG 平滑	0.8000	0.9530	0.9381	0.6667	0.9653	0.9321
	54	一阶导数	0.8333	0.9603	0.9479	0.6667	0.9722	0.9383
	43	二阶导数	0.8333	0.9639	0.9511	0.6111	0.9722	0.9321
	24	标准正态变量变换	0.8333	0.9567	0.9446	0.7222	0.9444	0.9198
	34	多元散射校正	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
CARS-QDA	29	无	0.8667	0.9278	0.9218	0.7778	0.9444	0.9259
	13	归一化	0.8333	0.9206	0.9121	0.7222	0.9306	0.9074
	27	去趋势	0.8667	0.9278	0.9218	0.7778	0.9514	0.9321
	27	基线校正	0.8667	0.9314	0.9251	0.7222	0.9444	0.9198
	48	SG 平滑	0.8667	0.9278	0.9218	0.7778	0.9583	0.9383
	54	一阶导数	0.8667	0.9350	0.9283	0.7222	0.9444	0.9198
	43	二阶导数	0.8333	0.9278	0.9186	0.7222	0.9583	0.9321
	24	标准正态变量变换	0.8333	0.9242	0.9153	0.8333	0.9444	0.9321
	34	多元散射校正	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
CARS-MDA	29	无	0.9000	0.8339	0.8404	0.7778	0.8819	0.8704
	13	归一化	0.9333	0.8339	0.8436	0.8333	0.8681	0.8642
	27	去趋势	0.9333	0.8592	0.8664	0.7778	0.8958	0.8827



(续表 5)

建模方法	变量数	光谱预处理	校正集			预测集		
			敏感性	特异性	总正确率	敏感性	特异性	总正确率
	27	基线校正	0.9000	0.8484	0.8534	0.8333	0.8889	0.8889
	48	SG 平滑	0.8667	0.8556	0.8567	0.7778	0.9028	0.8889
	54	一阶导数	0.9000	0.8520	0.8567	0.7778	0.8889	0.8765
	43	二阶导数	0.9000	0.8484	0.8534	0.7778	0.9306	0.9136
	24	标准正态变量变换	0.9000	0.8303	0.8371	0.8333	0.8889	0.8827
	34	多元散射校正	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

### 3 结论

本研究利用近红外光谱分析技术和模式识别方法对空苞山核桃进行快速无损识别。采用两种近红外检测装置采集山核桃的光谱,比较了8种不同预处理方法及CARS变量筛选后所建立的LDA、QDA和MDA分类模型的性能。结果表明,检测装置1的空苞山核桃分类模型性能优于检测装置2的分类模型,MSC为较优的预处理方法,建立的山核桃空苞分类模型的性能最优,LDA、QDA和MDA模型的特异性、敏感性和正确率均为100%。经CARS特征波长筛选后,建模所用变量数大幅减少,分类模型得到有效简化,而精度与全波长模型持平。本文为空苞山核桃的快速无损识别提供一种可行的方法,也为其它坚果类的内部缺陷识别研究提供新思路。

### 参 考 文 献

- [1] 郑万钧. 中国树木志[M]. 北京: 中国林业出版社, 1983: 2276-2283.  
ZHENG W J. Dendrography of China[M]. Beijing: China Forestry Press, 1983: 2276-2283.
- [2] 赵玉顺. 临安山核桃产品营养成分特性研究[D]. 杭州: 浙江农林大学, 2013.  
ZHAO Y S. Study on the characteristics of nutritional components of Pecan products in Lin'an[D]. Hangzhou: Zhejiang A & F University, 2013.
- [3] 吕秋菊, 沈月琴, 高宇列, 等. 山核桃产业的发展过程、动因及展望[J]. 浙江农林大学学报, 2012, 29(1): 97-103.  
LÜ Q J, SHEN Y Q, GAO Y L, et al. Development process, motivation and prospect of pecan in industry[J]. Journal of Zhejiang A & F University, 2012, 29(1): 97-103.
- [4] 郭传友, 黄坚钦, 方炎明. 山核桃研究综述及展望[J]. 经济林研究, 2004, 22(1): 61-63.  
GUO C Y, HUANG J Q, FANG Y M. Review and prospect of *Carya cathayensis* research[J]. Nonwood Forest Research, 2004, 22(1): 61-63.
- [5] 王田利. 核桃出现“空苞”的原因及防治对策[J]. 果农之友, 2014, 5: 16-17.  
WANG T L. Causes of 'empty bud' in walnut and its control measures[J]. Friends of Fruit Farmers, 2014, 5: 16-17.
- [6] 刘爽, 柴春祥. 近红外光谱技术在水产品检测中的应用进展[J]. 食品安全质量检测学报, 2021, 12(21): 8590-8596.  
LIU S, CHAI C X. Application of near infrared spectroscopy in aquatic products detection[J]. Food Safety and Quality Detection Technology, 2021, 12(21): 8590-8596.
- [7] 徐叶挺, 张校立, 庄红梅, 等. 基于近红外光谱反射技术的砀山酥梨花粉活力快速检测[J]. 新疆农业科学, 2022, 59(3): 617-624.  
XU Y T, ZHANG X L, ZUANG H M, et al. Rapid detection of pollen viability of Dangshan Pear Based on near infrared reflectance spectroscopy[J]. Xinjiang Agricultural Sciences, 2022, 59(3): 617-624.
- [8] 乔春艳, 张雪茹, 郑卫民, 等. 基于近红外光谱技术对市售羊奶粉品质指标独立模型与通用模型探究[J]. 食品工业科技, 2021, 42(22): 297-305.  
QIAO C Y, ZHANG X R, ZHENG W M, et al. Study on independent and general models of quality indexes of goat milk powder based on near infrared spectroscopy[J]. Science and Technology of Food Industry, 2021, 42(22): 297-305.
- [9] 雷雨梦, 闫国超, 杨静, 等. 青花菜中矿质元素的近

- 红外光谱快速测定[J]. 浙江农业学报, 2022, 34(3): 548-556.
- LEI Y M, YAN G C, YANG J, et al. Rapid determination of mineral elements in broccoli by near infrared spectroscopy[J]. Acta Agriculturae Zhejiangensis, 2022, 34(3): 548-556.
- [10] 徐赛, 陆华忠, 王旭, 等. 基于可见/近红外光谱的菠萝水心病无损检测[J]. 农业工程学报, 2021, 37(21): 287-294.
- XU S, LU H Z, WANG X, et al. Nondestructive detection of pineapple heart disease based on visible / near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(21): 287-294.
- [11] 张保霞. 基于主成分分析和改进支持向量机的猪肉品质识别[J]. 食品与机械, 2022, 38(1): 146-151.
- ZHANG B X. Pork quality recognition based on principal component analysis and improved support vector machine[J]. Food & Machinery, 2022, 38(1): 146-151.
- [12] 田喜, 陈立平, 王庆艳, 等. 全透射近红外光谱的苹果整果糖度在线检测模型优化[J]. 光谱学与光谱分析, 2022, 42(6): 1907-1914.
- TIAN X, CHEN L P, WANG Q Y, et al. Optimization of on-line detection model for sugar content of apple whole fruit by full transmission near infrared spectroscopy [J]. Spectroscopy and Spectral Analysis, 2022, 42(6): 1907-1914.
- [13] 蒋大鹏, 张冬妍, 李丹丹, 等. 运用近红外光谱技术对松子霉变的快速检测[J]. 东北林业大学学报, 2019, 47(5): 83-87, 99.
- JIANG D P, ZHANG D Y, LI D D, et al. Rapid detection of pine nut mildew by near infrared spectroscopy[J]. Journal of Northeast Forestry University, 2019, 47(5): 83-87, 99.
- [14] HU J Q, MA X C, LIU L L, et al. Rapid evaluation of the quality of chestnuts using near-infrared reflectance spectroscopy[J]. Food Chemistry, 2017, 231: 141-147.
- [15] 郑剑, 周竹, 仲山民, 等. 基于近红外光谱与CARS-PLS-LDA的褐变板栗识别[J]. 食品工业, 2016, 37(2): 134-138.
- ZHENG J, ZHOU Z, ZHONG S M, et al. Recognition of browning chestnut based on near infrared spectroscopy and CARS PLS LDA[J]. The Food Industry, 2016, 37(2): 134-138.
- [16] ROGEL -CASTILLO C, BOULTON R, OPAST-PONGKARN A, et al. Use of near-infrared spectroscopy and chemometrics for the nondestructive identification of concealed damage in raw almonds (*Prunus dulcis*)[J]. Journal of Agricultural and Food Chemistry, 2016, 64(29): 5958-5962.
- [17] PANNICO A, SCHOUTEN R, BASILE B, et al. Non-destructive detection of flawed hazelnut kernels and lipid oxidation assessment using NIR spectroscopy[J]. Journal of Food Engineering, 2015, 160: 42-48.
- [18] MOSCETTI R, MONARCA D, CECCHINI M, et al. Detection of mold-damaged chestnuts by near-infrared spectroscopy[J]. Postharvest Biology and Technology, 2014, 93: 83-90.
- [19] 余怀鑫. 基于可见-近红外光谱技术的柑橘品质无损检测研究[D]. 武汉: 华中农业大学, 2021.
- YU H X. Nondestructive detection of citrus quality based on visible near infrared spectroscopy [D]. Wuhan: Huazhong Agricultural University, 2021.
- [20] 刘硕, 瞿崇晓, 祝中科, 等. 基于MSR和AMSR的红外融合增强算法[J]. 应用科学学报, 2022, 40(3): 423-433.
- LIU S, QU C X, ZHU Z K, et al. Infrared fusion enhancement algorithm based on MSR and AMSR[J]. Journal of Applied Sciences, 2022, 40(3): 423-433.
- [21] 曾韦珊, 黄林旺, 吕欣欣, 等. 湿加松针叶儿茶素和表儿茶素总量近红外光谱预测模型的建立[J]. 中南林业科技大学学报, 2022, 42(4): 22-29.
- ZENG W S, HUANG L W, LÜ X X, et al. Establishment of near infrared spectral prediction model for total catechins and epicatechins in needle leaves of *Pinus elliottii*[J]. Journal of Central South University of Forestry & Technology, 2022, 42(4): 22-29.
- [22] 石文强, 许秀英, 张伟, 等. 近红外光谱的北方寒地土壤含水率预测模型研究[J]. 光谱学与光谱分析, 2022, 42(6): 1704-1710.
- SHI W Q, XU X Y, ZHANG W, et al. Study on prediction model of soil moisture content in northern cold region based on near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2022, 42(6): 1704-1710.
- [23] 冯瑞杰, 陈争光, 衣淑娟. 基于贝叶斯优化的SVM玉米品种鉴别研究[J]. 光谱学与光谱分析, 2022, 42(6): 1698-1703.

- FENG R J, CHEN Z G, YI S J. Study on maize variety identification based on bayesian optimization SVM[J]. Spectroscopy and Spectral Analysis, 2022, 42(6): 1698–1703.
- [24] 张延宁. 基于光谱技术的肉品掺假检测方法研究[D]. 杭州: 浙江大学, 2021.
- ZHANG Y N. Study on detection method of meat adulteration based on spectral technology[D]. Hangzhou: Zhejiang University, 2021.
- [25] 杨邵文. 基于多源光谱数据的土壤属性建模优化与实证研究[D]. 淮南: 安徽理工大学, 2021.
- YANG S W. Optimization and demonstration of soil attribute modeling based on multi-source spectral data[D]. Huainan: Anhui University of Science and Technology, 2021.
- [26] 张孝红, 蒋雪松, 沈飞, 等. 漫透射近红外光谱的便携式面粉品质安全检测仪的设计[J]. 光谱学与光谱分析, 2022, 42(4): 1235–1242.
- ZHANG X H, JIANG X S, SHEN F, et al. Design of portable flour quality safety detector based on diffuse transmission near infrared spectroscopy [J]. Spectroscopy and Spectral Analysis, 2022, 42(4): 1235–1242.
- [27] LI H, LIANG Y, XU Q, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica Chimica Acta, 2009, 648(1): 77–84.
- [28] 段青玲, 徐晓玲, 李道亮, 等. 基于 CHG-LDA 的鱼类新鲜度无损检测方法[J]. 农业机械学报, 2021, 52(10): 385–393.
- DUAN Q L, XU X L, LI D L, et al. Non-destructive testing method of fish freshness based on CHG-LDA[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(10): 385–393.

## Rapid Identification of Empty Bud *Carya cathayensis* Based on Near-infrared Spectroscopy

Yu Chuze, Weng Dingkan, Cao Shuosun, Sun Tong\*

(College of Optical, Mechanical and Electrical Engineering, Zhejiang A & F University, Hangzhou 311300)

**Abstract** Empty bud *Carya cathayensis* refer to *Carya cathayensis* with no seed kernels or stunted development, which will seriously affect the product quality of *Carya cathayensis*. To achieve rapid and non-destructive identification of empty bud *Carya cathayensis*, two NIR detection devices were used to collect spectra of shelled *Carya cathayensis* samples in the wavelength range of 200–1 160 nm. Eight different preprocessing methods were used for spectral preprocessing, and competitive adaptive reweighted sampling (CARS) was used to select the sensitive wave length variables of empty bud *Carya cathayensis*. Finally, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and mahalanobis discriminant analysis (MDA) were applied to establish the classification models of empty bud and normal *Carya cathayensis*. The results showed that the performance of the classification models of empty bud *Carya cathayensis* established using detection device 1 were better than that of the detection device 2. The classification models established after multivariate scattering correction (MSC) pretreatment showed the best recognition results. The specificity, sensitivity and correctness of LDA, QDA, and MDA models were all 1, which were better than the classification models established by other preprocessing methods. After CARS variable screening, the number of spectral variables used for modeling was significantly reduced, effectively simplifying the classification model. However, the model performance was still on par with that of the full-band model. This paper provides a feasible method for the rapid nondestructive identification of empty bud *Carya cathayensis* by near-infrared spectroscopy combined with chemometric methods.

**Keywords** *Carya cathayensis*; near-infrared spectroscopy; empty bud; competitive adaptive reweighted sampling