

近红外光谱技术结合偏最小二乘判别分析 检测三七品质

李颖¹, 马雨辰², 刘萌¹, 孙兆敏¹, 付才力², 李占明^{3*}

(1. 厦门海洋职业技术学院海洋生物学院, 厦门 361000; 2. 苏州工业园区新国大研究院, 苏州 215123;
3. 江苏科技大学粮食学院, 镇江 212100)

摘要: **目的** 通过近红外光谱技术实现不同等级三七样品的快速鉴别。**方法** 采集等级 A (20 头)、等级 B (30 头)、等级 C (40 头)、等级 D (60 头) 4 种不同等级三七样品的近红外光谱, 构建偏最小二乘判别分析 (partial least squares discriminant analysis, PLS-DA) 分类器模型鉴别 4 种等级的三七样品, 同时为了减近红外光谱中的冗余波长变量, 进一步优化模型的判别结果, 利用竞争自适应重加权采样 (competitive adaptive reweighted sampling, CARS) 算法提取近红外光谱中的特征变量。**结果** 所构建的 PLS-DA 分类器模型对等级 C 和等级 D 的三七样品, 鉴别准确率达到 100%, 但是对于等级 A 和等级 B 的三七样品因为存在误判, 鉴别准确率仅为 0% 和 20%。经过 CARS 算法提取近红外光谱特征变量后, 光谱变量数大幅减少, 从 1557 个变量下降到 78 个变量。以优选后的特征变量构建的 CARS-PLS-DA 分类器模型更加简化, 对 4 种等级三七样品的预测均方根误差均明显下降, 说明模型的预测分类变量更接近真实的分类变量, 鉴别结果更加准确。同时, 对 4 种等级三七样品的鉴别准确率显著上升, 其中对于等级 C 和等级 D 的鉴别准确率为 100%, 对于等级 B 的鉴别准确率从 20% 提升到 100%, 等级 A 鉴别准确率从 0% 提升到 75%。**结论** 所构建的 CARS-PLS-DA 分类器模型对 4 种等级的三七样品具有更好的鉴别效果, 可以实现不同等级三七的品质鉴定。

关键词: 三七; 近红外光谱; 品质鉴定; 偏最小二乘判别分析; 竞争自适应重加权采样

Combination of near-infrared spectroscopy and partial least squares discriminant analysis in detecting the quality of *Panax notoginseng*

LI Ying¹, MA Yu-Chen², LIU Meng¹, SUN Zhao-Min¹, FU Cai-Li², LI Zhan-Ming^{3*}

(1. College of Marine Biology, Xiamen Ocean Vocational College, Xiamen 361000, China; 2. National University of Singapore Suzhou Research Institute, Suzhou 215123, China; 3. School of Grain Science and Technology, Jiangsu University of Science and Technology, Zhenjiang 212100, China)

ABSTRACT: Objective To realize the rapid identification of different grades of *Panax notoginseng* samples by near-infrared spectroscopy technology. **Methods** Collecting the near infrared spectroscopy of 4 kinds of different grades of *Panax notoginseng*, including grade A (20 tou), grade B (30 tou), grade C (40 tou), grade D (60 tou), partial least squares discriminant analysis (PLS-DA) classifier model was used to rapid discriminate the quality of *Panax*

基金项目: 福建省教育厅项目 (JAT210801)、福州市科技计划项目 (AFZ2021K010003)

Fund: Supported by the Fujian Provincial Education Department Project (JAT210801), and the Fuzhou Science and Technology Project (AFZ2021K010003)

*通信作者: 李占明, 博士, 讲师, 主要研究方向为农产品品质控制。E-mail: lizhanming@just.edu.cn

*Corresponding author: LI Zhan-Ming, Ph.D, Lecturer, School of Grain Science and Technology, Jiangsu University of Science and Technology, No.666, Changhui Road, Dantu District, Zhenjiang 212100, China. E-mail: lizhanming@just.edu.cn

notoginseng. In order to reduce redundant wavelength variables of near infrared spectroscopy and optimize the discriminant results of model, competitive adaptive reweighted sampling (CARS) was used to extract characteristic wavelength variables of the near infrared spectroscopy. **Results** The constructed PLS-DA classifier model could be used to rapid discriminate the *Panax notoginseng* grade of C and grade D, with discriminant accuracy were both 100%. However, the discriminant accuracy was only 0% and 20% for the *Panax notoginseng* grade of grade A and grade B, because misjudgment was found. And the number of characteristic variables were reduced from 1557 to 78 by CARS. And then, the CARS-PLS-DA classifier model was built by those characteristic variables. As a result, the CARS-PLS-DA classifier model was more simple, and the root mean square errors of prediction of different grades of *Panax notoginseng* were decreased obviously, indicating that the prediction classification variables of the model were closer to the real classification variables and the identification results were more accurate. Besides, the discriminant accuracy of *Panax notoginseng* of different grades increased significantly, among which the discriminant accuracy of grade C and grade D were 100%, the discriminant accuracy of grade B increased from 20% to 100%, and the discriminant accuracy of grade A increased from 0% to 75%. **Conclusion** The CARS-PLS-DA classifier model has better identification effect on *Panax notoginseng*, of different grades, and can realize the quality identification of *Panax notoginseng* of different grades.

KEY WORDS: *Panax notoginseng*; near-infrared spectroscopy; quality discriminate; partial least squares discriminant analysis; competitive adaptive reweighted sampling

0 引言

三七 [*Panax notoginseng* (Burk.) F.H.Chen], 又名田七, 被誉为“中药中的阿司匹林”。云南省文山州是我国最重要的三七培育地址, 被誉为“三七之乡”^[1]。三七在我国已经有四百多年的种植历史, 是我国非常重要的药食同源作物, 其主要功效成分包括皂苷类、黄酮类、甾醇类、丹皮酸、多糖、氨基酸和有机酸等, 具有多种功效与作用, 包括抗氧化、止血、化瘀、消肿定痛等^[2-5]。

三七是五加科人参属植物三七的干燥根和根茎, 根据我国 GB/T 19086—2008《地理标志产品 文山三七》, 头是表示三七大小的专业规格单位, 指 500 g 干燥三七中的主根个数。一般地, 三七的头数与其价格、功效成分含量具有显著的联系, 头数越小, 功效成分含量越多, 价格越贵。因为随着三七栽培年份的不断增长, 三七的根茎中功效成分含量越多, 根茎的体积和重量也越大^[6-7], 在经过加工干燥处理后的三七主根的体积和质量也越大, 即 500 g 质量的干燥三七中的主根个数也越少。日常消费中, 由于干燥三七质硬, 水分仅为 3%, 无法直接口服, 因此消费者更青睐于购买方便煮制的三七粉末进行食用, 造成无法直接通过肉眼观察三七的头数来鉴别三七的品质。同时, 一些不法商家也借此以次充好牟利。三七的常规检测主要通过理化分析其中的营养物质成分, 如多糖、黄酮等化学成分的含量, 结果准确可靠, 但检测过程耗时长、操作复杂烦琐^[8-10]。

近红外光谱分析技术是通过测定样品组成成分中一

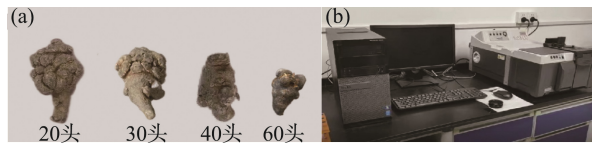
些有机分子中含氢基团在 780~2526 nm 的近红外波长范围内的特征吸收峰信息, 并结合化学计量学方法对光谱进行分析与处理的一种技术, 具有快速、高效、绿色环保、便捷等优点, 在食品、农产品、药品的掺伪鉴别、产地鉴别、有效成分测定等方面得到广泛应用^[11-14]。因此, 本研究利用近红外光谱分析技术结合偏最小二乘判别分析 (partial least squares discriminant analysis, PLS-DA) 模型实现快速鉴别不同等级的三七样品, 为维护消费者合法权益和保障中药三七的市场安全提供理论依据, 主要包括以下两方面内容: (1) 通过采集 4 种不同等级的三七样品的近红外光谱图, 运用 PLS-DA 分类器模型鉴别 4 种不同等级的三七品质; (2) 利用竞争自适应重加权采样 (competitive adaptive reweighted sampling, CARS) 算法提取光谱特征变量, 简化上述 PLS-DA 分类器模型, 减少冗余波长变量对模型影响, 提高模型鉴别效果。

1 材料与方法

1.1 材料与设备

采购不同头数的文山三七, 包括 20 头样品 (即 500 g 干燥三七中的主根个数约 20 个)、30 头样品、40 头样品、60 头样品, 图 1 (a) 为不同头数三七对比。不同头数三七样品分别对应不同等级, 20 头三七即为等级 A、30 头三七即为等级 B、40 头三七即为等级 C、60 头三七即为等级 D, 4 个等级三七样品分别收集 20 份, 利用 AF-20A 型中药超微粉碎机分别粉碎上述不同等级的三七样品, 将粉碎后的样品过 60 目筛网后放置于 60℃ 的电热恒温鼓风干燥箱中干

燥直至恒重,干燥后将每一份三七样品平均分为 4 份,最后用塑封袋(60 mm×40 mm)密封置于干燥器中保存备用。



注: (a)不同头数三七对比; (b)实验所用 ANTARISII型傅里叶变换近红外光谱分析仪。

图 1 实验材料及仪器环境

Fig.1 Materials and environmental instruments

用 ANTARISII型傅里叶变换近红外光谱分析仪采集不同等级三七样品的近红外光谱,图 1 (b)为实验仪器图,光谱仪配置高灵敏度和高稳定性的 InGaAs 检测器、积分球采样系统、迈克尔干涉仪、多种采样方式以及内置自动金箔背景采集方式。本研究设置的光谱采集参数:光谱范围 4000~10000 cm^{-1} ,分辨率为 8 cm^{-1} ,平均扫描次数为 32 次,采集环境为室温 25°C、空气湿度 60%。

AF-20A 中药超微粉碎机(奥力中药机械有限公司); 9070MBE 型号电热恒温鼓风干燥箱(厦门纳柏科技有限公司); ANTARISII型傅里叶变换近红外光谱分析仪(德国赛默飞世尔科技公司)。

1.2 实验方法

1.2.1 光谱采集

近红外光谱采集用空旋转样品杯作为实验的测量背景,按照编号顺序依次从干燥器中取干燥后的三七粉末样品(大约 3.0 g 左右)均匀地平铺在旋转样品杯中,按照设置的光谱采集参数采集不同等级三七样品的近红外光谱。将采集到的样品光谱划分为校正集和预测集,以校正集光谱训练模型的各项参数,模型建立后利用预测集光谱验证所建立分类器模型的鉴别效果。

1.2.2 分析模型构建

在构建模型前,通过主成分分析对所采集的光谱集合进行分析和简化,通过从高维数、复杂的光谱数据集合中提取能够代表原有变量信息的主成分^[15]。利用 Kennard-Stone 算法按照 1:1 的比例划分为训练集和预测集, Kennard-Stone 算法基本原理如下:给定训练集中样品光谱数目,先从所有光谱数据集合中选择欧氏距离最远的两个光谱数据进入训练集,然后依次计算比较剩下的每一个光谱数据到训练集内已选光谱数据之间的欧式距离,将欧氏距离最远和欧氏距离最近的两个样本选入训练集,重复上述计算方法直到光谱数据集合中训练集样本数量满足给定数目^[16]。

PLS-DA 分类器模型是一种有监督的判别分析方法,具有许多优点:如样本需要数目少,可以减少近红外光谱数据集合中波长变量之间存在的多重共线性的影响,是一

种广泛应用的定性分析方法^[17-18]。PLS-DA 分类器模型的原理与 PLS 回归模型类似,但与 PLS 回归模型不同, PLS-DA 分类器模型的性质矩阵 Y 不是给定的实验结果而是由人工给定的分类变量组成。模型的建立分为两个步骤:先通过留一交互检验法,以预测残差平方和(predicted residual sum of squares, PRESS)作为主要指标,当 PRESS 值最小时即可得最佳因子数, PRESS 计算公式如式(1):

$$\text{PRESS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

其中, y_i 为观察值, \hat{y}_i 为拟合值。然后通过所收集的近红外光谱数据集合矩阵 X 以及人工给定的分类变量矩阵 Y 建立 PLS-DA 分类器模型,主要过程如式(2)~(3):

$$X = TP^E + E \quad (2)$$

$$Y = UQ^E + F \quad (3)$$

其中, X 是 $n \times m$ 维的不同等级三七的近红外光谱数据矩阵, Y 认为是人工给定的 $n \times p$ 维的分类变量性质矩阵, P^E 、 T 分别是近红外光谱数据矩阵 X 的载荷矩阵和得分矩阵, Q^E 、 U 分别是分类变量性质 Y 的载荷矩阵和得分矩阵, E 和 F 是经过回归拟合后产生偏差值。将 T [式(4)]和 U [式(5)]矩阵进行分解得到关联参数矩阵 F 。

$$F = T \times U(TT)^{-1} \quad (4)$$

$$U = T \times F \quad (5)$$

通过训练集的载荷阵 P^E 和预测集样品光谱矩阵 T_w 运算从而得到预测集的预测分类变量 Y_w [式(6)]:

$$Y_w = T_w \times F \times Q^E \quad (6)$$

由于近红外光谱的变量数较多,冗余的无关光谱变量不仅降低模型的准确性,并且大大增加仪器成本,通过变量优选算法筛选近红外光谱中具有代表性的特征变量,以优选后的特征变量构建的分析模型,不但能够提高模型的准确性,且能减少便携式仪器的开发成本^[19-21]。实验运用 CARS 算法提取近红外光谱中的特征变量。CARS 算法是根据达尔文提出的“适者生存,优胜劣汰”的进化论观点为基础。CARS 算法提取光谱特征变量主要包括粗略提取和精细提取两个步骤,粗提取过程通过指数递减函数除去光谱中回归系数绝对值较低的特征变量,指数递减函数的计算公式如式(7):

$$r_i = ae^{-ki} \quad (7)$$

其中 i 是所采集光谱数据的数目, a 和 k 可以通过已知数据来计算。粗提取后利用自适应重加权采样(adaptive reweighted sampling, ARS)算法进一步除去冗余无关波长变量,算法主要过程如下:(1)通过蒙特卡罗采样法采集近红外光谱数据集合 N 次,每次都随机选取 80% 的光谱数据作为校正集,然后构建 PLS 回归模型;(2)通过指数递减函数将 b_i 值较小的变量除去;(3)利用 N 次 ARS 采样技术计算筛选出所构建 PLS 回归模型中回归系数绝对值较大的特征变量点;(4)以上述筛选出的光谱变量子集建立 PLS 回归模型,通过计算比较不同光谱变量子集的交叉检验均方根误

差值(root mean square error of cross validation, RMSECV), 当 RMSECV 最小时对应的光谱变量子集即为 CARS 算法优选出最优光谱特征变量集合^[22]。具体算法如式(8)~(10):

$$T=XW \quad (8)$$

$$y=Tc+e=XWc+e=Xb+e \quad (9)$$

$$b=Wc=[b_1, b_2, \dots, b_p]^T \quad (10)$$

其中 T 为得分矩阵, 是不同等级三七的近红外光谱矩阵 X 和 W 的线性组合。 c 是 y 对应得分矩阵 T 的回归系数, b 是 p 维的系数向量, e 是预测误差。 b 中第 i 个元素的绝对值 $|b_i|$ ($1 \leq i \leq p$) 代表了第 i 个波长对 y 的贡献, 第 i 个波长的贡献率 W_i 计算公式如式(11):

$$W_i = \frac{|b_i|}{\sum_{i=1}^p |b_i|}, i = 1, 2, 3, \dots, p \quad (11)$$

1.2.3 数据处理

所有数据分析处理用 MATLAB 程序(R2018a)平台完成。同时, 为了比较所建立模型对不同等级三七的鉴别效果, 用样品判别准确率和预测均方根误差(root mean square errors of prediction, RMSEP)作为分类器模型鉴别结果的评价指标。判别准确率和 RMSEP 计算公式如式(12)~(13):

$$\text{判别率}/\% = \frac{\text{模型能正确判别的样品个数}}{\text{样品总数目}} \times 100\% \quad (12)$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n_p}} \quad (13)$$

其中, y_i 为模型计算出的预测分类变量值, \hat{y}_i 为给定的分类变量值, n_p 是预测集的样本数目。

2 结果与分析

2.1 不同等级三七样品的近红外光谱解析

图 2 是 4 种不同等级三七样品的近红外光谱图。从图 2 中可以看出, 不同等级三七样品在 4000~10000 cm^{-1} 范围内的光谱吸收峰的位置基本一致, 大致在 4400、4800、5200、7000 及 8300 cm^{-1} 附近, 但不同吸收峰的最大强度有较为明显差异, 说明不同等级三七样品中的含氢有机物质的含量不同, 可能原因是不同等级三七样品中的主要生理活性物质, 如黄酮、多酚、多糖以及氨基酸等含量不同引起样品间吸收峰强度的差异^[23]。在近红外光谱中的几处吸收峰分别对应不同含氢基团的特征吸收, 在 4453 cm^{-1} 左右的光谱峰对应的是 C-H 组合频和 O-H 倍频的特征吸收, 4300~4320 cm^{-1} 左右的光谱峰为亚甲基中的 C-H 组合频的特征吸收, 5200~6000 cm^{-1} 左右的吸收峰是 C-H 的二级倍频的特征吸收, 6000~7000 cm^{-1} 左右的吸收峰是 O-H 或 N-H 二级倍频的特征吸收, 8300 cm^{-1} 左右的吸收峰是 C-H 拉伸三级倍频的特征吸收^[24-25]。总体上, 不同等级三七样品的光谱相似性较高, 通过肉眼分析光谱很难直接鉴别三七的品质, 需要通过化学计量学模型进一步挖掘分析不同

等级三七样品的近红外光谱。

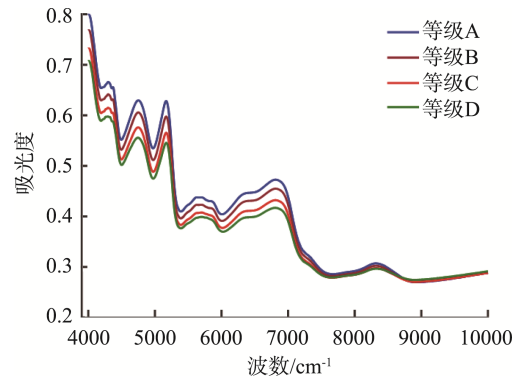


图 2 4 种不同等级三七样品的近红外光谱图
Fig.2 Near infrared spectroscopy of *Panax notoginseng* from 4 kinds of different grades

2.2 基于 PLS-DA 分类器模型的三七品质鉴别

构建模型前先采用主成分分析来分析不同等级三七的近红外光谱集合以便更加直观地展现不同等级三七样品的差异性。通过分析计算简化光谱变量可得, 光谱矩阵的前 5 个特征值分别为 0.9225、0.0049、0.0005、0.0001、0.00005, 计算得前 3 个特征值的累计贡献率分别为 99.40%、99.83%、99.98%, 表明前 3 个特征值已能够充分代表所采集到光谱矩阵^[26-27]。图 3 是以上述光谱矩阵的前 3 个特征值构建的主成分得分三维分布图。由图 3 可以看出, 不同等级的三七样品比较集中, 被分为 4 个区域, 等级 C 和等级 D 分界清晰, 但是等级 A 部分样品与等级 B 部分样品有交叉, 可能原因是等级 A 和等级 B 的样品之间差异不明显, 通过主成分分析无法使二者分离, 表明需要进一步借助化学计量学方法实现不同等级三七样品鉴别。

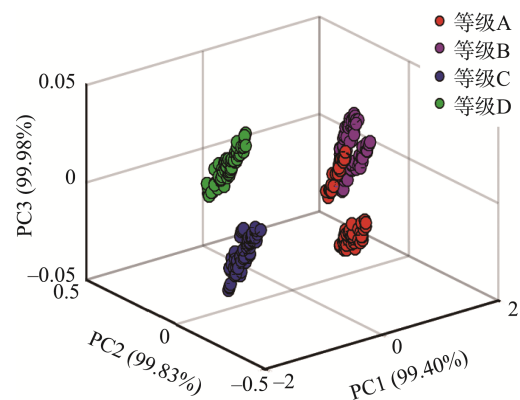


图 3 不同等级三七样品的三维主成分得分分布图
Fig.3 Score plot of three dimensional principal component of *Panax notoginseng* from different grades

将以上不同等级三七的近红外光谱集合利用 Kennard- Stone 算法按照 1:1 的比例划分为训练集和预测集, 以训练集构建 PLS-DA 分类器模型, 利用预测集检验模型的

鉴别效果。图 4 是训练集中不同等级三七样品的 PLS-DA 分类器模型中 PRESS 分布, 从图 4 中可知, 随着横坐标模型因子数的不断增加, PRESS 不断下降, 当模型因子数为 12 时, PRESS 的值达到最小, 为 0.1458, 如果模型的因子数继续增加, PRESS 几乎不变趋于稳定。因此, 构建不同等级三七样品的 PLS-DA 分类器模型所选择最佳因子数是 12。

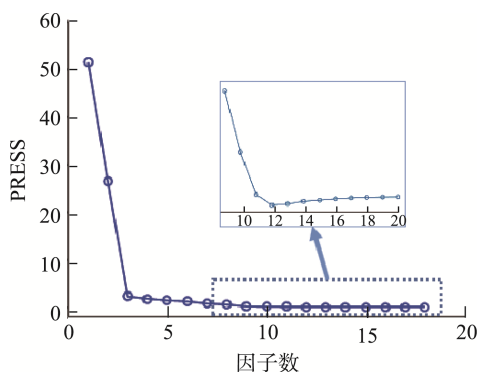


图 4 PLS-DA 分类器模型的 PRESS 值变化趋势图

Fig.4 PRESS distribution of PLS-DA model

在最佳因子数下建立不同等级三七样品的 PLS-DA 分类器模型, 模型对预测集的预测分类变量结果如图 5。在预测集中, 将 4 种不同等级的 180 个样品分别编号, 样品编号 1~40 号样品为等级 A 三七, 41~80 号样品为等级 B 三七, 81~120 号样品为等级 C 三七, 121~160 号样品为等级 D 三七, 同时分别给定数值 1.0、2.0、3.0、4.0 为等级 A、B、C 和 D 的三七样品所对应的准确分类变量。从图 5 可知, PLS-DA 分类器模型对等级 C 和等级 D 三七样品的预测分类变量值分别在 2.8~3.0 以及 3.9~4.1, 预测分类变量值均接近给定的分类变量, 样品判别准确率达 100%, 但等级 A 和等级 B 三七样品的预测分类变量值分别介于 1.5~1.6 和 2.5~2.6 之间, 相比于给定分类变量 1.0 和 2.0, PLS-DA 分类器模型将等级 A 的 40 个样品均误判为等级 B 样品, 将等级 B 中的 34 个样品误判为等级 C 样品, 说明所构建的 PLS-DA 分类器模型对等级 C 和等级 D 三七样品具有良好判别效果, 但对于等级 A 和等级 B 三七样品存在误判风险。因此, 将进一步优化判别模型, 提取红外光谱中的特征变量, 提高模型的准确性。

2.3 基于 CARS-PLS-DA 分类器模型的三七品质鉴别

为提高 PLS-DA 分类器模型的稳定性、准确性, 剔除光谱中影响模型准确度的冗余变量, 简化模型, 利用 CARS 算法提取近红外光谱中特征变量。图 6 为 CARS 优选特征变量步骤流程, 从图 6 中可知, 随着样品采集次数的增加, RMSECV 不断下降[图 6(b)], 不同等级三七样品的近红外光谱中冗余的变量数不断减少[图 6(a)], 当采样运行次数为 28 时, 光谱变量数下降速度趋于稳定, 回归系数路径剧烈变化[图 6(c)], RMSECV 达到最低点, 而后迅速上

升并趋于稳定, 表明光谱中的冗余的变量被有效去除^[28-30], 此时光谱的特征变量数从最初的 1557 下降为 78。

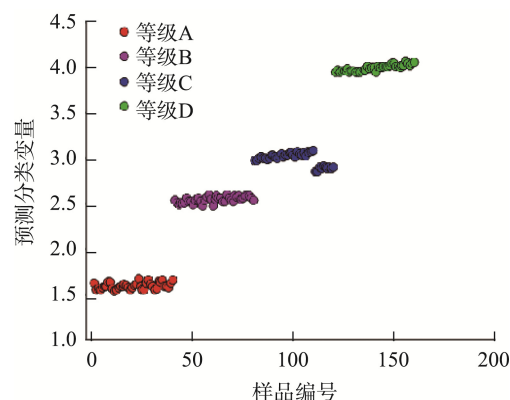
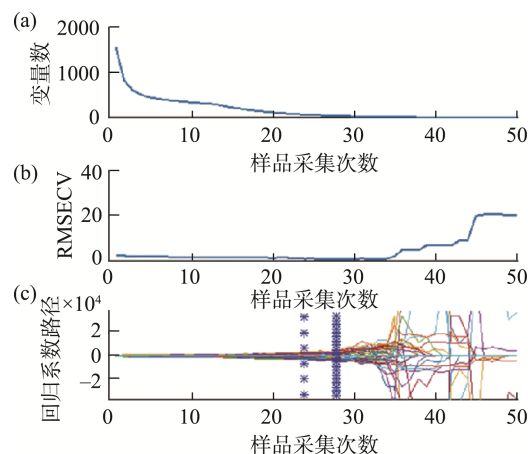


图 5 PLS-DA 分类器模型的预测分类变量值分布

Fig.5 Distribution of predicted categorical variable values of PLS-DA model



注: (a)为变量数随样品采集次数增加的趋势变化图; (b)为 RMSECV 随样品采集次数增加的趋势变化图; (c)为回归系数随样品采集次数增加的趋势变化图。

图 6 CARS 算法优选特征变量流程

Fig.6 Characteristic variables selection process of the CARS

图 7 为 CARS 算法在 4000~10000 cm^{-1} 范围内提取的光谱特征变量分布图, 光谱图中圆圈所对应横坐标即为提取特征变量。从图 7 可知, 利用 CARS 算法提取的特征变量集中在 4000~6000 cm^{-1} 和 7000~9500 cm^{-1} 范围内, 且光谱变量数大大减少, 从 1557 个光谱变量下降为 78 个光谱变量, 以优化提取的特征变量构建 CARS-PLS-DA 分类器模型。同理, PRESS 值最小为 0.1382, 对应模型的最佳因子数为 11。在模型最佳因子数下, 建立的 CARS-PLS-DA 分模型的预测结果如图 8。从图 8 可以看出, CARS-PLS-DA 分类器对等级 B、C 和 D 三七样品的预测分类变量值分别在 2.3~2.4、2.9~3.0、3.9~4.0 附近, 说明等级 B、C 和 D 的预测分类变量均在给定分类变量附近, 判别准确率为

100%, CARS-PLS-DA 分类器对等级 A 三七样品的预测分类变量值分别在 1.4~1.5 之间, 对于预测分类变量超过 1.5 的 10 个样品将被模型误判为等级 B 样品。表 1 是 PLS-DA 和 CARS-PLS-DA 模型对不同等级三七的判别效果, 从表 1 中可以看出, 相比于 PLS-DA 分类器模型, CARS-PLS-DA 分类器模型对 4 种不同等级三七样品的 RMSEP 均下降, 说明其预测分类变量更接近真实的分类变量, 鉴别结果更加准确。同时, 计算模型对不同等级三七样品的判别率可得, CARS-PLS-DA 分类器模型能够准确判别等级 C 和等级 D 三七样品, 鉴别准确率达 100%, 对于等级 B 三七样品的鉴别准确率从 20% 提升到 100%, 而对于等级 A 三七样品的鉴别准确率从 0% 上升为 75%。上述结果表明, 通过提取光谱特征变量后构建的 CARS-PLS-DA 分类器模型对不同等级的三七样品具有更好的鉴别效果, 并且模型所需光谱变量数较少。

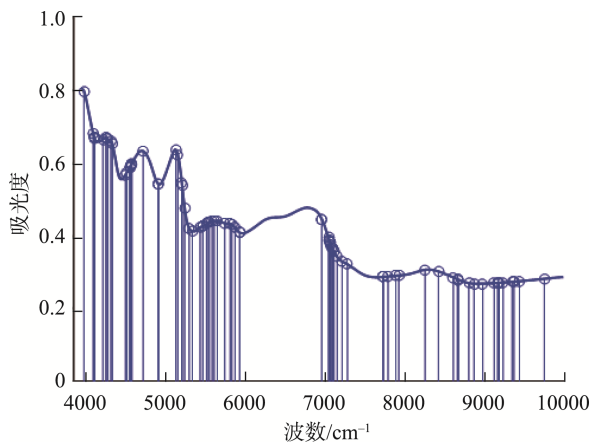


图 7 CARS 优选的特征变量分布图

Fig.7 Distribution of selected characteristic variables by CARS

表 1 PLS-DA 和 CARS-PLS-DA 分类器模型对不同等级三七的判别效果

Table 1 Discrimination results of different grades of *Panax notoginseng* by PLS-DA and CARS-PLS-DA classifier model

模型	等级 A		等级 B		等级 C		等级 D	
	判别率/%	RMSEP	判别率/%	RMSEP	判别率/%	RMSEP	判别率/%	RMSEP
PLS-DA	0	0.5982	20	0.5464	100	0.0667	100	0.0352
CARS-PLS-DA	75	0.4866	100	0.4159	100	0.0492	100	0.0313

参考文献

- [1] 阙祖亮, 庞丹清, 陈勇, 等. 三七的种植及采收加工现状[J]. 江苏农业科学, 2020, 48(4): 41-45.
QUE ZL, PANG DQ, CHEN Y, et al. Current situation of planting, harvesting and processing of panax notoginseng [J]. Jiangsu Agric Sci, 2020, 48(4): 41-45.
- [2] JIE W, XU J, ZHONG JB. Effect of *Radix notoginseng* saponins on platelet activating molecule expression and aggregation in patients with blood hyperviscosity syndrome [J]. Chin J Integr Tradit West Med, 2004, 24(4): 312-317.

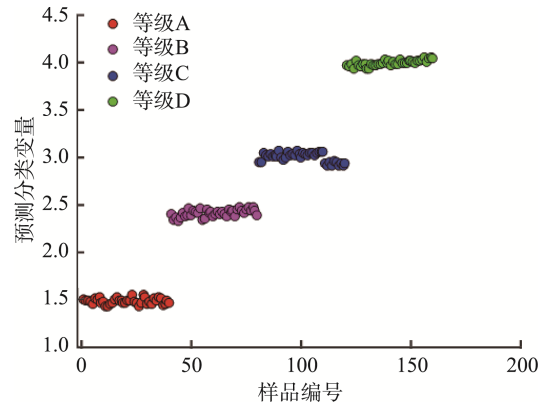


图 8 CARS-PLS-DA 模型的预测分类变量值分布

Fig.8 Distribution of predicted categorical variable values of CARS-PLS-DA model

3 结论

本研究利用近红外光谱结合化学计量学实现不同品质三七的无损、便捷、快速准确鉴定。实验结果表明, 以 CARS 算法提取得到的 78 个光谱特征变量构建的 CARS-PLS-DA 分类器模型对不同等级的三七样品具有更好的鉴别效果, 表现为 4 种不同等级三七样品的 RMSEP 均下降, 且对等级 B 样品的判别准确率从 20% 提升到 100%, 对等级 A 品的判别准确率从 0% 提高至 75%。因此, 近红外光谱技术结合偏最小二乘判别分析检测三七品质, 在实践中具有良好的可行性, 同时也为便携式仪器的开发提供一定理论基础。但上述研究对现实中隐蔽的三七品质鉴别问题: 如不同品质三七掺假问题、三七及其伪品掺假问题等, 还需要进一步探讨。

- [3] 张喜平, 齐丽丽, 刘达人. 三七及其有效成分的药理作用研究现状[J]. 医学研究杂志, 2007, 36(4): 96-98.
ZHANG XP, QI LL, LIU DR. Research status of pharmacological action of panax notoginseng and its active components [J]. J Med Res, 2007, 36(4): 96-98.
- [4] WEI X, SUN C, YAN Z, et al. Hypolipidemic and antioxidant activities of sanchi (*Radix notoginseng*) in rats fed with a high fat diet [J]. Phytomedicine, 2011, 18(6): 516-520.
- [5] 蔡树杏, 张慧晔, 邹江林, 等. 三七茎叶中皂苷类成分研究进展[J]. 食品与发酵科技, 2021, 57(3): 108-114.
CAI SX, ZHANG HY, ZOU JL, et al. Research progress of the saponins in

- the stems and leaves of panax notoginseng [J]. *Food Ferment Sci Technol*, 2021, 57(3): 108–114.
- [6] 王振峰, 高云涛, 张文斌, 等. 不同生长年限三七中总皂苷含量的变化特征[J]. *安徽农业科学*, 2012, 40(15): 8458–8459.
WANG ZF, GAO YT, ZHANG WB, *et al.* Variation characteristics of total saponins content in panpax notoginseng at different growth ages [J]. *J Anhui Agric Sci*, 2012, 40(15): 8458–8459.
- [7] 刘晓行, 李宏. 云南文山三七市场价格变动分析和产业可持续发展研究[J]. *农村经济与科技*, 2015, 26(9): 145–147.
LU XX, LI H. Research on price changes and industry sustainable development of panax notoginseng in Wenshan Yunnan [J]. *Rural Econ Sci Technol*, 2015, 26(9): 145–147.
- [8] 李彩凤, 胡欣, 金鹏飞, 等. HPLC 法同时测定三叶青块根中 6 个黄酮类成分的含量[J]. *中国药房*, 2019, 30(13): 1755–1758.
LI CF, HU X, JIN PF, *et al.* Simultaneous determination of 6 flavonoids in the roots of tetragium hemsleyanum by HPLC [J]. *China Pharm*, 2019, 30(13): 1755–1758.
- [9] 梁毅, 康炯. 粮油产品质量安全检测技术的研究进展与发展趋势[J]. *食品安全质量检测学报*, 2019, 10(4): 848–853.
LIANG Y, KANG J. Research progress and development trend of quality and safety detection technology for grain and oil product [J]. *J Food Saf Qual*, 2019, 10(4): 848–853.
- [10] GUAN C, LI D, LIU Q. Analysis of initiator content of prepreg by near-infrared spectroscopy [J]. *Rev Anal Chem*, 2022, 41(1): 74–82.
- [11] COSTA M, MORGANO MA, FERREIRA M, *et al.* Quantification of mineral composition of Brazilian bee pollen by near infrared spectroscopy and PLS regression [J]. *Food Chem*, 2019, 273: 85–90.
- [12] JAN UP, DIETMAR RK, REINHOLD C. On-line application of near infrared (NIR) spectroscopy in food production [J]. *Trend Food Sci Technol*, 2015, 46(2): 211–230.
- [13] KAZEMI A, MAHMOUDI A, VELADI H, *et al.* Rapid identification and quantification of intramuscular fat adulteration in lamb meat with VIS-NIR spectroscopy and chemometrics methods [J]. *J Food Meas Charact*, 2022, 16(3): 2400–2410.
- [14] TORRAGA MRG, GIUDICI R. Stabilizer-free dispersion copolymerization monitoring by in-line NIR spectroscopy [J]. *Ind Eng Chem Res*, 2020, 59(35): 15497–15505.
- [15] ABDI H, WILLIAMS LJ. Principal component analysis [J]. *Wiley Comput Stat*, 2010, 2(4): 433–459.
- [16] CLAEYS DD, VERSTRAELEN T, PAUWELS E, *et al.* Conformational sampling of macrocyclic alkenes using a Kennard-Stone-based algorithm [J]. *J Phys Chem A*, 2010, 114(25): 6879–6887.
- [17] LEE LC, LIONG CY, JEMAIN AA. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps [J]. *Analyst*, 2018, 143(15): 3526–3539.
- [18] AMINU M, AHMAD NA. Complex chemical data classification and discrimination using locality preserving partial least squares discriminant analysis [J]. *ACS Omega*, 2020, 5(41): 26601–26610.
- [19] BRENCHLEY JM, HORCHNER U, KALIVAS JH. Wavelength selection characterization for NIR spectra [J]. *Appl Spectrosc*, 1997, 51(5): 689–699.
- [20] SONG X, DU G, LI Q, *et al.* Rapid spectral analysis of agro-products using an optimal strategy: Dynamic backward interval PLS-competitive adaptive reweighted sampling [J]. *Anal Bioanal Chem*, 2021, 412(12): 2795–2804.
- [21] ZHENG K, FENG T, ZHANG W, *et al.* Variable selection by double competitive adaptive reweighted sampling for calibration transfer of near infrared spectra [J]. *Chemometr Intelligent Lab*, 2019, 191: 109–117.
- [22] ZHENG K, LI Q, WANG J, *et al.* Stability competitive adaptive reweighted sampling (SCARS) and its applications to multivariate calibration of NIR spectra [J]. *Chemometr Intelligent Lab*, 2012, 112: 48–54.
- [23] 贺焕玲. 活性三七饮片指纹图谱研究及单体皂苷含量测定[J]. *北方药学*, 2018, 15(1): 2.
HE HL. Study on Fingerprint of active sanqi pieces and determination of monomer saponin content [J]. *J North Pharm*, 2018, 15(1): 2.
- [24] JERRY W, LOIS W. Practical guide to interpretive near-infrared spectroscopy [M]. Boca Raton: CRC Press, 2007.
- [25] OUYANG Q, WANG L, PARK B, *et al.* Simultaneous quantification of chemical constituents in matcha with visible-near infrared hyperspectral imaging technology [J]. *Food Chem*, 2021, 350: 129–141.
- [26] DESTEFANIS G, BARGE MT, BRUGIAPAGLIA A. The use of principal component analysis (PCA) to characterize beef [J]. *Meat Sci*, 2000, 56(3): 255–259.
- [27] BARTH J, KATUMULLAGE D, YANG CY, *et al.* Classification of wines using principal component analysis [J]. *J Wine Econ*, 2021, 16(1): 56–67.
- [28] 刘国海, 夏荣盛, 江辉, 等. 一种基于 SCARS 策略的近红外特征波长选择方法及其应用[J]. *光谱学与光谱分析*, 2014, 34(8): 2094–2097.
LIU GH, XIA RS, JIANG H, *et al.* A wavelength selection approach of near infrared spectra based on SCARS strategy and its application [J]. *Spectrosc Spectr Anal*, 2014, 34(8): 2094–2097.
- [29] PEI M, HUANG XJ. Determination of trace phenolic acids in fruit juice samples using multiple monolithic fiber solid-phase microextraction coupled with high-performance liquid chromatography [J]. *Anal Methods*, 2016, 8(18): 3831–3838.
- [30] YAN J, HUANG XP, WU YY, *et al.* Determination of total sugar content in *Siraitia grosvenorii* by near infrared diffuse reflectance spectroscopy with wavelength selection techniques [J]. *J Food Meas Charact*, 2019, 13(4): 2713–2721.

(责任编辑: 于梦娇 韩晓红)

作者简介



李颖, 硕士, 助教, 主要研究方向为食品加工与检测。
E-mail: liying2015@qq.com



李占明, 博士, 讲师, 主要研究方向为农产品品质控制。
E-mail: lizhanming@just.edu.cn