高光谱成像技术结合机器学习和特征波段筛选的识别人参年限研究

李 梦[1,2]，周 聪[3]，王 慧[2]，杨 健[2]，张小波[2*]

1. 河南中医药大学，河南 郑州 450046
2. 中国中医科学院中药资源中心 道地药材品质保障与资源持续利用全国重点实验室，北京 100700
3. 江西省道地药材质量评价研究中心，江西 赣江 330000

摘 要：目的 为实现对人参 *Panax ginseng* 年限的准确、无损、低成本识别，拟建立一种基于高光谱成像技术结合机器学习的人参年限识别方法。**方法** 以 1～7 年生吉林通化地区人参为研究对象，分别在可见-近红外波段（visible-near infrared，VNIR）和短波红外波段（short-wave infrared，SWIR）范围内采集人参高光谱图像，共得到 84 份人参样品的高光谱图像及 1 680 个感兴趣区域的高光谱数据。在 VNIR、SWIR 和 VNIR＋SWIR 融合波段范围内，分别对人参样品高光谱数据进行多元散射校正（multiple scattering correction，MSC）、标准正态变化（standard normal variation，SNV）、Savitzky-Golay 平滑、一阶导（first-order derivative，FD）、二阶导（second-order derivative，SD）的预处理，然后分别结合偏最小二乘判别分析（partial least squares discriminant analysis，PLS-DA）、线性支持向量机（Linear SVC）判别分析方法，在以“药食”为区分的 2 分类识别，以大于、小于、等于 5 年为区分的 3 分类识别，以 7 个年份进行逐年区分的 7 分类识别的 3 个年限分类尺度，分别建立人参年限的识别模型。**结果** 在 VNIR 410～720 nm 内，同一波长下 1～7 年人参平均光谱反射率整体有依次降低的趋势。不同分类识别模型的混淆矩阵评估结果表明，SWIR 波段和融合波段经 FD 预处理后的 LinearSVC 模型在 3 个年限尺度下的分类效果较好，准确率较高，2、3、7 分类模型的预测集准确率分别为 99.60%、98.41%、95.24%。利用连续投影算法（continuous projection algorithm，SPA）筛选的特征波段建立的识别模型在 2 分类和 3 分类时精度较高，且所用波段更少，分类识别效率更高。**结论** 高光谱成像技术结合机器学习和特征波段筛选方法，可以较好实现对特定产地人参的年限识别，为实现该技术在人参年限识别和质量控制等实际应用方面提供参考。

关键词：高光谱成像；机器学习；特征波段；人参；年限识别

中图分类号：R282.1 **文献标志码：**A **文章编号：**0253 - 2670(2026)05 - 1887 - 09

DOI: 10.7501/j.issn.0253-2670.2026.05.025

Hyperspectral imaging technology combined with machine learning and characteristic band screening for *Panax ginseng* age identification research

LI Meng[1,2], ZHOU Cong[3], WANG Hui[2], YANG Jian[2], ZHANG Xiaobo[2]

1. Henan University of Chinese Medicine, Zhengzhou 450046, China
2. State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
3. Research Center for Quality Evaluation of Dao-di Herbs, Ganjiang 330000, China

Abstract: Objective To achieve accurate, nondestructive and low-cost identification of *Panax ginseng* age, a *P. ginseng* age identification method was established in this study based on hyperspectral imaging technology combined with machine learning. **Methods** Hyperspectral images of 84 *P. ginseng* samples and hyperspectral data of 1 680 regions of interest were obtained by acquiring *P. ginseng* hyperspectral images in the visible-near infrared (VNIR) and short-wave infrared (SWIR) bands from one to seven years old in Tonghua, Jilin, China, respectively. The hyperspectral data of ginseng samples were preprocessed with multiple scattering correction (MSC), standard normal variation (SNV), Savitzky-Golay smoothing, first-order derivative (FD) and second-order derivative (SD) in the VNIR, SWIR and VNIR + SWIR fusion bands, and then combined with partial least squares discriminant analysis

收稿日期：2025-09-02

基金项目：国家中医药管理局中医药创新团队及人才支持计划项目（ZYYCXTD-D-202005）；中国中医科学院科技创新工程（CI2021A03901）；中央本级重大增减支项目（2060302）

作者简介：李 梦，助理研究员，博士研究生，主要从事中药资源区划与中药质量评价研究。E-mail: limeng0642@163.com

***通信作者：**张小波，研究员，主要从事中药资源调查与区划等研究。E-mail: jack110007@163.com

(PLS-DA), linear Then, we combined PLS-DA and LinearSVC discriminant analysis methods to establish the identification models of *P. ginseng* years in two classification scales distinguished by "medicinal food", three classification scales distinguished by greater than, less than, and equal to five years, and seven classification scales distinguished by seven years, respectively. **Results** In the VNIR 410—720 nm band range, there was an overall trend of sequential decrease in the average spectral reflectance of *P. ginseng* from one year to seven years at the same wavelength. The results of confusion matrix evaluation of different classification recognition models showed that the LinearSVC model with FD preprocessing in SWIR band and fusion band had better classification and higher accuracy at three annual scales, and the prediction set accuracy of 2, 3 and 7 classification models were 99.60%, 98.41% and 95.24%, respectively. The recognition models built using the feature bands screened by the continuous projection algorithm (SPA) have higher accuracy at 2 and 3 classifications, and use fewer bands for more efficient classification and recognition. **Conclusion** Hyperspectral imaging technology combined with machine learning and feature band screening methods can better achieve the identification of the age of *P. ginseng* of specific origin, and provide a reference for realizing the practical applications of this technology in *P. ginseng* age identification and quality control.

Key words: hyperspectral imaging; machine learning; characteristic bands; *Panax ginseng* C. A. Meyer; age identification

人参 *Panax ginseng* C. A. Meyer 作为一种驰名中外的名贵中药材，早在秦汉时期已有应用，在《神农本草经》中被列为上品，具有“主补五脏，安精神，定魂魄，止惊悸，除邪气，明目，开心益智。久服，轻身延年”功效，常用作补药[1]，具有很高的药用价值。人参的生长年限是影响人参品质的重要因素。不同年限的人参在生长状况和物质代谢上存在很大的差距，近年来科研人员通过相关实验来研究测定不同年限人参中相关成分与指标间的差异。李向高[2]对不同生长年限的人参总皂苷含量进行研究，结果表明 4～6 年生人参皂苷随年生而增加。张万博等[3]对 5、8、18 年生人参皂苷含量进行测定，结果表明随着栽培年限的增加，各部位皂苷含量均有所提高。陈丽雪[4]研究结果表明 5 年和 6 年生人参对免疫缺陷小鼠的免疫效果要好于 3 年和 4 年生人参。

目前，人参的年限识别常采用性状鉴别法、显微鉴别法、色谱法、光谱法、质谱法、分子鉴别技术[5]等方法。如詹达琦等[6]将小波变换处理方法用于不同生长年限人参的二维红外相关光谱数据预处理中，进行人参生长年限的鉴别。余江锋等[7]利用高效液相色谱法进行人参中皂苷含量测定，根据人参中 8 种主要人参皂苷总量以及人参皂苷 Re、人参皂苷 Rb_1、人参皂苷 Rc、人参皂苷 Rd 单体含量与人参皂苷 Rg_1 比值的变化规律构建模型推测人参生长年限。崔绍庆[8]基于人工嗅觉系统鉴别不同年限人参，效果良好。目前的人参年限识别方法如性状和显微鉴别主观性相对较强，化学和分子鉴别等方法需要对人参样品进行有损处理，对仪器和操作方法有较高要求，成本也相对较高。因此，研究一种准确、无损、低成本的人参年限识别方法十分必要。

高光谱成像技术具有无损、快速、无污染等特点，测量前后被测样品的性质不发生任何变化，可同时获取样品的光谱信息和空间图像信息，能够准确反映出所测样品的物理和化学信息，可实现快速无损的鉴别和检测[9]，已广泛用于农业、食品、环境和医学等领域[10]。且由于高光谱具有较高分辨率，数据量较大，所以常筛选重要的特征波段来简化模型和提高精度。高光谱成像技术已在中药的真伪鉴别、无损检测、含量测定、产地区分、质量研究等方面有了研究和应用，如应用于甘草[11]、苦杏仁和桃仁[12]、人参[13]、枸杞[14]等中药的产地和真伪鉴别。在年限识别方面，已对陈皮[15]、白茶[16]、小麦种子[17]、玉米种子[18]、棉种[19]等食品和农作物的年份进行了研究。

基于高光谱成像技术在中药和年限识别等方面的研究应用实例，本研究利用高光谱成像技术结合机器学习和特征波段筛选的方法对人参年限进行识别，建立人参年限识别分类模型，试比较不同预处理、建模等方法对年限识别分类模型精度的影响，为人参的年限识别和质量控制提供新的参考方法。

1 材料

1.1 样品

本研究所用的人参样品均来自于吉林省通化市同一种植基地，以减少除年限外的其他因素干扰，在该基地同时采集 1～7 年的每个生长年限各 12 份人参样品。经中国中医科学院中药资源中心金艳副研究员鉴定为五加科植物人参 *P. ginseng* C. A. Meyer 的干燥根和根茎。

1.2 仪器

高光谱成像设备为 HySpex 系列高光谱成像光谱仪（Norsk Elektro Optikk A/S，Norway），其主要由 2 个卤钨灯（150W/12V，H-LAM Norsk Elektro Optikk，Oslo，Norway）、CCD 探测仪、移动平台、可见-近红外相机 SN0605 VNIR（410～990 nm，H-V16，Norsk Elektro Optikk，Oslo，Norway）与短波红外相机 N3124 SWIR（950～2 500 nm，H-S16，Norsk Elektro Optikk，Oslo，Norway）组成。

2 方法

2.1 数据获取及分析

2.1.1 高光谱图像采集和校正 2 个相机分别对放置在黑色背景中的人参样品通过线性扫描方式进行高光谱图像采集，将 2 个相机镜头与样品的距离设为 25 cm，积分时间（integration tume）和帧周期（frame period）分别设为 VNIR（3 800、18 000 μs）和 SWIR（4 500、46 928 μs）；平台移动速度为 1.5 mm/s，光谱分辨率均为 6 nm。将样品平行有间隔的方式整齐摆放在传送带上，同时保证样品处于相机扫描的有效范围内，将用于黑白校正的白板摆放在样品后方 5 cm 处，VNIR 相机获取 108 个波段的图像，SWIR 相机获取 288 个波段的图像。在高光谱图像数据采集完成后，为消除仪器、电流等外部因素对样品数据的影响，利用仪器自带 RAD 校正软件对原始高光谱图像进行 RAD 校正。随后使用 ENVI 5.3 软件（ITT Visual Information Solutions, Boulder，CO，美国）进行黑白板校正，黑白板校正为光谱数据处理中的一种常用方法，通常认定黑板（移动平台）反射率为 0，具有朗伯特性的标准白板反射率为 1。将样本与白板、黑板进行计算得到相对反射率。计算公式如下：

$$R=(I_R-I_B)/(I_w-I_B) \quad (1)$$

R 是校正后的反射率图像，I_R 是原始反射率图像，I_w 是白板图像，I_B 是黑板参考图像。

2.1.2 提取感兴趣区域（region of interest，ROI）的光谱数据 完成图像校正后，为保证获取样品准确的光谱信息，利用 ENVI 5.3 软件中的 ROI 进行数据提取。计算每个 ROI 内的平均相对反射率，作为样品的原始光谱数据，用于后续处理分析。本研究对感兴趣区域内各年限人参样品的高光谱数据均值进行方差分析，初步判断各年限人参样品的高光谱差异性。

2.1.3 数据预处理 本研究使用多元散射校正（multiplicative scatter correction，MSC）、Savitzky-Golay 平滑（SG 平滑）、标准正态变换（standard normalized variate，SNV）、一阶导数（first derivative，FD）和二阶导数（second derivative，SD）5 种常用的光谱预处理方法对所得人参的高光谱数据进行预处理。

2.1.4 构建分类识别模型 本研究使用偏最小二乘判别分析（partial least square-discriminant analysis，PLS-DA）、线性支持向量机分类器（linear support vector classification，LinearSVC）这 2 种分类方法建立人参年限识别模型。PLS-DA 是一种使用偏最小二乘回归（partial least squares regression，PLSR）的高维线性判别分类模型，属于多元分类模型。LinearSVC 是基于 liblinear 库实现，与基于 libsvm 库的支持向量机分类（SVC）相比，对较大训练集样本量能够很好的进行归一化。

为深入分析不同分类方式对人参年限识别结果的变化，本研究对人参年限识别进行了 3 种年份尺度划分，一是对 1～5 年和 6～7 年“药食”区分[20]的人参年限 2 分类识别，二是进一步明确以 5 年为界进行的年限小于 5 年、等于 5 年、大于 5 年的人参年限 3 分类识别，三是对 7 个年份进行逐年区分的识别，即人参年限 7 分类识别。计算得到每种年限尺度分类模型的识别精度。识别精度为验证测试样本集中识别正确的样本数，除以验证测试样本总数，以百分比表示。

2.1.5 特征波段筛选 高光谱波段范围广、数据量大，为降低模型运算时间和复杂度，本研究采用连续投影算法（successive projections algorithm，SPA）提取特征波段，通过前向迭代搜索特征变量算法最大限度地消除高光谱数据的共线信息和冗余信息[21]。

2.1.6 模型评价 建立分类模型后，本研究采用混淆矩阵（confusion matrix）对模型的分类性能进行评价。混淆矩阵又称误差矩阵，是机器学习中常用的分类模型评估方法。混淆矩阵是一种可视化的模型评价方法，矩阵中每一行代表各年限的预测值，每一列代表各年限的真实值，从混淆矩阵中可以准确直观的看出模型的预测结果与每一个类别的分类情况。

2.2 数据处理及分析所用软件

本研究所用的图像校正软件为高光谱成像系统 HySpex RAD 软件（Norsk Elektro Optikk，Oslo，

Norway），黑白板校正和ROI提取使用的软件为ENVI 5.3，数据预处理、分类识别模型构建及特征波段筛选等使用的软件为 Matlab 2020b 和 Python 3.7。

3　结果与分析

3.1　人参样品高光谱图像

将不同年限的人参样品分批摆放在平台上，通过 2 个相机采集并经过 RAD 校正，得到 VNIR 和 SWIR 在 2 个不同范围波段下的人参样品高光谱图像，见图 1。

图 1　人参样品高光谱图像示意图

Fig. 1　Hyperspectral image diagram of *P. ginseng* samples

3.2　人参样品 ROI 数据集及划分

为保证准确提取每份人参样本的光谱信息，手动提取不同波段范围内样本高光谱图像 ROI 内的高光谱数据，每份人参样品提取 20 个大小相近的 ROI，每个年限各 12 份人参样品，最终共获得 7 个年限人参样品共 1 680 个样本数据，根据随机划分法将数据集划分为训练集和预测集，其中训练集和预测集占比为 7∶3，即训练集 1 176 个数据，预测集 504 个数据。

3.3　原始光谱曲线

对不同相机采集的图像分别进行数据处理，为综合分析人参样品的高光谱曲线规律，合并 2 个相机得到的样品光谱数据，即得到覆盖 410～2 500 nm 的可见-短波红外波段（即融合波段）光谱数据。由于 2 个相机拍摄的灯光照射角度以及相机参数等因素的影响，在波长 950 nm 附近形成曲线断层，但每个样品的拍摄条件相同且在同一环境下进行扫描，数据具有可比性。不同年限人参原始数据光谱曲线图与平均光谱曲线图见图 2。

由图 2 可知，不同年限人参光谱曲线具有相似趋势，但各年限间有一些比较明显的差别。由平均

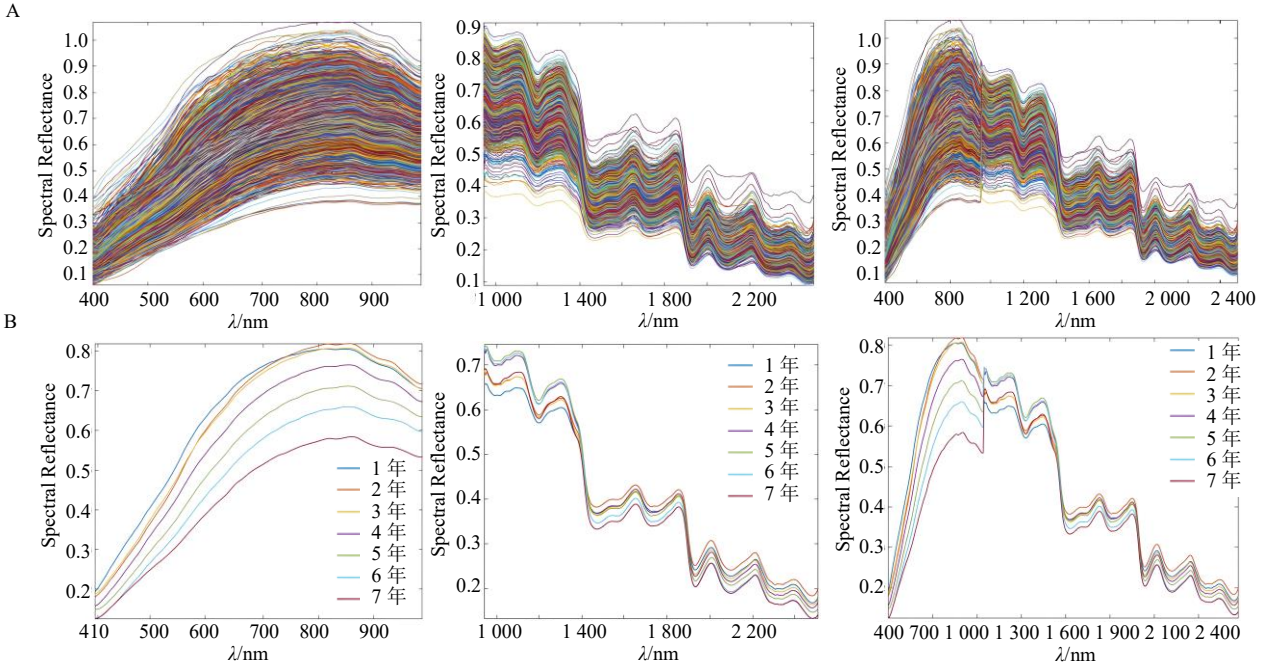

图 2　不同年限人参样品原始光谱 (A) 和平均光谱 (B) 曲线图

Fig. 2　Curves of original spectra (A) and average spectra (B) of *P. ginseng* samples of different ages

光谱曲线图可看出，在 VNIR 410～720 nm 内，整体上有比较明显表现出在同一波长从 1～7 年人参平均光谱反射率依次降低的趋势，其中 2 年和 3 年的人参平均光谱反射率差异比较小，通过对 1～7 年人参的平均光谱反射率值进行方差分析（ANOVA），得到 VNIR 范围每个波段下 7 个年限

人参样品 ANOVA 的统计量 *F* 值。由图 3 可证明该波段范围内 1 年至 7 年人参的平均光谱反射率具有显著性差异，猜测随着人参年份的增长，其内部化学成分含量不断累积形成差异，并在光谱反射率上得到显现。在 VNIR 410～990 nm 内，4～7 年人参平均光谱反射率均是依次降低，证明 4～7 年人参内部化学成分含量差异较为显著。在 SWIR 范围的 1 700 和 2 100 nm 附近，4～7 年人参也呈现出平均光谱反射率依次降低趋势，其他波长下无明显规律趋势。

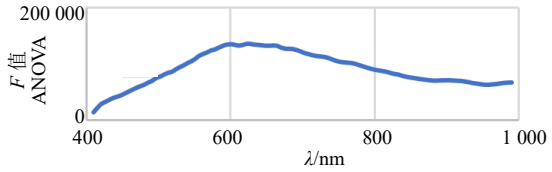

图 3 VNIR 波段下不同年限人参样品的 *F* 值

Fig. 3 *F*-values of *P. ginseng* samples of different ages at VNIR bands

3.4 光谱预处理曲线

使用 MSC、SG、SNV、FD 和 SD 这 5 种预处理方法，分别对不同年限人参样品的 VNIR、SWIR 以及融合波段的原始光谱数据进行预处理。以融合波段为例，预处理后得到的光谱曲线图见图 4。

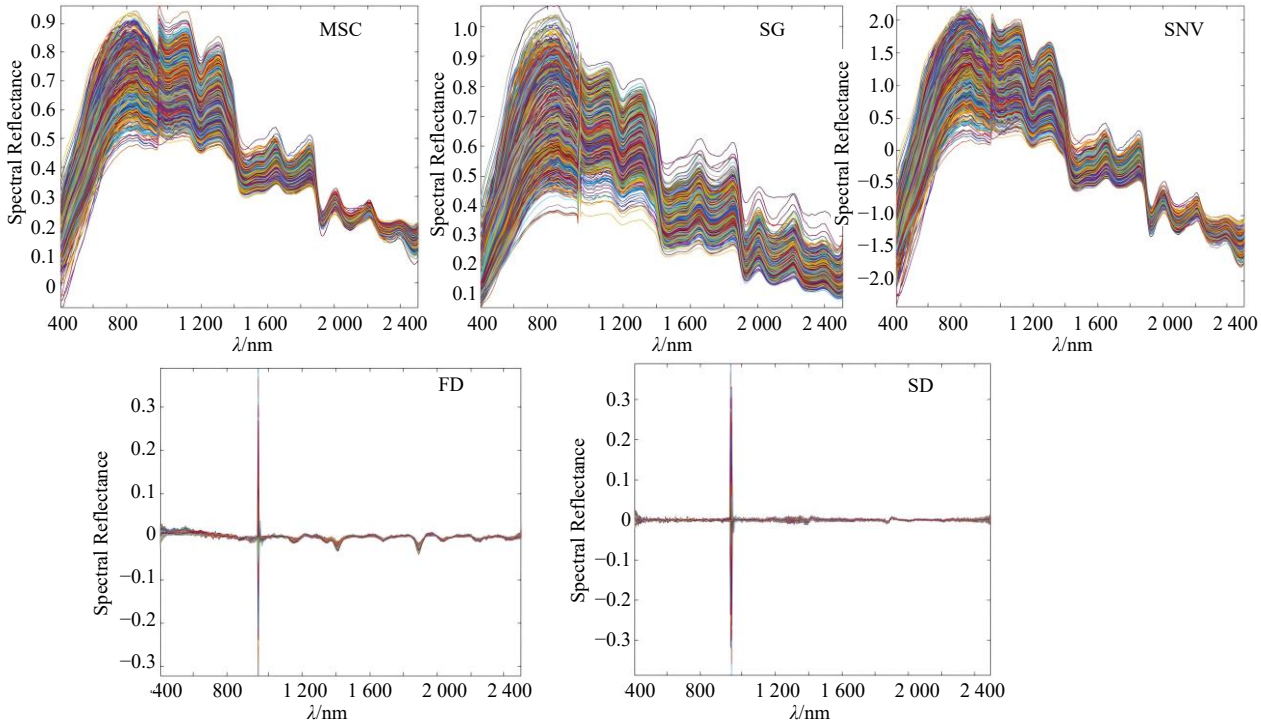

图 4 融合波段原始光谱数据分别经 MSC、SG、SNV、FD、SD 预处理后光谱曲线图

Fig. 4 Spectral curves of raw fusion-band data preprocessed by MSC, SG, SNV, FD, SD respectively

3.5 分类识别模型构建及结果

偏最小二乘判别分析（PLS-DA）是将训练集的数据转换为用于预测验证集类的中间潜在变量。由于适当数量的潜在变量可以完全描述数据，为了对不同类别的样本进行最好的区分，采用十次交叉验证法来获得最佳的潜在变量数量。过多的潜在变量会使模型无法拟合，因此在本研究中潜在变量的数量被限制在 15 以内。线性支持向量机分类器（LinearSVC）是使用 One-vs-All（也称 One-vs-Rest）实现多分类的算法，对数据量较大的模型有更好的表现，适合用于多分类模型。

将人参样品的原始光谱数据及经 MSC、SG、SNV、FD 和 SD 预处理后的高光谱数据，分别结合 PLS-DA、LinearSVC 这 2 种分类算法，以及 3 种分类尺度方法，建立分类识别模型，并分别计算不同分类识别模型预测集的准确率。建模结果见表 1、2。原始光谱数据经过预处理后，绝大部分都能够明显提高模型准确率，尤其是 MSC、SNV、FD、SD 几种预处理方法为佳。整体上以经 FD 预处理后的 LinearSVC 模型分类效果较好。SWIR 波段和融合

表 1 不同年限人参样本在各波段不同预处理的 PLS-DA 模型准确率

Table 1 Accuracy of PLS-DA models with different preprocessing methods of *P. ginseng* samples in each band of different ages

波段	预处理方式	2 分类准确率/%		3 分类准确率/%		7 分类准确率/%	
		训练集	预测集	训练集	预测集	训练集	预测集
VNIR	无	89.20	94.05	90.68	85.03	92.80	88.07
	MSC	96.03	96.35	90.34	87.52	91.75	89.39
	SG	94.74	93.28	90.42	86.37	90.71	87.31
	SNV	96.89	95.78	90.68	85.60	91.84	89.58
	FD	95.77	94.43	89.30	85.22	93.14	89.58
	SD	95.25	93.09	88.01	83.88	91.23	86.17
SWIR	无	99.22	98.85	94.13	92.32	94.97	91.67
	MSC	98.88	97.89	93.53	89.64	94.36	90.91
	SG	98.88	98.85	94.22	92.13	95.31	91.67
	SNV	98.71	97.31	94.13	88.48	94.44	90.72
	FD	99.65	98.46	96.72	91.94	96.96	89.77
	SD	96.72	93.09	91.20	83.11	93.40	79.55
全波段	无	99.22	99.42	95.69	94.05	96.44	93.18
	MSC	99.57	99.42	95.69	95.20	96.88	93.75
	SG	99.05	99.04	95.25	93.09	95.83	92.23
	SNV	99.31	99.23	95.94	94.63	96.44	92.99
	FD	99.91	99.42	97.93	94.05	99.22	94.32
	SD	99.31	96.93	95.08	87.91	97.83	87.31

表 2 不同年限人参样本在各波段不同预处理的 LinearSVC 模型准确率

Table 2 Accuracy of LinearSVC models for *P. ginseng* samples of different ages under various bands and preprocessing methods

波段	预处理方式	2 分类准确率/%		3 分类准确率/%		7 分类准确率/%	
		训练集	预测集	训练集	预测集	训练集	预测集
VNIR	无	89.54	85.52	80.87	82.54	65.22	61.90
	MSC	97.11	95.63	93.62	93.45	91.75	88.69
	SG	89.45	85.91	80.52	82.14	66.15	62.69
	SNV	97.02	96.23	94.30	92.26	91.50	85.51
	FD	99.57	98.21	98.72	96.23	98.98	93.65
	SD	99.82	95.44	99.23	92.26	99.66	78.57
SWIR	无	96.17	96.82	88.86	91.86	77.89	75.59
	MSC	99.65	98.41	98.81	96.61	97.02	93.25
	SG	95.83	95.83	89.03	87.10	76.61	77.78
	SNV	99.49	98.81	98.98	97.62	97.70	93.45
	FD	100.00	99.60	100.00	98.41	99.83	93.85
	SD	100.00	96.43	100.00	95.63	99.91	82.94
全波段	无	97.28	95.04	91.92	91.67	85.88	81.35
	MSC	99.32	99.21	97.53	95.63	94.60	89.85
	SG	96.94	96.63	92.26	90.47	84.70	82.34
	SNV	99.57	99.01	96.85	94.24	94.39	89.88
	FD	100.00	99.40	100.00	98.41	100.00	95.24
	SD	100.00	97.22	100.00	96.23	100.00	84.13

波段模型准确率相对较高。

在 VNIR 波段范围，经 FD 预处理后的 LinearSVC 模型最佳，各分类尺度预测集准确率分别为 98.21%、96.23%、93.65%；在 SWIR 波段范围，经 FD 预处理后的 LinearSVC 模型最佳，各分类尺度预测集准确率分别为 99.60%、98.41%、93.85%；在融合波段范围，经 FD 预处理后的 LinearSVC 模型最佳，各分类尺度预测集准确率分别为 99.40%、98.41%、95.24%。

在 2 分类尺度，SWIR-FD-LinearSVC 模型最佳，预测集准确率 99.60%；在 3 分类尺度，SWIR 和融合波段-FD-LinearSVC 模型均为最佳，预测集准确率 98.41%；在 7 分类尺度，融合波段-FD-LinearSVC 模型最佳，预测集准确率 95.24%。

因此，SWIR/融合波段-FD-LinearSVC 模型比较适合用于人参的年限分类识别。

3.6 SPA 特征波段筛选

经过预处理后的光谱数据可以有效提升模型预测能力，但由于高光谱波段之间具有强相关性，有大量与分类识别无关的光谱信息。且上述研究表明 FD 预处理后模型预测结果普遍更好，因此尝试将 SWIR 和融合波段经过 FD 预处理后的光谱数据作为输入，利用连续投影方法（SPA）筛选特征波段，筛选出的特征波段分布图见图 5。

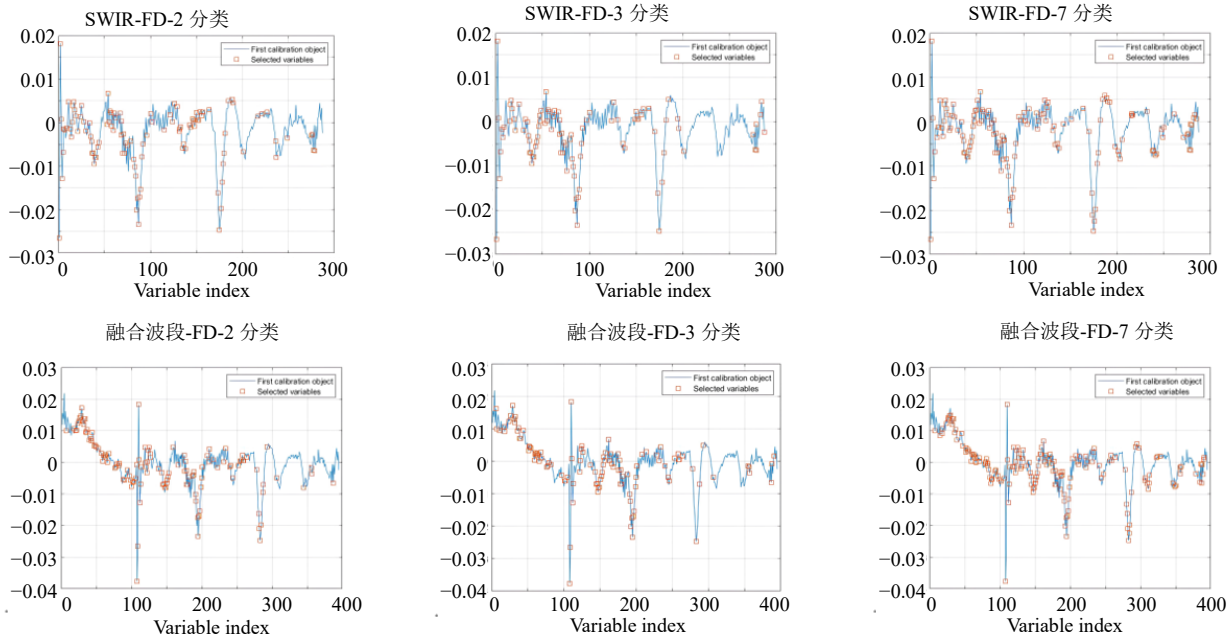

图 5 SWIR 波段及融合波段经 FD 预处理后在各年份分类尺度下特征波段分布图

Fig. 5 Distribution of characteristic bands under different age classification scales based on FD preprocessing in SWIR band and fusion band

在各波段下筛选出的特征波段数目及分布各不相同。2、3、7 分类 3 个年份分类尺度下，SWIR 波段筛选出的特征波段数目分别为 65、62、75。融合波段筛选出的特征波段数目分别为 139、125、113。SWIR 波段 FD 预处理后筛选出的特征波段数目分别为 107、93、134。融合波段 FD 预处理后筛选出的特征波段数目分别为 124、109、165。从筛选结果上看出，SWIR 波段利用 SPA 算法筛选特征波段后，波段数目明显减少，但经 FD 预处理后，特征波段数目反倒增加。对各波段筛选出的特征波段再次建立分类识别模型。建模结果见表 3。

筛选特征波段后的建模结果显示 3 个分类尺度的模型预测集准确率较好，LinearSVC 模型的预测集识别准确率均能达到 92%以上，模型所用波段减少为原始数据波段的 30%～40%，能够保证模型的准确性和精度，可一定程度上提高模型的运算效率，剔除不重要波段的影响。筛选后的 SWIR 波段与融合波段特征波段在 2 分类和 3 分类时精度较高，且所用波段更少，分类识别效率更高。但部分结果预测集准确率略有降低。可见，对于人参年限的识别，需不断调整参数优化模型，并根据建模结果判断是否需要进行特征波段的筛选，或选择其他特征波段筛选方法筛选得到更具代表性的波段。

表 3　特征波段年限识别分类模型准确率

Table 3　Accuracy of characteristic band year recognition classification model

分类方式	波段	数量（占比/%）	分类模型	准确率/%	
				训练集	预测集
2 分类	VS-FD	124（31%）	LinearSVC	99.91	99.21
			PLS-DA	99.55	98.86
	S-FD	107（37%）	LinearSVC	99.83	99.01
			PLS-DA	98.87	98.01
3 分类	VS-FD	109（27%）	LinearSVC	99.74	97.62
			PLS-DA	95.18	92.33
	S-FD	93（32%）	LinearSVC	99.06	97.82
			PLS-DA	93.37	90.91
7 分类	VS-FD	165（42%）	LinearSVC	99.83	94.44
			PLS-DA	96.76	92.05
	S-FD	134（46%）	LinearSVC	99.57	92.86
			PLS-DA	93.22	86.93

3.7　模型评价

采用混淆矩阵对人参年限分类模型进行评估，具体评价指标包括召回率（Recall）、精确度（Precision）、F1 分数（F1 Score）等，这些评价指标从不同的角度解释了分类模型的精度，具体公式如下：

召回率＝TP/(TP＋FN)　（1）

精确度＝TP/(TP＋FP)　（2）

F1 分数＝2×(召回率×精确度)/(召回率＋精确度)　（3）

其中，TP 代表真阳性，即预测值落到真实值上的样本个数；FN 代表假阴性，即预测值落到非真实值上的样本个数；FP 代表假阳性，即非预测值落到真实值上的样本个数。以融合波段-FD-LinearSVC 模型为例，其判别结果的混淆矩阵见图 6。

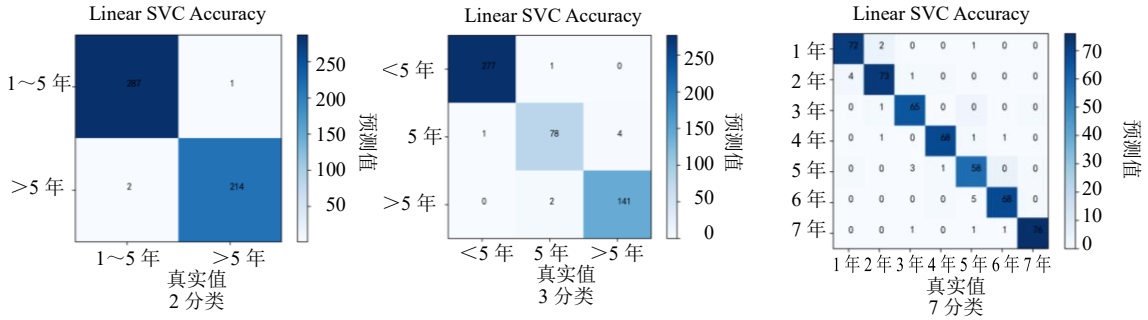

图 6　融合波段-FD-LinearSVC 模型判别结果的混淆矩阵

Fig. 6　Confusion matrix of discriminant result from fusion band-FD-LinearSVC model

融合波段-FD-LinearSVC 和 SWIR-FD-LinearSVC 分类预测模型在各年限尺度下的混淆矩阵预测精度均达到 92%以上，召回率和 F1 分数也均很高，表示分类模型均具有很好的性能。

4　讨论

本研究建立了一种基于高光谱成像技术结合机器学习和特征波段筛选对不同年限人参进行识别的方法，相比于其他传统及常用方法有更加无损、快速、低成本等优势，且对于非专业技术人员更简易方便，Kwon Yong-Kook 等[22]利用傅里叶变换红外光谱来鉴别人参叶片的栽培年限，指出光谱信息可以用来判断人参的生长年份。本研究主要以人参药材为研究对象，建立的方法更加适用于药材市场和流通等环节。通过对不同年限人参样品的平均光谱曲线分析，1～7 年人参的平均光谱反射率具有显著性差异，且在 VNIR 410～720 nm 内，7 个年限人参样品的平均光谱曲线整体上有比较明显的从 1～7 年人参平均光谱反射率依次降低的趋势。不同年限人参光谱间的明显差异，也进一步证明利用本方法进行人参年限识别的合理性和可靠性。

本研究采用 PLS-DA、LinearSVC 分类方法结合多种预处理方式和 SPA 特征波段筛选方法建立

的分类识别模型，可以对不同年限的人参进行高精度识别。同时本研究在 1～5 年和 6～7 年“药食”区分的人参年限 2 分类识别、以 5 年为界进行的年限小于 5 年、等于 5 年、大于 5 年的人参年限 3 分类识别、对 7 个年份进行逐年区分的 7 分类识别 3 个年限分类尺度的区分上，分别建立了识别模型，对比 3 个不同分类尺度的年限识别模型结果，可以发现 PLS-DA 和 LinearSVC 方法下人参年限 2 分类识别模型预测准确率均高于 3 分类和 7 分类，说明分类尺度不同导致识别模型预测难度也不同，2 分类识别由于人参年限的差异相比 7 分类识别的人参逐年之间差异更加明显，所以 2 分类识别难度相对较低，预测准确率更高。

经过 MSC、SNV、FD、SD 4 种方法预处理的样品，大部分能够明显提高模型准确率，尤其以 FD 预处理方法为佳。整体上以 SWIR 波段和融合波段经 FD 预处理后的 LinearSVC 模型分类效果较好，证明该方法组合更加适合进行人参的年限识别。在此基础上，利用 SPA 法筛选经 FD 预处理后的 SWIR 波段及融合波段的特征波段，并再次建立识别模型，所用波段数仅为原始波段的三分之一，可以在 2 分类和 3 分类尺度上实现更高效率的高精度识别。

本研究通过高光谱数据信息结合机器学习和特征波段筛选方法，可以在特定产地人参的年限识别方面得到较好的应用，而高光谱成像技术还包括丰富的图像空间信息，未来计划尝试将光谱信息和图像信息相结合，深入挖掘有效信息提高识别能力，为人参等药材的质量控制和轻量化检测仪器设备的开发提供基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 林仲凡. 有关人参的历史考证 [J]. 中国农史, 1985, 4(4): 78-84.
[2] 李向高. 人参加工炮制前后化学成分的变化 [J]. 中药通报, 1986, 11(4): 2-7.
[3] 张万博, 代月, 廉美兰, 等. 不同栽培年限人参不同部位中皂苷含量的分析 [J]. 延边大学农学学报, 2016, 38(1): 13-17.
[4] 陈丽雪. 不同年生人参根及 5 年生人参不同部位免疫活性的对比研究 [D]. 长春: 吉林农业大学, 2019.
[5] 线小云, 李葵秀, 李满桥, 等. 人参属药用植物种质资源研究进展 [J]. 中草药, 2025, 56(1): 360-373.
[6] 詹达琦, 张晓明, 孙素琴. 基于小波变换的二维红外相关光谱鉴别人参的生长年限 [J]. 光谱学与光谱分析, 2007, 27(8): 1497-1501.
[7] 余江锋, 李育平, 何伟, 等. 吉产不同生长年限人参中 8 种主要人参皂苷与人参皂苷 Rg_1 比值的变化规律研究 [J]. 中国药房, 2019, 30(1): 31-35.
[8] 崔绍庆. 基于不同纳米材料修饰的 QCM 气敏传感器的制备及人工嗅觉系统的实现 [D]. 杭州: 浙江大学, 2015.
[9] Tankeu S, Vermaak I, Chen W Y, *et al*. Differentiation between two “Fang Ji” herbal medicines, *Stephania tetrandra* and the nephrotoxic *Aristolochia fangchi*, using hyperspectral imaging [J]. *Phytochemistry*, 2016, 122: 213-222.
[10] 李梦, 李静, 张小波. 高光谱成像技术的发展现状及其在中药领域中的应用前景 [J]. 西部中医药, 2021, 34(10): 149-153.
[11] 殷文俊, 茹晨雷, 郑洁, 等. 基于高光谱成像技术融合光谱和图像特征鉴别不同产地的甘草 [J]. 中国中药杂志, 2021, 46(4): 923-930.
[12] 郑洁, 茹晨雷, 张璐, 等. 基于近红外高光谱成像技术对不同产地苦杏仁和桃仁药材的鉴别 [J]. 中国中药杂志, 2021, 46(10): 2571-2577.
[13] 李梦, 张小波, 刘绍波, 等. 部分可解释机器学习方法的高光谱人参产地识别和分析 [J]. 光谱学与光谱分析, 2022, 42(4): 1217-1221.
[14] 王磊, 覃鸿, 李静, 等. 近红外高光谱图像的宁夏枸杞产地鉴别 [J]. 光谱学与光谱分析, 2020, 40(4): 1270-1275.
[15] 鲍一丹, 吕阳阳, 朱红艳, 等. 陈皮年份的高光谱技术鉴别研究 [J]. 光谱学与光谱分析, 2017, 37(6): 1866-1871.
[16] 陈书媛, 张友超, 杨杰, 等. 基于高光谱成像技术的白茶储藏年份判别 [J]. 食品工业科技, 2021, 42(18): 276-283.
[17] 孙梅, 黄宇, 陈兴海, 等. 基于高光谱图像的小麦种子年份快速鉴别分析研究 [J]. 中国粮油学报, 2022, 37(1): 170-174.
[18] 王庆国, 黄敏, 朱启兵, 等. 基于高光谱图像的玉米种子产地与年份鉴别 [J]. 食品与生物技术学报, 2014, 33(2): 163-170.
[19] 段龙, 鄢天荥, 王江丽, 等. 结合高光谱成像和机器学习的棉种年份鉴别 [J]. 光谱学与光谱分析, 2021, 41(12): 3857-3863.
[20] 中华人民共和国国家卫生健康委员会 (原卫生部).关于批准人参 (人工种植) 为新资源食品的公告 (2012 年第 17 号) [EB/OL]. (2012-09-05) [2023-02-26]. http://www.nhc.gov.cn/sps/s7891/201209/e94e15f2d9384b6795597ff2b101b2f1.shtml.
[21] Zhang L, An D, Wei Y G, *et al*. Prediction of oil content in single maize kernel based on hyperspectral imaging and attention convolution neural network [J]. *Food Chem*, 2022, 395: 133563.
[22] Kwon Y K, Ahn M S, Park J S, *et al*. Discrimination of cultivation ages and cultivars of ginseng leaves using Fourier transform infrared spectroscopy combined with multivariate analysis [J]. *J Ginseng Res*, 2014, 38(1): 52-58.

[责任编辑 时圣明]