

应用高光谱鉴别黑枸杞和唐古特白刺果

赵凡, 闫昭如, 宋海燕

山西农业大学工学院, 山西 太谷 030801

摘要 黑枸杞含有花青素、多糖、氨基酸和微量元素等多种营养成分, 具有极高的经济和医药价值, 其市场价格很高。唐古特白刺果外观和黑枸杞极为相似, 其价格较低, 经常被用于冒充黑枸杞。高光谱图像技术结合图像和光谱于一体, 常用于食品检测和识别等领域。结合高光谱图像技术, 无损识别黑枸杞和唐古特白刺果。采集黑枸杞(180份)和唐古特白刺果(180份)的高光谱图像, 利用掩膜提取光谱, 光谱范围为900~1700 nm, 共254个波段, 去除前22个异常波段。采用Kennard-Stone法划分样品, 校正集: 预测集=2:1; 采用连续投影算法(SPA)法对光谱进行降维, 设定提取特征波长范围为0~30, 最终提取特征波长为20个; 分别将全光谱(FS)和SPA提取的20个特征波长作为模型输入, 建立支持向量机(SVM)和极限学习机(ELM)识别模型。结果表明, 基于FS和SPA建立的SVM模型识别率为100%; 基于FS和SPA建立的ELM模型识别率为100%; SPA法在不降低模型识别精度的情况下, 能减少模型输入, 输入仅为FS的8.62%, 大大降低模型运算量。此研究为识别黑枸杞和唐古特白刺果提供了参数。

关键词 高光谱; 模型; 鉴别; 黑枸杞; 唐古特白刺果

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)07-2240-05

引言

黑枸杞是中国西北地区特有的茄科枸杞类野生植物, 它在食品、医药等方面具有极高的经济价值, 拥有“沙漠软黄金”和“草原上的黑珍珠”等美誉^[1], 黑枸杞具有降血压、抗衰老、提高机体免疫等保健作用^[2-3]。其中含有丰富的多糖、黄酮、氨基酸、矿物质和微量元素(B1, B2, Ca, Mg, Cu, Zn和Fe)等各种营养成分; 黑枸杞含有花青素, 是人体天然抗氧化剂。和普通红枸杞相比, 黑枸杞的氨基酸含量是前者2倍多, Zn, Cu, Pb, Mg, Ni和Cr等微量元素含量和单位多糖、黄酮、多酚含量均明显高于红枸杞^[4]。

近些年来, 随着野生黑枸杞的人群青睐程度越来越高, 出现了人工种植黑枸杞。无论是野生黑枸杞, 还是种植黑枸杞, 市场价格都非常高。一些不法商家利用和黑枸杞无法区分的低价产品冒充黑枸杞, 例如唐古特白刺果, 主要分布于中国西北地区, 具有一定的医药作用。但是唐古特白刺果市场价格远远低于黑枸杞, 而外形、颜色、大小和黑枸杞非常相似, 仅凭肉眼很难区分, 往往冒充黑枸杞被售卖。鉴别黑枸杞和白刺果已成为黑枸杞市场急需解决的问题之一。

高光谱图像技术在农产品检测和识别方面有着非常广泛

的应用前景。Liu等^[5]对带真菌和有损伤的草莓进行识别; Dong等^[6]利用高光谱图像对不同浓度的猕猴桃膨大果进行识别; Yang^[7]利用高光谱图像技术检测熟牛肉储存过程中的新鲜度。目前高光谱图像技术鉴别黑枸杞和白刺果尚无人报道。

为研究高光谱图像技术鉴别黑枸杞和白刺果, 以黑枸杞干果和唐古特白刺干果为研究对象, 建立支持向量机(support vector machine, SVM)和极限学习机(extreme learning machine, ELM)识别模型, 并采用连续投影算法(successive projections algorithm, SPA)提取特征波长, 比较全光谱(full spectrum, FS)和连续投影算法对模型精度的影响。

1 实验部分

1.1 材料

实验所用黑枸杞和唐古特白刺果均由青海千拓贸易有限公司提供, 原产地为青海。黑枸杞根据颗粒大小分为特级(0.6 cm以上)、高级(0.5~0.6 cm)、中级(0.4~0.5 cm)三级, 本工作选用颗粒在0.4~0.5 cm范围的中级野生和种植黑枸杞作为实验材料; 同时选用颗粒与黑枸杞近似的0.4~0.5 cm范围的唐古特白刺果为实验材料。白刺果和黑枸杞如图1所示。实验材料去除杂质, 黑枸杞去除果柄。黑枸杞和

收稿日期: 2020-07-16, 修订日期: 2020-11-29

基金项目: 国家重点研发计划项目(2018YFD0700300)资助

作者简介: 赵凡, 女, 1989年生, 山西农业大学工学院教师 e-mail: 1140117238@qq.com

白刺果每份样本质量均为 (5 ± 0.1) g。野生黑枸杞和种植黑枸杞各 90 份,黑枸杞样品总数为 180 份;白刺果样品数为 180 份。样品总数为 360 份。



图 1(a) 唐古特白刺果

Fig. 1(a) Nitraria Tanggu



图 1(b) 黑枸杞

Fig. 1(b) Black Goji berries

1.2 仪器和设备

高光谱图像系统: GaiaSorter“盖亚”高光谱分选仪北京汉光卓立公司; 4 个 35 W 溴钨灯、电控平台和计算机等部件。图像光谱范围是 900~1 700 nm; 光谱分辨率为 3.19 nm; 设定曝光时间为 10 ms; 物距为 20 cm; 图像采集速率为 $7.2 \text{ mm} \cdot \text{s}^{-1}$ 。

1.3 光谱数据采集

仪器箱体内存在暗电流、光源分布不均匀, 这些因素会使采集到的高光谱图像含有较大噪音, 故需对高光谱图像进行黑白校正^[8-9]。见式(1)

$$R\% = \frac{R_0 - B}{W - B} \times 100\% \quad (1)$$

式(1)中: R_0 为反射光谱图像; W 为白板漫反射图像; B 为暗图像; R 为校正后漫反射光谱图像。

利用 ENVI4.8 软件建立掩膜提取高光谱图像。所有样品均选取第 130 波段处的图像进行阈值分割。当黑枸杞和白刺果阈值分别为 0.18 和 0.16 时, 能够分别提取完整的黑枸杞和白刺果图像, 因此分别设定阈值为 0.18 和 0.16 进行提取黑枸杞和白刺果图像。

将每份黑枸杞的平均光谱作为此样品的反射光谱; 将每份白刺果的平均光谱作为此样品的反射光谱。

1.4 样品划分和光谱降维

K-S 算法已经被证明在选择代表性样品方面的具有很好的效果, 例如朱晓琳采用 K-S 法对库勒勒香梨用进行划分^[10], Angelo 等采用 K-S 法对意大利生产的红蒜进行样品

划分^[11]。因此本研究采用 K-S 法划分样品。

SPA 法可以在高光谱庞大复杂的数据中去除冗余数据、提取特征波长数据^[12]。近些年, 许多国内外学者利用光谱分析技术检测农产品某些成分或者分类时常采用 SPA 法作特征波长的选取。本研究采用 SPA 法对光谱降维。

1.5 建模方法

(1) SVM 模型

SVM 是一种在特征空间上的间隔最大的线性分类器, 能够正确划分训练集, 求解几何间隔最大的分离超平面^[13], 从而来解决庞大复杂数据的分类问题。

(2) ELM 模型

ELM 是一种新型的快速学习算法, 对于单层神经网络, ELM 可随机初始化输入权重和偏执并得到相应的输出权重。其模型在保证学习精度的前提下, 比传统的神经网络算法速度更快^[14]。

2 结果与讨论

2.1 光谱预处理和样本划分

样品原始光谱共 254 个波段, 波长范围为 900~1 700 nm。由于噪音等干扰, 在 254 个波段中, 开端波段波长存在异常, 所以去除前 22 个波段。采用第 23~254 波段的光谱作为本研究所用光谱, 共 232 个波段。

黑枸杞和白刺果共 360 份样品光谱如图 2(a) 所示。从图中可看出, 1 100~1 700 nm 范围的光谱信息非常丰富。波长

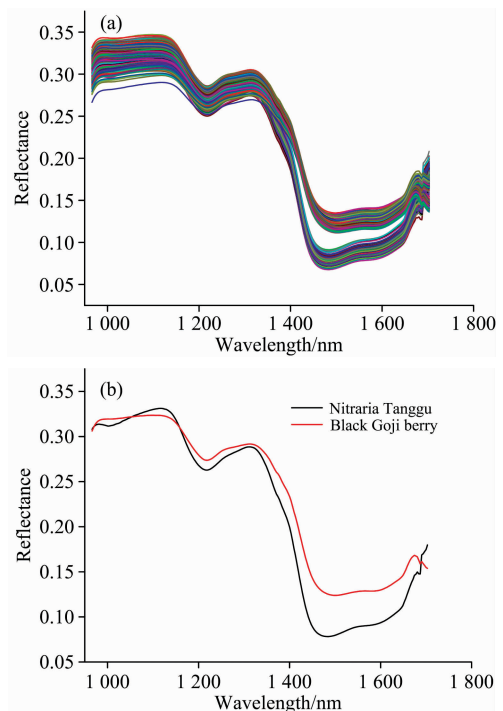


图 2 (a) 黑枸杞和唐古特白刺果的 232 个波段反射光谱; (b) 黑枸杞和唐古特白刺果各自的光谱平均值曲线

Fig. 2 (a) 232 reflectance spectral bands of black Goji berries and Nitraria Tanggu; (b) Average spectra of black Goji berries and Nitraria Tanggu

1 235 和 1 450~1 700 nm 范围有明显吸收峰。雷建刚对不同产地枸杞的研究结论：枸杞在 1 235 和 1 535 nm 处均有明显吸收峰相一致^[15]。根据图 2(b)，对比黑枸杞和白刺果平均光谱可知，两者光谱吸收峰基本一致，光谱曲线走势一致；但是在波长 1 450~1 700 nm 范围内唐古特白刺果吸收峰比黑枸杞明显高。

按照校正集和预测集样品数为 2 : 1 的比例，用 K-S 法划分 360 份样品，得到校正集 240 个(黑枸杞和白刺果各 120 份)。预测集 120 个(黑枸杞和白刺果各 60 份)。

2.2 光谱数据降维

设定 SPA 选择最多波长数为 30，均方根误差随特征波长数变化曲线如图 3 所示，随着 X 轴特征波长的增加，Y 轴均方根误差(RMSE)起初急剧降低，随后缓慢降低。用均方根误差确定最佳特征波常数，选取最佳特征波长数为 20，SPA 选取的特征波长分布如图 4 所示。

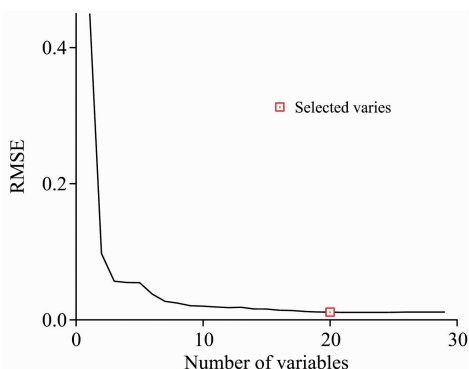


图 3 均方根误差随 SPA 中特征波长数变化曲线

Fig. 3 RMSE changed with the number of characteristic wavelength in SPA

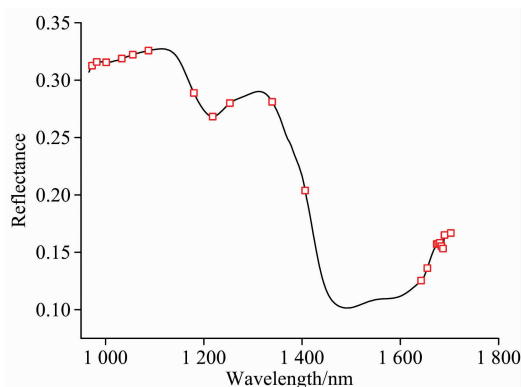


图 4 SPA 选取的是 20 个特征波长分布曲线

Fig. 4 20 characteristic wavelength selected by SPA

2.3 建模结果

分别将全光谱 232 个波段、经 SPA 提取的 20 个特征波长作为输入变量，建立 SVM 和黑枸杞和白刺果识别模型。图 5 和图 6 是两种模型对黑枸杞的识别结果；每个图纵坐标中，1.0 代表唐古特白刺果，2.0 代表黑枸杞。

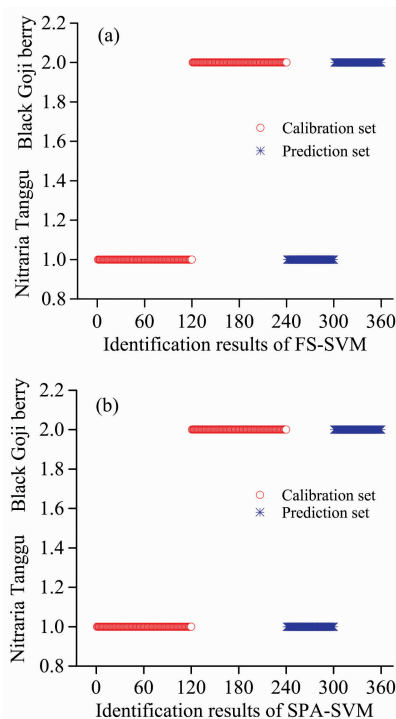


图 5 SVM 黑枸杞和唐古特白刺果识别结果
Fig. 5 Identification results of black Gojiberries and nitraria Tanguu by SVM

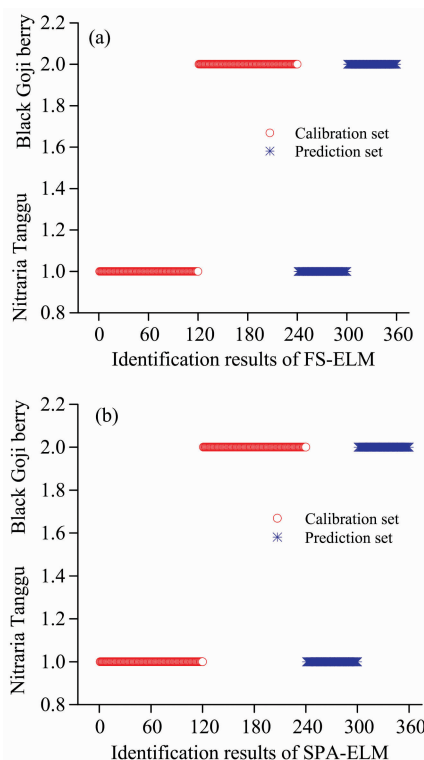


图 6 ELM 黑枸杞和唐古特白刺果识别结果
Fig. 6 Identification results of black Gojiberries and nitraria Tanguu by ELM

2.3.1 SVM 模型

采用径向基函数(radial base function, RBF)作为核函数,通过留一交叉验证方法(cross validation, CV)寻找最佳惩罚因子(c)、核函数参数(g),基于 FS 和 SPA 不同模型确定的 c 和 g 见表 1。

表 1 SVM 模型参数
Table 1 Parameters of SVM

特征波长选择	c	g
FS	0.003 906 3	0.003 906 3
SPA	0.062 5	0.189 46

SVM 模型对黑枸杞和唐古特白刺果识别结果如图 5 所示。FS-SVM 和 SPA-SVM 两种模型的校正集预测率均为 100%;同时,FS-SVM 和 SPA-SVM 两种模型的预测集预测

率均为 100%。

2.3.2 ELM 模型

在 ELM 模型中,采用“sigmoidal”函数作为激活函数,设置隐含层神经元个数为 1~100,步长为 1,确定 FS 和 SPA 的隐含层神经元个数为 6 和 5。ELM 模型对黑枸杞和唐古特白刺果识别结果如图 6 所示。由图可知,基于 FS 和 SPA 建立的 ELM 模型校正集和预测集识别率均为 100%。

3 结 论

(1)从简化模型方面,SPA 提取的特征波长数 20 个,仅为 FS 的 8.62%,大大降低了模型运算量。

(2)FS 和 SPA 的波长分别作为输入,SVM 和 LEM 识别模型识别率均达到了 100%。

References

- [1] ZHAO Xu, WANG Xin-ru, DUAN Chang-qing, et al(赵旭,王新茹,段长青,等). Food Science(食品科学), 2019, 40(8): 202.
- [2] Yahui L, Xiaobo Z, Tingting S, et al. Food Analytical Methods, 2016, 10(4): 1.
- [3] Tian Z, Aierken A, Pang H, et al. Journal of Liquid Chromatography & Related Technologies, 2016: 10826076. 2016. 1179201.
- [4] SUN Xiao-hong, WANG Tong, LÜ Kang-wen, et al(孙晓红,王潼,吕康文,等). Journal of Nutrition(营养学报), 2016, 38(5): 509.
- [5] Liu Q, Sun K, Peng J, et al. Food Analytical Methods, 2018, 11(5): 1518.
- [6] Dong J, Guo W, Zhao F, et al. Food Analytical Methods, 2017, 10(2): 477.
- [7] Yang D, He D, Lu A, et al. Applied Spectroscopy, 2017, 71(10): 2286.
- [8] ZHAO Fan, DONG Jin-lei, GUO Wen-chuan(赵凡,董金磊,郭文川). Modern Food Science and Technology(现代食品科技), 2016, 32(4): 223.
- [9] Dong J L, Guo W C, Wang Z W, et al. Food Analytical Methods, 2016, 9(5): 1087.
- [10] ZHU Xiao-lin, LI Guang-hui, ZHANG Meng(朱晓琳,李光辉,张萌). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(11): 3547.
- [11] D'Archivio A A, Foschi M, Aloia R, et al. Food Chemistry, 2019, 275: 333.
- [12] GUO Wen-chuan, ZHU De-kuan, ZHANG Qian, et al(郭文川,朱德宽,张乾,等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2020, 51(9): 350.
- [13] DONG Jin-lei, GUO Wen-chuan(董金磊,郭文川). Food Science(食品科学), 2015, 36(16): 101.
- [14] FENG Yu, CUI Ning-bo, et al(冯禹,崔宁博,等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(S1): 153.
- [15] LEI Jian-gang, LIU Dun-hua(雷建刚,刘敦华). Food Science(食品科学), 2013, 34(20): 148.

Hyperspectra Used to Recognize Black Goji Berry and Nitraria Tanggu

ZHAO Fan, YAN Zhao-ru, SONG Hai-yan

College of Engineering, Shanxi Agricultural University, Taigu 030801, China

Abstract Black Goji berry contains various nutrients such as cyanidin, polysaccharides, trace elements and so on, and has extremely high economic and medical value, the similar Nitraria Tanggu impersonates in the market. The market price of Nitraria Tanggu is low. Hyperspectral image technology combines image and spectrum in one, commonly used in food detection and recognition. This study combined with hyperspectral image technology to non-destructively identify Black Goji Berry and Nitraria Tanggu. Hyperspectral reflection spectra of Black Goji Berry (180) and Nitraria Tanggu (180) in the range of 900~1700 nm were collected respectively, a total of 254 bands. Removing the first 22 abnormal bands and using the last 232 bands as model inputs. Kennard-Stone method is used to divide samples, correction set : prediction set = 2 : 1. The successive projections algorithm (SPA) method is used for spectral dimensionality reduction, setting the characteristic wavelength range to 0~30, which extracts 20 characteristic wavelengths. The full spectrum and 20 characteristic wavelengths extracted by SPA are used as model inputs to establish support vector machine (SVM) and extreme learning machine (ELM) models to identify Black Goji Berry and Nitraria Tanggu. The results show that the recognition rates of the SVM model based on FS and SPA are both 100%, the recognition rates of the ELM model based on FS and SPA are both 100%, the SPA method can reduce model input without reducing the accuracy of model recognition. The input is only 8.62% of FS, which greatly reduces the number of model calculations. This study provides a theoretical basis for identifying Black Goji Berry and Nitraria Tanggu.

Keywords Hyperspectral; Model; Identify; Black Goji Berry; Nitraria Tanggu

(Received Jul. 16, 2020; accepted Nov. 29, 2020)

本刊声明

近期以来,一些不法分子假冒《光谱学与光谱分析》期刊社名义,以虚假网站等形式欺骗广大作者、读者。这些虚假网站公然假冒《光谱学与光谱分析》期刊名义进行大肆的征稿并骗取作者的审稿费和版面费。经部分作者及读者举报,现有关部门已就此介入调查。本刊将通过法律途径向假冒者追究相应的责任,维护本刊权利。

本刊官方网站已正式开通,网址为

<http://www.gpxygfx.com/>

在此郑重声明,本网址为《光谱学与光谱分析》期刊唯一开通运行的官方网站。本刊从未授权任何单位或个人以任何形式(包括网上网下)代理本刊征稿、审稿等业务。

希望广大读者和作者切实维护好自身的合法权益,防止受骗上当。

《光谱学与光谱分析》期刊社

2019年3月15日