



食品与发酵工业

Food and Fermentation Industries

ISSN 0253-990X, CN 11-1802/TS

《食品与发酵工业》网络首发论文

题目：基于高光谱成像技术的酿酒高粱品种分类
作者：孙婷，田建平，胡新军，罗惠波，黄丹，黄浩平
DOI：10.13995/j.cnki.11-1802/ts.025054
收稿日期：2020-07-14
网络首发日期：2020-09-22
引用格式：孙婷，田建平，胡新军，罗惠波，黄丹，黄浩平. 基于高光谱成像技术的酿酒高粱品种分类. 食品与发酵工业.
<https://doi.org/10.13995/j.cnki.11-1802/ts.025054>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于高光谱成像技术的酿酒高粱品种分类

孙婷¹, 田建平¹, 胡新军¹✉, 罗惠波^{2,3}, 黄丹^{2,3}, 黄浩平¹

1 (四川轻化工大学 机械工程学院, 四川 宜宾, 644000)

2 (四川轻化工大学 生物工程学院, 四川 宜宾, 644000)

3 (酿酒生物技术及应用四川省重点实验室, 四川 宜宾, 644000)

✉通信作者 胡新军, 讲师, E-mail: xjhu@suse.edu.cn

摘要 针对不同品种酿酒高粱分类难、分类精度有待提高等问题, 提出了一种结合光谱和图像信息的高光谱成像技术对酿酒高粱进行分类的方法。采集 11 类共 550 个高粱样本的高光谱数据, 运用连续投影算法从多元散射校正预处理后光谱中筛选出 48 个特征波长, 再提取图像的灰度共生矩阵作为图像特征, 利用纹理特征、全光谱、特征光谱和它们分别结合图像特征建立支持向量机、偏最小二乘判别和极限学习机分类模型, 最后再采集 220 个未参与建模样本对所建模型进行外部验证。结果显示基于特征光谱结合纹理特征建立的支持向量机模型效果最佳, 训练集和测试集的认识率分别为 96%、95.3%, 验证集的认识率达到 91.8%, 高于单一光谱数据建模效果, 说明了光谱和图像信息结合可以提高酿酒高粱的分类认识率。该方法为高粱品种的高精度分类和不同酿酒原料的快速无损检测提供了可行的方法。

关键词 高粱; 高光谱; 特征波长; 灰度共生矩阵; 机器学习

DOI: 10.13995/j.cnki.11-1802/ts.025054

Classification of liquor sorghum varieties based on hyperspectral imaging technology

SUN Ting¹, TIAN Jianping¹, HU Xinjun¹✉, LUO Huibo^{2,3}, HUANG Dan^{2,3}, HUANG Haoping¹

1 (College of Mechanical Engineering, Sichuan University of Science & Engineering, Yibin 644000, China)

2 (College of Biotechnology Engineering, Sichuan University of Science & Engineering, Yibin 644000 China)

3 (Sichuan Key Laboratory of Brewing Biotechnology and Application, Yibin 644000 China)

Abstract In order to solve problem in classifying liquor sorghum, and to improve the classification accuracy a method for classification of liquor sorghum using hyperspectral imaging technology combined with spectral and image information was proposed. The hyperspectral data of 550 sorghum samples from 11 classes were collected, 48 characteristic wavelengths were selected using the successive projection algorithm from the pre-processed spectrum of multiple scattering correction, and then the gray level co-occurrence matrix were extracted as image feature, the support vector machine (SVM), partial least squares discriminant analysis and extreme learning machine classification models were established using the texture feature, full spectrum, feature spectrum and their combined image feature. Finally, 220 non-participating modeling samples were collected for external verification of the built models. The results showed that SVM model based on the feature spectrum combined with texture feature had the best effect, and the recognition rate of the training set and testing set was 96% and 95.3%, respectively. The recognition rate of the verification set was 91.8%, which was higher than the modeling effect of single spectral data, the effects indicated that the combination of spectral and image information could improve classification recognition rate of liquor sorghum. This method provided a feasible method for high-precision classification of sorghum varieties and rapid non-destructive testing of different brewing materials.

收稿日期: 2020-07-14

基金项目: 四川省科技厅重大科技专项 (2018GZ0112); 四川轻化工大学研究生创新基金 (y2019003); 自贡市重点科技计划项目 (2018CXJD06)

作者简介: 孙婷, 硕士研究生

网络首发时间: 2020-09-22 16:40:04 网络首发地址: <https://kns.cnki.net/kcms/detail/11.1802.TS.20200922.1521.003.html>

Key words sorghum; hyperspectral; characteristic wavelength; grey level co-occurrence matrix; machine learning

高粱是我国重要的粮食作物之一，主要种植在以华北、东北为主的北方地区和以西南为主的南方地区，因其籽粒中含有丰富的营养物质在酿酒行业有着“好酒离不开红粮”的精辟论断，每年的需求量高达 2 000 万吨^[1]。目前酿酒高粱品种主要以高淀粉含量的泸州红、青壳洋、睿糯 7 号等糯高粱为主，由于酿酒高粱种类繁多，产地各不相同，籽粒中的淀粉、蛋白质、脂肪和单宁等含量有较大差异，导致酿造出的白酒香型、风格、品质和产酒率上也有很大的差别^[2]。由此可见，在高粱原材料分批入库前准确高效地对高粱品种进行鉴别，对酿造过程中泡粮时间、用水量、蒸粮时间等生产工艺控制和产出优质白酒具有十分重要的指导意义^[3]。传统的鉴别方式主要有人工经验判别和生物抽样检测^[4]，前者易受主观影响，效率低，难以形成统一的标准，后者操作繁琐、费时费力，都无法满足现代白酒企业对酿酒原料高粱的鉴别需求^[5]，因此急需寻找一种快速、准确且简便的高粱品种分类检测方法。

高光谱技术作为一种高效无损检测技术，可以同时获取对象的光谱信息和空间信息^[6]，通常光谱特征在农副产品的分类、质量检测和指标评价等领域运用更多^[7-9]，空间图像信息的应用较少，但是也有研究人员试图将高光谱的光谱信息和图像信息结合起来，以提高模型的准确性和可靠性^[10]。例如王彩霞等人利用特征波长、颜色特征以及纹理特征分别建立了偏最小二乘判别模型（Partial Least Squares Discriminant Analysis, PLS-DA）鉴别 5 个牛肉品种，结果显示特征波长结合纹理特征的模型识别效果最佳，预测集识别率为 93.55%，均高于纯光谱数据精度^[11]；Huang 等人结合玉米种子的光谱特征、形态特征和纹理特征分别建立了支持向量机模型（Support Vector Machine, SVM）来判别 17 个玉米品种，结果显示结合光谱信息和图像信息在测试集上的准确率达到 92%，高于只使用单一信息建模精度^[12]；邢素霞等人同样结合光谱信息和图像信息建立 RBF 神经网络对鸡肉品质进行分类，测试集分类正确率达到 100%^[13]。

本研究的目标是结合光谱信息和图像信息对 11 个高粱品种进行分类，同时通过比较及外部验证优化高光谱技术结合机器学习方法鉴别不同高粱品种，以提高该技术应用过程中的准确性和效率。主要对高粱高光谱数据进行预处理，选取特征波长，从图像中提取灰度共生矩阵（Gray Level Co-occurrence Matrix, GLCM），分别基于全光谱、特征波长、纹理特征（GLCM）和它们的融合数据建立高粱的分类模型，最后使用未参与建模的验证集数据对所建模型进行外部验证。

1 实验部分

1.1 样品准备

实验选用两个批次共 11 种不同品种的高粱，分别来自四川泸州、宜宾地区和山东济南地区，详细的产地、收获年份和批次如表 1 所示，其中矮抗 8 号、红塔 2 号和铁杆 1 号为白高粱，其余为红高粱。样本去除杂质，每次选取约 120 g 高粱装袋为 1 个样本，每种高粱取 50 个样本，并均匀平铺于 80 cm 直径的培养皿中，直至与培养皿边缘齐平，共计 550 个样本。使用 Kennard-Stone (KS) 算法将样本划分为 400 个训练集和 150 个测试集，此外，每种高粱再采集 20 个共计 220 个样本用于模型外部验证。

表1 高粱品种参数
(Tabel 1 Parameters of sorghum varieties)

品种	产地	收获年份	批次	品种	产地	收获年份	批次
澳洲高粱	四川泸州	2018	1	超级糯 1 号	山东济南	2019	2
红茅 6 号	四川泸州	2018	1	红冠 1 号	山东济南	2019	2
粳高粱	四川宜宾	2019	1	红塔 2 号	山东济南	2019	2
泸州红	四川泸州	2018	1	睿糯 7 号	山东济南	2019	2

宜糯红	四川宜宾	2019	1	铁杆 1 号	山东济南	2019	2
矮抗 8 号	山东济南	2019	2				

1.2 仪器与设备

采用芬兰 Specim 公司生产的高光谱采集系统，主要部件有：FX10E 型号的高光谱相机(Specim, 芬兰)，其外形尺寸为 150 mm×75 mm×85 mm，重量为 1.4 kg，探测器类型为 InGaAs，视场角 FOV 为 38°，相机输出为 12 bit，空间分辨率为 1 024×628 像素，2 组功率为 150 W 的卤素灯光源(OSRAM, 德国)，高精度电控载物台(Specim, 芬兰)、Lumo-scanner 专用软件(DELL, 美国)以及辅助支架。

为了保证样本光谱信息和图像信息的完整性，具体的参数设定如表 2 所示。

表2 高光谱采集参数设置
(Tabel 2 Hyperspectral acquisition parameter settings)

参数名称	设定值
光谱范围/nm	400~1 000
光谱分辨率/nm	1.31
光谱通道数	448
曝光频率/Hz	50
曝光时间/ms	8
平台移动速度 mm/s	10.84
白板起始位置/mm	起: 24.47 止: 34.87
物体起始位置/mm	起: 119.80 止: 256.23
R 通道	波段 191(650.96 nm)
G 通道	波段 117(551.57 nm)
B 通道	波段 55(469.28 nm)

注：表中 RGB 通道所选的波段是用于高光谱相机生成伪彩色图像的波段。

1.3 高光谱图像采集与标定

采集前调整好系统参数，尽量保持外部环境一致，按照样本编号将装满高粱的培养皿放置在载物台上依次扫描，得到 770 组高粱样本的原始高光谱数据。为了减少传感器暗电流噪声和光强变化对图像信号产生的影响，需要对获得的高光谱图像进行黑白校正^[14]，校正公式见式（1）。

$$R = \frac{I_0 - I_d}{I_w - I_d} \tag{1}$$

式中： R ，校正后的光谱图像； I_0 ，原始高光谱图像； I_d ，盖上镜头采集的暗参考图像； I_w ，采集的标准白板图像，这样高光谱图像的每一个像素点均对应一条光滑的光谱曲线。

2 理论与方法

2.1 数据预处理

感兴趣区域（Region of Interest, ROI）的选取对后续的建模分析至关重要，每个高粱样本统一选取一块 100×100 像素无杂质，亮度均匀的区域，按照校正公式（1）计算出每个像素的反射率，再对 ROI 区域内所有像素点反射率求平均值，就得到了样本的光谱曲线。为了提取高光谱图像中光谱有效信息，需要对光谱数据进行预处理来减少光谱信号中的光散射、高频噪声等干扰信息^[6]。多元散射校正（Multiplicative Scatter Correction, MSC）可以有效地消除散射带来的光谱差异，从而增强光谱与数据之间的相关性，该方法通过理想光谱修正光谱数据的基线平移和偏移现象^[15]，具体实现方法见式（2）~（4）。

①计算平均光谱作为标准光谱；

$$P_{i,j} = \frac{\sum_{l=1}^n S_{l,j}}{n} \tag{2}$$

②将每个样本的光谱与平均光谱进行线性回归运算，得到每个样本相对于标准光谱的尺度平移量和偏移量；

$$S_i = m_i P + b_i \tag{3}$$

③原始光谱减去求得的基线平移量后除以偏移量,得到校正后的光谱;

$$S_{imsc} = \frac{(S_i - b_i)}{m_i} \quad (4)$$

式中: $P_{i,j}$, 标准光谱矩阵; $S_{i,j}$, 原始光谱矩阵; n , 样本数量, j , 波段数量; S_i , 每个样本的光谱向量; m_i , 相对偏移系数; b_i , 平移量; S_{imsc} , i 样本的 MSC 校正光谱。

2.2 特征波长筛选

本文采集的 400~1 000 nm 范围内的高光谱数据有 448 个波长, 存在大量冗余和多重共线性问题, 需要从这些波段中选取最具代表性的特征波长。连续投影算法 (Successive Projection Algorithm, SPA) 是一种使矢量空间共线性最小化的前向变量选择算法^[16], 其运算过程为: 首先随机选择一个波长, 分别计算该波长与剩余向量的投影, 将最大投影值的波长选入特征波长; 再对没选入的波长重复以上过程, 直至获得特征波长数量则截止^[17]。选用此方法筛选特征波长。

2.3 图像信息提取

不同品种的高粱颗粒在外形、大小、颜色上面都有不同程度的差异, 这些差异会直接反映在图像信息中, 而目前的相关研究大部分只使用了光谱信息, 却忽略了图像信息。灰度共生矩阵 (GLCM) 用于计算相邻像素之间不同灰度值的组合数目, 以此反映在指定方向上的梯度信息^[18], 是一种描述图像信息的常用方法。先将高粱灰度图转换为固定级数的灰度图, 分别从 0°、45°、90°、135° 四个方向计算灰度图的能量 (Energy)、熵 (Entropy)、惯性矩 (Moment of inertia) 和相关性 (Correlation), 计算公式如下, 再以这 4 个参数的平均值和方差作为图像纹理特征的典型性信息。

$$Energy = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (P(i,j))^2 \quad (5)$$

$$Entropy = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j) \log(i,j) \quad (6)$$

$$MOI = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j)(i-j)^2 \quad (7)$$

$$COR = \frac{[\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} ijP(i,j)] - \mu_x \mu_y}{\delta_x \delta_y} \quad (8)$$

式中: $Energy$, 能量; $Entropy$, 熵; MOI , 惯性矩; COR , 相关性; L , 灰度级数; $P(i,j)$, 转换 L 级后的图像; i , 一个像素的灰度值; j , 距离 i 固定步长像素的灰度值; μ_x 和 μ_y , 平均值; δ_x 和 δ_y , 方差; 其中:

$$\mu_x = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} iP(i,j) \quad (9)$$

$$\mu_y = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} jP(i,j) \quad (10)$$

$$\delta_x^2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j)(i - \mu_x)^2 \quad (11)$$

$$\delta_y^2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j)(j - \mu_y)^2 \quad (12)$$

2.4 建模方法

支持向量机(SVM)是一种以结构风险最小化原则为基础的模式识别算法, 主要思路是把训练集数据从输入空间非线性地映射到一个多维度的特征空间, 然后在多维度的特征空间中通过最小化某种损失函数获得一个线性的拟合函数^[19], 寻找新的最优分类面作为决策面, 实现不同类数据的分离。偏最小二乘判别 (PLS-DA) 是多元分类模型的一种, 它是基于偏最小二乘回归的高维线性判别模型分类技术。该方法首先将样本类别进行数值标定并作为反应变量, 再运用 PLSR 建立反应变量与解释变量之间的关系模型, 比较模型的反应变量预测值大小, 从而确定各样本的类别^[20], 适合于变量数众多且存在多重共线性的情况, 是常用的有监督统计模式识别方法^[21]。极限学习机 (Extreme Learning Machine, ELM) 作为一种单隐层前馈神经网络, 主要由输入层、隐含层和输出层组成, 随机产生网络的输入权值和隐层节点偏置值, 将输入数据与权值矩阵作乘积运算再加上偏置矩阵的结果输入到激活函数, 输出层把激活函数的输出与连接权值作乘积运算, 从而得到最终的预测结果^[22]。训练过程无需过多的人为干预, 运算速度快, 对于数据量很大的情况能够有效的提高运算效率^[23]。

通过正确率（Accuracy）、运行时间及外部验证结果对预测模型进行综合评价分析^[24]，正确率的计算公式见式（2），

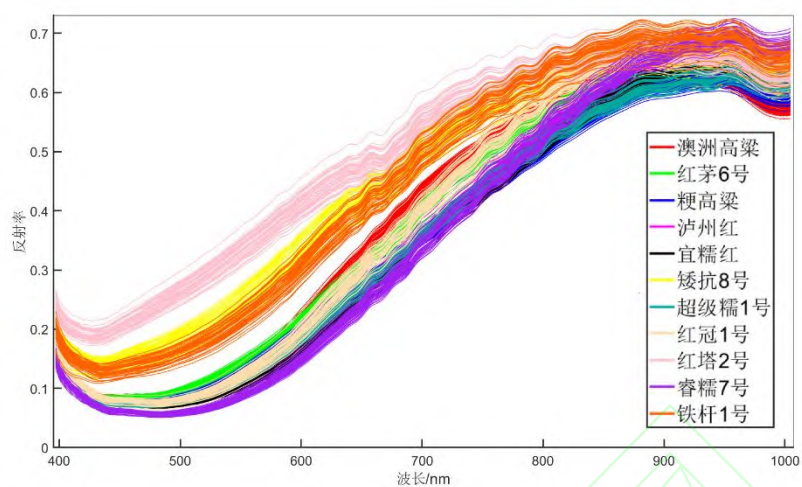
$$Accuracy = \frac{TP}{TP+FP} \quad (13)$$

式中Accuracy，分类的正确率；TP，正确分类的高粱类别；FP，错误分类的高粱类别。

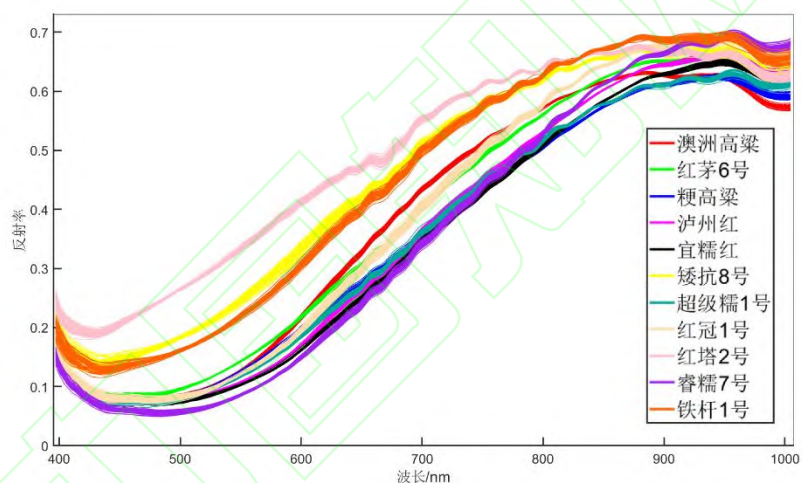
3 结果与分析

3.1 高粱光谱曲线分析

11类高粱共计550个样本的原始光谱曲线和MSC预处理后的光谱曲线如图1所示，每个颜色代表不同的类别。两幅图都可见在400~1 000 nm波长范围内高粱的光谱曲线总体趋势相似，均呈上升趋势，波峰波谷变化较平缓，且出现了部分曲线交叉重叠的情况。预处理后的曲线重叠明显减少，噪声也更少，各样本曲线的辨识度更高。其中白高粱矮抗8号、红塔2号和铁杆1号的光谱曲线均在其余高粱的上方，其在430~890 nm范围内的曲线走势也与其余高粱有明显差异，这些信息可以将红白高粱较容易地区分开。在400~500 nm范围内其余8类高粱的反射率大致相同，这会对分类造成困难；几乎对于所有品种，光谱均在600 nm、890 nm、950 nm和990 nm附近出现峰和谷，这可能与羧基官能团所带分子键的伸缩振动有关。以上的差异性为运用高光谱识别高粱品种提供了有效判别依据。



a- 原始光谱曲线

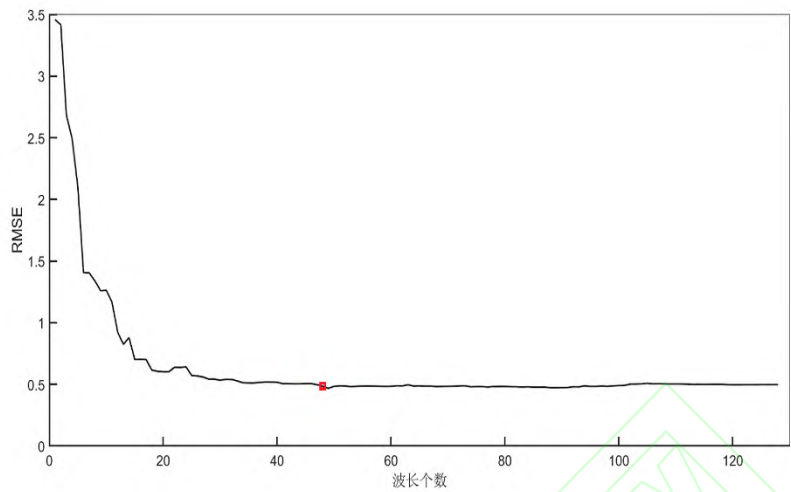


b- 经 MSC 预处理光谱

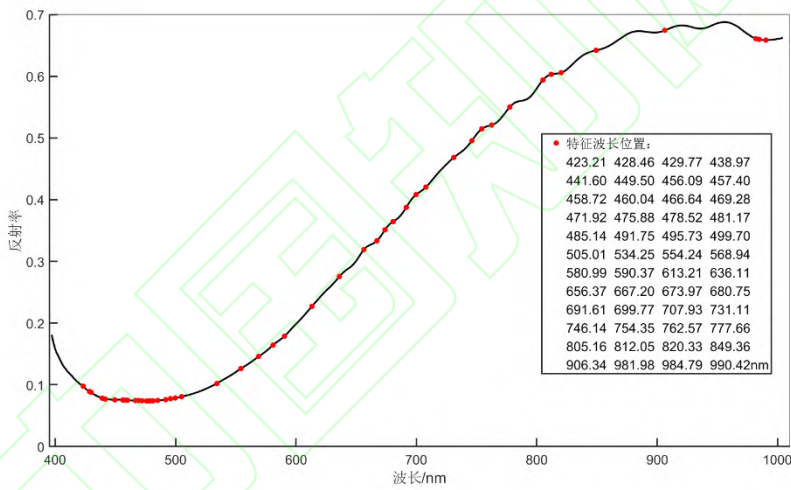
图 1 高粱样本的光谱曲线
(Fig.1 Spectral curve of sorghum samples)

3.2 最优特征波长选取

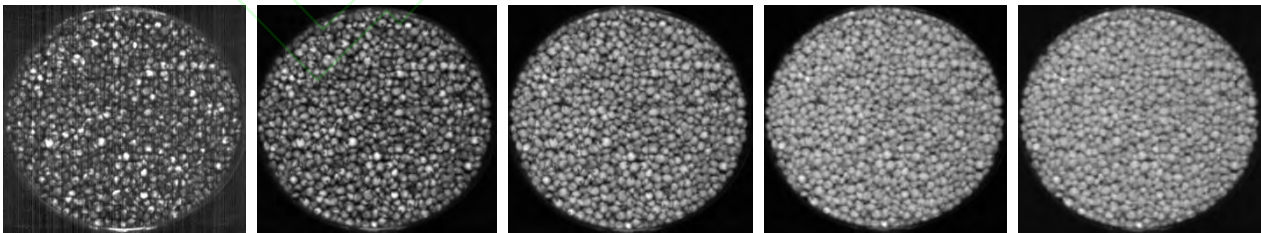
使用 SPA 算法从 448 个波长中筛选特征波长, 图 2 (a) 为均方根误差 (Root Mean Square Error, RMSE) 的变化趋势图, 可见随着波长数量不断增加, 误差逐渐减少, 在数量达到 48 之后误差基本保持稳定, 说明此时已没有冗余波长可筛除, 最终使用 RMSE 值为 0.485 3 对应的 48 个波长为最优特征波长。它们的具体位置如图 2 (b) 红色标记所示, 其中图 2 (c) 是澳洲高粱 5 个典型特征波长的图像, 它们能够代表大部分信息, 具有特征性。将特征波长存储在 550×48 矩阵中, 光谱减少量为 89.3%, 极大地缩短了计算时间。



a- 均方根误差变化趋势图



b- SPA 选择的最优波长位置



d- 460.04 nm e- 613.21 nm f- 691.61 nm g- 805.16 nm i- 990.42 nm

c- 典型特征波长下的高光谱图像

图 2 特征波长的提取过程
(Fig.2 The process of extracting characteristic wavelengths)

3.3 图像特征 (GLCM) 提取

将每个样本的灰度图像转换成 16 级灰度图，依次对 770 个高粱样本进行图像特征提取，距离参数取值 1，方向选取 0°、45°、90°和 135°，部分代表样本的纹理特征如表 3 所示，将其存储在 550×8

矩阵中作为图像特征。

表3 高粱纹理特征参数
(Table 3 Sorghum texture feature parameters)

样本编号	能量		熵		惯性矩		相关性	
	平均值	方差	平均值	方差	平均值	方差	平均值	方差
澳洲	0.480 2	0.001 8	1.475 8	0.037 9	0.139 0	0.036 7	0.333 4	0.002 3
红矛 6 号	0.493 4	0.001 7	1.407 8	0.035 7	0.125 2	0.031 9	0.399 0	0.002 9
粳高粱	0.499 2	0.002 0	1.384 0	0.035 0	0.113 1	0.028 1	0.412 6	0.002 7
泸州红	0.507 5	0.002 3	1.240 7	0.029 6	0.084 2	0.018 3	0.731 9	0.005 4
宜糯红	0.497 7	0.002 3	1.268 4	0.030 0	0.085 3	0.018 4	0.689 8	0.004 8
矮抗 8 号	0.583 2	0.001 2	1.208 8	0.027 3	0.119 0	0.025 3	0.285 2	0.001 2
超级糯 1 号	0.555 8	0.002 0	1.120 3	0.025 7	0.084 2	0.074 4	0.896 5	0.006 8
红冠 1 号	0.585 4	0.001 8	1.084 6	0.024 0	0.071 2	0.014 0	0.758 7	0.004 7
红塔 2 号	0.556 9	0.001 3	1.309 3	0.031 0	0.151 2	0.033 8	0.173 3	0.000 6
睿糯 7 号	0.589 4	0.002 7	0.969 9	0.022 6	0.055 3	0.010 5	1.545 7	0.014 7
铁杆 1 号	0.572 0	0.001 3	1.240 1	0.029 2	0.121 2	0.026 9	0.334 3	0.001 8

3.4 高粱分类建模效果

使用 11 类高粱的全光谱数据、48 个特征波长数据和它们融合纹理特征的数据分别建立 SVM、PLS-DA 和 ELM 分类模型。其中 SVM 使用 10 折交叉验证选取最佳的 *c*、*g* 值，核函数采用径向基函数（RBF）；PLS-DA 在建模前使用 10 折交叉验证选取最佳主成分数，主成分数根据建模输入数据不一，图 3 为 550 个样本全光谱建模前两个主成分 PC1（74.04%）、PC2（21.49%）的散点图，从图中可以看出不同品种的高粱分布较分散，比较容易区分；ELM 则随机产生输入层和隐含层之间的连接权值矩阵 *W*，隐含层偏置矩阵 *B*，激活函数选取 Sigmoid 函数。建模的参数设置和各模型的精度如表 4 所示。

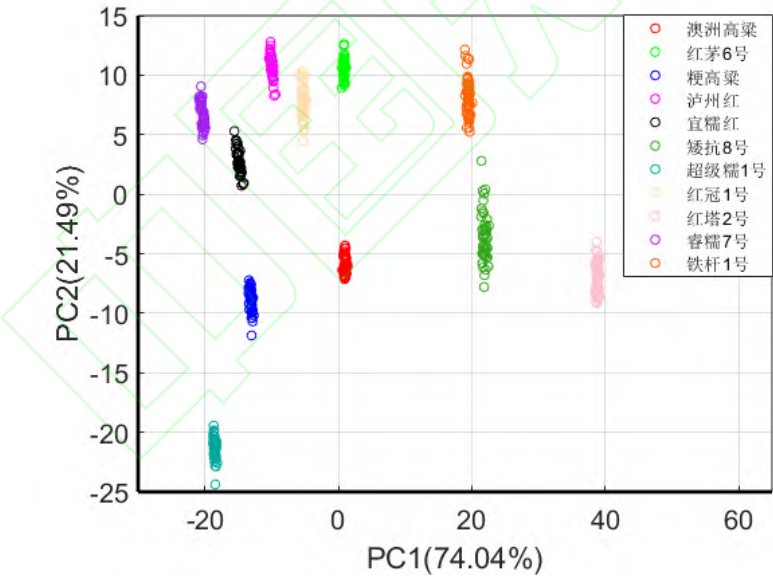


图 3 PLS-DA 全光谱建模前两个主成分图
(Fig.3 The first two principal component maps of PLS-DA full spectrum model)

表4 各模型参数及分类结果
(Table4 Parameters and classification results of each model)

建模数据	SVM				PLS-DA				ELM	
	训练集	测试集	时间/s	c 值	g 值	训练集	测试集	时间/s	主成分数	训练集测试集时间/s
纹理特征	83.3%	86%	2.622	78	16	81.8%	80.7%	0.087	11	87% 80.7% 0.066
全光谱	98.5%	93.3%	35.888	2.838	0.125	93.5%	84.7%	7.352	17	94.6% 82.7% 4.243
特征光谱	97.5%	92.7%	9.351	8.000	0.707	96.8%	94.7%	3.980	9	97.2% 93.3% 4.038
全光谱+纹理	95%	94.7%	41.869	11.314	0.088	87.5%	89.3%	7.319	19	93.3% 94.7% 4.190
特征光谱+纹理	96%	95.3%	9.561	11.310	0.350	94.5%	91.3%	3.773	13	94.5% 94.7% 3.739

对比不同模型在 5 类数据集上的表现, 整体来看 SVM 模型在训练集和测试集的正确识别率大于 PLS-DA 和 ELM, 但是运行时间也更长。使用纹理特征数据建模的精度均在 80%~87%之间, 相对于其它数据集精度较低, 说明单一的纹理特征信息不能全面地反映不同高粱品种的差异。比较使用全光谱和特征光谱建模时, 特征光谱的正确识别率根据数据集不同有时高于全光谱, 有时低于全光谱, 总体来看两者精度相差不大, 说明筛选出的特征光谱中包含了大量有效的光谱信息, 使用特征光谱对高粱进行识别是可行的, 且减少了计算时间。此外可知, 在结合了图像信息纹理特征以后, 全光谱和特征光谱的正确识别率均有所提升, 可以说明结合纹理特征后的数据更全面地表达了样本的分类信息, 可以提高高粱品种的识别率。在上述的所有模型中, 特征光谱结合纹理信息建立的 SVM 模型精度最高, 训练集和测试集的正确识别率分别为 96%、95.3%, 运行时间为 9.561 s。综上所述, 结合光谱信息和图像信息的高粱品种识别方法是可行的。

3.5 模型外部验证

为了进一步验证上述模型的准确性和稳定性, 采用外部验证方式验证模型效果, 即将未参与建模的 220 个验证样本组成的验证集代入模型进行预测, 代入模型的数据集以及模型参数与上述建立的模型保持一致, 表 5 为外部验证的正确分类率。由表 5 可知, 综合比较, SVM 模型的验证效果最好, ELM 次之, PLS-DA 相对来说最差, 其中特征光谱结合纹理特征的 SVM 模型验证集识别准确率达到 91.8%, 能够满足精度要求, 但是由于 SVM 分类器训练过程中需要寻找支持向量, 所以耗时相对较长, 未来可以与粗集理论等结合提高分类效率。

表5 各模型验证集分类结果
(Table 5 Classification results of validation set of each model)

分类精度	全光谱	特征光谱	全光谱+纹理特征	特征光谱+纹理特征
SVM	90.5%	88.2%	90.5%	91.8%
PLS-DA	86.4%	85.5%	85.6%	89.1%
ELM	89.6%	88.6%	91.8%	91.4%

4 结论

本文基于高光谱技术光谱和图像信息结合的方法对 11 个品种的高粱进行识别研究。采集高粱的高光谱图像, 使用 SPA 算法从 MSC 预处理后光谱中筛选出 48 个特征波长, 再提取图像的纹理特征, 分别基于纹理特征、全光谱、特征光谱和它们分别结合图像信息建立了 SVM、PLS-DA 和 ELM 分类模型, 最后使用未参与建模的数据进行外部验证。结果显示使用特征光谱结合纹理特征建立的 SVM 分类模型效果最佳, 测试集的正确识别率为 95.3%, 验证集的精度达到 91.8%, 可见光谱和图像结合的方法可以有效实现酿酒高粱的快速识别, 且提高了模型的识别精度, 这为不同酿酒原料的检测和酿酒自动化的实现提供了理论基础。今后可以从扩大波段范围, 或者增加图像的颜色和形状信息等入手挖掘更多加以研究。

参 考 文 献

- [1] 蒋兰. 酿酒高粱淀粉含量测定及性质研究[D]. 重庆: 重庆大学, 2013.
JIANG L. Determination of starch content and properties of brewing sorghum[D]. Chongqing: Chongqing University, 2013.
- [2] 毛祥, 温雪瓶, 黄丹, 等. 5 种常用酿酒高粱的主要成分及淀粉特性差异分析[J]. 中国酿造, 2020, 39 (3): 57-62.
MAO X, WEN X P, HUANG D, et al. Analysis of the difference in main components and starch characteristics of five commonly used brewing sorghum[J]. China Brewing, 2020, 39(3): 57-62.
- [3] 买书魁, 杨洋, 赵小波, 等. 基于 NIR 的白酒酿酒高粱中关键指标的定量分析[J]. 食品科技, 2019, 44 (2): 301-307.
MAI S K, YANG Y, ZHAO X B, et al. Quantitative analysis of key indexes in liquor sorghum based on NIR[J]. Food Science and Technology, 2019, 44(2): 301-307.
- [4] 王勇生, 李洁, 王博, 等. 基于近红外光谱扫描技术对高粱中粗脂肪、粗纤维、粗灰分含量的测定方法研究[J]. 中国粮油学报, 2020, 35 (3): 181-185.
WANG Y S, LI J, WANG B, et al. Study on the determination method of crude fat, crude fiber and crude ash content in sorghum based on near infrared spectroscopy scanning technology[J]. Journal of the Chinese Cereals and Oils Association, 2020, 35(3): 181-185.
- [5] 邵琦, 陈云浩, 杨淑婷, 等. 基于随机森林算法的玉米品种高光谱图像鉴别[J]. 地理与地理信息科学, 2019, 35 (5): 34-39.
SHAO Q, CHEN Y H, YANG S T, et al. Hyperspectral image identification of corn varieties based on random forest algorithm[J]. Geography and Geo-Information Science, 2019, 35(5): 34-39.
- [6] 赵鹏, 唐艳慧, 李振宇. 支持向量机复合核函数的高光谱显微成像木材树种分类[J]. 光谱学与光谱分析, 2019, 39 (12): 3776-

3782.
ZHAO P, TANG Y H, LI Z Y. Classification of wood tree species based on hyperspectral microscopic imaging of support vector machine compound kernel function[J]. Spectroscopy and Spectral Analysis, 2019, 39(12): 3776-3782.
- [7] ZHANG J, DAI L M, CHENG F. Classification of frozen corn seeds using hyperspectral VIS/NIR reflectance imaging[J]. Molecules, 2019, 24(1).
- [8] 刘翠玲, 林珑, 于重重, 等. 基于深度学习的花生高光谱图像分类方法研究[J]. 计算机仿真, 2020, 37(3): 189-192+283.
LIU C L, LIN N, YU C C, et al. Research on peanut hyperspectral image classification method based on deep learning[J]. Computer Simulation, 2020, 37(3): 189-192+283.
- [9] 韩仲志, 刘杰. 高光谱亚像元分解预测花生中的黄曲霉毒素 B₁[J]. 中国食品学报, 2020, 20(3): 244-250.
HAN Z Z, LIU J. Prediction of aflatoxin B₁ in peanut by hyperspectral subpixel decomposition[J]. Chinese Journal of Food Science, 2020, 20(3): 244-250.
- [10] 邹志勇, 吴向伟, 陈永明, 等. 低温冷冻和机械损伤条件下马铃薯高光谱图像特征响应特性研究[J]. 光谱学与光谱分析, 2019, 39(11): 3571-3578.
ZOU Z Y, WU X W, CHEN Y M, et al. Characteristic response characteristics of potato hyperspectral image under low temperature freezing and mechanical damage[J]. Spectroscopy and Spectral Analysis, 2019, 39(11): 3571-3578.
- [11] 王彩霞, 王松磊, 贺晓光, 等. 高光谱技术融合图像信息的牛肉品种识别方法研究[J]. 光谱学与光谱分析, 2020, 40(3): 911-916.
WANG C X, WANG S L, HE X G, et al. Research on beef variety recognition method based on hyperspectral technology and image information[J]. Spectroscopy and Spectral Analysis, 2020, 40(3): 911-916.
- [12] HUANG M, HE C J, ZHU Q B, et al. Maize seed variety classification using the integration of spectral and image features combined with feature transformation based on hyperspectral imaging [J]. Applied Sciences, 2016, 6(6).
- [13] 邢素霞, 王九清, 陈思, 等. 基于 K-means-RBF 的鸡肉品质分类方法研究[J]. 食品科学技术学报, 2018, 36(4): 93-99.
XING S X, WANG J Q, CHEN S, et al. Research on chicken quality classification method based on K-means-RBF[J]. Journal of Food Science and Technology, 2018, 36(4): 93-99.
- [14] 刘爽, 谭鑫, 刘成玉, 等. 高光谱数据处理算法的小麦赤霉病籽粒识别[J]. 光谱学与光谱分析, 2019, 39(11): 3540-3546.
LIU S, TAN X, LIU C Y, et al. Recognition of wheat head blight based on hyperspectral data processing algorithm[J]. Spectroscopy and Spectral Analysis, 2019, 39(11): 3540-3546.
- [15] HE P H, WU Y, WANG J J, et al. Detection of mites *Tyrophagus putrescentiae* and *Cheyletus eruditus* in flour using hyperspectral imaging system coupled with chemometrics[J]. Journal of Food Process Engineering, 2020, 43(6).
- [16] 黄平捷, 李宇涵, 俞巧君, 等. 基于 SPA 和多分类 SVM 的紫外-可见光光谱饮用水有机污染物判别方法研究[J]. 光谱学与光谱分析, 2020, 40(7): 2267-2272.
HUANG P J, LI Y H, YU Q J, et al. Research on the identification method of drinking water organic pollutants based on SPA and multi-class SVM[J]. Spectroscopy and Spectral Analysis, 2020, 40(7): 2267-2272.
- [17] 宋镇, 姬长英, 张波. 基于光谱与图像信息的杏鲍菇多糖含量检测[J]. 华南农业大学学报, 2019, 40(3): 104-110.
SONG Z, JI C Y, ZHANG B. *Pleurotus eryngii* polysaccharide content detection based on spectrum and image information[J]. Journal of South China Agricultural University, 2019, 40(3): 104-110.
- [18] 芦兵, 孙俊, 杨宁, 等. 基于荧光透射谱和高光谱图像纹理的茶叶病害预测研究[J]. 光谱学与光谱分析, 2019, 39(8): 2515-2521.
LU B, SUN J, YANG N, et al. Research on tea disease prediction based on fluorescence transmission spectrum and hyperspectral image texture[J]. Spectroscopy and Spectral Analysis, 2019, 39(8): 2515-2521.
- [19] 赵刘, 齐海军, 金秀, 等. 面向大米分类的高光谱特征波长提取方法[J]. 江苏农业科学, 2018, 46(17): 235-238.
ZHAO L, QI H J, JIN X, et al. Hyperspectral feature wavelength extraction method for rice classification[J]. Jiangsu Agricultural Sciences, 2018, 46(17): 235-238.
- [20] 黄亚伟, 张令, 王若兰. 新陈玉米的拉曼光谱快速判别研究[J]. 现代食品科技, 2014, 30(12): 149-152.
HUANG Y W, ZHANG L, WANG R L. Raman spectroscopy of newly-aged corn rapid discrimination research[J]. Modern Food Science and Technology, 2014, 30(12): 149-152.
- [21] 高彤, 吴静珠, 林珑, 等. 基于 NIR 和 PLS-DA 法的东北大米产地快速溯源方法研究[J]. 中国粮油学报, 2019, 34(7): 114-117; 124.
GAO T, WU J S, LIN L, et al. Research on the rapid traceability method of rice producing area in northeast china based on NIR and PLS-DA method[J]. Journal of the Chinese Cereals and Oils Association, 2019, 34(7): 114-117; 124.
- [22] 吴亚榕, 王欢, 李键红. 适用于数据分类的极限学习机优化算法[J]. 软件导刊, 2019, 18(6): 10-13+17.
WU Y R, WANG H, LI J H. Optimization algorithm of extreme learning machine suitable for data classification[J]. Software Guide, 2019, 18(6): 10-13+17.
- [23] 陆银标. 基于极限学习机的苹果叶片病害识别[D]. 杨凌: 西北农林科技大学, 2019.
LU Y B. Apple leaf disease identification based on extreme learning machine[D]. Yangling: Northwest A&F University, 2019.
- [24] 董小栋, 郭培源, 徐盼, 等. 融合高光谱和图像深度特征的腊肉分类与检索算法研究[J]. 食品工业科技, 2018, 39(23): 255-260; 266.
DONG X D, GUO P Y, XU P, et al. Research on bacon classification and retrieval algorithm combining hyperspectral and image depth features[J]. Food Industry Science and Technology, 2018, 39(23): 255-260; 266.