

基于高光谱小波特征的脱绒棉种活力检测

杜文玲¹, 郭 鹏^{1,2*}, 刘 笑¹

(1. 石河子大学理学院, 石河子 832003; 2. 绿洲城镇与山盆系统生态兵团重点实验室, 石河子 832003)

摘 要: 为了提取不同活力种子光谱特征的精细结构和复杂信息, 该研究探索了连续小波变换提取不同活力脱绒棉种的光谱信息, 并提出了一种基于相关性及特征重要性筛选小波特征 (wavelet features, WFs) 的方法。通过人工老化试验, 获取不同活力等级的脱绒棉种, 并采集其高光谱影像, 对原始光谱进行 Savitzky-Golay 平滑、多元散射校正、一阶微分、二阶微分等预处理。然后对比 gauss4、mexh 和 bior6.8 等小波基函数提取的 WFs。利用主成分分析对光谱特征 (spectral features, SFs) 与 WFs 降维, 并建立支持向量机 (support vector machines, SVM)、随机森林 (random forest, RF)、极限学习机 (extreme learning machines, ELM) 和反向传播神经网络 (back propagation neural network, BPNN) 等的种子活力检测模型, 对比 SFs 与 WFs 的建模精度。为了进一步提取出 WFs 中的精细光谱信息, 基于相关性分析和随机森林特征重要性评价, 提取了与种子活力的相关性在前 1% 的小波特征 (1%|R|-WFs)、在种子活力识别中特征重要性在前 1% 的小波特征 (1%Importance-WFs) 及二者融合的 1%|R|+1%Importance-WFs 3 个 WFs 特征集并带入上述机器学习模型。结果表明: 1) bior6.8 函数提取的不同活力脱绒棉种的 WFs 效果较好, 其他函数在提取 WFs 时, 出现明显的振铃效应。2) 在各品种的所有机器学习模型中, WFs 主成分的建模精度均高于 SFs 主成分的建模精度, 且基于 1%|R|+1%Importance-WFs 的准确率最高。3) 金科 21 与金科 20 种子活力检测的最优模型均为: 1%|R|+1%Importance-WFs+ELM; 新陆早 64 种子活力检测的最优模型为: 1%|R|+1%Importance-WFs+各机器模型与 PCA-WFs+ ELM/BPNN。金科 21 最优模型训练集和测试集的准确率分别为 99.63%、98.28%; 金科 20 与新陆早 64 最优模型训练集和测试集的准确率均为 100%。结果表明, 该研究提出的基于相关性及特征重要性的方法能够有效提取出不同活力脱绒棉种的光谱差异信息, 为种子活力高光谱检测提供一种新的光谱特征分析思路。

关键词: 连续小波变换; 相关性; 特征重要性; 高光谱; 种子活力检测

doi: 10.11975/j.issn.1002-6819.202406004

中图分类号: S127

文献标志码: A

文章编号: 1002-6819(2024)-20-0174-13

杜文玲, 郭鹏, 刘笑. 基于高光谱小波特征的脱绒棉种活力检测[J]. 农业工程学报, 2024, 40(20): 174-186. doi: 10.11975/j.issn.1002-6819.202406004 <http://www.tcsae.org>

DU Wenling, GUO Peng, LIU Xiao. Detection of delinted cotton seed vigor based on hyperspectral wavelet features[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(20): 174-186. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202406004 <http://www.tcsae.org>

0 引 言

新疆是中国棉花主要产区之一, 也是中国最重要的优质棉生产基地, 2023 年新疆棉花产量占全国 90% 以上^[1-2]。在提高产量的同时, 棉花的质量也在不断提升, 近年来, 新疆棉花的长度、断裂比强度等纤维质量均有所上升^[3-4]。筛选优质、高活力的种子对提高棉花产量、选育高品质棉花品种至关重要。依托人的感官识别种子的颜色、纹理和形状等外部特征信息难以实现种子活力变化的区分^[5-7]。传统的种子活力生理生化检测和发芽测试等方法测试周期长、工作量大、成本高、效率低^[8-10], 难以满足现代农业生产对自动化、高通量种子活力检测

的需求^[11-12]。

近年来, 随着高光谱技术的发展, 连续丰富的波谱内蕴含着大量表征种子内部物质变化的信息^[13]。随着存储过程中种子活力的变化, 其内部的蛋白质、蔗糖、脂肪和水分等物质也会随之改变^[6,13-14]。高通量的光谱信息使得依托种子内部物质对光谱的吸收特点来表征其含量变化进而实现种子活力的检测成为可能。快速获取数据、丰富的光谱信息、无损检测等优点, 使得高光谱技术在种子活力检测中得到了广泛应用^[5,13,15-16]。金文玲等^[12]设计水稻种子活力近红外吸收光谱检测系统, 并结合偏最小二乘判别分析 (partial least squares discriminant analysis, PLS-DA) 模型, 实现了水稻种子活力无损检测。孙俊等^[5]将高光谱技术与深度学习相结合检测水稻种子活力等级, 使用堆叠自动编码器 (stacked auto-encoder, SAE) 提取光谱特征, 并构建支持向量机模型, 经灰狼优化算法 (grey wolf optimizer, GWO) 优化后模型精度为 98.75%。杨冬凤等^[17]、YUAN 等^[18]采用连续投影算法 (successive

收稿日期: 2024-06-02 修订日期: 2024-09-23

基金项目: 国家自然科学基金项目 (U2003109)

作者简介: 杜文玲, 研究方向为高光谱遥感在农业领域的应用。

Email: duwenling@xjshzu.com

※通信作者: 郭鹏, 博士, 教授, 研究方向为农业信息化和遥感创新应用。

Email: gp163@163.com

projection algorithm, SPA)、竞争性自适应重加权 (competitive adaptive reweighted sampling, CARS) 等特征筛选算法结合平均影响值 (mean impact value, MIV)、相关系数筛选光谱特征, 并采用机器学习算法对人工老化的玉米种子进行识别, 模型准确率均在 99% 以上。ZHAO 等^[19]、QI 等^[20] 将高光谱技术与深度学习结合, 实现了不同活力等级的水稻种子与糯玉米种子识别。CUI 等^[21] 基于高光谱数据的一阶微分特征, 实现了不同活力玉米种子根长的预测。LIU 等^[22] 提出空间分辨光谱 (spatially resolved spectroscopy, SRS) 和光谱组合分析的光谱比方法来检测不同玉米种子活力, 准确率达到 91.67%。目前, 基于高光谱技术检测种子活力的研究大多基于传统的光谱特征分析方法, 对不同活力种子之间光谱特征的复杂信息和精细结构还有待进一步的挖掘。

凭借着时频局域性、多尺度分析、有效分离信号和噪声以及丰富的小波基函数等优点, 连续小波变换 (continuous wavelet transform, CWT) 能够有效揭示光谱数据中的精细结构, 捕捉复杂的光谱变化, 帮助研究人员更好理解和解释光谱数据中的复杂信息^[23-24]。在实际应用中, 连续小波变换已被证明能够揭示光谱与水分、土壤有机质、土壤氮元素、植物叶绿素、植物冠层成分含量等物质之间的相关性^[25-26], 并通过结合作物冠层高光谱反射率数据估算作物参数和检测和鉴定作物病虫害等生理变化^[26-28]。已有研究^[29] 将离散小波变换 (discrete wavelet transformation, DWT) 应用于种子活力检测, 并取得了较好的结果。DWT 根据一定的规则对尺度和位移进行离散采样^[30], 一定程度上减小了小波变换系数的冗余度, 但也会导致某些细节信息的丢失, 此外, 相关研究表明, 连续小波变换具有比离散小波变换更加细致的尺度空间, 在信号奇异性检测方面比离散小波变换更具有优越性^[31]。目前, 在小波变换中, 小波系数的提取通常依赖于相关性方法, 该方法大多基于目标变量与小波系数之间的线性关系^[25,27-28], 提取不同尺度和位置的特征, 忽略了特征之间的交互作用, 小波系数的局部模式和细节信息等尚未被充分挖掘。

为了填补 CWT 在种子活力检测中的空缺, 同时进一步挖掘小波系数的精细结构和复杂信息, 本研究对新陆早 64、金科 20、金科 21 三类脱绒棉种分别进行人工老化试验并采集其高光谱数据, 提取单粒棉种平均光谱, 并对其进行多个尺度的连续小波变换, 对比不同小波基函数在分解和重构不同活力等级脱绒棉种光谱特征的能力。并提出一种基于相关性及特征重要性提取小波特征的方法, 综合考虑小波系数特征与目标变量之间的关联、非线性关系、特征交互作用及特征重要性等, 充分挖掘小波系数特征的细节信息和局部特征, 为种子活力高光谱检测提供了一种新的光谱特征分析思路。

1 材料与方法

1.1 样本制备与数据获取

以高产、优产、抗逆、抗病性强在石河子种植广泛的新陆早 64、金科 20、金科 21 三个品种脱绒棉种为试验对象, 选取颗粒饱满、大小均匀、无损伤霉变的脱绒棉种各 400 粒, 将各品种种子四等分, 其中三份采用 50% 甲醇溶液对其分别进行 30、60、90 min 老化处理^[32], 并采集其高光谱影像。试验所用的高光谱影像采集设备 (图 1) 为芬兰奥卢公司生产的 Specim IQ 便携式高光谱相机, 采集的波谱范围为 397.32~1 003.58 nm, 共 204 个通道, 光谱分辨率为 7 nm。由于棉种表层覆盖一层包衣, 难以区辨种胚的位置, 参考相关研究, 对高光谱影像中的单粒棉种绘制感兴趣区, 单个棉种感兴趣区大小在 120~200 个像素范围内, 计算感兴趣区内所有像素的光谱均值作为棉种的平均光谱^[16]。



图 1 高光谱数据采集及感兴趣区提取
Fig.1 Hyperspectral data acquisition and region of interest extraction

1.2 标准发芽试验与活力等级划分

对 4 个老化梯度 (0、30、60、90 min) 的脱绒棉种采用砂发芽法, 以砂床为介质, 将发芽盒放置在 25 ℃、光照环境下进行种子萌发 12 d 标准发芽试验, 以 GB/T3543.4《农作物种子发芽检验规程》中对正常幼苗的鉴定标准统计并计算不同处理后的种子发芽势、发芽率及活力指数^[33-34]。

由表 1 可知, 随着老化时间的加长, 各活力指标总体上呈现下降趋势, 其中活力指数综合反映了种子发芽和幼苗生长情况, 且在不同老化梯度中表现出显著差异, 反应了不同老化梯度下种子的活力水平差异^[7,17,35-37]。对活力指数与老化时间进行相关性分析, 各品种的活力指数与老化时间都呈现显著负相关 (各品种的相关系数 R 分别为 -0.930 2 (JK21)、-0.999 4 (JK20)、-0.980 3 (XLZ64))。因此, 以活力指数差异性为依据, 进行各品种脱绒棉种活力等级划分。具体划分情况如表 1。

其中, XLZ64 在 T3 处理下种子的发芽势、发芽率和活力指数均反常高于 T2, 在许思^[35]、温增莲^[36]、刘锋^[38] 等的研究中同样出现了老化时间长种子的活力指标高于老化时间短的种子的情况。其原因可能在于 XLZ64 经 90 min 老化处理后, 种子内对发芽有抑制作用的可溶性酚类化合物开始进行了转化^[39], 导致其活力有所提升, 因此将其剔除。

表 1 不同老化处理下脱绒棉种活力情况

Table 1 Vigor of delinted cotton seeds under different aging treatments

品种 Variety	编号 No.	老化时间 Ageing time/min	老化梯度 Aging gradient	发芽势 Germination potential/%	发芽率 Germination rate/%	活力指数 Vigor index	活力等级 Energy level
金科 21	JK21	0	T0	60.00±8.49a	76.50±4.24a	110.31±6.15a	A
		30	T1	50.50±1.41ab	64.00±1.41b	74.25±6.72b	B
		60	T2	45.00±1.41b	58.00±1.41bc	65.52±1.10c	C
		90	T3	38.00±2.83b	51.50±2.83c	56.73±2.31 d	D
金科 20	JK20	0	T0	18.00±2.83a	45.00±7.07a	36.81±0.96a	A
		30	T1	16.50±1.41a	33.50±4.24ab	32.13±1.23b	B
		60	T2	17.00±1.41a	33.00±1.41ab	26.59±0.41c	C
		90	T3	12.00±2.83a	29.00±1.41b	21.96±1.19 d	D
新陆早 64	XLZ 64	0	T0	17.50±2.83a	57.00±12.73a	37.28±5.06a	A
		30	T1	7.00±2.83b	18.50±1.41b	14.12±2.94b	B
		60	T2	0.00±0.00b	7.00±1.41b	2.95±1.39c	C
		90	T3	3.00	16.00	7.21	

注：表中不同小写字母表示同一指标 $P<0.05$ 差异显著。

Note: Different lowercase letters in the table indicate that the same index is significant different at $P<0.05$.

1.3 研究方法

1.3.1 光谱预处理及异常样本剔除

采用 SG 平滑、多元散射校正、一阶微分、二阶微分等方法对提取的感兴趣区的种子平均光谱进行降噪，由于两端噪声较大，选择 400~1 000 nm 范围内的光谱。同时，为剔除异常光谱样本，采用马氏距离法^[40]对预处理后的光谱数据进行异常样本剔除，各品种每个老化梯度均检测出 4 个异常样本。最终共剔除所有品种的 1 100 个样本中的 44 个异常样本。

对剔除异常值的样本以 7:3 的比例划分训练集和测试集，各数据集的样本数如表 2 所示。

表 2 试验棉种样本数

Table 2 Number of test cotton seed samples

品种 Variety	编号 No.	样本数量 Number of samples	训练集样本数量 Number of train set samples	测试集样本数量 Number of test set samples
金科 21	JK21	384	268	116
金科 20	JK20	384	268	116
新陆早 64	XLZ64	288	201	87

1.3.2 小波特征提取

连续小波变换能够同时提取时间和频率的局部信息^[24,41]，细化不同尺度和位置的函数或信号，与傅里叶变换相比，在处理非平稳信号时更具优势。CWT 通过缩放和平移小波基函数并与其进行卷积运算获取信号在不同尺度和位置上的局部特征^[24,30,42]。不同尺度上能量系数能够更加全面的分析光谱数据的特性，提取光谱中的弱信息。CWT 的计算式如下：

$$\begin{cases} W_f(a,b) \leq f \\ \psi_{a,b} \geq \int_{-\infty}^{+\infty} f(\lambda) \psi_{a,b}(\lambda) d\lambda \end{cases} \quad (1)$$

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi \frac{\lambda - b}{a} \quad (2)$$

式中 a 为尺度因子； b 为平移因子； λ 为光谱波段（400~1 000 nm）； $f(\lambda)$ 为种子光谱反射率； $\psi_{a,b}$ 为小波基函数； $W_f(a,b)$ 为小波系数，是由分解尺度和光谱波段组成的二维矩阵。为了方便计算同时不影响 CWT 的准确性，仅保留二进分解尺度（即 2^n ， $n = 1, 2, \dots, 10$ ）^[24,43-44]，记为第 1 尺度（CWT-1），第 2 尺度（CWT-2），...，第 10 尺度（CWT-10）。

1.3.3 特征筛选

1) 主成分分析

主成分分析（principal components analysis, PCA）是最常用的降维方法之一，其原理是将原始变量进行线性组合，在减少信息损失的前提下，将多个指标转化为较少的互不相关的综合指标，即主成分，能够有效地实现数据压缩与消除冗余^[45]。

2) 基于相关性及特征重要性的小波特征提取

基于相关性分析，提取目标变量敏感小波特征的研究较为成熟^[25-27,46]，在此基础上，本研究探索了基于随机森林特征重要性评价筛选小波特征的可行性，并综合考虑小波特征与种子活力等级的相关性及其在种子活力识别中的贡献度，提取了与种子活力的相关性在前 1% 的小波特征（1%|R|-WFs）、在种子活力识别中特征重要性在前 1% 的小波特征（1%Importance-WFs）及二者融合的 1%|R|+1%Importance-WFs 3 个 WFs 特征集。

① 1%|R|-WFs

对各尺度小波系数特征与种子活力等级进行相关性分析，得到相关系数 R ，并对 R 取绝对值，得到小波系数特征与种子活力等级相关系数绝对值在不同尺度和不同波段的分布图，对 $|R|$ 降序排序，为了降低冗余，提取 $|R|$ 在前 1% 的小波特征构建 1%|R|-WFs 特征集^[24,46]。

② 1%Importance-WFs

随机森林通过构建决策树能够捕捉数据中的非线性关系，同时能够综合考虑特征之间的交互作用及特征重要性^[47]，使其更好地提取小波特征的精细结构和复杂信息。

利用 RF 的特征重要性评价，基于随机森林的袋外数据（out of bag, OOB）误差，评估各尺度小波系数特征的重要性，衡量各尺度的小波系数特征在识别种子活力等级过程中的贡献度^[47-48]，得到各尺度小波系数特征在不同波段的特征重要性分布图，对特征重要性进行降序排序，提取特征重要性在前 1% 的小波特征构建 1%Importance-WFs 特征集。

③ 1%|R|+1%Importance-WFs

为了进一步地提高模型的性能，增强模型的可解释性，将特征与目标变量之间较强相关性的 1%|R|-WFs 特征集与识别目标变量过程中贡献度较高的 1%Importance-

WFs 特征集相融合，并剔除二者重叠的特征，得到 $1\%|R|+1\%Importance$ -WFs 特征集。

1.3.4 模型构建与检验

为了探究小波特征在种子活力检测中的适应性，分别对预处理后的光谱特征及小波特征进行主成分分析降维，基于降维后的光谱特征主成分和小波特征主成分建立支持向量机 (support vector machines, SVM) [49]、随机森林 (random forest, RF) [50]、极限学习机 (extreme learning machines, ELM) [51] 和反向传播神经网络 (back propagation neural network, BPNN) [52] 等多种机器学习模型对不同活力等级的脱绒棉种进行识别。此外，将各单一尺度小波特征、相关性分析及特征重要性分析提取的 3 个小波特征集也输入到上述种子活力等级识别模型，并计算准确率 (accuracy)、总体精度 (overall accuracy, OA) 和 Kappa 系数等检验各模型精度。

2 结果与分析

2.1 特征提取

2.1.1 原始光谱特征提取

分析不同活力等级脱绒棉种的平均光谱曲线 (图 2)，其中 JK21 经 T1、T2、T3 老化处理的 B、C、D 级种子平均光谱曲线与 T0 处理的 A 级种子平均光谱曲线差异明显，在 600~750 nm 范围内光谱反射率明显降低，其原因可能在于受到老化胁迫后的种子色素含量发生变化，在此波段对光的吸收增强，导致反射率降低 [53]，但 B、C、D 级种子平均光谱曲线总体走势相似，且在不同波段内存在重叠，难以从光谱内直接区分种子的不同活力等级。JK20 经 T3 老化处理后的 C 级种子在 420~550 nm 范围内的反射率明显高于 A、B 级种子的光谱反射率，但 A、B 级种子光谱曲线相似度较高，难以区分。XLZ64 的 A、B、C 各活力等级种子平均光谱的整体趋势与波峰、波谷位置均高度相似。各品种种子的光谱反射率与不同活力等级之间均无显著相关关系。

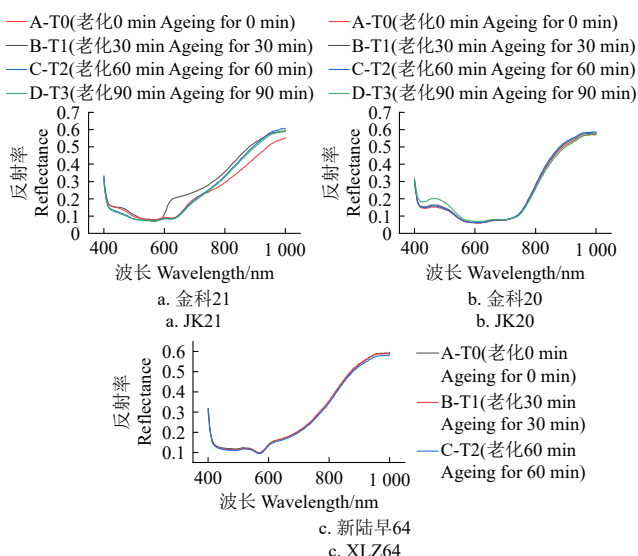


图 2 不同活力等级种子平均光谱曲线

Fig.2 Average spectral curves of seeds with different vigor levels

2.1.2 光谱预处理

本研究采用 SG 平滑、多元散射校正、一阶微分、二阶微分等方法对光谱进行预处理，并基于 SVM 建立种子活力识别模型，各模型训练集和测试集的准确率如表 3 所示。由表可知，JK21 和 JK20 原始光谱经二阶微分预处理后的效果最好；而 XLZ64 经多元散射校正预处理后的效果最好。

表 3 不同预处理方法各品种活力预测准确率

Table 3 The prediction accuracy of vigor of different varieties under different pretreatment methods %

预处理方法 Pretreatment method	金科 21 JK21		金科 20 JK20		新陆早 64 XLZ64	
	训练集 Train set		训练集 Train set		训练集 Train set	
	Test set	Test set	Test set	Test set	Test set	Test set
原始光谱 Original spectral	73.00	72.00	59.70	59.48	52.00	46.39
多项式平滑 Savitzky-Golay smoothing	72.00	73.00	61.57	60.34	51.04	52.08
多元散射校正 Multiplicative scatter correction	75.00	75.00	67.91	65.52	87.67	82.43
一阶微分 First-order differential	80.00	79.00	74.25	73.27	73.61	65.62
二阶微分 Second-order differentiation	93.33	93.00	91.42	90.52	82.29	73.96

2.1.3 小波特征提取

分别对 JK21、JK20、XLZ64 三个品种最优预处理后的光谱进行连续小波变换，对比分析了 gauss4、mexh 和 bior6.8 等不同小波基函数分解提取不同活力等级脱绒棉种光谱信息的能力，研究发现，gauss 4 和 mexh 分解各品种脱绒棉种不同活力等级种子光谱信息的效果较差，提取第 6~10 等较高尺度的光谱的低频信息时，在信号突变点出现明显振铃效应，说明该小波基函数与光谱信息的局部特征不匹配，出现了过拟合现象 [54]。基于 bior6.8 函数进行连续小波变换时，该现象明显改善，因此，选取 bior6.8 函数提取小波特征，得到各尺度小波系数曲线如图 3、4、5 所示。由图可知，各品种不同活力等级的小波特征在不同尺度上存在明显差异，在较低尺度上，光谱特征更加突出，随着尺度的增大，光谱分解能力变差，小波分解的光谱变化变得粗糙 [24]。

图 3s 中 JK21 在第 1 到第 5 尺度上，小波分解的光谱变化更为精细，在可见光的蓝光波段和近红外波段的波峰、波谷等光谱特征更为突出，第 3 到第 5 尺度的 600 nm 附近范围内，A 级种子光谱特征与 B、C、D 级种子光谱特征差异更加明显，但 B、C、D 的种子光谱特征高度相似，难以区分。

图 4 中 JK20 各尺度的小波系数特征的整体趋势和波峰、波谷等光谱特征与 JK21 相似，但未经老化处理的 A 级种子光谱特征与老化处理后的 B、C、D 级种子光谱特征未见明显差异。

图 5 中 XLZ64 在前几个尺度下经小波分解和重构的光谱特征更加突出，但各活力等级种子的光谱特征较为相似，难以区分。在第 9、10 等更高尺度上，不同活力

等级的光谱差异更加明显,但整体光谱变化变得粗糙。

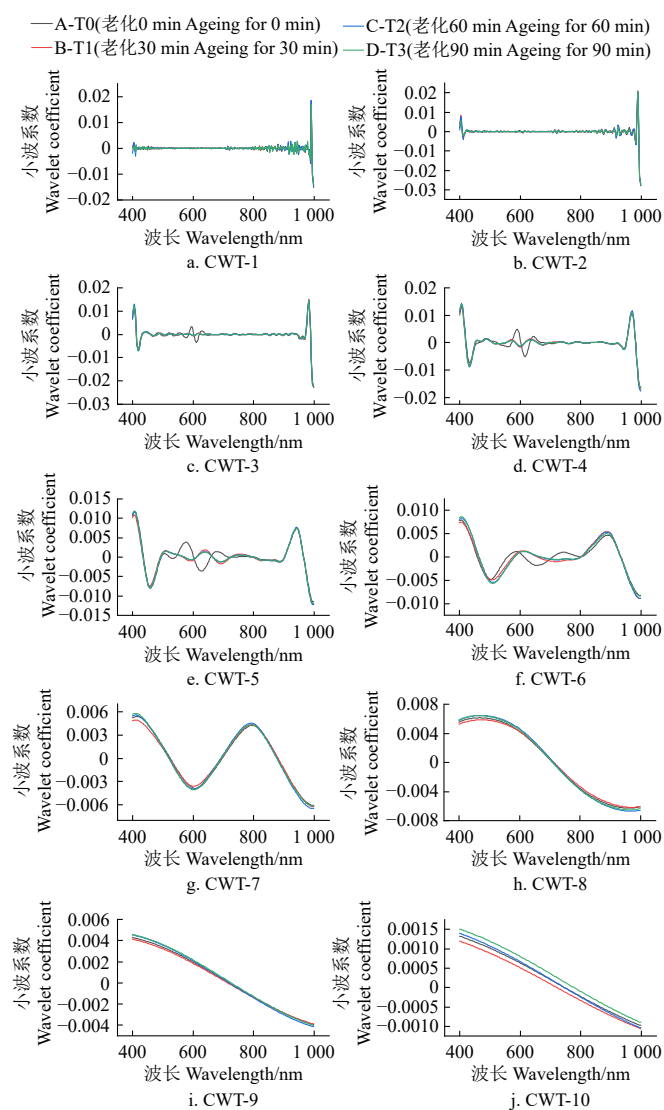


图3 JK21不同尺度下小波系数曲线

Fig.3 Wavelet coefficient curves at different scales of JK21

2.2 特征筛选

2.2.1 主成分分析

对预处理后的光谱特征 (spectral features, SFs) 及小波特征 (wavelet features, WFs) 分别进行 PCA 降维, 选择累积贡献率 90% 以上的主成分, 最终各处理的主成分及累积贡献率如表 4。

由各品种预处理光谱及小波特征的主成分得分散点图 6 可知, 对预处理光谱特征提取的主成分 (PCA-SFs), 各活力等级样本均存在不同程度的重叠, 而小波特征主成分 (PCA-WFs) 中, 样本的差异性更加明显, 其中, JK21 的 A 级种子与 B、C、D 级种子的样本分离性较好, 但 B、C、D 级种子的样本之间重叠较为严重; JK20 的 B、C、D 级之间的样本分离性较好, A、B 级的样本重叠较严重; XLZ64 各活力等级的样本差异性最好, 各样本之间完全分离, 没有重叠区域。

2.2.2 相关性及特征重要性分析

基于相关性和特征重要性对各品种种子不同尺度的

WFs 进行筛选, 提取 1%|R|-WFs 与 1%Importance-WFs 各 20 个特征, 图 7~9 中深红色波段为提取的 1%|R|-WFs 与 1%Importance-WFs, 在各尺度和波段中的分布如下。

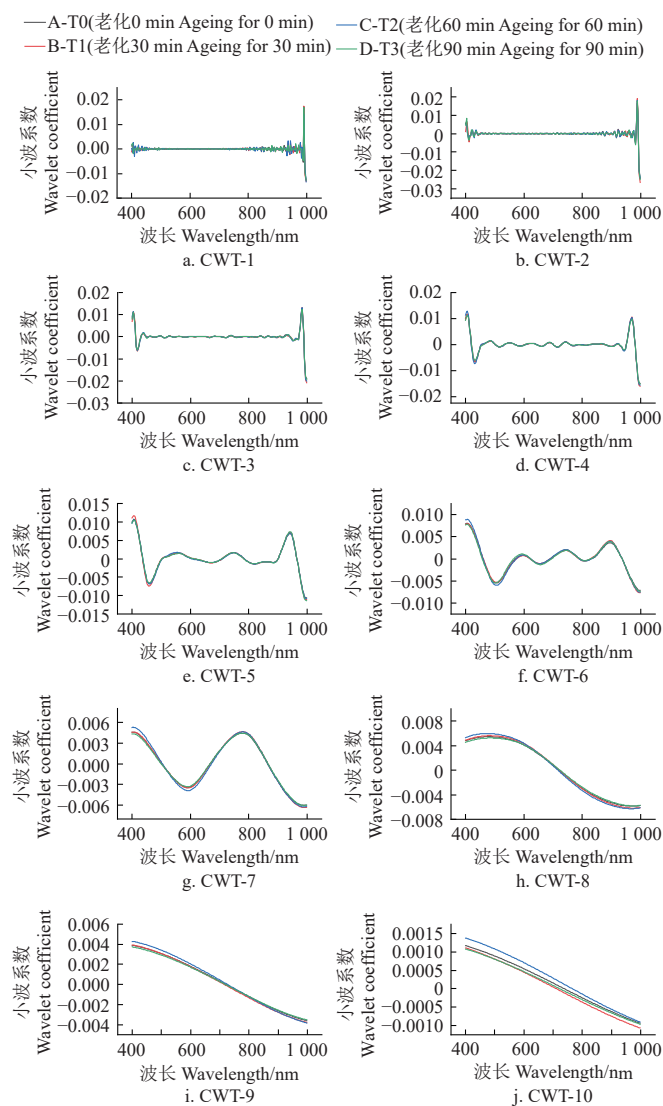


图4 JK20不同尺度下小波系数曲线

Fig.4 Wavelet coefficient curves at different scales of JK20

图 7 中, 对 JK21 提取的 1%|R|-WFs 主要在第 4 和第 6 尺度, 特征集中在 550~650 nm 的可见光波段, 充分考虑了未经老化处理的 A 级种子与老化处理的 B、C、D 级种子在该范围内的特征差异。1%Importance-WFs 在尺度分布上主要位于前 3 个尺度和第 7 尺度, 在光谱分布上位于波峰、波谷等光谱特征较为突出的 400~600 nm 范围的可见光及 900 nm 附近的近红外波段。

图 8 中, JK20 相关性及特征重要性提取的 WFs 均主要分布在光谱特征较为突出的前 4 个尺度, 其中 1%|R|-WFs 较为分散地分布在可见光到近红外波段, 1%Importance-WFs 则主要分布在 450 nm、600 nm 附近的可见光及 750~780 nm 附近的近红外波段。

图 9 中, 对 XLZ64 提取的 1%|R|-WFs 分布在样本可分性较好的第 8、9 尺度, 特征集中分布在 450 及 580 nm 附近。1%Importance-WFs 主要分布在前 2 个尺度及第 7、

8 尺度，特征分布在光谱特征突出的 400~600 nm 范围的可见光波段及 900~1 000 nm 范围的近红外波段。

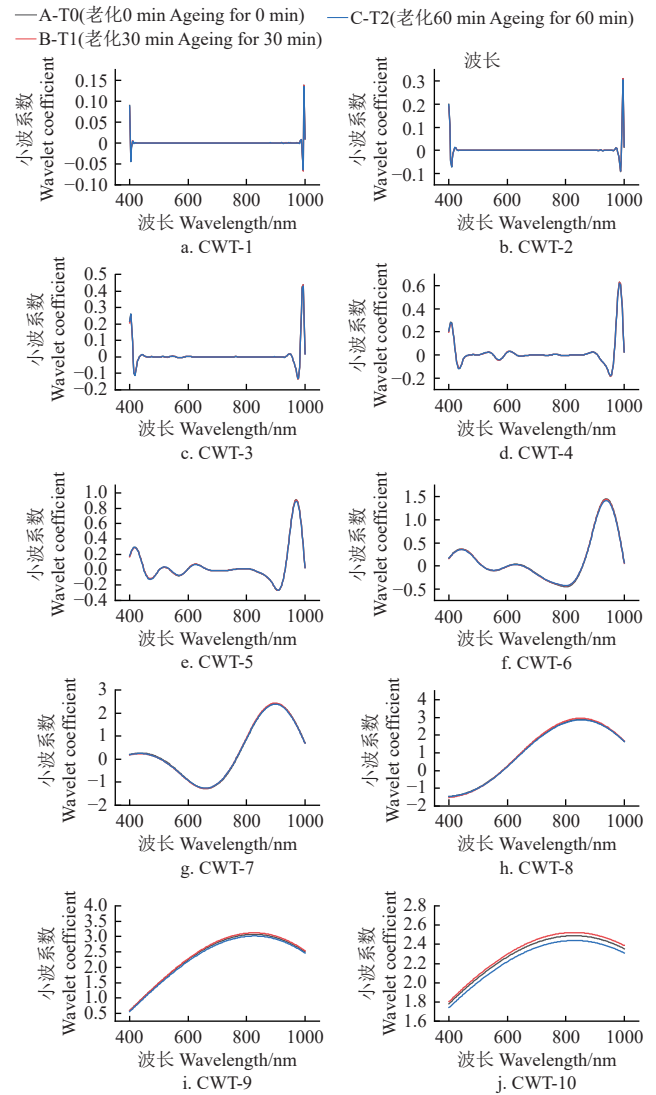


图 5 XLZ64 不同尺度下小波系数曲线
Fig.5 Wavelet coefficient curves at different scales of XLZ64

表 4 光谱特征及小波特征主成分及累积贡献率
Table 4 Principal component and cumulative contribution rate of spectral features and wavelet features

品种 Seed varieties	金科 21 JK21		金科 20 JK20		新陆早 64 XLZ64	
	SFs	WFs	SFs	WFs	SFs	WFs
主成分 Principal component	16	9	10	7	4	2
累积贡献率 Accumulative contribution rate/%	90.64	90.61	90.72	90.99	95.68	93.00

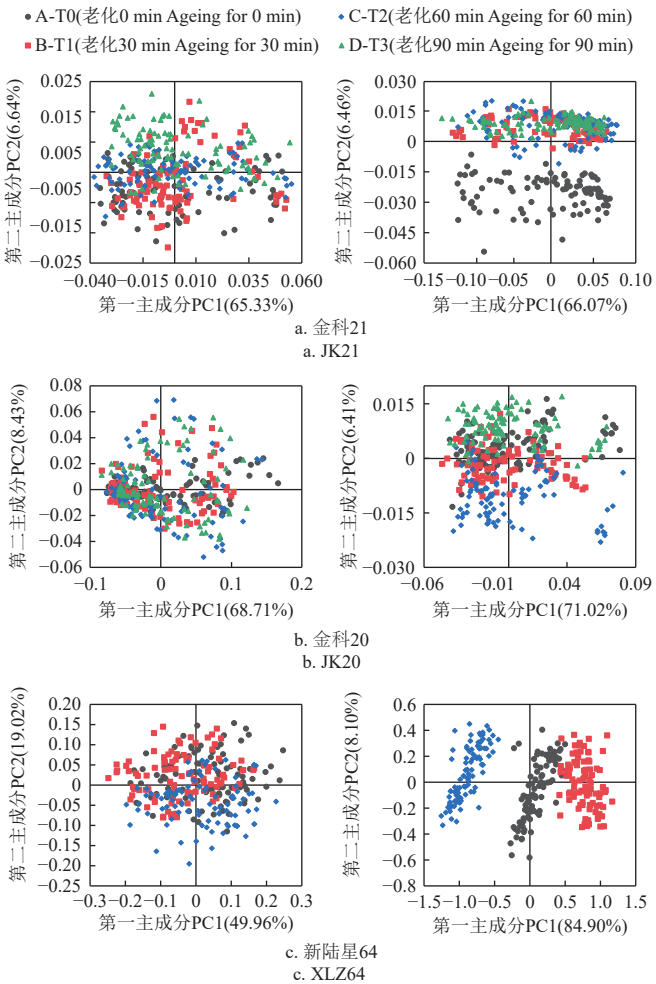
注：SFs 为光谱特征 (spectral features, SFs)，WFs 为小波特征 (wavelet features, WFs)。
Note: SFs are spectral features (SFs), and WFs are wavelet features (WFs).

2.3 活力检测模型建立及评价

2.3.1 基于 RF 构建单尺度 WFs 脱绒棉种活力检测模型

对所有单个尺度的 WFs 分别构建基于 RF 算法的脱绒棉种活力检测模型，得到的总体精度 (overall accuracy, OA) 和 Kappa 系数如图 10 所示。随着尺度的增大，小波分解得到的光谱变化特征变得粗糙，检测种子活力的能力下降，各品种种子活力检测的总体精度和

Kappa 系数均有所下降。而 XLZ64 不同活力的种子光谱在第 9、10 尺度上差异明显，因此在较高尺度上，检测精度又有所上升。该结果再一次证明了较低尺度小波特征在 JK21、JK20、XLZ64 种子活力检测中的优越性，及较高尺度小波系数特征在 XLZ64 种子活力检测中的可分性。



注：左图为预处理光谱，右图为小波系数特征。
Note: The left image is the pretreatment spectra, and the right image is the wavelet coefficient feature.

图 6 各品种预处理光谱与小波系数特征主成分得分散点图
Fig.6 The scatter plot of principal component scores of pretreatment spectra and wavelet coefficient feature of each variety

2.3.2 基于 PCA-SFs 与 PCA-WFs 的脱绒棉种活力检测

分别基于 PCA-SFs 与 PCA-WFs 构建 SVM、RF、ELM 和 BPNN 种子活力检测模型，对 3 个品种脱绒棉种活力进行识别，得到各模型的精度如表 5。由于 PCA 在降维过程中不可避免损失了部分数据信息，这些丢失的信息可能影响分类的精度^[55-56]，因此 JK21、JK20 主成分降维后的精度略低于降维前光谱特征的建模精度。

光谱特征与小波特征建模对比分析可知，3 个品种各模型的 PCA-WFs 建模精度均优于 PCA-SFs 建模精度，其中 JK21 和 JK20 的提升效果较为显著，而 XLZ64 由于其 PCA-SFs 的建模精度已经较高，PCA-WFs 的提升效果并不十分显著，表明 CWT 在识别精细的种子活力

等级中具有一定的优越性。PCA-WFs 模型的高精度与 PCA-WFs 主成分得分图中的样本差异性, 表明 CWT 用于种子活力高光谱检测具备可行性。

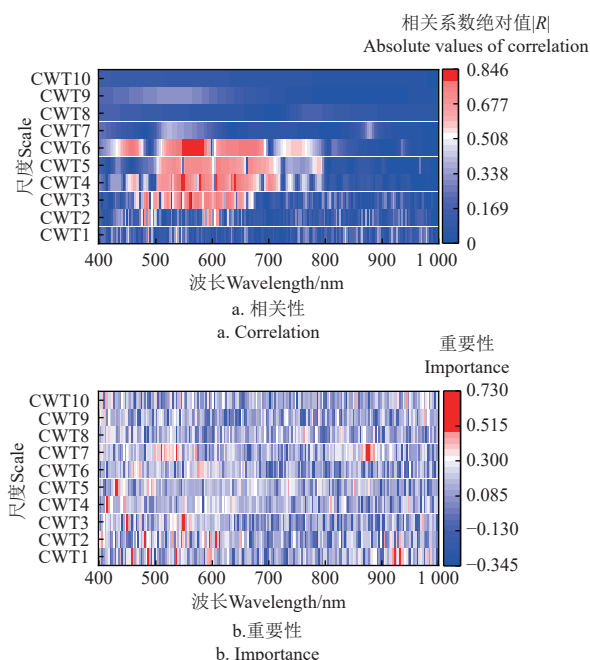


图7 金科21不同尺度下小波系数特征的相关性及特征重要性

Fig.7 The correlation and feature importance of wavelet coefficient characteristics at different scales in JK21

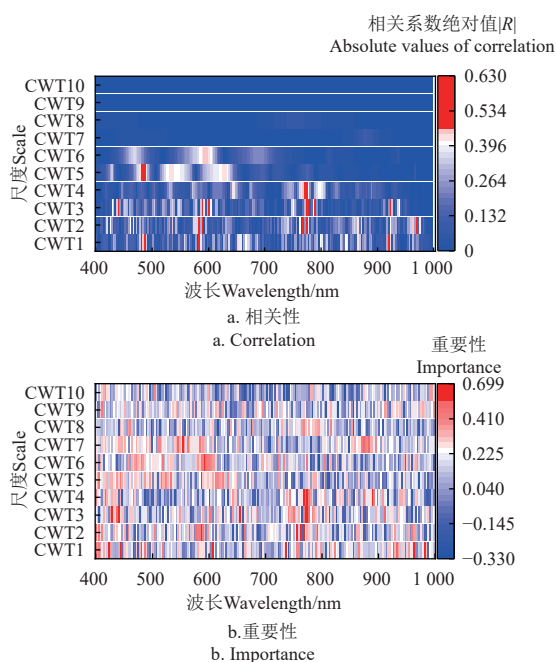


图8 金科20不同尺度下小波系数特征的相关性及特征重要性

Fig.8 The correlation and feature importance of wavelet coefficient characteristics at different scales in JK20

从各模型精度对比分析, 各品种基于 PCA-SFs 与 PCA-WFs 的建模结果中, 表现最好的模型均为 ELM, BPNN 与 RF 次之, SVM 的精度最低。ELM 和 BPNN 模型能够通过激活函数捕捉复杂特征, 对复杂数据有较强的适应能力, 在处理非线性数据时表现良好^[57-59]。RF 通过构建多棵决策树, 同样能够有效捕捉数据中的复杂非

线性关系^[60]。因此, 在具有明显的非线性特征的高维数据集中, ELM、BPNN 和 RF 模型能更好地泛化, 具有较高的分类精度。SVM 通过寻找最佳超平面来进行分类, 计算复杂度较高, 在数据分布不均或特征维度较高时, 可能会出现过拟合或欠拟合^[61]。

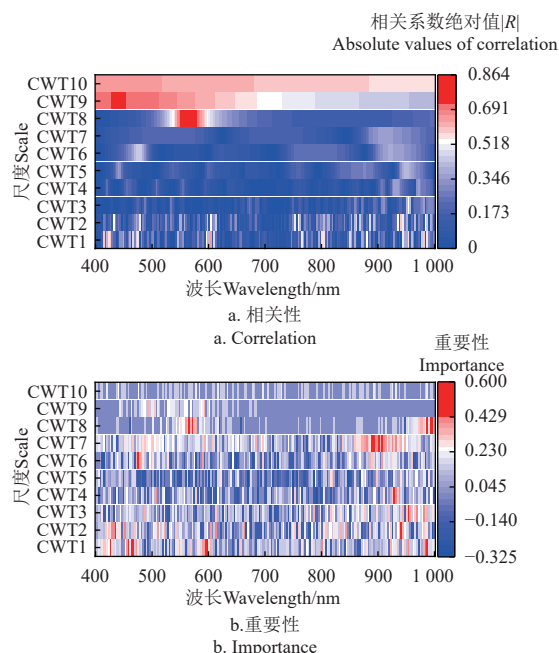


图9 新陆早64不同尺度下小波系数特征的相关性及特征重要性

Fig.9 The correlation and feature importance of wavelet coefficient characteristics at different scales in XLZ64

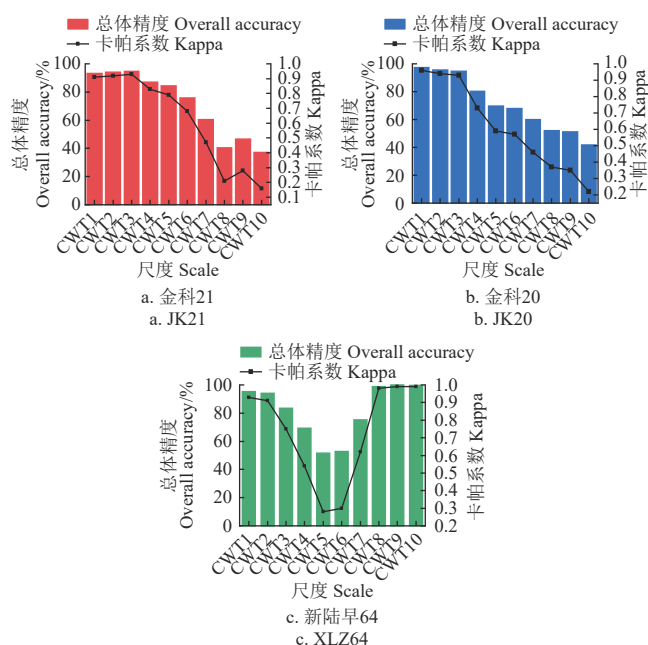


图10 单个分解尺度下小波系数特征 RF 建模精度

Fig.10 RF modeling accuracy of wavelet coefficient characteristics at a single decomposition scale

2.3.3 基于相关性及特征重要性小波特征的脱绒棉种活力检测

分别将 1%|R|-WFs、1%Importance-WFs 及融合相关性和特征重要性的小波系数特征 (1%|R|+1%Importance-

WFs) 代入上述机器学习模型, 对比各特征及模型组合的建模精度(表 6-表 8)。从特征筛选来看, 1%|R|+1%Importance-WFs 在各品种种子活力检测模型中均表现最优, 其训练集和测试集的准确率均高于 1%|R|-WFs 和 1%Importance-WFs 的准确率, 1%Importance-WFs 建模精度次之, 1%|R|-WFs 建模精度最低。对比各机器学习模型精度, 基于同一小波系数特征筛选方法, 总体上, 在各品种种子活力检测模型中 ELM 的准确率最高, BPNN 和 RF 次之, SVM 的准确率最低。对比同一机器

学习模型不同特征筛选方法, ELM 和 BPNN 模型较为稳定, RF 次之, SVM 模型的稳定性最差。不同品种种子活力检测模型中, JK21 的 1%|R|+1%Importance-WFs 与 ELM 模型组合的精度最高, 训练集和测试集的精度达到 99.63% 和 98.28%。JK20 的 1%|R|+1%Importance-WFs 与 ELM 精度最高, 训练集和测试集的准确率均达到 100%。XLZ64 种子活力检测模型中, 基于 1%|R|+1%Importance-WFs 的特征筛选方法与各机器模型组合的训练集和测试集的准确率均达到 100%。

表 5 光谱特征及小波系数特征主成分建模精度

Table 5 Modeling accuracy of principal components of spectral features and wavelet coefficient features %

品种 Variety		支持向量机 SVM		随机森林 RF		极限学习机 ELM		反向传播神经网络 BPNN	
		训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set
金科 21 JK21	PCA-SFs	89.55	86.21	99.63	87.93	96.94	91.38	90.29	87.07
	PCA-WFs	93.96	92.54	100	93.1	97.76	96.55	95.52	93.1
金科 20 JK20	PCA-SFs	88.43	87.07	92.16	91.38	95.15	93.66	92.91	93.10
	PCA-WFs	94.78	95.52	97.02	96.55	99.25	97.41	98.13	96.55
新陆早 64 XLZ64	PCA-SFs	98.99	97.67	100	97.67	100	100	100	100
	PCA-WFs	99.49	98.84	100	99.89	100	100	100	100

表 6 JK21 相关性及特征重要性小波系数特征建模精度

Table 6 Modeling accuracy of JK21 correlation and feature importancer of wavelet coefficient features %

指标 Index	支持向量机 SVM		随机森林 RF		极限学习机 ELM		反向传播神经网络 BPNN	
	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set
1%Importance	96.27	97.41	100.00	97.41	99.63	97.1	98.88	93.96
1% R	72.76	75.86	100.00	87.07	97.76	93.97	95.9	93.97
1%Importance+1% R	97.39	95.69	100.00	96.55	99.63	98.28	99.25	94.83

表 7 JK20 相关性及特征重要性小波系数特征建模精度

Table 7 Modeling accuracy of JK20 correlation and feature importancer of wavelet coefficient features %

指标 Index	支持向量机 SVM		随机森林 RF		极限学习机 ELM		反向传播神经网络 BPNN	
	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set
1%Importance	97.76	98.25	100.00	97.73	99.63	98.25	98.51	97.37
1% R	87.68	85.09	99.25	92.21	96.27	93.68	94.40	90.35
1%Importance+1% R	99.25	99.12	100.00	98.25	100.00	100.00	99.63	98.25

表 8 XLZ64 相关性及特征重要性小波系数特征建模精度

Table 8 Modeling accuracy of XLZ64 correlation and feature importancer of wavelet coefficient features %

指标 Index	支持向量机 SVM		随机森林 RF		极限学习机 ELM		反向传播神经网络 BPNN	
	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set	训练集 Train set	测试集 Test set
1%Importance	99.00	98.89	100.00	97.47	100.00	98.82	99.5	98.82
1% R	97.00	96.47	100.00	98.82	100.00	100.00	100.00	99.5.00
1%Importance+1% R	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

3 讨 论

将小波特征与目标变量进行相关性分析, 提取敏感波段的研究相对成熟, 但在本研究中基于相关性提取脱绒棉种活力的敏感小波特征建模精度未见明显提升, 而基于随机森林特征重要性评价的方法提取敏感特征的建模精度更优, 此外, 本研究发现综合考虑相关性与特征重要性提取的 WFs 在尺度分布上基本上与基于各单一尺度小波特征建模精度较高的尺度相符; 在光谱分布上与小波特征分析的突出特征较为一致。如 JK21 基于相关性提取的 1%|R|-WFs 充分考虑了未经老化处理的种子与老化处理后种子在第 4 尺度上 550~650 nm 范围内的光谱差异, 基于特征重要性提取的 1%Importance-WFs 主要分布单个尺度 WFs 建模精度较高前 3 个尺度上, 同时在光谱

分布上考虑了各活力等级种子在较低尺度上可见光及近红外的波峰、波谷特征。JK20 提取的 1%|R|-WFs、1%Importance-WFs 在尺度分布上与 JK20 单个尺度 WFs 建模精度较高的模型在尺度上的分布相一致, 在光谱分布上, 与低尺度波峰波谷特征波段相符。XLZ64 提取的 1%|R|-WFs 和 1%Importance-WFs 包含了 XLZ64 单个尺度 WFs 建模精度较高的高尺度 WFs, 1%Importance-WFs 在第 7 尺度上的特征充分考虑了各活力种子样本之间表现显著差异的 880~900 nm 范围的特征, 在第 1-3 尺度上, 考虑了 WFs 在 950~1 000 nm 范围的近红外波峰、波谷特征。说明本研究提出的基于相关性及特征重要性筛选小波系数特征方法具备一定的合理性, 能够充分挖掘种子活力等级识别的有效光谱信息。

在本研究中还存在以下不足：在研究对象上，仅采集了不同活力等级的脱绒棉种高光谱影像来分析 CWT 结合高光谱检测种子活力的可行性，对于其他种子未作进一步的研究，本研究所选取的 bior6.8 小波基函数是否适用于其他种子活力检测有待进一步研究。此外，对于 XLZ64 在 T3 处理下种子的各活力指标均反常高于 T2 的情况的原因未进行深入的探究分析，在接下来的研究中，本文将基于以上不足，做出进一步的研究与完善。

4 结 论

本研究以 JK21、JK20、XLZ64 三个品种老化处理的脱绒棉种为对象，采集各品种不同活力等级的高光谱影像，并对比了 gauss 4、mexh、bior6.8 等小波基函数提取小波特征 (wavelet features, WFs) 的能力，筛选出适用于脱绒棉种活力检测的小波基函数，并基于该函数提取了 10 个尺度的 WFs，对光谱特征 (spectral features, SFs) 与 WFs 进行主成分分析，并建立种子活力检测模型对比二者的精度。最后分析了基于 1%|R|-WFs、1%Importance-WFs 及 1%|R|+1%Importance-WFs 3 个特征集与支持向量机 (support vector machines, SVM)、随机森林 (random forest, RF)、极限学习机 (extreme learning machines, ELM) 和反向传播神经网络 (back propagation neural network, BPNN) 等机器学习模型组合的建模精度，并得出以下结论：

1) 对不同活力等级种子的光谱特征进行连续小波变换时，gauss4、mexh 等小波基函数提取的 WFs 在较高尺度提取光谱低频信息时，出现了明显的振铃效应，而基于 bior6.8 函数提取的 WFs 则未出现该现象，说明 bior6.8 函数的局部性和对称性可以有效地增强不同活力脱绒棉种高光谱信息的局部特征。

2) 对比 PCA-SFs 与 PCA-WFs 构建的多种机器学习模型精度，基于 PCA-WFs 的模型精度均优于基于 PCA-SFs 的模型精度，对比相关性和特征重要性筛选的 3 个小波特征集构建的各品种种子活力检测模型，基于 1%|R|+1%Importance-WFs 模型精度优于基于 1%Importance-WFs 的模型精度，基于 1%|R|-WFs 的模型精度最差，对比各机器学习模型，ELM 准确率最高、BPNN 与 RF 次之，SVM 准确率最低。

3) 对比各品种所有特征集与机器学习模型组合，JK21 种子活力检测模型的最优组合为 1%|R|+1%Importance-WFs+ELM，训练集和测试集的准确率分别为 99.63%、98.28%；JK20 最优组合为：1%|R|+1%Importance-WFs+ELM，训练集和测试集的准确率均为 100%；XLZ64 的最优组合为：PCA-WFs+ELM 或 BPNN 及 1%|R|+1%Importance-WFs+所有机器学习模型，训练集和测试集的准确率均为 100%。

[参 考 文 献]

- [1] 刘翔, 托亚. 国家级棉花棉纱交易中心缘何落子新疆[N]. 新疆日报 (汉). 2023-12-20 (002).
- [2] 刘毅. 2023 年新疆棉花总产 511.2 万吨占全国总产超九成[J]. 棉花科学, 2023, 45(2): 1.
- [3] 魏诗雯. 新疆棉花向“质”而行[N]. 中国纺织报. 2024-4-8 (002).
- [4] 任华. 去年新疆棉花质量达近年来最好水平[N]. 新疆日报 (汉), 2024-03-24 (002).
- [5] 孙俊, 张林, 周鑫, 等. 采用高光谱图像深度特征检测水稻种子活力等级[J]. 农业工程学报, 2021, 37(14): 171-178.
SUN Jun, ZHANG Lin, ZHOU Xin, et al. Detection of rice seed vigor level by using deep feature of hyperspectral images[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2021, 37(14): 171-178. (in Chinese with English abstract)
- [6] 夏宇, 孟静武, 罗斌, 等. 高光谱结合理化参数跨品种识别玉米种子贮藏年份 [J/OL]. 农业工程学报, 2024, 40 (16): 261-268.
XIAO Yu, MENG Jingwu, LUO Bin, et al. Effects of different sorting methods on seed vigor of rice[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40 (16): 261-268. (in Chinese with English abstract)
- [7] 杨波, 段明磊, 杨童. 基于高光谱成像技术的西瓜种子活力等级分类方法研究[J]. 河南农业科学, 2022, 51(9): 151-158.
YANG Bo, DUAN Minglei, YANG Tong. Research on the classification method of watermelon seed vigor level based on hyperspectral imaging technology[J]. Journal of Henan Agricultural Sciences, 2022, 51(9): 151-158. (in Chinese with English abstract)
- [8] 卢伟, 张孜谔, 蔡苗苗, 等. 基于光声光谱和 TCA 迁移学习的稻种活力检测[J]. 农业工程学报, 2020, 36(22): 341-348.
LU Wei, ZHANG Zixu, CAI Miaomiao, et al. Detection of rice seeds vigor based on photoacoustic spectrum combined with TCA transfer learning[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2020, 36(22): 341-348. (in Chinese with English abstract)
- [9] 丁子予, 岳学军, 曾凡国, 等. 基于机器学习和深度学习的玉米种子活力光谱检测[J]. 华中农业大学学报, 2023, 42(3): 230-240.
DING Ziyu, YUE Xuejun, ZENG Fanguo, et al. Spectral detection of maize seed vigor based on machine learning and deep learning[J]. Journal of Huazhong Agricultural University, 2023, 42(3): 230-240. (in Chinese with English abstract)
- [10] 彭晓伟, 张爱军, 王楠, 等. 高光谱成像技术在作物种子方面的应用[J]. 国土资源遥感, 2020, 32(4): 23-32.
PENG Xiaowei, ZHANG Aijun, WANG Nan, et al. Application of hyperspectral imaging technology in crop seeds[J]. Remote Sensing for Natural Resources, 2020, 32(4): 23-32. (in Chinese with English abstract)

- [11] HU QY, LU WY, GUO YX, et al. Vigor detection for naturally aged soybean seeds based on polarized hyperspectral imaging combined with ensemble learning algorithm [J]. *Agriculture*, 2023, 13(8): 1499.
- [12] 金文玲, 曹乃亮, 朱明东, 等. 基于近红外超连续激光光谱的水稻种子活力无损分级检测研究[J]. *中国光学*, 2020, 13(5): 1032-1043.
- JIN Wenling, CAO Nailiang, ZHU Mingdong, et al. Nondestructive grading test of rice seed activity using nearinfrared super-continuum laser spectrum[J]. *Chinese Optics*, 2020, 13(5): 1032-1043. (in Chinese with English abstract)
- [13] XING MY, YUAN L, WANG QY, et al. Physiological alterations and nondestructive test methods of crop seed vigor: A comprehensive review [J]. *Agriculture*, 2023, 13(3): 527.
- [14] GOJLO E. Activity of α -d-galactosidase in long-stored seeds of *vicia hirsuta* [J]. *Agriculture*, 2023, 13(7): 1306.
- [15] 石睿, 罗斌, 张晗, 等. 种子活力性状无损速测技术研究进展[J]. *江苏农业科学*, 2024, 52(7): 1-10.
- SHI Rui, LUO Bin, ZHANG Han, et al. Research progress on non-destructive rapid measurement technology for seed vigor[J]. *Jiangsu Agricultural Sciences*, 2024, 52(7): 1-10. (in Chinese with English abstract)
- [16] 尤佳, 李景彬, 黄勇, 等. 基于高光谱图像技术的脱绒棉种活力检测[J]. *江苏农业科学*, 2018, 46(13): 231-235.
- YOU Jia, LI Jingbin, HUANG Yong, et al. Study on detection technology of delinted cottonseed vigor based on hyperspectral image technology[J]. *Jiangsu Agricultural Sciences*, 2018, 46(13): 231-235. (in Chinese with English abstract)
- [17] 杨冬风, 李爱传, 刘金明, 等. 耦合平均影响值-连续投影算法优化种子活力近红外检测模型[J]. *光谱学与光谱分析*, 2022, 42(10): 3135-3142.
- YANG Dongfeng, LI Aichuan, LIU Jinming, et al. Optimization of seed vigor near-infrared detection by coupling mean impact value With successive projection algorithm[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(10): 3135-3142. (in Chinese with English abstract)
- [18] YUAN P, PANG L, WANG LM, et al. Application of hyperspectral imaging to discriminate waxy corn seed vigour after aging[J]. *International Food Research Journal*, 2022, 29(2): 397-405.
- [19] ZHAO X, PANG L, WANG L, et al. Deep convolutional neural network for detection and prediction of waxy corn seed viability using hyperspectral reflectance imaging[J]. *Mathematical and Computational Applications*, 2022, 27(6): 109.
- [20] QI H, HUANG Z, SUN Z, et al. Rice seed vigor detection based on near-infrared hyperspectral imaging and deep transfer learning[J]. *Frontiers in Plant Science*, 2023, 14: 1283921.
- [21] CUI H, BING Y, ZHANG X, et al. Prediction of maize seed vigor based on first-order difference characteristics of hyperspectral data[J]. *Agronomy*, 2022, 12(8): 1899.
- [22] LIU W, LUO B, KANG K, et al. Non-destructive detection of single corn seed vigor based on visible/near-infrared spatially resolved spectroscopy combined with chemometrics[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2024, 312: 124089.
- [23] 于雷, 洪永胜, 周勇, 等. 连续小波变换高光谱数据的土壤有机质含量反演模型构建[J]. *光谱学与光谱分析*, 2016, 36(5): 1428-1433.
- YU Lei, HONG Yongsheng, ZHOU Yong, et al. Inversion of soil organic matter content using hyperspectral data based on continuous wavelet transformation[J]. *Spectroscopy and Spectral Analysis*, 2016, 36(5): 1428-1433. (in Chinese with English abstract)
- [24] CHENG X, FENG Y, GUO A, et al. Detection of rubber tree powdery mildew from leaf level hyperspectral data using continuous wavelet transform and machine learning[J]. *Remote Sensing*, 2023, 16(1): 105.
- [25] 杨锡震, 陈俊英, 张秋雨, 等. 基于小波特征和冬小麦生理参数的土壤水分高光谱模型优化[J]. *农业工程学报*, 2023, 39(10): 66-75.
- YANG Xizhen, CHEN Junying, ZHANG Qiuyu, et al. Optimization of the soil moisture model based on hyperspectral inversionby integrating wavelet features and growth parameters of winter wheat[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2023, 39(10): 66-75. (in Chinese with English abstract)
- [26] DOU S, ZHANG W, DENG Y, et al. Comparison of citrus leaf water content estimations based on the continuous wavelet transform and fractional derivative methods[J]. *Horticulturae*, 2024, 10(2): 177.
- [27] 玉米提·买明, 王雪梅. 连续小波变换的土壤有机质含量高光谱估测[J]. *光谱学与光谱分析*, 2022, 42(4): 1278-1284.
- YUMITI Maiming, WANG Xuemei. Hyperspectral estimation of soil organic matter content based on continuous wavelet transformation[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(4): 1278-1284. (in Chinese with English abstract)
- [28] 刘杨, 孙乾, 冯海宽, 等. 基于小波分析的马铃薯地上生物量估算[J]. *光谱学与光谱分析*, 2021, 41(4): 1205-1212.
- LIU Yang, SUN Qian, FENG Haikuan, et al. Estimation of above-ground biomass of potato based on wavelet analysis[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(4): 1205-1212. (in Chinese with English abstract)
- [29] 杨冬风, 尹淑欣, 姜丽, 等. 玉米种子活力近红外光谱智能检测方法研究[J]. *核农学报*, 2013, 27(7): 957-961.
- YANG Dongfeng, YIN Shuxin, JIANG Li, et al. Research on maize vigor intelligent detection based on near infrared spectroscopy[J]. *Journal of Nuclear Agricultural Sciences*, 2013, 27(7): 957-961. (in Chinese with English abstract)
- [30] 龙岳红, 秦建新, 贺新光, 等. 洞庭湖流域植被动态变化的小波多分辨率分析[J]. *地理学报*, 2015, 70(9): 1491-1502.
- LONG Yuehong, QIN Jianxin, HE Xinguang, et al. Wavelet

- multi-resolution analysis of vegetation dynamic change in Dongting Lake Basin[J]. *Acta Geographica Sinica*, 2015, 70(9): 1491-1502. (in Chinese with English abstract)
- [31] 袁建虎, 何竹青, 龙云飞, 等. 低信噪比信号的连续与离散小波变换比较研究[J]. *自动化仪表*, 2009, 30(8): 8-10.
YUAN Jianhu, HE Zhuqing, LONG Yunfei, et al. Comparative study on continuous and discrete wavelet transform for low signal-to-noise ratio signal[J]. *Process Automation Instrumentation*, 2009, 30(8): 8-10. (in Chinese with English abstract)
- [32] 孟超敏, 姬俊华, 邵丹阳. 甲醇老化处理对棉花种子发芽和幼苗生长的影响[J]. *种子*, 2017, 36(3): 10-11, 15.
MENG Chaomin, JI Junhua, SHAO Danyang. Effect of aging treatment with methanol on cotton seed germination and seedling growth[J]. *Seed*, 2017, 36(3): 10-11, 15. (in Chinese with English abstract)
- [33] 李润枝. 棉花种子活力测定技术手册[M]. 北京: 中国农业大学出版社, 2018: 4-7.
- [34] 付坤. 人工老化对水稻种子活力的影响[J]. *中国农业文摘-农业工程*, 2024, 36(3): 65-69.
- [35] 许思, 赵光武, 邓飞, 等. 基于高光谱的水稻种子活力无损分级检测[J]. *种子*, 2016, 35(4): 34-40.
XU Si, ZHAO Guangwu, DENG Fei, et al. Research on detection technology of rice seed vigor based on hyperspectral[J]. *Seed*, 2016, 35(4): 34-40. (in Chinese with English abstract)
- [36] 温增莲, 罗火林, 郭夏丽, 等. 电喷雾萃取及电离质谱法分析大豆水提液快速鉴别种子活力[J]. *农业工程学报*, 2016, 32(16): 273-278.
WEN Zenglian, LUO Huolin, GUO Xiali, et al. Rapid identification of seed vigor of soybean water extracts by electrospray ionization mass spectrometry[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2016, 32(16): 273-278. (in Chinese with English abstract)
- [37] 吴静珠, 李晓琪, 刘翠玲, 等. 基于 THz-TDS 反射成像技术的玉米种子活力无损检测研究[J]. *光谱学与光谱分析*, 2020, 42(9): 2840-2844.
WU Jingshu, LI Xiaoqi, LIU Cuiling, et al. Research on nondestructive testing of corn seed vigor based on THz-TDS reflection imaging[J]. *Spectroscopy and Spectral Analysis*, 2020, 42(9): 2840-2844. (in Chinese with English abstract)
- [38] 刘锋, 胡华兵, 王荣华, 等. 甲醇老化处理对甜菜种子发芽的影响[J]. *中国农学通报*, 2022, 38(33): 28-33.
LIU Na, HU Huabing, WANG Ronghua, et al. Methanol aging treatment: effect on germination of sugar beet seeds[J]. *Chinese Agricultural Science Bulletin*, 2022, 38(33): 28-33. (in Chinese with English abstract)
- [39] 张晓洁, 隋洁, 胡苏汶. 棉花种子成熟度与活力状况分析[J]. *中国棉花*, 2004(12): 6-7.
- [40] 陈秀莲, 黄梅珍, 富雨超. 具有活体检测功能的手背静脉身份识别方法研究[J]. *光学学报*, 2021, 41 (6): 90-97.
CHEN Xiulian, HUANG Meizhen, FU Yuchao. Recognition method of dorsal hand vein with liveness detection function[J]. *Acta Optica Sinica*, 2021, 41 (6): 90-97. (in Chinese with English abstract)
- [41] 许中林, 李国禄, 董天顺, 等. 声发射信号分析与处理方法研究进展[J]. *材料导报*, 2014, 28(9): 56-60, 73.
XU Zhonglin, LI Guolu, DONG Tianshun, et al. Overview on development of acoustic emission signal analysis technique and processing[J]. *Materials Reports*, 2014, 28(9): 56-60, 73. (in Chinese with English abstract)
- [42] 尼格娜热·阿曼太, 丁建丽, 葛翔宇, 等. 1960—2017 年艾比湖流域实际蒸散量与气象要素的变化特征[J]. *地理学报*, 2021, 76(5): 1177-1192.
AMANTAI Nigenare, DING Jianli, GE Xiangyu, et al. Variation characteristics of actual evapotranspiration and meteorological elements in the Ebinur Lake basin from 1960 to 2017[J]. *Acta Geographica Sinica*, 2021, 76(5): 1177-1192. (in Chinese with English abstract)
- [43] 孟珊, 李新国. 基于高光谱小波能量特征向量估算湖滨绿洲表层土壤有机碳含量[J]. *光谱学与光谱分析*, 2023, 43(12): 3853-3861.
MENG Shan, LI Xinguo. Estimation of surface soil organic carbon content in lakeside oasis based on hyperspectral wavelet energy feature vector[J]. *Spectroscopy and Spectral Analysis*, 2023, 43(12): 3853-3861. (in Chinese with English abstract)
- [44] 郭斌, 白昊睿, 张波, 等. 基于 RF 和连续小波变换的露天煤矿土壤锌含量高光谱遥感反演[J]. *农业工程学报*, 2022, 38(10): 138-147.
GUO Bin, BAI Haorui, ZHANG Bo, et al. Inversion of soil zinc contents using hyperspectral remote sensing based on random forest and continuous wavelet transform in an opencast coal mine[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2022, 38(10): 138-147. (in Chinese with English abstract)
- [45] 胡会强, 位云朋, 徐华兴, 等. 基于高光谱成像技术和主成分分析对粉葛年限的鉴别[J]. *光谱学与光谱分析*, 2023, 43(6): 1953-1960.
HU Huiqiang, WEI Yunpeng, XU Huaxing, et al. Identification of the age of puerariae thomsonii radix based on hyperspectral imaging and principal component analysis[J]. *Spectroscopy and Spectral Analysis*, 2023, 43(6): 1953-1960. (in Chinese with English abstract)
- [46] 江远东, 李新国, 杨涵. 基于连续小波变换的表层土壤有机碳含量的高光谱估算[J]. *江苏农业学报*, 2023, 39(1): 118-125.
WANG Yuandong, LI Xinguo, YANG Han. Hyperspectral estimation of organic carbon content in surface soils based on continuous wavelet transform[J]. *Jiangsu Journal of Agricultural Sciences*, 2023, 39(1): 118-125. (in Chinese with English abstract)

- [47] 魏令港, 黄靓, 曾令宏. 基于改进特征筛选的随机森林算法对锂渣混凝土强度的预测研究[J]. 材料导报, 2024, 38(9): 160-165.
WEI Linggang, HUANG Liang, ZENG Linghong. Using random forest with improved variable selection to predict the compressive strength of concrete with lithium slag[J]. Materials Reports, 2024, 38(9): 160-165. (in Chinese with English abstract)
- [48] 张磊, 宫兆宁, 王启为, 等. Sentinel-2 影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报, 2019, 23 (2): 313-326.
ZHANG Lei, GONG Zhaoning, WANG Qiwei, et al. Wetland mapping of yellow river delta wetlands based on multi-feature optimization of Sentinel-2 images [J]. National Remote Sensing Bulletin, 2019, 23 (2): 313-326. (in Chinese with English abstract)
- [49] 张佳薇, 支佳豪, 管雪梅, 等. 应用高光谱图像技术对林下作物质量等级鉴别方法——以黄芪为例[J]. 东北林业大学学报, 2024, 52(6): 79-84.
ZHANG Jiawei, ZHI Jihao, GUAN Xuemei, et al. Identification method of understory crops quality grade using hyperspectral image technology: A case of astragalus membranaceus[J]. Journal of Northeast Forestry University, 2024, 52(6): 79-84. (in Chinese with English abstract)
- [50] HONG Y, CHEN S, CHEN Y, et al. Comparing laboratory and airborne hyperspectral data for the estimation and mapping of topsoil organic carbon: Feature selection coupled with random forest[J]. *Soil and Tillage Research*, 2020, 199: 104589.
- [51] 张伏, 王新月, 崔夏华, 等. 高光谱结合极限学习机的玉米品种鉴别[J]. 光谱学与光谱分析, 2023, 43(9): 2928-2934.
ZHANG Fu, WANG Xinyue, CUI Xiahua, et al. Identification of maize varieties by hyperspectral combined with extreme learning machine[J]. Spectroscopy and Spectral Analysis, 2023, 43(9): 2928-2934. (in Chinese with English abstract)
- [52] 吴旭阳, 管港云, 刘志伟, 等. 基于 IPSO-BP 神经网络的 CO₂ 红外吸收光谱定量分析[J/OL]. 光学学报, 1-20[2024-09-29]. <http://kns.cnki.net/kcms/detail/31.1252.o4.20240412.1347.010.html>.
WU Xuyang, GUAN Gangyun, LIU Zhiwei, et al. Quantitative Analysis of CO₂ Infrared Absorption Spectrum Based on IPSO-BP Neural Network[J]. Acta Optica Sinica, 1-20[2024-09-29]. <http://kns.cnki.net/kcms/detail/31.1252.o4.20240412.1347.010.html>. (in Chinese with English abstract)
- [53] 刘爽, 于海业, 陈美辰, 等. 大豆不同水氮胁迫响应规律的高光谱分析[J]. 光谱学与光谱分析, 2020, 40(5): 1575-1580.
LIU Suang, YU Haiye, CHEN Meichen, et al. Study on response of water and nitrogen stresses in soybean based on hyperspectral analysis[J]. Spectroscopy and Spectral Analysis, 2020, 40(5): 1575-1580. (in Chinese with English abstract)
- [54] 刘斌, 李昕. 基于 CNN 和融合目标的三通道小波滤波器组识别[J]. 计算机应用与软件, 2024, 41(2): 209-215, 285.
LIU Bin, LI Xin. The recognition of three-channel wavelet filter banks based on CNN and fusion target[J]. Computer Applications and Software, 2024, 41(2): 209-215, 285. (in Chinese with English abstract)
- [55] 洪子衿, 张艺严, 马静, 等. 拉曼光谱与机器学习结合的微塑料精确识别研究[J/OL]. 中国无机分析化学, 2024, 14 (8): 1047-1057.
HONG Zijin, ZHANG Yiyan, MA Jing, et al. Accurate Microplastic Recognition through Combined Raman Spectroscopy and Machine Learning [J]. Chinese Journal of Inorganic Analytical Chemistry, 2024, 14 (8): 1047-1057. (in Chinese with English abstract)
- [56] 高峰, 邢雅阁, 罗华平, 等. 基于可见/近红外光谱与化学计量学的杏品种无损鉴别方法[J]. 光谱学与光谱分析, 2024, 44(1): 44-51.
GAO Feng, XING Yage, LUO Huaping, et al. Nondestructive identification of apricot varieties based on visible/near infrared spectroscopy and chemometrics methods[J]. Spectroscopy and Spectral Analysis, 2024, 44(1): 44-51. (in Chinese with English abstract)
- [57] 祝志慧, 洪琪, 吴林峰, 等. 基于紫外-可见透射光谱技术和极限学习机的早期鸡胚雌雄识别[J]. 光谱学与光谱分析, 2019, 39(9): 2780-2787.
ZHU Zhihui, HONG Qi, WU Linfeng, et al. Early identification of male and female embryos based on UV/Vis transmission spectroscopy and extreme learning machine[J]. Spectroscopy and Spectral Analysis, 2019, 39(9): 2780-2787. (in Chinese with English abstract)
- [58] 朱晨光, 刘亚军, 李鑫星, 等. 凡纳滨对虾新鲜度的高光谱检测方法[J]. 光谱学与光谱分析, 2023, 43(1): 107-110.
ZHU Chenguang, LIU Yajun, LI Xinxing, et al. Detection method of freshness of penaeus vannamei based on hyperspectral[J]. Spectroscopy and Spectral Analysis, 2023, 43(1): 107-110. (in Chinese with English abstract)
- [59] 刘秀英, 余俊茹, 王世华. 光谱特征变量和 BP 神经网络构建油用牡丹种子含水率估算模型[J]. 农业工程学报, 2020, 36(22): 308-315.
LIU Xiuying, YU Junru, WANG Shihua. Estimation of moisture content in peony seed oil using spectral characteristic variables and BP neural network[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2020, 36(22): 308-315. (in Chinese with English abstract)
- [60] 刘卫东, 唐志鹏, 夏炎, 等. 中国碳强度关键影响因子的机器学习识别及其演进[J]. 地理学报, 2019, 74(12): 2592-2603.
LIU Weidong, TANG Zhipeng, XIA Yan, et al. Identifying the key factors influencing Chinese carbon intensity using machine learning, the random forest algorithm, and evolutionary analysis[J]. Acta Geographica Sinica, 2019, 74(12): 2592-2603.

(in Chinese with English abstract)

[61] 刘磊, 王乐, 张凯南, 等. 新疆东天山黄山地区遥感岩性自动分类[J/OL]. 地质通报, 1-11[2024-09-29]. <http://kns.cnki.net/kcms/detail/11.4648.P.20240412.1453.002.html>.

LIU Lei, WANG Le, ZHANG Kainan, et al. Automatic

lithology classification using remote sensing data in Huangshan area, East Tianshan, Xinjiang[J]. Geological Bulletin of China, 1-11[2024-09-29]. <http://kns.cnki.net/kcms/detail/11.4648.P.20240412.1453.002.html>. (in Chinese with English abstract)

Detection of delinted cotton seed vigor based on hyperspectral wavelet features

DU Wenling¹, GUO Peng^{1,2*}, LIU Xiao¹

(1. College of Science, Shihezi University, Shihezi 832003, China; 2. Key Laboratory of Oasis Town and Basin System Ecological Corps, Shihezi 832003, China)

Abstract: The purpose of this study is to explore the feasibility of using continuous wavelet transform to extract spectral difference information of different vigor desiccated cotton species. A method of filtering wavelet features (WFs) based on correlation and feature importance is proposed to extract the fine structure and complex information of spectral features of seeds with different vigor. Different vigor classes of desiccated cotton seeds were obtained through artificial aging experiments. And its high spectral image was collected. The raw spectra are preprocessed with Savitzky-Golay smoothing, multivariate scattering correction, first-order differentiation, and second-order differentiation. Then, the WFs extracted by wavelet basis functions such as gauss4, mexh and bior6.8 were compared. Spectral wavelet features (SFs) and WFs were downsampled using principal component analysis. Based on machine learning algorithms such as support vector machines (SVM), random forest (RF), extreme learning machines (ELM), and back propagation neural network (BPNN), a seed vigor detection model was developed for SFs principal components and WFs principal components. The accuracy of the seed vigor detection model was compared between SFs principal components and WFs principal components. The fine spectral information in WFs was further extracted based on correlation analysis and random forest feature importance evaluation. Including the 1% |R| -WFs feature set with the correlation with seed vigor at the top 1%, the 1% Importance-WFs feature set with the feature importance at the top 1% in seed vigor recognition, and the 1% |R| + 1% Importance-WFs feature set with the combination of the two, and bring these three WFs feature sets into the above machine learning model. The results showed that: 1) The bior6.8 function extracted better WFs for different vigor desiccated cotton species. Other wavelet basis functions show a clear ringing effect when extracting WFs. 2) The modeling accuracy of the WFs principal components is higher than that of the SFs principal components in all machine learning models for each species. The model based on 1% |R| + 1% Importance-WFs has the highest accuracy. 3) The optimal models for seed vigor detection of Jinke 21 and Jinke 20 were: 1% |R| + 1% Importance-WFs + ELM. The optimal model of Xinluzao 64 seed vigor detection is: 1% |R| + 1% Importance-WFs + any machine model and PCA-WFs + ELM / BPNN. The accuracies of the training set and test set of the optimal model of Jinke 21 are 99.63% and 98.28%, and the accuracies of the training set and test set of the optimal model of Jinke 20 and Xinluzao 64 are both 100%. The results indicate that the method proposed in this paper based on correlation and feature importance can effectively extract spectral difference information of different vitality dried cottonseeds, providing a new spectral characterization approach for seed vitality hyperspectral detection.

Keywords: CWT; correlation; features importance; hyperspectral technology; seed vigor detection