

基于高光谱深度特征的油菜叶片锌含量检测

周鑫^{1,4,5}, 王坚¹, 赵春江^{1,2,3}, 孙俊¹, 史磊¹

(1. 江苏大学电气信息工程学院, 镇江 212013; 2. 北京市农林科学院信息技术研究中心, 北京 100097;
3. 国家农业信息化工程技术研究中心, 北京 100097; 4. 江苏大学智能农机装备理论与技术重点实验室, 镇江 212013;
5. 省部共建现代农业装备与技术协同创新中心, 镇江 212013)

摘要: 为了实现油菜叶片锌含量的快速无损检测, 该研究采用一种基于高光谱成像技术结合深度迁移学习算法的高精度检测方法, 通过无土栽培的方式, 设置 10 个不同胁迫类别 (2 种不同硅浓度环境结合 5 个不同锌胁迫梯度), 获取无硅环境和有硅环境中重金属锌胁迫下总计 4 000 个油菜叶片样本。利用高光谱成像设备采集油菜叶片样本高光谱图像信息, 并将整个叶片作为感兴趣区域获取其平均光谱信息。通过对比不同预处理后光谱对硅作用下油菜叶片锌含量预测性能, 确立标准正态变量变换 (standard normalized variable, SNV) 算法作为最佳预处理方法, 并对 SNV 处理的光谱数据进行进一步分析。利用堆叠自编码器 (stacked auto-encoder, SAE) 对预处理后的最佳光谱数据进行降维, 并与传统的降维算法进行比较。最后, 对最优 SAE 深度学习网络进行迁移学习, 得到迁移堆叠自编码器 (transfer stacked auto-encoder, T-SAE) 模型, 验证无硅环境和有硅环境中深度学习模型之间的可迁移性。结果表明, 基于 SAE 提取深度特征的支持向量机回归 (support vector machine regression, SVR) 模型对无硅环境或有硅环境中油菜叶片中锌含量的预测效果较好。无硅环境和有硅环境中所建立的 SNV-SAE-SVR 模型性能较佳, 预测集的决定系数 (R_p^2)、均方根误差 (RMSEP) 和相对分析误差 (RPD) 分别为 0.850 7、0.034 66 mg/kg 和 2.607, 0.876 6、0.028 54 mg/kg 和 2.732。此外, 基于 T-SAE 提取深度特征的 SVR 模型能有效实现无硅环境和有硅环境中锌含量的预测, 最佳 SNV-T-SAE-SVR 模型预测集的 R_p^2 、RMSEP 和 RPD 分别为 0.881 0、0.027 48 mg/kg 和 2.966。研究结果表明, 深度迁移学习方法结合高光谱成像无损检测技术能够有效实现油菜叶片锌含量检测。

关键词: 无损检测; 重金属; 硅环境; 深度学习; 迁移学习; 高光谱图像

doi: 10.11975/j.issn.1002-6819.202404222

中图分类号: S126

文献标志码: A

文章编号: 1002-6819(2024)-19-0262-10

周鑫, 王坚, 赵春江, 等. 基于高光谱深度特征的油菜叶片锌含量检测[J]. 农业工程学报, 2024, 40(19): 262-271. doi:

10.11975/j.issn.1002-6819.202404222 <http://www.tcsae.org>

ZHOU Xin, WANG Jian, ZHAO Chunjiang, et al. Non-destructive determination of zinc content in oilseed rape leaves based on hyperspectral depth characteristics[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(19): 262-271. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202404222 <http://www.tcsae.org>

0 引言

近年来, 农业土壤重金属污染已成为一个亟待解决的严重环境问题^[1]。锌 (Zinc, Zn) 不但是农作物必需的微量元素, 而且是危害人类健康的重金属。当土壤中的锌含量超过一定限度时, 将严重威胁着农业生产的安全^[2]。植物修复是一种利用植物从土壤中提取、运输和固定重金属以恢复其完整性的技术^[3]。现有研究表明, 油菜植株能有效吸收土壤环境中的锌, 适当的锌 (一般作物叶片中锌的浓度在 10~100 mg/kg) 可以促进作物生长, 而过量的锌 (一般作物叶片中锌的浓度大于 100 mg/kg) 会影响作物的生理活性、生长速度和细胞结构^[4]。此外,

硅在重金属污染土壤的修复中也得到了广泛的应用。多项研究表明, 在低硅浓度的环境下, 作物对重金属的吸收会受到明显的阻碍, 导致其对重金属的积累能力降低; 相反, 当环境中的硅浓度较高时, 作物对重金属的吸收能力则会显著增强, 从而可能增加作物中重金属的含量^[5-7]。以中国自然环境为例, 土壤中有效硅质量分数在 2.1~681.4 mg/kg 之间^[8]。目前, 国内外农作物重金属锌检测方法主要包括火焰原子吸收光谱法^[9]、电感耦合等离子体发射光谱法^[10]和电感耦合等离子体质谱法^[11]等, 这些化学检测方法虽然能够实现重金属锌的痕量检测, 但是操作要求高、成本大且对待测样具有破坏性。因此, 寻求一种快速、有效、无损的无硅环境和有硅环境中油菜叶片锌含量检测技术和方法, 有利于监测作物逆境胁迫和修复农业土壤重金属。

可见光-近红外高光谱成像技术 (visible near-infrared hyperspectral imaging, Vis-NIR HSI) 是一种前沿的农作物信息检测技术, 集光谱信息和图像信息于一体, 能够实现农作物信息的无损检测^[4,9]。目前, 该技术在重金属

收稿日期: 2024-04-30 修订日期: 2024-07-01

基金项目: 国家自然科学基金青年基金项目 (32201653); 中国博士后科学基金项目 (2021M701479); 江苏大学农业工程学部项目 (NGXB20240202); 江苏大学大学生科研立项项目 (Y23A093)

作者简介: 周鑫, 博士, 博士后, 副教授, 硕士生导师, 研究方向为农作物信息的无损检测。Email: zhouxin_21@ujs.edu.cn

胁迫下作物信息分析方面已取得一定进展。黄富荣等采用 Vis-NIR 高光谱成像技术对菜心铜胁迫梯度进行了分类检测研究, 所建立的偏最小二乘判别分析 (partial least squares discriminant analysis, PLS-DA) 模型对铜胁迫梯度类别检测准确率达到 95%^[12]。CAO 等采用 Vis-NIR 高光谱成像技术对油菜叶片铅含量检测, 所建立的哈里斯鹰优化算法优化支持向量机回归 (harris hawks optimization support vector regression, HHO-SVR) 模型能够实现铅含量的高精度检测^[13]。FENG 等利用 Vis-NIR 高光谱图像技术结合化学计量学方法对芒草中镉浓度进行测定, 研究发现采用竞争性自适应重加权算法 (competitive adaptive reweighed sampling, CARS) 结合偏最小二乘回归模型 (PLSR) 能准确检测芒草中的镉含量^[14]。YU 等采用 Vis-NIR 高光谱成像系统结合机器学习算法对烟草汞胁迫进行了分类研究, 结果表明 Vis-NIR HSI 能够有效实现烟草汞胁迫的定性判别分析^[15]。此外, 研究学者们将深度学习 (deep learning, DL) 方法引入到高光谱图像信息处理中, 并取得了令人瞩目的成果^[16-18]。针对不同作物信息检测时, 传统的深度学习方法需要从头开始学习每个任务, 运算周期长且效率低, 所建立的检测模型欠缺在新目标检测上的泛化能力。迁移学习 (transfer learning, TL) 可以利用任务之间的相似性, 对一个或多个源任务模型进行参数微调, 将源任务中学到的有用信息迁移到目标任务中, 从而加速学习过程并提升模型的泛化能力^[19]。堆叠自编码器 (stacked auto-encoder, SAE) 是一种深度学习算法, 它通过组合多个自编码器层来构建深度神经网络模型, 从而提取更高级、更抽象的特征^[20]。而迁移堆叠自编码器 (transfer stacked auto-encoder, T-SAE) 是在 SAE 算法的基础上, 引入迁移学习方法理念, 利用源域任务中已构建的模型来改进目标域任务的学习效果^[21]。因此, 本文采用 T-SAE 算法对高光谱深度特征模型进行迁移学习, 目前将该方法应用于无硅环境和有硅环境中油菜叶片锌含量无损检测研究中还未见报道。

本文采用高光谱图像无损检测技术, 结合化学计量学方法和深度迁移学习方法, 针对无硅环境和有硅环境中油菜叶片重金属锌含量开展定量检测研究, 以期论证基于高光谱图像无损检测技术实现无硅环境和有硅环境中油菜叶片锌含量测定的可行性和利用深度迁移学习模型提高无硅环境和有硅环境中油菜叶片锌含量检测精度的有效性。

1 材料与方法

1.1 样本制备

试验品种为秦油 10 号油菜, 采用珍珠岩袋培方式进行油菜样本培育。样本培育地点在江苏大学校本部 Venlo 型温室中进行。将长势形状相近的油菜分为 2 组 (Z 组和 ZS 组), 每组有 5 个胁迫试剂类别, 每个胁迫试剂类别选取油菜中间叶位叶片样本 400 个, 每组采集油菜叶片样本 2 000 个, 总计 4 000 个油菜叶片样本。其中,

Z 组为无硅环境中 Zn 胁迫下油菜样本分组, 包括 Z1、Z2、Z3、Z4 和 Z5 分别表示 Si 环境浓度均为 0, Zn 胁迫浓度为 100、200、300、400 和 500 mg/L, 以 $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 形式进行试剂配制; ZS 组为有硅环境中 Zn 胁迫下油菜样本分组, 包括 ZS1、ZS2、ZS3、ZS4 和 ZS5 分别表示 Si 环境浓度均为 600 mg/L, 而 Zn 胁迫浓度为 100、200、300、400 和 500 mg/L, 以 $\text{Na}_2\text{SiO}_3 \cdot 9\text{H}_2\text{O}$ 和 $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 形式进行试剂配制。油菜植株生长期间浇灌的营养液为霍格兰营养液。每个类别胁迫试剂配制量为 600 mL, 每天于 7:00、12:00 和 18:00 三个时段进行胁迫试剂浇灌, 每次 20 mL, 连续浇灌 10 d。油菜叶片样本采集是在胁迫试剂浇灌完成 7 d 后进行, 油菜叶片样本采集完成后做好标签, 立刻送往实验室进行样本高光谱图像信息采集。

1.2 Vis-NIR 高光谱图像信息采集

Vis-NIR 高光谱图像信息采集系统主要由高光谱成像仪 (ImSpector V10E, Spectral Imaging Ltd., Oulu, 芬兰)、CCD 相机 (Zyla 4.2 Plus, Andor Technology, Inc., 英国)、2 个 150W 光纤卤素灯、电动位移平台 (SC30021A, Zolix, 北京) 和暗箱等组成。该系统能够有效获取 431.05~962.45 nm 波段范围内的 618 个光谱通道信息, 光谱分辨率为 2.8 nm。CCD 相机对样本的曝光时间设置为 50 ms, 电动位移平台速度为 1.26 mm/s。为了校正由于相机的暗电流效应对图像的影响, 通过用盖子覆盖镜头和使用标定用的白板 (反射率接近 100%) 分别获得全黑标定图像 R_b 和全白标定图像 R_w , 再将油菜叶片放在样本台上, 扫描获取原始高光谱图像 R_0 , 最后按式 (1) 进行校正, 从而得到校正图像 R 。

$$R = \frac{R_0 - R_b}{R_w - R_b} \quad (1)$$

采用 ENVI 4.5 软件环境对 Vis-NIR 高光谱成像信息进行处理, 并将整个油菜叶片样本作为感兴趣区域 (region of interest, ROI)。Vis-NIR 高光谱 ROI 数据采集过程如图 1 所示, 其主要步骤如下: 首先, 通过对比背景和样本区域的光谱, 得到两个明显不同的光谱波段 (543.16 nm 和 673.25 nm), 并利用两者之间的比值变换得到比值图像; 其次, 利用阈值分割法对比值图像进行处理, 获得二值化掩模图像, 最小阈值设置为 1.5。最后, 将二值化掩模图像应用在归一化处理后的油菜叶片的高光谱图像上, 得到掩模后油菜叶片高光谱图像信息。通过计算得到 ROI 内所有像元的平均光谱信息, 并将其作为输入进行进一步处理。

1.3 重金属锌含量的测定

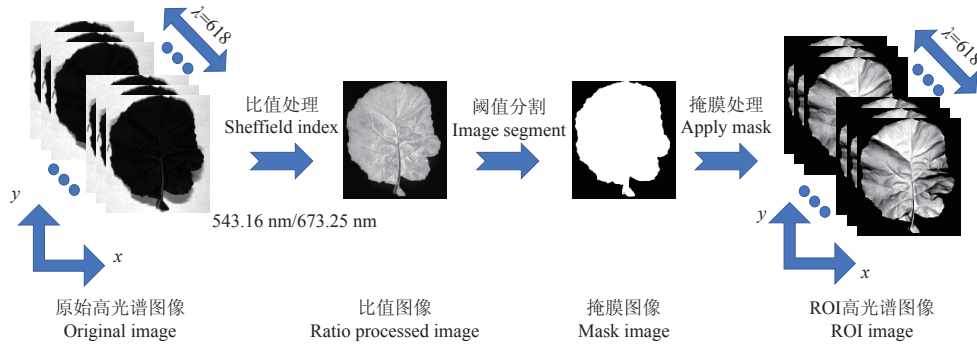
Vis-NIR 高光谱图像信息采集完成后, 采用去离子水对油菜叶片进行 3 次清洗。将油菜叶片样本在 120 °C 干燥温度下干燥至恒量后研磨成粉末, 每个油菜叶片样本粉末称取量为 0.01 g 用于火焰原子吸收光谱法 (flame atomic absorption spectrometry, FAAS) 测定 Zn 含量。火焰原子吸收光谱法 (FAAS) 测定 Zn 含量对照样品粉末称取的量为 0.05 g。根据国家标准《GB 509.14-2017》^[22]

中 FAAS 法测定 Zn 含量, 油菜叶片样本粉末和对照样品粉末先经微波消解处理, 对照样品定容后通过火焰原子吸收光谱仪 (ZEEnit®700P) 完成重金属 Zn 标准曲线的制作, 通过自动注入油菜叶片消解定溶液进行样本 Zn 含量检测。必要时需要进一步稀释油菜叶片消解定溶液, 完成重金属 Zn 含量的检测。其中, 火焰原子吸收光谱法仪器性能指标设置为: 光路模式选择高光通量的单光束, 波长为 213.9 nm、光谱带宽 0.2 nm、灯电流 3 mA、燃烧

头高度 3 mm、空气流量 9 L/min 和乙炔流量 2 L/min。 C_{Zn} 含量 (mg/kg) 的计算公式如下:

$$C_{Zn} = \frac{(d_1 - d_0) \cdot V}{1000 m} \quad (2)$$

式中 d_1 表示待测样消化液中 Zn 的含量, ng/mL。 d_0 表示空白溶液中 Zn 的含量, ng/mL。 V 表示消化液的总体积, mL。 m 表示干燥前样品的质量, g。1000 表示换算系数。



注: λ 为光谱波长数。

Note: λ is the number of spectral wavelengths.

图 1 Vis-NIR 高光谱 ROI 数据采集过程

Fig. 1 Vis-NIR hyperspectral ROI data acquisition processed

1.4 光谱信息处理算法

为了实现光谱数据的有效预处理, 本文采用 Savitzky-Golay (SG) 平滑算法^[23]、标准正态变量变换 (standard normalized variable, SNV)^[24]、一阶微分 (first derivative, FD) 和二阶微分 (second derivative, SD) 算法^[25] 对 ROI 光谱进行处理。

采用主成分分析法 (principal component analysis, PCA)^[26]、连续投影算法 (successive projections algorithm, SPA)^[27]、迭代保留信息变量法 (iteratively retains informative variables, IRIV)^[28] 和变量空间迭代收缩法 (variable iterative space shrinkage approach, VISSA)^[29] 对最佳预处理后光谱数据进行数据降维。利用堆叠自编码器 (stacked auto-encoder, SAE) 和迁移堆叠自编码器 (transfer stacked auto-encoder, T-SAE) 提取最佳预处理后光谱数据的深度特征。

1.5 SAE 算法

自动编码器 (auto encoder, AE) 主要由三层网络结构组成包括输入层、隐含层和输出层^[30]。编码过程 (Encoding) 为输入层至隐含层之间的映射关系, 解码过程 (Decoding) 为隐含层至输出层之间的映射关系。堆叠自编码器 (Stacked auto-encoder, SAE) 是由简单 AE 结构叠加而来的深层神经网络, 即 SAE 是由多个 AE 堆叠级联形成。SAE 采用无监督贪婪逐层预训练的方式, 再将网络展开并通过有监督的微调整整个网络结构^[31]。第一个 AE 的隐含层的输出直接作为第二个 AE 的隐含层输入, 依此类推, 将最后一个 AE 的隐含层的输出作为 SAE 提取得到的特征信息。

1.6 T-SAE 算法

迁移堆叠自编码器 (transfer stacked auto-encoder, T-

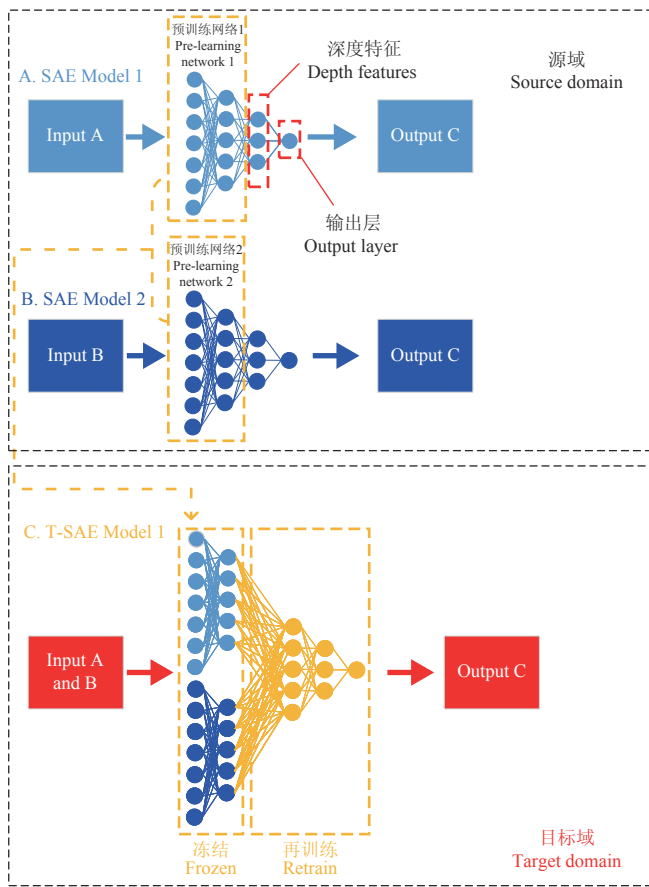
SAE) 包含两种迁移方式单模型迁移堆叠自编码器 (single-model T-SAE) 和双模型迁移堆叠自编码器 (dual-model T-SAE), 本文采用双模型迁移堆叠自编码器 (dual-model T-SAE) 的方式^[21]。在构建好的 A 和 C、B 和 C 的深度网络关系模型 (源域) 基础上 (图 2), 即源域中已构建好的 A 和 C 的深度网络关系模型为 SAE Model 1, B 和 C 的深度网络关系模型为 SAE Model 2, 基于迁移学习的 T-SAE 算法用于建立 A、B 和 C 三者之间的深层网络关系模型 (目标域)。本文中 A 为无硅环境中油菜叶片的 Vis-NIR 高光谱数据, B 为有硅环境中油菜叶片的 Vis-NIR 高光谱数据, C 为油菜叶片中 Zn 的含量。

对 SAE Model 1 和 SAE Model 2 模型网络层数是否相等进行判定, 如果相同, 直接进行 T-SAE 迁移学习; 如果不相同, 则需要对 T-SAE 迁移网络进行平移和扩展。假设 SAE Model 1 网络层数为 m , SAE Model 2 网络层数为 n , $m < n$ 。为了保持 T-SAE 迁移学习过程中 SAE Model 1 和 SAE Model 2 网络层数相同, 需要保持 SAE Model 1 网络的深度特征层参数不变的前提下, 将 SAE Model 1 网络层数从 m 层扩展到 n 层。其核心计算方法为: 以 SAE Model 1 网络从第 k 层扩展到第 $k+1$ 层为例, 第 $k+1$ 层第 i 节点 Y_i 的计算公式如下^[32]:

$$Y_i = \sum_{p=1}^l S_f (\omega_{ij} X_p + b_{ij}) \quad (3)$$

式中 X_p 为 SAE Model 1 网络第 k 层第 p 节点的值, ω_{ij} 为 SAE Model 1 网络第 k 层第 i 节点到第 $k+1$ 层第 j 节点的权重系数, b_{ij} 为 SAE Model 1 网络第 k 层第 i 节点到第 $k+1$ 层的第 j 个节点的偏置, S_f 为 SAE Model 1 网络的激活函数。 $i=1,2,3,\dots,l$; $j=1,2,3,\dots,l$ 。 p 为 SAE Model 1

网络中深度特征层的节点数。



注：A 为无硅环境中油菜叶片的 Vis-NIR 高光谱数据，B 为有硅环境中油菜叶片的 Vis-NIR 高光谱数据，C 为油菜叶片中 Zn 的含量。
Note: A is the Vis-NIR hyperspectral data of oilseed rape leaves in the silicon-free environment, B is the Vis-NIR hyperspectral data of oilseed rape leaves in the silicon environment, and C is the content of Zn in oilseed rape leaves.

图 2 T-SAE 深度迁移学习模型构建

Fig.2 T-SAE deep transfer learning model construction

为保证 SAE Model 1 网络深度特征层参数不变， ω_{ij} 取值和 b_{ij} 取值如下：

$$\omega_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \quad (4)$$

$$b_{ij} = 0 \quad (5)$$

以 SAE Model 1 网络从第 k 层扩展到第 $k+1$ 层的方法为参考，一直持续到将 SAE Model 1 网络层数从 m 层扩展到 n 层。完成 SAE Model 1 网络层数扩展后，进行目标域的 T-SAE 深度迁移学习模型构建 (T-SAE Model 1)，流程如图 2 所示。其中，T-SAE Model 1 模型迁移的方式为源域模型 (SAE Model 1 和 SAE Model 2 模型) 中预训练网络层的保存冻结状态，保留主模型中预训练网络层的权重，并使用来自目标域数据对 T-SAE Model 1 模型重新训练并实现参数微调，进而获得新的深度特征层。

1.7 SVR 算法

本文基于预处理前后光谱数据、最佳预处理后特征光谱数据和最佳预处理后深度特征数据分别建立不同支持向量机回归 (support vector machine regression, SVR) 模型^[33]，SVR 算法选用的核函数为 RBF 核函数。其中，

2 000 个无硅环境中采集的油菜叶片 Vis-NIR 光谱数据放在数据集 Z 中，2 000 个有硅环境中采集的油菜叶片 Vis-NIR 光谱数据放在数据集 ZS 中，4 000 个无硅环境和有硅环境中采集的油菜叶片 Vis-NIR 光谱数据放在数据集 ZZS 中。将三组数据按照按 3:1 的比例分成训练集和预测集，无硅环境中数据集 Z 和有硅环境中数据集 ZS，训练集和预测集样本数划分均分别为 1 500 个和 500 个。无硅环境和有硅环境中数据集 ZZS，训练集和预测集样本数分别为 3 000 个和 1 000 个。其中，为了保证所建立的模型具有较强的泛化能力，数据集 ZZS 中划分训练集样本均来自于数据集 Z 和数据集 ZS 的训练集样本，数据集 ZZS 剩余数据将被放置于预测集中。

2 结果与分析

2.1 油菜叶片不同锌浓度检测及平均光谱分析

无硅环境和有硅环境中不同 Zn 浓度胁迫下油菜叶片平均 Zn 含量试验结果如图 3 所示。从图 3 中可以看出，在无硅环境和有硅环境中随着 Zn 胁迫浓度的增加，油菜叶片中平均 Zn 含量均上升。在相同的铅胁迫浓度下，高浓度硅环境中油菜叶片中 Zn 含量均高于无硅环境中油菜叶片中 Zn 含量。当 Zn 胁迫浓度分别为 100、200、300、400 和 500 mg/L 时，有硅环境和无硅环境中油菜叶片实际平均 Zn 含量之比分别为 1.34、2.53、3.10、2.67 和 3.31。从结果可以看出，有硅环境和无硅环境中油菜叶片实际平均 Zn 含量之比变化总体上呈现先减少后增加再减少又增加的变化趋势，这一变化趋势与重金属铅胁迫下有硅和无硅环境中油菜叶片铅含量检测的相关研究结果相一致^[32]。

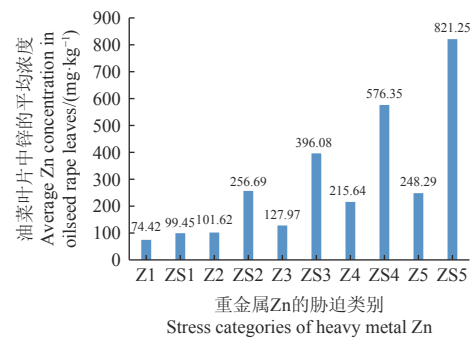


图 3 无硅环境和有硅环境中 Zn 胁迫下油菜叶片的平均 Zn 含量

Fig.3 Average Zn content in oilseed rape leaves under Zn stress in both silicon-free and silicon environments

不同 Zn 胁迫浓度下无硅环境和有硅环境中油菜叶片的平均光谱曲线如图 4 所示。从图 4 中可以看出，所有的 Vis-NIR 平均光谱都呈现出相同的变化趋势，相同 Zn 胁迫浓度下无硅环境和有硅环境中油菜叶片的 Vis-NIR 光谱反射率存在差异。对比分析图 4 可以看出，不同 Zn 胁迫浓度下无硅环境和有硅环境中油菜叶片的平均光谱曲线存在差异。因此，利用 Vis-NIR 光谱能够有效实现无硅环境和有硅环境中油菜叶片 Zn 含量的测定。

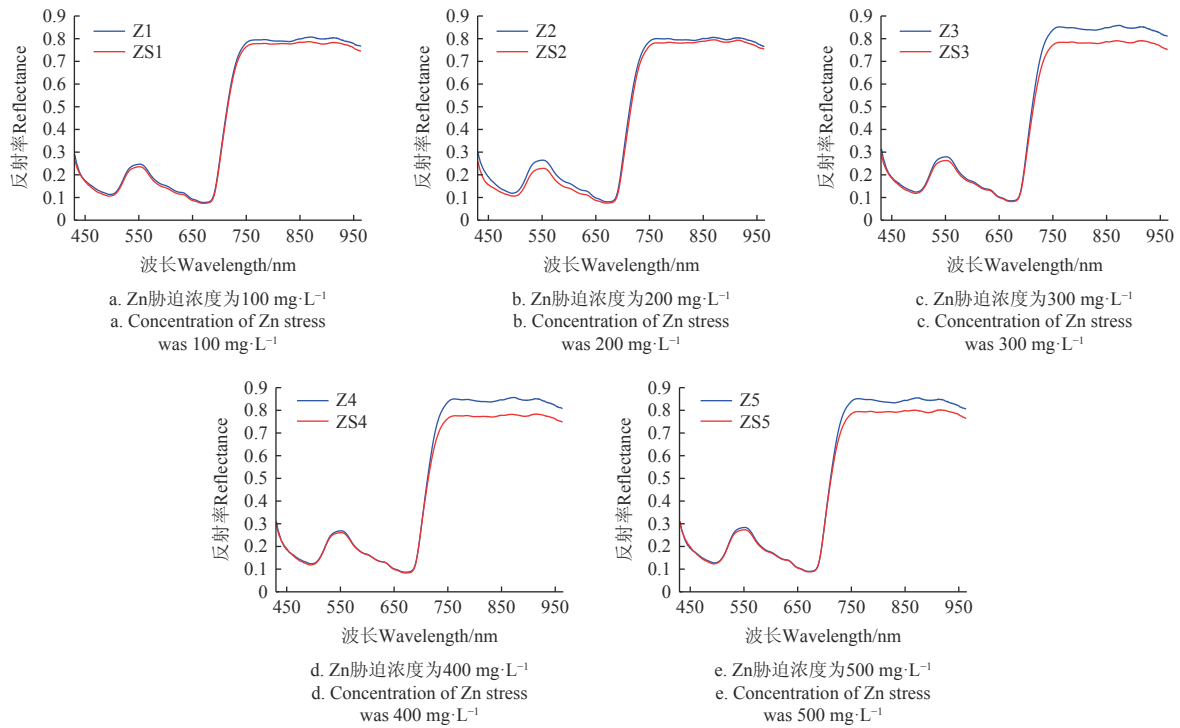


图4 不同 Zn 胁迫浓度下无硅环境和有硅环境中油菜叶片的平均光谱曲线

Fig.4 Average spectral curves of oilseed rape leaves in silicon-free and silicon environments under different Zn stress concentrations

2.2 Vis-NIR 光谱数据预处理

不同预处理后 Vis-NIR 光谱数据如图5所示,从图5可以看出不同预处理后光谱曲线存在显著差异性。不同预处理后 Vis-NIR 光谱数据建立的 SVR 模型结果如表1所示。从表1可以看出,基于不同预处理方法的 Vis-NIR 光谱数据所建立的 Zn 浓度预测 SVR 模型的性能存在差异。其中,采用 SNV 预处理后的无硅环境中 Vis-NIR 光谱数据所构建的 SVR 模型性能较佳,预测集的决定系数 (R_p^2)、

均方根误差 (RMSEP) 和相对分析误差 (RPD) 分别为 0.7657、0.041 19 mg/kg 和 2.309。采用 SNV 预处理后的有硅环境中 Vis-NIR 光谱数据所构建的 SVR 模型表现较佳,预测集的 R_p^2 为 0.851 8, RMSEP 为 0.036 41 mg/kg, RPD 为 2.649。利用 SNV 处理后的无硅和有硅环境中 Vis-NIR 光谱数据所构建的 SVR 模型性能仍为较佳, R_p^2 为 0.878 1, RMSEP 为 0.028 86 mg/kg, RPD 为 2.740。因此,本文将对 SNV 处理后的 Vis-NIR 光谱数据进行进一步分析处理。

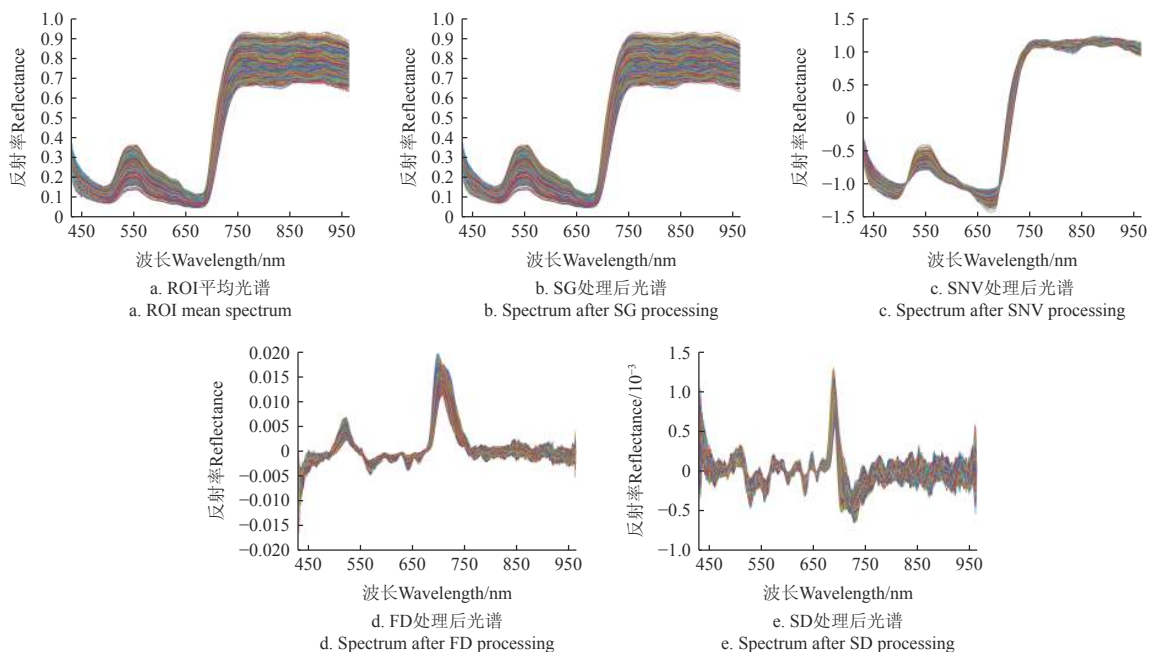


图5 预处理前后油菜叶片光谱曲线

Fig.5 Spectral curves of rapeseed leaves before and after pretreatment

表 1 基于不同预处理后 Vis-NIR 光谱数据的 SVR 模型结果
Table 1 Results of SVR model based on different pre-processed Vis-NIR spectral datas

数据集 Data set	样本数 Sample number	预处理 Pretreatment	训练集 Calibration set		预测集 Prediction set		相对分析误差 RPD
			训练集决定系数 R_c^2	训练集均方根误差 RMSEC/(mg·kg ⁻¹)	预测集决定系数 R_p^2	预测集均方根误差 RMSEP/(mg·kg ⁻¹)	
Z	2 000	Raw	0.792 9	0.039 33	0.638 3	0.051 01	1.132
		SG	0.771 7	0.041 11	0.630 1	0.051 54	1.157
		SNV	0.886 9	0.029 36	0.765 7	0.041 19	2.309
		1 st Der	0.817 8	0.037 09	0.732 2	0.044 24	2.088
		2 nd Der	0.847 3	0.033 90	0.744 7	0.042 99	2.239
ZS	2 000	Raw	0.832 3	0.035 16	0.659 3	0.058 26	1.411
		SG	0.810 5	0.039 32	0.653 4	0.059 21	1.390
		SNV	0.966 2	0.016 79	0.851 8	0.036 41	2.649
		1 st Der	0.862 6	0.035 08	0.809 6	0.038 13	2.491
		2 nd Der	0.898 6	0.024 97	0.831 3	0.037 63	2.526
ZZS	4 000	Raw	0.850 8	0.037 80	0.749 5	0.042 49	2.244
		SG	0.688 8	0.056 17	0.642 7	0.064 07	1.299
		SNV	0.986 4	0.026 52	0.878 1	0.028 86	2.740
		1 st Der	0.913 8	0.030 88	0.807 8	0.036 43	2.485
		2 nd Der	0.914 5	0.030 22	0.857 1	0.035 83	2.687

注：Z、ZS 和 ZZS 分别表示无硅环境中油菜叶片 VIS-NIR 光谱数据、有硅环境中油菜叶片 VIS-NIR 光谱数据和无硅和有硅环境中油菜叶片 VIS-NIR 光谱数据，下同。Note: Z, ZS and ZZS represent VIS-NIR spectral data of oilseed rape leaves in the silicon-free environment, VIS-NIR spectral data of oilseed rape leaves in the silicon environment and VIS-NIR spectral data of oilseed rape leaves in the silicon-free and silicon environment, respectively. The same below.

2.3 数据降维

本文采用 PCA、SPA、IRIV 和 VISSA4 种降维算法对 SNV 预处理后的 VIS-NIR 光谱数据进行降维处理，原始数据集包括数据集 Z、ZS 和 ZZS。PCA 对数据集 Z、ZS 和 ZZS 降维后不同主成分数下的累积贡献率如图 6 所示。由图 6 可知，PCA 选择的无硅环境中数据集 Z、有硅环境中数据集 ZS 和无硅和有硅环境中数据集 ZZS 的主成分数分别为 5、15 和 11，累积贡献率分别为 97.11%、99.68% 和 98.12%。

3 种降维算法对数据集 Z、ZS 和 ZZS 降维后提取的特征分布如图 7 所示。从图 7 中可以看出，无硅环境中经过 SNV 处理后，3 种特征提取算法提取的 Zn 相关特征光谱数分别为 15、66 和 178。有硅环境中经过 SNV 处理后，3 种特征提取算法提取的 Zn 相关特征光谱数分别为 22、63 和 169。无硅环境和有硅环境中经过 SNV 处理后，3 种特征提取算法提取的 Zn 相关特征光谱数分别为 53、113 和 189。三种特征提取算法得到的 Vis-NIR 光谱特征主要分布在以 450、550、670、760、910 和 960 nm 为中心波长附近。其中，450、550 和 670 nm

左右的可见光谱区域出现的振动主要是由作物叶片中色素（叶绿素、类胡萝卜素、花青素和叶黄素等）所引起的^[34]，760 nm 中心波长附近为与水有关的 O-H 官能团的第三倍频的延伸^[35]，而 910 和 960 nm 为中心波长附近分别对应 C-H 和 O-H 官能团的第三倍频和第二倍频^[21]。

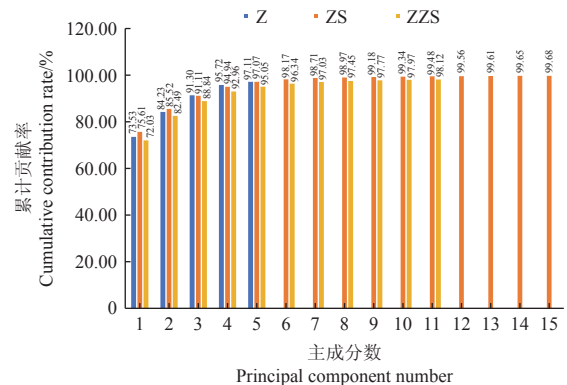


图 6 PCA 处理后不同主成分数下的累积贡献率
Fig.6 Cumulative contribution rate under different principal component numbers after PCA treatment

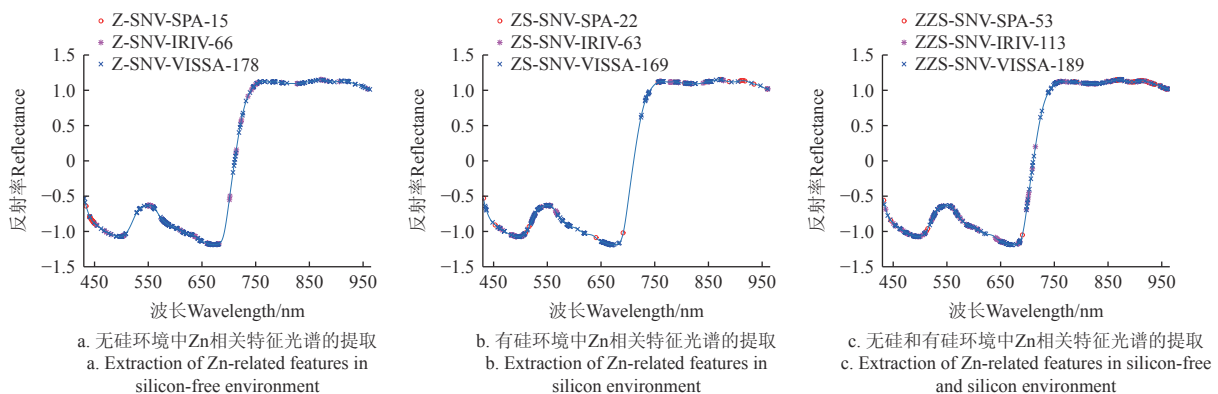


图 7 三种特征提取算法得到不同环境中 Zn 相关特征数据
Fig.7 Zn correlation feature extracted by three feature extraction algorithms in different environments

2.4 数学建模

4 种数学降维处理后数据建立的 SVR 模型结果如表 2 所示。从表 2 可以看出, 基于不同数据降维方法处理后的 Vis-NIR 光谱数据所建立的 Zn 浓度预测 SVR 模型的性能存在差异。其中, 采用 SNV-SPA 处理后的无硅环境中 Vis-NIR 光谱数据所构建的 SVR 模型性能最佳, 预测集的 R_p^2 、RMSEP 和 RPD 分别为 0.781 8、0.040 37 mg/kg 和 2.417。采用 SNV-PCA 处理后的有硅环境中

Vis-NIR 光谱数据所构建的 SVR 模型表现最佳, 预测集的 R_p^2 为 0.859 3, RMSEP 为 0.034 91 mg/kg, RPD 为 2.655。利用 SNV-VISSA 处理后的无硅和有硅环境中 Vis-NIR 光谱数据所构建的 SVR 模型性能仍为最佳, R_p^2 为 0.856 8, RMSEP 为 0.043 61 mg/kg, RPD 为 2.652。基于特征数据的 SVR 模型对无硅和有硅环境中油菜叶片 Zn 含量的预测性能有待进一步提升。因此, 有必要探索替代算法来提高模型的有效性。

表 2 基于特征光谱数据的 SVR 模型结果

Table 2 SVR model results based on spectral data after dimensionality reduction

数据集 Data set	样本数 Sample number	处理算法 Pretreatment	训练集 Calibration set		预测集 Prediction set		
			训练集决定系数 R_c^2	训练集均方根误差 RMSEC/(mg·kg ⁻¹)	预测集决定系数 R_p^2	预测集均方根误差 RMSEP/(mg·kg ⁻¹)	相对分析误差 RPD
Z	2000	SNV-PCA	0.823 8	0.035 69	0.755 5	0.042 61	2.263
		SNV-SPA	0.907 4	0.026 44	0.781 8	0.040 37	2.417
		SNV-IRIV	0.862 0	0.039 07	0.728 4	0.044 82	2.060
		SNV-VISSA	0.879 1	0.038 92	0.770 7	0.040 84	2.359
ZS	2000	SNV-PCA	0.929 8	0.016 75	0.859 3	0.034 91	2.655
		SNV-SPA	0.898 1	0.023 21	0.828 7	0.038 12	2.511
		SNV-IRIV	0.924 1	0.016 98	0.812 3	0.038 96	2.494
		SNV-VISSA	0.900 1	0.022 75	0.848 5	0.035 04	2.592
ZZS	4 000	SNV-PCA	0.887 2	0.033 14	0.853 8	0.043 83	2.650
		SNV-SPA	0.960 8	0.013 83	0.838 6	0.049 03	2.535
		SNV-IRIV	0.963 8	0.013 57	0.842 6	0.047 92	2.588
		SNV-VISSA	0.974 4	0.012 35	0.856 8	0.043 61	2.652

本文采用 SAE 算法对 SNV 算法处理后的光谱数据进行深度特征的获取, 基于 SAE 提取的深度特征所建立的 SVR 模型结果如表 3 所示。从表 3 可以看出, 对于单一环境(无硅环境或有硅环境中)光谱数据, 基于 SAE 提取的深度特征的 SVR 模型性能得到了进一步的增强。其中, 对于数据集 Z(无硅环境中油菜叶片样本), 所建立的 SNV-SAE-SVR 模型表现最佳, 预测集的 R_p^2 、RMSEP 和 RPD 分别为 0.850 7、0.034 66 mg/kg 和 2.607, 最佳模型尺度为 618-481-381。模型尺度表示 SAE 每层神经元的数量。例如 [618-481-381] 表示输入层共有 618 个神经元(即原始输入波长数为 618 个), 网络中包含两个隐藏层的神经元数量分别为 481 个和 381 个。最后一个

隐含层(91 个节点)作为 SAE 网络的输出(即 SVR 模型的输入)。对于数据集 ZS(有硅环境中油菜叶片样本), 所建立的 SNV-SAE-SVR 模型表现最佳, 预测集的 R_p^2 、RMSEP 和 RPD 分别为 0.876 6、0.028 54 mg/kg 和 2.732, 最佳模型尺度为 618-531-441。而对于数据集 ZZS(无硅环境和有硅环境中油菜叶片样本), 所建立的 SNV-SAE-SVR 模型表现最佳, 预测集的 R_p^2 、RMSEP 和 RPD 分别为 0.839 4、0.036 35 mg/kg 和 2.549, 最佳模型尺度为 618-511-391, 该模型的性能仍然落后于单一环境(无硅环境或有硅环境中)深度特征所建立的 SVR 模型性能。因此, 本研究采用 T-SAE 算法以期提高在无硅环境和有硅环境中油菜叶片中锌含量检测的准确性。

表 3 基于 SAE 提取的深度特征的 SVR 模型结果

Table 3 Results of SVR model based on deep features extracted by SAE

数据集 Data set	模型尺度 Model scale	训练集 Calibration set		预测集 Prediction set		
		训练集决定系数 R_c^2	训练集均方根误差 RMSEC/(mg·kg ⁻¹)	预测集决定系数 R_p^2	预测集均方根误差 RMSEP/(mg·kg ⁻¹)	相对分析误差 RPD
Z-SNV	618-481	0.912 9	0.025 38	0.809 5	0.036 74	2.490
	618-481-381	0.939 3	0.018 22	0.850 7	0.034 66	2.607
	618-481-381-273	0.928 7	0.023 64	0.820 1	0.036 94	2.509
	618-481-381-273-260	0.906 2	0.029 18	0.804 6	0.037 07	2.463
ZS-SNV	618-531	0.953 2	0.014 57	0.866 2	0.032 52	2.661
	618-531-441	0.963 4	0.012 97	0.876 6	0.028 54	2.732
	618-531-441-271	0.938 6	0.017 31	0.840 9	0.036 29	2.583
	618-531-441-271-191	0.919 7	0.023 16	0.825 8	0.036 70	2.517
ZZS-SNV	618-511	0.887 8	0.033 26	0.764 3	0.042 97	2.312
	618-511-391	0.912 8	0.024 87	0.839 4	0.036 35	2.549
	618-511-391-269	0.814 7	0.039 11	0.749 1	0.043 78	2.245
	618-511-391-269-229	0.813 3	0.040 15	0.738 6	0.046 86	2.198

利用 T-SAE 模型(双模型迁移堆叠自编码器(dual-model T-SAE))对源域中已构建好的无硅环境中深度网络关系模型为 SAE Model 1 和有硅环境中的深度网络关系模型为 SAE Model 2 进行深度特征学习迁移, 完成 T-

SAE Model 1 模型构建。其中, 无硅环境中不同 Zn 浓度预测的最佳预训练学习网络模型尺度为 618-481, 有硅环境中不同 Zn 浓度预测的最佳预训练学习网络模型尺度为 618-531, 则 T-SAE 模型的初始迁移网络模型尺度为

1 236-1 012。基于 T-SAE 提取的深度特征的 SVR 模型结果如表 4 所示。由表 4 可知, 对于无硅环境和有硅环境中的油菜叶片样本, 所建立的 SNV-T-SAE-SVR 模型对 Zn 含量预测性能最佳, 预测集的 R_p^2 、RMSEP 和 RPD 分别为 0.881 0、0.027 48 mg/kg 和 2.966, 最佳模型尺度为 1 236-1 012-812-571。从结果可以看出, 深度迁移学习模型能显著提高有硅环境和无硅环境中油菜叶片 Zn 含量的检测, 这一研究结果与深度迁移学习模型在油菜植

株中重金属镉含量检测^[21]和硅作用下油菜叶片 Pb 含量检测^[32]中应用的结果相一致。深度迁移学习算法能够共享源域(单一无硅环境或有硅环境下重金属 Zn 检测的深度学习 SAE 模型)浅层特征, 在有监督学习方式下对深层网络参数进行微调, 搭建基于深度学习和迁移学习的目标域(有硅和无硅环境中重金属 Zn 检测)学习框架, 提高硅作用下油菜叶片重金属 Zn 检测模型的精度和泛化能力。

表 4 基于 T-SAE 提取的深度特征的 SVR 模型结果
Table 4 Results of SVR model based on deep features extracted by T-SAE

模型尺度 Model scale	训练集 Calibration set		预测集 Prediction set		相对分析误差 RPD
	训练集决定系数 R_c^2	训练集均方根误差 RMSEC/(mg·kg ⁻¹)	预测集决定系数 R_p^2	预测集均方根误差 RMSEP/(mg·kg ⁻¹)	
1 236-1 012	0.949 6	0.016 04	0.835 6	0.031 39	2.522
1 236-1 012-812	0.956 7	0.015 28	0.857 9	0.030 73	2.634
1 236-1 012-812-571	0.970 5	0.012 04	0.881 0	0.027 48	2.966
1 236-1 012-812-571-308	0.969 8	0.012 82	0.868 0	0.030 28	2.697
1 236-1 012-812-571-308-223	0.943 4	0.016 39	0.774 7	0.040 57	2.378

3 结 论

在本研究中, 深度迁移学习算法迁移堆叠自编码器 T-SAE (transfer stacked auto-encoder) 结合 Vis-NIR 高光谱成像技术 (visible near-infrared hyperspectral imaging) 成功地实现了无硅环境和有硅环境中油菜叶片锌含量的较高精度检测, 所建立的支持向量机回归 SVR (support vector machine regression) 模型对无硅环境和有硅环境中的油菜叶片 Zn 含量预测性能较佳, 该模型预测集的决定系数 R_p^2 (the coefficient of determination) 和均方根误差 RMSEP (root mean square error) 分别为 0.839 4 和 0.036 35 mg/kg。本文所采用的深度迁移学习模型为无硅环境和有硅环境中油菜叶片锌含量无损检测提供了新思路, 为更好地监测农作物逆境胁迫和修复农业土壤重金属提供了强有力的技术支持。

[参 考 文 献]

- [1] 王敏捷, 盛光遥, 王锐. 土壤重金属污染修复植物处置技术进展[J]. 农业资源与环境学报, 2021, 38(2): 151-159. WANG Minjie, SHENG Guangyao, WANG Rui. Progress in disposal technologies for plants polluted with heavy metals after phytoextraction[J]. Journal of Agricultural Resources and Environment, 2021, 38(2): 151-159. (in Chinese with English abstract)
- [2] MATHIMANI T, RENE E, DEVANESAN S, et al. Removal of zinc by *Selenastrum* sp. and simultaneous biodiesel production from metal-resistant biomass for energy and environmental sustainability[J]. Algal Research-Biomass Biofuels and Bioproducts, 2023, 73: 103157.
- [3] GERHARDT K, GERWING P, GREENBERG B. Opinion: Taking phytoremediation from proven technology to accepted practice[J]. Plant, 2017, 256: 170-185.
- [4] FU L, SUN J, WANG S, et al. Nondestructive evaluation of Zn content in rape leaves using MSSAE and hyperspectral imaging[J]. Spectrochimica Acta Part A-molecular And Biomolecular Spectroscopy, 2023, 281: 121641.
- [5] 戴宏博, 陈克云, 杨京民, 等. 富硅稻壳灰施用对土壤镉活性和不同品种水稻镉积累的影响[J]. 农业工程学报, 2023, 39(4): 200-207. DAI Hongbo, CHEN Keyun, YANG Jingmin, et al. Effects of silica-rich rice hull ash on soil Cd activity and Cd accumulation in different varieties of rice[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(4): 200-207. (in Chinese with English abstract)
- [6] CEN G, SUN T, CHEN Y, et al. Characterization of silicon transporter gene family in *Saccharum* and functional analysis of the ShLsi6 gene in biotic stress[J]. Gene, 2022, 822: 146331.
- [7] RIZWAN M, ALI S, ADREES M, et al. Cadmium stress in rice: Toxic effects, tolerance mechanisms, and management: A critical review[J]. Environmental Science & Pollution Research, 2016, 23(18): 17859-17879.
- [8] 刘丽君, 黄张婷, 孟赐福, 等. 中国不同生态系统土壤硅的研究进展[J]. 土壤学报, 2021, 58(1): 31-41. LIU Lijun, HUANG Zhangting, MENG Cifu, et al. Research progress on soil silicon in different ecosystems in China[J]. Acta Pedologica Sinica, 2021, 58(1): 31-41. (in Chinese with English abstract)
- [9] ZHOU X, ZHAO C, SUN J, et al. Classification of heavy metal Cd stress in lettuce leaves based on WPCA algorithm and fluorescence hyperspectral technology[J]. Infrared Physics & Technology, 2021, 119: 103936.
- [10] VIJENDRA S, ACHLESH D. Effects of sophorolipids augmentation on the plant growth and phytoremediation of heavy metal contaminated soil[J]. Journal of Cleaner Production, 2021, 280: 124406.
- [11] 李亚丽, 刘蜜, 郭伟, 等. 利用电感耦合等离子体质谱法测定肥料重金属砷、镉、铅、铬含量[J]. 中国土壤与肥料, 2024(2): 235-243. LI Yali, LIU Mi, GUO Wei, et al. The determination of arsenic, cadmium, lead and chromium in fertilizer by inductively coupled plasma mass spectrometer method[J]. Soils and

- Fertilizers Sciences in China*, 2024(2): 235-243. (in Chinese with English abstract)
- [12] 黄富荣, 周泳欣, 郑海平, 等. 基于高光谱成像技术的菜心铜胁迫快速检测[J]. *现代农业装备*, 2022, 43(1): 17-23. HUANG Furong, ZHOU Yongxin, ZHENG Haiping, et al. Rapid detection of copper stress in flowering chinese cabbage based on hyperspectral imaging technology[J]. *Modern Agricultural Equipment*, 2022, 43(1): 17-23. (in Chinese with English abstract)
- [13] CAO Y, SUN J, YAO K, et al. Nondestructive detection of lead content in oilseed rape leaves based on MRF-HHO-SVR and hyperspectral technology[J]. *Journal of Food Processing Engineering*. 2021, 44(9): e13793.
- [14] FENG X, CHEN H, CHEN Y, et al. Rapid detection of cadmium and its distribution in *Miscanthus sacchariflorus* based on visible and near-infrared hyperspectral imaging[J]. *Science of the Total Environment*, 2019, 659: 1021-1031.
- [15] YU K, FANG S, ZHAO Y. Heavy metal Hg stress detection in tobacco plant using hyperspectral sensing and data-driven machine learning methods[J]. *Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy*, 2020, 245: 118917.
- [16] WEI L, WANG K, LU Q, et al. Crops fine classification in airborne hyperspectral imagery based on multi-feature fusion and deep learning[J]. *Remote Sensing*, 2021, 13(15): 2917.
- [17] 张凡, 王文秀, 王春山, 等. 高光谱图像结合卷积神经网络的马铃薯干腐病潜育期识别[J]. *光谱学与光谱分析*, 2024, 44(2): 480-489. ZHANG Fan, WANG Wenxiu, WANG Chunshan, et al. Study on hyperspectral detection of potato dry rot in gley stage based on convolutional neural network[J]. *Spectroscopy and Spectral Analysis*, 2024, 44(2): 480-489. (in Chinese with English abstract)
- [18] MISHRA P, SADEH R, BINO E, et al. Complementary chemometrics and deep learning for semantic segmentation of tall and wide visible and near-infrared spectral images of plants[J]. *Computers and Electronics in Agriculture*, 2021, 186: 106226.
- [19] NALEPA J, MYLLER M, KAWULOK M. Transfer learning for segmenting dimensionally reduced hyperspectral images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(7): 1228-1232.
- [20] 孙俊, 靳海涛, 芦兵, 等. 基于高光谱图像及深度特征的大米蛋白质含量预测模型[J]. *农业工程学报*, 2019, 35(15): 295-303. SUN Jun, JIN Haitao, LU Bing, et al. Prediction model of rice protein content based on hyperspectral image and deep feature[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(15): 295-303. (in Chinese with English abstract)
- [21] ZHOU X, ZHAO C, SUN J, et al. Detection of lead content in oilseed rape leaves and roots based on deep transfer learning and hyperspectral imaging technology[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2023, 290: 122288.
- [22] 国家卫生和计划生育委员会, 国家食品药品监督管理总局. GB 5009.14-2017 食品安全国家标准 食品中锌的测定[S]. 北京: 中国标准出版社, 2017: 1-4.
- [23] 孙俊, 周鑫, 毛罕平, 等. 基于荧光光谱的生菜农药残留检测[J]. *农业工程学报*, 2016, 32(19): 302-307. SUN Jun, ZHOU Xin, MAO Hanping, et al. Detection of pesticide residues in lettuce based on fluorescence spectra[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2016, 32(19): 302-307. (in Chinese with English abstract)
- [24] 毛继华, 赵恒谦, 金倩, 等. 河北铅锌尾矿库区土壤重金属含量高光谱反演方法对比[J]. *农业工程学报*, 2023, 39(22): 144-156. MAO Jihua, ZHAO Hengqian, JIN Qian, et al. Comparative study on the hyperspectral inversion methods for soil heavy metal contents in Hebei lead-zinc tailings reservoir areas[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(22): 144-156. (in Chinese with English abstract)
- [25] 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)
- [26] 孙俊, 周鑫, 李青林, 等. 乐果农药对生菜叶片微观结构的作用机理研究[J]. *农业机械学报*, 2017, 48(5): 355-360. SUN Jun, ZHOU Xin, LI Qinglin, et al. Mechanism of different concentrations of dimethoate residue on microstructure of lettuce leaves[J]. *Transactions of the Chinese Society for Agricultural Machinery*. 2017, 48(5): 355-360. (in Chinese with English abstract)
- [27] 章海亮, 谢潮勇, 田彭, 等. 基于可见/近红外光谱和数据驱动的机器学习方法测量土壤有机质和总氮[J]. *光谱学与光谱分析*, 2023, 43(7): 2226-2231. ZHANG Hailiang, XIE Chaoyong, TIAN Peng, et al. Measurement of soil organic matter and total nitrogen based on visible/near infrared spectroscopy and data-driven machine learning method[J]. *Spectroscopy and Spectral Analysis*. 2023, 43(7): 2226-2231. (in Chinese with English abstract)
- [28] YUN Y, WANG W, TAN M, et al. A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration[J]. *Analytica Chimica Acta*, 2014, 807: 36-43.
- [29] DENG B, YUN Y, LIANG Y, et al. A novel variable selection approach that iteratively optimizes variable space using weighted binary matrix sampling[J]. *Analyst*, 2014, 139(19): 4836-4845.
- [30] ZHOU X, ZHAO C, SUN J, et al. A deep learning method for predicting lead content in oilseed rape leaves using fluorescence hyperspectral imaging[J]. *Food Chemistry*, 2023, 409: 135251.

- [31] 岑海燕, 朱月明, 孙大伟, 等. 深度学习在植物表型研究中的应用现状与展望[J]. 农业工程学报, 2020, 36(9): 1-16. CEN Haiyan, ZHU Yueming, SUN Dawei, et al. Current status and future perspective of the application of deep learning in plant phenotype research[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(9): 1-16. (in Chinese with English abstract)
- [32] ZHOU X, ZHAO C, SUN J, et al. Determination of lead content in oilseed rape leaves in silicon-free and silicon environments based on deep transfer learning and fluorescence hyperspectral imaging[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2024, 311: 123991.
- [33] SUN J, TIAN Y, WU X, et al. Nondestructive detection for moisture content in green tea based on dielectric properties and VISSA-GWO-SVR algorithm[J]. *Journal of Food Process Engineering*, 2020, 43(8): e14421.
- [34] SHAO Y, JI S, XUAN G, et al. Soluble solids content monitoring and shelf life analysis of winter jujube at different maturity stages by Vis-NIR hyperspectral imaging[J]. *Postharvest Biology and Technology*, 2024, 210: 112773.
- [35] 张文祥, 潘嘹, 卢立新. 包装生鲜牛肉 pH 值的高光谱无损检测方法[J]. 食品与发酵工业, 2023, 49(22): 273-278. ZHANG Wenxiang, PAN Liao, LU Lixin. Hyperspectral non-destructive testing method for pH value of packaged fresh beef[J]. *Food and Fermentation Industries*, 2023, 49(22): 273-278. (in Chinese with English abstract)

Non-destructive determination of zinc content in oilseed rape leaves based on hyperspectral depth characteristics

ZHOU Xin^{1,4,5}, WANG Jian¹, ZHAO Chunjiang^{1,2,3}, SUN Jun¹, SHI Lei¹

(1. School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China; 2. Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; 3. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China; 4. Key Laboratory for Theory and Technology of Intelligent Agricultural Machinery and Equipment of Jiangsu University, Zhenjiang 212013, China; 5. Jiangsu Province and Education Ministry Co-sponsored Synergistic Innovation Center of Modern Agricultural Equipment, Zhenjiang 212013, China)

Abstract: Non-destructive testing can be expected to rapidly and accurately detect the zinc content in oilseed rape leaves. In this study, a high-precision detection was realized to combine with deep transfer learning using hyperspectral imaging technology. Oilseed rape with similar growth shape was divided into two groups (Group Z and Group ZS) by soilless cultivation, each of which included the five types of stress reagents. Therefore, 400 oilseed rape leaf samples were selected for each type of stress reagent, while 2 000 oilseed rape leaf samples were collected for each group, leading to 4000 oilseed rape leaf samples in total. The hyperspectral image information of oilseed rape leaf samples was obtained by hyperspectral imaging equipment. The whole blade was taken as the region of interest. The average spectral information was obtained in the region of interest after calculation. Firstly, the predictive performance of different pre-treated spectra was compared for the zinc content in oilseed rape leaves under the action of silicon. Standard normalized variable (SNV) was established as the best pre-processing. The spectral data was processed by SNV for further analysis. A stacked auto-encoder (SAE) was used to reduce the dimensionality of the best pre-processed spectral data, compared with the traditional. Finally, transfer stacked auto-encoder (T-SAE) was performed on the optimal SAE deep learning network. The transfer learning model was obtained to verify the portability between the deep learning models in silicon-free and silicon environments. The results showed that the support vector machine regression (SVR) model with SAE extraction depth features shared the best prediction on zinc content in the oilseed rape leaves under silicon-free or silicon environments. The best performance was achieved in the SNV-SAE-SVR model under a silicon-free environment. The coefficient of determination (R_c^2) and root mean square error (RMSEC) of the calibration set were 0.939 3 and 0.018 22 mg/kg, respectively, while the coefficient of determination (R_p^2), root mean square error (RMSEP), and residual predictive deviation (RPD) of prediction set were 0.850 7, 0.034 66 mg/kg and 2.607, respectively. The R_c^2 , RMSEC, R_p^2 , RMSEP, and RPD of the prediction set were 0.963 4, 0.012 97 mg/kg, 0.876 6, 0.028 54 mg/kg and 2.732, respectively. In addition, the SVR model with T-SAE extraction depth features performed the best prediction on zinc content in both silicon-free and silicon environments, where the R_c^2 , RMSEC, R_p^2 , RMSEP, and RPD of the optimal SNV-T-SAE-SVR model prediction set were 0.970 5, 0.012 04 mg/kg, 0.881 0, 0.027 48 mg/kg and 2.966, respectively. Deep transfer learning combined with hyperspectral imaging technology can effectively detect the zinc content in oilseed rape leaves under both silicon-free and silicon environments.

Keywords: non-destructive testing; heavy metals; silicon environment; deep learning; transfer learning; hyperspectral image