

基于同步辐射 X 射线荧光光谱与机器学习的非靶标金属组学方法区分暴露于不同形态汞的水稻

敖飘雪¹, 魏超杰², 解宏鑫³, 费钰茜⁴, 崔丽巍⁵, 王伟², 涂成龙¹, 商立海^{1,6}, 李柏³, 李玉锋³

1. 贵州医科大学公共卫生与健康学院/环境污染与疾病监控教育部重点实验室, 贵州 贵阳 561113
2. 中国农业大学工学院/全国金属组学创新研究中心, 北京 100083
3. 中国科学院高能物理研究所/中国科学院-香港大学金属组学与健康和环境联合实验室/中国科学院纳米生物效应与安全性重点实验室/北京金属组学平台/全国金属组学创新研究中心, 北京 100049
4. 河南农业大学国际教育学院, 河南 郑州 450002
5. 中国科学院大学资源与环境学院, 北京 100049
6. 中国科学院地球化学研究所/环境地球化学国家重点实验室, 贵州 贵阳 550081

摘要：

[背景] 汞是一种全球性污染物, 严重威胁人类健康。不同形态汞的毒性不同, 建立区分暴露于不同形态汞的样品的方法有助于针对性开展汞的治理, 为降低人类汞暴露风险提供依据。

[目的] 建立基于同步辐射 X 射线荧光 (SRXRF) 光谱与机器学习相结合的非靶标金属组学方法, 从而区分暴露于无机汞 (IHg) 或甲基汞 (MeHg) 的水稻。

[方法] 水稻种子分别暴露于超纯水 (对照组)、0.1 mg·L⁻¹ 的 IHg (IHg 组) 或 MeHg (MeHg 组) 溶液中, 种子发芽后继续培养 21 d, 收集水稻叶片、烘干、称重、压片。利用 SRXRF 测定各组水稻叶中金属组的含量。采用不同机器学习模型如软独立建模聚类分析 (SIMCA)、最小二乘判别分析 (PLS-DA) 和逻辑回归 (LR) 对不同组叶片的 SRXRF 全光谱进行分类和识别, 筛选出区分效果最优的模型以区分暴露于 IHg 或 MeHg 的水稻。进一步利用特征元素作为输入参数以提升运算速度, 减少模型计算量, 优化模型。

[结果] SRXRF 显示, 对照组、IHg 组和 MeHg 组的 SRXRF 光谱强度各不相同, 提示 IHg 或 MeHg 暴露可干扰水稻叶中金属组的稳态平衡。将 SRXRF 光谱进行主成分分析 (PCA), 发现对照组能与汞暴露组很好区分, 但无法区分 IHg 组和 MeHg 组。利用 PLS-DA、SIMCA 和 LR 三个模型进行区分, 发现训练集的准确率都高于 98%, 验证集的准确率都高于 95%, 交叉验证集的准确率都高于 94%, 其中 LR 模型的准确率均高于 PLS-DA 模型和 SIMCA 模型。以线性模型 LR 挑选出的 K、Ca、Mn、Fe、Zn 为特征元素区分 IHg 组和 MeHg 组的预测准确率为 92.05%。与全光谱模型相比, 利用特征光谱预测模型虽然预测准确率下降, 但模型输入参数减少了 99.51%, 且精确度、召回率和 F1 得分在 84.48% 以上, 同样可用于区分暴露于不同形态汞的水稻。

[结论] 基于 SRXRF 和机器学习的非靶标金属组学方法可快速识别暴露于不同形态汞的水稻, 减少人体摄入汞的风险。

关键词： 无机汞; 甲基汞; 水稻; 同步辐射 X 射线荧光光谱; 机器学习; 非靶标金属组学

Non-targeted metallomics based on synchrotron radiation X-ray fluorescence spectroscopy and machine learning for screening inorganic or methylmercury-exposed rice plants AO Piaoxue¹, WEI Chaojie², XIE Hongxin³, FEI Yuqian⁴, CUI Liwei⁵, WANG Wei², TU Chenglong¹, SHANG Lihai^{1,6}, LI Bai³, LI Yufeng³ (1. School of Public Health/ Key Laboratory of Environmental Pollution Monitoring and Disease Control, Ministry of Education, Guizhou Medical University, Guiyang, Guizhou 561113, China; 2. College of Engineering/ National Consortium for Excellence in Metallomics, China Agricultural University, Beijing 100083, China; 3. CAS-HKU Joint Laboratory of Metallomics on Health and Environment/ CAS Key Laboratory for Biomedical Effects of Nanomaterials and Nanosafety/ Beijing Metallomics Facility/ National Consortium for Excellence in Metallomics/ Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China; 4. College



DOI 10.11836/JEOM24253

组稿专家
陈汉清 (首都医科大学公共卫生学院),
E-mail: chenhq@ccmu.edu.cn

基金项目
国家自然科学基金 (11975247); 中国科学院大学生创新实践训练计划 (IHEP202305)

作者简介
敖飘雪 (1999—), 女, 硕士生;
E-mail: aopx@ihep.ac.cn

通信作者
涂成龙, E-mail: chenglongtu@163.com
商立海, E-mail: shanglihai@vip.skleg.cn
李玉锋, E-mail: liyf@ihep.ac.cn

作者中包含编委会成员 无
伦理审批 不需要
利益冲突 无申报
收稿日期 2024-06-01
录用日期 2024-08-27

文章编号 2095-9982(2024)10-1095-08
中图分类号 R11
文献标志码 A

►引用
敖飘雪, 魏超杰, 解宏鑫, 等. 基于同步辐射 X 射线荧光光谱与机器学习的非靶标金属组学方法区分暴露于不同形态汞的水稻 [J]. 环境与职业医学, 2024, 41(10): 1095-1102.

►本文链接
www.jeom.org/article/cn/10.11836/JEOM24253

Funding
This study was funded.

Correspondence to
TU Chenglong, E-mail: chenglongtu@163.com
SHANG Lihai, E-mail: shanglihai@vip.skleg.cn
LI Yufeng, E-mail: liyf@ihep.ac.cn

Editorial Board Members' authorship No
Ethics approval Not required
Competing interests None declared
Received 2024-06-01
Accepted 2024-08-27

► To cite
AO Piaoxue, WEI Chaojie, XIE Hongxin, et al. Non-targeted metallomics based on synchrotron radiation X-ray fluorescence spectroscopy and machine learning for screening inorganic or methylmercury-exposed rice plants [J]. Journal of Environmental and Occupational Medicine, 2024, 41(10): 1095-1102.

► Link to this article
www.jeom.org/article/en/10.11836/JEOM24253

of International Education, Henan Agricultural University, Zhengzhou, Henan 450002, China; 5. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China; 6. State Key Laboratory of Environmental Geochemistry/ Institute of Geochemistry, Chinese Academy of Sciences, Guiyang, Guizhou 550081, China)

Abstract:

[Background] Mercury, as a global heavy metal pollutant, poses a serious threat to human health. The toxicity of mercury depends on its chemical form. Distinguishing the forms of mercury in the environment is of great significance for mercury management and reducing human mercury exposure risks.

[Objective] To establish a non-targeted metallomics method based on synchrotron radiation X-ray fluorescence (SRXRF) spectroscopy combined with machine learning to screen inorganic mercury (IHg) or methylmercury (MeHg) exposed rice plants.

[Methods] Rice seeds were exposed to ultra-pure water (control group), 0.1 mg·L⁻¹ IHg (IHg group) or MeHg (MeHg group) solutions, respectively. After germination, the seedlings were cultured for 21 d, and rice leaves were collected, dried, weighed, and pressed. The content of metallome in rice leaves was determined by SRXRF. Machine learning models including soft independent modeling cluster analysis (SIMCA), partial least squares discriminant analysis (PLS-DA), and logistic regression (LR) were used to classify the SRXRF full spectra of different groups and find the best model to distinguish rice exposed to IHg or MeHg. Besides, characteristic elements were selected as input parameters to optimize the model by improving computing speed and reducing model calculation.

[Results] The SRXRF spectral intensities of the control group, IHg group, and MeHg group were different, indicating that exposure to IHg and MeHg can interfere the homeostasis of metallome in rice leaves. The results of principal component analysis (PCA) of SRXRF spectra showed that the control group could be well distinguished from the mercury exposed groups, but the IHg group and the MeHg group were mostly overlapped. The accuracy rates of the three models (PLS-DA, SIMCA, and LR) were higher than 98% for the training set, higher than 95% for the validation set, and higher than 94% for the cross-validation set. Besides, the accuracy of the LR model was higher than that of the PLS-DA model and the SIMCA model. Furthermore, the accuracy was 92.05% when using characteristic elements K, Ca, Mn, Fe, and Zn selected by LR to distinguish the IHg group and the MeHg group. Compared with the full spectra model, although the prediction accuracy of the characteristic spectral model decreased, the input parameters of the model decreased by 99.51%, and precision, recall, and F1 score were above 84.48%, indicating that the model could distinguish rice exposed to different mercury forms.

[Conclusion] Non-targeted metallomics method based on SRXRF and machine learning can be applied for high-throughput screening of rice exposed to different forms of mercury and thus decrease the risks of people being exposed to mercury.

Keywords: inorganic mercury; methylmercury; rice; synchrotron radiation X-ray fluorescence; machine learning; non-targeted metallomics

汞 (mercury, Hg) 是一种全球性污染物, 具有很强的致癌性、致畸性、神经毒性和发育毒性等, 可随大气在全球范围内进行长距离迁移, 对人类和生态系统带来不利影响^[1-2]。根据联合国环境规划署《2018 年联合国全球汞评估》报告, 全球每年汞排放量约为 8000 t, 对人类和环境造成了巨大威胁。环境中汞包括无机汞和有机汞。汞的毒性、生物可利用性、迁移性和再迁移性与其化学形态密切相关^[3]。不同形态的汞的毒性作用不同, 人类长期暴露于无机汞 (inorganic mercury, IHg), 会对肾脏造成损伤^[4]; 有机汞特别是甲基汞 (methylmercury, MeHg), 具有生物蓄积和生物放大能力, 可通过食物链富集到人体内, 严重损害神经系统, 并可穿过胎盘屏障影响胎儿神经发育^[5]。水稻是非常重要的粮食作物, 研究发现汞矿区稻米甲基汞含量占水稻植株中总甲基汞含量的 77% 以上, 食用被甲基汞污染的大米是人体摄入甲基汞的重要来源^[6-9]。全球超过 114 个国家种植水稻, 是世界一半人口的主食。据预测, 到 2025 年, 稻米产量要超过 8 亿 t 才能满足世界人口的需求^[10]。研究表明汞可影响包括水稻在内

的各种农作物的生长, 导致产量下降, 且不同形态汞对植物的毒性不同^[11], 因此, 建立筛查暴露于不同形态汞农作物的方法, 可为环境汞的治理以及降低人类摄入汞的风险提供支持。目前虽然已有方法可以直接检测稻米中汞的形态, 但需要将样品进行化学前处理、色谱分离后再通过电感耦合等离子体质谱仪等仪器进行形态分析。这些方法通常非常耗时且分析通量较低, 不利于开展大规模样品筛查。

金属组学是继基因组学、蛋白质组学和代谢物组学之后提出的系统研究金属/类金属及其与生物体内基因、蛋白质、代谢物和其他生物分子的相互作用和功能联系的一门新兴学科^[12-14]。健康的生物体系中存在金属组的稳态平衡, 而环境污染物暴露可能会扰乱生物体系中金属组的稳态平衡。如暴露于纳米聚对苯二甲酸乙二酯会影响植物体内的金属组稳态平衡^[15]。另外, 重金属也可引起植物应激, 导致植物离子/细胞稳态失衡^[16]。因此, 汞暴露也可能引起生物体内金属组的变化, 而不同形态的汞引起生物体内金属组的变化差异也可能不同。

同步辐射技术已被广泛应用于不同研究领域^[17]。同步辐射 X 射线荧光(synchrotron radiation X-ray fluorescence, SRXRF)是一种重要的多元素无损/微损分析方法。课题组研究表明,通过基于 SRXRF 和机器学习的非靶标金属组学方法,可利用血清筛查癌症患者^[18-19]。此外,基于 SRXRF 和深度学习算法的非靶标金属组学方法也已成功用于筛查暴露于不同尺寸微塑料的水稻^[20]。鉴于汞污染水稻对人体健康的影响,区分不同汞形态暴露的水稻,有助于降低人体接触汞的健康风险。

本研究拟以水稻幼苗为例,发展 SRXRF 结合机器学习的非靶标金属组学方法以区分暴露于不同形态汞的水稻。本工作可通过识别含不同形态汞的水稻从而减少人体摄入含汞稻米的风险,也可为筛查环境及人体中其他新兴污染物如持久性有机污染物、人工纳米材料等提供借鉴。

1 材料与方法

1.1 水稻培养及处理

挑选颗粒饱满的水稻种子(Y 两优 900)使用 1% (体积分数)次氯酸钠(NaClO)消毒 15 min,用超纯水(18.2 MΩ·cm)洗净后平铺在装有超纯水(对照组)、0.1 mg L⁻¹ IHg(IHg 组)或 0.1 mg L⁻¹ MeHg(MeHg 组)溶液的 15 cm 的培养皿中避光发芽(每个培养皿中放置 50 粒水稻种子,每组三个平行样)。0.1 mg L⁻¹ 为中国农业土壤中总汞的平均水平。发芽后,在培养条件为白天温度 28 °C,光照时间 14 h,晚上温度 20 °C,黑暗时间 10 h,相对湿度 50%~70%,光照 300~350 lux 的培养间继续培养 21 d。培养结束后收集水稻叶片,置于烘箱中,在 40 °C 条件下烘至恒重,剪碎。准确称取 20 mg 样品放入压片机中压成直径为 6 mm 的圆片,固定在聚酯薄膜上用于后续 SRXRF 数据采集。

1.2 SRXRF 数据采集

SRXRF 数据采集于北京同步辐射装置(Beijing Synchrotron Radiation Facility, BSRF)的 4W1B 线站。BSRF 的运行能量为 2.5 GeV,束流强度为 150~250 mV。入射 X 射线是能量为 10~18 KeV 能量的多色光。X 射线荧光强度利用四元 Hitachi Vortex-ME4®硅漂移探测器(日本 Hitachi High Technologies)与 Quantum Detectors Xspress3 多通道分析仪系统(英国 Quantum Detectors)进行检测。束斑大小为 50 μm×50 μm,采集时间为 1 s^[20]。水稻压片样品安装在计算机控制的 XYZ 平移台上进行 SRXRF 光谱数据收集^[21]。对照组、IHg 组和 MeHg 组

三组样品共获得 576 条光谱数据,每组样品获得 192 条光谱。

1.3 数据分析

首先,通过主成分分析(principal components analysis, PCA)对采集的 SRXRF 数据进行线性变换,实现数据降维。在 PCA 的基础上,采用软独立建模聚类分析(soft independent modeling class analog, SIMCA)、最小二乘判别分析(partial least squares discriminant analysis, PLS-DA)和逻辑回归(logistic regression, LR)等机器学习算法,建立判别分析模型,对数据进行分类和识别。通过比较训练集、交叉验证集和验证集准确率来评价模型,筛选出效果最好的模型。本次实验将获取后的水稻叶片压片后,获取 SRXRF 光谱。将获取的 576 条光谱数据按照 7:3 的比例随机分配训练集与验证集。训练集样本数为 403,验证集为 173。采用五折交叉验证产生交叉验证集,将数据集随机划分五份,其中四份为训练集,剩余一份为验证集,依次循环验证集,共获取五个检测准确率,取均值为交叉验证集的结果。

结合 SRXRF 全光谱数据,LR 模型见公式(1)~(3)。

$$Z_1 = a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,1024}x_{1024} + b_1 \quad (1)$$

$$Z_2 = a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,1024}x_{1024} + b_2 \quad (2)$$

$$Z_3 = a_{3,1}x_1 + a_{3,2}x_2 + \cdots + a_{3,1024}x_{1024} + b_3 \quad (3)$$

$x_1, x_2, \dots, x_{1024}$ 表示通道为 1~1024 的 SRXRF 光谱。

Z 表示载荷与通道乘积的累计和, a 表示线性模型的载荷系数。

Z_1 、 Z_2 、 Z_3 分别表示对照组、IHg 组和 MeHg 组的载荷与通道乘积的累计和。

使用线性模型预测, P_1 、 P_2 、 P_3 分别为对照组、IHg 组、MeHg 组的预测值,见公式(4)。

$$P_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \quad (4)$$

i 分别取 1、2、3(分别对应对照组、IHg 组和 MeHg 组)。公式 4 中 Z_i 分别取公式(1)~(3)中的 Z_1 、 Z_2 和 Z_3 , j 取 3。

精确度表示预测为正例中正确的比例,召回率表示预测为正例中正确个数占实际正例比例,F1 得分表示精确度与召回率的调和平均数,准确率为所有的预测正确(正例负例)的占总的比重。计算公式(5)~(8)如下所示:

$$\text{精确度} = TP / (TP + FP) \quad (5)$$

$$\text{召回率} = TP / (TP + FN) \quad (6)$$

$$F1 = (2 \times \text{精确度} \times \text{召回率}) / (\text{精确度} + \text{召回率}) \quad (7)$$

$$\text{准确率} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (8)$$

式中, TP: 被正确地划分为正例的个数, 即实际为正例且被分类器划分为正例的实例数(样本数); FP: 被错误地划分为正例的个数, 即实际为负例但被分类器划分为正例的实例数; FN: 被错误地划分为负例的个数, 即实际为正例但被分类器划分为负例的实例数; TN: 被正确地划分为负例的个数, 即实际为负例且被分类器划分为负例的实例数。

为了优化模型, 减少数据计算时间, 从高维线性相关的 SRXRF 光谱中挑选影响模型分类结果的特征光谱, 建立特征光谱优化模型。

Pycharm 是一种集成开发环境, 能够提供包含编辑、调试与测试的功能与工具, 较大程度提高开发效率。Scikit-learn 是 Python 中最广泛使用的机器学习库之一, 非机器学习专家也可以访问, 并且可以跨学科和应用领域^[22]。因此本实验数据处理与模型开发在基于 PyCharm 环境和 Scikit-learn 算法库的 Windows 系

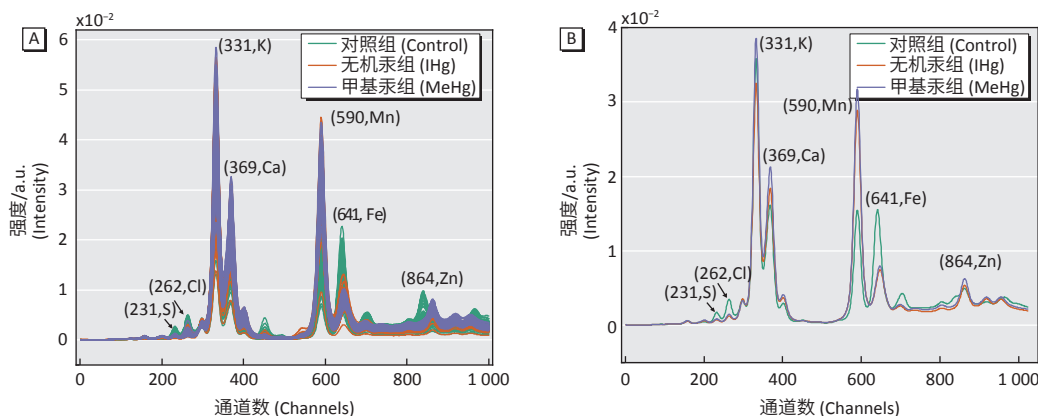
统中进行。

2 结果

2.1 暴露 IHg 或 MeHg 的水稻叶片的 SRXRF 图谱

图 1 显示了通道数在 1~1024 范围内的对照组、IHg 组和 MeHg 组的原始 SRXRF 光谱(图 1A)及均值光谱(图 1B)。根据波峰位置和光谱强度可以判断样品中的各元素含量的差异情况。由图 1 可知, 三组样品的光谱强度不同, 考虑到样品质量相同, 这说明不同组中对应的元素含量不同。

在 231、262、641 通道中, 对照组的光谱强度均高于 IHg 组和 MeHg 组, 经拟合, 对应为硫(S)、氯(Cl)及铁(Fe)元素。在 369、590、864 通道中, 对照组的光谱强度均低于 IHg 组和 MeHg 组, 拟合对应为钙(Ca)、锰(Mn)及锌(Zn)元素。在 331 通道中, 对应为钾(K)元素, 光谱强度 MeHg 组 > 对照组 > IHg 组; 在所有通道中, MeHg 组的光谱强度均高于 IHg 组。



[注] A: 原始光谱, B: 均值光谱。

[Note] A: Raw spectra, B: Averaged spectra.

图 1 暴露 IHg 或 MeHg 的水稻叶的 SRXRF 光谱

Figure 1 SRXRF spectra of rice leaves exposed to IHg or MeHg

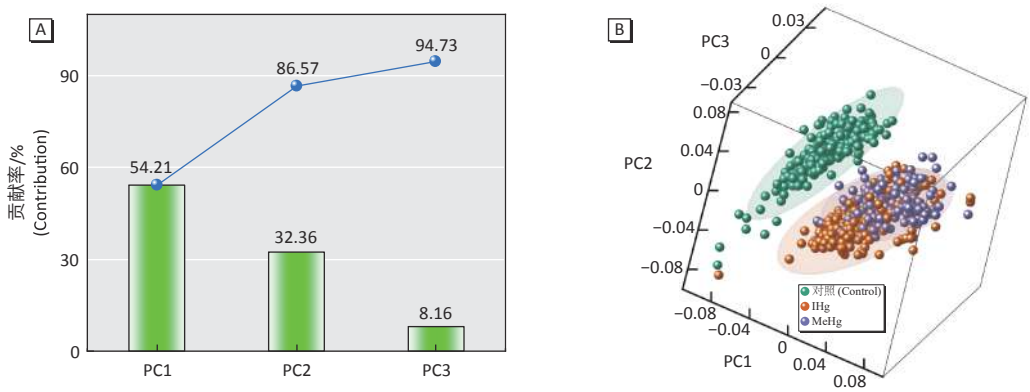
2.2 PCA 聚类分析

为进一步区分 IHg 和 MeHg 暴露的水稻, 对 SRXRF 光谱进行了 PCA 分析。PCA 分析是一种对数据降维的非监督聚类方法, 通过 PCA 得到的少量 PC 变量可以解释大部分原始信息, PC 贡献率越高, 越能更好地解释原始信息。如图 2A 所示, PC1 为第一主成分, 贡献率为 54.21%; PC2 为第二主成分, 贡献率为 32.36%; PC3 为第三个主成分, 贡献率为 8.16%; PC1、PC2、PC3 的累积贡献率已达到 94.73%, 能够表达原始光谱中绝大部分的信息, 清晰地看出不同样品的光谱差异, 因此选用 PCA 得到的前 3 个 PCs 进行了聚类分析, 得到如图 2B 的散点图。在 PCA 散点图中, 数个样本的点聚在一起, 那么就说明这几个样本之间的相

似性非常高, 反之, 如果几个样本的点非常分散, 则说明这几个样本之间的相似性比较低。根据图 2B 可知, 对照组(绿色)的光谱在前 3 个主成分分布上有明显的偏离, 汞暴露组与对照组的光谱特征存在一定可区分度, 但是 IHg 组(橙色)和 MeHg 组(紫色)的光谱有部分存在叠加现象。

2.3 基于 SRXRF 全光谱区分 IHg 或 MeHg 暴露水稻

在 PCA 分析的基础上, 对 SRXRF 的全光谱进行建模分析。基于 SRXRF 全光谱的模型区分不同样品的结果如表 1 所示, 三个模型训练集的准确率都高于 98%, 验证集的准确率都高于 95%, 交叉验证集的准确率都高于 94%, 说明三个模型都有很好的预测效果, 其中 LR 模型预测效果最好。



[注] A: 前 3 个 PC 的贡献率。B: 前 3 个 PC 的得分散点图。
[Note] A: The contribution rate of the first three PCs. B: Score plot of the first three PCs.

图 2 SRXRF 光谱的 PCA 处理结果
Figure 2 Classification of SRXRF spectra by PCA

表 1 基于 SRXRF 全光谱的机器学习模型准确率
Table 1 The accuracies of machine learning models based on full spectra of SRXRF

模型 (Model)	变量个数 (No. of variable)	潜变量 (Latent variable)	准确率(Accuracy)/%		
			训练集 (Training set)	验证集 (Validation set)	交叉验证集 (Cross-validation set)
PL-SDA	1000	16	98.27	95.78	96.37
SIMCA	1000	17	98.35	96.53	94.44
LR	1000	/	99.20	97.57	95.44

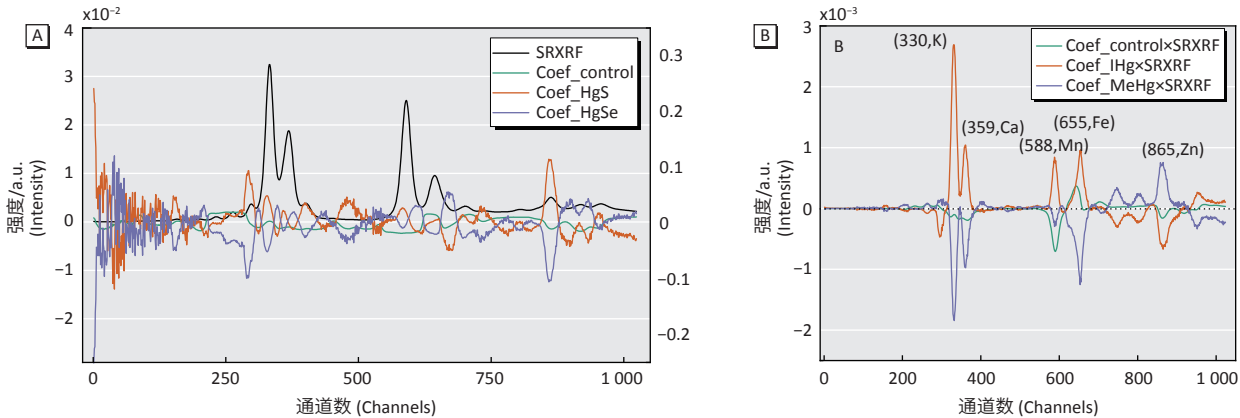
2.4 基于特征波长的 LR 模型区分 IHg 和 MeHg 暴露水稻

为了提升运算速度,减少模型计算量,需从高维线性相关的 SRXRF 光谱中挑选影响模型分类结果的特征光谱,建立特征光谱优化模型。在本研究中,基于 SRXRF 全光谱的 LR 模型分类效果最好。每个分类的系数权重结果图如图 3A 所示。考虑到光谱数据值不同,需考虑数据的系数和权重。因此,通过解析 LR 模型,然后根据每个通道的均值光谱和权重的乘积进行求

和,挑选出特征光谱。每个通道所在位置的权重与所在光谱乘积绘制曲线如图 3B 所示,330、359、588、655、865 通道的 IHg 与 MeHg 谱图有明显差异。通过峰面积拟合后,这些通道分别对应 K、Ca、Mn、Fe、Zn 元素。因此,选择 330、359、588、655、865 为特征通道建立特征光谱优化的 LR 模型。

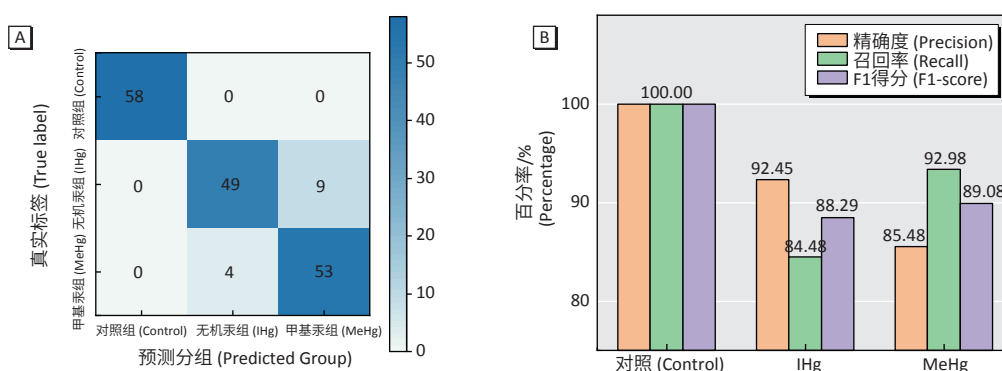
图 4A 展示了特征光谱分类模型的混淆矩阵与特征谱模型的预测准确性。在混淆矩阵中,特征光谱分类模型能够准确预测对照组,但 IHg 组与 MeHg 组存在误判的情况,有 9 个 IHg 组样本被误判为 MeHg 组,4 个 MeHg 组样本被误判为 IHg 组。模型准确率为 92.05%。

特征光谱分类模型的评价结果如图 4B 所示,模型区分对照组的精确度、召回率和 F1 得分都为 100%,虽然区分 IHg 组和 MeHg 组的精确度、召回率和 F1 得分都较对照组低,但都高于 84.48%,可用于筛查暴露于不同形态汞的水稻。



[注] A: 每个分类的系数权重。B: 均值光谱和权重的乘积曲线图。
[Note] A: Coefficient weight for each classification. B: Product curve of mean spectra and weight.

图 3 线性模型的详细描述
Figure 3 Detailed information of LR model



[注] A: 混淆矩阵, B: 分类效果图。

[Note] A: Confusion matrix, B: Classification rendering.

图4 基于 SRXRF 特征光谱的 LR 模型的预测准确度

Figure 4 Performance of LR model with characteristic SRXRF spectra

3 讨论

前人研究表明, IHg 与 MeHg 对水稻的毒性有明显差异, 包括种子发芽率、株高、穗重及叶绿素含量差异等^[23-24]。这些往往也会导致植物体内元素含量(金属组)变化^[25-27]。因此, 本研究以水稻幼苗为例, 建立 SRXRF 结合机器学习的非靶标金属组学方法, 从而区分暴露于不同形态汞的水稻。SRXRF 光谱研究结果显示, 在 Hg 暴露后, 水稻中 S、Cl 和 Fe 元素含量减少。研究发现, 水稻 Hg 暴露后, 水稻叶绿素含量减少, 从而影响光合作用^[28]。S、Cl 和 Fe 元素均参与植物光合作用^[25-26, 29], 而 Hg 阻碍了植物中 S 和 Fe 的获取, 引起植物光合作用紊乱^[27]。因此, Hg 暴露后, 水稻中 S、Cl 和 Fe 元素含量减少。除此以外, 在 Hg 暴露后, 水稻中的 Ca、Mn 和 Zn 元素含量升高。Ca²⁺在植物生长发育以及外界信号响应方面起着信使作用, 当植物受到外界环境刺激时, 植物 Ca²⁺浓度出现特异性升高, 细胞间可通过这些钙信号进行信息传递, 从而产生正确的应答反应^[30]。Zn 和 Mn 是植物生长发育过程中重要的微量营养元素, 都具有氧化应激的缓解作用, 对植物抗逆性起着重要作用^[31-32]。而汞诱导的氧化应激是汞植物毒性的代表性特征^[33], 因此, 当水稻 Hg 暴露后, 水稻可能通过摄取更多的 Zn 和 Mn 来缓解汞暴露导致的氧化损伤。MeHg 组的光谱强度均高于 IHg 组表明 MeHg 引起金属组的更大改变, 同样表明 IHg 与 MeHg 的植物毒性差异。综上所述, 通过研究水稻植株的金属组含量变化情况, 可区分环境中暴露于不同形态汞的水稻。

本研究采用了不同的机器学习模型, 以选出最佳预测模型。SIMCA 分析是一种基于 PCA 的有监督模式识别方法。首先利用先验分类知识, 对训练集中的每

一类样本建立一个 PCA 模型; 然后将未知样本逐一去拟合各类的主成分模型, 从而对数据进行判别归类^[34]。PLS-DA 是一种有监督的判别分析方法, 主要是通过构建自变量与分类变量之间的回归模型, 然后有效地提取与分类相关的特征变量, 实现数据的分类和识别, 能够解决线性或弱非线性数据的分类和判别问题, 是最常用的分类算法之一^[35]。LR 模型属于线性回归方法, 模型基于二项分类, 能够对定性变量的准确性进行有效的分析^[36]。结果显示 SIMCA、PLS-DA、LR 模型预测训练集和验证集的准确率均达到 95%, 都有很好的预测效果, LR 模型训练集和验证集的准确率均高于 PLS-DA 模型和 SIMCA 模型, 可见 LR 这种构建预测模型的方法, 在区分不同形态汞对植物的毒性作用时能够发挥较好的作用。另外, 样本划分和排序随机性对模型有一定的影响, 仅选择 1 组训练集与测试集, 可能导致训练所得的预测准确率偏离真实值, 存在评估可靠性不足的隐患^[37]。因此本次实验引入交叉验证, 对数据集进行多次切分, 形成多组训练集和验证集以提高样本信息利用率, 使预测结果稳定而真实。结果显示, SIMCA、PLS-DA、LR 三个模型的交叉验证准确率均高于 94%, 模型评估可靠性较好。

虽然 LR 模型预测准确率在 95%以上, 但是输入全光谱数据量庞大, 另外, SRXRF 图谱显示, 输入的 1~1024 通道的光谱中, 三个组绝大多数元素光谱是没有差异的, 因此为了提升运算速度, 减少模型计算量, 本研究从高维线性相关的 SRXRF 光谱中挑选影响模型分类结果的特征光谱, 建立特征光谱优化模型。通过解析线性模型, 共挑选出 K、Ca、Mn、Fe、Zn 五个特征元素, 并建立特征光谱模型, 模型准确率可达 92.05%, 精确度、召回率和 F1 得分均高于 84.48%。虽然该方法比

全光谱模型准确率下降,但模型输入参数量下降了99.51%,从而有助于实现高通量快速筛查。

本研究也存在一定的局限性。本实验为室内水培实验,环境和人为因素的影响可控,而真实自然环境复杂,影响因素多,Ca、Mn、Zn及K都是水稻的常量元素,容易受到如施肥、土壤调理剂或土地的盐碱水平等自然及人为因素影响,因此后续还应收集自然环境下的样品进行分析。另外本研究仅基于SIMCA、PLS-DA、LR预测模型进行结果分类,后续研究中还可以采用其他机器学习方法进行模型拟合,以期提高预测效能并对本研究相关结论进行验证。

汞的毒性取决于其化学形态^[2]。研究发现,ihg毒性作用主要表现为肾损伤,而MeHg主要引起神经系统损伤,可通过胎盘屏障影响胎儿神经发育,可通过食物链蓄积对人类健康产生危害。如1956年日本熊本县爆发的水俣病事件、1972年伊拉克发生的食用MeHg杀菌剂拌种后的种子造成的中毒事件等,给人类和生态环境带来了严重灾难^[38-39]。人群摄入MeHg的途径除了食用鱼类及其他水产品外,还可通过食用被汞污染的大米,如贵州典型汞矿区的大部分大米中MeHg含量(4.40~14.6 ng·g⁻¹)远高于对照区大米MeHg的平均含量(2.5 ng·g⁻¹),严重威胁着当地人群的健康^[2]。因此识别甲基汞暴露的水稻对降低人类健康风险有着重要意义。除直接检测稻米中的汞形态外,本研究所发展的方法无需对样品进行化学处理,可快速准确区分暴露于不同形态汞的水稻,有利于开展大规模筛查。考虑到同步辐射机时有限,课题组正在利用商用X射线荧光机器开展相关研究。本研究也可为筛查环境或人体中其他污染物提供借鉴。

参考文献

- [1] BECKERS F, RINKLEBE J. Cycling of mercury in the environment: sources, fate, and human health implications: a review[J]. *Crit Rev Environ Sci Technol*, 2017, 47(9): 693-794.
- [2] MAHBUB KR, KRISHNAN K, NAIDU R, et al. Mercury toxicity to terrestrial biota[J]. *Ecol Indic*, 2017, 74: 451-462.
- [3] 冯琳, 赵媚, 冯新斌, 等. 食用大米人群甲基汞暴露健康风险及摄入量限值研究进展[J]. *安全与环境工程*, 2022, 29(5): 5-12.
FENG L, ZHAO M, FENG XB, et al. Research progress on health risk and intake limit of methylmercury exposure among rice consumption population[J]. *Saf Environ Eng*, 2022, 29(5): 5-12.
- [4] CHEN B, DONG S. Mercury contamination in fish and its effects on the health of pregnant women and their fetuses, and guidance for fish consumption—a narrative review[J]. *Int J Environ Res Public Health*, 2022, 19(23): 15929.
- [5] WANG W, CHEN F, ZHANG L, et al. Neurotransmitter disturbances caused by methylmercury exposure: microbiota-gut-brain interaction[J]. *Sci Total Environ*, 2023, 873: 162358.
- [6] ZHANG H, FENG X, LARSEN T, et al. In inland China, rice, rather than fish, is the major pathway for methylmercury exposure[J]. *Environ Health Perspect*, 2010, 118(9): 1183-1188.
- [7] MENG M, LI B, SHAO JJ, et al. Accumulation of total mercury and methylmercury in rice plants collected from different mining areas in China[J]. *Environ Pollut*, 2014, 184: 179-186.
- [8] FENG L, LI P, FENG X. Methylmercury bioaccumulation in rice and health effects: a systematic review[J]. *Curr Opin Environ Sci Health*, 2021, 23: 100285.
- [9] LIU M, ZHANG Q, CHENG M, et al. Rice life cycle-based global mercury biotransport and human methylmercury exposure[J]. *Nat Commun*, 2019, 10(1): 5164.
- [10] ABDULLAH-ZAWAWI MR, GOVENDER N, MUHAMMAD NA N, et al. Genome-wide analysis of sulfur-encoding biosynthetic genes in rice (*Oryza sativa* L.) with *Arabidopsis* as the sulfur-dependent model plant[J]. *Sci Rep*, 2022, 12(1): 13829.
- [11] 瞿爱权, 东惠如, 李俊国. 汞对水稻、油菜影响的研究初报[J]. *环境科学*, 1980(6): 50-52,13.
QU AQ, DONG HR, LI JG. Preliminary study on the effects of mercury on rice and rapeseed[J]. *Environ Sci*, 1980(6): 50-52,13.
- [12] 何丽娜, SCHOTT E, 吴一楠. 金属组学方法识别水中络合态重金属形态及其吸附净化技术研究进展[J]. *中国无机分析化学*, 2024, 14(1): 46-55.
HE LN, SCHOTT E, WU YN. Research progress on identification of heavy metal complexes in water by metallomics and its purification technologies by adsorption[J]. *Chin J Inorg Anal Chem*, 2024, 14(1): 46-55.
- [13] XIE H, WEI X, ZHAO J, et al. Size characterization of nanomaterials in environmental and biological matrices through non-electron microscopic techniques[J]. *Sci Total Environ*, 2022, 835: 155399.
- [14] 李玉峰, 孙红哲, 陈春英, 等. 金属组学[M]. 北京: 科学出版社, 2016: 1-5.
LI Y F, SUN H Z, CHEN C Y, et al. *Metallomics*[M]. Beijing: Science Press, 2016: 1-5.
- [15] 解宏鑫, 费钰茜, 曹浩楠, 等. 纳米聚对苯二甲酸乙二醇酯对水稻幼苗光合作用系统和金属组稳态平衡的影响[J]. *中国无机分析化学*, 2024, 14(08): 1058-1065.
XIE HX, FEI YQ, CAO HN et al. Effects of nano poly(ethylene terephthalate) on photosynthesis system and metallome homeostasis of rice seedlings[J]. *Chin J Inorg Anal Chem*, 2024, 14(08): 1058-1065.
- [16] JAMLA M, KHARE T, JOSHI S, et al. Omics approaches for understanding heavy metal responses and tolerance in plants[J]. *Curr Plant Biol*, 2021, 27: 100213.
- [17] 史雪倩, 李红, 白旭, 等. 同步辐射技术在环境纳米金属组学研究中的应用[J]. *生态毒理学报*, 2019, 14(2): 1-11.
SHI XQ, LI H, BAI X, et al. Review of studies on the environmental nanometallomics using synchrotron radiation techniques[J]. *Asian J Eco-toxicol*, 2019, 14(2): 1-11.
- [18] 魏超杰, 李超, 解宏鑫, 等. 基于同步辐射X射线荧光光谱与一维卷积神经网络的癌症筛查方法[J]. *中国无机分析化学*, 2024, 14(1): 104-111.
WEI CJ, LI C, XIE HX, et al. Research on cancer screening method based on synchrotron radiation X-ray fluorescence spectroscopy and one-dimensional convolutional neural network[J]. *Chin J Inorg Anal Chem*, 2024, 14(1): 104-111.
- [19] HE L, LU Y, LI C, et al. Non-targeted metallomics through synchrotron radiation X-ray fluorescence with machine learning for cancer screening using blood samples[J]. *Talanta*, 2022, 245: 123486.

- [20] XIE H, WEI C, WANG W, et al. Screening the phytotoxicity of micro/nanoplastics through non-targeted metallomics with synchrotron radiation X-ray fluorescence and deep learning: taking micro/nano polyethylene terephthalate as an example [J]. *J Hazard Mater*, 2024, 463: 132886.
- [21] CUI L, TIAN X, XIE H, et al. *Cardamine violifolia* as a potential Hg hyperaccumulator and the cellular responses [J]. *Sci Total Environ*, 2023, 863: 160940.
- [22] ABRAHAM A, PEDREGOSA F, EICKENBERG M, et al. Machine learning for neuroimaging with scikit-learn [J]. *Front Neuroinform*, 2014, 8: 14.
- [23] ZHAO J, LI Y, LI Y, et al. Selenium modulates mercury uptake and distribution in rice (*Oryza sativa* L.), in correlation with mercury species and exposure level [J]. *Metallomics*, 2014, 6(10): 1951-1957.
- [24] 高大翔, 郝建朝, 李子芳, 等. 汞胁迫对水稻生长及幼苗生理生化的影响 [J]. *农业环境科学学报*, 2008(1): 58-61.
- GAO DX, HAO JC, LI ZF, et al. Effects of Hg stress on growth and physiological and biochemical characteristics of rice seedlings [J]. *J Agro-Environ Sci*, 2008(1): 58-61.
- [25] SHAH SH, ISLAM S, MOHAMMAD F. Sulphur as a dynamic mineral element for plants: a review [J]. *J Soil Sci Plant Nutr*, 2022, 22(2): 2118-2143.
- [26] 李俊成, 于慧, 杨素欣, 等. 植物对铁元素吸收的分子调控机制研究进展 [J]. *植物生理学报*, 2016, 52(6): 835-842.
- LI JC, YU H, YANG SX, et al. Research progress of molecular regulation of iron uptake in plants [J]. *Plant Physiol J*, 2016, 52(6): 835-842.
- [27] EL-SHEHAWI A M, RAHMAN A, ELSEEHY M M, et al. Mercury toxicity causes iron and sulfur deficiencies along with oxidative injuries in alfalfa (*Medicago sativa*) [J]. *Plant Biosyst-Int J Dealing Aspects Plant Biol*, 2022, 156(1): 284-291.
- [28] HUANG Y, LI F, YI J, et al. Transcriptomic and physio-biochemical features in rice (*Oryza sativa* L.) in response to mercury stress [J]. *Chemosphere*, 2022, 309: 136612.
- [29] 程明芳, 金继运, 李春花, 等. 氯离子对作物生长和土壤性质影响的研究进展 [J]. *浙江农业科学*, 2010(1): 12-14.
- CHENG MF, JIN JY, LI CH, et al. Research progress in the effect of chloride on crop growth and soil characteristics in China [J]. *J Zhejiang Agric Sci*, 2010(1): 12-14.
- [30] 郑远, 陈兆进. 植物细胞器钙信号研究进展 [J]. *植物生理学报*, 2015, 51(8): 1195-1203.
- ZHENG Y, CHEN ZJ. Organellar calcium signaling in plants [J]. *Plant Physiol J*, 2015, 51(8): 1195-1203.
- [31] ALEJANDRO S, HÖLLER S, MEIER B, et al. Manganese in plants: from acquisition to subcellular allocation [J]. *Front Plant Sci*, 2020, 11: 300.
- [32] GUPTA N, RAM H, KUMAR B. Mechanism of Zinc absorption in plants: uptake, transport, translocation and accumulation [J]. *Rev Environ Sci Biol/ Technol*, 2016, 15(1): 89-109.
- [33] CHEN J, YANG ZM. Mercury toxicity, molecular response and tolerance in higher plants [J]. *Biometals*, 2012, 25(5): 847-857.
- [34] 窦琳琳, 张淹, 刘海滨, 等. 便携式近红外光谱仪用于东阿阿胶鉴定及特征波长筛选研究 [J]. *中草药*, 2023, 54(9): 2925-2930.
- DOU LL, ZHANG Y, LIU HB, et al. Identification of Dong'e Ejiao and its characteristic wavelength selecting using portable NIR spectrometers [J]. *Chin Tradit Herb Drugs*, 2023, 54(9): 2925-2930.
- [35] YUAN R, LIU G, HE J, et al. Classification of Lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis [J]. *Comput Electron Agric*, 2021, 182: 106043.
- [36] 李信, 薛桂澄, 柳长柱, 等. 基于信息量模型和信息量-逻辑回归模型的海南岛中部山区地质灾害易发性研究 [J]. *地质力学学报*, 2022, 28(2): 294-305.
- LI X, XUE GC, LIU CZ, et al. Evaluation of geohazard susceptibility based on information value model and information value-logistic regression model: a case study of the central mountainous area of Hainan Island [J]. *J Geomech*, 2022, 28(2): 294-305.
- [37] 张钧博, 何川, 严健, 等. 基于交叉验证的XGBoost算法在岩爆烈度分级预测中的适用性探讨 [J]. *隧道建设*, 2020, 40(S1): 247-253.
- ZHANG JB, HE C, YAN J, et al. Discussion on the applicability of XGBoost algorithm based on cross validation in prediction of rockburst intensity classification [J]. *Tunnel Construction*, 2020, 40(S1): 247-253.
- [38] HARADA M. Congenital Minamata disease: Intrauterine methylmercury poisoning [J]. *Teratology* 1978, 18(2): 285-288.
- [39] BAKIR F, DAMLUJI S F, AMIN-ZAKI L et al. Methylmercury poisoning in Iraq [J]. *Science*. 1973, 181: 230-241.

(英文编辑: 汪源; 责任编辑: 陈姣)