

基于EIS-CARS的微塑料FTIR特征光谱选择及ANN识别

时嘉悦¹, 张双圣^{1*}, 赵延杰², 强 静³, 薛 娇⁴, 丁 鹏⁴, 高懿婷¹

¹徐州工程学院环境工程学院, 江苏 徐州

²优艺(聊城)水处理有限公司, 山东 聊城

³中国矿业大学数学学院, 江苏 徐州

⁴江苏举世检测有限公司, 江苏 宿迁

收稿日期: 2026年3月17日; 录用日期: 2026年6月23日; 发布日期: 2026年7月17日

摘 要

针对运用中红外光谱建立微塑料(MPs)识别的人工神经网络(ANN)建模效率和精度低的问题, 提出了一种等间隔采样(EIS)耦合竞争自适应重加权采样算法(CARS)的特征光谱选择方法(EIS-CARS)。以20种典型MPs为研究对象, 讨论了EIS-CARS与CARS选择特征光谱的差异性, 并对比分析了特征光谱和全光谱条件下不同ANN模型对MPs的识别性能。试验结果表明: 基于EIS-CARS算法共选择44个特征波数点, 仅为CARS算法的33.85%, 但是特征光谱的选择时间减少了97.25%, 而且交互验证均方根误差(RMSECV)也略有下降; 特征波数点主要集中在1802~1385 cm⁻¹、1307~1164 cm⁻¹、1073~956 cm⁻¹、774~500 cm⁻¹四个波数段, 与MPs主要官能团的波数分布对应; 基于特征光谱构建的单层神经网络(SNN)与多层神经网络(MNN)模型均表现出较快的收敛速率和较高的识别精度, 其中MNN模型对待测MPs的正确识别率达到99.17%。基于EIS-CARS能够在不损失识别精度的情况下, 显著提升了建模效率, 特别是大幅缩减了计算成本高昂的超参数寻优时间。研究结果可为塑料的识别分类、研究MPs的环境行为及制定防治措施提供技术参考。

关键词

微塑料, 人工神经网络, 等间隔采样, 竞争自适应重加权采样, 特征光谱

Selection of FTIR Characteristic Spectra of Microplastics Based on EIS-CARS and ANN Recognition

Jiayue Shi¹, Shuangsheng Zhang^{1*}, Yanjie Zhao², Jing Qiang³, Jiao Xue⁴, Peng Ding⁴, Yiting Gao¹

*通讯作者。

文章引用: 时嘉悦, 张双圣, 赵延杰, 强静, 薛娇, 丁鹏, 高懿婷. 基于 EIS-CARS 的微塑料 FTIR 特征光谱选择及 ANN 识别[J]. 环境保护前沿, 2026, 16(7): 1211-1222. DOI: 10.12677/aep.2026.167123

¹College of Environmental Engineering, Xuzhou University of Technology, Xuzhou Jiangsu²Youyi (Liaocheng) Water Treatment Co., Ltd., Liaocheng Shandong³School of Mathematics, China University of Mining and Technology, Xuzhou Jiangsu⁴Jiangsu Jushi Testing Co., Ltd., Suqian Jiangsu

Received: March 17, 2026; accepted: June 23, 2026; published: July 17, 2026

Abstract

Aiming at the low efficiency and accuracy of artificial neural network (ANN) modeling for microplastic (MP) identification using mid-infrared spectroscopy, this study proposes a novel feature spectral selection method called EIS-CARS, which combines equal interval sampling (EIS) with competitive adaptive reweighted sampling (CARS). Using 20 typical MPs as research objects, the differences between EIS-CARS and CARS in feature spectral selection were explored, and the recognition performance of various ANNs under full and feature spectra conditions was compared. Results show that EIS-CARS selects 44 feature wavenumbers, just 33.85% of CARS, cuts feature selection time by 97.25%, and slightly reduces RMSECV. The chosen feature wavenumbers mainly focus on four ranges: 1802~1385 cm^{-1} , 1307~1164 cm^{-1} , 1073~956 cm^{-1} , and 774~500 cm^{-1} , corresponding to the wavenumber distribution of MPs' main functional groups. Both single neural network (SNN) and multilayer neural network (MNN) models built on the feature spectra show fast convergence and high recognition accuracy. The MNN model achieves a correct MP recognition rate of 99.17%. The ANN model based on EIS-CARS-selected feature spectra matches the full-spectrum ANN in correct recognition rate but with higher modeling efficiency, making it an excellent MP identification method. This study offers technical insights for plastic identification and classification, and is highly valuable for understanding MP environmental behavior and developing preventive measures.

Keywords

Microplastics, Artificial Neural Networks, Equidistant Interval Sampling, Competitive Adaptive Reweighted Sampling, Feature Spectrum

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

Open Access

1. 引言

塑料制品因其密度低、化学稳定性高、可塑性强及成本低廉等突出优势，已渗透到化妆品与个人护理品、纺织纤维、食品包装、农膜、建材和电子产品等几乎所有民生与工业领域[1]。然而，全球塑料循环利用率不足 9%，约 79% 被遗弃至自然系统，形成“累积型”污染源[2]。在光照、氧化、风化等耦合作用下，大塑料逐步碎化为粒径 $< 5 \text{ mm}$ 的微塑料(microplastics, MPs) [3]。自 Thompson 等[4]在海洋水体中提出“微塑料”概念以来，在河流、湖泊、深海，甚至两极，以及土壤、地下水和许多生物群落等不同环境介质中均检测到了微塑料的存在[5]-[9]，其可作为载体富集重金属、持久性有机污染物和病原微生物，诱发氧化应激、内分泌干扰及食物链传递风险[10] [11]。因此，快速、准确识别环境 MPs 的聚合物类型，是评估其源-汇过程、环境行为和健康风险的前提，也是实现塑料高值回收与精准管控的技术瓶颈[12] [13]。

目前 MPs 识别主要包括目视法、热分析法与光谱法[14]。目视法受主观经验与颗粒尺寸限制，误判

率高达 70% [15]; 热分析可提供聚合物单体信息, 但前处理复杂、破坏样品且耗时[16]。光谱法因无损、微区、快速等优势成为主流, 其中拉曼光谱对荧光敏感, 近红外光谱易受水分影响, 而中红外具有操作简单、灵敏度高、波数准确、重复性好等特点, 在微塑料识别检测中受到广泛应用[17], 该方法成为欧盟海洋垃圾专家组推荐的微塑料识别检测方法[18]。随着计算机科学的发展, 光谱技术与机器学习的联合使用在微塑料等物质识别检测领域得到了广泛应用[16] [19]。例如, Kaihara 等[20]基于近红外光谱构建了塑料决策树, 实现了对 18 种塑料的分类, 误差仅为 3%。然而, 中红外全谱维度高, 存在信息冗余、共线及基底漂移, 直接输入机器学习模型将导致“维度灾难”、训练耗时及过拟合风险[21]。

变量选择算法可有效去掉无用和冗余信息, 并保留关键原始波数点, 有效提升建模效率, 其中竞争自适应重加权采样(Competitive Adaptive Reweighted Sampling algorithm, CARS)因结合蒙特卡洛采样(Monte Carlo sampling, MCS)与偏最小二乘法(Partial Least Squares, PLS)回归系数, 被广泛用于土壤重金属、中药材及塑料分类[22], 但 CARS 仍需对整个光谱遍历迭代, 计算负荷随样本量与维度呈指数增长, 因此如何在保留全光谱主要信息的基础上高效选择特征光谱, 且不降低识别模型的精度是构建微塑料识别检测模型面临的一个重要问题。

基于此, 本文提出一种等间隔采样(Equal Interval Sampling, EIS)耦合竞争自适应重加权采样算法(EIS-CARS)的特征光谱选择方法, 并以 20 种典型微塑料为研究对象, 研究人工神经网络(Artificial Neural Network, ANN)模型对微塑料的识别性能。本研究的目的: ① 选择微塑料的特征光谱; ② 建立基于 ANN 及超参数优化的微塑料识别模型; ③ 评价不同 ANN 模型对微塑料的识别性能。本文旨在形成基于傅里叶红外光谱和 ANN 的微塑料高效精准识别策略, 为塑料的识别分类、研究微塑料的环境行为及制定防治措施提供技术参考。

2. 材料与方法

2.1. 实验材料与光谱采集

试验选取生活中常见的 20 种典型塑料类型, 包括丙烯腈-丁二烯-苯乙烯共聚物(Acrylonitrile Butadiene Styrene, ABS)、乙烯醋酸乙酯共聚物(Ethylene Vinyl Acetate, EVA)、聚对苯二甲酸丁二醇酯(Polybutylene Terephthalate, PBT)、聚碳酸酯(Polycarbonate, PC)、聚己内酯(Polycaprolactone, PCL)、聚乙烯(Polyethylene, PE)、聚醚砜(Polyethersulfone, PES)、聚对苯二甲酸乙二酯(Polyethylene Terephthalate, PET)、聚乳酸(Polylactic Acid, PLA)、聚甲基丙烯酸甲酯(Polymethyl Methacrylate, PMMA)、聚甲醛(Polyoxymethylene, POM)、聚丙烯(Polypropylene, PP)、聚苯醚(Polyphenylene Oxide, PPO)、聚苯硫醚(Polyphenylene Sulfide, PPS)、聚苯乙烯(Polystyrene, PS)、聚四氟乙烯(Polytetrafluoroethylene, PTFE)、聚乙烯醇(Polyvinyl Alcohol, PVA)、聚氯乙烯(Polyvinyl Chloride, PVC)、苯乙烯-丁二烯嵌段共聚物(Styrene Butadiene Styrene, SBS)、热塑性聚氨酯(Thermoplastic Polyurethane, TPU)等, 全部购自东莞特塑朗化工原料经营部, 粒径 200 目, 纯度 99.99%。

通过傅里叶红外光谱仪(Nicolet iS10, 美国尼高力公司)对微塑料进行扫描, 测试范围为 4000 cm^{-1} ~ 400 cm^{-1} , 分辨率为 4 cm^{-1} , 每个样品扫描 9 次, 即得 180 组光谱数据。

2.2. 特征光谱选择

以光谱的横坐标轴(波数)为特征光谱的选择对象, 首先, 分步采用 EIS 和 CARS 方法选择特征波数, 并通过特征波数获取对应的特征透过率, 特征波数与特征透过率构成特征光谱。然后, 对特征透过率进行标准正态变换[19], 采用基于自适应矩估计(Adaptive Moment Estimation, Adam)算法[23]的 ANN 模型进行权重和偏差的优化。

Step 1 原始光谱的等间隔采样(EIS)

原始红外光谱的横坐标波数向量为 $\mathbf{r}^0 = [r_1^0, r_2^0, \dots, r_p^0]^T$, 纵坐标透过率向量为 $\mathbf{x}^0 = [x_1^0, x_2^0, \dots, x_p^0]^T$, 其中 r_i^0 表示红外光谱中横坐标轴的一个波数点, cm^{-1} ; x_i^0 表示光谱中与 r_i^0 对应的透过率, %。在波数等间隔采样时, 假定间隔值为 N , 即以第一个波数点 r_1^0 为起始点, 间隔 N 个点后, 取 r_{N+2}^0 为第二个波数点, 依次采样, 进而构建新的波数向量 $\mathbf{r}'$ 。为了确定最优的间隔值 N , 以 $\mathbf{r}'$ 对应的透过率为输入, 以微塑料类型为输出, 建立 PLS 模型[24][25], 并进行 k 折交叉验证[26], 以交互验证均方根误差(Root Mean Square Error of Cross Validation, RMSECV)为评价指标, 确定最优的间隔值。RMSECV 最小值时对应的间隔透过率向量记为 $\mathbf{x}' = [x'_1, x'_2, \dots, x'_q]^T$ 。

k 折交叉验证是将 c 组样本数据随机且均匀地分成 k 个子集(记为 $C_1, C_2, \dots, C_k$), 依次选择其中一个子集作为测试集, 其余子集作为训练集, 建立 k 次 PLS 模型。 k 折交叉验证的 RMSECV 的计算表达式为:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^k \sum_{j=1}^{c/k} (\mathbf{d}_j^i - \mathbf{u}_j^i)^T (\mathbf{d}_j^i - \mathbf{u}_j^i)}{c \times m}}$$

当测试集为 $C_i (i=1, 2, \dots, k)$ 时, 模型输出记为 $[\mathbf{u}_1^i, \mathbf{u}_2^i, \dots, \mathbf{u}_{c/k}^i]^T$, 正确输出记为 $[\mathbf{d}_1^i, \mathbf{d}_2^i, \dots, \mathbf{d}_{c/k}^i]^T$, $\mathbf{u}_j^i$ 与 $\mathbf{d}_j^i$ 均表示 m 维列向量。

Step 2 基于 CARS 算法选择特征光谱

CARS 算法可以自适应地选择具有重要信息的样本, 并对每个样本进行加权, 以提高模型的性能和可解释性[22], 其核心过程包括: 采用指数衰减函数(EDF)剔除弱贡献变量, 结合自适应重加权采样(ARS)与 PLS 建模进行迭代优化, 最终依据最小 RMSECV 确定特征波长, 具体过程参考文献[27][28]。

RMSECV 最小值时对应的特征波数记为 $\mathbf{r}'' = [r_1'', r_2'', \dots, r_n'']^T$, 特征透过率向量记为 $\mathbf{x}'' = [x_1'', x_2'', \dots, x_n'']^T$ 。

2.3. 神经网络建模策略

为实现 ANN 建模, 以标准正态变换后的特征透过率为输入, 微塑料类型为输出建立特征光谱数据集。根据微塑料类型, 将特征光谱数据集随机分成 2 组, 命名为特征训练集和特征测试集。为减少模型的过拟合风险, 并提高模型的泛化能力, 本文设定特征测试集样本数量是特征训练集的 2 倍。

分别采用单层神经网络(Single-layer Neural Network, SNN)[29]、多层神经网络(Multi-layer Neural Network, MNN)[30][31]建立微塑料的识别模型。

(1) SNN 模型

将标准正态变换后的特征透过率向量作为输入, 微塑料类型为输出(下同), 采用 SNN 构建模型关系 $\mathbf{y} = \varphi(\mathbf{W}\mathbf{x} + \mathbf{b})$, 其中 $\mathbf{x}$ 表示 n 维特征透过率列向量; $\mathbf{y}$ 表示 m 维微塑料类型列向量; $\mathbf{W}$ 表示 $m \times n$ 的权重矩阵; $\mathbf{b}$ 为 m 维的偏差列向量; $\varphi(\cdot)$ 表示激活函数, 常用的激活函数有 Sigmoid 函数、Relu 函数、Softmax 函数等。

SNN 训练采用平方和误差函数, 通过增量规则调整权重以减小输出误差[32]。具体使用 Adam 算法迭代优化: ① 初始化权重矩阵 $\mathbf{W}$ 、偏差 $\mathbf{b}$ 及超参数(α , β_1 , β_2 , ε); ② 计算模型输出 $\mathbf{y}$ 与正确输出 $\mathbf{d}$ 之间的误差 $\mathbf{e} = \mathbf{d} - \mathbf{y}$ 与节点增量 $\boldsymbol{\delta} = \varphi'(\mathbf{W}\mathbf{x} + \mathbf{b}) \circ \mathbf{e}$, 其中 $\circ$ 表示 Hadamard 积; ③ 基于梯度的一阶和二阶矩估计动态更新权重 $\mathbf{W}$ 与偏差 $\mathbf{b}$; ④ 重复至 $\text{RMSE} \leq 0.0001$ 收敛。

(2) MNN 模型

MNN 训练采用反向传播算法, 网络结构含输入层、隐藏层和输出层, 其中 1 个隐藏层为浅层网络,

2 个及以上为深度网络[33]。以浅层网络为例(设隐藏层节点数为 m_1 , 输入层到隐藏层权重矩阵为 W_1 , 隐藏层到输出层为 W), 训练过程为: ① 初始化权重矩阵 W_1 、 W , 初始化偏差 b_1 和 b 及 Adam 算法参数; ② 前向传播计算隐藏层输出 $y_1 = \varphi_1(W_1x + b_1)$ 和模型输出 $y = \varphi_2(Wy_1 + b)$; ③ 反向传播, 先计算输出层增量 $\delta = \varphi_2'(Wy_1 + b) \circ (d - y)$, 再计算隐藏层增量 $\delta_1 = \varphi_1'(W_1x + b_1) \circ W^T \delta$; ④ 采用 Adam 算法更新规则计算 $\Delta W, \Delta W_1, \Delta b$ 和 Δb_1 , 并调整参数; ⑤ 重复迭代直至 $RMSE \leq 0.0001$ 。

为了提高 ANN 模型的性能和泛化能力, 避免过度拟合, 采用遗传算法对 ANN 超参数进行优化。

2.4. 性能评价指标

特征训练集及特征测试集均采用 RMSE 评价 ANN 模型的精度, 采用正确识别率(Accuracy)评价 ANN 模型的识别性能。以训练集为例给出各指标的计算公式:

$$RMSE = \sqrt{\frac{\sum_{i=1}^a (d_i - y_i)^T (d_i - y_i)}{a \times m}}$$

式中, $Y = [y_1, y_2, \dots, y_a]^T$ 表示训练集中 a 个样本的模型输出, $D = [d_1, d_2, \dots, d_a]^T$ 表示训练集的正确输出, y_i 与 d_i 均表示 m 维列向量;

$$Accuracy = \frac{TP}{TP + FP}$$

式中, TP 和 FP 分别代表训练集中被正确识别的样本个数与被错误识别的样本个数。

3. 结果与讨论

3.1. 原始光谱分析

各微塑料样品的平均中红外光谱如图 1。大部分样品在 $2800 \sim 3200 \text{ cm}^{-1}$ 存在芳香族和脂肪族 C-H 的伸缩振动, 但也表现出一定的差异性。在大于 3500 cm^{-1} 和小于 500 cm^{-1} 波数段内, 各样品包含的信息量较少, 而在 $3500 \sim 500 \text{ cm}^{-1}$ 波数段内含有丰富的光谱信息, 故仅在 $3500 \sim 500 \text{ cm}^{-1}$ 波数段内选择特征光谱。

3.2. 特征光谱选择

以红外光谱透过率为输入, 微塑料类型为输出, 建立 180 组光谱数据集。为了便于计算误差, 按照 1.1 节中 20 种微塑料的顺序进行编号, 将微塑料类型名称转换为编号。在 $3500 \sim 500 \text{ cm}^{-1}$ 波数段内, 波数间隔为 1.4464 cm^{-1} , 共计 2074 个波数点, 采取 EIS-CARS 选择特征光谱。

3.2.1. 等间隔采样(EIS)

设置间隔 N 的取值为 $0 \sim 19$, 在 2074 个波数点内, 分别可构建 20 组新的波数向量 $r'_0, r'_1, r'_2, \dots, r'_{19}$ 。基于 180 组光谱数据集, 以新的波数向量对应的透过率为输入, 以微塑料类型为输出, 分别建立 PLS 模型。不同间隔值下, 采用“5 折交叉验证”, RMSECV 结果见图 2。当 $N = 8$ 时, RMSECV 最小(0.0954), 因此最佳间隔值为 8, 所得样本的透过率数据共计 231 个。

3.2.2. 基于 CARS 算法选择特征光谱

基于 EIS 所得 231 个透过率数据, 采用 CARS 算法选择特征透过率。采用 MCS 进行 60 次采样, 在此过程中 RMSECV 的变化如图 3(a), 在第 22 次采样时, RMSECV 最小(0.0904), 共选择出 44 个特征波数, 其特征红外光谱如图 3(b)。在选择特征光谱过程中, 等间隔采样用时 4.190 s, 基于 CARS 选择特征光谱用时 5.906 s, 合计用时 10.096 s (计算机配置如下: Intel Core i7-6700 CPU @ 3.40 GHz, RAM 8 GB)。

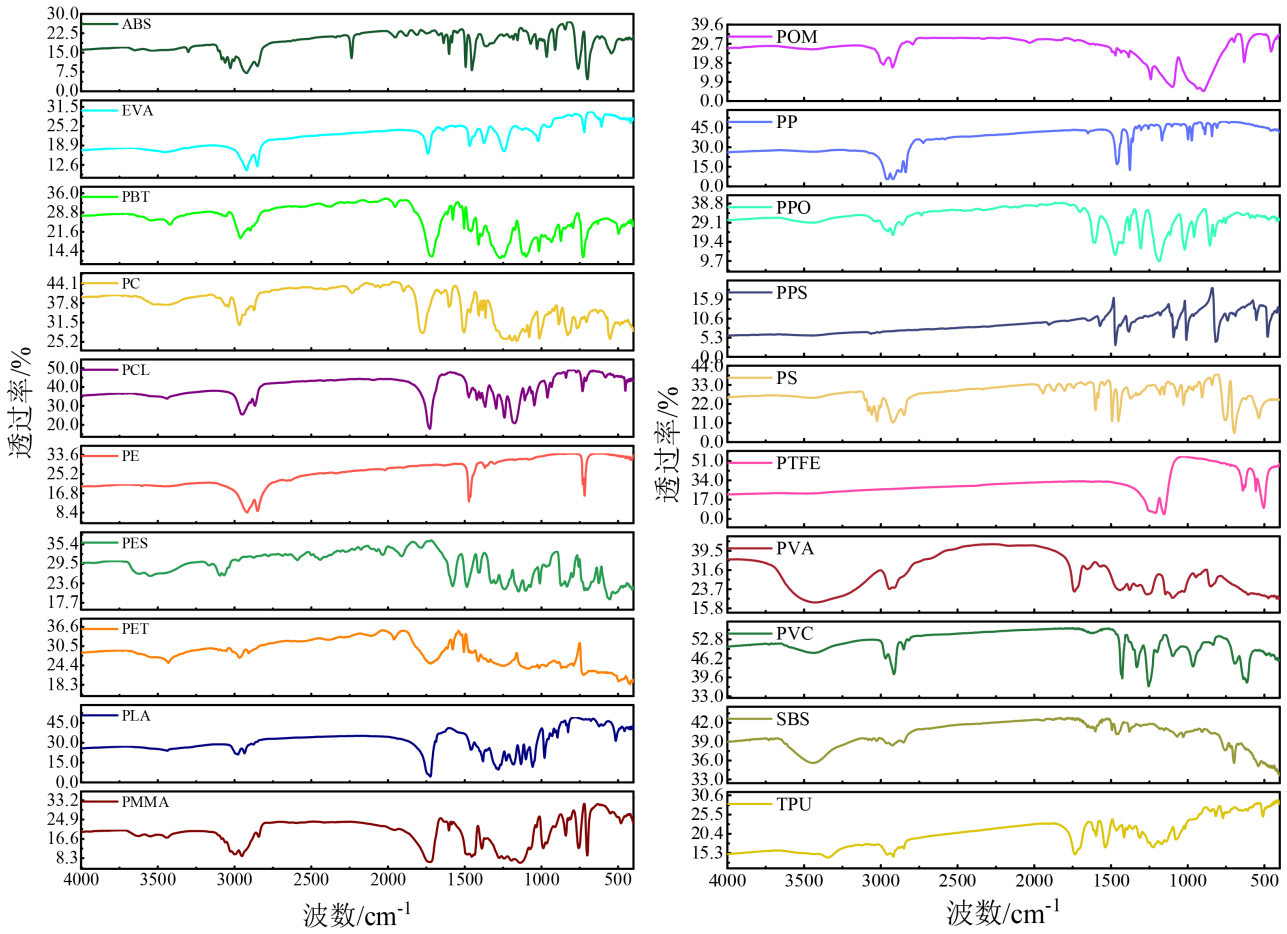


Figure 1. Average mid-infrared spectrum of microplastic samples
图 1. 微塑料样品的平均中红外光谱

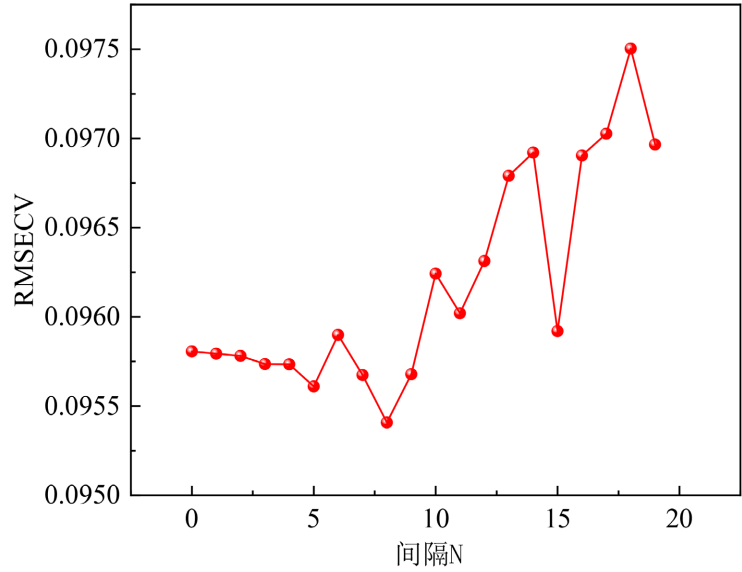


Figure 2. RMSECV varies with the change of interval N
图 2. RMSECV 随间隔 N 的变化

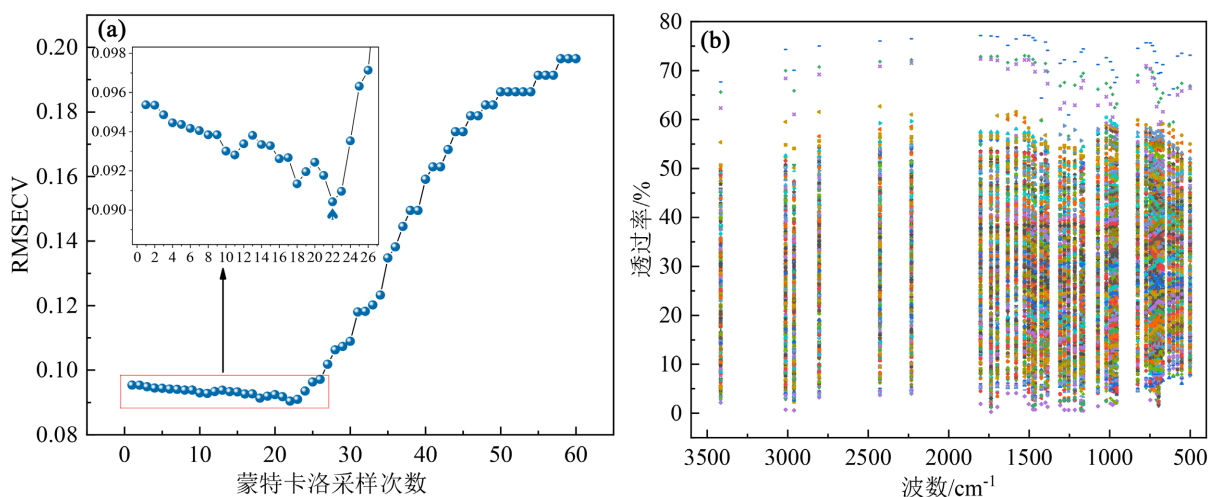


Figure 3. RMSECV variation (a) in EIS-CARS algorithm sampling process and selected feature spectra (b)

图 3. EIS-CARS 算法采样过程中 RMSECV 变化(a)及选择的特征光谱(b)

同时, 以原始红外光谱的 2074 个波数点为基础, 直接采用 CARS 算法选择特征光谱, 在 MCS 采样 60 次过程中 RMSECV 的变化如图 4(a), 在第 25 次采样时, RMSECV 最小(0.0933), 共选择出 130 个特征波数, 其特征红外光谱如图 4(b)。基于 CARS 算法选择特征光谱用时 367.570 s, 是 EIS-CARS 算法用时的 36.41 倍。

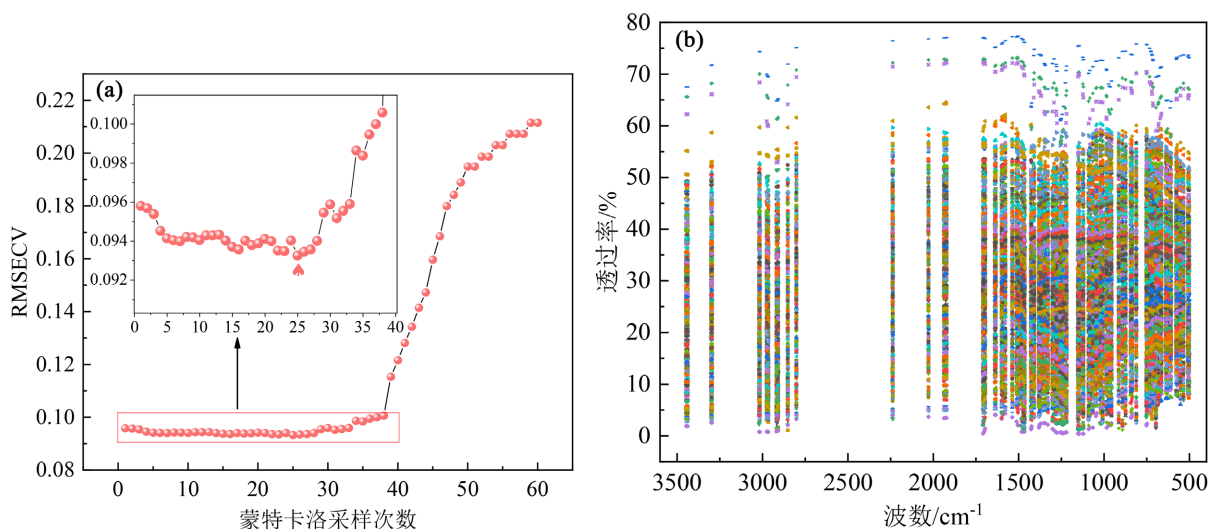


Figure 4. RMSECV variation (a) in CARS algorithm sampling process and selected feature spectra (b)

图 4. CARS 算法采样过程中 RMSECV 变化(a)及选择的特征光谱(b)

EIS-CARS 算法和 CARS 算法在最优采样条件下的 RMSECV 均小于 0.095, 且 EIS-CARS 算法的 RMSECV 更小, 表明两种方法选择的特征光谱均较可靠, 但是 CARS 算法选择的特征波数是 EIS-CARS 算法的近 3 倍, 这意味着利用 EIS-CARS 算法选择的特征光谱进行 ANN 建模的效率更高。

在特征波数分布方面, CARS 算法选择的特征波数分布相对均匀, 尤其在 1700~500 cm^{-1} 之间分布密集。EIS-CARS 算法选择的特征波数主要集中在 1802~1385 cm^{-1} 、1307~1164 cm^{-1} 、1073~956 cm^{-1} 、774~500 cm^{-1} 四个波数段, 与图 1 对比, 这四个波数段内的官能团的类型和数量丰富, 包含的信息量最大, 均是

20 种微塑料特征官能团(如 C=O、C-O-C、C-O、C=C、CF₂、S=O、C-Cl 等)所在的区域。因此 EIS-CARS 算法选择的光谱信息更具有针对性和特征性。

综上, 从 RMSECV、计算效率和官能团分布三个角度分析, 采用 EIS-CARS 算法选择特征光谱的效果更优(见表 1)。

Table 1. Feature spectra extracted by different algorithms
表 1. 不同算法选择的特征光谱

方法	特征波数(cm ⁻¹)
EIS-CARS	500.4371、552.5057、578.54、604.5743、630.6087、669.6602、682.6773、695.6945、708.7116、721.7288、734.7459、747.7631、773.7974、825.866、956.0375、969.0547、982.0718、995.089、1021.123、1073.192、1164.312、1177.329、1216.381、1255.432、1281.466、1307.501、1385.604、1398.621、1424.655、1463.707、1476.724、1502.758、1528.792、1580.861、1632.929、1698.015、1737.067、1802.153、2231.719、2426.976、2804.473、2960.679、3012.748、3416.28
CARS	501.8835、506.2226、507.6689、535.1495、552.5057、571.3083、574.201、575.6473、577.0938、607.4671、611.8062、627.716、634.9478、636.3941、637.8405、663.8748、666.7675、671.1065、679.7846、681.231、684.1237、695.6945、713.0507、714.497、718.8361、720.2824、724.6215、726.0678、733.2996、749.2094、750.6558、752.1021、811.4025、835.9905、837.4368、838.8832、843.2222、873.5956、892.3981、914.0934、956.0375、963.2693、970.501、973.3937、984.9645、986.4109、1008.106、1018.231、1019.677、1021.123、1042.819、1044.265、1047.158、1065.96、1083.316、1090.548、1120.921、1123.814、1145.509、1148.402、1152.741、1219.273、1222.166、1226.505、1238.076、1245.308、1251.093、1252.539、1253.986、1256.878、1275.681、1278.574、1280.02、1293.037、1301.715、1329.196、1345.106、1346.552、1349.445、1350.891、1368.247、1369.694、1375.479、1376.926、1379.818、1381.265、1408.745、1411.638、1424.655、1426.101、1427.548、1465.153、1468.046、1473.831、1482.509、1504.204、1505.651、1512.882、1537.47、1577.968、1593.878、1599.663、1634.376、1638.715、1696.569、1700.908、1705.247、1708.14、1922.199、1926.539、1936.663、2027.783、2237.504、2238.95、2801.581、2852.203、2908.611、2910.057、2917.289、2966.465、2967.911、2972.25、2975.143、3015.64、3018.533、3019.979、3296.232、3300.572、3437.975、3446.653

3.3. ANN 模型识别及泛化能力

3.3.1. 模型训练及超参数优化

利用 EIS-CARS 算法选择的 44 个特征波数, 分别从 2.2 节的 180 组光谱数据集中选择特征透过率, 对其进行标准正态变换, 将变换后的透过率作为 ANN 模型的输入, 20 种微塑料的类型作为输出, 构成特征光谱数据集, 即输入层节点数 $n = 44$, 输出层节点数 $m = 20$ 。从 180 组特征光谱数据集中随机选择 60 组作为特征训练集, 用于训练 ANN 模型; 剩余 120 组作为特征测试集, 测试模型对微塑料的正确识别率。分别采用 SNN 和 MNN 模型建立微塑料识别模型。

(1) SNN 模型

ANN 超参数值对模型性能和训练速度均有一定的影响。在 SNN 模型中, 需优化的参数包括 seed、bsize、 α 和 β_1 等 4 个参数, 超参数优化模型如下: $\max f(\mathbf{z})$, 其中 $\mathbf{z} = [\text{seed}, \text{bsize}, \alpha, \beta_1]^T$, 各参数取值范围设定为 $1 \leq \text{seed} \leq 100$, $1 \leq 60/\text{bsize} \leq 6$, $0 < \alpha \leq 1$, $0 < \beta_1 \leq 1$, 且 seed 和 60/bsize 取整数。 $f(\mathbf{z})$ 表示参数取 $\mathbf{z}$ 时, SNN 经过训练后在特征测试集中的正确识别率。在遗传算法中设定群体大小为 20, 进化代数数为 100, 交叉概率为 0.8, 变异概率为 0.3, 计算得到最优参数组合为: seed = 54, bsize = 12, $\alpha = 0.7718$, $\beta_1 = 0.0662$ 。运用这组参数, 建立 SNN 模型。SNN 模型的训练过程见图 5。在训练轮数 epoch 为 6 时, SNN 模型达到收敛, 耗时 2.466 s。

(2) MNN 模型

MNN 设置 2 个隐藏层, 与 SNN 需优化的参数相比, MNN 需优化的参数增加隐藏层的节点个数 m_1

和 m_2 ，因此 MNN 需优化的参数包括 seed、 m_1 、 m_2 、bsize、 α 和 β_1 等 6 个参数。

全连接 MNN 参数优化模型如下： $\max f(\mathbf{z})$ ，其中 $\mathbf{z} = [\text{seed}, m_1, m_2, \text{bsize}, \alpha, \beta_1]^T$ ，各参数取值范围设定为 $1 \leq \text{seed} \leq 100$ ， $1 \leq m_1 \leq 200$ ， $1 \leq m_2 \leq 200$ ， $1 \leq 60/\text{bsize} \leq 6$ ， $0 < \alpha \leq 1$ ， $0 < \beta_1 \leq 1$ ，且 seed， m_1 ， m_2 和 $60/\text{bsize}$ 取整数。 $f(\mathbf{z})$ 表示参数取 $\mathbf{z}$ 时，MNN 经过训练后在特征测试集中的正确识别率。由于需优化的参数增多，因此在遗传算法中设定群体大小为 40，进化代数为 200，交叉概率为 0.8，变异概率为 0.3，计算得到最优参数组合为：seed = 7， $m_1 = 94$ ， $m_2 = 73$ ，bsize = 15， $\alpha = 0.0297$ ， $\beta_1 = 0.0934$ 。运用这组参数，建立 MNN 模型。MNN 模型的训练过程见图 5。在训练轮数 epoch 为 15 时，MNN 模型达到收敛，耗时 2.619 s。

通过选择特征光谱，超参数优化后两种模型均表现出较快的收敛速率，而且对特征训练集和特征测试集中微塑料的正确识别率均达到 100% (表 2)，这说明两种模型的质量较高。

3.3.2. 模型泛化能力评估

为验证两种 ANN 模型的泛化能力，对 20 种微塑料分别重新取样，作为待测样本。重复前述的原始光谱扫描(分辨率为 4 cm^{-1} ，每个样品扫描 6 次，截取光谱范围为 $3500 \sim 500 \text{ cm}^{-1}$)，得到 120 组光谱数据。通过 EIS-CARS 算法选择的特征波数(表 1)获取特征透过率，然后对其进行标准正态变换。将变换后的透过率作为上述 ANN 模型的输入，运行模型可得到对应的微塑料类型，并与真实微塑料类型进行对比，计算正确识别率(表 2)。

3.4. 全光谱建模对比

为体现基于特征光谱的 ANN 模型优越性，采用全光谱进行 ANN 建模，且训练集对应特征训练集的全光谱，测试集对应特征测试集的全光谱。对全光谱 2074 个波数点对应的透过率进行标准正态变换，并将其作为 ANN 模型的输入，即输入层节点数 $n = 2074$ ，以 20 种微塑料的类型作为输出，即输出层节点数 $m = 20$ ，分别采用 SNN 和 MNN 建立微塑料识别模型，超参数的选择、遗传算法参数的设置参考 2.3.1 节。

(1) SNN 模型

经超参数优化后，最优参数组合为：seed = 89，bsize = 10， $\alpha = 0.0350$ ， $\beta_1 = 0.5274$ 。运用这组参数，建立 SNN 模型。SNN 模型的训练过程见图 5。在训练轮数 epoch 为 20 时，SNN 模型达到收敛，耗时 5.756 s，SNN 模型对训练集和测试集中微塑料的正确识别率均达到 100% (表 2)。

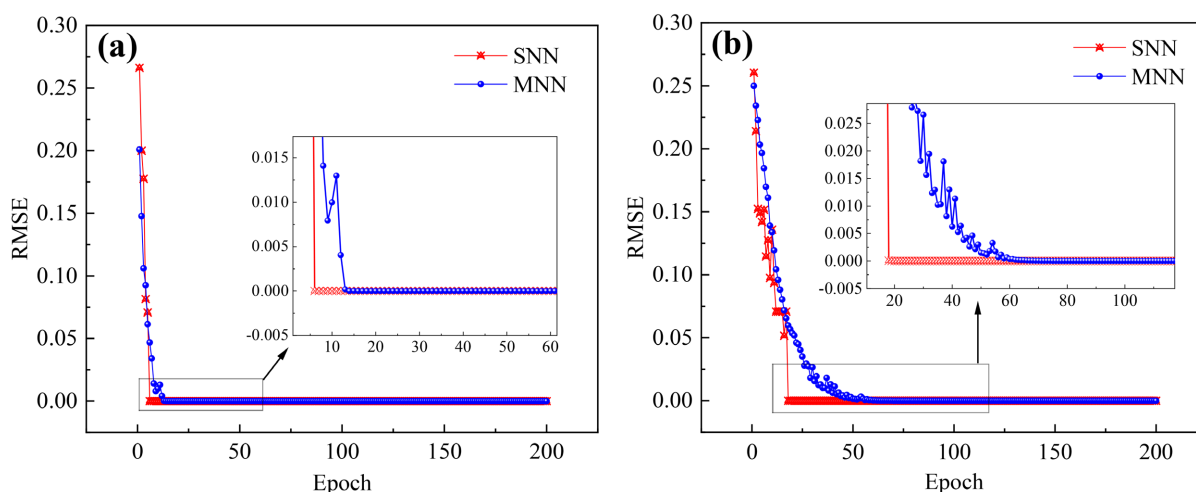


Figure 5. Training process of ANN model using characteristic spectra (a) and full spectra (b)

图 5. 特征光谱(a)与全光谱(b)条件下 ANN 模型的训练过程

(2) MNN 模型

经超参数优化后，最优参数组合为： $\text{seed} = 17$ ， $m_1 = 33$ ， $m_2 = 108$ ， $\text{bsize} = 60$ ， $\alpha = 0.0066$ ， $\beta_1 = 0.0407$ 。运用这组参数，建立 MNN 模型。MNN 模型的训练过程见图 5。在训练轮数 epoch 为 61 时，MNN 模型达到收敛，耗时 14.996 s，MNN 模型对训练集和测试集中微塑料的正确识别率均达到 100% (表 2)。

为验证全光谱条件下 ANN 模型的泛化能力，将 2.3.2 节 120 组待测样本全光谱的透过率经标准正态变换后作为 ANN 模型的输入，运行模型可得到对应的微塑料类型，并与真实微塑料类型进行对比，计算正确识别率(表 2)。

Table 2. Recognition performance and modeling efficiency of ANN models for microplastics using different data sources
表 2. 不同数据源下 ANN 模型对微塑料的识别性能及建模效率

建模数据	ANN	特征训练集			特征测试集		待测样本		建模 时间(h)
		RMSE	Accuracy (%)	训练时长 (s)	RMSE	Accuracy (%)	RMSE	Accuracy (%)	
特征光谱	SNN	0	100	2.466	0	100	0.0406	98.33	1.37
	MNN	0	100	2.619	0.0100	100	0.0288	99.17	5.82
全光谱	SNN	0	100	5.756	0	100	0.0408	98.33	3.20
	MNN	0	100	14.996	0.0064	100	0.0314	99.17	33.32

由表 2 可知，在超参数优化后，特征光谱与全光谱条件下 SNN 与 MNN 模型均表现出较快的收敛速率，而且在训练集和测试集中的正确识别率均达到 100%，表明两种神经网络模型均具备良好的识别性能。对于待测样本，特征光谱与全光谱条件下两种 ANN 模型的正确识别率相等，其中 SNN 为 98.33%，MNN 为 99.17%，表现出优异的泛化能力。基于特征光谱的 ANN 模型可达到与全光谱模型相同的识别性能，这可能与以下因素有关：(1) 通过 EIS-CARS 算法选择特征光谱可有效筛选光谱中的有效信息，保证 ANN 模型的输入达到与全光谱相同的输入效果；(2) 特征测试集设置为特征训练集的 2 倍，并采用遗传算法对超参数进行优化，有效防止了模型的过拟合，并提升其泛化能力。

模型的训练和预测效率也是模型的重要衡量标准[34]。表 2 中全光谱条件下 SNN、MNN 模型的训练时间分别是特征光谱条件下相应模型的 2.33 倍和 5.73 倍。在超参数优化过程中，特征光谱条件下，SNN、MNN 模型的求解耗时分别为 1.37 h 和 5.82 h，而在全光谱条件下两种模型的耗时分别高达 3.20 h 和 33.32 h，计算代价非常大。因此基于特征光谱建立 ANN 模型的效率更高。

综上，从识别精度和计算效率两方面综合分析，基于 EIS-CARS 选择特征光谱的 ANN 模型对微塑料的识别性能更优，并可根据识别精度与效率综合评判选择 ANN 模型。

4. 结论

(1) EIS-CARS 算法在保证原光谱主要信息的条件下，有效避免了对整个光谱数据集的遍历采样，减少计算量，在不损失识别精度的情况下，显著提升了建模效率，特别是大幅缩减了计算成本高昂的超参数寻优时间。

(2) 将特征测试集设置为特征训练集的 2 倍，采用遗传算法对超参数进行优化，可提升模型拟合效果和泛化能力。基于特征光谱构建的 SNN 和 MNN 模型对待测样本的正确识别率可达到 99.83%和 99.17%。

(3) 基于特征光谱的 ANN 模型可达到与全光谱条件下 ANN 模型相同的正确识别率，但是全光谱条件下超参数优化和模型训练的计算代价非常大，因此基于 EIS-CARS 选择特征光谱的 ANN 模型是一种

优异的微塑料识别检测方法。

基金项目

徐州工程学院大学生创新创业训练计划项目(xcx2025071); 江苏省产学研合作项目(BY20251002)。

参考文献

- [1] Wang, C., Liu, Y., Chen, W., Zhu, B., Qu, S. and Xu, M. (2021) Critical Review of Global Plastics Stock and Flow Data. *Journal of Industrial Ecology*, **25**, 1300-1317. <https://doi.org/10.1111/jiec.13125>
- [2] Geyer, R., Jambeck, J.R. and Law, K.L. (2017) Production, Use, and Fate of All Plastics Ever Made. *Science Advances*, **3**, E1700782. <https://doi.org/10.1126/sciadv.1700782>
- [3] Auta, H.S., Emenike, C.U. and Fauziah, S.H. (2017) Distribution and Importance of Microplastics in the Marine Environment: A Review of the Sources, Fate, Effects, and Potential Solutions. *Environment International*, **102**, 165-176. <https://doi.org/10.1016/j.envint.2017.02.013>
- [4] Thompson, R.C., Olsen, Y., Mitchell, R.P., Davis, A., Rowland, S.J., John, A.W.G., et al. (2004) Lost at Sea: Where Is All the Plastic? *Science*, **304**, 838-838. <https://doi.org/10.1126/science.1094559>
- [5] Corradini, F., Casado, F., Leiva, V., Huerta-Lwanga, E. and Geissen, V. (2021) Microplastics Occurrence and Frequency in Soils under Different Land Uses on a Regional Scale. *Science of The Total Environment*, **752**, Article 141917. <https://doi.org/10.1016/j.scitotenv.2020.141917>
- [6] Wang, J., Li, J., Liu, S., Li, H., Chen, X., Peng, C., et al. (2021) Distinct Microplastic Distributions in Soils of Different Land-Use Types: A Case Study of Chinese Farmlands. *Environmental Pollution*, **269**, Article 116199. <https://doi.org/10.1016/j.envpol.2020.116199>
- [7] Niu, L., Wang, Y., Li, Y., Lin, L., Chen, Y. and Shen, J. (2023) Occurrence, Degradation Pathways, and Potential Synergistic Degradation Mechanism of Microplastics in Surface Water: A Review. *Current Pollution Reports*, **9**, 312-326. <https://doi.org/10.1007/s40726-023-00262-x>
- [8] Liu, Y.Y., Ptacek, C.J., Baldwin, R.J., Cooper, J.M. and Blowes, D.W. (2020) Application of Zero-Valent Iron Coupled with Biochar for Removal of Perfluoroalkyl Carboxylic and Sulfonic Acids from Water under Ambient Environmental Conditions. *Science of The Total Environment*, **719**, Article 137372. <https://doi.org/10.1016/j.scitotenv.2020.137372>
- [9] 杨光蓉, 陈历睿, 林敦梅. 土壤微塑料污染现状、来源、环境命运及生态效应[J]. 中国环境科学, 2021, 41(1): 353-365.
- [10] Li, Y., Tao, L., Wang, Q., Wang, F., Li, G. and Song, M. (2023) Potential Health Impact of Microplastics: A Review of Environmental Distribution, Human Exposure, and Toxic Effects. *Environment & Health*, **1**, 249-257. <https://doi.org/10.1021/envhealth.3c00052>
- [11] Zeb, A., Liu, W., Ali, N., Shi, R., Wang, Q., Wang, J., et al. (2023) Microplastic Pollution in Terrestrial Ecosystems: Global Implications and Sustainable Solutions. *Journal of Hazardous Materials*, **461**, Article 132636. <https://doi.org/10.1016/j.jhazmat.2023.132636>
- [12] 吴雪, 冯巍巍, 蔡宗岐. 近红外光谱的海水微塑料快速识别[J]. 光谱学与光谱分析, 2022, 42(11): 3501-3506.
- [13] 孙彦敏, 高慧敏, 徐春祥, 等. 环境中微塑料检测技术研究进展[J]. 生态与农村环境学报, 2024, 40(1): 36-43.
- [14] 骆永明, 施华宏, 涂晨, 等. 环境中微塑料研究进展与展望[J]. 科学通报, 2021, 66(13): 1547-1562.
- [15] Hidalgo-Ruz, V., Gutow, L., Thompson, R.C. and Thiel, M. (2012) Microplastics in the Marine Environment: A Review of the Methods Used for Identification and Quantification. *Environmental Science & Technology*, **46**, 3060-3075. <https://doi.org/10.1021/es2031505>
- [16] 白润昊, 范瑞琪, 刘琪, 等. 机器学习在微塑料识别与环境风险评估中的应用研究进展[J]. 环境科学, 2024, 45(2): 1185-1195.
- [17] 李晓英, 邓文怡, 武影影, 等. 基于红外光谱分析的塑料识别方法[J]. 北京信息科技大学学报, 2011, 26(3): 49-52.
- [18] Gago, J., Galgani, F., Maes, T. and Thompson, R.C. (2016) Microplastics in Seawater: Recommendations from the Marine Strategy Framework Directive Implementation Process. *Frontiers in Marine Science*, **3**, Article ID: 319. <https://doi.org/10.3389/fmars.2016.00219>
- [19] 胡斌, 付浩, 王文斌, 等. 基于红外光谱的城市生活垃圾高值化利用深度分选[J]. 光谱学与光谱分析, 2022, 42(5): 1353-1360.
- [20] Kaihara, M., Satoh, M. and Satoh, M. (2007) Systematization Method for Distinguishing Plastic Groups by Using NIR Spectroscopy. *Analytical Sciences*, **23**, 921-924. <https://doi.org/10.2116/analsci.23.921>

-
- [21] Lu, X., Li, F., Yang, W., Zhu, P. and Lv, S. (2023) Quantitative Analysis of Heavy Metals in Soil by X-Ray Fluorescence with Improved Variable Selection Strategy and Bayesian Optimized Support Vector Regression. *Chemometrics and Intelligent Laboratory Systems*, **238**, Article 104842. <https://doi.org/10.1016/j.chemolab.2023.104842>
- [22] Li, H., Liang, Y., Xu, Q. and Cao, D. (2009) Key Wavelengths Screening Using Competitive Adaptive Reweighted Sampling Method for Multivariate Calibration. *Analytica Chimica Acta*, **648**, 77-84. <https://doi.org/10.1016/j.aca.2009.06.046>
- [23] Salem, H., Kabeel, A.E., El-Said, E.M.S. and Elzeki, O.M. (2022) Predictive Modelling for Solar Power-Driven Hybrid Desalination System Using Artificial Neural Network Regression with Adam Optimization. *Desalination*, **522**, Article 115411. <https://doi.org/10.1016/j.desal.2021.115411>
- [24] Krishnan, A., Williams, L.J., McIntosh, A.R. and Abdi, H. (2011) Partial Least Squares (PLS) Methods for Neuroimaging: A Tutorial and Review. *NeuroImage*, **56**, 455-475. <https://doi.org/10.1016/j.neuroimage.2010.07.034>
- [25] Saidi, F., Khetari, S., Yahia, I.S., Zahran, H.Y., Hidouri, T. and Ameer, N. (2022) The Use of Principal Component Analysis (PCA) and Partial Least Square (PLS) for Designing New Hard Inverse Perovskites Materials. *Computational Condensed Matter*, **31**, e00667. <https://doi.org/10.1016/j.cocom.2022.e00667>
- [26] Wong, T.T. and Yeh, P.Y. (2020) Reliable Accuracy Estimates from K-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, **32**, 1586-1594. <https://doi.org/10.1109/tkde.2019.2912815>
- [27] 李跑, 周骏, 蒋立文, 等. 窗口竞争性自适应重加权采样策略的近红外特征变量选择方法[J]. 光谱学与光谱分析, 2019, 39(5): 1428-1432.
- [28] 路皓翔, 张静, 李灵巧, 等. 最小角回归结合竞争性自适应重加权采样的近红外光谱波长选择[J]. 光谱学与光谱分析, 2021, 41(6): 1782-1788.
- [29] Martínez-Rego, D., Fontenla-Romero, O. and Alonso-Betanzos, A. (2012) Nonlinear Single Layer Neural Network Training Algorithm for Incremental, Nonstationary and Distributed Learning Scenarios. *Pattern Recognition*, **45**, 4536-4546. <https://doi.org/10.1016/j.patcog.2012.05.009>
- [30] Jayas, D.S., Paliwal, J. and Visen, N.S. (2000) Review Paper (Ae—Automation and Emerging Technologies): Multi-Layer Neural Networks for Image Analysis of Agricultural Products. *Journal of Agricultural Engineering Research*, **77**, 119-128. <https://doi.org/10.1006/jaer.2000.0559>
- [31] Abuqaddom, I., Mahafzah, B.A. and Faris, H. (2021) Oriented Stochastic Loss Descent Algorithm to Train Very Deep Multi-Layer Neural Networks without Vanishing Gradients. *Knowledge-Based Systems*, **230**, Article 107391. <https://doi.org/10.1016/j.knosys.2021.107391>
- [32] Roy, D., Panda, P. and Roy, K. (2020) Tree-CNN: A Hierarchical Deep Convolutional Neural Network for Incremental Learning. *Neural Networks*, **121**, 148-160. <https://doi.org/10.1016/j.neunet.2019.09.010>
- [33] Ososkov, G. and Goncharov, P. (2017) Shallow and Deep Learning for Image Classification. *Optical Memory and Neural Networks*, **26**, 221-248. <https://doi.org/10.3103/s1060992x1704004x>
- [34] Zhang, Q., Li, F. and Yang, W. (2024) A Deep Spectral Prediction Network to Quantitatively Determine Heavy Metal Elements in Soil by X-Ray Fluorescence. *Journal of Analytical Atomic Spectrometry*, **39**, 478-490. <https://doi.org/10.1039/d3ja00392b>