

王星程,王好雨,潘欣语,等.机器学习在环境微塑料领域的应用[J].中国环境科学,2025,45(6):3428-3440.

Wang X C, Wang H Y, Pan X Y, et al. Applications of machine learning in the environmental microplastics studies [J]. China Environmental Science, 2025,45(6):3428-3440.

## 机器学习在环境微塑料领域的应用

王星程<sup>1</sup>,王好雨<sup>1</sup>,潘欣语<sup>2</sup>,陶芳<sup>1\*</sup>,郑兴兴<sup>3</sup>,曹爽<sup>1\*\*</sup> (1.中国计量大学能源环境与安全工程学院,浙江杭州 310018; 2.山东省淄博市张店区生态环境技术服务中心,山东淄博 255000; 3.聚光科技(杭州)股份有限公司,浙江杭州 310052)

**摘要:**本文综述系统地探讨了机器学习技术在微塑料领域的应用,涵盖分类识别、定量分析与吸附性能预测.经梳理近年文献,发现如卷积神经网络(CNN)和支持向量机(SVM)等技术,对提升微塑料检测精度与效率意义重大.在分类识别中,CNN模型可精准辨别微塑料类型与形状;定量分析时,借助图像和光谱数据,机器学习能快速确定微塑料浓度.在吸附性能预测方面,基于非混合定量结构-性质关系(QSPR)的模型展现出比传统模型更高的准确性和鲁棒性.然而,当前还面临数据质量不佳、收集标注困难及模型可解释性欠缺等挑战.未来研究应聚焦数据集多元化及增强模型可解释性,以推动机器学习技术在微塑料研究中的进一步应用.

**关键词:** 机器学习; 微塑料; 分类识别; 定量分析; 吸附性能

中图分类号: X131 文献标识码: A 文章编号: 1000-6923(2025)06-3428-13

**Applications of machine learning in the environmental microplastics studies.** WANG Xing-cheng<sup>1</sup>, WANG Hao-yu<sup>1</sup>, PAN Xin-yu<sup>2</sup>, TAO Fang<sup>1\*</sup>, ZHENG Xing-xing<sup>3</sup>, CAO Shuang<sup>1\*\*</sup> (1.College of Energy Environment and Safety Engineering, China Jiliang University, Hangzhou 310018, China; 2.Ecological Environment Technology Service Center of Zhangdian District, Zibo 255000, China; 3.Focused Photonics Technology Co, Ltd. Hangzhou 310052, China). *China Environmental Science*, 2025,45(6): 3428-3440

**Abstract:** This review systematically explores the application of machine learning technology in the field of microplastics, covering classification and identification, quantitative analysis, and prediction of adsorption properties. By combing through recent literature, it has been found that technologies such as convolutional neural networks (CNN) and support vector machines (SVM) are of great significance for improving the accuracy and efficiency of microplastic detection. In classification and identification, CNN models can accurately distinguish the types and shapes of microplastics; during quantitative analysis, machine learning can quickly determine the concentration of microplastics with the help of image and spectral data. In terms of predicting adsorption properties, models based on quantitative structure-property relationships (QSPR) have shown higher accuracy and robustness than traditional models. However, there are currently challenges such as poor data quality, difficulties in collection and annotation, and a lack of model interpretability. Future research should focus on diversifying datasets and enhancing model interpretability to promote the further application of machine learning technology in microplastic research.

**Key words:** machine learning; microplastic; classification identification; quantitative analysis; adsorption performance

塑料自19世纪70年代被创造发明至今,凭借其价格低廉、重量轻、耐用性高以及延展性好等优点,迅速融入了人类生活的各个领域.到2020年,全球塑料产量达约3.61亿t<sup>[1]</sup>,预计到2050年,将有大约12000t塑料垃圾进入垃圾填埋场或自然环境<sup>[2]</sup>.然而,塑料在自然环境中难以完全降解,物理、化学以及生物老化作用只能将其分解成不计其数的微小塑料碎片,其中尺寸小于5mm的塑料碎片被定义为微塑料(MPs)<sup>[3]</sup>.微塑料的主要来源分为初级微塑料与次级微塑料.初级微塑料是指在生产过程中专门针对某些特定用途所制造的微小塑料颗粒,包括个人护

理产品、医疗领域、水性涂料、粘合剂和电子产品等<sup>[3]</sup>;次级微塑料指的是较大的塑料碎片通过紫外线辐射、热老化、生物摄食及机械作用等老化方式破碎成粒径较小的塑料碎片<sup>[4]</sup>.这些微塑料常见于陆地<sup>[5]</sup>、海洋<sup>[6]</sup>、大气<sup>[7]</sup>以及淡水环境<sup>[8]</sup>中,对生态系统与人类健康造成一系列影响<sup>[9-10]</sup>.例如,研究发现微塑料在鱼肠道中的积累会导致粘膜损伤、渗

收稿日期: 2024-12-15

基金项目: 浙江省自然科学基金资助项目(LTGS24B070005);浙江省“尖兵领雁+X”研发攻关计划项目(2024C03133)

\* 责任作者, 讲师, taofang@cjlu.edu.cn; \*\* 副教授, caoshuang@cjlu.edu.cn



异显著等特点,导致其来源与分布较为广泛<sup>[25]</sup>,因此机器学习数据的采集需要多种方法,其中主要包括实地采样、卫星遥感、摄像机拍摄以及水下探测等,此外还包括文献数据的采集。

首先,实地采样是常见的数据采集方法之一。科研人员通常会选取河流、湖泊、海洋、沙滩等特定地点采集样品,随后对样品进行初步筛选、分类和计数等前处理工作。例如,Kedzierski等<sup>[26]</sup>在2014年使用4.4m长的蝠鲼网(manta net,网孔尺寸:333 $\mu$ m;网口:16 $\times$ 60cm)从地中海水域的120个地点采集微塑料样本,经实验室二次取样后,使用衰减全反射傅里叶变换红外光谱仪(ATR-FTIR)对微塑料颗粒进行光谱分析,并将光谱数据纳入机器学习的数据库中用于训练模型<sup>[26]</sup>。虽然这种方法需要消耗大量的人力、物力和时间投入,且其覆盖范围和时间跨度相对有限,但实地采样仍被认为是一种较为可靠的数据采集方法,可以提供直接的观测和测量数据,获得更详细、更全面的信息,且具有较高的精确度。

其次,文献中的数据采集也是获取微塑料相关信息的重要方式。科研人员可利用学术数据库,如Web of Science、CNKI、PubMed等,通过设置“微塑料”、“microplastic\*”以及与研究相关的关键词,如特定环境介质(海水、淡水、土壤等)、微塑料特征(粒径、类型等),检索并筛选出符合研究需求的文献。获取文献数据后,需对其进行整理和筛选。由于不同文献的研究方法、实验条件存在差异,数据质量参差不齐,所以要根据研究目的和标准,对数据进行评估。对于数据不完整、测量方法不可靠的文献数据,需谨慎使用或予以排除。整理过程中,还需将分散在不同文献中的同类数据进行整合,使其便于后续分析。在机器学习模型的训练和验证中,文献数据发挥着重要作用。一方面,当实地采样数据不足时,文献数据可作为补充,扩大数据集规模,提高模型训练的可靠性和泛化能力。另一方面,文献数据可用于验证模型的准确性。将模型预测结果与文献中实际观测数据进行对比,若两者相符,可增强对模型的信心;若存在差异,则有助于发现模型存在的问题,进而对模型进行优化。

再次,摄像机拍摄也可以用于水体中塑料数据采集,这对于水体中微塑料数据的采集及后续数据处理也有借鉴意义。通过拍摄图像,并利用人工或计算机识别技术,可以提取相关数据并用于机器学习

模型的训练。例如,Politikos等<sup>[27]</sup>通过拖曳相机图像得到海底垃圾图像数据,并基于区域的卷积神经网络(CNN)自动检测真实世界环境中的海底海洋垃圾。与实地采样相比,该方法能够深入水下区域探测,但由于硬件设施以及环境因素(如光照条件、光吸收、光折射等)的限制,其更适用于水体透明度较高的区域。为了解决此问题,Zocco等<sup>[28]</sup>基于生成式对抗网络开发了一种水下图像增强模型,解决了一定水下成像较差的问题。摄像机拍摄具有灵活性,可以针对特定区域采集图像,并在合适的环境条件下可以获得分辨率较高的图像。但是,该方法只能单次获取特定时间和地点的图像,缺乏连续监测的能力。

最后,卫星遥感是另一种常见的数据采集方法。卫星、飞机和无人机等遥感平台上搭载了各种有源或无源机载传感器,可以对地球表面进行监测。许多研究将光学传感器(如高分辨率成像技术、多光谱与超光谱技术、荧光技术和拉曼技术)搭载于卫星、飞机和无人机等平台,获取光谱数据,用于微塑料的分类识别<sup>[29]</sup>。进一步将遥感数据与先进的机器学习相结合,可以得到更具有成本效益的解决方案。例如,Sannigrahi等利用开放的遥感数据,并结合支持向量(SVM)和随机森林(RF)两种机器学习模型进行分类分析,创建了海洋塑料的光谱特征剖面,用于区分漂浮塑料和其他海洋垃圾<sup>[30]</sup>。卫星遥感技术可以覆盖大范围地区,并进行连续、长期的监测,但其缺点也很明显,包括低分辨率和较高的数据获取成本。

综上所述,在数据采集过程中,应根据具体需求综合运用不同的方法,卫星遥感技术适用于大范围监测,摄像机拍摄便于灵活控制并获取特定区域的图像,实地采样则具有高精确度和详细信息,而文献数据采集能提供丰富的历史和多地区数据。根据实际情况和资源条件,选择一种合适的方法或结合多种方法进行数据采集,从而获得高效、准确的数据。

## 1.2 数据的预处理

从各种数据源中采集的数据可能是分散的、不完整的,甚至是存在错误的。因此,对采集的数据进行预处理是必要的步骤。预处理的目的是清理和纠正输入的数据,得到高效简洁的数据集,以便在机器学习的过程中可以更快、更准确地应用,此过程是机器学习中一个重要的阶段<sup>[31]</sup>。数据预处理主要包括数据清洗和特征工程两个方面。其中数据清洗包括数

据转换和归一化、数据集划分以及数据增强等,特征工程包括特征选择和特征提取等<sup>[32-34]</sup>。

数据清洗的目的是去除无效数据、处理缺失值、检测异常值、去除重复数据以及检查数据真实性,从而为模型的构建做好准备。Hosseinzadeh 等<sup>[35]</sup>对数据清洗技术进行了比较,定义了四个关键参数:可伸缩性、效率、数据质量和准确性以及可用性。描述了众多技术的优缺点,以帮助选择最佳的数据清理算法。数据转换和归一化旨在更好地适应于不同的模型和算法,对提高机器学习的性能有重要作用<sup>[36]</sup>。数据集划分是将原始数据集划分为训练集、验证集和测试集,以便更好地评估模型的性能和泛化能力。数据增强则通过对原始数据进行变换,如对图像进行旋转、翻转、裁剪等操作,来增强数据集的规模和多样性,从而提高模型的鲁棒性和泛化能力<sup>[37]</sup>。

特征选择和提取是对原始数据中的特征进行筛选和提取,剔除无关的特征并提取有用的特征,从而降低模型的复杂度并提高预测精度。值得注意的是, Jia 等<sup>[38]</sup>对当前主流的特征选择和提取方法进行了分析,包括基于 SVM 的启发式搜索策略方法、基于搜索策略的分支定界搜索算法以及波束搜索算法等。对于每种算法都给出了应用实例,并评价了各自的优缺点,为方法选择提供了参考价值<sup>[38]</sup>。

不同的预处理步骤对机器学习算法的准确性和最终结果有显著影响,因此在机器学习中不能忽视数据预处理的重要性。研究显示, Huang 等<sup>[37]</sup>系统地评估了数据预处理技术对最大似然法的有效性,详细分析了各种数据预处理技术及其组合的优缺点。例如,对于缺失值较多的数据集,运用平均插补法可以扩大所有机器学习方法的平衡相对误差的平均值,起到了较好的效果;而对于缺失值较少的数据集,在分类与回归树的方法下,平均插补法在平衡相对误差平均值的准确性和稳健性方面通常比列表删除法差<sup>[37]</sup>。根据研究结果发现,数据预处理技术可能会显著影响最终的预测,有时还可能对机器学习方法的预测性能有负面影响<sup>[37]</sup>。因此,为了避免错误减少误差,需要根据数据类型、任务类型以及应用场景来进行选择和调整。

综上所述,数据采集和预处理是微塑料领域研究的重要基础,机器学习的应用可以帮助人们

更好地分析和利用这些数据。通过合理的数据清洗和特征工程,可以准备出适用于机器学习的高质量数据集,并通过机器学习算法对数据进行分析 and 预测。

## 2 机器学习算法在微塑料领域的应用

### 2.1 微塑料分类识别方面的应用

微塑料的分类识别是指将采集到的微塑料样本进行分类,以便更好地了解微塑料对环境、生物体以及生态系统的影响。但由于其粒径较小、形状各异以及数量较多等原因,导致微塑料的分类识别是一个较为复杂的过程。传统的方法会先使用体式显微镜对样品进行初步分析,然后通过对选定的颗粒进行光谱测量,从而进一步得到微塑料的详细信息。在此过程中无论是体视显微镜的筛查还是光谱测量,都是耗时且费力的工作。引入机器学习算法能够对微塑料进行自动化分析,显著降低了过程的复杂程度,从而减少了人力物力的消耗。然而,环境介质的差异可能显著影响基于图像的微塑料分类识别效果<sup>[39]</sup>。例如,在海洋环境中,盐度、光线反射和水下浊度的变化可能降低图像质量,从而影响模型的识别精度<sup>[40]</sup>。相较于清洁的实验室条件,野外图像采集需要结合去噪与增强技术,以提高机器学习模型的鲁棒性<sup>[39]</sup>。Chen 等<sup>[17]</sup>的研究表明,基于光谱数据的分类模型在海洋环境中的表现受到盐度变化的显著影响。Navarro 等<sup>[15]</sup>则指出,通过引入环境校正模型,可以大幅提升复杂野外条件下的分类识别准确性。这对于基于近红外光谱或荧光染色的分类方法尤为重要,因为不同介质中的光谱响应可能发生变化,因此需要特定的校正模型。本节将对机器学习在微塑料分类识别过程的应用进行阐述,表 1 机器学习在不同模型在微塑料分类识别方面的应用表 1 总结了不同机器学习模型在微塑料分类识别方面的应用。

在微塑料分类识别中,人工神经网络(ANN)、卷积神经网络(CNN)、SVM 等机器学习方法,有助于实现自动化分类识别。其应用形式主要有两种:一是基于计算机视觉与机器学习;二是基于特定设备与机器学习。

首先,基于计算机视觉的分类识别是将计算机视觉技术和机器学习算法相结合,只需要将微塑料

放置在合适背景下进行简单的拍摄,在等待几秒钟后就能获得一份关于其形态、大小和数量等信息的报告.Lorenzo-Navarro 等<sup>[15]</sup>基于深度学习网络架构,利用分辨率为 1600 万像素或更高的数码相机或手机拍摄的照片,自动统计并分类微塑料颗粒,将其分为三种类型:碎片、颗粒和线条.此架构中包含了两个阶段:第一阶段是使用 U-net 神经网络对图像中的粒子进行分割,提取与图像中粒子相对应的像素;第二阶段对每个粒子都使用 VGG16 神经网络进行分类<sup>[15]</sup>.相比于人工识别的方法,此方法能更快得到结果,显著减少了人力物力的消耗.但此方法也有一定的局限性:一是拍摄环境会影响识别效果;二是只能将微小颗粒从图片中分类识别出,无法将塑料制品与非塑料制品分开;三是在拍摄前要对样品进行预处理,需要去除非塑料材料.Massarelli 等<sup>[16]</sup>也开发了基于 K-NN 算法的模型对微塑料进行计数和分类,将其分为四种形态(颗粒、碎片、线、纤维)和四种尺寸(500 $\mu\text{m}$ 、500~1000 $\mu\text{m}$ 、1000~2000 $\mu\text{m}$ 、2000~5000 $\mu\text{m}$ ),该方法分类微塑料碎片的精度、召回率和 F1 分数高于 0.90,表明 K-NN 分类器的良好性能.然而,这种方法也存在缺陷:在正确分类某些特定类型的微塑料时,需要使用较高比例的训练数据才能获得良好的结果,增加了一定的工作量<sup>[16]</sup>.另外,拍摄相机的分辨率越大,就可以检测到尺寸越小的微塑料,避免了细节的丢失,但较高的分辨率也会导致噪声的影响增强,因此在选择拍摄仪器时需要对此影响进行考虑<sup>[16]</sup>.例如,Navarro 等<sup>[41]</sup>用像素 4800 $\times$ 6900~9700 $\times$ 13800 的扫描仪获取图像信息,图像中产生的噪音和伪影可能与某些微塑料混淆,影响分类识别准确度.基于计算机视觉对微塑料进行分类识别的算法较为简单,且对于手机、相机及无人机拍摄的图像有很强的适应性,但由于其难以对塑料与非塑料进行辨别,需要较为繁琐的前处理,故需要进一步的研究,增强其在实际环境中的应用,从而扩大机器学习的应用范围.

其次,在基于特定设备的分类识别中,常用的设备包括近红外反射光谱(NIR)、激光诱导击穿光谱(LIBS)、扫描电子显微镜(SEM)以及衰减全反射-傅里叶变换红外光谱仪(ATR-FTIR)等.例如,Höppener 等<sup>[42]</sup>首次将机器学习与基于 SEM 的阴极发光(CL)

光谱相结合,训练 ANN 分类模型成功对 6 种微塑料进行分类.Höppener 等<sup>[42]</sup>前期为 6 种类型微塑料制作了一个聚合光谱,从聚合光谱中获取微塑料的特征并将其用于训练模型.因此,数据库中缺少关于老化微塑料的光谱特征,这是其较为明显的缺陷<sup>[42]</sup>.由于样品的颜色对光谱质量没有影响,故能对种类复杂的微塑料进行分类,且准确率较高.但在真实环境中,微塑料的环境老化会改变其特征光谱,从而导致分类准确率显著下降.针对老化影响光谱质量的问题,Renner 等<sup>[43]</sup>于 2017 年提出了一种基于光谱特征选择算法的数据库搜索方法,用于识别微塑料的光谱.该方法首先利用光谱的一阶导数来检测振动带,再通过光谱曲线来消除噪声,最终通过峰面积与数据库进行比对,识别微塑料类型<sup>[43]</sup>.此外,Michel 等<sup>[44]</sup>比较了 4 种光谱技术与机器学习结合时对商用与海洋微塑料识别的准确性,发现 ATR-FTIR、NIR、LIBS 和 X 射线荧光光谱(XRF)对商用和海洋微塑料的识别成功率分别为 99%、91%、97%、70%和 99%、81%、76%、66%.未来研究可以将 Renner 等<sup>[43]</sup>提出的基于最相关振动带的自动曲线拟合方法应用于老化微塑料光谱识别的研究中,既能使光谱不受颜色、添加剂等因素的影响,又排除环境老化对光谱质量的影响.

相较于第一种方式,使用光谱分析的优点是可以更加准确地对微塑料进行分类识别,并且能够实现更进一步的分类;然而,这种方法对设备有一定要求,更适用于未经过环境老化的微塑料,且速度较慢,成本较高.而作为代替方案——荧光染色方法可作为高通量筛查微塑料,且成本较低的一种分类识别的方法.Meyers 等<sup>[45]</sup>结合尼罗红(NR)染色颗粒的图像分析与监督机器学习分类树模型,开发了一种直接且可靠的方法来检测和识别微塑料.此模型主要分为两步:首先,通过荧光显微镜获得塑料与非塑料颗粒图像,利用 CART(分类和回归树)算法生成决策树,高精度预测颗粒是塑料还是非塑料;其次,图像采集分析在获得的 RGB 颜色信息后,使用 CART 算法生成的决策树识别 7 种聚合物类型,包括尼龙、PE、PET、PP、PS、PUR 和 PVC,两个步骤生成的决策树各不相同,但其规则均基于 NR 染色颗粒的发射光谱特征进行推断.此方法将方式一与方式二相结合,在荧光显微镜上配备有照相机,弥补了计算机视觉

技术难以将塑料与非塑料颗粒区分的缺陷,同时继承了红外光谱检测的准确性,并保留了计算机视觉检测速度快、成本效益高的优点.未来的研究可以将计算机视觉的技术进一步延伸至荧光染色方法,

使模型在检测识别微塑料的基础上,将其形状大小进行分类,同时也需要进一步研究真实环境中的自然老化微塑料的应用,从而使机器学习在微塑料分类识别中的应用更加广泛.

表 1 机器学习的不同模型在微塑料分类识别方面的应用

Table 1 Application of different models of machine learning for microplastic classification and identification

模型	微塑料类型	输入	输出	表现	目的	特征	参考文献
K 近邻 (K-NN)		相机拍摄的微塑料图像	微塑料 4 种形状类别	准确度 =92.2%	快速、自动地按颗粒、碎片、线、纤维四种形态和四种尺寸对微塑料进行计数和分类		[16]
		FTIR 光谱	微塑料类型识别结果	准确率 =90.5%	运用 K-NN 方法自动识别微塑料 FTIR 光谱	对识别 PE 等聚合物的光谱非常有效,同时需要不常见的微塑料光谱来增强学习数据库,才能达到较好识别效果.	[26]
卷积神经网络(CNN)	聚碳酸酯(PC)、聚乙烯(PE)、聚对苯二甲酸乙二酯(PET)、聚丙烯(PP)、聚苯乙烯(PS)、聚氯乙烯(PVC)	表面增强拉曼光谱(SERS)数据	微塑料混合物的鉴定结果	准确率 =99.54%	同时识别和分析 6 种常见微塑料混合物 SERS 光谱中的每个成分	未经预处理的光谱数据经过 CNN 训练,微塑料的平均识别精度高达 99.54%.CNN 能够学习复杂、大量数据的特征.局限性:微塑料种类不足、只对制备样品时混合比例相同的微塑料进行了鉴定.	[46]
SVM				准确率 =93.85%	通过分析高光谱数据	1D-CNN 模型总体表现尤其出色,分类精度方面具有显著优势.土壤中微塑料浓度较低时,SVM 在分类方面表现良好.	
BPNN	PE PP PVC	高光谱数据	不同类型农田土壤中微塑料的成分	准确率 =94.14%	分类和检测土壤中三种微塑料:PE、PP 和 PVC,以及 PP+PE、PVC+PE 等组合	局限性:研究使用的数据集较小,且土壤样品中存在的其他成分的光谱可能会影响微塑料土壤样品的光谱特征,从而影响模型	[47]
1D-CNN				准确率 =95.00%		的分类精度.	
VGG16		使用 U-Net 神经网络进行分割后的相机或手机拍摄的照片	微塑料的碎片、颗粒、线条三种形状类别	准确率 =98.11%	对数码相机或手机分辨率为 1600 万像素或更高的照片中 1-5mm 范围内的微塑料颗粒进行自动计数和分类	需要对样品进行预处理,去除有机物或沙子等非塑料材料.所提出的架构比作者在此前提出的 SMACC 系统要快得多.	[15]
DT				准确率 =87.9%	通过分析高光谱数据	CNN 的整体分类精度优于其他快速、无损地分类识别	
SVM	PE PP PVC	高光谱数据	不同类型农田土壤中微塑料成分	准确率 =85.6%	土壤中的 PE、PP、PVC	2 种,分类精度达到 90%以上.与 FT-IR 和拉曼光谱相比,该方法	[48]
CNN				准确率 =92.6%	3 种微塑料,以及 PE+PP、PP+PVC 等组合	提高了客观性并减少了大量的检测和分析时间.	
VGG16		微塑料的 SEM 图像	微塑料的纤维、碎片、珠子三种形状类别	准确率 =98.33%	根据微塑料的形状对微塑料进行分类	具有预训练卷积层和微调 FC 层的 VGG16 模型可以实现较高的形状分类分数.	[49]
SIMCA	PC、PE、PET、PP、PS、PVC、聚酰胺	傅立叶变换红外	微塑料类别分类	平均灵敏度 =0.85	对全球生产的 9 种最	SIMCA 模型对所有塑料类型都具有出色的特异性.与 SIMCA 模型相比,PLS-D 模型具有更好的	[50]
PLS-DA	酸甲酯(PMMA)、聚氨酯(PU)	高光谱图像	结果	平均灵敏度 >0.85	常见聚合物进行分类	灵敏度.	
RF	PE、PP、PMMA、PS、聚丙烯腈(PAN)	FTIR 图像	微塑料类别分类结果		使用基于 RF 模型的分	RF 分类器在保持高吞吐率的同时产生了高精度.	[51]
					类来快速识别大型 HSI 数据集中的 MP		

续表 1

模型	微塑料类型	输入	输出	表现	目的	特征	参考文献
PLS-DA	PVC、PC、PS、PP、低密度聚乙烯(LDPE)、高密度聚乙烯(HDPE)	衰减全反射中红外(ATR-MIR)光谱	微塑料类别分类结果,在鱼类中识别不同微塑料种类	掺假样品和对照样品的分类正确率为100%	对人工掺假鱼类和海鲜样品中不同类型微塑料的存在进行分类	当单独分析掺假鱼样品时(分析中不使用对照样品),只有受HDPE 污染的样品才能100%正确分类.	[52]
SIMCA	PA-6、PE、PP、PET 和 PS	近红外高光谱成像(HSI-NIR)	微塑料类别分类结果,从沙子中分类识别不同微塑料种类	灵敏度>99%	通过最少的样品制备自动识别海滩沙子中的微塑料	颗粒的尺寸、颜色、风化、结晶度和密度等因素对光谱变异性有很大影响,从而影响分类性能.通过使用光谱技术(NIR)消除了目视检查的主观性,并使用SIMCA 提供快速识别.	[53]

最后,识别微塑料的训练模型有两种训练方法,一是监督学习,可以获得准确度较高的预测结果;二是无监督学习,可能可以揭示一些隐藏信息.例如,在Massarelli 等<sup>[16]</sup>开发的模型中,首先将监督学习与专家提供的分类结果进行比较,发现通过 K-NN 算法进行的自动分类计数与专家所提供的结果基本一致,且准确率始终高于 0.9,取得了令人满意的结果<sup>[16]</sup>.接着,将无监督学习与专家提供的分类结果进行比较,发现 K-均值算法在分类过程中倾向于采用与监督学习完全不同的方式,这种差异可能揭示了一些潜在的隐藏信息<sup>[16]</sup>.因此,在模型训练时,结合监督学习和无监督学习的结果进行对比分析,可能会为微塑料识别提供新的反馈,进一步提升分类精度和揭示数据中潜在的模式.

综上所述,微塑料的分类识别旨在通过高效准确的分析技术,揭示其形态、大小和种类等特征,进而评估其环境和生态影响.传统显微镜和光谱方法虽能提供详细信息,但耗时费力,而机器学习技术通过计算机视觉或光谱数据结合算法如 ANN、CNN、SVM、K-NN 等实现自动化分类,大幅降低了复杂度.基于图像的识别方法具有高效性,但易受拍摄环境和样品预处理影响;基于光谱的识别方法则更为精准,适合未老化样品但成本较高.结合荧光染色和机器学习的方法综合了速度与准确性的优势,同时克服了部分技术局限.未来研究需针对真实环境中老化微塑料的光谱特征和自然噪声,优化模型的鲁棒性,并结合监督与无监督学习,挖掘潜在模式,以进一步提升微塑料分类识别的性能和适用性.

## 2.2 微塑料定量分析方面的应用

微塑料的定量分析指的是通过科学的方法确定微塑料数量、大小等参数,从而进行详细分析和预测.

机器学习在微塑料定量分析方面的应用本质上是在分类识别的基础上再对其光谱、图像等信息进行进一步处理分析,得到需要的定量分析结果.相较于微塑料的分类识别,微塑料的准确计数更具挑战性,通常需结合多种算法进行分析,才能获取更接近真实值的结果.微塑料在不同环境介质中的浓度和形态分布特征存在显著差异<sup>[54]</sup>.在淡水和海洋环境中,这种差异尤为突出.例如,在淡水环境中,微塑料粒径较小且表面较光滑;而在海洋环境中,因盐析效应可能发生团聚,导致模型在定量分析时误差增大<sup>[55]</sup>.Chen 等<sup>[17]</sup>的研究指出,在不同介质条件下训练独立模型有助于提高预测精度.因此,模型训练中需引入介质特异性数据集,并优化算法以适应复杂环境条件,从而提升模型的实际应用价值.本节将对机器学习在微塑料定量分析过程的应用进行阐述.

首先,利用机器学习算法对各类微塑料图像进行分析是一种可以快速准确计算微塑料的数量和尺寸分布的方法.微塑料图像主要包括普通图像、显微镜图像以及光谱图像等.使用显微镜根据微塑料的颜色、尺寸、亮度、形态对其进行量化是微塑料研究中常用的技术<sup>[56-57]</sup>,因此,基于图像分析的自动化计数十分重要,简化了在量化微塑料时繁琐的体力劳动,提升研究效率.

由于近年来机器学习的快速发展,机器学习的方法已经广泛应用于生物医学领域的显微镜图像分析,但其在微塑料领域的应用相对较少<sup>[58]</sup>.在微塑料领域,样品分析的第一步常常是通过视觉识别或借助光学显微镜对收集到的微塑料进行分类识别.然而,由于微塑料粒径小、透明度高等特点,使用上面两种方式进行识别时误报率较高,而通过 SEM 图像能更准确高效地对微塑料进行分类识别,这也为

微塑料的定量分析奠定了良好的基础<sup>[59-61]</sup>。常见的方法是先通过 SEM 对微塑料进行成像,然后使用机器学习算法进行图像分析。例如,Shi 等<sup>[49]</sup>通过 SEM 对微塑料进行成像后,使用 U-Net 和 MultiResUNet 算法进行语义分割,从而实现对微塑料的识别,但对于缠结的塑料纤维,无法准确计算接触、重叠或密集微塑料的数量。因此,他们进一步使用 Mask-RCNN 算法进行实例分割,获得准确的数量。

然而,在从共享边界框中检测多个对象并生成准确掩模时存在问题。为解决此问题,借鉴 Chen 等<sup>[62]</sup>的方法,在实例分割中使用了像素嵌入。通过带有像素嵌入的双头 U-Net 实例分割框架来进行分割,最终得到 161 个微塑料纤维,与真实值 112 个相当,高估主要是由不准确的嵌入向量引起的<sup>[49]</sup>,这可以通过更多的训练数据集或改进模型算法来降低误差。在微塑料定量分析中,图像分割是实现精准量化的关键基础,只有准确分割出微塑料,才能有效开展后续的量化工作。为了提高图像分割的准确性,有众多算法及方法都得到了应用。例如,Khadangi 等<sup>[63]</sup>提出的 EM-net 算法;Haberl 等<sup>[64]</sup>提出的一种基于云的深度卷积神经网络的即用型图像分割解决方案—CDeep3M;以及 Matuszewski 等<sup>[65]</sup>通过参数化架构并减少 CNN(U-Net)的可训练权重数量来提高其泛化能力;Wegmayr 等<sup>[66]</sup>则利用深度像素嵌入来分离缠结的纤维,其性能比 Mask-RCNN 更好;另外,Lorenzo-Navarro 等<sup>[15]</sup>将大图像分割成重叠的块,通过 U-Net 进行分割,然后连接成最终结果。这种利用机器学习算法对显微镜下的微塑料图像进行分析的方法,可以提供准确的微塑料数量和尺寸分布数据,并且有望通过不断的算法优化和更多的训练数据集,进一步提高分析的准确性。

其次,基于光谱的定量分析通常非常具有挑战性,这主要是由于光谱数据的复杂性和多样性。在微塑料光谱图像的定量分析中,常见的方法之一是 FTIR 光谱。这些方法通常先使用成像技术对微塑料光谱进行成像,然后使用机器学习算法对获得的光谱图像进行分析。举例来说,da Silva 等<sup>[50]</sup>基于焦平面阵列(FPA)的傅里叶红外光谱成像技术,获得高光谱图像后,先使用偏最小二乘判别分析(PLS-DA)与类比软独立建模(SIMCA)模型,对聚合物进行了分类,其次根据 PLS-DA 模型预测的塑料混合物图像样本进一步对

微塑料的粒子丰度和大小进行了量化。然而,da Silva 等<sup>[50]</sup>也发现光谱质量在朝向粒子边缘方向有所衰减,原因是颗粒厚度通常会随颗粒长度变化而改变光路,同时颗粒的不规则形状和容易受污染等特性也会影响光谱的质量。光谱质量对于微塑料粒子的边缘轮廓的识别至关重要。因此,光谱质量的优劣会直接影响定量分析结果。为了降低此类因素的影响,Yan 等<sup>[67]</sup>在对现有机器学习的微塑料识别方法进行系统评估的基础上,提出了一种集成学习算法,通过神经网络聚合 5 种 ML 算法的输出,避免了单独的 ML 模型的性能较差的问题。研究显示,基于 FPA 的 FTIR 成像技术可以加快光谱分析速度,同时记录多达数千个光谱,并能快速对大范围区域进行化学成像<sup>[68]</sup>,因此在微塑料的定量分析中得到广泛应用<sup>[51,69]</sup>。例如,Simon 等<sup>[69]</sup>提出了一种使用 FTIR 高光谱成像量化废水处理厂的微塑料质量及其去除率的方法。Hufnagl 等<sup>[51]</sup>也基于 FPA-FTIR 成像技术,通过应用随机决策森林分类器区分了不同的聚合物类型,高精度地测量它们的丰度和尺寸分布。与通过数据库搜索方式识别微塑料光谱并与现有数据库进行对比的方法不同<sup>[43]</sup>,Hufnagl 等<sup>[51]</sup>提出的识别方法是一种更快捷的替代方案,即基于模型对 FTIR 图像进行分类从而快速识别微塑料。其与数据库搜索之间的主要区别在于使用实际数据的多变量模型来决定类型的从属关系,而不是仅使用参考数据。

综上所述,微塑料的定量分析通过科学方法确定其数量和尺寸等参数,其中机器学习的应用主要集中在图像和光谱信息的处理上。基于显微镜或 SEM 的图像分割技术是实现微塑料定量分析的关键,近年来广泛采用如 U-Net、Mask-RCNN 和 EM-net 等深度学习算法,但颗粒重叠等问题仍需通过模型优化和扩展数据集来解决。在光谱分析方面,FTIR 技术结合偏 PLS-DA 和随机森林等机器学习方法,已被用于聚合物分类、数量和尺寸分布的高效分析,但光谱质量对边缘识别的影响需通过多变量模型和集成学习优化。总体而言,借助机器学习和成像技术,微塑料的图像与光谱分析能够显著提升定量分析的效率 and 准确性,未来随着技术和算法的不断进步,该领域研究将更加精准可靠。

### 2.3 微塑料分布以及吸附性能的应用

在当前科技迅速发展的背景下,利用机器学习模

型来预测微塑料的丰度、性质以及分布情况等方面已经成为前沿研究领域.与上文所述的微塑料定量分析的内容有所不同,前者侧重于利用微塑料的成像信息对其进行分析,而后者重点更多的是研究输入变量参数(如温度、环境介质、pH值等)与机器学习模型的结合.环境介质的物理化学条件(如pH值、盐度、温度等)对微塑料的吸附性能具有显著影响,这在模型训练中至关重要<sup>[70]</sup>.例如,Zhu等<sup>[19]</sup>研究发现,海水中高盐度条件可能显著降低PVC对疏水性污染物的吸附能力,而

在淡水中,吸附能力随着pH值变化显著增强.为适应不同介质的影响,机器学习模型可通过引入多变量交互项(如pH值与温度的联合效应)优化预测性能.此外,Chen等<sup>[17]</sup>提出的基于集成学习的算法框架成功整合多种环境因子,提高了模型的鲁棒性和适用性.因此,在吸附性能预测中,应单独构建针对不同环境介质的分配系数预测模型,并通过多变量交互分析明确关键驱动因素.表2对机器学习在微塑料来源分布以及吸附性能方面的应用进行了总结.

表2 机器学习在微塑料分布及吸附能力方面的应用

Table 2 Application of machine learning to microplastic distribution and adsorption capacity

模型	输入	输出	表现	目的	特征	参考文献
RF			准确率=67.8%		XGBoost在该数据集上的表现优于GBDT和RF,这是因为	
GBDT	以海洋微塑料样本的16个特征为输入	6类微塑料的丰度区间	准确率=69.2%	基于机器学习预测海洋表层水中微塑料丰度	boosting树本质上对过度拟合具有鲁棒性,并且与GBDT相比,XGBoost优化了目标函数.但由于缺乏深层海水的数	[71]
XGBoost			准确率=71.9%		据,研究依然存在局限性	
多层感知器(MLP)	原始反应性数据(温度和化学描述符)	不同温度下常见微塑料和环境介质(水或海水)之间的分配系数	使用线性模型足以可靠地预测微塑料的吸附能力	基于非混合定量结构-性质关系(QSPR)的模型来预测不同条件下的微塑料-环境介质吸附系数	模型涵盖了更多数量和更多样化的有机化合物,从而能够更全面的应用到不同领域	[19]
SVM-线性						
RF	特定pH值条件下的正辛醇/水分布系数和7个化学描述符(包括正辛醇/水分配系数)			基于机器学习的预测模型确定PE、PP和PS在各种特定水环境中对有机污染物的吸附能力	无论为模型开发选择什么变量,总有一些机器学习模型可以提供良好的结果.但输入变量的增加并不意味着模型结果的改善	[20]
SVM	logD、共价酸度 $\epsilon_a$ 、共价碱度 $\epsilon_b$ 、相对分子质量 $M_w$ 、分子体积 $V$ 、平均分子极化率与分子体积之比 $\pi$ 、最负原子电荷 $q^-$ )	不同的微塑料/水分配系数				
ANN						
RF	分布在海洋中的海洋微塑料的数据	全球微塑料分布图	准确率<93.33%	基于9445个样本的机器学习模型创建一张用于预测海洋微塑料污染的全球地图	使用十倍交叉验证可以来估计随机森林算法的性能并避免过度拟合.样本数量有限阻碍了模型的预测	[17]
TPOT					使用H <sub>2</sub> O AutoML算法得到的	
GAMA	消化时间、微塑料类型、微塑料直径、微塑料浓度、底物类型和浓度	微塑料厌氧消化过程中累积的甲烷产量	AutoML算法根据收集的微塑料数据集成功预测了甲烷的产量	应用4种著名的AutoML算法自动生成优化的ML模型,用于根据微塑料和基	基于梯度提升机(GBM)的模型,在预测甲烷产量方面取得了高质量的结果.自动机器学习进一步减轻了	[72]
AutoGluon					质数据预测甲烷产量	
H <sub>2</sub> O AutoML					研究所耗费的人力、物力以及时间,降低了成本	
Ward's method 聚类	每种聚合物浓度与总MP浓度	MP的主要来源		基于聚类分析模型对微塑料的来源进行简单高效的预测	基于聚类算法对微塑料的来源进行分析,能够为制定环境保护策略打好基础	[18]
Mask-RCNN	微塑料纤维的显微镜图像	微塑料的尺寸及数量	平均相对计数差异=(7%±2%) 平均相对长度差异=(11%±3%)	基于机器学习模型成功对微塑料的数量及长度进行分析	人工计数在纤维计数方面的优势较为明显,而在测量纤维长度时则几乎不显著	[66]

首先,大多数微塑料丰度的预测仍然依赖于统计学和基础的模型.然而,Zhen 等<sup>[71]</sup>首次将机器学习用于微塑料丰度预测中,开发并比较了 RF、梯度提升决策树(GBDT)和极限梯度提升(XGBoost)3 种集成学习模型预测不同海域表层水中微塑料丰度分布.结果显示,XGBoost 的模型的预测性能最好,总准确率为 0.719,为微塑料丰度分布的预测提供了新的思路<sup>[71]</sup>.此外,Chen 等<sup>[17]</sup>提供了另一种可行的机器学习模型来预测微塑料的分布情况.Chen 等<sup>[17]</sup>基于 9445 个样本使用 RF 模型创建了用于预测海洋微塑料的全球地图,模型将紫外线(UV)辐射的变化、海平面上升(SLR)和地表温度上升作为预测模型的重要因素,从而达到更贴近实际的预测结果.但由于样本数量和调查区域的有限,不同研究的采样和分析方法不一致,使得不同来源的数据进行比较变得困难,这是未来需要去解决的问题.

其次,预测微塑料的吸附能力通常涉及微塑料与环境介质之间的分配系数预测.Zhu 等<sup>[19]</sup>开发了一种基于非混合定量结构-性质关系(QSPR)的模型,使用 8 种机器学习算法来预测有机污染物在固相(PVC、PP、PE)和水相(海水、淡水)中的分配系数,其中包括多层感知器(MLP)和极限学习机(ELM),通过其预测不同温度下的分配系数,并运用遗传算法(GA)、粒子群优化算法(PSO)和 AdaBoost 算法对模型进行优化,提出了一种更准确预测分配系数的混合智能模型.尽管在这些研究中使用了多种建模方法模拟了不同类型的化合物,但结果依然显示出相似的预测能力.例如,Astray 等<sup>[20]</sup>开发了 RF、SVM 和 ANN 三种机器学习模型,使用不同的输入配置,包括分子质量、正辛醇/水的分配系数等参数,预测不同微塑料(PE、PS、PP)在不同水域中的吸附能力,对比了两种输入配置下 3 种机器学习模型的预测效果,结果显示,无论为模型开发选择的变量如何,总有一些机器学习模型可以提供良好的结果,并且一味增加输入变量并不意味着模型的改善.

综上所述,利用机器学习模型预测微塑料的性质、丰度以及分布情况是当前研究的热点之一.不同于传统定量分析,这类研究侧重输入变量与模型的结合及其对吸附性能的影响.研究表明,理化条件显著影响吸附性能,多变量交互项的引入可优化模型表现.基于集成学习的算法进一步提升了预测的鲁

棒性和适用性.在吸附能力预测中,多种机器学习模型(如 RF、SVM、ANN)均表现良好,但增加输入变量未必会改善效果.此研究有望为微塑料研究提供更准确的预测结果,进而为环境保护和可持续发展提供科学依据.然而,数据收集和数据分析标准化等问题仍需深入探讨与解决,以确保研究成果的可靠性和可持续性.

### 3 机器学习在微塑料领域的优势和挑战

有关机器学习在微塑料领域的优势总结为以下四点:首先,高效处理大量数据,机器学习可以快速处理大规模的复杂数据集,通过对大量样本进行训练和学习,能够有效地识别和分类微塑料颗粒,这有助于快速分析和评估微塑料在不同环境中的分布和影响;其次,高精度的分析,机器学习算法利用各种数据特征和模式,通过训练模型可以实现对微塑料的高精度分类、识别和定量分析;第三,自动化的数据处理,利用机器学习算法可以实现对微塑料的图像、光谱以及其他相关数据自动化处理和分析,节省了大量的时间和人力,提高了研究效率;最后,高适应性,机器学习可以应用到不同的微塑料研究场景中,包括水下探测、卫星遥感、野外采样等<sup>[73-74]</sup>.

同样,机器学习在微塑料领域的应用也面临一些挑战,主要包括以下 3 点:首先是数据质量和多样性,采集到的微塑料数据质量参差不齐,可能存在噪声和不确定性,这会影响机器学习的准确性和泛化能力;其次是数据的收集和标注,机器学习需要大量的标记数据进行训练,但样本的收集和标记是一项复杂且耗时的任务,且目前微塑料数据的采集和处理还相对不充分,需要更多的数据支持和完善,以保证机器学习算法的训练和应用<sup>[75]</sup>;最后是模型的可解释性,机器学习算法通常被认为是“黑盒子”,即其内部工作机制难以被解释清楚,这可能限制了其在微塑料领域的广泛应用<sup>[76]</sup>.

综上所述,机器学习在微塑料领域的应用具有处理大量数据、高精度的分类和识别以及自动化数据处理等优势.但同时也面临数据收集和标注的困难、数据质量和多样性的挑战,以及模型解释性和可解释性的需求.未来,随着技术的进一步发展和数据的积累,机器学习在微塑料领域的应用有望取得更大的突破和应用.

## 4 结论及展望

### 4.1 结论

**4.1.1 分类识别** 在分类识别方面,SVM、PLS-DA、RF、KNN等算法适用于样本小、数据稀疏的场景,通过有效的数据预处理,能显著提升分类准确率.其中,VGG16算法凭借高准确率和快速处理速度,在微塑料形状分类中优势明显;CNN算法则在光谱数据鉴定时,展现出远超SVM、BPNN、K-NN等算法的分类精度,为微塑料类型的精准判断提供了有力支持.

**4.1.2 定量分析** 在定量分析领域,机器学习聚焦微塑料来源、丰度和尺寸参数.图像分析方面,虽存在对缠结塑料纤维计数误差,但已为定量分析奠定了基础;光谱分析中,基于FPA的FTIR成像技术结合机器学习算法,虽受光谱质量影响,但能实现微塑料粒子丰度和大小量化,不过整体研究数量有限,需创新方法提高分析精度与效率.

**4.1.3 吸附性能预测** 基于非混合定量结构-性质关系(QSPR)的模型在处理复杂化学描述符和预测有机污染物分配系数时,展现出比RF、SVM和ANN等传统模型更高的预测准确性和鲁棒性.此外,尽管RF和ANN等模型在吸附能力预测中表现良好,但其泛化能力和对变量选择的依赖性较强.然而,QSPR模型的改进空间主要体现在以下几点:一是需引入更多环境介质特征(如pH值、盐度和温度)以提高模型对复杂环境的适应性;二是优化特征工程方法,挖掘新的关键化学描述符;三是结合深度学习和集成学习技术,如XGBoost或AutoML框架,以进一步提升预测性能和模型稳定性.通过这些改进,QSPR模型有望在吸附性能预测中发挥更重要的作用,为微塑料污染治理提供可靠的科学依据.

### 4.2 展望

**4.2.1 数据质量提升** 当前微塑料数据质量参差不齐,严重影响机器学习模型的准确性与泛化能力.后续应建立统一的数据采集标准和规范,加强对数据采集过程的质量控制,减少噪声和不确定性.同时,采用先进的数据清洗和增强技术,扩充数据的多样性,构建高质量的数据集,为模型训练提供坚实的数据基础.

**4.2.2 模型可解释性突破** 机器学习算法“黑盒

子”的特性限制了其在微塑料领域的广泛应用.未来需开发新型可解释性模型,或探索有效的模型解释工具,如基于特征重要性分析、模型可视化等方法,深入理解模型决策过程,提高模型的可信度和可解释性,促进其在实际应用中的推广.

**4.2.3 环境介质适应性优化** 鉴于环境介质对微塑料行为影响显著,应针对淡水、海水、土壤等不同介质的物理化学特性,构建具有针对性的机器学习模型.深入研究不同介质中微塑料的特征变化规律,开发高鲁棒性算法,提升模型在复杂环境下的适应性和预测精度,为准确评估微塑料的环境风险和制定有效的污染治理策略提供技术保障.

**4.2.4 技术融合创新** 目前机器学习应用滞后于研究,亟待发展与机器学习相结合的高通量检测识别和定量技术.探索将机器学习与纳米技术、光谱技术、成像技术等深度融合,实现对微塑料的快速、精准检测和分析,为全面深入理解微塑料的环境行为提供有力手段,推动环境保护政策的科学制定与有效实施.

### 参考文献:

- [1] Yang H, Yumeng Y, Yu Y, et al. Distribution, sources, migration, influence and analytical methods of microplastics in soil ecosystems [J]. *Ecotoxicology and Environmental Safety*, 2022,243:114009.
- [2] Geyer R, Jambeck J R, Law K L. Production, use, and fate of all plastics ever made [J]. *Science Advances*, 2017,3(7):e1700782.
- [3] Lin Z, Jin T, Zou T, et al. Current progress on plastic/microplastic degradation: fact influences and mechanism [J]. *Environmental Pollution*, 2022,304:119159.
- [4] Dissanayake P D, Kim S, Sarkar B, et al. Effects of microplastics on the terrestrial environment: a critical review [J]. *Environmental Research*, 2022,209:112734.
- [5] Kim S K, Kim J S, Lee H, et al. Abundance and characteristics of microplastics in soils with different agricultural practices: importance of sources with internal origin and environmental fate [J]. *Journal of Hazardous Materials*, 2021,403:123997.
- [6] Li J, Gao F, Zhang D, et al. Zonal distribution characteristics of microplastics in the southern indian ocean and the influence of ocean current [J]. *Journal of Marine Science and Engineering*, 2022, 10(2):290.
- [7] Ding J, Sun C, He C, et al. Atmospheric microplastics in the northwestern pacific ocean: distribution, source, and deposition [J]. *Science of The Total Environment*, 2022,829:154337.
- [8] Tsering T, Sillanpää M, Sillanpää M, et al. Microplastics pollution in the brahmaputra river and the indus river of the indian himalaya [J]. *Science of The Total Environment*, 2021,789:147968.
- [9] Barboza L G A, Dick Vethaak A, Lavorante B R B O, et al. Marine microplastic debris: an emerging issue for food security, food safety and human health [J]. *Marine Pollution Bulletin*, 2018,133:336-348.

- [10] Zhang Z, Gao S H, Luo G, et al. The contamination of microplastics in china's aquatic environment: occurrence, detection and implications for ecological risk [J]. *Environmental Pollution*, 2022,296:118737.
- [11] Qiao R, Deng Y, Zhang S, et al. Accumulation of different shapes of microplastics initiates intestinal injury and gut microbiota dysbiosis in the gut of zebrafish [J]. *Chemosphere*, 2019,236:124334.
- [12] Li J, Yang D, Li L, et al. Microplastics in commercial bivalves from china [J]. *Environmental Pollution*, 2015,207:190-195.
- [13] Pandey B, Pathak J, Singh P, et al. Microplastics in the ecosystem: an overview on detection, removal, toxicity assessment, and control release [J]. *Water*, 2022,15(1):51.
- [14] Lee S, Kang K K, Sung S E, et al. Toxicity study and quantitative evaluation of polyethylene microplastics in icr mice [J]. *Polymers*, 2022,14(3):402.
- [15] Lorenzo-Navarro J, Castrillón-Santana M, Sánchez-Nielsen E, et al. Deep learning approach for automatic microplastics counting and classification [J]. *Science of The Total Environment*, 2021,765:142728.
- [16] Massarelli C, Campanale C, Uricchio V F. A handy open-source application based on computer vision and machine learning algorithms to count and classify microplastics [J]. *Water*, 2021,13 (15):2104.
- [17] Chen B, Zhang Z, Wang T, et al. Global distribution of marine microplastics and potential for biodegradation [J]. *Journal of Hazardous Materials*, 2023,451:131198.
- [18] Kameda Y, Yamada N, Fujita E. Source- and polymer-specific size distributions of fine microplastics in surface water in an urban river [J]. *Environmental Pollution*, 2021,284:117516.
- [19] Zhu T, Tao C, Cheng H, et al. Versatile in silico modelling of microplastics adsorption capacity in aqueous environment based on molecular descriptor and machine learning [J]. *Science of The Total Environment*, 2022,846:157455.
- [20] Astray G, Soria-Lopez A, Barreiro E, et al. Machine learning to predict the adsorption capacity of microplastics [J]. *Nanomaterials*, 2023,13(6):1061.
- [21] Zou H H, He P J, Peng W, et al. Rapid detection of colored and colorless macro- and micro-plastics in complex environment via near-infrared spectroscopy and machine learning [J]. *Journal of Environmental Sciences*, 2025,147:512-522.
- [22] Goyetche R, Kortazar L, Amigo J M. Issues with the detection and classification of microplastics in marine sediments with chemical imaging and machine learning [J]. *TrAC Trends in Analytical Chemistry*, 2023,166:117221.
- [23] Su J, Zhang F, Yu C, et al. Machine learning: next promising trend for microplastics study [J]. *Journal of Environmental Management*, 2023, 344:118756.
- [24] Lin J Y, Liu H T, Zhang J. Recent advances in the application of machine learning methods to improve identification of the microplastics in environment [J]. *Chemosphere*, 2022,307:136092.
- [25] Wang C, Zhao J, Xing B. Environmental source, fate, and toxicity of microplastics [J]. *Journal of Hazardous Materials*, 2021,407:124357.
- [26] Kedzierski M, Falcou-Préfol M, Kerros M E, et al. A machine learning algorithm for high throughput identification of ftir spectra: Application on microplastics collected in the Mediterranean Sea [J]. *Chemosphere*, 2019,234:242-251.
- [27] Politikos D V, Fakiris E, Davvetas A, et al. Automatic detection of seafloor marine litter using towed camera images and deep learning [J]. *Marine Pollution Bulletin*, 2021,164:111974.
- [28] Li J, Skinner K A, Eustice R M, et al. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images [J]. *IEEE Robotics and Automation Letters*, 2017:387-394.
- [29] Salgado-Hernanz P M, Bauzá J, Alomar C, et al. Assessment of marine litter through remote sensing: recent approaches and future goals [J]. *Marine Pollution Bulletin*, 2021,168:112347.
- [30] Sannigrahi S, Basu B, Basu A S, et al. Development of automated marine floating plastic detection system using sentinel-2imagery and machine learning models [J]. *Marine Pollution Bulletin*, 2022,178:113527.
- [31] Haider S N, Zhao Q, Meran B K. Automated data cleaning for data centers: a case study[C]//2020 39th Chinese Control Conference (CCC). Shenyang, China: IEEE, 2020:3227-3232.
- [32] Ruiz-Chavez Z, Salvador-Meneses J, Garcia-Rodriguez J. Machine learning methods based preprocessing to improve categorical data classification [M]//Yin H, Camacho D, Novais P, et al. *Intelligent Data Engineering and Automated Learning - IDEAL 2018:Vol.11314*. Cham: Springer International Publishing, 2018:297-304.
- [33] Li C. Preprocessing methods and pipelines of data mining: an overview [J]. *arXiv preprint*, 2019:1-7.
- [34] Tawakuli A, Kaiser D, Engel T. Synchronized preprocessing of sensor data [C]//2020IEEE International Conference on Big Data. Atlanta, GA, USA: IEEE, 2020:3522-3531.
- [35] Hosseinzadeh M, Azhir E, Ahmed O H, et al. Data cleansing mechanisms and approaches for big data analytics: a systematic study [J]. *Journal of Ambient Intelligence and Humanized Computing*, 2023,14(1):99-111.
- [36] Narita M, Igarashi T. Programming-by-example for data transformation to improve machine learning performance [C]//Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion. Marina del Ray California: ACM, 2019:113-114.
- [37] Huang J, Li Y F, Xie M. A Systematic analysis of data preprocessing for machine learning-based software cost estimation [J]. *Information and Software Technology*, 2015,67:108-127.
- [38] Jia W, Sun M, Lian J, et al. Feature dimensionality reduction: a review [J]. *Complex & Intelligent Systems*, 2022,8(3):2663-2693.
- [39] Cowger W, Gray A, Christiansen S H, et al. Critical review of processing and classification techniques for images and spectra in microplastic research [J]. *Applied Spectroscopy*, 2020,74(9):989-1010.
- [40] Gambin A F, Angelats E, Gonzalez J S, et al. Sustainable marine ecosystems: deep learning for water quality assessment and forecasting [J]. *IEEE Access*, 2021,9:121344-121365.
- [41] Lorenzo-Navarro J, Castrillon-Santana M, Santesarti E, et al. Smacc: a System for microplastics automatic counting and classification [J]. *IEEE Access*, 2020,8:25249-25261.
- [42] Höppener E M, Shahmohammadi M (Sadegh), Parker L A, et al. Classification of (micro)plastics using cathodoluminescence and machine learning [J]. *Talanta*, 2023,253:123985.
- [43] Renner G, Schmidt T C, Schram J. A new chemometric approach for automatic identification of microplastics from environmental compartments based on ft-ir spectroscopy [J]. *Analytical Chemistry*, 2017,89(22):12045-12053.
- [44] Michel A P M, Morrison A E, Preston V L, et al. Rapid identification of marine plastic debris via spectroscopic techniques and machine learning classifiers [J]. *Environmental Science & Technology*, 2020, 54(17):10630-10637.
- [45] Meyers N, Catarino A I, Declercq A M, et al. Microplastic detection

- and identification by Nile red staining: towards a semi-automated, cost- and time-effective technique [J]. *Science of The Total Environment*, 2022,823:153441.
- [46] Luo Y, Su W, Xu D, et al. Component identification for the IR spectra of microplastics mixture with convolutional neural network [J]. *Science of The Total Environment*, 2023,895:165138.
- [47] Xu L, Chen Y, Feng A, et al. Study on detection method of microplastics in farmland soil based on hyperspectral imaging technology [J]. *Environmental Research*, 2023,232:116389.
- [48] Ai W, Liu S, Liao H, et al. Application of hyperspectral imaging technology in the rapid identification of microplastics in farmland soil [J]. *Science of The Total Environment*, 2022,807:151030.
- [49] Shi B, Patel M, Yu D, et al. Automatic quantification and classification of microplastics in scanning electron micrographs via deep learning [J]. *Science of The Total Environment*, 2022,825:153903.
- [50] Da Silva V H, Murphy F, Amigo J M, et al. Classification and quantification of microplastics (<100 $\mu$ m) using a focal plane array-fourier transform infrared imaging system and machine learning [J]. *Analytical Chemistry*, 2020,92(20):13724-13733.
- [51] Hufnagl B, Steiner D, Renner E, et al. A methodology for the fast identification and monitoring of microplastics in environmental samples using random decision forest classifiers [J]. *Analytical Methods*, 2019,11(17):2277-2285.
- [52] Owen S, Cureton S, Szuhán M, et al. Microplastic adulteration in homogenized fish and seafood - a mid-infrared and machine learning proof of concept [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2021,260:119985.
- [53] Vidal C, Pasquini C. A comprehensive and fast microplastics identification based on near-infrared hyperspectral imaging (HSI-NIR) and chemometrics [J]. *Environmental Pollution*, 2021,285:117251.
- [54] Ivleva N P. Chemical Analysis of Microplastics and Nanoplastics: challenges, advanced methods, and perspectives [J]. *Chemical Reviews*, 2021,121(19):11886-11936.
- [55] Yang M, Zhang B, Chen X, et al. Transport of microplastic and dispersed oil co-contaminants in the marine environment [J]. *Environmental Science & Technology*, 2023,57(14):5633-5645.
- [56] Sørensen L, Groven A S, Hovsbakken I A, et al. UV degradation of natural and synthetic microfibers causes fragmentation and release of polymer degradation products and chemical additives [J]. *Science of The Total Environment*, 2021,755:143170.
- [57] Yan Y, Zhu F, Zhu C, et al. Dibutyl phthalate release from polyvinyl chloride microplastics: Influence of plastic properties and environmental factors [J]. *Water Research*, 2021,204:117597.
- [58] Xing F, Xie Y, Su H, et al. Deep learning in microscopy image analysis: a survey [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018,29(10):4550-4568.
- [59] Song Y K, Hong S H, Jang M, et al. A comparison of microscopic and spectroscopic identification methods for analysis of microplastics in environmental samples [J]. *Marine Pollution Bulletin*, 2015,93(1/2):202-209.
- [60] Shim W J, Hong S H, Eo S E. Identification methods in microplastic analysis: a review [J]. *Analytical Methods*, 2017,9(9):1384-1391.
- [61] Wang Z M, Wagner J, Ghosal S, et al. SEM/EDS and optical microscopy analyses of microplastics in ocean trawl and fish guts [J]. *Science of The Total Environment*, 2017,603-604:616-626.
- [62] Chen L, Strauch M, Merhof D. Instance segmentation of biomedical images with an object-aware embedding learned with local constraints [J]. *Medical Image Computing and Computer Assisted Intervention*, 2020:451-459.
- [63] Khadangi A, Boudier T, Rajagopal V. EM-net: deep learning for electron microscopy image segmentation[C]//2020 25th International Conference on Pattern Recognition (ICPR). 2021:31-38.
- [64] Haberl M G, Churas C, Tindall L, et al. CDeep3M—plug-and-play cloud-based deep learning for image segmentation [J]. *Nature Methods*, 2018,15(9):677-680.
- [65] Matuszewski D J, Sintorn I M. Reducing the U-net size for practical scenarios: virus recognition in electron microscopy images [J]. *Computer Methods and Programs in Biomedicine*, 2019,178:31-39.
- [66] Wegmayr V, Sahin A, Samundsson B, et al. Instance segmentation for the quantification of microplastic fiber images [C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass Village, CO, USA: IEEE, 2020:2199-2206.
- [67] Yan X, Cao Z, Murphy A, et al. An ensemble machine learning method for microplastics identification with FTIR spectrum [J]. *Journal of Environmental Chemical Engineering*, 2022,10(4):108130.
- [68] Siri Wong S, Tanthanuch W, Srisamut D, et al. Performance evaluation of focal plane array (FPA)-FTIR and synchrotron radiation (SR)-FTIR microspectroscopy to classify rice components [J]. *Microscopy and Microanalysis*, 2022,28(6):2059-2068.
- [69] Simon M, Van Alst N, Vollertsen J. Quantification of microplastic mass and removal rates at wastewater treatment plants applying Focal plane array (FPA)-based Fourier transform infrared (FTIR) imaging [J]. *Water Research*, 2018,142:1-9.
- [70] Sun Y, Yuan J, Zhou T, et al. Laboratory simulation of microplastics weathering and its adsorption behaviors in an aqueous environment: a systematic review [J]. *Environmental Pollution*, 2020,265:114864.
- [71] Zhen Y, Wang L, Sun H, et al. Prediction of microplastic abundance in surface water of the ocean and influencing factors based on ensemble learning [J]. *Environmental Pollution*, 2023,331:121834.
- [72] Xu R Z, Cao J S, Ye T, et al. Automated machine learning-based prediction of microplastics induced impacts on methane production in anaerobic digestion [J]. *Water Research*, 2022,223:118975.
- [73] Cowger W, Gray A, Christiansen S H, et al. Critical review of processing and classification techniques for images and spectra in microplastic research [J]. *Applied Spectroscopy*, 2020,74(9):989-1010.
- [74] Qi Y, Hu D, Jiang Y, et al. Recent progresses in machine learning assisted Raman spectroscopy [J]. *Advanced Optical Materials*, 2023,11(14):2203104.
- [75] Weisser J, Pohl T, Heinzinger M, et al. The identification of microplastics based on vibrational spectroscopy data - a critical review of data analysis routines [J]. *TrAC Trends in Analytical Chemistry*, 2022,148:116535.
- [76] Lei L, Pang R, Han Z, et al. Current applications and future impact of machine learning in emerging contaminants: a review [J]. *Critical Reviews in Environmental Science and Technology*, 2023,53(20):1817-1835.

**作者简介:** 王星程(2000-),男,新疆伊犁人,中国计量大学硕士研究生,研究方向为基于机器学习的污染物环境行为研究.1360698353@qq.com.