黄瑞潇,卢永琦,郑志敏,等.典型固体废物的 XRF-XRD 耦合光谱指纹特征识别 [J]. 中国环境科学, 2025,45(10):5585-5595.
Huang R X, Lu Y Q, Zheng Z M, et al. Identification of XRF-XRD coupled spectral fingerprint characteristics of typical solid waste [J]. China Environmental Science, 2025,45(10):5585-5595.

典型固体废物的 XRF-XRD 耦合光谱指纹特征识别

黄瑞潇,卢永琦,郑志敏,杨玉飞,杨金忠*,黄启飞 (中国环境科学研究院,环境基准与风险管控全国重点实验室,固体废物污染控制技术研究所,北京 100012)

摘要：针对常见的 10 类固体废物开展 X 射线荧光(XRF)与 X 射线衍射(XRD)实验,通过非参数检验提取固体废物的 XRF 与 XRD 指纹特征,并采用归一化的方式结合两种指纹特征形成 XRF-XRD 耦合光谱指纹特征,以三种指纹特征为数据集训练机器学习分类模型,发现随机森林模型分类识别效果最好,验证集分类精确度、召回率、准确率分别达 98.0%、97.5%、97.6%,证实耦合光谱指纹特征分类效果优于单一光谱指纹特征.形成了固体废物 XRF-XRD 多维指纹特征提取方法,为固体废物指纹特征识别奠定方法基础.

关键词：指纹特征；固体废物；XRF；XRD
中图分类号：X705 **文献标识码**：A **文章编号**：1000-6923(2025)10-5585-11

Identification of XRF-XRD coupled spectral fingerprint characteristics of typical solid waste. HUANG Rui-xiao, LU Yong-qi, ZHENG Zhi-min, YANG Yu-fei, YANG Jin-zhong*, HUANG Qi-fei (Research Institute of Solid Waste Management, State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environment Sciences, Beijing 100012, China). *China Environmental Science*, 2025,45(10)：5585~5595

Abstract：This study conducted X-ray fluorescence (XRF) and X-ray diffraction (XRD) experiments on 10 common types of solid waste. By employing non-parametric tests, the XRF and XRD fingerprint characteristics of solid waste were extracted. A normalized approach was adopted to combine the two fingerprint characteristics to form XRF-XRD coupled spectral fingerprint characteristics. A machine learning classification model was trained using the three fingerprint characteristics as a dataset. It was found that the random forest model exhibited the best classification recognition performance, with classification accuracy, recall rate, and precision rate in the validation set reaching 98.0%, 97.5%, and 97.6%, respectively. This confirmed that the classification performance of coupled spectral fingerprint characteristics was superior to that of single spectral fingerprint characteristics. A method for extracting multi-dimensional fingerprint characteristics of solid waste using XRF-XRD was developed, laying a methodological foundation for the identification of fingerprint characteristics of solid waste.

Key words：fingerprint features；solid waste；XRF；XRD

随着我国工业能力发展,固体废物产生量逐年增加,部分企业为降低处理成本,会选择在野外违规倾倒固体废物,据统计,目前中国固体废物累计堆存量约 600 亿 t,年新增堆存量近 30 亿 t,尤其是钢渣等固体废物回收利用率较低,占用大量土地资源,易造成环境污染[1].野外堆存固体废物无法通过肉眼快速识别分类,难以选择最佳的回收或处理方式,造成资源浪费并提高因处理不当造成的环境风险.

固体废物的指纹特征是固体废物的指标及其描述的组合,通过构建不同固体废物的指纹特征可以实现固体废物快速分类.X 射线荧光(XRF)与 X 射线衍射(XRD)常用于检测固体废物的主要元素含量及物相[2-7],两种检测方式都有成熟的便携式快速检测装置[8-9],具有实现指纹特征提取与快速分类检测的潜力.在固体废物 XRF 指纹特征提取领域,研究团队[10]以固体废物的 XRF 数据构建了黄河流域七类一般工业固体废物的指纹特征,以该指纹特征构建的分类识别模型可达 83.3%的分类准确率;在固体废物 XRD 指纹特征提取领域,相关研究[11]以有色金属冶炼行业为例,分析了不同单元产生固体废物的物相差异,初步确定了有色金属冶炼行业的固体废物 XRD 指纹特征,但未根据指纹特征训练分类识别模型,除固体废物外,通过建立 XRD 对照图谱和确认共有峰等步骤可以建立不同产地的中药 XRD 特征

收稿日期：2025-02-26
基金项目：国家重点研发计划项目(2024YFC3906401);黄河流域生态保护和高质量发展联合研究项目(2022-YRUC-01-0303)
* 责任作者, 副研究员, yang.jinzhong@craes.org.cn

图谱,为中药类别鉴定提供有效参照方法[12-13].

识别未知固体废物种类对于固体废物的回收、利用和风险防控至关重要.然而,现有的方法存在一定的适用性不足的问题,特别是复杂的野外堆存场景中难以高效识别固体废物种类.为进一步提高固体废物指纹特征可用性以及提升分类识别模型分类准确率,使其能够应用于野外堆存固体废物的快速分类识别,本文在固体废物 XRF 指纹特征研究的基础上,建立固体废物 XRD 指纹特征提取方法,形成固体废物的 XRF-XRD 耦合光谱指纹特征,并结合机器学习算法筛选固体废物分类识别模型,验证指纹特征可用性,形成更加准确的多维指纹提取方法及固体废物指纹特征,支撑固体废物精细化风险管控.

1 材料与方法

1.1 供试材料与仪器

1.1.1 供试材料 针对野外堆存未知固体废物快速分类的实际需要,本文通过资料查找选定铝灰、锂渣、炉渣、电镀污泥、氰渣、钢渣、钴渣、飞灰、磷石膏与岩屑[1,14-20]10 类常被非法倾倒的典型固体废物,自相关工厂进行样品收集,最终获取 103 个固体废物样品.

1.1.2 仪器 X 射线粉末衍射仪(Rigaku Ultima Ⅳ,日本岛津公司),X 射线荧光光谱仪(Panalytical Axios,荷兰帕纳科公司),电热恒温干燥箱(101-2A,北京科委永兴仪器有限公司),手动等静压压片机(NL-12J,天津诺雷信达科技有限公司).

1.2 实验方法

1.2.1 样品制备 将样品置于电热恒温干燥箱中干燥约 12h,采用四分法对干燥后的样品取样约 50g 左右研磨过 200 目筛得到粉末状样品[21].过筛所得粉末状样品分为两份,其中一份使用 X 射线粉末衍射仪直接测量,另一份使用压片机在 25MPa 压力下持续压制 0.5min,制成厚度为 10mm 的薄片后用 X 射线荧光光谱仪测量.

1.2.2 实验条件 X 射线荧光光谱仪测量条件:工作电压 60kV;工作电流 50mA;X 射线最大输出功率≥4kW;使用靶材 Rh;测量时间 30s;分析谱线 Kα;测试范围 11Na-92U.

X 射线粉末衍射仪测量条件:工作电压 40kV;工作电流 40mA;X 射线最大输出功率≥3kW;光管功率≥2kW;发射狭缝 DS=1°;接收狭缝 RS=0.15mm;使用靶材 Cu;扫描方式连续扫描;扫描速度 2°/min;扫描范围 5°~90°(2θ);步长 0.02°.

1.3 特征提取方法

1.3.1 指纹指标的初选 基于《HJ 1211-2021 固体废物 无机元素的测定 波长色散 X 射线荧光光谱法》中的固体废物 XRF 数据处理步骤,使用 SuperQ4.0D 软件并选择 Kα 谱线处理样品 XRF 数据,将样品的特征 X 射线波长和能量换算为氧化物含量数据[22],剔除 10 类固体废物中含量为零或较少的氧化物含量数据,这类数据统计学意义较小,无法作为指纹特征提取的基本因子;使用 MDI Jade 6.0 软件处理样品 XRD 数据,将衍射数据与国际晶体衍射数据库(Inorganic Crystal Structure Database,ICSD)中的标准 PDF 卡片对比进行物相检索标识出主要特征峰及代表物相[23-24],将同类固体废物衍射数据叠加平均初步获取固体废物的平均 XRD 图谱,在图谱中根据强度选取两种主要物相分别代表的特征峰,多个特征峰代表同一物相时选取强度最高的特征峰,剔除 10 类固体废物选出的重复特征峰后剩余特征峰及强度作为用于提取指纹特征的基本因子.

1.3.2 指纹特征提取 非参数检验法是进行特征提取常用方法[25-26],该方法可以提取组间差异显著的因子作为指纹因子构建指纹特征,用于多个独立样本的非参数检验方法主要有 Kruskal-Wallis H(K-W H)检验、Mood 中位数检验与 Jonckheere-Terpstra(JT)检验 3 种.

K-W H 检验是一种秩检验方法,在数据不满足正态分布时要优于方差分析[27],在固体废物指纹因子提取过程中该检验的零假设为不同种固体废物的因子数据相同,将计算所得的 H 值与临界卡方值对比,如果 H 值较大则认为该指纹因子在各类固体废物之间存在显著差异,可以作为区分不同固体废物的指纹因子,临界卡方值可通过自由度与显著性水平有关,通过查询卡方分布表得知[28],本研究检验显著性水平均取 0.01.H 值计算公式如下[29]:

$$H=\frac{12}{N(N+1)}\sum_{i=1}^{m}\frac{R_i^2}{n_i}-3(N+1) \quad (1)$$

式中:H 表示检验统计量;R_i 表示第 i 种固体废物类别的秩和;m 表示固体废物类别数量;n_i 表示第 i 种固体

废物类别样本个数;N 表示固体废物样本总数.

Mood 中位数检验是一种特殊的卡方检验, 将单个因子的所有样本数据混合求取平均值 M,统计各类固体废物样本数据大于 M 和小于 M 的个数汇总为列联表,采用卡方检验计算上述列联表的实际观测值与理论推断值之间的偏离程度并与临界卡方值对比,判断该因子在组间是否存在显著差异.卡方统计量计算公式如下[30-31]:

$$X^2=\sum_{i=1}^{k}\frac{(f_i-np_i)^2}{np_i} \tag{2}$$

式中:X^2 表示卡方统计量;f_i 表示真实频数;np_i 表示预测频数;k 表示区间个数;i 表示第 i 个区间.

JT 检验的基本思想与两独立样本的 Mann−Whitney U 检验类似,在分组变量有序时效能显著优于 K−W H 检验[32],该检验零假设为同一因子在不同固体废物类别的分布无显著差异,在单个因子的 k 个独立样本中取出两个样本:X_{is} 和 X_{jt}(其中 i=1,2,…,k;s=1,2,…,n_j;j=1,2,…,k;t=1,2,…,n_i)进行对比,计算第 i 类固体废物样本观测值小于第 j 类固体废物样本观测值的对数 U_{ij},并对所有 U_{ij} 求和获取 JT 检验统计量,与 JT 检验临界值 $J\alpha$ 对比判断是否拒绝零假设.JT 检验临界值需对比 2019 年出版的《非参数统计(第 5 版)》中的 Jonckheere−Terpstra 检验临界值 $J\alpha$ 表获取[33],对数 U_{ij} 与 JT 检验统计量公式如下所示[34]:

$$U_{ij}=\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}D_{st} \tag{3}$$

式中:U_{ij} 表示观测值对数;n_i 与 n_j 分别表示第 i 类固体废物和第 j 类固体废物的样本容量;D_{st} 表示调整系数,在 $X_{is}=X_{jt}$ 时,D_{st}=0.5;$X_{is}>X_{jt}$ 时,D_{st}=0;$X_{is}<X_{jt}$ 时,D_{st}=1.

$$J=\sum_{i<j}U_{ij} \tag{4}$$

式中:J 表示 JT 检验统计量.

将固体废物 XRF 与 XRD 基本因子数据通过上述三种非参数检验法检验后,提取所有检验方法均判定为差异显著(P<0.01)的因子作为指纹因子构建各类固体废物的 XRF 与 XRD 光谱指纹特征.

1.4　基于指纹特征的分类模型训练

机器学习算法在分类领域显著优于传统算法,目前常用于分类识别的机器学习算法有 Fisher 判别分析[35]、随机森林(RF)[36]与 k−最近邻(KNN)[37]等.Fisher 判别分析是基于方差分析思想的线性分类判别方法,逐步判别分析通过 Wilks 统计量最小原则逐步引入新指纹因子并去除无用指纹因子[38],筛选出最佳指纹因子组并根据该因子组的协方差矩阵构建线性判别函数;随机森林是基于决策树的集成算法,通过在原始数据集中随机抽样训练多个决策树分类模型,统计每颗决策树的分类结果投票选择出最终分类结果[39];k−最近邻是一种较为简单成熟的机器学习分类算法,该算法通过在一定的特征空间内统计未知样本周围 k 个已知样本的类别归类未知样本[37].为避免出现过拟合现象,在分类模型构建时需提前划分出不参与模型训练的数据构成验证集,以模型对验证集的分类准确率判断模型优劣.

本研究通过上述三种机器学习算法分别基于固体废物的 XRF 指纹特征、XRD 指纹特征与 XRF−XRD 耦合光谱指纹特征构建快速分类识别模型,通过验证集分类识别精确度(Precision)、召回率(Recall)和准确率(Accuracy)综合评价分类模型的性能,验证耦合光谱指纹特征是否优于单指纹特征.

2　结果与分析

2.1　指纹指标的初选

由样品 XRF 数据处理所得的氧化物数据共有二氧化硅(SiO_2)、三氧化二铝(Al_2O_3)、三氧化二铁(Fe_2O_3)、氧化钾(K_2O)、氧化钠(Na_2O)、氧化钙(CaO)、氧化镁(MgO)、三氧化二砷(As_2O_3)、氧化钡(BaO)、五氧化二钒(V_2O_5)、氧化锌(ZnO)、三氧化铬(CrO_3)、氧化铜(CuO)、氧化锰(MnO)、氧化镍(NiO)、五氧化二磷(P_2O_5)、氧化铅(PbO)、氧化钴(CoO)、二氧化钛(TiO_2)、三氧化硫(SO_3)、氧化锶(SrO)、二氧化锆(ZrO_2) 22 项氧化物含量数据, CuO、Cd_2O_3 等 10 项基本因子在多数固体废物样本中未达到检出限,因此这些基本因子在较多种类的固体废物中会表现出相同的特征,不能形成有效区分不同类别固体废物的指纹特征因子,不具有统计学意义,因此这 10 项基本因子被去除,保留 Al_2O_3、MgO 等 12 项基本因子作为指纹提取待选因子.

由样品 XRD 数据处理得 103 个固体废物 XRD 图谱,将同类固体废物 XRD 图谱叠加平均后标识出主要特征峰及其代表物相.以氰渣为例,将 9 个氰渣样品的 XRD 图谱(图 1(a))在数值上叠加平均获取氰

渣平均图谱(图 1(b)),在 MDI Jade 6.0 软件中分析图谱可知氰渣主要物相为二氧化硅(SiO_2)与碳酸钙镁($CaMg(CO_3)_2$)并记录特征峰位作为基本因子.

提取 10 种固体废物的主要特征峰并去除重复特征峰,获取可作为指纹提取待选因子的 15 项基本因子 如表 1 所示,在特征峰选取结果中有两个峰位都代表物相 SiO_2,为方便区分将峰位 20.84°与 26.64°的物相标为 SiO_2(Ⅰ)与 SiO_2(Ⅱ).

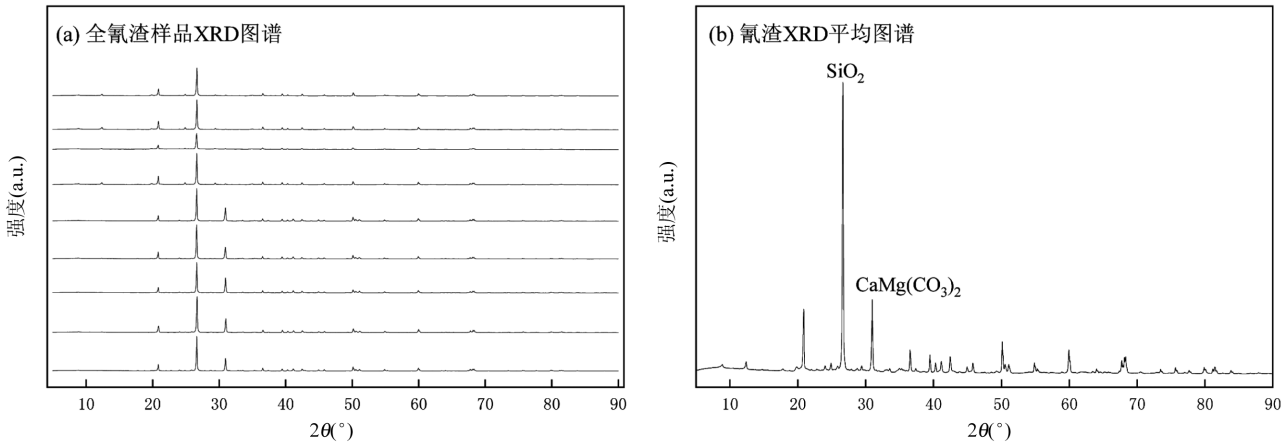


图 1 氰渣 XRD 特征峰标识过程

Fig.1 Identification process of XRD characteristic peaks of cyanide residue

表 1 XRD 指纹提取待选因子

Table 1 XRD fingerprint extraction of candidate factors

来源固体废物	代表物相	名称	2θ(°)
电镀污泥	$CaSO_4·2H_2O$	硫酸钙	11.62
炉渣	SiO_2(Ⅰ)	二氧化硅	20.84
锂渣	$Na_6Ca_2Al_6Si_6O_{24}(SO_4)_2$	钙霞石	23.88
钴渣	$CaSO_4$	硫酸钙	25.48
炉渣	$Al_6Si_2O_{13}$	莫来石	26.24
氰渣	SiO_2(Ⅱ)	二氧化硅	26.64
锂渣	$CaAl_2Si_2O_8$	钙长石	27.88
飞灰	KCl	氯化钾	28.36
磷石膏	$CaPO_3(OH)·2H_2O$	二水磷酸氢钙	29.12
岩屑	$CaCO_3$	碳酸钙	29.34
钢渣	Ca_2SiO_4	硅酸二钙	29.70
氰渣	$CaMg(CO_3)_2$	碳酸钙镁	30.96
飞灰	NaCl	氯化钠	31.64
铝灰	Al	铝	38.46
钢渣	CaF_2	氟化钙	47.06

2.2 指纹特征提取

通过软件 IBM SPSS 26.0 对 10 种固体废物的 XRF 基本因子与 XRD 基本因子进行非参数检验,各因子判别能力检验结果如表 2、3 所示.在三种非参数检验方法中 JT 检验指纹因子提取效果要显著优于 K-W H 检验与中位数检验,所有 XRF 与 XRD 基本因子在 K-W H 检验与中位数检验下均表现出显著的组间差异,无法有效筛选出可以有效区分不同类别固体废物的指纹因子,因此各项基本因子仅通过 JT 检验即可判定差异是否显著.在 12 项 XRF 基本因子的 JT 检验中,Na_2O、SO_3 与 $TiO_2$3 项基本因子显著性大于 0.01,组间差异不显著,其余 Al_2O_3、MgO 等 9 项基本因子显著性小于 0.01,组间差异显著;在 15 项 XRD 基本因子的 JT 检验中, $CaSO_4·2H_2O$、$Na_6Ca_2Al_6Si_6O_{24}(SO_4)_2$ 等 7 项基本因子显著性大于 0.01,组间差异不显著,其余 SiO_2(Ⅰ)、$Al_6Si_2O_{13}$ 等 8 项基本因子显著性小于 0.01,组间差异显著.最终选取 9 项 XRF 指纹因子与 8 项 XRD 指纹因子分别构建 XRF 指纹特征与 XRD 指纹特征.

使用 9 项 XRF 指纹因子构建的 10 类典型固体废物指纹特征模型如图 2 所示.岩屑[40]与氰渣[41]的特征相似,SiO_2 含量较高;飞灰[42]与磷石膏[43]的特征相似,CaO 含量较高;电镀污泥[44]与钴渣[45]特征相似,CaO、Fe_2O_3 与 SiO_2 三种氧化物含量较高;铝灰[46]的特征较为明显,其 Al_2O_3 含量显著高于其他固体废物;锂渣[47]、炉渣[48]、钢渣[49]三种固体废物的特征为都含有 CaO、Al_2O_3、SiO_2 三种氧化物,但炉渣 CaO 与钢渣 Al_2O_3 含量较低.因此,对于固体废物的 XRF 指纹特征来说,铝灰、锂渣、炉渣、钢渣四种固体废

物较易区分,其余六种固体废物出现分类错误的可　　能性较大.

表 2　XRF 基本因子非参数检验结果

Table 2　Non parametric test results of XRF basic factors

指纹因子	名称	K−W H 检验		中位数检验		JT 检验	
		检验统计量	显著性	检验统计量	显著性	检验统计量	显著性
Na_2O	氧化钠	71.239	0.000*	58.101	0.000*	−1.432	0.152
MgO	氧化镁	84.711	0.000*	87.710	0.000*	−2.679	0.007*
Al_2O_3	氧化铝	93.740	0.000*	90.510	0.000*	−6.009	0.000*
SiO_2	二氧化硅	83.622	0.000*	75.383	0.000*	3.270	0.001*
P_2O_5	五氧化二磷	87.025	0.000*	71.357	0.000*	4.036	0.000*
SO_3	三氧化硫	88.310	0.000*	92.732	0.000*	1.775	0.076
K_2O	氧化钾	89.515	0.000*	87.164	0.000*	−3.615	0.000*
CaO	氧化钙	90.872	0.000*	82.141	0.000*	4.110	0.000*
TiO_2	二氧化钛	84.136	0.000*	67.584	0.000*	−1.931	0.053
MnO	氧化锰	72.896	0.000*	62.666	0.000*	−2.722	0.006*
Fe_2O_3	氧化铁	86.484	0.000*	82.141	0.000*	3.301	0.001*
ZnO	氧化锌	90.735	0.000*	89.783	0.000*	−4.498	0.000*

注:"*"表示在0.01水平上显著.

表 3　XRD 基本因子非参数检验结果

Table 3　Non parametric test results of XRD basic factors

指纹因子	名称	K−W H 检验		中位数检验		JT 检验	
		检验统计量	显著性	检验统计量	显著性	检验统计量	显著性
$CaSO_4 \cdot 2H_2O$	生石膏	66.849	0.000*	60.093	0.000*	2.363	0.018
SiO_2(Ⅰ)	石英	94.478	0.000*	90.599	0.000*	3.267	0.001*
$Na_6Ca_2Al_6Si_6O_{24}(SO_4)_2$	钙霞石	83.880	0.000*	74.869	0.000*	1.001	0.317
$CaSO_4$	硫酸钙	59.115	0.000*	56.226	0.000*	1.233	0.217
$Al_6Si_2O_{13}$	莫来石	81.403	0.000*	74.569	0.000*	2.834	0.005*
SiO_2(Ⅱ)	石英	89.553	0.000*	74.007	0.000*	3.103	0.002*
$CaAl_2Si_2O_8$	钙长石	86.411	0.000*	65.841	0.000*	0.608	0.543
KCl	氯化钾	69.733	0.000*	48.431	0.000*	−2.498	0.012
$CaPO_3(OH)\cdot 2H_2O$	二水磷酸氢钙	79.661	0.000*	74.672	0.000*	3.881	0.000*
$CaCO_3$	方解石	46.421	0.000*	28.891	0.001*	4.449	0.000*
Ca_2SiO_4	钙橄榄石	61.837	0.000*	63.555	0.000*	2.948	0.003*
$CaMg(CO_3)_2$	菱镁矿	69.189	0.000*	55.472	0.000*	1.540	0.124
NaCl	氯化钠	72.708	0.000*	63.096	0.000*	−3.301	0.001*
Al	铝	75.668	0.000*	67.105	0.000*	−2.951	0.003*
CaF_2	萤石	75.220	0.000*	55.840	0.000*	−1.526	0.127

注:"*"表示在0.01水平上显著.

使用 8 项 XRD 指纹因子构建的 10 类典型固体废物指纹特征模型如图 3 所示.锂渣[47]与钢渣[49]的特征相似,Ca_2SiO_4特征峰信号强度较高;岩屑[40]与氰渣[41]的特征具有相似性,两种固体废物的 SiO_2(Ⅰ)与 SiO_2(Ⅱ)特征峰信号强度较高,但岩屑的 $CaCO_3$特征峰信号强度要高于氰渣;炉渣、电镀污泥、钴渣 3 种固体废物特征不明显,无明显主要特征峰[44-45,48];铝灰、飞灰、磷石膏 3 种固体废物均存在明显的独有特征峰,铝灰含有 Al 特征峰[46],飞灰含有 NaCl 特征峰[42],磷石膏含有 $CaPO_3(OH)\cdot 2H_2O$ 特征峰[43].因此,对于固体废物的 XRD 特征来说,岩屑、氰渣、铝灰、飞灰、磷石膏 5 种固体废物较易区分,其余 5 种固体废物出现分类错误的可能性较大.

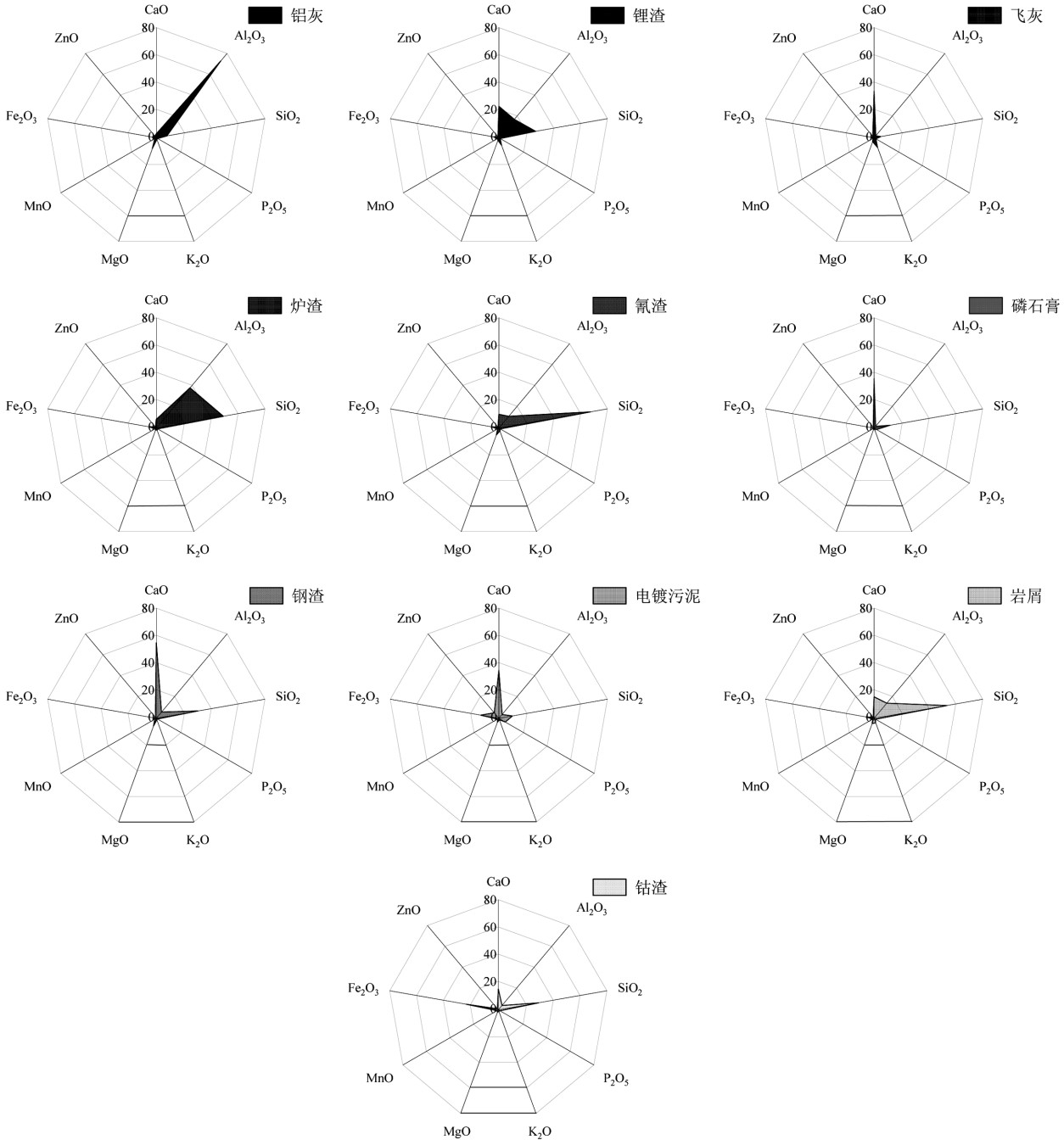
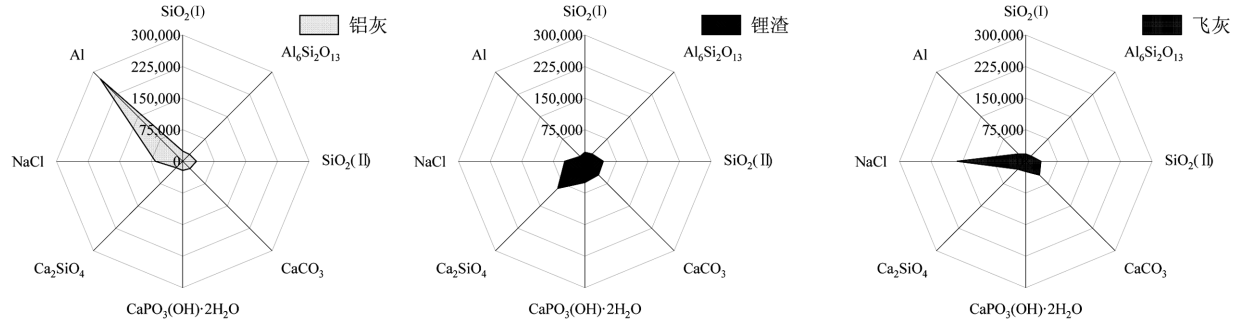


图 2 典型固体废物 XRF 指纹特征

Fig.2 Typical XRF fingerprint characteristics of solid waste



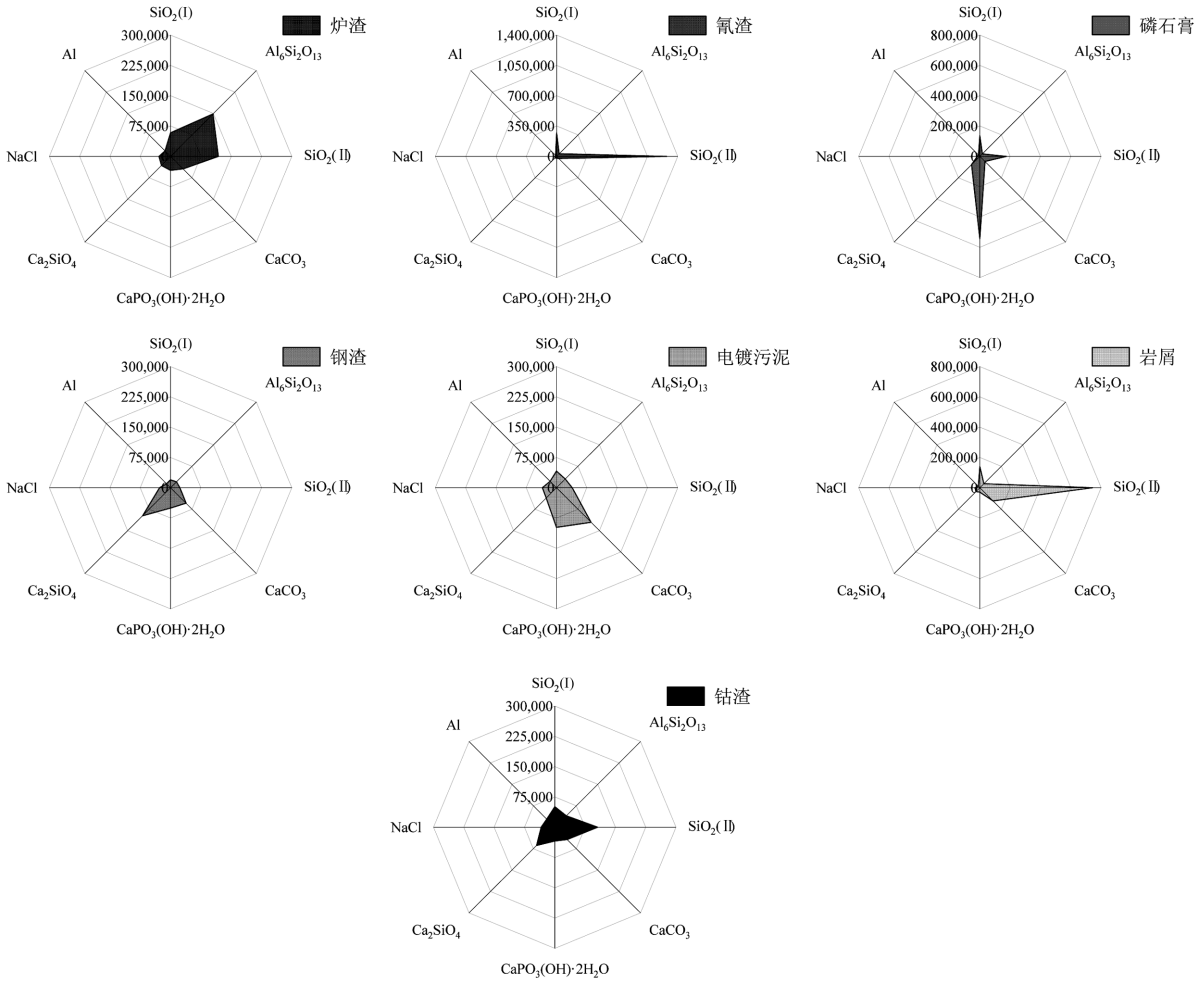


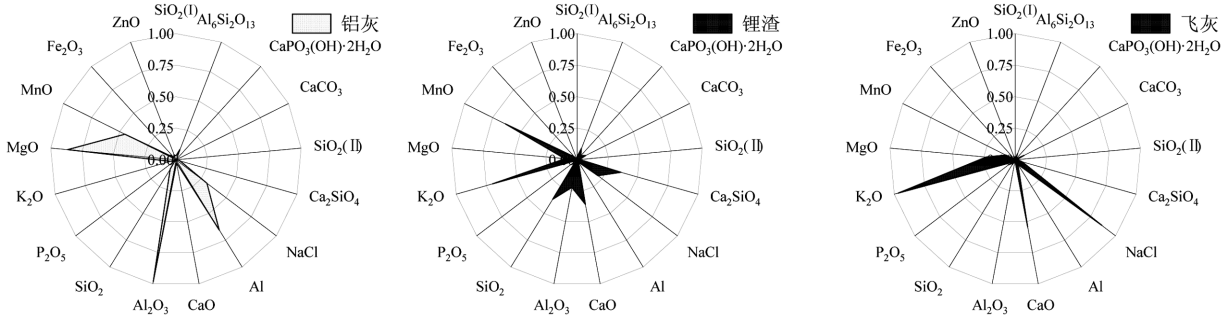
图 3　典型固体废物 XRD 指纹特征

Fig.3　Typical XRD fingerprint characteristics of solid waste

构建 XRF−XRD 耦合光谱指纹特征需通过归一化统一 XRF 指纹特征与 XRD 指纹特征的数量级,归一化公式如下所示:

$$y=\frac{x-\min x}{\max x-\min x} \tag{5}$$

式中:y 为归一化特征值;x 为某固体废物样本的待归一化特征因子;min x 为该特征因子所有样本的最小值;max x 为该特征因子所有样本的最大值.构建的 XRF−XRD 耦合光谱指纹特征如图 4 所示,对于特定类别的固体废物仅依靠单一指纹特征进行分类识别出现错误的可能性较高,采用耦合指纹特征可以通过结合 XRF 指纹特征与 XRD 指纹特征的优势避免该现象.例如,飞灰与磷石膏的 XRF 特征相似,区分难度较高,但两种固体废物的 XRD 特征存在较大的差别;锂渣与钢渣的 XRD 特征相似,但根据两种固体废物的 XRF 特征可以有效区分.在分类模型训练时耦合光谱指纹特征可以提供更多样的特征因子组合,提高模型分类识别准确率.



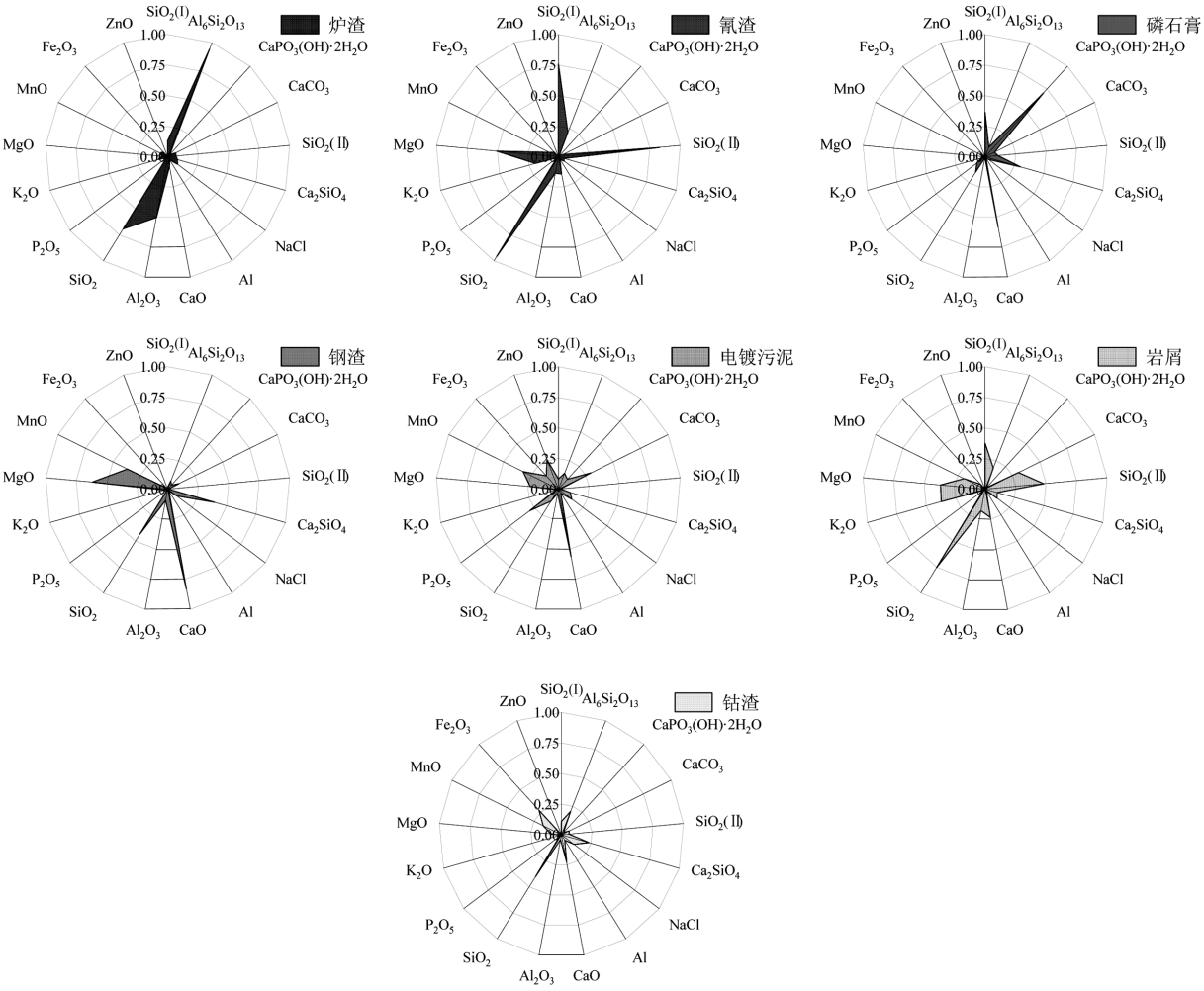


图 4 典型固体废物 XRF−XRD 耦合光谱指纹特征

Fig.4 XRF−XRD Coupled Spectral Fingerprint Characteristics

2.3 基于指纹特征的分类模型训练

本研究开发环境基于 Python 3.10,采用软件 Pycharm 2022 作为集成开发环境构建 Fisher 判别分析、随机森林、K−最近邻三种机器学习分类算法,分别以上文获取的 XRF 指纹特征、XRD 指纹特征以及耦合光谱指纹特征作为原始数据集训练分类识别模型.进行模型训练时随机按照 0.6:0.4 的比例将原始数据集划分为训练集与验证集,将所有机器学习算法的随机数种子设置为 42 以确保多次程序运行过程中出现的随机数序列相同, 以验证集分类准确率来判断分类识别效果.

各机器学习分类模型分类识别结果如表 4 所示,从三个评价指标来看随机森林模型均显著优于其余两种机器学习分类模型,最高训练集分类精确度、召回率、准确率分别达 98.0%、97.5%、97.6%,Fisher 判别分析属于线性函数分析,通过判别函数计算未知样本与标准样本的距离判断未知样本类别,在处理较多类别的固体废物数据时会出现维数灾难,降低分类效果[50];K−最近邻同样需要计算未知样本与已知标准样本的距离作为分类标准,在处理高维数据时也会出现维数灾难问题,但其参考的已知样本数量要多于判别分析,最终分类效果优于判别分析[51];随机森林在构建每颗决策树时仅选择部分特征,通过多颗决策树组合分析能够有效处理非线性数据,随机森林分类模型适用于处理高维数据,避免维数灾难[52].

从依靠不同指纹特征作为基础数据集训练模型的分类识别结果来看,随机森林与 K−最近邻两种分类模型采用耦合光谱指纹特征的效果要优于单一指纹特征,判别分析模型采用耦合光谱指纹特征训练效能反而会降低,判别分析的分类能力主要依赖从待选特征中选出特征组合构建线性判别函数,特征数量较多时无法有效选择出最佳特征组合,影

响判别函数的判别能力,判别分析在固体废物分类识别领域有较大的局限性[53].

表 4　机器学习分类识别模型训练集识别结果

Table 4　Machine learning classification recognition model training set recognition results

特征	Fisher 判别分析(%)			随机森林(%)			K−最近邻(%)		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy
XRF 指纹特征	71.7	80.0	81.0	95.7	92.5	92.9	80.0	82.5	83.3
XRD 指纹特征	75.1	75.0	73.8	94.9	90.8	90.5	84.6	86.1	85.7
耦合光谱指纹特征	80.0	70.8	73.8	98.0	97.5	97.6	81.3	90.0	90.5

以训练集分类准确率最高的随机森林分类识别模型为例,采用不同特征训练模型的验证集错误分类个案结果如表 5 所示,与特征分析猜测相同,在 XRF 指纹特征下钴渣与电镀污泥两类固体废物之间较易出现分类错误,在 XRD 指纹特征下钴渣与电镀污泥可以被明显区分,避免两类固体废物的分类错误;在 XRD 指纹特征下锂渣与钢渣两类固体废物之间较易出现分类错误,电镀污泥、钴渣无明显特征,较易识别为其他类别固体废物,在 XRF 指纹特征下钴渣与铝灰、电镀污泥与钢渣、锂渣与钢渣可以被明显区分,避免上述类别固体废物分类的错误.因此,采用耦合光谱指纹特征可以有效避免单一指纹特征的缺点,显著提升分类准确率.

表 5　随机森林分类识别模型训练集错误分类个案类别及数量

Table 5　Random forest classification and recognition model training set misclassified case categories and quantities

错误分类个案类别	XRF 指纹特征	XRD 指纹特征	耦合光谱指纹特征
钴渣→电镀污泥	3	0	1
锂渣→钢渣	0	1	0
电镀污泥→钢渣	0	1	0
钴渣→铝灰	0	2	0

3　结论

3.1　基于 10 类典型固体废物的 XRF 与 XRD 数据,采用不同的非参数检验方法筛选固体废物指纹特征因子,发现 K−W H 检验与中位数检验不能有效识别固体废物的指纹特征,依靠 JT 检验成功提取了不同类别固体废物的 XRF 指纹特征与 XRD 指纹特征并通过归一化结合两种指纹特征形成 XRF−XRD 耦合光谱指纹特征.

3.2　分别将 10 类固体废物的 XRF 指纹特征、XRD 指纹特征与耦合光谱指纹特征作为基础数据集,采用 Fisher 判别分析、随机森林、K−最近邻三种机器学习算法训练分类识别模型,结果发现 Fisher 判别分析模型在固体废物分类识别领域有较大的局限性,随机森林算法在识别未知类别固体废物时表现出色,分类精确度、召回率、准确率分别达 98.0%、97.5%、97.6%,并证实采用耦合光谱指纹特征进行固体废物分类识别要优于单一光谱指纹特征.

4　展望

本研究提出了一种提取固体废物指纹特征并基于指纹特征训练分类识别模型的方法,但仍然存在较大的改进空间.一是同类固体废物样本数量较少,所提取出的指纹特征可能无法有效表征所有该类别的固体废物.二是所涉及的固体废物类别相对较少,分类识别模型在泛用性上存在较大的提升空间.在后续研究中将会增加用于提取指纹特征的同类固体废物样本数量与增加更多的固体废物种类,提高指纹特征的代表性和分类模型的泛用性.

参考文献:

[1] 国家发展改革委.关于“十四五”大宗固体废弃物综合利用的指导意见 [EB/OL]. https://www.gov.cn/zhengce/zhengceku/2021−03/25/content_5595566.htm/2021−03−18.
National Development and Reform Commission. Guiding opinions on the comprehensive utilization of bulk solid waste during the 14th Five Year Plan Period [EB/OL]. https://www.gov.cn/zhengce/zhengceku/2021−03/25/content_5595566.htm/2021−03−18.

[2] De Matteis C, Pollastri S, Mantovani L, et al. Potentially toxic elements speciation in bottom ashes from a municipal solid waste incinerator: A combined SEM−EDS, μ−XRF and μ−XANES study [J]. Environmental Advances, 2024,15:100453.

[3] Pfandl K, Küppers B, Scheiber S, et al. X−ray fluorescence sorting of non−ferrous metal fractions from municipal solid waste incineration bottom ash processing depending on particle surface properties [J]. Waste management & research, 2020,38(2):111−21.

[4] Bedoya-Henao C A, Carmona-Ramírez J D, Betancur-Granados N, et al. Characterization and potential uses of municipal solid waste incineration remnants in San Andrés Islands – Colombia [J]. Case Studies in Chemical and Environmental Engineering, 2025,11:101047.

[5] Wang J, Xiang L, Ren C, et al. Quantitative determination of quaternary solid waste-based binders and its hydrates by XRD [J]. Construction and Building Materials, 2024,425:135888.

[6] Tiberg C, Sjöstedt C, Karlfeldt Fedje K. Speciation of Cu and Zn in bottom ash from solid waste incineration studied by XAS, XRD, and geochemical modelling [J]. Waste Management, 2021,119:389-98.

[7] Ohbuchi A, Koike Y, Nakamura T. Quantitative phase analysis of fly ash of municipal solid waste by X-ray powder diffractometry/Rietveld refinement [J]. Journal of Material Cycles and Waste Management, 2019,21(4):829-37.

[8] Li Y, Hu J, Li C, et al. Magnetic covalent organic framework for efficient solid-phase extraction of uranium for on-site determination by portable X-ray fluorescence spectrometry [J]. Analytical Chemistry, 2024,96(15):5757-62.

[9] Possenti E, Catrambone M, Colombo C, et al. Potential of synchrotron X-ray diffraction computed tomography (XRDCT) for a 3D non-destructive and/or non-invasive characterisation of cultural heritage geomaterials (CHG) [J]. Journal of Cultural Heritage, 2024,68:28-37.

[10] 黄瑞潇,杨玉飞,李海涛,等.黄河流域典型固体废物的指纹特征识别 [J]. 环境科学研究, 2024,37(8):1810-1818.
Huang R, Yang Y, Li H, et al. Fingerprint feature recognition of typical solid waste in the yellow river basin [J]. Research of Environmental Sciences, 2024,37(8):1810-1818.

[11] Li X, Yang Y, Die Q, et al. Characteristics of solid waste from common generation source in nonferrous smelting industry reveal a new classification method [J]. Heliyon, 2023,9(10):e20545.

[12] 吴怡青,刘 静,刘 新,等.矿物类中药滑石粉的 XRD Fourier 图谱研究 [J]. 中药材, 2019,42(12):2856-2860.
Wu Y, Liu J, Liu X, et al. Study on XRD Fourier spectra of mineral based traditional Chinese medicine talc powder [J]. Journal of Chinese Medicinal Materials, 2019,42(12):2856-2860.

[13] Qian X, Wang Y, Liu Z, et al. Establishment of XRD fourier fingerprint identification method of realgar decoction pieces and its anti-tumor activity in tumor-in-situ transplanted mice [J]. Journal of Ethnopharmacology, 2024,331:118303.

[14] 范 彬,陈佳亮,王 林,等.新形式非法倾倒铝灰案件生态环境损害鉴定评估研究 [J]. 环境生态学, 2023,5(12):35-9,46.
Fan B, Chen J, Wang L, et al. Research on identification and assessment of environmental damage of a new type event of illegal dumping of aluminum ash [J]. Environmental Ecology, 2023,5(12): 35-9,46.

[15] 董凡超.以“法”提质向“新”而行 [N]. 法治日报, 2024-10-24(5).
Dong F. To elevate the quality of 'law' and move towards 'novelty' [N]. 2024-10-24(5).

[16] Chen J, Zhang L, Song L, et al. Effects of illegal solid waste dumping on the structure of soil bacterial communities: A case study in China [J]. Toxics, 2025,13(1):20.

[17] Shan L, Xiaoyi W, Jun L, et al. Impact of metallurgy on the environment: based on an illegal mining pollution case in Beijing, China [J]. International Journal of Environmental Science and Technology, 2023,20(3):2857-2868.

[18] Otwong A, Jongmeewasin S, Phenrat T. Legal obstacles for the circular economy in Thailand: Illegal dumping of recyclable hazardous industrial waste [J]. Journal of Cleaner Production, 2021,302:126969.

[19] Kazamias G, Zorpas A A. Drill cuttings waste management from oil & gas exploitation industries through end-of-waste criteria in the framework of circular economy strategy [J]. Journal of Cleaner Production, 2021,322:129098.

[20] Wang P, Hu Y, Cheng H. Municipal solid waste (MSW) incineration fly ash as an important source of heavy metal pollution in China [J]. Environmental Pollution, 2019,252:461-475.

[21] Sampling R W. Draft Technical Guidance Planning, Implementation, and Assessment [R]: EPA530-D-02-002, United States Environmental Protection Agency, 2002.

[22] 王 娜,吴 磊,王家松,等.光释光测年中铀钍钾含量 ICP-MS 和 XRF 测量方法 [J]. 岩矿测试, 2024,43(5):793-801.
Wang N, Wu L, Wang J, et al. Determination of uranium, thorium and polassium for optically stimulated luminescence dating by ICP-MS and XRF [J]. Rock and Mineral Analysis, 2024,43(5):793-801.

[23] Chen K, Han S, Meng F, et al. Acid controlled washing of municipal solid waste incineration fly ash: Extraction of calcium inhibiting heavy metals and reaction kinetics [J]. Science of The Total Environment, 2024,909:168599.

[24] Hamada H M, Shi J, Abed F, et al. Recycling solid waste to produce eco-friendly ultra-high performance concrete: A review of durability, microstructure and environment characteristics [J]. Science of The Total Environment, 2023,876:162804.

[25] Yang J M, Yu P T, Kuo B C. A Nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data [J]. IEEE Transactions on Geoscience and Remote Sensing, 2010,48(3):1279-1293.

[26] Oh S Y, Chung K Y. Target speech feature extraction using non-parametric correlation coefficient [J]. Cluster Computing, 2014,17(3): 893-899.

[27] Macfarland T W, Yates J M. Kruskal–Wallis H-Test for oneway analysis of variance (ANOVA) by Ranks [M]//MACFARLAND T W, YATES J M. Introduction to nonparametric statistics for the biological sciences using R. Cham; Springer International Publishing. 2016:177-211.

[28] Albertus M. Asymptotic Z and chi-squared tests with auxiliary information [J]. Metrika, 2022,85(7):859-883.

[29] Poleto C, Merten G H, Minella J P. The identification of sediment sources in a small urban watershed in southern Brazil: An application of sediment fingerprinting [J]. Environmental Technology, 2009,30 (11):1145-1153.

[30] Nahm F S. Nonparametric statistical tests for the continuous data: The basic concept and the practical use [J]. Kja, 2016,69(1):8-14.

[31] Momeni F, Gildeh B S. Nonparametric tests for median in fuzzy environment [J]. International Journal of Fuzzy Systems, 2016,18 (1):130-139.

[32] Metsämuuronen J. Directional nature of Goodman–Kruskal gamma and some consequences: Identity of Goodman–Kruskal gamma and

Somers delta, and their connection to Jonckheere–Terpstra test statistic [J]. Behaviormetrika, 2021,48(2):283−307.

[33] 吴喜之.非参数统计 [M]. 5 版.北京:中国统计出版社, 2019:72−75.
Wu X. Nonparametric statistics [M]. 5th edition. Beijing: China Statistics Press, 2019:72−75.

[34] 王伏虎,赵喜仓.多个独立样本非参数检验中的 Jonckheere−Terpstra 方法 [J]. 统计与决策, 2009,(24):158−159.
Wang F, Zhao X. Jonckheere terpstra method in non parametric testing of multiple independent samples [J]. Statistics & Decision, 2009,(24):158−159.

[35] Xanthopoulos P, Pardalos P M, Trafalis T B. Linear discriminant analysis [J]. Robust data mining, 2013:27−33.

[36] 花 旗,夏庆霖,刘奇锋.基于数据驱动的南岭地区花岗岩岩体含矿性判别 [J]. 地质科技通报, 2023:1−17.
Hua Q, Xia Q, Liu Q. Ore−bearing discrimination of granite rock masses in the Nanling area via data−driven models [J]. Bulletin of Geological Science and Technology, 2025,44(1):332−345.

[37] Carpenter K A, Huang X. Machine learning−based virtual screening and its applications to Alzheimer's drug discovery: A review [J]. Current Pharmaceutical Design, 2018,24(28):3347−3358.

[38] Franz C, Makeschin F, Weiß H, et al. Sediments in urban river basins: Identification of sediment sources within the Lago Paranoá catchment, Brasilia DF, Brazil – using the fingerprint approach [J]. Science of The Total Environment, 2014,466−467:513−523.

[39] Uddin S, Khan A, Hossain M E, et al. Comparing different supervised machine learning algorithms for disease prediction [J]. BMC Medical Informatics and Decision Making, 2019,19(1):281.

[40] 肖再亮,王 波,李茂川,等.油基岩屑脱油残渣制备复合胶凝材料研究 [J]. 四川环境, 2023,42(5):8−14.
Xiao Z, Wang B, Li M, et al. Study on preparation of composite cementitious material from Oil. based drilling cuttings pyrolysis Residues [J]. Sichuan Environment, 2023,42(5):8−14.

[41] Aparicio C, Filip J, Machala L. From Prussian blue to iron carbides: high−temperature XRD monitoring of thermal transformation under inert gases. Powder Diffraction. 2017,32(S1):S207−S212.

[42] Qin J, Zhang Y, Heberlein S, et al. Characterization and comparison of gasification and incineration fly ashes generated from municipal solid waste in Singapore [J]. Waste Management, 2022,146:44−52.

[43] 马彦涛,房小健,黄绪泉,等.矿渣−硅灰−磷石膏基骨料的制备及性能 [J]. 化工环保, 2025,42(5):255−260.
Ma Y, Fang X, Huang X, et al. Preparation and properties of slag−silica fume−phosphogypsum−based aggregate [J]. Environmental Protection of Chemical Industry, 2025,42(5):255−260.

[44] Zhang M, Chen C, Mao L, et al. Use of electroplating sludge in production of fired clay bricks: Characterization and environmental risk evaluation [J]. Construction and Building Materials, 2018,159:27−36.

[45] 史齐勇,齐鹏宇,宁 超,等.钴冶炼废渣在冶金球团中的利用及球团性能提升机理 [J]. 安徽工业大学学报(自然科学版), 2025,42(1):1−8.
Shi Q, Qi P, Ning C, et al. Utilization of cobalt smelting slag in metallurgical pellets and the mechanism of pellet performance enhancement [J]. ournal of Anhui University of Technology(Natural Science), 2025,42(1):1−8.

[46] Zhang J, Liu B, Shen H, et al. Theoretical and experimental on the thermodynamic, kinetic and phase evolution characteristics of secondary aluminum ash [J]. Journal of Materials Research and Technology, 2022,19:3857−3866.

[47] Yiren W, Dongmin W, Yong C, et al. Micro−morphology and phase composition of lithium slag from lithium carbonate production by sulphuric acid process [J]. Construction and Building Materials, 2019,203:304−313.

[48] Hoshino S, Yamada K, Hirao H. XRD/Rietveld analysis of the hydration and strength development of slag and limestone blended cement [J]. Journal of Advanced Concrete Technology, 2006,4(3):357−367.

[49] Ren Q, Zhang Y, Long Y, et al. Crystallisation behaviour of blast furnace slag modified by adding fly ash [J]. Ceramics International, 2018,44(10):11628−11634.

[50] Shi Y, Huang W, Luo J, et al. Detection and discrimination of pests and diseases in winter wheat based on spectral indices and kernel discriminant analysis [J]. Computers and Electronics in Agriculture, 2017,141:171−180.

[51] Park C H, Kim S B. Sequential random k−nearest neighbor feature selection for high−dimensional data [J]. Expert Systems with Applications, 2015,42(5):2336−2342.

[52] Darst B F, Malecki K C, Engelman C D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data [J]. BMC Genetics, 2018,19(1):65.

[53] Kemsley E K. Discriminant analysis of high−dimensional data: a comparison of principal components analysis and partial least squares data reduction methods [J]. Chemometrics and Intelligent Laboratory Systems, 1996,33(1):47−61.

作者简介：黄瑞潇(2000−),男,山东济南人,中国环境科学研究院硕士研究生,主要研究方向为固体废物指纹特征时别.发表论文 1 篇.huangruixiao22@mails.ucas.ac.cn.