

智能算法在多环芳烃检测中的研究进展

刘园园^{1, 2*}

1. 商丘工学院信息与电子工程学院, 河南商丘, 476000

2. 江苏师范大学物理与电子工程学院, 江苏徐州, 221116

摘要 | 多环芳烃 (PAHs) 具有环境持久性、生物富集性及潜在致癌风险, 痕量快速准确检测PAHs对环境监测与风险评估具有重要意义。针对PAHs光谱检测中谱峰重叠、结构相似、基质干扰及低浓度信号波动等问题, 综述智能算法在PAHs分类识别、浓度预测和混合组分分析中的研究进展。介绍了机器学习和深度学习在PAHs检测中的研究进展, 分析了不同算法在学习机制、数据依赖性、光谱特异性和跨基质泛化能力等方面的本质差异及对PAHs智能检测的准确性、稳健性和现场应用能力。进一步分析PAHs智能检测的关键瓶颈与研究路线, 为算法选择和现场应用提供参考。未来将加强标准化数据集构建、跨域迁移及模型适用性研究, 提高PAHs智能检测的可靠性、可解释性及复杂场景适应能力。

关键词 | 多环芳烃; 表面增强拉曼光谱; 机器学习; 深度学习

Copyright © 2026 by the author(s).

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

<https://creativecommons.org/licenses/by-nc/4.0/>



1 引言

多环芳烃 (Polycyclic aromatic hydrocarbons, PAHs) 是由两个及以上苯环组成的疏水性持久污染物, 具有化学稳定性高、环境滞留时间长和易发生生物富集等特征^[1-4]。如图1所示, PAHs主要来源于煤炭、石油及其他有机物的不完全燃烧, 通过大气沉降、地表径流和工业排放等途径进入水体、土壤、沉积物及生物体, 对生态与人体健康造成潜在风险^[5-9]。PAHs具有致癌性、致突变性, 部分高环 (≥ 4 环) 组分在体内代谢后生成更具致癌活性产物, 可诱发DNA损伤并干扰细胞结构与生理功能^[10-11]。因此, 快速、准确且高灵敏的痕量PAHs检

测技术对环境污染进行早期预警、风险评估与精准治理具有重要意义。

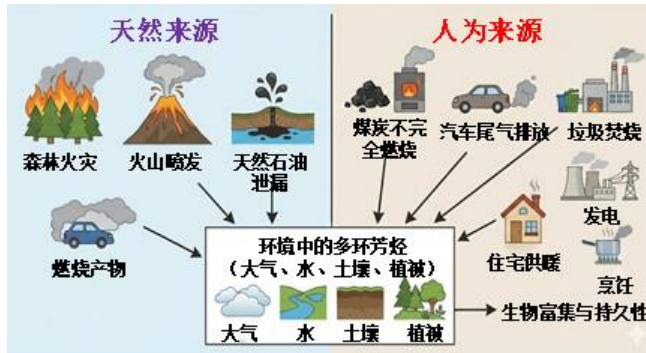


图1 多环芳烃来源^[作者自绘]

Figure 1 Sources of PAHs^[Illustrated by the author]

基金项目: 江苏省研究生科研与实践创新计划项目 (SJCX25_1504)。

作者简介: 刘园园, 商丘工学院信息与电子工程学院讲师, 硕士, 研究方向为 SERS 检测与深度学习。

文章引用: 刘园园. 智能算法在多环芳烃检测中的研究进展[J]. 计算机与智能教育技术, 2026, 1(1): 29-39.

<https://doi.org/10.68070/606ank63>

目前, PAHs检测方法主要包括气相色谱^[12-13]、高效液相色谱^[14-15]、气相色谱-质谱^[16-17]和荧光光谱法^[18]。这些方法在实验室条件下具有较高的定性和定量能力, 但通常需要萃取、净化、浓缩和分离等前处理步骤, 难以同时满足快速检测、低成本分析和现场应用需求。表面增强拉曼光谱(Surface Enhanced Raman Scattering, SERS)^[19]是一种具有高灵敏度和分子指纹特征的拉曼增强技术, 用于目标分子的结构识别, 可在较短时间内获得清晰谱峰信息, 为痕量PAHs快速检测提供了新的技术手段。

尽管SERS显著增强PAHs的特征拉曼信号, 但在实际检测中仍受到谱峰重叠、分子结构相似、基底吸附取向差异以及信噪比不足等因素影响, 导致传统方法难以实现稳定的识别分类与定量^[20]。人工智能算法通过降维、判别建模和端到端特征学习提升复杂光谱分析能力, 在PAHs的分类识别与浓度回归预测中发挥关键作用, 其技术演进由“人工特征工程与浅层模型”向“自动特征学习与深层网络”递进。因此, 本文阐述了传统机器学习主要包括支持向量机^[21]、随机森林^[22]、偏最小二乘回归^[23]等研究现状, 探讨卷积神经网络^[24]、长短期记忆网络^[25]和注意力机制^[26]等深度学习算法性能和应用以及未来发展趋势。

2 传统机器学习模型

机器学习于1956年由Arthur Samuel首次提出^[27], 是计算机训练大量的输入数据并对其建模, 再将新数据输入到模型中, 通过优化模型参数对新数据进行预测或分类^[28]。常用算法有支持向量机(Support vector machine, SVM)、随机森林

(Random forest, RF)及偏最小二乘回归(Partial least squares regression, PLSR)等。传统机器学习通过光谱预处理、特征提取及监督建模, 建立PAHs光谱信息与物质类别或浓度之间的映射关系, 是光谱智能分析的重要技术路径。

2.1 主成分分析

主成分分析(Principal component analysis, PCA)是通过正交变换将原始高维变量压缩为少数主成分, 并使前几个主成分保留数据的总体变异信息^[29-30]。在PAHs检测中, PCA通常用于压缩高维光谱、减少变量共线性、识别异常样本并观察样品之间聚类趋势^[31-32]。因此, PCA通常作为后续LDA、SVM或回归模型的前置步骤。

Chen等^[33]利用碳量子点复合薄膜产生交叉响应指纹, 结合PCA和LDA实现了16种PAHs的识别和定量, 分类准确率为100%, 最低检测限达到57 nM, 表明PCA能够综合不同传感单元和激发条件下的响应差异, 适合处理交叉响应型传感阵列数据。Ghohestani等^[34]进一步将纸基微流控荧光阵列、智能手机成像与PCA-LDA相结合, 通过压缩RGB多变量信息实现10种PAHs的便携式识别, 训练集准确率为100%, 在5.0 mg/L条件下交叉验证准确率为93%, 体现了PCA在低成本现场检测中的应用潜力。Wang等^[35]通过Au与聚离子液体(PIL)纳米碗组装作为基底, 并结合PCA和SVM算法, 在真实河水中实现结构相近PAHs及其二元、三元混合物的稳健区分, 如图2所示。该模型准确率达到100%, 且二元混合物的主成分空间距离与组分摩尔比呈显著线性关系($R^2=0.998$), 表明该模型具有良好的混合组分识别与比例表征能力。

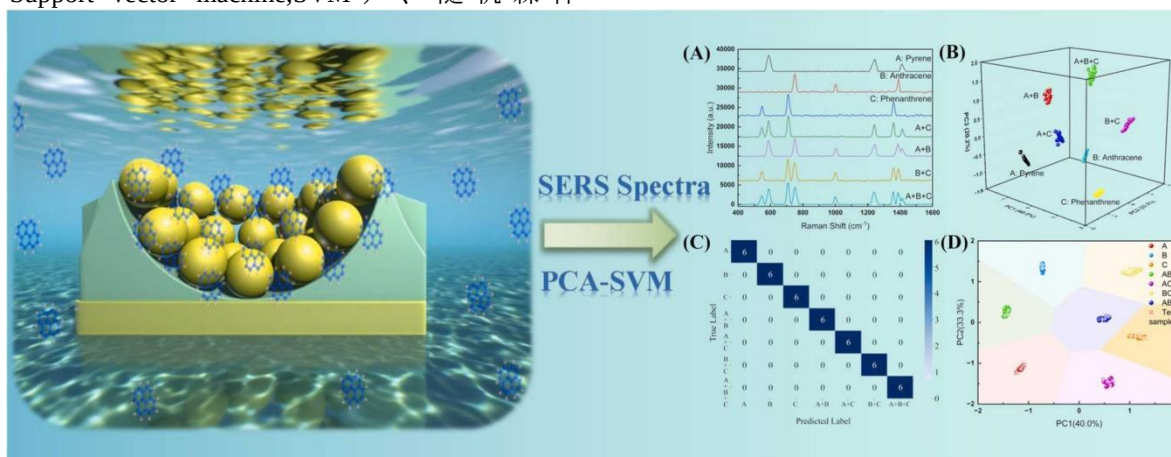


图2 SERS传感器中机器学习算法应用^[35]

Figure 2 Applications of machine learning algorithms in SERS sensors^[35]

PCA数据依赖性较低,适合样本量有限时的数据探索、异常识别和后续建模的前置降维。但其光谱特异性相对较弱,主成分优先保留总体方差较大的方向,无法保留关键信息,难以描述PAHs光谱中的非线性类别边界,不能自动学习多层次光谱特征^[36-37]。当样品基质、SERS基底批次或仪器背景变化产生的方差大于目标物差异时,PCA首先表征基质而非PAHs本身,跨基质泛化能力因而较弱。故PCA更适合作为探索性分析和前置特征压缩工具,不宜单独承担复杂真实样品的最终判别或定量任务。

2.2 支持向量机

支持向量机(support vector machine, SVM)是一种基于统计学习理论的监督学习方法,通过在特征空间中构建最优分类超平面实现类别判定,并可借助核函数将非线性问题映射到高维空间;用于连续变量预测时则扩展为支持向量回归(SVR)。

王书涛等人^[38]利用三维荧光光谱法获取3种PAHs混合物的重叠荧光响应,将基于遗传算法(GA)优化的支持向量机(SVM)应用于PAHs混合物种类的检测中,准确率达95.42%,表明GA-SVM模型能够从重叠光谱中提取稳定的判别信息,准确区分不同PAHs混合物类别,如图3所示。吴兴等人^[39]采用平行因子结合粒子群(PSO)优化SVM算法实现水中PAHs单质物质(乙萘、氟烃和萘)检测,分类准确率为100%。乙萘、氟烃和萘的回收率分别为 $100.45\% \pm 6.25\%$ 、 $100.10\% \pm 6.39\%$ 和 $95.07\% \pm 7.46\%$,该模型通过参数优化与多维光谱分解提高了SVM对复杂荧光信号的分类性能。

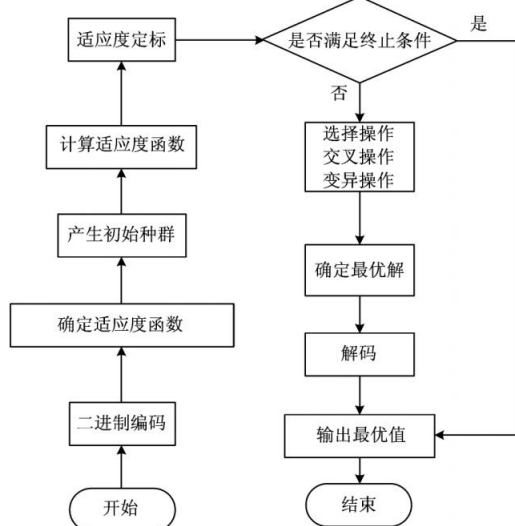


图3 GA 优化 SVM 参数算法流程图^[38]

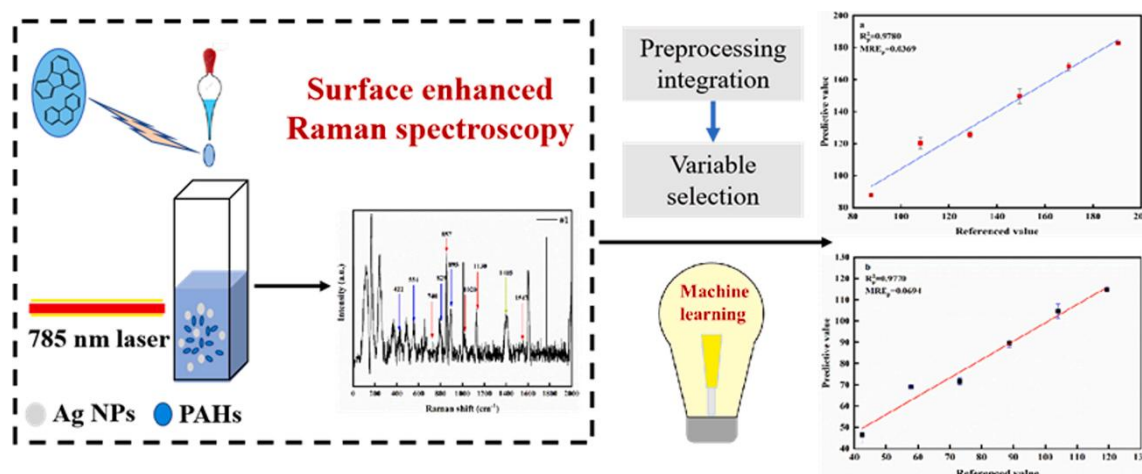
Figure 3 Flowchart of the GA-optimized SVM parameter algorithm^[38]

SVM在小样本条件下具有较好的泛化能力,适用于高维数据,并能够通过核函数处理非线性问题^[40]。相较PCA, SVM利用类别或浓度标签,光谱判别特异性更强,适用于结构相近PAHs的分类以及有限样本条件下的回归。其局限在于模型性能对核函数及惩罚参数较为敏感;当基质、基底或仪器导致光谱分布整体偏移时,原有决策边界失效,跨基质泛化需重新标定、参数优化或迁移校正。当样本规模和类别数量增大时,训练成本增加。此外, SVM虽然能够建立非线性判别边界,但难以直观揭示具体谱峰对预测结果的贡献。

2.3 随机森林

随机森林(Random forest, RF)是一种基于自主重采样技术的非线性数据处理算法^[41],根据信息增益比或基尼指数选择最优特征进行分割。RF模型通过构建多棵具有差异性的决策树,利用非线性阈值组合筛选具有代表性的峰位和谱段,适合处理变量多、噪声强及谱峰交互明显的的数据。该算法具有高预测精度、易于克服过拟合缺陷以及强大的抗噪声能力的优势,在定量分析中被广泛使用^[42-43]。此外,结合预处理方法或变量选择可以突出数据矩阵与浓度之间的关系,从光谱中提取最有效的信息,进一步提升射频校准模型的预测性能和鲁棒性。

Guo等人^[44]以Ag NPS为基底,结合RF校准模型对水中菲和芴进行定量分析,获得最低平均相对预测误差(MREP)和最大决定系数(R²),如图4所示。菲和芴的R²分别为0.978和0.977, MREP分别为0.0369和0.0694,该方法有效区分谱峰重叠,能够快速且简便地监测水体中PAHs的含量。Douglas等人^[45]采用RF结合可见光-近红外(vis-NIR)光谱技术,对污染土壤中的PAHs含量进行快速原位评估,土壤光谱与PAHs含量之间呈现良好的非线性关系,预测性能(R²=0.71, RMSEP=0.99, RPD=1.99)优于PLSR模型,该模型在PAHs快速筛查和现场原位检测方面具有一定应用潜力。Zhang等人^[46]将SERS技术与RF校准模型相结合,评估了复杂水体基质中PAHs定量分析及致癌风险预测的可行性。其湖水、自来水和蒸馏水中的菲、苯并[a]芘和荧蒽检测的MREP分别为0.0646、0.0949和0.0992, R²分别为0.9658、0.9537和0.9551,表明RF模型能够在不同水体基质条件下实现PAHs的有效定量分析,为复杂环境水体中PAHs污染水平及其潜在致癌风险评估提供数据支撑。

图4 水中 PAHs 定量分析方法原理图^[44]Figure 4 Schematic diagram of a method for the quantitative analysis of PAHs in water^[44]

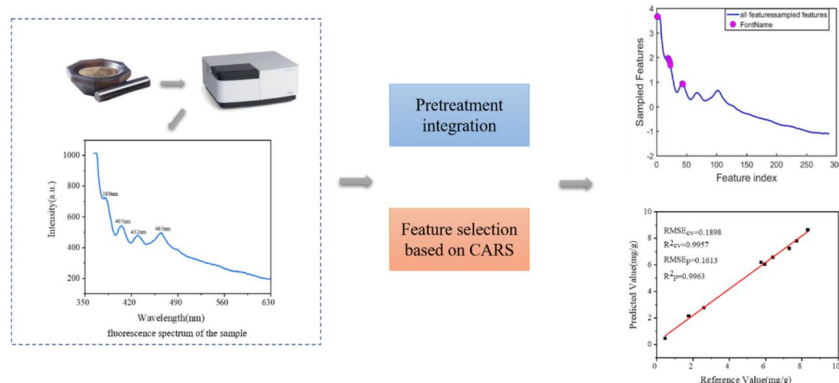
RF模型具有较强的非线性建模能力及抗噪声和抗过拟合性能, 适合处理变量较多、谱峰重叠明显的SERS数据^[47]。然而, 随着决策树数量和模型规模增加, 其计算与存储成本也会提高; 对于密集且高度相关的光谱变量, 重要性变量指标难以准确区分各变量的独立贡献^[48]。此外, RF主要依据单个或多个变量的阈值组合进行预测, 对跨波段序列依赖和多层次光谱特征的自动学习能力有限, 且无法消除SERS基底批次差异造成的光谱分布变化。因此, RF更适合作为非线性、小中等规模数据的稳健基线模型。

2.4 偏最小二乘回归模型

偏最小二乘回归 (partial least-squares regression, PLSR) 通过考虑自变量矩阵与目标变量之间的相关关系, 提取潜在变量并建立回归模型, 该方法适用于变量多及变量之间存在明显共线性的情况, 在光谱定量分析中应用广泛^[49]。

Li 等人^[50]将竞争性自适应重加权采样

(CARS) 与偏最小二乘回归 (PLSR) 结合, 检测土壤中菲的含量, CARS-PLSR较原始光谱-偏最小二乘校准模型 (RAW-PLSR) 和小波变换-标准正态变异 (WT-SNV) 校准模型表现出更优预测性能, 校正集和预测集的 R^2 分别为0.9957和0.9963, RMSE分别为18.98%和16.13%。研究表明CARS变量筛选能够有效减少冗余信息和干扰变量, 提高PLSR模型对土壤中菲含量的定量预测能力。Feng等人^[51]融合拉曼光谱与红外光谱, 采用PLSR对土壤中蒽和荧蒽进行定量分析。该模型较单一光谱模型具有更高的预测精度, 蒽和荧蒽的 R^2 分别为0.9514和0.9609, RMSE分别为0.9143和0.8614, RPD均高于4, 表明多源光谱融合PLSR模型能够有效整合不同光谱技术的互补信息, 提升了PLSR模型在土壤中PAHs的定量预测能力。史晓凤^[52]采用偏最小二乘法结合神经网络 (PLSR-ANN) 的方法对水中萘、菲和芘三种PAHs进行了定量分析, 获得三种PAHs的平均相对误差分别是29.4%、27.8%和21.5%, 实现了混合溶液中各PAHs的定量检测。

图5 CARS-PLSR 模型检测土壤中菲含量原理图^[50]Figure 5 Schematic diagram of the CARS-PLSR model for detecting phthalate levels in soil^[50]

PLSR模型具有结构简洁、计算效率高、适合小样本数据以及结果具有一定化学解释性等优点，常被用作基准模型或特征筛选工具^[53-55]。然而，该模型性能依赖光谱预处理、变量筛选和潜变量数的人工优化，竞争吸附、非线性SERS增强及复杂基质会破坏既有协方差结构，使模型对跨基底、跨仪器和跨基质泛化能力受到限制。

综上所述，基于SERS光谱的传统机器学习方法

具有计算快、训练样本需求较小、对硬件要求低的优点，但这类方法易受基质干扰、依赖人工特征设计，流程繁琐且对参数敏感，当数据规模变大、维度更高，模型性能受到影响，如表1所示的机器学习方法对比。因此，研究正逐步转向深度学习端到端建模或解析特征峰提取等策略，以减少特征工程并提高鲁棒性。

表 1 传统机器学习方法在 PAHs 光谱检测比较^[作者自绘]

Table 1 Comparison of traditional machine-learning methods for PAHs spectral detection^[Illustrated by the author]

方法	数据依赖性	光谱特异性	跨基质泛化能力
PCA	低	保留大方差方向、不保证对应目标峰	易把基质、背景方差作为主成分
SVM	低—中	利用标签形成判别边界、缺少谱段结构先验	决策边界依赖训练分布，域偏移时需再标定
RF	中	筛选变量关键谱段、共线波段贡献易分散	覆盖多基质训练时较稳健，系统性分布漂移明显
PLSR	低—中	线性定量下目标相关性强、可解释	协方差结构随基质、仪器变化而改变

3 智能深度学习算法模型

近年来，深度学习^[56]逐渐应用在SERS光谱分析，通过多层非线性网络将特征提取与分类、回归整合为端到端过程，能够从高维光谱中逐级学习峰位、峰强、峰形及跨谱段组合关系，其性能优于传统机器学习，包括CNN侧重局部峰形与邻域模式，LSTM侧重沿光谱序列的长程依赖，注意力机制和Transformer进一步建模全局关联并突出关键谱段。

3.1 卷积神经网络

卷积神经网络（CNN）是通过卷积运算自动学习输入数据局部特征和高层抽象特征的前馈神经网络，由输入层、卷积层、池化层、全连接层和输出层组成。与传统全连接神经网络相比，CNN不依赖大量人工特征设计，通过多层网络结构自动完成特征提取与表达，同时借助权值共享减少参数数量，提高训练效率和泛化能力。

Atta等^[57]人制备金纳米星（GNSs）包裹十六烷基三甲基溴化铵（CTAB）的SERS基底，结合卷积神经网络（CNN）检测污染水中PAHs，如图6所示，该模型准确率为90%，F1为94%，为复杂实际环境中分子的SERS检测提供了一种灵敏、选择性强的方法。Qiu等^[58]针对1-羟基芘这一典型羟基化PAHs暴露生物标志物，建立了β-CD包覆金纳米颗

粒自组装膜SERS检测方法，并结合卷积神经网络（CNN）和遗传算法（GA）实现了尿液中1-OHPyr快速定量，如图7所示。其R2和RMSEP分别为0.9639和0.6327，优于其他对比模型，该模型能够快速、灵敏、准确地检测尿液中的1-OHPyr，为评估PAHs暴露的健康风险提供参考。

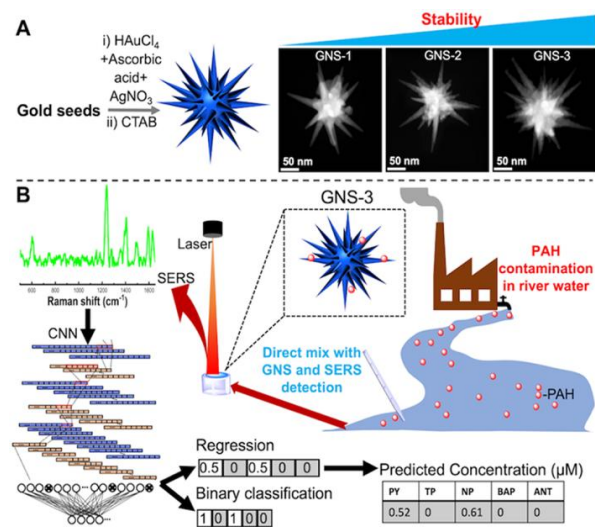


图 6 基于 CTAB 修饰金纳米星与深度学习的环境水体

PAHs 的 SERS 智能检测平台^[57]

Figure 6 A smart SERS detection platform for PAHs in environmental water based on CTAB-modified gold nanostars and deep learning^[57]

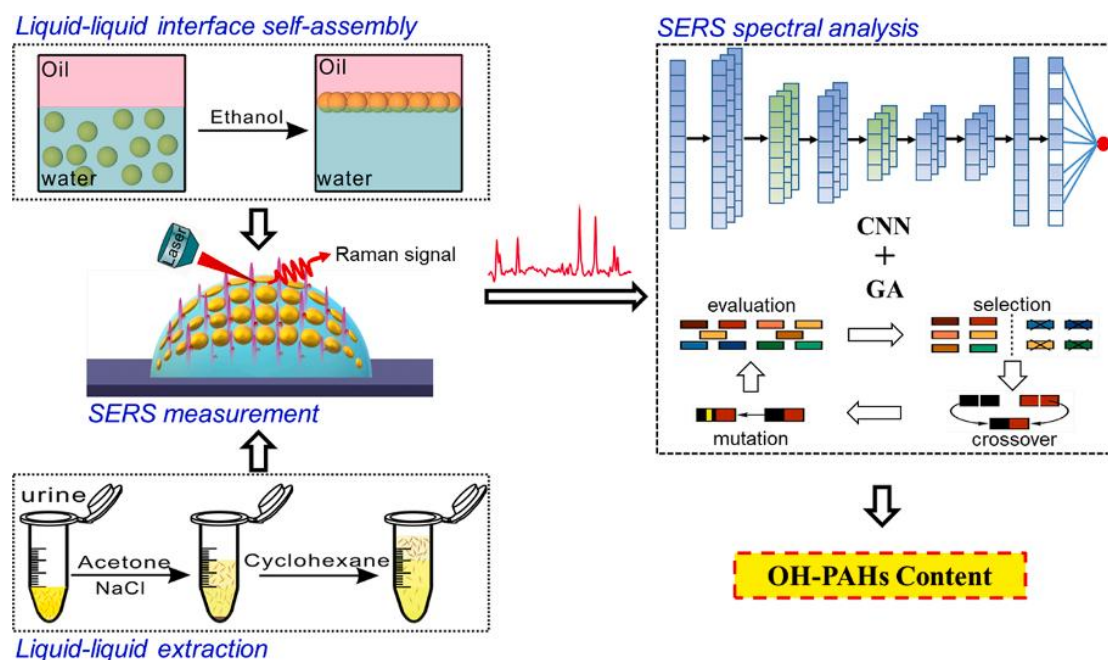


图7 基于液-液界面自组装 SERS 薄膜与深度学习的人体尿液中1-OHPyr 检测流程图^[58]

Figure 7 Flowchart for the detection of 1-OHPyr in human urine using a SERS film based on liquid-liquid interface self-assembly and deep learning^[58]

CNN通过局部卷积核和权值共享自动提取峰位邻域、峰形及相邻峰组合特征，因而对局部光谱结构具有较高特异性，并适用于结构相似、谱峰重叠明显PAHs的分类与浓度回归。与传统机器学习相比，CNN减少了人工特征设计，但数据依赖性较高，需要较多代表性的训练样本，并且对训练集与测试集之间的仪器条件、基底批次和样品基质差异较为敏感。同时，卷积层更关注局部特征，对相距较远谱段之间的关联建模能力有限，更适合局部峰形主导与样本量较充足的任务。

3.2 长短期记忆神经网络

长短期记忆网络（long short-term memory, LSTM）是通过遗忘门、输入门、输出门和细胞状态控制信息保留与更新，可缓解普通循环神经网络在长序列训练中的梯度消失问题^[59]。这种门控机制使LSTM能够同时捕捉短期和长期依赖关系，使其成为时间序列预测和基于序列的回归等任务的理想选择。

Adade等人^[60]通过LSTM捕获光谱序列特征，实现二苯并(a,h)蒽（DbA）和苯并(b)荧蒽（BbF）的定量检测，如图8所示。相比传统机器学习模型，LSTM模型表现最佳，DbA和BbF的RMSE值最低（RMSE<0.09），R2值最高（R2>0.99），RPD值同样最高（DbA为6.63，BbF为5.83），均有显著提升，表明LSTM模型能够捕获和利用序列数据中的长期依赖关系提高SERS定量性能。徐桂烽^[61]将变分模态分解（VMD）、样本熵（SE）与LSTM相结

合，构建了VMD-SE-LSTM混合模型，用于河流中PAHs含量预测。对比5种模型，VMD-SE-LSTM的平均绝对误差分别降低82.12%、81.32%、70.02%、66.39%和13.00%，表明多尺度分解与LSTM深度模型的结合有效提高河流PAHs浓度预测准确性。

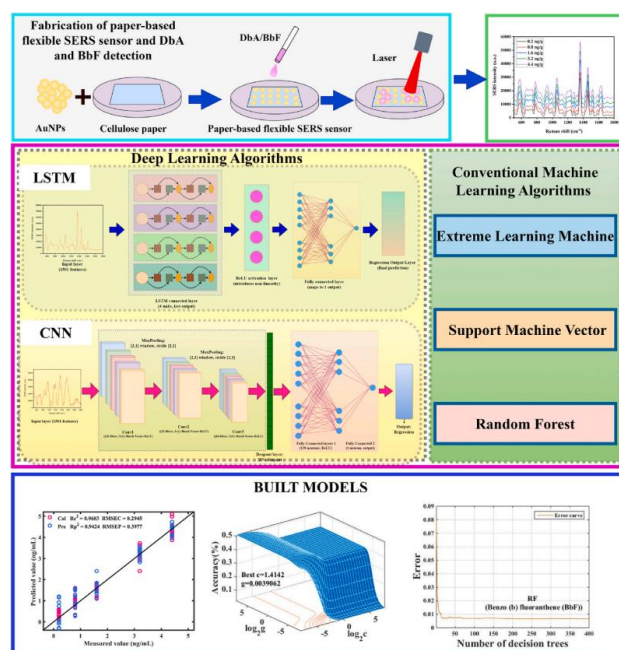


图8 纸基柔性 SERS 传感器深度学习算法预测 DbA 和 BbF 方法原理图^[60]

Figure 8 Schematic diagram illustrating the principle of the deep learning algorithm for predicting DbA and BbF using a paper-based flexible SERS sensor^[60]

相比传统循环神经网络, LSTM能够在较长序列中稳定传递关键特征, 减缓梯度消失导致的长期依赖难以学习问题, 使重要信息在多次递推后仍可被有效利用并作用于后续预测。但面对结构复杂的时空数据时, 难以有效突出序列中的关键信息, 从而误判PAHs物质。当基质变化改变整体峰强和序列模式时, 模型出现分布偏移, 跨基质泛化能力偏低。故LSTM更适用于具有明显长程相关的PAHs预测。

3.3 注意力机制

注意力机制 (attention mechanism) 依据输入特征对当前任务的贡献动态分配权重, 使模型能够突出关键谱段并抑制冗余背景^[62-63]。注意力机制及基于自注意力的Transformer结构可以建立不同谱段之间的关联, 能够同时表征弱特征峰、峰位偏移和跨谱段组合关系。注意力权重反映模型内部的统计关联, 并不等同于化学因果机制; 若训练数据存在

批次偏差, 模型关注与目标无关的背景特征^[64]。

Yu等^[63]构建了融合双通道CNN、Transformer自注意力机制和自适应阈值的SpecRecFormer模型, 以单组分拉曼光谱为基础识别混合PAHs组分。该模型兼顾局部峰形提取与跨波段全局关联, 在3个真实混合光谱数据集上的识别准确率分别为93.75%、89.21%及93.63%, 显著优于传统神经网络模型, 表明其在混合样本有限条件下具有较好的迁移识别能力。Tian等人^[65]构建了麻雀搜索算法优化的CNN-LSTM-Attention模型, 并结合磁驱动富集和环糊精修饰SERS基底, 实现了水中痕量PAHs检测, 如图9所示。该模型利用CNN提取局部谱峰、峰形和峰强特征, LSTM学习沿拉曼位移方向的序列依赖关系, Attention模块突出高判别价值特征, SSA模型优化超参数, 其测试集分类准确率超98%, 检测限低至10-8 mol/L, 为结构相近PAHs的峰重叠和类别判定问题提供了技术思路。

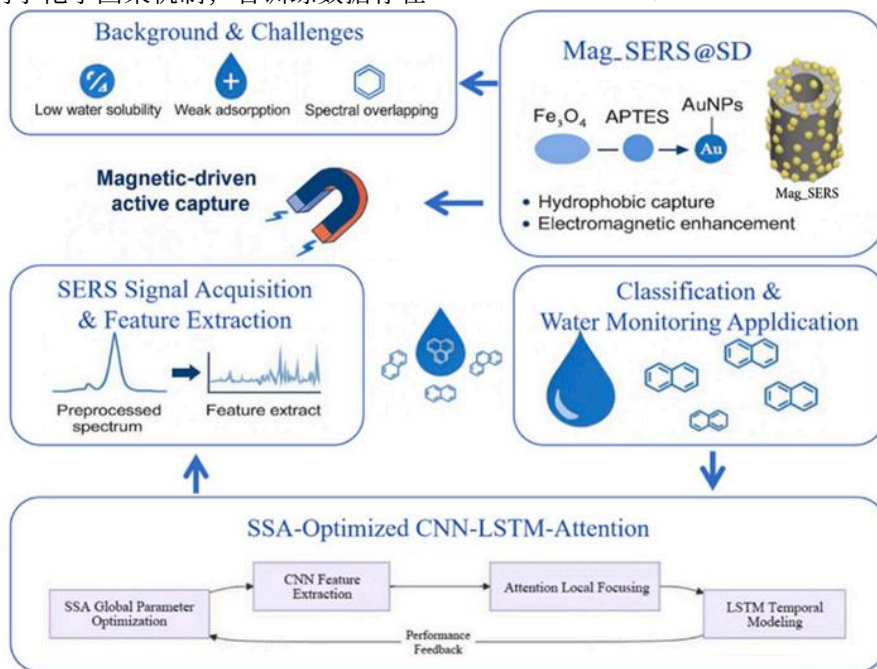


图9 PAHs特异性基底与深度学习算法协同检测原理图^[65]

Figure 9 Schematic diagram illustrating the principle of collaborative detection using a PAHs-specific substrate and a deep learning algorithm^[65]

注意力模型能够突出弱特征峰和跨波段关联信息, 在缓解谱峰重叠、背景干扰及混合组分识别困难方面具有优势, 同时减少了对人工特征选择的依赖。然而, 此类模型通常结构复杂、参数量较大, 对训练样本规模、超参数设置及数据分布较为敏感, 且注意力模型理论上可整合更广泛的上下文信息, 但无法自动消除基质、基底和仪器引起的域偏移。因此, 此类模型更适合样本量较充足、需要全局谱段关联和混合组分识别的任务。

综上所述, SERS技术能够有效增强PAHs的分

子指纹信号, 深度学习能够有效挖掘SERS光谱中的局部峰形特征、序列相关信息和非线性浓度响应关系, 对解决PAHs同系物结构相似、谱峰重叠和复杂基质干扰问题具有重要意义。二者结合为痕量PAHs的快速、灵敏和智能化检测提供了新的技术路径。深度学习模型在PAHs光谱特征提取与建模能力上各具优势, 其数据依赖性、光谱特异性及跨基质泛化能力存在明显差异, 如表2所示。当测试域超出训练分布时, 深度学习模型学习到非目标背景, 影响检测结果。因此, 后续研究应重点关注多域数据覆

盖、化学先验约束及模型不确定性评估，并进一步用可行性。
结合实际检测设备验证模型在资源受限条件下的应

表 2 深度学习方法在 PAHs 光谱检测中的比较^[作者自绘]

Table 2 Comparison of deep-learning methods for PAHs spectral detection^[Illustrated by the author]

方法	数据依赖性	光谱特异性	跨基质泛化能力
CNN	高	善于峰位邻域、峰形及局部组合	易学习基底、背景捷径特征
LSTM	高	利用跨波段、时间顺序关系	序列模式随基质改变时易失配
注意力/Transformer	高—很高	适合弱峰、多峰协同和混合组分	具备全局建模潜力，不能自动消除域偏移

4 总结与展望

PAHs具有环境持久性、生物富集性和潜在致癌风险，快速、准确检测PAHs对环境监测和风险评估具有重要意义。随着光谱检测技术与人工智能算法的深度融合，人工智能算法能够提高PAHs光谱检测中重叠谱峰解析、结构相近组分识别及非线性定量能力。

PCA、SVM、RF和PLSR等传统机器学习方法在小样本、低计算成本及定量建模方面具有一定优势，而CNN、LSTM及注意力/Transformer等深度学习方法能够进一步挖掘局部峰形、跨波段序列关系和全局谱段关联，在结构相近PAHs识别、谱峰重叠解析、混合组分判别及非线性浓度预测等方面表现出较大潜力。

复杂基质、不同SERS基底批次及光谱分布漂移，低浓度和混合组分样本不足等因素制约模型在真实环境中应用。因此，后续研究将沿“标准化数据—跨域建模—可信验证—现场部署”的技术路线，构建覆盖多基质、多批次的标准化数据集，将特征峰、分子振动和吸附机制等化学先验嵌入模型，并引入不确定性评估提高结果的可解释性与可信度；最终以检测限、回收率等为综合指标推进模型现场设备集成。未来PAHs智能检测的发展由单一数据集上的预测精度评价逐步转向对跨域泛化、检测限、回收率及实际应用稳定性的综合考察，并进一步在便携式检测平台上验证算法的运行效率和适用性，从而推动构建可迁移、可解释和具有现场应用潜力的智能检测体系。

致谢

本工作得到了江苏省研究生科研与实践创新计划项目（SJCX25_1504）的支持。

数据可用性

数据将根据请求提供。

利益冲突声明

作者声明他们没有已知的可能影响本文所报告工作的竞争性财务利益或个人关系。

参考文献

[1] A. B. Patel, S. Shaikh, K. R Jain, et al. Polycyclic aromatic hydrocarbons: sources, toxicity, and remediation approaches[J]. *Frontiers in Microbiology*, 2020, 11: 562813.

[2] 李花, 赵立坤, 包仕钰, 等. 多环芳烃降解菌及其应用研究进展[J]. *环境工程技术学报*, 2023, 13(5): 1663-1676.

[3] J. Wilson, U. Pöschl, M. Shiraiwa, et al. Non-equilibrium interplay between gas-particle partitioning and multiphase chemical reactions of semi-volatile compounds: mechanistic insights and practical implications for atmospheric modeling of polycyclic aromatic hydrocarbons[J]. *Atmospheric Chemistry and Physics*, 2021, 21: 6175-6198.

[4] A. Mojiri, I. Zhou, A. Ohashi, et al. Comprehensive review of polycyclic aromatic hydrocarbons in water sources, their effects and treatments[J]. *Science of the Total Environment*, 2019, 696: 133971.

[5] S. Chimjarn, O. Delhomme, M. Millet. Temporal distribution and gas/particle partitioning of polycyclic aromatic hydrocarbons (PAHs) in the atmosphere of Strasbourg, France[J]. *Atmosphere*, 2021, 12(3): 337.

[6] B.P. Singh, K. Kumar, V.K. Jain. Distribution of ring PAHs in particulate/gaseous phase in the urban city of Delhi, India: Seasonal variation and cancer risk assessment[J]. *Urban Climate*, 2021, 40: 101010.

[7] 王东梅, 陈千喜, 张雪琪, 等. 我国生态环境中多环

- 芳烃污染状况研究进展[J]. 环境监测管理与技术, 2022, 34(5): 10-15.
- [8] L. Famiyeh, K. Chen, J. Xu, et al. A review on analysis methods, source identification, and cancer risk evaluation of atmospheric polycyclic aromatic hydrocarbons[J]. *Science of the Total Environment*, 2021, 789: 147741.
- [9] N. Sun, S. Yu, Z. Cai, et al. Inhibition of polycyclic aromatic hydrocarbon(PAHs) release from sediments in an integrated rice and crab coculture system by rice straw biochar[J]. *Journal of Cleaner Production*, 2022, 367: 133058.
- [10] 黄丹, 黄勇, 安永龙, 等. 北运河流域(北京段)表层土壤多环芳烃空间分布特征及来源解析[J]. *水文地质工程地质*, 2023, 50 (3): 159-171.
- [11] Y. Zhou, Q. Zou, M. Fan, et al. Highly efficient anaerobic co-degradation of complex persistent polycyclic aromatic hydrocarbons by a bioelectrochemical system[J]. *Journal of Hazardous Materials*, 2020, 381: 120945.
- [12] 韩永辉, 侯博, 安国荣. 环境中多环芳烃分析方法研究进展[J]. *化学分析计量*, 2025, 34 (5): 141-148.
- [13] 顾芳芳. GC×GC-FID定性定量分析柴油中多环芳烃[J]. *山西化工*, 2024, 44 (6): 102-103, 110.
- [14] 李沁怡. 基于等离子共振热点可调的SERS分析研究环境中多环芳香族化合物[D]. 上海: 上海师范大学, 2025.
- [15] N. Akvan, G. Azimi, H. Parastar. Chemometric assisted determination of 16 PAHs in water samples by ultrasonic assisted emulsification microextraction followed by fast high-performance liquid chromatography with diode array detector[J]. *Microchemical Journal*, 2019, 150: 104056.
- [16] 郑轩. 基于车载测试的重型柴油车黑碳与多环芳烃排放特征研究[D]. 北京: 清华大学, 2017.
- [17] R. Pissinatti, C.M. Nunes, A.G.D. Souza, et al. Simultaneous analysis of 10 polycyclic aromatic hydrocarbons in roasted coffee by isotope dilution gas chromatography-mass spectrometry: Optimization, in-house method validation and application to an exploratory study[J]. *Food Control*, 2015, 51: 140-148.
- [18] 刘婷婷. 基于三维荧光光谱法对水中多环芳烃定性定量分析[D]. 秦皇岛: 燕山大学, 2018.
- [19] Y.Y. Liu, Y.P. Tian, S. Chen, et al. Rapid detection and mechanistic analysis of 3,4-Benzo(a)pyrene by surface-enhanced Raman scattering[J]. *Surface Innovations*, 2025, 14(1): 3-12.
- [20] X.Y. Li, R. Wang, X.Y. Lan, et al. Portable and in situ rapid detection of benzo[a]pyrene: a review[J]. *Food Chemistry*, 2026, 498(Pt 1): 147130.
- [21] R. Pal, L. Arcamo, R. Farnood. Predicting the occurrence of substituted and unsubstituted, polycyclic aromatic compounds in coking wastewater treatment plant effluent using machine learning regression[J]. *Chemosphere*, 2024, 361: 142476.
- [22] M.H. Roy, D. Larocque. Robustness of random forests for regression[J]. *Journal of Nonparametric Statistics*, 2012, 24 (4): 993-1006.
- [23] M. Hou, Y. Huang, L. Ma, et al. Quantitative analysis of single and mix food antiseptics basing on SER S spectra with PLSR method[J]. *Nanoscale Research Letters*, 2016, 11 (1): 1-8.
- [24] 韩允童, 王靖岳, 侯兴达, 等. 基于改进CNN-LSTM模型的在役轴承寿命预测方法[J]. *机械强度*, 2026, 48 (2): 40-46.
- [25] 高士娟, 陈伟, 丁娟. 基于CNN-LSTM模型的高重频脉冲激光信号实时识别方法[J]. *激光杂志*, 2026, 47 (1): 232-238.
- [26] 雷柯松, 吐松江·卡日, 伊力哈木·亚尔买买提, 等. 基于WGAN-GP和CNN-LSTM-Attention的短期光伏功率预测[J]. *电力系统保护与控制*, 2023, 51 (09): 108-118.
- [27] J.W. Tang, Q.H. Liu, X.C. Yin, et al. Comparative analysis of machine learning algorithms on surface enhanced Raman spectra of clinical *Staphylococcus* species[J]. *Frontiers in Microbiology*, 2021, 12: 696921.
- [28] S. Pahlow, S. Meisel, D. Cialla-May, et al. Isolation and identification of bacteria by means of Raman spectroscopy[J]. *Advanced Drug Delivery Reviews*, 2015, 89: 105-120.
- [29] J.R. Beattie, F.W.L. Esmonde-white. Exploration of principal component analysis: Deriving principal component analysis visually using spectra[J]. *Applied Spectroscopy*, 2021, 75(4): 361-375.
- [30] A.S. Souza, M.A. Bezerra, U.M.F.W. Cerqueira, et al. An introductory review on the application of principal component analysis in the data exploration of the chemical analysis of food samples[J]. *Food Science and Biotechnology*, 2024, 33(6): 1323-1336.
- [31] W.J. Peveler, A. Roldan, N. Hollingsworth, et al. A sensor array for the discrimination of polycyclic aromatic hydrocarbons using conjugated polymers and the inner filter effect[J]. *Chemical Science*, 2019, 10: 10034-10040.
- [32] W. Liu, S.G. Sun, Y. Liu, et al. Determination of benzo(a)pyrene in peanut oil based on Raman spectroscopy and machine learning methods[J]. *Spectrochim*

- mica Acta Part A: Molecular and Biomolecular Spectroscopy, 2023, 299: 122806.
- [33] L.F. Chen, X.K. Tian, Y. Li, et al. Rapid and sensitive screening of multiple polycyclic aromatic hydrocarbons by a reusable fluorescent sensor array[J]. Journal of Hazardous Materials, 2022, 424: 127694.
- [34] E. Ghohestani, J. Tashkhourian, H. Sharifi, et al. A poly(arylene ethynylene)-based microfluidic fluorescence sensor array for discrimination of polycyclic aromatic hydrocarbons[J]. Analyst, 2022, 147(19): 4266-4274.
- [35] S.Q. Wang, H.Y. Jiang, Y. Jin, et al. Trace-level discrimination and detection of polycyclic aromatic hydrocarbons via hybrid photonic-plasmonic SERS sensors integrated with machine learning algorithms[J]. Journal of Hazardous Materials, 2025, 498: 140013.
- [36] 葛琨, 何致峰, 林佳娜, 等. 表面增强拉曼光谱结合主成分分析快速筛查食品接触材料中多环芳烃[J]. 分析测试学报, 2021, 40(11): 1588-1595.
- [37] 刘思思, 张波, 李星颖, 等. 机器学习在环境分析检测中的应用研究进展[J]. 分析测试学报, 2024, 43(8): 1105-1116.
- [38] 王书涛, 刘娜, 程琪, 等. 三维荧光光谱结合GA-SVM对多环芳烃的分类鉴别[J]. 光谱学与光谱分析, 2020, 40(4): 1149-1155.
- [39] 王书涛, 吴兴, 朱文浩, 等. 平行因子结合支持向量机对多环芳烃的荧光检测[J]. 光学学报, 2019, 39(5): 397-405.
- [40] 王立国, 赵亮, 刘丹凤. SVM在高光谱图像处理中的应用综述[J]. 哈尔滨工程大学学报, 2018, 39(6): 973-983.
- [41] 吕红燕, 冯倩. 随机森林算法研究综述[J]. 河北省科学院学报, 2019, 36(3): 37-41.
- [42] 向宇, 李茂刚, 闫春华, 等. 近红外光谱结合Light GBM的含油污泥多环芳烃含量快速定量分析方法研究[J]. 分析测试学报, 2025, 44(8): 1602-1611.
- [43] M.G. Li, Y.Y. Xu, J. Men, et al. Hybrid variable selection strategy coupled with random forest(RF) for quantitative analysis of methanol in methanol-gasoline via Raman spectroscopy[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 251: 119430.
- [44] M. Guo, M. Li, H. Fu, et al. Quantitative analysis of polycyclic aromatic hydrocarbons (PAHs) in water by surface-enhanced Raman spectroscopy (SERS) combined with random forest[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2023, 287(Pt 1): 122057.
- [45] R.K. Douglas, S. Nawar, M.C. Alamar, et al. Rapid detection of alkanes and polycyclic aromatic hydrocarbons in oil-contaminated soil with visible near-infrared spectroscopy[J]. European Journal of Soil Science, 2019, 70(1): 140-150.
- [46] R.L. Zhang, M.J. Guo, M.G. Li, et al. Surface-enhanced Raman spectroscopy combined with chemometrics for quantitative analysis and carcinogenic risk estimation of polycyclic aromatic hydrocarbons in water with complex matrix[J]. Chemometrics and Intelligent Laboratory Systems, 2025, 257: 105293.
- [47] 张笑寒, 孟祥添, 唐海涛, 等. 优化光谱输入量的土壤有机质随机森林预测模型[J]. 农业工程学报, 2023, 39(2): 90-99.
- [48] 宋述芳, 何入洋. 基于随机森林的重要性测度指标体系[J]. 国防科技大学学报, 2021, 43(2): 25-32.
- [49] 陶培峰, 王建华, 李志忠, 等. 基于高光谱的土壤养分含量反演模型研究[J]. 地质与资源, 2020, 29(1): 68-75, 84.
- [50] H. Li, M. Li, H. Tang, et al. Quantitative analysis of phenanthrene in soil by fluorescence spectroscopy coupled with the CARS-PLS model[J]. RSC Advances, 2023, 13(14): 9353-9360.
- [51] Y. Feng, T. Zhang, M. Zhao, et al. Raman-infrared spectral fusion combined with partial least squares (PLS) for quantitative analysis of polycyclic aromatic hydrocarbons in soil[J]. Analytical Methods, 2020, 12: 1203-1211.
- [52] 史晓凤. 应用表面增强拉曼光谱 (SERS) 分析探测水中多环芳烃 (PAHs) 的研究[D]. 青岛: 中国海洋大学, 2012.
- [53] X. Yan, X. Shi, J. Ma. Fast quantification of the mixture of polycyclic aromatic hydrocarbons using surface-enhanced Raman spectroscopy combined with PLS-GA-BP network[J]. Journal of Ocean University of China, 2021, 20: 1451-1458.
- [54] C.F. Ma, Q. Zhang, J. Liang, et al. Quantitative analysis of four PAHs in oily sludge by surface-enhanced Raman spectroscopy combined with partial least squares regression based on a novel nano-silver-silicon coupling substrate[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2024, 318: 124531.
- [55] S. Fornasaro, F. Alsamad, M. Baia, et al. Surface enhanced Raman spectroscopy for quantitative analysis: Results of a large-scale European multi-instrument interlaboratory study[J]. Analytical Chemistry, 2020, 92(5): 4053-4064.
- [56] 牛猛. 基于深度学习的多环芳烃时间序列分析[D]. 武汉: 中国地质大学, 2024.

- [57] S. Atta, J. Q. Li, T. Vo-Dinh, Multiplex SERS detection of polycyclic aromatic hydrocarbon (PAH) pollutants in water samples using gold nanostars and machine learning analysis[J]. *Analyst*, 2023, 148(20): 5105-5116.
- [58] M.Q. Qiu, S.G. Zheng, P. Li, et al. Detection of 1-OHPyr in human urine using SERS with injection under wet liquid-liquid self-assembled films of β -CD-coated gold nanoparticles and deep learning[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2023, 290: 122238.
- [59] J.X. Gu, Z.H. Wang, J. Kuen, et al. Recent advances in convolutional neural networks[J]. *Pattern Recognition*, 2018, 77: 354-377.
- [60] S.Y.S. S. Adade, H. Lin, X. Nunekeku, et al. Flexible paper-based AuNP sensor for rapid detection of diabenz(a,h)anthracene(DbA) and benzo(b)fluoranthene(BbF) in mussels coupled with deep learning algorithms[J]. *Food Control*, 2025, 168: 110966.
- [61] 徐桂烽. 基于LSTM混合模型的河流中多环芳烃含量预测研究[D]. 重庆: 重庆大学, 2022.
- [62] 汤英杰, 刘媛华. 基于预训练模型融合深层特征词向量的中文文本分类[J]. *上海理工大学学报*, 2023, 45(2): 189-197.
- [63] X.N. Yu, T.Y. Liu, L.L. Kong, et al. SpecRecFormer: deep learning-driven adaptive component identification of PAH mixtures based on single-component Raman spectra[J]. *Analytical Chemistry*, 2025, 97(18): 9876-9883.
- [64] C.Z. Meng, L. Trinh, N. Xu, et al. Interpretability and fairness evaluation of deep learning models on MIMIC-IV datasets[J]. *Scientific Reports*, 2022, 12: 7166.
- [65] Y.P. Tian, Z.Q. Geng, Z.D. Su, et al. Trace-level detection of free polycyclic aromatic hydrocarbons based on magnetic driving and deep learning-assisted recognition[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2026, 352: 127531.

Research Progress on Intelligent Algorithms for the Detection of Polycyclic Aromatic Hydrocarbons

Liu Yuanyuan^{1,2}

1. School of Information and Electronic Engineering, Shangqiu Institute of Technology, Shangqiu 476000
2. School of Physics and Electronic Engineering, Jiangsu Normal University, Xuzhou 221116

Abstract: Polycyclic aromatic hydrocarbons (PAHs) are environmentally persistent, bioaccumulative, and pose a potential carcinogenic risk. Therefore, the rapid and accurate detection of trace amounts of PAHs is of great significance for environmental monitoring and risk assessment. Addressing issues such as peak overlap, structural similarity, matrix interference, and signal fluctuations at low concentrations in the spectral detection of PAHs, this review summarizes research progress on intelligent algorithms for the classification and identification of PAHs, concentration prediction, and the analysis of mixed components. This paper reviews recent research advances in machine learning and deep learning for PAHs detection, analyzes the fundamental differences among various algorithms in terms of learning mechanisms, data dependency, spectral specificity, and cross-matrix generalization capabilities, and evaluates their impact on the accuracy, robustness, and field applicability of intelligent PAHs detection. To further analyze the key bottlenecks and research pathways in the intelligent detection of PAHs, providing guidance for algorithm selection and field applications. In the future, we will strengthen efforts in building standardized datasets, cross-domain transfer, and research on model applicability to improve the reliability, interpretability, and adaptability of intelligent PAHs detection in complex scenarios.

Keywords: Polycyclic aromatic hydrocarbons; Surface-enhanced Raman spectroscopy; Machine learning; Deep learning