DOI: 10.19812/j.cnki.jfsq11-5956/ts.20251230001
引用格式: 王金花, 张蓉, 冯鑫, 等. 人工智能在食品新污染物分析中的应用潜力研究进展[J]. 食品安全质量检测学报, 2026, 17(8): 3–11.
WANG JH, ZHANG R, FENG X, *et al*. Research progress on the application potential of artificial intelligence in the analysis of emerging contaminants in food [J]. Journal of Food Safety & Quality, 2026, 17(8): 3–11. (in Chinese with English abstract).

人工智能在食品新污染物分析中的应用潜力研究进展

王金花, 张　蓉, 冯　鑫, 别　玮, 李　祯, 周熙成, 张晓龙*

(中国海关科学技术研究中心, 北京　100026)

摘　要: 持久性有机污染物(persistent organic pollutants, POPs)、微塑料(microplastics, MPs)、内分泌干扰物、抗生素等目前全球关注的新污染物(emerging contaminants, ECs)已成为威胁食品安全、损害公众健康的重大风险因素, 而传统食品安全分析方法在应对未知、复杂的新污染物方面面临技术瓶颈。近年来, 机器学习(machine learning, ML)等人工智能(artificial intelligent, AI)技术由于数据处理能力强、模式识别性能优异, 因而为解决该难题提供了全新、有力的工具。本文阐述了 AI 技术在优化样品前处理流程、提高仪器分析效率、构建筛查数据库、挖掘确证信息等检测分析关键技术环节中的突出优势, 评估了其在食品中新污染物分析的应用潜力。尽管应用前景广阔, 该领域仍面临数据质量、模型可解释性及标准化等技术与制度和机制挑战。总体而言, AI 技术正在切实推动食品中新污染物的检测分析向智能化、可预测的方向发展。

关键词: 食品安全; 人工智能; 新污染物; 检测分析

Research progress on the application potential of artificial intelligence in the analysis of emerging contaminants in food

WANG Jin-Hua, ZHANG Rong, FENG Xin, BIE Wei, LI Zhen,
ZHOU Xi-Cheng, ZHANG Xiao-Long*

(*Science and Technology Research Center of China Customs*, *Bejing* 100026, *China*)

ABSTRACT: Persistent organic pollutants (POPs), microplastics (MPs), endocrine disruptors, antibiotics and other emerging contaminants (ECs) are currently major risk factors threatening food safety and harming public health. Traditional food safety analytical methods face technical bottlenecks in addressing unknown and complex new contaminants. In recent years, artificial intelligence (AI) technologies, such as machine learning (ML), have provided a novel and powerful tool to tackle this challenge due to their strong data processing capabilities and excellent pattern recognition performance. This article elaborated on the prominent advantages of AI technology in optimizing sample pre-treatment processes, improving instrumental analysis efficiency, constructing screening databases and extracting confirmatory information in key analytical processes, and assessed its potential applications in the analysis of

收稿日期: 2025-12-30
基金项目: 海科中心自立科研课题项目(2023HZ07)
第一作者: 王金花(1968—), 女, 博士, 研究员, 主要研究方向为食品质量与安全。E-mail: wangjh2008@aliyun.com
***通信作者:** 张晓龙(1972—), 男, 博士, 研究员, 主要研究方向为食品安全。E-mail: zhzhsh2001@126.com

emerging contaminants in food. Despite broad prospects, this field still faced challenges related to data quality, model interpretability, standardisation and institutional and procedural mechanisms. Overall, AI technology is effectively promoting the development of intelligent and predictive analysis of emerging contaminants in food.

KEY WORDS: food safety; artificial intelligence; emerging contaminants; detection and analysis

0 引 言

由于近年来全球化学品生产和使用量剧增，大量结构新颖、危害尚不明确、尚未纳入常规监测体系的化学物质进入环境，并通过食物链逐级富集于食品中，因此对食品安全及公众健康构成了严峻而长久的威胁。这些物质被统称为“新污染物”，其范畴广泛，包括持久性有机污染物(persistent organic pollutants, POPs)[如全氟及多氟烷基物质(per- and polyfluoroalkyl substances, PFAS)、多氯联苯(polychlorinated biphenyls, PCBs)]、内分泌干扰物(如双酚A、邻苯二甲酸酯)、微塑料(microplastics, MPs)、溴代阻燃剂、抗生素和抗生素耐药基因(antibiotics resistance genes, ARGs)及纳米材料等[1−6]。更值得关注的是，多数新污染物具有高毒性、难降解、易生物累积、长距离迁移等特性，故即使在极低浓度下也可能对人体免疫、内分泌、神经和生殖等系统造成不可逆的损害[7−13]。目前，国内外广泛关注的新污染物主要是国际公约管控的 POPs、MPs、内分泌干扰物、抗生素等[14−15]。这些污染物浓度普遍较低，检测难度远高于常规污染物，现有检测标准和方法尚不能支撑新污染物的筛查、风险评估及有效管控工作。

食品新污染物的分析技术正在从依赖经验的模式转向数据智能驱动的模式，由此引发不同技术路线在效率、准确性上呈现出显著差异。具体而言，在样品前处理优化方面，传统方法多采用正交实验设计(orthogonal experimental design, OED)或单因素优化(one-factor-at-a-time, OFAT)，难以充分、可靠地处理复杂无序的多维变量关系。相比之下，机器学习方法则具有更强的系统优化能力。例如，MA 等[16]的研究系统对比了传统 OED 与基于径向基函数神经网络(radial basis function neural network, RBFNN)的优化方案，得出明确结论：RBFNN 结合非支配排序遗传算法 II (non-dominated sorting genetic algorithm II, NSGA-II)不仅大幅缩短优化周期，还将活性成分的提取效率提高了 30.11%，证明了人工智能(artificial intelligent, AI)在优化复杂提取过程中的优势。在数据分析层面，算法的选择直接关系到所建模型的通用性与稳定性。对于结构化特征数据(如分子描述符)，基于树模型的算法[如 XGBoost、随机森林(random forest, RF)]已具有极好的鲁棒性。FENG 等[17]的研究证实，利用机器学习预测农药保留时间(retention time, RT)时，XGBoost 和 RF 模型对处理 log*P*、分子量等离散描述符的预测偏差(ΔRT)小于 1.0 min 的比例超过 90%。但毋庸讳言，基于规则的传统机器学习方法在处理非结构化的原始质谱图像或复杂光谱数据时，其特征提取能力尚无法与卷积神经网络(convolutional neural networks, CNN)等深度学习(deep learning, DL)模型相提并论。这一点在 GLOAGUEN 等[18]开发的 NeatMS 工具中得到了极好地体现，其采用 CNN 直接分析峰形，而 CHETNIK 等[19]提出的 MetaClean 分类器基于 AdaBoost 算法，二者均在降低假阳性率和处理高噪背景方面大大优于传统的 XCMS 或 MZmine 等峰检测算法。

虽然 AI 技术发展前景广阔，但现有研究也揭示了不同算法的局限性。KOVÁCS 等[20]明确论证了不同算法对输入特征集的敏感程度不同，故训练数据某些特定特征会导致 ΔRT，进而直接干扰食品安全监管措施的制定。因此，如何就不同目标(结构预测、光谱识别、风险评估等)设计既预测能力优良又可解释性强的智能分析体系，是 AI 技术应用于食品新污染物检测领域时的核心且迫切的挑战。

AI 本质上是通过模仿人类思考、认知、决策和行动的过程[21−22]，因此算法是 AI 技术的核心。而目前在种类繁多的 AI 算法中，机器学习(machine learning, ML)和 DL 是两大关键领域。二者的根本目标都是让计算机从数据中自主学习、归纳规律，进而自行完成推理、做出决策。具体而言，ML 算法是一种典型的基于数据的 AI 技术，其让计算机挖掘数据中潜藏的规律和模式，以此对未知数据进行预测与分类。DL 是 ML 的一个重要分支，它以深度神经网络为基础，通过构建具有多个隐藏层的模型，自动从海量数据中学习复杂的模式及高级特征表示[23]。ML 与 DL 的关系及区别见图 1。

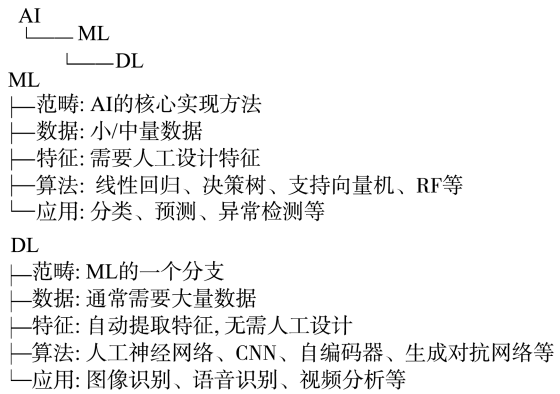


图 1 ML 与 DL 的关系与区别框架图

Fig.1 Framework diagram of the relationship and differences between ML and DL

近年来, 以高分辨质谱(high-resolution mass spectrometry, HRMS)为核心的非靶向/可疑筛查技术为未知污染物的发现带来极大机遇[24–26], 但同时也不可避免地产生了海量、多维、非结构化的复杂数据, 给数据解析带来了前所未有的挑战[27–28]。传统的人工解析或简单的数据库匹配方法效率极低, 常规方法分析质谱图时, 超过 90%的污染物及代谢物尚不能可靠确证[4]。正因如此, AI 特别是 ML 和 DL 技术, 凭借其处理高维数据、识别复杂模式和自主学习的能力[29–32], 正在促使食品中新污染物的检测分析模式发生革命性变化: AI 已不再是简单的辅助工具, 而是作为真正的数据引擎, 贯穿于样品处理、仪器分析、数据解析、风险评估的诸环节, 极大地提升了检测的通量、准确性、智能化水平和未知物发掘能力。本文旨在综述 AI 技术在食品中新污染物检测分析中, 从样品前处理、仪器分析到数据采集与处理等各关键环节的应用潜能, 探讨其在优化流程、构建数据库、挖掘确证信息等方面的突出优势, 同时客观分析当前面临的技术瓶颈、制度和机制难题以及发展前景, 以期为未来食品中新污染物智能防控体系的构建提供参考。

1　人工智能助力食品新污染物分析的潜能

1.1　优化前处理流程

食品样品基质十分复杂, 含有油脂、蛋白质、糖类等多种干扰物, 而新污染物常以痕量形式存在。因此提取效率是前处理过程中最关键、最直接影响后续检测结果可靠性的步骤。这一过程通常包括提取、净化和浓缩等环节, 目前分析食品中新污染物时, 技术人员通常依赖经验进行烦琐的试错实验来优化条件, 即采用 OED 或 OFAT 方法, 虽然二者逻辑清晰, 但在面对多变量叠加的食品基质时, 往往需要大量组合实验, 耗时耗力, 所得局部最优解未必能兼顾整体效率与成本。相比之下, 利用 AI 技术手段进行优化时, 构建非线性回归模型能够显著提升优化效率。通过优化提取参数、简化前处理步骤, AI 技术将为实现前处理流程的智能化、自动化和高效便捷开辟新的途径。

优化提取参数。MA 等[16]的研究提供了借鉴; 以中药复方“白芨五味子颗粒”为例, 把传统 OED 与 RBFNN 结合 NSGA-II 的 AI 优化方案在提取工艺上的表现作对比, 结果表明了两种技术路线在关键指标上的差异(表 1)。

结果显示, 与传统 OED 方法相比, AI 优化的方案使活性成分的提取效率最高提升了 30.11%。这不仅证实了 ML 能够有效预测和分析复杂的提取过程, 更证明了其在突破传统经验限制、挖掘潜在最优工艺参数方面的巨大潜力。对于食品新污染物检测而言, 这种高效率、低成本的 AI 优化路径思路, 能够帮助实验室技术人员更快建立复杂基质中痕量污染物高效提取的方案, 切实提升整体检测效能。

简化前处理步骤。目前食品中抗生素残留多采用液相色谱-质谱/质谱法(liquid chromatography-tandem mass spectrometry, LC-MS/MS)作为确证和定量方法, 如 GB 31656.24—2025《食品安全国家标准 水产品中喹诺酮类药物残留量的测定 液相色谱-串联质谱法》和 GB/T 21315—2007《动物源性食品中青霉素族抗生素残留量检测方法 液相色谱-质谱/质谱法》, 以及行业标准 SN/T 1988—2023《出口动物源性食品中头孢类抗生素残留量的测定 液相色谱-质谱/质谱法》等。以新发布的 GB 31656.24—2025 为例, 样品中抗生素残留经酸化乙腈提取, 正己烷脱脂后还需固相萃取柱净化, 净化包括活化小柱、上样、洗脱、氮吹和复溶等步骤, 前处理过程复杂。HUANG 等[33]利用一种改进的光谱装置短时间内获取多种抗生素特征光谱图像, 通过 ML 技术, 将数千幅特征光谱图像与浓度标签关联, 从而实现对多种抗生素浓度的准确预测, 简化了样品前处理过程。WANG 等[34]基于分子印迹聚合物(molecular imprinting technology, MIPs)构建了一种伏安电子舌传感器阵列。研究人员利用主成分分析(principal component analysis, PCA)等机器学习算法对这些复杂的响应模式进行处理和分析, 实现了对药品

表 1　AI 优化与传统方法在优化提取参数中的对比分析

Table 1　Comparison analysis of AI optimization and traditional methods in optimizing extraction parameters

参数	传统方法(正交实验/单因素)	AI 优化方法(RBFNN + NSGA-II)	AI 优势对比
耗时	长: 需逐级筛选或进行全因子实验, 实验周期长	短: 通过模型模拟替代大量实体实验, 快速确定最优区域	研发周期大幅缩短
成本	高: 依赖大量标准品和试剂消耗	低: 减少了无效实验次数, 降低了试剂与标准品消耗	经济性能更高
稳定性	中: 依赖人为经验判断, 结果易受主观因素影响	高: 算法自动寻优, 减少了人为干预, 结果重现性好	结果更客观稳定
适用范围	窄: 适用于线性关系明显的简单体系	宽: 擅长处理高度非线性的复杂多维变量关系	更适应复杂基质
优化效果	基于现有经验, 提升空间有限	提取效率提升 30.11%	挖掘出潜在最优解

和生物样品中环丙沙星、左氧氟沙星和莫西沙星等 3 种氟喹诺酮类抗生素的同时、高选择性检测，检出限(limit of detection, LOD)为 1 μmol/L。这种集成 MIPs 与 ML 的传感平台，能够在简化或省略液液萃取、固相萃取等烦琐前处理步骤的情况下，直接对复杂样品进行分析，大大提高了检测速度和便利性。

1.2 优化仪器分析参数

仪器分析，特别是色谱-质谱联用法，是新污染物检测的主流分析手段。然而，要获得最佳分离度、最高灵敏度和最稳定的信号，需要对流动相梯度、RT、离子源温度、碰撞能量等大量仪器参数进行细致调节。而目前此类参数高度依赖专家经验，不同实验室、不同处理批次之间存在差异。因此，AI 技术通过学习历史数据或实时反馈来自动优化参数，是分析过程智能化和标准化自然而然的选择。

RT 是色谱行为预测中一个十分重要的正交参数。FENG 等[17]系统地考察了使用 ML 预测液相色谱-高分辨质谱法 (liquid chromatography-high-resolution mass spectrometry, LC-HRMS)中农药及其转化产物的 RT。分别使用 321 种和 77 种农药作为训练集和验证集，比较了多种 ML 算法的性能。研究发现，基于 log*P*、分子量、拓扑极性表面积等大量化学描述符训练的 XGBoost、RF 和轻量级梯度提升机算法(light gradient boosting machine, lightGBM)模型在 RT 预测上表现出最佳的精度，验证集中至少 90%的农药 ΔRT<1.0 min，完全符合非靶向筛查(non-targeted screening, NTS)的应用需求。SILLÉ 等[35]进一步明确论证了，RT 感知注释对新污染物 NTS 的价值，既有利于提高暴露组学数据的可解释性和可重复性，又将 RT 预测、定量结构-保留关系(quantitative structure retention relationship, QSRR)建模和 AI 3 者自然整合到非靶向代谢组学流程中，同时实现定性筛查和定量测定。

在质谱参数优化方面，AI 的优势更加凸显。SINGH 等[36]在考察合成农用化学品对人类健康的影响时，设计了一个严谨、有层次的预测框架，该框架整合了 RF、lightGBM 和 CatBoost 等 ML 模型，并结合互信息增益和递归特征消除等特征选择方法，借助多级特征选择、混合集成学习、沙普利加和解释(shapley additive explanations, SHAP)和自定义损失函数(CustomLoss)逐层优化，提高预测精度，最终获得极好的预测性能。其中，lightGBM-粒子群算法(particle swarm optimization, PSO)+CustomLoss 的集成在预测的准确度和精确性上分别高达 98.87%和 98.59%。该研究的方法论逻辑与优化质谱数据采集和处理流程的目标高度契合，非常适用于最大化目标物信号强度和检出率。此外，MELNIKOV 等[37]开发出名为“peakonly”的算法，专门用于解决 LC-HRMS 数据中的峰检测难题，能够精确区分真实的化合物峰和低强度噪声峰，在检测真实阳性峰方面精度极高，从而直接提升了质谱数据的质量。该算法经过调整后也可用于气相色谱-质谱法(gas chromatography-mass spectrometry, GC-MS)峰分析，克服传统机器学习(非 DL)方法在 GC-MS 代谢组学数据优化中长期存在的局限性，实现更高精度的特征提取。

更先进的应用体现在多技术融合的智能分析系统。MU 等[38]将离子淌度谱(ion mobility spectrometry, IMS)-HRMS 与机器学习技术相结合进行 PFAS 的 NTS。由于 IMS-HRMS 能直接给出质荷比(*m*/*z*)和 RT 以及碰撞截面(collision cross section, CCS)等目标物分析的特征信息，故研究人员可用 ML 模型预测目标物的 RT 和 CCS，并将其作为额外的筛选维度，同时匹配实验测得的 *m*/*z*、RT、CCS，极大地提高了 PFAS 筛查的准确性和定性鉴定的可靠性，假阳性结果显著减少。在实际应用于废水样本时，该方法共鉴定出 56 种潜在 PFAS，其中 45 种是参考标准品之外的新发现，充分证明了采用 AI 技术对多维参数开展协同分析，在复杂基质中发现未知污染物具有极强的优势。

1.3 提升数据分析效能

由于基于 HRMS 分析的 NTS 是发现未知新污染物的有力工具[39–40]，但其核心难点在于从海量、复杂的 HRMS 数据中“大海捞针”般地找到真正有价值的信号。因此 AI 技术已然成为解决该难题的关键，其在数据预处理、特征提取和初步筛选等环节均有极好的应用。

在数据预处理与特征提取阶段，AI 算法能够自动、高效地完成传统上耗时费力的手动工作。GLOAGUEN 等[18]开发了名为 NeatMS 的 DL 工具，用于辅助大规模液相色谱-质谱法代谢组学分析中的峰校正。该工具利用 CNN 分析峰的形状、强度和周围环境，自动判断一个峰是真实的化合物信号，还是由噪音或基质干扰引起的假峰。相较于目前常用的传统数据预处理软件包 XCMS 和 MZmine，峰检测的假阳性率大大降低。更难得的是，NeatMS 软件已在 GitHub 上以开源形式公开。CHETNIK 等[19]则提出了 MetaClean，即一种基于 AdaBoost 算法的 ML 分类器，用于过滤 LC-MS/MS 非靶向代谢组学数据分析中假阳性峰。该分类器整合了 11 个先前发表的峰值质量评估指标，经过训练后，能在独立测试集上有效去除低质量的峰积分，其准确率可达 81%，具有良好的跨平台适用性。MELNIKOV 等[37]提出的“peakonly” DL 框架同样属于此类应用，它应用 DL 模型直接、准确地进行峰检测，克服了传统方法的种种不足。

在 NTS 的智能化发展方面，AI 使得“伪靶向筛查”(pseudo-targeted screening, PTS)成为可能。传统的非靶向分析因缺乏标准品和可靠的参考数据库，难以对未知物进行有效确证。而 PTS 策略结合 ML，则可在没有标准品的情况下进行有效筛选[3]。其基本思路是利用 ML 模型(如前文提及的 RT 预测模型)先预测一批“可疑”化合物在给定分析条件下的色谱和质谱行为(即 RT 及 MS/MS 碎片模式)，

然后在分析样品时自动搜寻预测位置上出现、质谱特征匹配的信号, 因此可极大缩小需要人工审核的候选未知物名单范围, 真正从“广撒网”转向“精准捕捞”。FENG 等[17]利用训练好的 RT 预测混合模型对 120 个草莓样品进行了筛查, 初步筛出 20 种农药及 12 种代谢物, 其中大部分得到了标准品的直接确证, 验证了该方法的有效性。

综合分析现有文献, 可就 AI 峰检测算法与传统方法在主要性能指标上进行系统性对比(表 2)。

上述对比表明, AI 驱动的峰检测方法在准确性和效率两方面都明显优于传统方法。例如, NeatMS 把质谱图当做图像处理, 用 CNN 自动提取峰形轮廓, 有效解决了传统方法在低丰度峰识别中的瓶颈问题; 而 MetaClean 则基于 AdaBoost 算法构建强分类器, 又以多个弱分类器的加权集成来提高在复杂基质中抗干扰的能力。

1.4　构建数据库

全面、准确的数据库是进行化合物鉴定的基础。然而, 对于新污染物而言, 现有公共数据库(如 PubChem, ChEMBL)中缺乏足够的高质量参考数据, 尤其是有关理化性质、质谱碎片规律、环境行为和毒性的数据。AI 技术不仅能够帮助更好地利用现有数据库, 更能主动构建、完善新的有实际应用价值的数据库。

CHENG 等[41]利用 ML 方法对经济合作与发展组织(Organisation for Economic Co-operation and Development, OECD)清单中所列 3486 种 PFAS 的生物活性进行了系统分类预测。先使用 GC、1-MNN 等不同 ML 模型分别预测了这些 PFAS 在 23 种不同生物靶标(如酶、受体)上的潜在活性, 再据此绘制出预测热图, 为后续的风险优先级排序和针对性测试提供了数据基础, 相当于构建了一个虚拟的、前瞻性的“PFAS 生物活性数据库”。KWON 等[42]利用新的无监督/半监督 ML 模型, 自动预测 PFAS 对多种人类生物靶点(酶、基因、蛋白质和细胞系)的生物活性, , 其中半监督学习模型被巧妙地用来预测 OECD 最新报告列表中 4730 种广泛应用于行业及消费产品中的 PFAS 的生物活性, 不仅开创性地开展结构−活性关系研究, 而且极大地扩展了预测范围。

AI 能很好地揭示化学空间中各要素的内在联系, 帮助理解庞杂的污染物家族。SU 等[43]开发的“PFAS-Atlas”平台基于深度无监督学习模型, 把数以万计的 PFAS 进行了亚组分类并予以可视化。通过聚类相似的化学结构, 清晰、直观地呈现不同 PFAS 类别之间的关系, 显著减少了“未分类”PFAS 的数量。更重要的是, 该平台能高效筛选 PFAS 的化学结构与其持久性、生物富集性、毒性数据之间的关系, 辅助决策者确定优先监测 PFAS 的类别, 合理分配监管资源。PARK 等[44]对 PFAS 污染源分类中所用的 RF、梯度增强决策树、支持向量机和人工神经网络等 ML 模型做了详实分析, 证实其准确率已达 96%以上; 而 XGBooste 则被证明可用于预测 PFAS 的浓度。所有这些 AI 方法都直接且充分地克服了传统 PFAS 检测和处理方法耗时、成本高、范围小的重大缺陷。

AI 技术促进了专用数据库的构建。XIE 等[45]为自动识别纳米塑料首先建立了纳米塑料的拉曼光谱数据库, 然后在此基础上构建 RF 模型, 很好地实现了水中纳米塑料的高精度识别(98.8%)。该研究说明 AI 可驱动建立新污染物筛查专用数据库。但也明确说明, 没有高质量、结构

表 2　AI 峰检测算法与传统方法性能对比

Table 2　Performance comparison of AI peak detection algorithm and traditional methods

指标	传统方法(XCMS/MZmine)	AI 方法(NeatMS/MetaClean)	说明与依据
假阳性率	较高(>50%)	显著降低(<10%)	XCMS 和 MZmine 依赖手动设定信噪比(*S*/*N*)阈值, 在低浓度区域易将噪声误判为峰; NeatMS 利用 CNN 学习真实峰形特征, 有效区分噪声, 将假阳性率控制在更低水平[18]
峰识别准确率/已知标准品查全率	较低(<50%)	NeatMS 提升至大于 90%; MetaClean 为 36%~81%	传统方法在峰边界划分和重叠峰拆分上存在明显不足; NeatMS 用质谱数据图像化处理配合 DL 做像素级分割, 显著提升识别精度[18]; MetaClean 用 AdaBoost 集成分类器优化特征选择, 进一步提高准确率[19]
处理速度	较慢(依赖串行计算与人工校正)	更快且可并行	XCMS/MZmine 通常需要多步流程(峰检测→对齐→去噪), 每步均需参数优化; 而 NeatMS 支持批量推理, 一旦模型训练完成, 单样本处理时间可缩短至秒级, 适合高通量筛查场景
自动化程度	较低(需大量人工干预与参数调优)	较高(端到端自动识别, 减少人为偏差)	如部署 AI 模型, 可实现“输入原始数据→输出纯净峰表”的全流程自动化, 减少对专家经验的依赖

化的数据，就不可能训练出高性能的模型。同样，MEYERS 等[46]研究探讨了基于尼罗红染色的半自动化微塑料检测与识别方法，公开了微塑料分析 RGB 模型代码和数据库。

1.5 挖掘确证信息

在初步筛选出可疑污染物后，接下来的关键一步是对其进行确证，即用充分证据确定其化学结构，最终明确究竟是哪种污染物，而此过程通常需要对精确质量数、同位素分布、MS/MS 碎片、RT 以及生物效应数据等信息的联合分析。因此，AI 技术在整合、挖掘确证信息时扮演着“侦探”的角色。

在质谱碎片解析方面，AI 可以很好地预测化合物的 MS/MS 裂解路径。BECK 等[47]讨论了 ML 方法在质谱学分支学科蛋白质组学中的应用，即用于肽段的测序和结构解析。可以预见，类似的 DL 模型(例如循环神经网络)将会被用来预测小分子污染物的碎片模式，为实验测得的 MS/MS 图谱提供可靠的理论参照，从而加速未知物的结构推断。

在多维信息融合与置信度评估方面，AI 已成为实现精准确证最有力、最成熟的工具。如前所述，把预测的 RT 和 CCS 与实验值匹配，即可构建多层次、有逻辑；连接的证据链，因此“正交参数”的应用能切实提高注释的置信度。具体而言，NTS 中一个化合物只有在 m/z、RT 和 MS/MS 各个层面都得到良好匹配，才能被赋予较高的确证置信度等级。基于 QSRR 并由 AI 增强的液相色谱和气相色谱 RT 预测是目前未被充分挖掘的正交参数。故将 RT 预测、QSRR 建模及 AI 3 者整合到非靶向代谢组学流程中，可实现精准确证[35]。

ARANGO-ARGOTY 等[48]提出的“Deep ARG”方法，利用 DL 从宏基因组数据中预测 ARGs，相比传统的“最佳匹配”方法能显著降低假阴性率，并识别更多样化的 ARGs，更加全面有效地监控抗生素耐药性，该方法本身就可用作食品中抗生素残留的间接“确证”方法。

此外，AI 对污染物检测新技术的发展具有十分明确、有力的促进作用。BIANCO 等[49]把 ML 与全息成像结合，实现了对水中微塑料的无标记识别。TIAN 等[50]对表面增强拉曼光谱(surface-enhanced Raman spectroscopy，SERS)技术在检测奶酪和酸奶等乳品时所遇基质效应做了充分分析，即该技术虽快速、灵敏，但易受基质干扰，因此他们提出了将 SERS 与 ML 相结合来校正信号偏差、消除背景干扰、提高信噪比，由此切实提升 SERS 在乳品样品中检测抗生素残留(兽药残留)的准确性。目前越来越多的先进的机器学习方法，诸如支持向量机、人工神经网络及 ML 模型，已被系统引入到 SERS 数据分析工作流程中，用以实现实时解析。这些算法显著提高了污染物检测的准确性，使乳制品生产商能够立即做出决策，主动、快速应对潜在的质量安全问题[51]。

2 人工智能在食品新污染物分析应用中的挑战

2.1 技术瓶颈

尽管 AI 在食品新污染物检测中展现出巨大潜力，但在实际应用上还处于起步阶段，其广泛应用面临着一系列挑战。当前最主要的技术瓶颈集中在数据质量和数量上。ML 的成功应用一定要以高质量、大规模且标注良好的数据集为前提[52]，而目前新污染物相关的实验数据普遍存在以下问题：(1)数据稀缺且分布不均，许多新污染物缺乏系统的实验室分析和监测数据；(2)数据异质性高，不同实验室、不同采样方法、不同分析平台产生的数据难以直接比较和整合，导致“数据孤岛”[53]现象；(3)数据标注成本高，给质谱峰或质谱图正确赋予化学身份需要昂贵的标准品，并高度依赖经验丰富的专家判断。ML 模型很容易出现泛化能力差、过拟合风险高的问题[41]；(4)模型的可解释性。DL 模型常被形象地称为“黑箱”，其决策过程难以解析。而这是食品安全监管中极其重要的障碍，因为监管决策必然要求有明确、可靠的科学依据及透明的逻辑。故而，虽然 SHAP、局部可解释性模型诊断解释(local interpretable model-agnostic explanations，LIME)等可解释性 AI 方法已被提出[6,54−55]，但如何将这些方法融入模型开发流程，做到既准确又可解释，仍是巨大挑战。例如，SHAP 方法在 PFAS 污染物检测及去除模型中已表现出极强的解释力，它既验证了“pH”“电荷积”等传统经验参数的科学性，又定量厘清了其对去除效率的具体贡献，而 LIME 方法则作为补充，确保了模型局部预测的透明度[51]。但是，SHAP、LIME 方法的有效性归根结底取决于底层数据的质量。在 PFAS 治理领域，由于缺乏对中间产物和复杂基质干扰的足够数据，目前的可解释性分析尚不能充分涵盖整个环境系统的复杂性，即期解释结果仅适用于所分析的数据片段，而不适用于所有环境场景。并且二者只能解释模型从数据中学到的规律，不能直接识别或弥补数据本身的缺陷。(5)算法的泛化能力及鲁棒性。不同算法对输入特征集敏感度差异极大，故若没有合适的设计，模型的通用性和稳定性都难以保证。KOVÁCS 等[20]在预测化学反应时发现，不同机器学习算法基于训练数据中不同特征(如不同试剂组合或底物结构)所得的 ΔRT 不同。这表明算法并非绝对客观，其稳定性与通用性本质上由输入特征的代表性和算法对这些特征的处理方式决定，若特征选择不当，模型就会产生难以识别的系统性错误。

2.2 制度和机制障碍

除了技术上的问题，还存在制度和机制层面的障碍。(1)缺乏统一的应用标准和规范。目前，从样品采集、前处

理、仪器分析到数据处理和结果报告, 各环节都缺乏统一的 AI 应用标准及操作规程。(2)数据共享机制缺失。科研机构和检测机构之间存在数据壁垒, 大家都不愿主动分享原始数据及所开发的模型, 而食品安全监测数据本身又具有极强的敏感性, 导致数据难以互通。建议突破传统的数据集中存储模式, 有计划地采用基于“联邦学习”的跨机构协同训练机制。各节点(实验室)在本地训练模型, 仅上传加密的梯度参数进行聚合, 在保护原始数据隐私的前提下实现高泛化能力 AI 模型的联合优化, 打破数据壁垒。(3)跨学科合作薄弱, 人才断层明显。AI 技术在食品安全中的应用是典型的交叉学科问题, 需要化学、生物学、计算机科学等学科领域的专家及政策制定者通力合作。但现实中存在“懂化学的不懂代码, 懂算法的不懂机制”的突出矛盾。建议建立联合研发机制, 让算法工程师深入实验室, 把色谱保留规律、质谱裂解机理等试验知识直接、合理地转化为算法的约束条件, 避免单纯据驱动的盲目拟合。更重要的是, 应开发低代码的专用分析平台, 将复杂的机器学习流程设计成为可视化的“积木模块”, 降低一线检测人员使用 AI 工具的门槛。(4)监管框架及标准滞后。现行食品安全法规和标准体系主要适用于传统检测方法, 难以适应 AI 技术的飞速发展。建议监管部门引入“算法审计”与“沙盒测试”机制, 一方面制定 AI 模型的验证标准, 要求所有模型必须通过对抗样本测试和回溯验证; 另一方面强制要求关键筛查模型具备可解释性输出能力(如 SHAP 分析), 确保 AI 判定的每一个“新污染物”都有化学特征依据, 为监管执法提供透明、可靠、可信赖的技术支撑。如何将基于 AI 的创新检测技术恰当地纳入监管体系, 建立与此匹配的验证、认证程序, 是监管部门当前面临的全新而重要的课题。

3　应用前景

AI 在食品新污染物分析中的应用未来或将聚焦在以下 3 个方面。(1)全流程智能化。陈晓红等[6]提出了“新污染物大模型”框架, 未来的趋势是建设覆盖 “筛选(动态筛查与智能识别)-评估(数据驱动风险评估模型)-控制(智能化监测、技术优选、动态调控)”全链条的智能系统, 系统中将融合化学知识图谱、物理化学性质、环境暴露、毒理学数据、文献知识等多源异构信息, 具备语义理解、、逻辑推理的能力, 能真正动态、可靠识别高风险污染物, 给出个性化治理方案并做到闭环管理。(2)无代码/低代码 AI 工具的普及。开发无代码/低代码 ML 工具有利于提高 AI 技术的可及性[2], 让不具备编程背景的食品安全检测人员也能轻松使用先进的 AI 方法, 加速推广和落地。(3)多源数据融合。已有大量研究[52–53]清楚论证了将光谱、色谱、高光谱成像、电子鼻等多种分析技术的数据予以融合, 并利用 AI 进行综合分析, 可以获得比单一传感器更完整、更准确的检测结果。此外, 边缘计算与物联网的结合正在切实推进实时在线监测的智能化发展, 实现从实验室的线下分析到现场快速筛查的实质性跨越。

4　结束语

ML 等 AI 技术正以前所未有的深度和广度变革食品中新污染物的检测分析模式。研究表明, AI 不仅具备优化提取参数、提高前处理效率的潜能, 更能通过预测 RT、优化质谱参数、精确提取特征峰来提升新污染物仪器分析的效能。在大数据时代, AI 已然成为新污染物 NTS 的核心技术手段, 很好地解决了海量数据解析的难题, 也实现了从“已知筛查”到“未知发掘”的重大跃迁。同时, AI 在构建新污染物预测性数据库、揭示新污染物化学空间规律、整合多维分析参数进行高准确性识别及确证方面都展现出强大潜力。然而, AI 技术在食品中新污染物分析领域的应用尚受制于技术、制度和机制诸多障碍。要让 AI 真正发挥其分析食品中新污染物的巨大潜能, 就必须凝聚多方力量, 共同致力于建立高质量的新污染物共享数据生态、开发透明可靠的智能算法、制定统一可行的行业标准规范, 最终建成完整成熟的食品中新污染物智能化分析及风险防控体系。

参考文献

[1] 陈玲, 杨潇, 张琳钰, 等. 食品中环境新污染物危害管控研究[J]. 中国工程科学, 2022, 24(6): 99–106.
CHEN L, YANG X, ZHANG LY, *et al.* Hazard and management of emerging environmental pollutants in food of China [J]. Strategic Study of Chinese Academy of Engineering, 2022, 24(6): 99–106.

[2] 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.

[3] 王亚韡, 张秋瑞, 于南洋, 等. 新污染物[J]. 化学进展, 2024, 36(11): 1607–1784.
WANG YW, ZHANG QR, YU NY, *et al.* Emerging pollutants [J]. Progress in Chemistry, 2024, 36(11): 1607–1784.

[4] 穆洪新, 张后虎, 陈玲, 等. 机器学习技术在新污染物可疑和非靶向筛查分析中的应用[J]. 生态与农村环境学报, 2025, 41(9): 1180–1187.
MU HX, ZHANG HH, CHEN L, *et al.* Application of machine learning techniques in non-targeted screening analysis of emerging contaminants [J]. Journal of Ecology and Rural Environment, 2025, 41(9): 1180–1187.

[5] 王晴, 刘烁, 吕俊岗. 机器学习在新污染物研究中的应用进展[J]. 生态环境损害研究, 2025, 1(2): 1–14.
WANG Q, LIU S, LV JG. Research progress on the application of machine learning in new pollutants studies [J]. Research on Eco-Environmental Damage, 2025, 1(2): 1–14.

[6] 陈晓红, 刘晓亮, 袁依格, 等. 基于人工智能技术的新污染物治理策略与路径研究[J]. 中国工程科学, 2025, 27(3): 152–163.
CHEN XH, LIU XL, YUAN YG, *et al.* Strategies and pathways for

emerging pollutant governance based on artificial intelligence technology [J]. Strategic Study of Chinese Academy of Engineering, 2025, 27(3): 152–163.

[7] LAI TT, KUNTZ D, WILSON AK. Molecular screening and toxicity estimation of 260000 perfluoroalkyl and polyfluoroalkyl substances (PFASs) through machine learning [J]. Journal of Chemical Information and Modeling, 2022, 62(19): 4569–4578.

[8] BONATO M, CORRÀ F, BELLIO M, *et al.* PFAS environmental pollution and antioxidant responses: An overview of the impact on human field [J]. International Journal of Environmental Research and Public Health, 2020, 17(21): 8020.

[9] AHN C, JEUNG EB. Endocrine-disrupting chemicals and disease endpoints [J]. International Journal of Molecular Sciences, 2023, 24(6): 5342.

[10] ZHAO B, REHATI P, YANG Z, *et al.* The potential toxicity of microplastics on human health [J]. Science of the Total Environment, 2024, 912: 168946.

[11] FEITEIRO J, MARIANA M, CAIRRÃO E. Health toxicity effects of brominated flame retardants: From environmental to human exposure [J]. Environmental Pollution, 2021, 285: 117475.

[12] QU T, SUN Q, TAN B, *et al.* Integration of network toxicology and transcriptomics reveals the novel neurotoxic mechanisms of 2,2',4,4'-tetrabromodiphenyl ether [J]. Journal of Hazardous Materials, 2025, 486: 136999.

[13] AMARASIRI M, SANO D, SUZUKI S. Understanding human health risks caused by antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARG) in water environments, current knowledge and questions to be answered [J]. Critical Reviews in Environmental Science and Technology, 2020, 50: 2016–2059.

[14] 国务院. 新污染物治理行动方案[EB/OL]. (2022-05-24) [2025-12-21]. https: //www.gov.cn/zhengce/content/2022-05/24/content_5692059.htm
General Office of the State Council. Action plan for the control of new pollutants [EB/OL]. (2022-05-24) [2025-12-21]. https://www.gov.cn/zhengce/content/2022-05/24/content_5692059.htm

[15] 庞国芳, 常巧英, 邵兵. 食品安全监管科技及发展方向分析[J]. 质量安全与检验检测, 2023, 33(1): 1–6.
PANG GF, CHANG QY, SHAO B. Analysis of regulatory technology and major research directions for food safety [J]. Quality Safety Inspection and Testing, 2023, 33(1): 1–6.

[16] MA JM, YAO JL, REN XY, *et al.* Machine learning-assisted data-driven optimization and understanding of the multiple stage process for extraction of polysaccharides and secondary metabolites from natural products [J]. Green Chemistry, 2023, 25: 3057–3068.

[17] FENG C, XU Q, QIU XL, *et al.* Evaluation and application of machine learning-based retention time prediction for suspect screening of pesticides and pesticide transformation products in LC-HRMS [J]. Chemosphere, 2021, 271: 129447.

[18] GLOAGUEN Y, KIRWAN JA, BEULE D. Deep learning-assisted peak curation for large-scale LC-MS metabolomics [J]. Analytical Chemistry, 2022, 94(12): 4930–4937.

[19] CHETNIK K, PETRICK L, PANDEY G. MetaClean: A machine learning-based classifier for reduced false positive peak detection in untargeted LC-MS metabolomics data [J]. Metabolomics, 2020, 16(11): 117.

[20] KOVÁCS DP, MCCORKINDALE W, LEE AA. Quantitative interpretation explains machine learning models for chemical reaction prediction and uncovers bias [J]. Nature Communications, 2021, 12: 1695.

[21] 孙名楷. 人工智能技术在现代食品检测中的应用与展望[J]. 现代食品, 2024, 30(22): 159–161.
SUN MK. Application and prospect of artificial intelligence technology in modern food detection [J]. Modern Food, 2024, 30(22): 159–161.

[22] 左敏, 纪慧卓, 苏礼君, 等. 人工智能在食品安全中的最新应用及进展[J]. 中国食品学报, 2024, 24(10): 1–13.
ZUO M, JI HZ, SU LJ, *et al.* Advancements and modern applications of artificial intelligence in ensuring food safety [J]. Journal of Chinese Institute of Food Science and Technology, 2024, 24(10): 1–13.

[23] 王霞, 祁兵, 王敏, 等. 人工智能算法在食品质量与安全中的应用研究进展[J]. 食品安全质量检测学报, 2025, 16(13): 180–188.
WANG X, QI B, WANG M, *et al.* Research progress on the application of artificial intelligence algorithms in food quality and safety [J]. Journal of Food Safety & Quality, 2025, 16(13): 180–188.

[24] SZABO D, FALCONER TM, FISHER CM, *et al.* Online and offline prioritization of chemicals of interest in suspect screening and non-targeted screening with high-resolution mass spectrometry [J]. Analytical Chemistry, 2024, 96(9): 3707–3716.

[25] CASEY JS, JACKSON SR, RYAN J, *et al.* The use of gas chromatography-high resolution mass spectrometry for suspect screening and non-targeted analysis of per- and polyfluoroalkyl substances [J]. Journal of Chromatography A, 2023, 1693: 463884.

[26] SAPOZHNIKOVA Y, NUÑEZ A. Non-targeted analysis with liquid chromatography-high resolution mass spectrometry for the identification of food packaging migrants [J]. Journal of Chromatography A, 2022, 1676: 463215.

[27] MANZ KE, FEERICK A, BRAUN JM, *et al.* Non-targeted analysis (NTA) and suspect screening analysis (SSA): A review of examining the chemical exposome [J]. Journal of Exposure Science and Environmental Epidemiology, 2023, 33(4): 524–536.

[28] PAN H, ZHAO G, TAN Y, *et al.* Liquid chromatography coupled to high-resolution mass spectrometry for high-throughput pesticide residue detection: Methodological advances and insights for herbal medicine analysis [J]. Critical Reviews in Analytical Chemistry, 2025, 1: 1–15.

[29] XIONG Y, LIU J, YU J, *et al.* OPEs-ID: A software for non-targeted screening of organophosphate esters based on liquid chromatography-high-resolution mass spectrometry [J]. Journal of Hazardous Materials, 2024, 465: 133275.

[30] WANG R, LU M, AN S, *et al.* 3D-MSNet: A point cloud-based deep learning model for untargeted feature detection and quantification in profile LC-HRMS data [J]. Bioinformatics, 2023, 39(5): 195.

[31] SHENG Y, XUE Y, WANG J, *et al.* Fast screening and identification of illegal adulteration in dietary supplements and herbal medicines using molecular networking with deep-learning-based similarity algorithms [J]. Analytical And Bioanalytical Chemistry, 2023, 415(16): 3285–3293.

[32] ARTURI K, HOLLENDER J. Machine learning-based hazard-driven prioritization of features in nontarget screening of environmental high-resolution mass spectrometry data [J]. Environmental Science & Technology, 2023, 57(46): 18067–18079.

[33] HUANG Y, CHEN J, DUAN Q, *et al.* A fast antibiotic detection method for simplified pretreatment through spectra-based machine learning [J]. Frontiers of Environmental Science & Engineering, 2022, 16: 38.

[34] WANG M, CETÓ X, DEL-VALLE M. A sensor array based on molecularly imprinted polymers and machine learning for the analysis of fluoroquinolone antibiotics [J]. ACS Sensors, 2022, 7(11): 3318–3325.

[35] SILLÉ FCM, PRASSE C, LUECHTEFELD T, *et al.* AI redefines mass spectrometry chemicals identification: retention time prediction in metabolomics and for a human exposome project [J]. Front Public Health. 2025, 13: 1687056.

[36] SINGH S, KAUR P, KAUR I, *et al.* A predictive framework using advanced machine learning approaches for measuring and analyzing the impact of synthetic agrochemicals on human health [J]. Scientific Reports, 2025, 15: 15544.

[37] MELNIKOV AD, TSENTALOVICH YP, YANSHOLE VV. Deep learning for the precise peak detection in high-resolution LC-MS Data [J]. Analytical Chemistry, 2020, 92(1): 588–592.

[38] MU HX, YANG ZC, CHEN L, *et al.* Suspect and nontarget screening of per-and polyfluoroalkyl substances based on ion mobility mass spectrometry and machine learning techniques [J]. Journal of Hazardous Materials, 2024, 461: 132669.

[39] 王振聚, 罗忻, 梁生康, 等. 高分辨质谱技术在全氟多氟化合物分析中的应用研究进展[J]. 分析测试学报, 2022, 41(6): 931–940.
WANG ZJ, LUO X, LIANG SK, *et al.* Research progress on application of high-resolution massspectrometry in analysis of PFAS [J]. Journal of Instrumental Analysis, 2022, 41(6): 931–940.

[40] DUDÁŠOVÁS, BERGER U, SEIWERT B, *et al.* Retrospective identification of novel and legacy per- and polyfluoroalkyl substances in German archived fish livers using a combined high-resolution mass spectrometry approach [J]. Environmental Science & Technology, 2025, 59(25): 12865–12877.

[41] CHENG WX, NG CA. Using machine learning to classify bioactivity for 3486 per- and polyfluoroalkyl substances (PFASs) from the OECD list [J]. Environmental Science & Technology, 2019, 53(23): 13970–13980.

[42] KWON H, ALI ZA, WONG BM. Harnessing semi-supervised machine learning to automatically predict bioactivities of per- and polyfluoroalkyl substances (PFASs) [J]. Environmental Science & Technology Letters. 2023, 10: 1017–1022.

[43] SU A, CHENG YY, ZHANG CW, *et al.* An artificial intelligence platform for automated PFAS subgroup classification: A discovery tool for PFAS screening [J]. Science of the Total Environment, 2024, 921: 171229.

[44] PARK J, BAIK JH, ADJEI-NIMOH S, *et al.* Advancements in artificial intelligence-based technologies for PFAS detection, monitoring, and management [J]. Science of the Total Environment, 2025, 980: 179536.

[45] XIE L, LUO S, LIU Y, *et al.* Automatic identification of individual nanoplastics by Raman spectroscopy based on machine learning [J]. Environmental Science & Technology, 2023, 57(46): 18203–18214.

[46] MEYERS N, CATARINO AI, DECLERCQ AM, *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: 153441–153452.

[47] BECK AG, MUHOBERAC M, RANDOLPH CE, *et al.* Recent developments in machine learning for mass spectrometry [J]. ACS Measurement Science Au, 2024, 4: 233–246.

[48] ARANGO-ARGOTY G, GARNER E, PRUDEN A, *et al.* Deep ARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data [J]. Microbiome, 2018, 6: 23.

[49] BIANCO V, MEMMOLO P, CARCAGNÌ P, *et al.* Microplastic identification via holographic imaging and machine learning [J]. Advanced Intelligent Systems, 2020, 2(2): 1900153.

[50] TIAN JQ, TANGLY, LU G, *et al.* Exploring the role of surface-enhanced Raman spectroscopy (SERS) in ensuring dairy quality and safety: Recent progress and future prospects [J]. Critical Reviews in Food Science and Nutrition, Critical Reviews in Food Science and Nutrition, 2025,1: 1–26.

[51] QI Y, HU D, JIANG Y, *et al.* Recent progresses in machine learning assisted Raman spectroscopy [J]. Advanced Optical Materials, 2023, 11: 2203104.

[52] LIU X, LU D, ZHANG A, *et al.* Data-driven machine learning in environmental pollution: Gains and problems [J]. Environmental Science & Technology, 2022, 56(4): 2124–2133.

[53] 段森, 唐葵, 哈兰. 供应链协同下制造业数据孤岛问题治理研究[J]. 信息技术与标准化, 2025(11): 37–41.
DUAN M, TANG K, HA L. Research on the governance of data island issues under supply chain collaboration in manufacturing [J]. Information Technology & Standardization, 2025(11): 37–41.

[54] 白明霞, 宋潮潮, 张圆正, 等. 机器学习赋能的水中全氟/多氟烷基物质去除技术[J]. 中国科学: 化学, 2025, 55(11): 3141–3154.
BAI MX, SONG CC, ZHANG YZ, *et al.* Machine learning-enabled technologies for the removal of per- andpolyfluoroalkyl substances from water [J]. Science China (Chemistry), 2025, 55(11): 3141–3154.

[55] ADADE SSYS, LIN H, JOHNSON NAN, *et al.* Advanced food contaminant detection through multi-source data fusion: Strategies, applications, and future perspectives [J]. Trends in Food Science & Technology, 2025, 156: 104851.

(责任编辑: 蔡世佳 安香玉)