

Machine Learning-Assisted Surface-Enhanced Raman Spectroscopy Detection for Environmental Applications: A Review

Sonali Srivastava, Wei Wang, Wei Zhou, Ming Jin, and Peter J. Vikesland*

Cite This: https://doi.org/10.1021/acs.est.4c06737



ACCESS	III Metrics & More	Article Recommendations	s Supporting Information	

ABSTRACT: Surface-enhanced Raman spectroscopy (SERS) has gained significant attention for its ability to detect environmental contaminants with high sensitivity and specificity. The cost-effectiveness and potential portability of the technique further enhance its appeal for widespread application. However, challenges such as the management of voluminous quantities of high-dimensional data, its capacity to detect low-concentration targets in the presence of environmental interferents, and the navigation of the complex relationships arising from overlapping spectral peaks have emerged. In response, there is a growing trend toward the use of machine learning (ML) approaches that encompass multivariate tools for effective SERS data analysis. This comprehensive review delves into the detailed steps needed to be considered when applying ML techniques for SERS analysis. Additionally, we explored a range of environmental applications where different ML tools were integrated with SERS for the detection



of pathogens and (in)organic pollutants in environmental samples. We sought to comprehend the intricate considerations and benefits associated with ML in these contexts. Additionally, the review explores the future potential of synergizing SERS with ML for real-world applications.

KEYWORDS: Surface-Enhanced Raman Spectroscopy, Machine Learning, Environmental Pollutants

1. INTRODUCTION

Comprehensive environmental monitoring, spanning wastewater, drinking water, surface water, and air quality, has the potential for early detection of contaminants and pollutants. Increased integration of public health and environmental monitoring is essential for safeguarding health, preserving ecosystems, and upholding regulatory standards. Environmental engineers have diligently worked to detect contaminants, including pathogenic bacteria and viruses, microplastics, per- and polyfluorinated substances (PFAS), and heavy metals within environmental matrices.^{2,3} Detection and monitoring of these contaminants, whether in water, wastewater, or air, is crucial and aids in understanding contaminant fate and transport while offering substantial benefits, including community-level surveillance and the potential for effective epidemic management.^{4,5} To achieve rapid monitoring in environment matrices, there is a need for techniques that offer high sensitivity, cost-effectiveness, and field adaptability. Conventional gold standard methods for pathogen detection, such as enzyme-linked immunosorbent assay (ELISA) and polymerase chain reaction (PCR), or for (in)organic pollutants, such as chromatography-based and mass spectrometry (MS)-based techniques, have long been staples of environmental analysis.⁶⁻⁸ While these methods are recognized for their accuracy and sensitivity, they often come with high costs and are challenging to adapt to resource-limited

settings.⁹ Additionally, diverse types of contaminants necessitate different detection techniques and often require specialized personnel and resource-intensive methods, thus limiting their universal application.

Surface-enhanced Raman spectroscopy (SERS) has emerged as an alternative approach with the capacity to detect a wide range of analytes from biological pathogens to emerging contaminants such as PFAS, and microplastics.^{10,11} SERS enhances the intrinsic Raman scattering of molecules that are associated with noble-metal nanoparticles or nanostructures.¹² This enhancement facilitates the detection and identification of trace quantities of analytes at concentrations that can approach the single-molecule level.¹³ With rapid advances in nanofabrication and synthesis, SERS has gained increasing attention for environmental analysis.^{14–16} SERS stands out as a valuable tool for advancing environmental monitoring due to its capacity to offer cost-effective, highly sensitive, and adaptable approaches that are suitable under challenging environmental conditions.^{17–19}

Received: July 3, 2024 Revised: October 21, 2024 Accepted: October 22, 2024

In the second se

However, as the technology has evolved, SERS analysis has exhibited certain challenges, such as managing vast volumes of spectral data, addressing overlapping Raman peaks, and dealing with spectral artifacts.^{20–22} The acquisition speed and the information density contained with SERS spectra are notable. SERS measurements can rapidly accumulate substantial amounts of spectral data (ranging from megabytes to gigabytes) from a single sample within minutes and this can challenge data handling and limit comprehensive analysis. Traditionally, researchers have primarily focused on individual high-intensity SERS peaks, thereby constraining the comprehensive and impartial interpretation of the data.²³ Moreover, the complexity of environmental matrices, replete with various impurities and molecules that may share similar or closely overlapped SERS peaks, presents challenges in distinguishing and interpreting the data.¹⁹ To tackle these complexities and make full use of the information obtained through SERS, researchers are turning to data analysis and machine learning (ML). While there have been several excellent review papers on the application of ML in SERS, $^{24-27}$ there remains a significant gap in the literature specifically addressing the fusion of these techniques for environmental pollutant analysis. Despite substantial advancements in SERS and ML, integrating these fields as standardized tools for environmental monitoring requires further exploration and review.

To bridge this gap, we review recent studies that used SERS in combination with ML (SERS-ML) methods, including multivariate analysis within environmental settings. We begin by providing an overview of SERS and ML and outlining detailed steps to consider when employing ML for SERS analysis. We aim to offer a comprehensive and comparative analysis of different ML techniques applied to enhance the use of SERS for the detection of pathogens-bacteria, viruses, or their nucleic acids; organic pollutants-polyaromatic hydrocarbons (PAHs), organophosphorus pesticides (OPPs); and inorganic pollutants-anions and heavy metals and an emerging pollutant: microplastics. Finally, we outline prospective advancements for using SERS with ML in environmental settings. We discuss future scope, limitations, and key considerations for deploying SERS-based sensors in realworld applications.

2. INTRODUCTION TO SERS

SERS is a powerful analytical technique for chemical and biochemical detection, with the potential for single-molecule sensitivity.²⁸ SERS enhances the intrinsic Raman signal of target molecules by several orders of magnitude using metallic nanoparticles or nanostructures. The mechanisms responsible for SERS include electromagnetic field amplification and charge transfer processes and a detailed discussion can be found elsewhere.^{29,30} The combined effect of these enhancement mechanisms underscores SERS' remarkable capability to detect molecules with ultrahigh sensitivity.^{31,32} SERS provides distinct molecular vibration fingerprinting patterns associated with molecular constituents, chemical bonds, and macromolecular configurations.¹² As a molecule-specific approach for analyte detection, SERS enables the identification of diverse targets without requiring extensive sample pretreatment.

SERS approaches for molecule detection can be classified as label-free or labeled.²⁸ In label-free SERS, the target molecule directly interacts with the SERS-active substrate and results in production of unique Raman spectra. Premasiri et al. demonstrated the advantages of SERS over bulk Raman in

differentiating Gram-positive Bacillus strains and Gramnegative bacteria such as Escherichia coli and Salmonella typhimurium.³³ Using a gold nanoparticle-covered SiO₂ substrate, they achieved a $\sim 10^4$ enhancement factor, enabling species and strain distinction at the single-cell level. Additionally, SERS effectively minimizes fluorescence interference, a major issue in bulk Raman spectra, particularly for biological samples. The fluorescence quenching effect occurs due to the energy transfer from the excited chromophores to the metal surface that shortens the fluorescence lifetime, allowing the Raman signal to dominate and resulting in cleaner, more accurate spectral data.^{34,35} Numerous studies have explored label-free SERS detection of molecules, encompassing a broad range from biological entities to emerging pollutants such as PFAS, microplastics, illicit drugs, and more.^{36–38} However, the major limitation is multiple peaks overlapping in SERS spectra, which complicates the identification of specific molecules. Table S1 summarizes the peak assignments for various environmental contaminants. Since multiple groups often share overlapping Raman bands, distinguishing them for multianalyte identification can be challenging. Additionally, SERS enhancement mechanisms are highly distance- and orientation-dependent, with only molecular components within ~ 10 nm of the substrate being preferentially enhanced.^{32,39} This spatial sensitivity leads to variability in SERS spectra across different substrates, further complicating standardization of the technique.

Although label-free detection has several advantages, it is also limited in its ability to identify molecules that exhibit weak or negligible Raman signals, such as viruses and heavy metals. In these instances, SERS tags have been developed by attaching intrinsically strong Raman scattering molecules (also known as Raman reporters) to the surfaces of nanoparticles, creating a distinct SERS spectrum of the Raman reporter.⁴⁰ The incorporation of biorecognition elements, such as antibodies or aptamers, into the SERS tags enables them to bind specifically to targeted molecules. This is called labeled SERS and is particularly effective for biomolecular identification and quantification, including viruses and their antigens, pathogenic bacteria, and biomolecules.^{41–43}

3. DECODING MACHINE LEARNING: UNDERSTANDING THE ESSENTIALS

ML is a subset of artificial intelligence that concentrates on developing algorithms and statistical models.⁴⁴ These models empower computers to learn and enhance their performance based on data or experiences, thus enabling them to make predictions, classifications, and decisions.⁴⁵ Within the domain of ML, we work with both training and test data sets. The training data set is instrumental in instructing the model, as algorithms generate a model that minimizes errors and provides an optimal fit.⁴⁴ Subsequently, the model undergoes testing with test data to evaluate its accuracy. In this context, data sets can be classified into supervised and unsupervised learning. In supervised learning, algorithms learn to make predictions based on prelabeled data.46 These training data sets consist of input features along with their corresponding output features. The primary goal is for the algorithm to discern robust relationships between the input features and the target outputs, thereby enabling accurate predictions on new, unseen data. Prominent examples of supervised learning algorithms are regression, decision trees, and neural networks. In contrast,

Table 1. An C Their Applica	Dverview witl tion to SERS	n Comparative Analysis of ML Tools—Including Unsupervised Learni Measurements	ıg, Supervised Learning, And Deep Learı	ning Methods with Examples for
Model	Category	Description	Examples of SERS with ML	Remarks
Principal Com-	Unsupervised	Multivariate analysis Dimension reduction method: Used to reduce the dimension of large data sets while preserving	Sensing the illicit drug fentanyl in wastewater at 0.8 ppb	Valuable technique for dimensionality re-
ponent Analy- sis (PCA) ⁴⁸		as much data variance as possible. Principal components (PC) are new variables constructed as a linear combination of the initial variables	concentration. Multiplex SERS detection of PAHs, ⁴⁹ microplastics, ³⁰ metricides ^{51–53} and nohrchlorinoted thionols ⁵⁴	duction, noise reduction, and data visual- ization. Assumes that the underlying data relation- chine are linear
		PCs explain the maximum amount of variance in the data.	Differentiation of pathogens: bacteria, ^{55–58} viruses. ⁵⁹	Sensitive to outliers. Mostly works for continuous numerical data.
Partial Least Square Regres- sion (PLSR) ⁶⁰	Supervised	Deals with data sets with multiple independent variables and one or more dependent variables.	Predicting concentration of the bacteriophage Phi6 using <i>Pseudomonas syringae</i> , ⁶¹ viral loads of Hepatitis B serum. ⁶²	Effective when there is collinearity (high correlation) among the independent variables.
				Also, useful when independent variables are large compared to the number of obser- vations.
		Performs dimensionality reduction by creating a new latent component that captures maximum covariance.		
K-means cluster- ing ⁶³	Unsupervised	Primarily used for pattern recognition and clustering tasks.	Identification of methicillin-resistant and methicillin- sensitive bacteria; ⁵⁸ cancer cells; ⁶⁴ cancer bio-	Easy to implement and computationally efficient, suitable for large data sets.
)		Operates on the principle that similar data points tend to be close to each other in a feature space.	markers. ⁶⁵	Assumes that clusters have equal variance.
		Each data point is assigned to the majority class among its K nearest neighbors based on a chosen distance metric (e.g., Euclidean distance).		Sensitive to outliers.
		Sensitive to the choice of K and the distance metric.		Requires specifying the number of clusters (K) in advance.
Hierarchical Clustering ⁶⁶	Unsupervised	Aims to group data points into clusters based on their similarity.	Differentiation of <i>Bacillus</i> strains, often used with PCA. ^{55,56}	Provides a visual representation of cluster- ing
)		Unlike other clustering methods, hierarchical clustering organizes data in a hierarchical tree-like structure called a dendrogram.		No need for predefined clusters.
		Can be either agglomerative (bottom-up), which starts with individual data points as clusters and merges them iteratively, or can be divisive (top-down), which starts with all data points in one cluster and splits them recursively.		Computationally intensive.
		Machine Learning (ML) model	ţ	
Random Forest (RF) ⁶⁷	Supervised	Ensemble ML technique that combines multiple decision trees to make robust predictions.	Multiplex detection of antibiotic resistance genes; 68 detection of $PAHs^{69}$ in water	Robust to overfitting. Can handle linear and nonlinear data.
				For large data sets, RFs can be computa- tionally intensive and may require tuning of hyperparameters.
		RF identifies the most influential feature by measuring its importance for prediction.		
Support Vector	Supervised	Can be used for both classification and regression. Works by finding an optimal hyperplane that best separates different classes in the data while	Detection of PAHs. ⁷¹	Highly effective for both linear and non-
Machine (SVM) ⁷⁰	-	maximizing the margin between them. A kernel function transforms input data into a higher- dimensional space, allowing SVM to find linear or nonlinear decision boundaries.	Diagnosis of viral infection to bacterial growth; ⁷² detections of different stains of <i>Escherichia. coli</i> ; ⁷³ detections of SARS-CoV- 2^{74}	linear classification. Regularization and margin maximization prevent overfitting However, the choice of kernel and tunino narameters is crucial
		Particularly effective in high-dimensional spaces and complex data distributions.		0
		Ability to handle nonlinear relationships through kernel functions, thus making it versatile.		
Logistic Regres- sion (Logit model) ⁷⁵	Supervised	Used for binary classification problems.	Classification of miRNAs; ^{76,77} SARS-COV-2 variants. ⁷⁸	Straightforward and easy-to-implement al- gorithm. Provides interpretable results.

Model	Category	Description	Examples of SERS with ML	Remarks
		Models the relationship between a binary outcome (0 or 1) and one or more independent variables by estimating probabilities.	Detection of microplastics. ⁷⁹	Assumes a linear relationship between the predictor variable and the log odds of the outcome.
		Uses the logistic function to transform a linear combination of input features into a probability score.		
		Deep learning algorithms		
Artificial Neural Network	Supervised	Class of ML models inspired by the human brain's neural structure.	Monitoring cellular drug responses. ⁸¹	ANNs can approximate complex functions, making them versatile for diverse tasks.
(ANNs) ³⁰		Consists of layers of interconnected nodes, organized into input, hidden, and outer layers.	Industrial wastewater source tracing. ⁸²	Interpretability can be challenging due to the complex, nonlinear nature of ANNs.
		Key features of ANNs include their ability to learn complex, nonlinear relationships from data through the training process.	Single-molecule detection using SERS. ⁸³	Need large volumes of labeled data to perform well.
		Training involves adjusting the weights and biases of connections between neurons to minimize prediction errors.		
		Often requires substantial amounts of data and computational resources for training.		
Convolutional Neural Net-	Supervised	Use convolutional layers to detect patterns in the input data. Convolution involves applying filters to small overlapping regions of the input.	Multiplex detection of PAHs, ⁸⁵ cancer detection, ⁸⁶ Salmonella serotype identification. ⁸⁷	Automatically learns local features and spatial hierarchies.
work (CNN)		CNNs typically have one or more fully connected layers.		Requires large amounts of labeled training data.
		CNNs use weight sharing to reduce the number of parameters.		Computationally intensive.
Multilayer Per-	Supervised	MLP is a feedforward ANN.	Determination of Mycobacterium tuberculosis infection	Not all ANNs are MLP. The distinction lies
ceptron (MLP) ⁸⁸		Information flows through the network in a forward direction.	and drug resistance.»	in the specific architecture and connec- tivity patterns within the network.
Residual Net- work (Re- sNet) ⁹⁰	Supervised	The core idea is to use residual blocks consisting of a shortcut connection (identity mapping) that bypasses one or more layers.	Identification of methicillin-resistant and -susceptible Staphylococcus aureus, ⁹¹ multiplex detection of organophosphorus pesticides in environmental water. ⁹²	Introduced to address the challenge of training very deep neural networks. Effectively maintain previous gradient in- formation by shortcut connection.

Table 1. continued

pubs.acs.org/est

Critical Review



Figure 1. Step-by-step guide for harnessing SERS-ML analysis model from SERS data collection to preprocessing and further ML model development and performance analysis.

unsupervised learning represents an alternative approach within ML.⁴⁷ Here, algorithms are designed to uncover patterns within unlabeled data. The training data comprises input features and lacks corresponding target outputs. The algorithm's main goal is to unveil inherent patterns and relationships within the data, which can be leveraged to cluster similar data points together. An extensive overview of multivariate analysis techniques, ML, and deep learning methods, along with examples of their application in SERS is provided in Table 1.

3.1. Building and Evaluating ML Models in SERS Analysis. Following the collection of SERS data, the process of constructing an ML model encompasses three major steps: data preprocessing, model development, and performance analysis. Figure 1 outlines a step-by-step process to develop a ML model for SERS data. Data preprocessing and cleaning is a fundamental step for SERS-ML models. Raw SERS spectra require cosmic ray removal, graph smoothening, and background subtraction of fluorescence or electromagnetic interference, substrate impurities, or electronic noise.93 Preprocessing also includes handling outliers and addressing inconsistencies or errors in the data set. Normalization or standardization is crucial for SERS data to eliminate variability and ensure accurate comparison. Additionally, feature selection or dimensionality reduction using PCA or latent vectors improves model efficiency and interpretability.

After preprocessing the data, the next step in supervised ML model development involves dividing the data set into training, validation, and test sets. Hundreds to thousands of SERS spectra can be obtained from replicates and through large-area scans in SERS. These data can be used for training and testing. Following this step, an appropriate ML algorithm or model architecture is selected based on factors such as the nature of

the problem (e.g., classification, regression), the characteristics of the data (e.g., structured, unstructured), and the desired outcome (e.g., accuracy, interpretability). Although there is no definitive method for selecting a specific algorithm over others, exploring various approaches is appropriate for selecting the best fit. Using the selected algorithm, the model undergoes training on the training set, iteratively adjusting parameters to minimize a specified loss function, commonly achieved through gradient descent optimization.⁹⁴ Model performance is evaluated on the validation set using metrics such as accuracy, precision, recall, F1 score, or mean squared error to gauge generalization to unseen data. Hyperparameters are adjusted, or alternative models are explored to improve results. Final performance is tested on the test set for an unbiased realworld estimate, with iterative refinements based on validation and test outcomes.

Choosing a machine learning model for large, complex data sets involves more than performance metrics. For real-life applications, researchers should also consider algorithm efficiency, balancing accuracy with the time and space needed for training and deployment. It is also essential to evaluate the complexity of data relationships (whether linear or nonlinear) and the size of the data set to prevent overfitting or underfitting. This can be achieved through bias-variance analysis and by examining learning curves.⁹⁵ Effective hyperparameter tuning, using techniques such as GridSearchCV or RandomSearchCV, ensures optimal performance without excessive computation.⁹⁶ Finally, interpretability is crucial for understanding how models make decisions, particularly when detecting environmental contaminants. Tools for feature importance analysis such as Partial Dependence Plots, and SHAP (SHapely Additive exPlanations) values can help clarify



Figure 2. Machine Learning tools for the classification of bacteria using SERS. **A.** Schematic of the Experimental Setup for Label-Free SERS Detection of Bacterial Strains Using Gold Nanoparticles. **B.** Average SERS spectra for 19 bacterial strains, normalized using the peak at 1326 cm⁻¹. **C.** Confusion matrix for 19 bacterial strains. Diagonal entries indicate class accuracies (out of 400 spectra), while off-diagonal entries represent misclassifications for each strain. Adapted with permission from ref 73. Copyright 2022 American Chemical Society. **D.** Three-dimensional PCA plot using first three PCs for classification of five Gram-positive *Bacillus* spores (*B. atrophaeus, B. anthracis sterne, B. thuringiensis, B. anthracis sterne killed, and B. anthracis ames killed*) and *Pantoea agglomerans* based on their SERS spectra. Adapted with permission from ref 56. Copyright 2008 Optica Publishing Group.

each feature's contribution, ensuring greater transparency in the decision-making process.⁹⁷

3.2. Why ML for SERS Analysis? SERS generates abundant data, posing substantial challenges in terms of comprehension and analysis. Within the span of a few seconds to minutes, comprehensive SERS data sets can be collected for individual molecules. Additionally, multicapture SERS platforms excel at identifying the unique vibrational fingerprints associated with target molecules, offering detailed spectral information.⁹⁸ However, when employed for real environmental matrices, SERS encounters significant challenges. The

presence of contaminants in environmental samples is characterized by extremely low concentrations, thus posing challenges for detection through analytical methods alone. Another major issue arises from the diverse array of coexisting compounds in real samples, potentially masking the signals of target molecules and making it difficult to discern subtle changes through visual inspection. This necessitates a robust analytical tool capable of not only handling voluminous data but also conducting multiplex analysis with high sensitivity and identifying complex relationships. Another significant advancement in SERS-ML analysis lies in considering multiple peaks for multiplex data sets. Whether labeled or label-free, SERS requires the comparison of spectral intensities against target concentration for subsequent analysis and quantification. The conventional method for analyzing SERS data often focuses solely on a univariate assessment of a single vibrational mode peak intensity and examining its trend across varying target concentrations. In cases with multiple targets or reporter molecules, peaks corresponding to each molecule are selected. Although this approach is effective in simple conditions with minimal background interference, it may be insufficient when multiple targets are present, as variations in peak intensity may no longer be solely influenced by changes in concentration. Choosing multiple peaks for the same molecule allows for more robust quantification compared to relying on a single peak.^{92,99} The multiple peaks assignment requires multivariate analysis and paves the way for employing ML for more comprehensive analysis.

Conventional methods often focus solely on a univariate assessment of a single vibrational mode, examining intensity changes that correlate with alterations in concentration. While this approach demonstrates effectiveness in simple scenarios involving one or two analytes, its limitations become evident in complex systems. ML enables analysis using multiple or entire spectra, thereby providing a more accurate and robust analysis.¹⁰⁰ Furthermore, the integration of ML into a data analysis flow may aid in uncovering hidden trends. As such, ML is a promising method for extracting previously undisclosed patterns from acquired data, a task that is impossible manually. Additionally, ML can extract peak features crucial for developing models that effectively discriminate between contaminants, thereby expanding the analytical capabilities of SERS in environmental monitoring and beyond. Monitoring pathogens or contaminants in various environmental matrices, such as drinking water and wastewater, using SERS creates opportunities for prompt prevention and enables real-time decision-making for appropriate treatment. When detecting target analytes within environmental matrices, SERS intensities from the background or other molecules can interfere with target spectra, making it difficult to detect the target. The use of ML has shown potential to tackle all these problems, making it versatile and applicable to SERS under complex environmental conditions. Herein, we focus on the detection of biological contamination such as bacteria, viruses, emerging (in)organic contaminants, and microplastics using SERS in combination with ML methods.

4. SERS-ML IN THE DETECTION OF BIOLOGICAL CONTAMINANTS

This section delves into the application of SERS-ML for the detection and characterization of biological contaminants in water and wastewater, encompassing (1) bacteria, (2) viruses, and (3) their nucleic acids (DNA/RNA). Through the synergy of SERS's unique spectral fingerprints and the analytical power of ML algorithms, researchers have made significant progress in augmenting the speed, accuracy, and sensitivity of microbial detection, thereby advancing our understanding of environmental health and safety.

4.1. Bacteria. SERS has emerged as a rapid and reliable approach for bacteria detection that leverages its capability to provide unique spectral fingerprints attributed to the cellular components of bacteria, including nucleic acids, cellular walls, and membranes (Table S1). However, the similarity of the collected SERS spectra makes it difficult for effective

differentiation. For example, Rahman et al. reported SERS spectra for 19 different bacteria strains by mixing them with gold nanoparticles. (Figure 2A,B).⁷³ The prominent peaks associated with bacteria typically arise from the ring vibrations of adenine and guanine, occurring at 600–735 and 1325 cm⁻¹.⁷³ The variations in peak ratios, intensities, and shifts in wavenumbers are too subtle to discern manually. Hence, the complexities of SERS spectra necessitate the application of machine learning for thorough and accurate analysis.

By harnessing the power of ML models, we can unravel subtle nuances in SERS spectra, enabling us to not only distinguish between different bacteria but also quantify their presence more accurately. Over the years, SERS-ML has become popular in the identification of bacteria at the species and strain level. Classification ML models such as PCA, SVM, and HCA have been extensively used for the diagnosis and differentiation of biological samples. Several studies have used PCA for classifying closely related strains such as cereus group Bacillus strains,⁵⁵ anthracis sterne live or killed,⁵⁶ serotypes of salmonella,¹⁰¹ and many more. In a recent study, Arslan et al. discriminated Cryptosporidium parvum oocysts, E. coli, and *Staphylococcus aureus* using PCA and HCA.⁵⁸ By using medium Gaussian SVM model with 10-fold cross-validation, Rahman et al. were able to predict bacterial strains with 87.7% accuracy as shown in the confusion matrix (Figure 2C).⁷³ Guicheteau et al. classified Bacillus spores using three PCs, which accounted for greater than 90% of the variance in the SERS spectra as shown in Figure 2D.⁵⁶ Gram-negative bacterium Pantoea agglomerans showed good separation from Gram-positive Bacillus spores. However, among Gram-positive Bacillus spores, there is an overlap between the three strains of Anthracis as they share the same exosporium layer.⁵⁶ It is interesting to note from abovementioned studies that PCA was effective in the classification of Gram-positive and Gram-negative bacteria. The reason is the large variance in their SERS spectra because of different cell wall structures. Gram-negative bacteria such as E. coli possess a lipid-rich outer cell well and a thin peptidoglycan layer. However, Gram-positive bacteria such as S. aureus have thicker peptidoglycan cell walls than Gram-positive bacteria.⁵⁶ Parasites such as C. parvum oocysts have much different cell structures than bacteria⁵⁸ and contain high carbohydrate components within the wall structure. Here, PCA and HCA are effective in classification since the linear combination of spectra features differentiates the targets. However, while the differentiation of two bacteria from the same categories has very similar spectral features, PCA does not do a good job of classifying complex and nonlinear data. Most of the time, PCA is used for dimension reduction, and the PCs are then used as inputs for advanced machine learning techniques such as SVM.

Wang et al. used multiplexing to detect waterborne pathogens using labeled SERS with SVM.¹⁰² Three bioconjugated gold nanoparticles were used to specifically bind *E. coli* 0157:H7 cells at different epitopes. To differentiate between positive signals and negative signals, SVM was employed using the input of the first 58 PCs. A linear kernel was used in the SVM and showed clear differentiation between positive and negative spectra with a limit of detection of 10 CFU/mL. Another study classified three antibiotic resistance isolates of *E. coli* ATCC25922, *E. coli* ST131:075, and *E. coli* ST1193:025 following direct deposition on a gold nanoparticle substrate. Despite having very similar SERS spectra, the SVM-PCA-assisted model showed excellent performance in classifying these strains.¹⁰³



Figure 3. Detection of DNA using SERS **A.** Illustration of iodide-modified silver nanoparticles along with MgSO₄ to enhance DNA binding. **B.** SERS spectra of a series of oligonucleotides containing varying proportions of adenine (A) and cytosine (C), where A% = A/(A + C). All spectra are normalized to the peak intensity at 1087 cm⁻¹, corresponding to PO₂⁻. **C.** Plot of relative peak intensity ratio of 723 cm⁻¹ to 1087 cm⁻¹ as a function of A% in the oligonucleotides. Adapted with permission from ref 116. Copyright 2015 American Chemical Society. **D.** Schematic illustration of the core–satellite SERS sensor for detection of multiple eARGs (extracellular antibiotic resistance genes). **E.** SERS spectra of multiple eARGs in the range from 1 aM to 1 pM. F. Comparison of *sul1* concentration for actual vs predicted. Black dots represent the calibration set (80% of the data set) and red dots represent the test set, comprising the remaining 20%. Adapted with permission from ref 68. Copyright 2022 Elsevier.

SERS-ML promises to revolutionize the field of bacterial detection, offering a faster, more cost-effective, and more precise alternative to traditional methods, thereby advancing our ability to understand and combat microbial threats in diverse environmental samples.

4.2. Virus. Wastewater monitoring for virus detection has been done since the outbreak of poliovirus.¹⁰⁴ Viruses shed in the fluids of symptomatic or asymptomatic patients enter wastewater systems and may remain infectious for up to 30 days.¹⁰⁵ To prevent the outbreak of viral disease, it is crucial to detect viruses early and take preventive actions. Moreover, since late 2019, the coronavirus SARS-CoV-2 has spread quickly worldwide thus demanding rapid point-of-care testing (POCT) methods to facilitate early diagnosis outside the laboratory. Conventional methods such as ELISA and reverse transcription PCR (RT-PCR) analysis have been used to diagnose viruses, and typically have high sensitivity (for SARS-CoV-2 RT-PCR can achieve 500-1000 copies/mL of viral RNA).¹⁰⁶ Alternatively, without compromising high specificity and sensitivity, SERS has been used for detecting viruses at concentrations as low as 80 copies/mL within an hour.¹⁰⁷

Zhang et al. employed bromide-coated silver nanoparticles and Ca(II) ions as aggregates to enhance hotspot formation, ideal for detecting viruses of 100 nm diameter.⁵⁹ SARS-CoV-2, Human Adenovirus 3, and H1N1 Influenza virus exhibited similar SERS spectra and were more challenging to differentiate in complex media such as serum and saliva, the authors utilized PCA to facilitate effective discrimination. By using PCA, they successfully detected viruses at a concentration of 100 copies per test, within 2 min. While PCA captures maximum variance in the data, it does not consider class labels unlike supervised ML models, and may not be very effective in maximizing separation between different classes.⁴⁸ A similar study led by Garg et al. used a supervised PCA-LDA model to differentiate different enveloped viruses.¹⁰⁸ Their approach involved utilizing PCA for feature extraction and dimension reduction, followed by LDA to identify a subspace that optimally separates different classes within the data. They were able to differentiate SARS-CoV-2, Zika, and Influenza A viruses within an environmental dust background with 86% accuracy. While these models demonstrated satisfactory performance, their limitations include their capability to detect viruses at

very low concentrations (<100 copies) and the absence of cross-validation.

To detect extremely low concentrations of viruses, Yang et al. developed a hierarchical array of gold nanoneedles (GNAs) functionalized with angiotensin-converting enzyme 2 (ACE2).¹⁰⁷ ACE2, known for its high specificity to SARS-CoV-2, effectively captured and enriched the viruses near the GNA hotspot region. By incorporating PC-discriminant analysis of SERS signals, they achieved the detection of viral loads as low as 80 copies/mL in under 5 min. While the ACE2based method demonstrates high specificity for SARS-CoV-2 it exhibits limitations in detecting viral variants. A study led by Moitra et al. developed a set of DNA probes, specifically an antisense oligonucleotide (ASO) capable of interacting with genetic sequences of SARS-CoV-2 regardless of its mutations. Thiolated ASOs targeting the N gene of SARS-CoV-2 were attached to gold nanoparticles, allowing for the examination of viral RNA attachment to ASOs and producing strong SERS signals. With the combination of SERS with the PCA-SVM model, they were able to predict positive samples with 100% sensitivity and 90% specificity at concentrations up to 63 copies/mL of RNA. This demonstrates that utilizing the indigenous design of SERS probes alongside ML techniques enables real-time detection of viral variants without the need for sophisticated instrumentation.

An alternative method for virus or bacteriophage detection involves analyzing alterations in SERS signals emitted by bacteria or their metabolites postinfection. In a study led by Wang et al., they determined that dimethyl disulfide (DMDS), a volatile sulfide metabolite, accumulated in the headspace of a sealed Petri dish.⁷² Accumulation of DMDS was monitored using silver nanoparticle-based SERS tape affixed to the dish cover, and subsequent SERS spectra were obtained. Notably, during virus infection (Phi6) of P. syringae, changes in the SERS intensity of the DMDS peak were observed. To discern the statistical distinction between infected and noninfected bacteria, a PCA-SVM classification model was employed, utilizing a quadratic kernel and the first nine principal components (accounting for 95% variance). The classification confusion matrix revealed an overall accuracy of 93%, with sensitivity and specificity exceeding 92%. These findings suggest that instead of solely examining virus-specific SERS peaks, bacterial peaks or metabolites can serve as indicators of viral presence. Despite subtle changes, ML models offer improved classification capabilities.

4.3. Nucleic Acids. Extensive research has been conducted in the identification of nucleic acids to characterize microbial communities in drinking water, wastewater, and soil both as a means to understand the impact of environmental parameters on microbial communities¹⁰⁹ as well as to detect pathogenic DNA/RNA¹¹⁰ and antimicrobial resistance genes.^{111,112} Non-enzymatic methods, including DNA microarrays, nanopores, and mass spectrometry, have emerged as rapid and effective tools, enhancing the sensitivity and specificity of DNA sequence detection. However, the noninvasive detection of SERS surpasses these methods in terms of sensitivity, speed, and simplicity.

Nucleic acids exhibit distinctive fingerprint information that reflects their breathing and ring skeleton vibration modes and they are well-suited for label-free SERS detection.¹¹³ The Bell group^{114,115} has successfully detected DNA and RNA using silver nanoparticles and MgSO₄. The presence of MgSO₄ as the aggregation agent induces the formation of nanoparticle/

nucleotide aggregates, resulting in an increased number of hotspots. Consequently, such natural trapping of nucleic acids significantly enhances the signal strength. Xu et al. used iodidemodified silver nanoparticles and MgSO4 to neutralize the surface charge and enhance DNA binding as shown in Figure 3A.¹¹⁶ The authors demonstrated the effectiveness of using the phosphate backbone (PO_2^{-}) as an internal standard for singlebase DNA analysis through SERS. By synthesizing oligonucleotides with varying adenine (A) and cytosine (C) ratios, they normalized the SERS spectra based on the intensity of PO_2^- (Figure 3B). This revealed a clear, linear trend in the characteristic adenine band intensity against increasing concentration. Moreover, plotting the relative SERS intensity of the adenine peak (723 cm⁻¹) to the PO₂⁻ peak (1087 cm⁻¹) against the A/(A + C) ratio resulted in a remarkably linear relationship (Figure 3C). This method demonstrates high precision and sensitivity for single-base discrimination. However, label-free detection encounters a significant challenge stemming from the similarity of the collected SERS spectra that can be attributed to the common phosphate/sugar backbone shared among nucleic acids. Raman peaks derived from phosphate, notably those at 815/860, 1087, and 1230 cm⁻¹ represent the symmetric bend, symmetric stretch, and asymmetric stretch modes of PO₂⁻ and contribute to the overall spectral resemblance of numerous nucleic acids.¹¹⁶ Typically the only readily detectable difference lies in the relative intensities of various nucleobases. Consequently, distinguishing between nucleotides is challenging, especially at lower concentrations. One effective approach to overcome this challenge is to employ multivariate analysis. PCA in combination with DA has been used to classify different nucleic acids.¹¹⁷ Furthermore, achieving single-base sensitivity or distinguishing mismatches poses a challenge with SERS alone. Kang et al.⁹⁹ addressed this issue by employing treebased multiclass support vector machine (Tr-SVM) classifiers to differentiate SERS spectra of gene sequences with 2-10 base mismatches. A tree-based decision rule was utilized to group correlated classes, offering multiple classifiers based on one of two decision levels. SVM was applied to maximize the margin between different classes using an optimal hyperplane. Through 10-fold cross-validation, this adaptable discriminatory tool accurately identified antibiotic resistance genes with a prediction accuracy of 90%.

Alternative to label-free detection, an indirect method that consists of the hybridization of nucleic acids can also be used to detect DNA/RNA. In this approach, SERS probes containing a complementary strand and a Raman reporter hybridize with a target, increasing the Raman reporter's intensity.¹¹⁸ Research employing such a hybridization technique with a SERS probe can identify the presence of nucleic acids at concentrations in the picomolar¹¹⁹ to the femtomolar range.¹²⁰ However, concentrating solely on a change in the intensity of a single peak in the presence of a target gene may be affected by background interference, particularly in complex matrices. Lu et al. identified extracellular antibiotic resistance genes (eARGs) including sull, tetA, and bla_{TEM} through a hybridization approach using the Raman reporters 4-mercaptobenzoic acid (4-MBA), 2mercaptopyrimidine (4-MPY), and 4-nitro blue tetrazolium chloride (4-NBT) in environmental samples (Figure 3D).68 Figure 3E illustrates SERS spectra for eARGs across concentrations ranging from 1 aM to 1 pM. The sensitivity of this method was compromised by the complexity and signal



Figure 4. Detection of organophosphorus pesticides (OPPs) in water using a deep-learning-based multicapture SERS platform. **A.** Schematic diagram of PVP, MBA, and Cys stabilized Ag nanocubes at the air/water interface and **B**. their corresponding SERS spectra for each OPPs (MAP, DMT, GLA, EPN, PT, and Pho). **C.** PCA plot showing cluster separation with multicapturer Ag-PVP + Ag-MBA + Ag-Cys. **D.** A schematic illustrating the detection of OPPs in environmental water samples, combining SERS spectra collection with a developed deep learning model. **E.** Confusion matrix showing the regression results for OPP concentration predictions in environmental samples. Adapted from ref 92. Copyright 2022 American Chemical Society.

attenuation induced within environmental samples such as wastewater treatment plant effluent, aquaculture water from cattle farms, and groundwater. To reduce noise interference, unsupervised ML for multivariate analysis was used by selecting multiple characteristic peaks of Raman reporters from the total SERS spectra hence, increasing robustness compared to univariable analysis. Among SVM, RF, PLS, and multilayer perceptron (MLP), RF was most suitable with the lowest root-mean-square error (RMSE) and highest regression coefficient (R^2) shown in Figure 3F for *sul1*. The concentration of eARGs in environmental samples, determined through RF, was comparable to that of ddPCR with no statistically significant difference (*p-value* > 0.05).

5. SERS-ML FOR THE DETECTION OF ORGANIC POLLUTANTS

This section explores the application of SERS-ML for detecting and characterizing organic pollutants, with an exclusive focus on PAHs and OPPs. Both classes of compounds are toxic, carcinogenic, and persistent in various ecosystems, making their detection critical for environmental monitoring. Given the complexity of environmental matrices, detecting PAHs and OPPs at the trace level presents significant analytical challenges. By focusing on these pollutants, we aim to highlight the effectiveness of SERS coupled with ML in overcoming these challenges and provide precise classification and concentration of pollutants.

5.1. PAHs. PAHs are a hazardous class of chemicals whose structure consists of multiple fused benzene rings.¹²¹ PAHs have sparked significant environmental concern due to their carcinogenic and mutagenic characteristics, and their propensity to readily pollute essential natural resources such as drinking and river water.^{122,123} These compounds arise through the incomplete combustion of coals and fuels,^{124,125} resulting in their existence not as isolated chemicals, but as

complex mixtures. This complexity poses a challenge for the on-site identification of PAHs. The detection of PAHs using SERS poses several challenges. One of them is that the hydrophobic nature of PAHs inhibits their adsorption to citrate-stabilized Au or Ag colloids due to their incompatibility with the surface chemistry.¹²⁶ Attempts have been made to overcome this limitation by modifying the substrate surface with colloidal hydrophobic films.^{127,128} Surface modifications such as thiol-modified and oleate-modified Fe₃O₄@Ag microspheres resulted in a 10⁻⁸ mol/L limit of detection (LOD).¹²⁶ Another challenge lies in accurately quantifying trace concentrations of PAHs, as spectra often contain multiple analytes with overlapping peaks, varying signal-to-noise ratios, and significant background interference. Incorporating ML methods with SERS has shown the possibility of detecting PAHs as low as 5 nmol/L in PAH mixtures because of its ability to handle the nonlinear relationship between concentration and spectrum intensity.¹²⁹ Atta et al. employed two one-dimensional CNNs for the multiclass classification and regression analysis of SERS spectra.⁸⁵ The CNN models, optimized through training on a calculated data set, demonstrate high precision (97%), F1 score (94%), and accuracy (90%) in classifying pollutants. The CNN regression model effectively predicted pollutant concentrations, achieving a combined RMSE_{spectrum} of 5.92 \times 10 $^{-2}$ and RMSE_{conc} of 1.07 \times 10^{-1} (μ M). In complex scenarios, where spectral data from the same class varies over time and overlapping peaks (as seen here), more sophisticated algorithms such as CNNs are more effective than basic algorithms. The effectiveness arises from their ability to not only extract features, but also capture various patterns and possibilities, resulting in improved performance in complex analyses.²⁵ In addressing the challenge of overlapping spectra arising from structurally similar benzene structures in PAHs, a recent investigation led by Bajomo et al. introduced an innovative unsupervised machine learning

method named Characteristic Peak Extraction (CaPE) algorithm for dimension reduction to extract distinctive SERS peaks of PAH mixtures.¹³⁰ By analyzing SERS spectra from complex mixtures, particularly for varying concentration ratios, CaPE efficiently identified and extracted spectra of individual components, subsequently matching them against an SERS spectral library for identification. By integrating chemical sensing with the CaPE algorithm, they were able to effectively address challenges such as incomplete libraries and frequency shifts in SERS peaks. The SERS-ML tandem methodology exhibits significant potential for rapid, on-the-field identification and detection of chemicals based on molecular structures, outperforming conventional demixing algorithms.

5.2. OPPs. Another emerging contaminant class in environmental systems, primarily sourced from agricultural runoff, are OPPs such as methamidophos (MAP), dimethoate (DMT), parathion, diazinon, and others. The potential toxicity of OPPs poses serious health risks, including acute and chronic neuropathy, reproductive toxicity, and endocrinopathy, highlighting the need for vigilant environmental monitoring. Several studies have been done on the detection of OPPs using SERS -ML either using a label-free strategy¹³¹⁻¹³³ or by leveraging the interaction between OPPs and reporter molecules (e.g., 4-MBA, L-cysteine) attached to nanoparticles, leading to a change in SERS spectra. Li et al. conducted a study using plasmonic nanocube metasurfaces (NCMs) to detect various OPPs, including MAP, DMT, glufosinate ammonium (GLA), ethyl para-nitro-phenyl (EPN), parathion (PT), and phosmet (Pho), achieving multiplex determination (Figure 4).92 Potential affinity agents for OPPs consist of poly-(vinylpyrrolidone), 4-MBA, and L-cysteine assembled on Ag nanocubes self-assembled at liquid/liquid interface (Figure 4A). The combined SERS spectra of OPPs were reconstructed (Figure 4B), enhancing the spectral variations for each OPP. SERS spectral variances before and after the capture of OPPs on the modified NCMs were complex and could not be distinguished manually. Hence, PCA was used to extract the spectral variances (Figure 4C) and as an input for a ResNetdeep learning model. The model demonstrated a classification accuracy exceeding 96% and a regression accuracy surpassing 92%. Furthermore, the model successfully identified all six OPPs spiked in environmental water samples (farm, river, and fishpond water), highlighting the capability to exclude interference from other matrices in real environmental samples (Figure 4D). The regression confusion matrix for concentration predictions demonstrated an accuracy exceeding 92% (Figure 4E).

6. SERS-ML IN THE DETECTION OF INORGANIC POLLUTANTS

In this section, we explore the intersection of water quality monitoring and advanced analytical techniques, focusing on the detection of inorganic anions and cations.

6.1. Anions. Regular monitoring of water systems is essential due to the occurrence of problematic inorganic nitrogen and sulfur species, such as nitrates, nitrites, and sulfates. When these compounds are present in elevated concentrations, they can pose substantial risks to aquatic life and other organisms. For instance, nitrates and nitrites can induce eutrophication and disturb aquatic ecosystems, whereas sulfates may aid in the creation of sulfuric acid, leading to pH declines in water bodies and subsequent disruptions.¹³⁴

Anthropogenic sources of nitrates in water and wastewater primarily reflect agricultural discharge coming from pesticides resulting in eutrophication¹³⁵ and elevated concentrations of nitrite can cause blue baby syndrome in infants.¹³⁶ SERS being a less destructive and noninvasive technique can help with the in situ detection of these anions. The inherent negative charge of anions elicits charge-charge repulsion, however, limiting the application of SERS. To address this, cation-coated nanoparticles and SERS substrates have proven effective in mitigating the repulsive forces and enhancing SERS capabilities. Moiser-Boss et al. employed a cationic silver substrate to detect nitrates and sulfates via solid-phase extraction.¹³⁷ They successfully detected nitrates at concentrations below 100 ppm. Kuster utilized nanostructured gold substrates without surface functionalization for the in situ detection of nitrate.¹³⁸ PCA was employed to time-averaged SERS spectra over the range of 1000 cm^{-1} to 1350 cm^{-1} to distinguish varying concentrations of nitrates in water through SERS. This method proved effective in achieving improved differentiation, considering the spectral noise measurements.¹³⁹ Gajaraj and colleagues identified the presence of nitrate in water and wastewater at a low concentration of 1 mg/L.¹⁴⁰ However, when analyzing actual wastewater samples, the absorption of nitrates in the "hotspot" region was impeded due to the interference of other ions, namely chloride and phosphate. This interference resulted from the competitive interaction among these ions due to their identical charge and ionic radii.¹³⁷ To eliminate chloride adsorption, one method involves applying a coating to the SERS substrate with enhanced selectivity toward nitrates and sulfates.¹⁴¹

Until now, limited research has combined ML with SERS for inorganic pollutant detection. While SERS has successfully achieved low limits of detection with single peak observation for anions, addressing complex samples with diverse anions requires more than manual differentiation based on a single peak. Integrating ML holds promise for facilitating multiplex detection of ionic anions in water and wastewater within minutes, streamlining procedures, and significantly improving time efficiency.

6.2. Heavy Metal Cations. Expanded industrial activities have led to widespread heavy metal pollution in both water and soil.¹⁴² Accumulation of heavy metals in water and crops poses a significant threat since these materials are potentially toxic to living organisms.¹⁴³ The need to detect metal ions in water has driven advancements in analytical techniques. Direct detection of monatomic metal ions using SERS is challenging due to their small Raman scattering cross-section and the trace concentrations present in water.¹⁴⁴ To overcome this limitation, metal ions can be detected through chemical binding to organic receptor ligands that induce perturbations in SERS intensities upon the formation of a coordination complex.^{145–147} Chelating agents such as Schiff base,¹⁴⁸ cyanide,¹⁴⁷ terpyridine derivatives¹⁴⁶ have been used for the detection of Co(II),¹⁴⁶ Cd,¹⁴⁹ Hg(II),¹⁵⁰ As(III),¹⁵¹ and $Cu(II)^{145}$ with LODs (μ g/L) lower than the WHO defined limits. Docherty et al.¹⁴⁸ identified toxic metal ions including Ni(II), Cu(II), Mn(II), and Co(II) through SERS using [O, N, N, O] tetradentate bis-Schiff (salen) base ligand. The interaction between the salen ligand and various metal ions led to modifications in the intensity and frequency of several bands that was influenced by the size, mass, and coordination bond strength of the metal ions.¹⁵² Notably, alterations in the C=N stretch peak around 1600 cm⁻¹ of the salen ligand were



Figure 5. A. Schematic illustration of an AuNP/L-cysteine SERS substrate used for detecting As(V). **B.** SERS spectra of As(V) solutions at varying concentrations on the AuNP/L-cysteine substrate. **C.** PCA scatter plot showing the first two principal components for different concentrations of As(V). Adapted from ref 154. Copyright from 2023 American Chemical Society. **D.** Schematic diagram illustrating heavy metal detection using SERS spectra of key metabolites of *E. coli*. **E.** Averaged SERS spectra obtained from *E. coli* cultured in media with K₂Cr₂O₇ and NaAsO₂, varying in concentration. **F.** SVM model used to classify between Cr(VI) and As(III) for concentration range 0.68 pM to 0.68 μ M and 0.5 pM to 0.5 μ M respectively and tSNE clustering analysis for different concentrations of Cr(VI) and As(III) in red and blue, respectively. Adapted from ref 156. Copyright from 2023 Proceedings of the National Academy of Science.

observed introducing different metal ions. It is essential to highlight that although subtle changes are discernible, the identification of peak shifts and intensity changes in complex sample matrices is challenging. The application of PCA proved effective in distinguishing between different metal ions, where each metal ion formed distinct clusters, facilitating clear differentiation. However, it is worth noting that predicting low concentrations of metal ions in real samples posed difficulties due to interfering signals from organic matter.¹⁵³ To overcome the interference from the background signal, Fang et al.¹⁵⁴ integrated CNN with SERS for predicting the As(V) concentration in water using the chelating agent L-cysteine as shown in Figure 5A,B. PCA was applied to highlight the spectral differences, revealing clear clustering of the spectral data, with the first two PCs accounting for 99% of the variance (Figure 5C). However, significant overlap in clusters was observed at lower concentrations of As(V). When PCA-PLS was used for prediction, the low coefficient of determination $(R^2 = 0.884)$ and high RMSE of 18.34 indicated poor prediction accuracy, likely due to the complex variables in the SERS spectra. In contrast, combining CNN with SERS provided rapid and accurate predictions of As(V) ions down to 1 ppm, achieving an R^2 of 0.991 and a much lower RMSE of 3.36. The superior performance of CNNs compared to PCA-PLS may be attributed to the ability of CNNs to handle complex patterns and nonlinear relationships within highdimensional SERS data.⁸³ It is noteworthy that traditional methods like ICP-MS for detecting metal ions can achieve detection levels in the ng/L range,¹⁵⁵ making them superior to SERS in terms of sensitivity, given that SERS typically has a limit of detection around $\mu g/L$. Nevertheless, the SERS method provides a much cheaper and simpler alternative capable of detecting the metal ions at μ g/L levels and with ML rapid analysis within a few seconds.

Another compelling study to detect the presence of heavy metals in drinking and wastewater led by Wei et al.,¹⁵⁶ used bacterial metabolic response transduced by heavy metals into chemical (metabolite) signals using SERS with ML algorithms (Figure 5D). Changes in the SERS spectra of metabolites, specifically the nucleotides ATP, uracil, and adenine, were observed in E. coli cultures upon exposure to Cr(VI) and As(III) (Figure 5E). The integration of SERS with the SVM model demonstrated a detection limit of 6.8 pM for Cr(VI) and 0.5 pM for As(III), achieving sensitivities and specificities exceeding 97% (Figure 5F). Notably, this approach achieved a LOD 6 orders of magnitude lower than traditional growth inhibition methods relying on optical density. Moreover, the CNN model successfully detected the concentration of heavy metal ions in tap water and wastewater samples, exhibiting the same LOD as the SVM model. This demonstrates that the reliable performance of the SERS model, aided by machine learning, remained stable irrespective of the impurities present in the samples.

In essence, ensuring the quality of drinking water and discharged water necessitates portable, rapid, and highly sensitive detection techniques, a need well-addressed by the combination of SERS with ML.

7. SERS-ML IN THE DETECTION OF MICROPLASTICS

Microplastics, tiny plastic particles less than 5 mm in size, typically originating from the breakdown of larger plastic containing items have emerged as a new type of pollutant, and have been detected in air, tap water, river, and sea-water.^{37,157–159} The ubiquitous nature of microplastics has raised concern due to their pervasive presence and their potential ecological and human health risks. As these tiny particles infiltrate aquatic ecosystems, soils, and air, detection techniques must adapt to the complexity of plastic sources,

pubs.acs.org/est



Figure 6. Detection of microplastics using SERS-based logistic regression model. **A.** Schematic illustration of a 3D-PGNP system integrated with a syringe filter for microplastic filtration. **B.** SERS maps for PS at concentrations ranging from 0 to 250 μ g/mL, along with the corresponding digital SERS maps after applying ML model. Black pixels represent the absence (0) of PS, while blue pixels indicate its presence (1). **C.** Schematic illustration of microplastic classification using the logistic regression method, including data training, data input, and quantification. **D.** Quantitative digital counts and the sum of positive Raman intensities (PRI) alongside the corresponding calibration curve. Adapted from ref 79. Copyright from 2023 Wiley Online Library.

their varying sizes, and their concentrations. Several SERSbased direct and indirect-based microplastics detection methods have been published that were able to detect microplastics in concentration ranges from 0.97 ng/mL to 6.5 μ g/mL.^{160–162} The similarity in SERS profiles among different microplastics has led to the use of ML methods with the potential to significantly expedite experimental analysis and computation. Luo et al.¹⁶³ introduced an innovative approach combining CNN with SERS for the rapid identification and classification of microplastic mixtures. Six types of microplastics were dispersed in water, creating 6300 sets of SERS spectra for training and testing. The confusion matrix results of each MP of the trained CNN model show that polycarbonate, polystyrene, and polyethylene terephthalate were all identified correctly. The CNN model achieved a remarkable identification accuracy of 99.54% without extensive spectral preprocessing, demonstrating its robustness in handling unprocessed SERS spectra for rapid and accurate identification of complex MP mixtures.

Recently, Kim et al. developed 3D-plasmonic gold nanopocket (3D-PGNP) nanostructures integrated with a syringe filter to filter and detect polystyrene (PS) and polyethylene (PE) microplastics as shown in Figure 6A.⁷⁹ Using SERS spectra of microplastics along with background noise, a logistic regression model was trained to convert map data into a digital format by classifying pixels as MP-positive or MP-negative (Figure 6B,C). By using digital counts, they were able to detect PS at a concentration of 2.5 μ g/mL. When quantifying positive pixels in the digital data, a logarithmic trend is observed between the digital count and the concentration of PS (Figure 6D). Additionally, they applied the developed model to complex environmental samples by spiking microplastics into tap water, river water, and seawater, successfully detecting low concentrations of PS with a high recovery rate, even in the presence of interferences.

8. FUTURE SCOPE: TRANSLATING SERS-ML POTENTIAL INTO REAL-WORLD SOLUTIONS

The integration of SERS with ML holds exceptional promise to revolutionize various fields by offering rapid, sensitive, and robust analytical methods. In environmental monitoring, SERS-ML platforms integrated with sensor networks have the potential to transform pollution monitoring with source tracing, thus enabling targeted mitigation and remediation strategies. Similarly, in biothreat detection, healthcare biosensing, and food safety, SERS-ML technologies offer significant potential. Despite such potential, the widespread adoption of SERS-based sensors as a standard recognition tool faces several challenges, necessitating concerted efforts to address these limitations and unlock future opportunities.

One of the primary hurdles is the lack of standardized protocols for SERS-ML applications. SERS spectra can be influenced by various factors such as substrate morphology, experimental conditions, and complex backgrounds, leading to challenges in interpretation and reproducibility. To address this issue, the development of standardized data sets and protocols, along with incorporation of internal standards or the establishment of calibration curves, is crucial. Identifying relevant spectral features and designing effective feature extraction methods can further improve SERS-ML models.

Addressing fundamental challenges in ML is also pivotal for advancing SERS-ML capabilities. Generalizability remains a significant concern, as models trained on lab-generated data sets may stumble when applied to real-world samples due to complex matrix effects and variations in background interference. Techniques such as meta-learning, adversarial robust ML, and domain transfer can enhance model robustness and out-of-distribution generalization.¹⁶⁴ Furthermore, the interpretability of ML models is essential for understanding the underlying chemical and physical processes captured in SERS spectra. While complex models such as deep neural networks offer high predictive power, their black-box nature poses interpretability challenges. Leveraging techniques such as explainable artificial intelligence, and visualization tools can help both ML practitioners and domain experts extract meaningful insights from complex spectra, further enhancing trustworthiness. Moreover, transitioning from a model-centric to a data-centric AI approach in SERS-ML is key to tackling real-world challenges more effectively. While conventional methods prioritize models, they often stumble in real-world scenarios due to data complexities. A data-centric focus prioritizes data quality, preprocessing tailored to SERS specifics, and robust validation. This ensures models are trained on top-notch data sets, boosting performance and adaptability. Embracing this shift not only deepens our understanding of data but also equips us with more resilient ML algorithms for practical SERS applications.

While resource-intensive, another hurdle is the requirement for rigorous validation against established analytical techniques on large sample sets as a means to validate the reliability and robustness of SERS-ML methods. Exploring synergies between SERS and other analytical techniques such as mass spectroscopy, infrared spectroscopy, and PCR can enrich chemical analysis and increase the reliability of SERS analysis. Integrating multimodal data fusion approaches can enhance the robustness and accuracy of SERS-based chemical sensing platforms. Collaborative interdisciplinary efforts are key to realizing the transformative impact of SERS-ML technologies on pressing societal and environmental issues.

Finally, the transformation of SERS-ML into a portable, field-ready tool necessitates comprehensive technological advancements at the system level. This entails the creation of compact SERS devices that seamlessly integrate with portable Raman spectroscopy tools and streamlined sample collection and handling systems. The development of efficient ML algorithms, along with optimized distributed controllercomputing-communication systems tailored for resource-constrained environments, and the exploration of hardware acceleration solutions, are pivotal needs for enabling on-site, real-time detection and decision-making driven by SERS-ML technology.

9. REFLECTING ON THE POTENTIAL: A CALL TO CONTEMPLATE SERS-ML INTEGRATION

While the authors acknowledge the significant potential of SERS-ML and foresee a promising future for integration, we present key considerations regarding the application of SERS-ML in real-time scenarios. Additionally, we offer a grand vision for transforming the field through the collaborative integration of SERS with ML. The first aspect to contemplate is democratized sensing and decision-making. Advancements in robust AI and hyperspectral analytics stand at the forefront of revolutionizing water quality monitoring. These advancements will enable the deployment of multiparameter aerial and autonomous aquatic robotic systems, equipped with SERS technology, across diverse geographical regions. By incorporating SERS-ML integration, these systems will have the capacity to efficiently detect and analyze water contaminants in real time, offering comprehensive insights into water quality. Costeffective, miniaturized sensors coupled with interpretable ML software can provide decentralized, real-time insights into water quality for local communities and municipalities.

Continuous scientific discoveries are also anticipated. The integration of vast data sets from distributed monitoring, along

with online deep learning and explainable models, may unveil new interdependencies and biogeochemical mechanisms. This continuous enhancement of process and fate knowledge regarding emerging pollutants is crucial for sustaining usable resources and has the potential to establish a feedback loop for diagnostics and future solutions. Moreover, evidence-based interventions and adaptations are envisioned. By combining spatiotemporal high-resolution contaminant predictions with human and ecological risk models, dynamic and costoptimizing policy interventions can be customized for local contexts. This approach aims to improve human and environmental health equitably through adaptive and targeted safeguards. In summary, the synergistic integration of spatial analytics, real-time AI, and systems modeling is envisaged to democratize, discover, and decide, thereby securing water futures effectively.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c06737.

Additional information on SERS band assignments for bacteria, polystyrene, nitrate, and PFAS (PDF)

AUTHOR INFORMATION

Corresponding Author

Peter J. Vikesland – Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; Virginia Tech Institute of Critical Technology and Applied Science (ICTAS) Sustainable Nanotechnology Center (VTSuN), Blacksburg, Virginia 24061, United States; o orcid.org/0000-0003-2654-5132; Email: vikesland@vt.edu

Authors

- Sonali Srivastava Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; Virginia Tech Institute of Critical Technology and Applied Science (ICTAS) Sustainable Nanotechnology Center (VTSuN), Blacksburg, Virginia 24061, United States
- Wei Wang Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; Virginia Tech Institute of Critical Technology and Applied Science (ICTAS) Sustainable Nanotechnology Center (VTSuN), Blacksburg, Virginia 24061, United States; • orcid.org/0000-0002-6231-2216
- Wei Zhou Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; o orcid.org/0000-0002-5257-3885
- Ming Jin Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c06737

Author Contributions

S. Srivastava directed the research effort, did the literature review, wrote the manuscript, and provided oversight of the research team. W. Wang, W. Zhou, and M. Jin aided in the writing of the manuscript. P. J. Vikesland obtained the research funding, provided oversight of the project, and collaborated in the writing and editing of the manuscript. All authors

contributed to the development of the manuscript and its revision.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the Sustainable Nanotechnology Interdisciplinary Graduate Program (VTSun IGEP) funded by the Virginia Tech Graduate School and through funding provided by the Virginia Tech Fralin Life Sciences Institute (FLSI).

REFERENCES

(1) Mao, K.; Zhang, K.; Du, W.; Ali, W.; Feng, X.; Zhang, H. The Potential of Wastewater-Based Epidemiology as Surveillance and Early Warning of Infectious Disease Outbreaks. *Curr. Opin. Environ. Sci. Heal.* **2020**, *17*, 1–7.

(2) Zulkifli, S. N.; Rahim, H. A.; Lau, W.-J. Detection of Contaminants in Water Supply: A Review on State-of-the-Art Monitoring Technologies and Their Applications. *Sensors Actuators B Chem.* 2018, 255, 2657–2689.

(3) Richardson, S. D. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* **2009**, *81* (12), 4645–4677.

(4) Jiang, G.; Liu, Y.; Tang, S.; Kitajima, M.; Haramoto, E.; Arora, S.; Choi, P. M.; Jackson, G.; D'Aoust, P. M.; Delatolla, R.; Zhang, S.; Guo, Y.; Wu, J.; Chen, Y.; Sharma, E.; Prosun, T. A.; Zhao, J.; Kumar, M.; Honda, R.; Ahmed, W.; Meiman, J. Moving Forward with COVID-19: Future Research Prospects of Wastewater-Based Epidemiology Methodologies and Applications. *Curr. Opin. Environ. Sci. Heal.* **2023**, *33*, No. 100458.

(5) Wade, M. J.; Lo Jacomo, A.; Armenise, E.; Brown, M. R.; Bunce, J. T.; Cameron, G. J.; Fang, Z.; Farkas, K.; Gilpin, D. F.; Graham, D. W.; Grimsley, J. M. S.; Hart, A.; Hoffmann, T.; Jackson, K. J.; Jones, D. L.; Lilley, C. J.; McGrath, J. W.; McKinley, J. M.; McSparron, C.; Nejad, B. F.; Morvan, M.; Quintela-Baluja, M.; Roberts, A. M. I.; Singer, A. C.; Souque, C.; Speight, V. L.; Sweetapple, C.; Walker, D.; Watts, G.; Weightman, A.; Kasprzyk-Hordern, B. Understanding and Managing Uncertainty and Variability for Wastewater Monitoring beyond the Pandemic: Lessons Learned from the United Kingdom National COVID-19 Surveillance Programmes. *J. Hazard. Mater.* **2022**, *424*, No. 127456.

(6) Brigmon, R. L.; Zam, S. G.; Bitton, G.; Farrah, S. R. Detection of Salmonella Enteritidis in Environmental Samples by Monoclonal Antibody-Based ELISA. *J. Immunol. Methods* **1992**, *152* (1), 135–142. (7) Zhang, S.; Li, X.; Wu, J.; Coin, L.; O'Brien, J.; Hai, F.; Jiang, G. Molecular Methods for Pathogenic Bacteria Detection and Recent Advances in Wastewater Analysis. *Water* **2021**, *13* (24), 3551.

(8) Kempińska, D.; Kot-Wasik, A. The Use of RP-HPLC-Q-TOF-MS as a Powerful Tool for Wastewater Composition Profiling and Selection of Water Pollution Marker Specific to Wastewater Contamination. *Monatshefte für Chemie - Chem. Mon.* **2018**, 149 (9), 1595–1604.

(9) Yang, S.; Rothman, R. E. PCR-Based Diagnostics for Infectious Diseases: Uses, Limitations, and Future Applications in Acute-Care Settings. *Lancet Infect. Dis.* **2004**, *4* (6), 337–348.

(10) Dey, T. Microplastic Pollutant Detection by Surface Enhanced Raman Spectroscopy (SERS): A Mini-Review. *Nanotechnol. Environ. Eng.* **2023**, *8* (1), 41–48.

(11) Feng, Y.; Dai, J.; Wang, C.; Zhou, H.; Li, J.; Ni, G.; Zhang, M.; Huang, Y. Ag Nanoparticle/Au@Ag Nanorod Sandwich Structures for SERS-Based Detection of Perfluoroalkyl Substances. *ACS Appl. Nano Mater.* **2023**, *6* (15), 13974–13983.

(12) Wei, H.; Hossein Abtahi, S. M.; Vikesland, P. J. Plasmonic Colorimetric and SERS Sensors for Environmental Analysis. *Environ. Sci. Nano* **2015**, *2* (2), 120–135.

(13) Kneipp, K.; Wang, Y.; Kneipp, H.; Perelman, L. T.; Itzkan, I.; Dasari, R. R.; Feld, M. S. Single Molecule Detection Using SurfaceEnhanced Raman Scattering (SERS). Phys. Rev. Lett. 1997, 78 (9), 1667-1670.

(14) Sharma, B.; Frontiera, R. R.; Henry, A.-I.; Ringe, E.; Van Duyne, R. P. SERS: Materials, Applications, and the Future. *Mater. Today* **2012**, *15* (1–2), 16–25.

(15) Terry, L. R.; Sanders, S.; Potoff, R. H.; Kruel, J. W.; Jain, M.; Guo, H. Applications of Surface-enhanced Raman Spectroscopy in Environmental Detection. *Anal. Sci. Adv.* **2022**, *3* (3–4), 113–145.

(16) Li, D.-W.; Zhai, W.-L.; Li, Y.-T.; Long, Y.-T. Recent Progress in Surface Enhanced Raman Spectroscopy for the Detection of Environmental Pollutants. *Microchim. Acta* **2014**, *181* (1–2), 23–43. (17) Xu, K.; Zhou, R.; Takei, K.; Hong, M. Toward Flexible Surface-Enhanced Raman Scattering (SERS) Sensors for Point-of-Care

Ennanced Raman Scattering (SERS) Sensors for Point-or-Care Diagnostics. Adv. Sci. 2019, 6 (16), No. 1900925.
(18) Halvorson, R. A.; Vikesland, P. J.Surface-enhanced Raman

spectroscopy (SERS) for environmental analyses. ; ACS Publications, 2010, 77497755.

(19) Zong, C.; Xu, M.; Xu, L.-J.; Wei, T.; Ma, X.; Zheng, X.-S.; Hu, R.; Ren, B. Surface-Enhanced Raman Spectroscopy for Bioanalysis: Reliability and Challenges. Chem. Rev. 2018, 118 (10), 4946-4980. (20) Langer, J.; Jimenez de Aberasturi, D.; Aizpurua, J.; Alvarez-Puebla, R. A.; Auguié, B.; Baumberg, J. J.; Bazan, G. C.; Bell, S. E. J.; Boisen, A.; Brolo, A. G.; Choo, J.; Cialla-May, D.; Deckert, V.; Fabris, L.; Faulds, K.; García de Abajo, F. J.; Goodacre, R.; Graham, D.; Haes, A. J.; Haynes, C. L.; Huck, C.; Itoh, T.; Käll, M.; Kneipp, J.; Kotov, N. A.; Kuang, H.; Le Ru, E. C.; Lee, H. K.; Li, J.-F.; Ling, X. Y.; Maier, S. A.; Mayerhöfer, T.; Moskovits, M.; Murakoshi, K.; Nam, J.-M.; Nie, S.; Ozaki, Y.; Pastoriza-Santos, I.; Perez-Juste, J.; Popp, J.; Pucci, A.; Reich, S.; Ren, B.; Schatz, G. C.; Shegai, T.; Schlücker, S.; Tay, L.-L.; Thomas, K. G.; Tian, Z.-Q.; Van Duyne, R. P.; Vo-Dinh, T.; Wang, Y.; Willets, K. A.; Xu, C.; Xu, H.; Xu, Y.; Yamamoto, Y. S.; Zhao, B.; Liz-Marzán, L. M. Present and Future of Surface-Enhanced Raman Scattering. ACS Nano 2020, 14 (1), 28-117.

(21) Li, D.; Yue, W.; Gao, P.; Gong, T.; Wang, C.; Luo, X. Surface-Enhanced Raman Spectroscopy (SERS) for the Characterization of Atmospheric Aerosols: Current Status and Challenges. *TrAC Trends Anal. Chem.* **2024**, *170*, No. 117426.

(22) Zhang, Y.; Zhao, S.; Zheng, J.; He, L. Surface-Enhanced Raman Spectroscopy (SERS) Combined Techniques for High-Performance Detection and Characterization. *TrAC Trends Anal. Chem.* **2017**, *90*, 1–13.

(23) Plou, J.; Valera, P. S.; García, I.; de Albuquerque, C. D. L.; Carracedo, A.; Liz-Marzán, L. M. Prospects of Surface-Enhanced Raman Spectroscopy for Biomarker Monitoring toward Precision Medicine. ACS Photonics **2022**, 9 (2), 333–350.

(24) Masson, J.-F.; Biggins, J. S.; Ringe, E. Machine Learning for Nanoplasmonics. *Nat. Nanotechnol.* **2023**, *18* (2), 111–123.

(25) Lussier, F.; Thibault, V.; Charron, B.; Wallace, G. Q.; Masson, J.-F. Deep Learning and Artificial Intelligence Methods for Raman and Surface-Enhanced Raman Scattering. *TrAC Trends Anal. Chem.* **2020**, *124*, No. 115796.

(26) Bi, X.; Lin, L.; Chen, Z.; Ye, J. Artificial Intelligence for Surface-Enhanced Raman Spectroscopy. *Small Methods* **2024**, 8 (1), No. 2301243.

(27) Horta-Velázquez, A.; Arce, F.; Rodríguez-Sevilla, E.; Morales-Narváez, E. Toward Smart Diagnostics via Artificial Intelligence-Assisted Surface-Enhanced Raman Spectroscopy. *TrAC Trends Anal. Chem.* **2023**, *169*, No. 117378.

(28) Le Ru, E.; Etchegoin, P. Principles of Surface-Enhanced Raman Spectroscopy: And Related Plasmonic Effects; Elsevier, 2008.

(29) Qian, X.-M.; Nie, S. M. Single-Molecule and Single-Nanoparticle SERS: From Fundamental Mechanisms to Biomedical Applications. *Chem. Soc. Rev.* **2008**, 37 (5), 912–920.

(30) Schatz, G. C.; Young, M. A.; Van Duyne, R. P. Electromagnetic Mechanism of SERS. In *Surface-enhanced Raman scattering: physics and applications*; Springer, 2006; pp 19–45.

(31) Yang, S.; Dai, X.; Stogin, B. B.; Wong, T.-S. Ultrasensitive Surface-Enhanced Raman Scattering Detection in Common Fluids. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (2), 268–273. (32) Kneipp, J.; Kneipp, H.; Kneipp, K. SERS—a Single-Molecule and Nanoscale Tool for Bioanalytics. *Chem. Soc. Rev.* 2008, 37 (5), 1052.

(33) Premasiri, W. R.; Moir, D. T.; Klempner, M. S.; Krieger, N.; Jones, G.; Ziegler, L. D. Characterization of the Surface Enhanced Raman Scattering (SERS) of Bacteria. *J. Phys. Chem. B* **2005**, *109* (1), 312–320.

(34) Pineda, A. C.; Ronis, D. Fluorescence Quenching in Molecules near Rough Metal Surfaces. *J. Chem. Phys.* **1985**, 83 (10), 5330–5337. (35) Chance, R. R.; Prock, A.; Silbey, R. Molecular Fluorescence and Energy Transfer near Interfaces. *Adv. Chem. Phys.* **1978**, 37, 1–65.

(36) Zhang, B.; Hou, X.; Zhen, C.; Wang, A. X. Sub-Part-per-Billion Level Sensing of Fentanyl Residues from Wastewater Using Portable Surface-Enhanced Raman Scattering Sensing. *Biosensors* **2021**, *11* (10), 370.

(37) Lv, L.; He, L.; Jiang, S.; Chen, J.; Zhou, C.; Qu, J.; Lu, Y.; Hong, P.; Sun, S.; Li, C. In Situ Surface-Enhanced Raman Spectroscopy for Detecting Microplastics and Nanoplastics in Aquatic Environments. *Sci. Total Environ.* **2020**, *728*, No. 138449.

(38) Huang, T.; McClelland, A.; Zeng, T. H. Trace PFAS Detection in Water Sources Using Silver Nanoparticles for Surface-Enhanced Raman Spectroscopy (SERS). In 2022 IEEE 22nd International Conference on Nanotechnology (NANO). 2022; pp 342–345.

(39) Moskovits, M. Surface-Enhanced Spectroscopy. *Rev. Mod. Phys.* **1985**, 57 (3), 783.

(40) Wang, Z.; Zong, S.; Wu, L.; Zhu, D.; Cui, Y. SERS-Activated Platforms for Immunoassay: Probes, Encoding Methods, and Applications. *Chem. Rev.* 2017, *117* (12), 7910–7963.

(41) Wang, W.; Srivastava, S.; Garg, A.; Xiao, C.; Hawks, S.; Pan, J.; Duggal, N.; Isaacman-VanWertz, G.; Zhou, W.; Marr, L. C.; et al. Digital Surface-Enhanced Raman Spectroscopy-Lateral Flow Test Dipstick: Ultrasensitive, Rapid Virus Quantification in Environmental Dust. *Environ. Sci. Technol.* **2024**, *58* (11), 4926–4936.

(42) Li, Y.; Lu, C.; Zhou, S.; Fauconnier, M.-L.; Gao, F.; Fan, B.; Lin, J.; Wang, F.; Zheng, J. Sensitive and Simultaneous Detection of Different Pathogens by Surface-Enhanced Raman Scattering Based on Aptamer and Raman Reporter Co-Mediated Gold Tags. *Sensors Actuators B Chem.* **2020**, *317*, No. 128182.

(43) Chuong, T. T.; Pallaoro, A.; Chaves, C. A.; Li, Z.; Lee, J.; Eisenstein, M.; Stucky, G. D.; Moskovits, M.; Soh, H. T. Dual-Reporter SERS-Based Biomolecular Assay with Reduced False-Positive Signals. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (34), 9056–9061.

(44) El Naqa, I.; Murphy, M. J. What Is Machine Learning? In *Machine Learning in Radiation Oncology*; Springer International Publishing: Cham, 2015; pp 3–11. DOI: 10.1007/978-3-319-18305-3 1.

(45) Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2* (3), 160.

(46) Kotsiantis, S. B.; Zaharakis, I.; Pintelas, P.; et al. Supervised Machine Learning: A Review of Classification Techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **2007**, *160* (1), 3–24.

(47) Mahesh, B. Machine Learning Algorithms-a Review. Int. J. Sci. Res. (IJSR).[Internet] 2020, 9 (1), 381–386.

(48) Ringnér, M. What Is Principal Component Analysis? Nat. Biotechnol. 2008, 26 (3), 303-304.

(49) Montes-García, V.; Gómez-González, B.; Martínez-Solís, D.; Taboada, J. M.; Jiménez-Otero, N.; De Una-Álvarez, J.; Obelleiro, F.; García-Río, L.; Pérez-Juste, J.; Pastoriza-Santos, I. Pillar[5]Arene-Based Supramolecular Plasmonic Thin Films for Label-Free, Quantitative and Multiplex SERS Detection. ACS Appl. Mater. Interfaces **2017**, 9 (31), 26372–26382.

(50) Luo, Y.; Qi, F.; Gibson, C. T.; Lei, Y.; Fang, C. Investigating Kitchen Sponge-Derived Microplastics and Nanoplastics with Raman Imaging and Multivariate Analysis. *Sci. Total Environ.* **2022**, *824*, No. 153963.

(51) Li, X.; Yang, T.; Song, Y.; Zhu, J.; Wang, D.; Li, W. Surface-Enhanced Raman Spectroscopy (SERS)-Based Immunochromatographic Assay (ICA) for the Simultaneous Detection of Two Pyrethroid Pesticides. *Sensors Actuators, B Chem.* **2019**, 283, 230–238. (52) Raveendran, J.; Docoslis, A. Detection and Quantification of Toxicants in Food and Water Using Ag–Au Core-Shell Fractal SERS Nanostructures and Multivariate Analysis. *Talanta* **2021**, 231 (April),

No. 122383. (53) Sahin, F.; Celik, N.; Camdal, A.; Sakir, M.; Ceylan, A.; Ruzi, M.; Onses, M. S. Machine Learning-Assisted Pesticide Detection on a Flexible Surface-Enhanced Raman Scattering Substrate Prepared by Silver Nanoparticles. *ACS Appl. Nano Mater.* **2022**, 5 (9), 13112– 13122.

(54) Ge, K.; Li, Y.; Wu, Q.; Gu, Y. Simultaneous and Rapid Detection of Polychlorinated Phenols in Water Samples by Surface-Enhanced Raman Spectroscopy Combined with Principal Component Analysis. *Anal. Bioanal. Chem.* **2022**, *414* (7), 2385–2395.

(55) Patel, I. S.; Premasiri, W. R.; Moir, D. T.; Ziegler, L. D. Barcoding Bacterial Cells: A SERS-Based Methodology for Pathogen Identification. *J. Raman Spectrosc.* **2008**, 1660–1672.

(56) Guicheteau, J.; Argue, L.; Emge, D.; Hyre, A.; Jacobson, M.; Christesen, S. Bacillus Spore Classification via Surface-Enhanced Raman Spectroscopy and Principal Component Analysis. *Appl. Spectrosc.* **2008**, *62* (3), 267–272.

(57) Yang, B.; Jin, S.; Guo, S.; Park, Y.; Chen, L.; Zhao, B.; Jung, Y. M. Recent Development of SERS Technology: Semiconductor-Based Study. *ACS Omega* **2019**, *4* (23), 20101–20108.

(58) Arslan, A. H.; Ciloglu, F. U.; Yilmaz, U.; Simsek, E.; Aydin, O. Discrimination of Waterborne Pathogens, Cryptosporidium Parvum Oocysts and Bacteria Using Surface-Enhanced Raman Spectroscopy Coupled with Principal Component Analysis and Hierarchical Clustering. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 2022, 267, No. 120475.

(59) Zhang, Z.; Li, D.; Wang, X.; Wang, Y.; Lin, J.; Jiang, S.; Wu, Z.; He, Y.; Gao, X.; Zhu, Z.; Xiao, Y.; Qu, Z.; Li, Y. Rapid Detection of Viruses: Based on Silver Nanoparticles Modified with Bromine Ions and Acetonitrile. *Chem. Eng. J.* **2022**, *438*, No. 135589.

(60) Ergon, R.; Granato, D.; Ares, G. Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR). *Math. Stat. Methods Food Sci. Technol.*; Wiley Blackwell: Chichester, 2014; pp 121–142.

(61) Wang, W.; Rahman, A.; Kang, S.; Vikesland, P. J. Investigation of the Influence of Stress on Label-Free Bacterial Surface-Enhanced Raman Spectra. *Anal. Chem.* **2023**, *95* (7), 3675–3683.

(62) Ahmad, S.; Majeed, M. I.; Nawaz, H.; Javed, M. R.; Rashid, N.; Abubakar, M.; Batool, F.; Bashir, S.; Kashif, M.; Ali, S.; et al. Characterization and Prediction of Viral Loads of Hepatitis B Serum Samples by Using Surface-Enhanced Raman Spectroscopy (SERS). *Photodiagnosis Photodyn. Ther.* **2021**, *35*, No. 102386.

(63) MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*; 1967; Vol. 1, pp 281–297.

(64) Reokrungruang, P.; Chatnuntawech, I.; Dharakul, T.; Bamrungsap, S. A Simple Paper-Based Surface Enhanced Raman Scattering (SERS) Platform and Magnetic Separation for Cancer Screening. *Sensors Actuators, B Chem.* **2019**, 285, 462–469.

(65) Banaei, N.; Moshfegh, J.; Mohseni-Kabir, A.; Houghton, J. M.; Sun, Y.; Kim, B. Machine Learning Algorithms Enhance the Specificity of Cancer Biomarker Detection Using SERS-Based Immunoassays in Microfluidic Chips. *RSC Adv.* **2019**, *9* (4), 1859– 1868.

(66) Malik, A.; Tuckfield, B. Applied Unsupervised Learning with R: Uncover Hidden Relationships and Patterns with k-Means Clustering, Hierarchical Clustering, and PCA; Packt Publishing Ltd, 2019.

(67) Liaw, A.; Wiener, M.; et al. Classification and Regression by RandomForest. *R News* **2002**, *2* (3), 18–22.

(68) Lu, S.; Yin, Z.; Zhang, L.; Du, J.; Jing, C. On-Site Detection of Multiple Extracellular Antibiotic Resistance Genes Using SERS. Sensors Actuators B Chem. 2022, 369 (May), No. 132262. (69) Guo, M.; Li, M.; Fu, H.; Zhang, Y.; Chen, T.; Tang, H.; Zhang, T.; Li, H. Quantitative Analysis of Polycyclic Aromatic Hydrocarbons (PAHs) in Water by Surface-Enhanced Raman Spectroscopy (SERS) Combined with Random Forest. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2023**, 287 (P1), No. 122057.

(70) Soman, K. P.; Loganathan, R.; Ajay, V. Machine Learning with SVM and Other Kernel Methods; PHI Learning Pvt. Ltd., 2009.

(71) Jian, C.; Boyan, J.; Ying, Z.; Zhenyu, W. Determining the Concentration of Polycyclic Aromatic Hydrocarbons in Water Using Surface Enhanced Raman Spectroscopy and Kernel Principal Components Analysis Combined with Support Vector Regression. *J. Appl. Spectrosc.* **2021**, *88* (1), 225–232.

(72) Wang, W.; Kang, S.; Vikesland, P. J. Surface-Enhanced Raman Spectroscopy of Bacterial Metabolites for Bacterial Growth Monitoring and Diagnosis of Viral Infection. *Environ. Sci. Technol.* **2021**, *55* (13), 9119–9128.

(73) Rahman, A.; Kang, S.; Wang, W.; Huang, Q.; Kim, I.; Vikesland, P. J. Lectin-Modified Bacterial Cellulose Nanocrystals Decorated with Au Nanoparticles for Selective Detection of Bacteria Using Surface-Enhanced Raman Scattering Coupled with Machine Learning. *ACS Appl. Nano Mater.* **2022**, *5* (1), 259–268.

(74) Li, Y.; Lin, C.; Peng, Y.; He, J.; Yang, Y. High-Sensitivity and Point-of-Care Detection of SARS-CoV-2 from Nasal and Throat Swabs by Magnetic SERS Biosensor. *Sensors Actuators B Chem.* **2022**, 365 (April), No. 131974.

(75) Wright, R. E. Logistic Regression. 1995.

(76) Zeng, X.; Liu, Y.; Liu, W.; Yuan, C.; Luo, X.; Xie, F.; Chen, X.; de la Chapelle, M. L.; Tian, H.; Yang, X.; Fu, W. Evaluation of Classification Ability of Logistic Regression Model on SERS Data of MiRNAs. J. Biophotonics **2022**, 15 (12), 1–12.

(77) Moisoiu, T.; Dragomir, M. P.; Iancu, S. D.; Schallenberg, S.; Birolo, G.; Ferrero, G.; Burghelea, D.; Stefancu, A.; Cozan, R. G.; Licarete, E.; Allione, A.; Matullo, G.; Iacob, G.; Bálint, Z.; Badea, R. I.; Naccarati, A.; Horst, D.; Pardini, B.; Leopold, N.; Elec, F. Combined MiRNA and SERS Urine Liquid Biopsy for the Point-of-Care Diagnosis and Molecular Stratification of Bladder Cancer. *Mol. Med.* **2022**, DOI: 10.1186/s10020-022-00462-z.

(78) Qin, J.; Tian, X.; Liu, S.; Yang, Z.; Shi, D.; Xu, S.; Zhang, Y. Rapid Classification of SARS-CoV-2 Variant Strains Using Machine Learning-Based Label-Free SERS Strategy. *Talanta* **2024**, *267*, No. 125080.

(79) Kim, J. Y.; Koh, E. H.; Yang, J. Y.; Mun, C.; Lee, S.; Lee, H.; Kim, J.; Park, S. G.; Kang, M.; Kim, D. H.; Jung, H. S. 3D Plasmonic Gold Nanopocket Structure for SERS Machine Learning-Based Microplastic Detection. *Adv. Funct. Mater.* **2024**, *34*, 1–11.

(80) Chen, W.-H.; Hsu, S.-H.; Shen, H.-P. Application of SVM and ANN for Intrusion Detection. *Comput. Oper. Res.* 2005, 32 (10), 2617–2634.

(81) Nam, W.; Chen, H.; Ren, X.; Agah, M.; Kim, I.; Zhou, W. Nanolaminate Plasmonic Substrates for High-Throughput Living Cell SERS Measurements and Artificial Neural Network Classification of Cellular Drug Responses. *ACS Appl. Nano Mater.* **2022**, *5* (8), 10358–10368.

(82) Huang, Y.; Yuan, B.; Wang, X.; Dai, Y.; Wang, D.; Gong, Z.; Chen, J.; Shen, L.; Fan, M.; Li, Z. Industrial Wastewater Source Tracing: The Initiative of SERS Spectral Signature Aided by a One-Dimensional Convolutional Neural Network. *Water Res.* **2023**, *232*, No. 119662.

(83) Thrift, W. J.; Ragan, R. Quantification of Analyte Concentration in the Single Molecule Regime Using Convolutional Neural Networks. *Anal. Chem.* **2019**, *91* (21), 13337–13342.

(84) O'Shea, K.; Nash, R. An Introduction to Convolutional Neural Networks. *arXiv*, November 26, 2015. DOI: 10.48550/arXiv.1511.08458.

(85) Atta, S.; Li, J. Q.; Vo-Dinh, T. Multiplex SERS Detection of Polycyclic Aromatic Hydrocarbon (PAH) Pollutants in Water Samples Using Gold Nanostars and Machine Learning Analysis. *Analyst* 2023, 148 (20), 5105–5116. (86) Erzina, M.; Trelin, A.; Guselnikova, O.; Dvorankova, B.; Strnadova, K.; Perminova, A.; Ulbrich, P.; Mares, D.; Jerabek, V.; Elashnikov, R.; Svorcik, V.; Lyutakov, O. Precise Cancer Detection via the Combination of Functionalized SERS Surfaces and Convolutional Neural Network with Independent Inputs. *Sensors Actuators B Chem.* **2020**, *308*, No. 127660.

pubs.acs.org/est

(87) Ding, J.; Lin, Q.; Zhang, J.; Young, G. M.; Jiang, C.; Zhong, Y.; Zhang, J. Rapid Identification of Pathogens by Using Surface-Enhanced Raman Spectroscopy and Multi-Scale Convolutional Neural Network. *Anal. Bioanal. Chem.* **2021**, *413* (14), 3801–3811.

(88) Popescu, M.-C.; Balas, V. E.; Perescu-Popescu, L.; Mastorakis, N. Multilayer Perceptron and Neural Networks. *WSEAS Trans. Circuits Syst.* **2009**, 8 (7), 579–588.

(89) Wang, L.; Zhang, X.-D.; Tang, J.-W.; Ma, Z.-W.; Usman, M.; Liu, Q.-H.; Wu, C.-Y.; Li, F.; Zhu, Z.-B.; Gu, B. Machine Learning Analysis of SERS Fingerprinting for the Rapid Determination of Mycobacterium Tuberculosis Infection and Drug Resistance. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 5364–5377.

(90) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016; pp 770–778.

(91) Tseng, Y.-M.; Chen, K.-L.; Chao, P.-H.; Han, Y.-Y.; Huang, N.-T. Deep Learning–Assisted Surface-Enhanced Raman Scattering for Rapid Bacterial Identification. *ACS Appl. Mater. Interfaces* **2023**, *15* (22), 26398–26406.

(92) Li, R.; Wang, Z.; Zhang, Z.; Sun, X.; Hu, Y.; Wang, H.; Chen, K.; Liu, Q.; Chen, M.; Chen, X. Deep Learning-Based Multicapturer SERS Platform on Plasmonic Nanocube Metasurfaces for Multiplex Detection of Organophosphorus Pesticides in Environmental Water. *Anal. Chem.* **2022**, *94* (46), 16006–16014.

(93) Wang, C.; Xiao, L.; Dai, C.; Nguyen, A. H.; Littlepage, L. E.; Schultz, Z. D.; Li, J. A Statistical Approach of Background Removal and Spectrum Identification for SERS Data. *Sci. Rep.* **2020**, *10* (1), 1460.

(94) Kotsiantis, S. B.; Zaharakis, I. D.; Pintelas, P. E. Machine Learning: A Review of Classification and Combining Techniques. *Artif. Intell. Rev.* **2006**, *26* (3), 159–190.

(95) Belkin, M.; Hsu, D.; Ma, S.; Mandal, S. Reconciling Modern Machine-Learning Practice and the Classical Bias–Variance Trade-Off. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (32), 15849–15854.

(96) Bartz, E.; Bartz-Beielstein, T.; Zaefferer, M.; Mersmann, O. Hyperparameter Tuning for Machine and Deep Learning with R: A Practical Guide; Springer Nature, 2023.

(97) Li, Z. Extracting Spatial Effects from Machine Learning Model Using Local Interpretation Method: An Example of SHAP and XGBoost. *Comput. Environ. Urban Syst.* **2022**, *96*, No. 101845.

(98) Kasera, S.; Herrmann, L. O.; Barrio, J. d.; Baumberg, J. J.; Scherman, O. A. Quantitative Multiplexing with Nano-Self-Assemblies in SERS. *Sci. Rep.* **2014**, *4* (1), 6785.

(99) Kang, S.; Kim, I.; Vikesland, P. J. Discriminatory Detection of SsDNA by Surface-Enhanced Raman Spectroscopy (SERS) and Tree-Based Support Vector Machine (Tr-SVM). *Anal. Chem.* **2021**, 93 (27), 9319–9328.

(100) Kang, S.; Nam, W.; Zhou, W.; Kim, I.; Vikesland, P. J. Nanostructured Au-Based Surface-Enhanced Raman Scattering Substrates and Multivariate Regression for PH Sensing. ACS Appl. Nano Mater. 2021, 4, 5768.

(101) Sundaram, J.; Park, B.; Hinton, A.; Lawrence, K. C.; Kwon, Y. Detection and Differentiation of Salmonella Serotypes Using Surface Enhanced Raman Scattering (SERS) Technique. *J. Food Meas. Charact.* **2013**, *7* (1), 1–12.

(102) Wang, C.; Madiyar, F.; Yu, C.; Li, J. Detection of Extremely Low Concentration Waterborne Pathogen Using a Multiplexing Self-Referencing SERS Microfluidic Biosensor. J. Biol. Eng. 2017, 11 (1), 1–11.

(103) Ding, Y.; Sun, Y.; Liu, C.; Jiang, Q. Y.; Chen, F.; Cao, Y. SERS-Based Biosensors Combined with Machine Learning for Medical Application. *ChemistryOpen* **2023**, *12* (1), No. e202200192.

(104) Schmidt, C. Watcher in the Wastewater. *Nat. Biotechnol.* **2020**, 38 (8), 917–920.

(105) Jones, D. L.; Baluja, M. Q.; Graham, D. W.; Corbishley, A.; McDonald, J. E.; Malham, S. K.; Hillary, L. S.; Connor, T. R.; Gaze, W. H.; Moura, I. B.; Wilcox, M. H.; Farkas, K. Shedding of SARS-CoV-2 in Feces and Urine and Its Potential Role in Person-to-Person Transmission and the Environment-Based Spread of COVID-19. *Sci. Total Environ.* **2020**, *749*, No. 141364.

(106) Wang, J.; Cai, K.; Zhang, R.; He, X.; Shen, X.; Liu, J.; Xu, J.; Qiu, F.; Lei, W.; Wang, J.; Li, X.; Gao, Y.; Jiang, Y.; Xu, W.; Ma, X. Novel One-Step Single-Tube Nested Quantitative Real-Time PCR Assay for Highly Sensitive Detection of SARS-CoV-2. *Anal. Chem.* **2020**, 92 (13), 9399–9404.

(107) Yang, Y.; Peng, Y.; Lin, C.; Long, L.; Hu, J.; He, J.; Zeng, H.; Huang, Z.; Li, Z. Y.; Tanemura, M.; Shi, J.; Lombardi, J. R.; Luo, X. Human ACE2-Functionalized Gold "Virus-Trap" Nanostructures for Accurate Capture of SARS-CoV-2 and Single-Virus SERS Detection. *Nano-Micro Lett.* **2021**, *13* (1), 1–13.

(108) Garg, A.; Hawks, S.; Pan, J.; Wang, W.; Duggal, N.; Marr, L. C.; Vikesland, P.; Zhou, W. Machine Learning-Driven SERS Fingerprinting of Disintegrated Viral Components for Rapid Detection of SARS-CoV-2 in Environmental Dust. *Biosens. Bioelectron.* **2024**, 247, No. 115946.

(109) Oh, S.; Hammes, F.; Liu, W.-T. Metagenomic Characterization of Biofilter Microbial Communities in a Full-Scale Drinking Water Treatment Plant. *Water Res.* **2018**, *128*, 278–285.

(110) Ramos-Mandujano, G.; Salunke, R.; Mfarrej, S.; Rachmadi, A. T.; Hala, S.; Xu, J.; Alofi, F. S.; Khogeer, A.; Hashem, A. M.; Almontashiri, N. A. M.; Alsomali, A.; Shinde, D. B.; Hamdan, S.; Hong, P.; Pain, A.; Li, M. A Robust, Safe, and Scalable Magnetic Nanoparticle Workflow for RNA Extraction of Pathogens from Clinical and Wastewater Samples. *Glob. Challenges* **2021**, DOI: 10.1002/gch2.202000068.

(111) Hiller, C. X.; Hübner, U.; Fajnorova, S.; Schwartz, T.; Drewes, J. E. Antibiotic Microbial Resistance (AMR) Removal Efficiencies by Conventional and Advanced Wastewater Treatment Processes: A Review. *Sci. Total Environ.* **2019**, *685*, 596–608.

(112) Pazda, M.; Kumirska, J.; Stepnowski, P.; Mulkiewicz, E. Antibiotic Resistance Genes Identified in Wastewater Treatment Plant Systems – A Review. *Sci. Total Environ.* **2019**, 697, No. 134023.

(113) Garcia-Rico, E.; Alvarez-Puebla, R. A.; Guerrini, L. Direct Surface-Enhanced Raman Scattering (SERS) Spectroscopy of Nucleic Acids: From Fundamental Studies to Real-Life Applications. *Chem. Soc. Rev.* **2018**, 47 (13), 4909–4923.

(114) Papadopoulou, E.; Bell, S. E. J. Label-Free Detection of Single-Base Mismatches in DNA by Surface-Enhanced Raman Spectroscopy. *Angew. Chemie - Int. Ed.* **2011**, *50* (39), 9058–9061.

(115) Bell, S. E. J.; Sirimuthu, N. M. S. Surface-Enhanced Raman Spectroscopy (SERS) for Sub-Micromolar Detection of DNA/RNA Mononucleotides. J. Am. Chem. Soc. 2006, 128, 15580–15581.

(116) Xu, L. J.; Lei, Z. C.; Li, J.; Zong, C.; Yang, C. J.; Ren, B. Label-Free Surface-Enhanced Raman Spectroscopy Detection of DNA with Single-Base Sensitivity. J. Am. Chem. Soc. 2015, 137 (15), 5149–5154. (117) Czaplicka, M.; Kowalska, A. A.; Nowicka, A. B.; Kurzydłowski, D.; Gronkiewicz, Z.; Machulak, A.; Kukwa, W.; Kamińska, A. Raman Spectroscopy and Surface-Enhanced Raman Spectroscopy (SERS) Spectra of Salivary Glands Carcinoma, Tumor and Healthy Tissues and Their Homogenates Analyzed by Chemometry: Towards Development of the Novel Tool for Clinical Diagnosis. Anal. Chim. Acta 2021, 1177, No. 338784.

(118) Bailey, M. R.; Pentecost, A. M.; Selimovic, A.; Martin, R. S.; Schultz, Z. D. Sheath-Flow Microfluidic Approach for Combined Surface Enhanced Raman Scattering and Electrochemical Detection. *Anal. Chem.* **2015**, *87* (8), 4347–4355.

(119) Su, J.; Wang, D.; Nörbel, L.; Shen, J.; Zhao, Z.; Dou, Y.; Peng, T.; Shi, J.; Mathur, S.; Fan, C.; Song, S. Multicolor Gold-Silver Nano-Mushrooms as Ready-to-Use SERS Probes for Ultrasensitive and Multiplex DNA/MiRNA Detection. *Anal. Chem.* **2017**, *89* (4), 2531–2538.

(120) Panigrahi, S.; Pal, T. Nanoparticles with Raman Spectroscopic Fingerprints for DNA and RNA Detection. *Chemtracts* **2004**, *17* (4), 200–205.

(121) Stading, R.; Gastelum, G.; Chu, C.; Jiang, W.; Moorthy, B. Molecular Mechanisms of Pulmonary Carcinogenesis by Polycyclic Aromatic Hydrocarbons (PAHs): Implications for Human Lung Cancer. *Semin. Cancer Biol.* **2021**, *76*, 3–16.

(122) Tang, G.; Liu, M.; Zhou, Q.; He, H.; Chen, K.; Zhang, H.; Hu, J.; Huang, Q.; Luo, Y.; Ke, H.; Chen, B.; Xu, X.; Cai, M. Microplastics and Polycyclic Aromatic Hydrocarbons (PAHs) in Xiamen Coastal Areas: Implications for Anthropogenic Impacts. *Sci. Total Environ.* **2018**, *634*, 811–820.

(123) Sharma, B. M.; Melymuk, L.; Bharat, G. K.; Přibylová, P.; Sáňka, O.; Klánová, J.; Nizzetto, L. Spatial Gradients of Polycyclic Aromatic Hydrocarbons (PAHs) in Air, Atmospheric Deposition, and Surface Water of the Ganges River Basin. *Sci. Total Environ.* **2018**, *627*, 1495–1504.

(124) Idowu, O.; Carbery, M.; O'Connor, W.; Thavamani, P. Speciation and Source Apportionment of Polycyclic Aromatic Compounds (PACs) in Sediments of the Largest Salt Water Lake of Australia. *Chemosphere* **2020**, *246*, No. 125779.

(125) Fellet, G.; Pošćić, F.; Licen, S.; Marchiol, L.; Musetti, R.; Tolloi, A.; Barbieri, P.; Zerbi, G. PAHs Accumulation on Leaves of Six Evergreen Urban Shrubs: A Field Experiment. *Atmos. Pollut. Res.* **2016**, 7 (5), 915–924.

(126) Huang, J.; Zhou, T.; Zhao, W.; Cui, S.; Guo, R.; Li, D.; Reddy Kadasala, N.; Han, D.; Jiang, Y.; Liu, Y.; Liu, H. Multifunctional Magnetic Fe3O4/Cu2O-Ag Nanocomposites with High Sensitivity for SERS Detection and Efficient Visible Light-Driven Photocatalytic Degradation of Polycyclic Aromatic Hydrocarbons (PAHs). J. Colloid Interface Sci. **2022**, 628, 315–326.

(127) Gu, H.-X.; Hu, K.; Li, D.-W.; Long, Y.-T. SERS Detection of Polycyclic Aromatic Hydrocarbons Using a Bare Gold Nanoparticles Coupled Film System. *Analyst* **2016**, *141* (14), 4359–4365.

(128) Péron, O.; Rinnert, E.; Lehaitre, M.; Crassous, P.; Compère, C. Detection of Polycyclic Aromatic Hydrocarbon (PAH) Compounds in Artificial Sea-Water Using Surface-Enhanced Raman Scattering (SERS). *Talanta* **2009**, *79* (2), 199–204.

(129) Yan, X.; Shi, X.; Ma, J. Fast Quantification of the Mixture of Polycyclic Aromatic Hydrocarbons Using Surface-Enhanced Raman Spectroscopy Combined with PLS-GA-BP Network. *J. Ocean Univ. China* **2021**, *20* (6), 1451–1458.

(130) Bajomo, M. M.; Ju, Y.; Zhou, J.; Elefterescu, S.; Farr, C.; Zhao, Y.; Neumann, O.; Nordlander, P.; Patel, A.; Halas, N. J. Computational Chromatography: A Machine Learning Strategy for Demixing Individual Chemical Components in Complex Mixtures. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, DOI: 10.1073/pnas.2211406119.

(131) Zhu, C.; Meng, G.; Zheng, P.; Huang, Q.; Li, Z.; Hu, X.; Wang, X.; Huang, Z.; Li, F.; Wu, N. A Hierarchically Ordered Array of Silver-Nanorod Bundles for Surface-Enhanced Raman Scattering Detection of Phenolic Pollutants. *Adv. Mater.* **2016**, *28* (24), 4871–4876.

(132) Weng, S.; Li, M.; Chen, C.; Gao, X.; Zheng, S.; Zeng, X. Fast and Accurate Determination of Organophosphate Pesticides Using Surface-Enhanced Raman Scattering and Chemometrics. *Anal. Methods* **2015**, 7 (6), 2563–2567.

(133) Lin, S.; Fang, X.; Fang, G.; Liu, F.; Dong, H.; Zhao, H.; Zhang, J.; Dong, B. Ultrasensitive Detection and Distinction of Pollutants Based on SERS Assisted by Machine Learning Algorithms. *Sensors Actuators B Chem.* **2023**, *384*, No. 133651.

(134) Verma, P.; Ratan, J. K. Assessment of the Negative Effects of Various Inorganic Water Pollutants on the Biosphere—an Overview. In *Inorganic Pollutants in Water*; Elsevier, 2020; pp 73–96. DOI: 10.1016/B978-0-12-818965-8.00005-6.

(135) Smith, V. H.; Schindler, D. W. Eutrophication Science: Where Do We Go from Here? *Trends Ecol. Evol.* **2009**, *24* (4), 201–207.

(136) Majumdar, D. The Blue Baby Syndrome. Resonance 2003, 8 (10), 20-30.

(137) Mosier-Boss, P. A.; Lieberman, S. H. Detection of Anionic Nutrients by SERS of Cationic-Coated Silver Substrates. Effect of Chloride Ion. *Appl. Spectrosc.* **2001**, *55* (10), 1327–1336.

(138) Küster, T.; Bothun, G. D. In Situ SERS Detection of Dissolved Nitrate on Hydrated Gold Substrates. *Nanoscale Adv.* **2021**, *3* (14), 4098–4105.

(139) Küster, T.; Bothun, G. D. In Situ SERS Detection of Dissolved Nitrate on Hydrated Gold Substrates. *Nanoscale Adv.* **2021**, *3* (14), 4098–4105.

(140) Gajaraj, S.; Fan, C.; Lin, M.; Hu, Z. Quantitative Detection of Nitrate in Water and Wastewater by Surface-Enhanced Raman Spectroscopy. *Environ. Monit. Assess.* **2013**, *185* (7), 5673–5681.

(141) Kitaw, S. L.; Birhan, Y. S.; Tsai, H.-C. Plasmonic Surface-Enhanced Raman Scattering Nano-Substrates for Detection of Anionic Environmental Contaminants: Current Progress and Future Perspectives. *Environ. Res.* **2023**, *221*, No. 115247.

(142) Wu, Q.; Leung, J. Y. S.; Geng, X.; Chen, S.; Huang, X.; Li, H.; Huang, Z.; Zhu, L.; Chen, J.; Lu, Y. Heavy Metal Contamination of Soil and Water in the Vicinity of an Abandoned E-Waste Recycling Site: Implications for Dissemination of Heavy Metals. *Sci. Total Environ.* **2015**, *506–507*, 217–225.

(143) Ashiq, H. L.; Eugenia, P. L.; Sasya, T.; S, A. G.; Mohammad, S. W.; Ani, K.; Sajad, H. W.; Fayaz, A. W. Accumulation of Heavy Metals on Soil and Vegetable Crops Grown on Sewage and Tube Well Water Irrigation. *Sci. Res. Essays* **2013**, *8* (44), 2187–2193.

(144) Alvarez-Puebla, R. A.; Liz-Marzán, L. M. SERS Detection of Small Inorganic Molecules and Ions. *Angew. Chemie Int. Ed.* **2012**, *51* (45), 11214–11223.

(145) Dugandžić, V.; Kupfer, S.; Jahn, M.; Henkel, T.; Weber, K.; Cialla-May, D.; Popp, J. A SERS-Based Molecular Sensor for Selective Detection and Quantification of Copper(II) Ions. *Sensors Actuators, B Chem.* **2019**, 279, 230–237.

(146) Tsoutsi, D.; Guerrini, L.; Hermida-Ramon, J. M.; Giannini, V.; Liz-Marzán, L. M.; Wei, A.; Alvarez-Puebla, R. A. Simultaneous SERS Detection of Copper and Cobalt at Ultratrace Levels. *Nanoscale* **2013**, 5 (13), 5841–5846.

(147) Kim, K.; Lee, J. W.; Shin, K. S. Cyanide SERS as a Platform for Detection of Volatile Organic Compounds and Hazardous Transition Metal Ions. *Analyst* **2013**, *138* (10), 2988–2994.

(148) Docherty, J.; Mabbott, S.; Smith, E.; Faulds, K.; Davidson, C.; Reglinski, J.; Graham, D. Detection of Potentially Toxic Metals by SERS Using Salen Complexes. *Analyst* **2016**, *141* (20), 5857–5863.

(149) Du, J.; Jing, C. Preparation of Fe3O4@Ag SERS Substrate and Its Application in Environmental Cr(VI) Analysis. *J. Colloid Interface Sci.* **2011**, 358 (1), 54–61.

(150) Makam, P.; Shilpa, R.; Kandjani, A. E.; Periasamy, S. R.; Sabri, Y. M.; Madhu, C.; Bhargava, S. K.; Govindaraju, T. SERS and Fluorescence-Based Ultrasensitive Detection of Mercury in Water. *Biosens. Bioelectron.* **2018**, *100*, 556–564.

(151) Li, J.; Chen, L.; Lou, T.; Wang, Y. Highly Sensitive SERS Detection of As3+ Ions in Aqueous Media Using Glutathione Functionalized Silver Nanoparticles. *ACS Appl. Mater. Interfaces* **2011**, 3 (10), 3936–3941.

(152) Socrates, G. Infrared and Raman Characteristic Group Frequencies: Tables and Charts; John Wiley & Sons, 2004.

(153) Zeng, Y.; Ren, J.; Shen, A.; Hu, J. Field and Pretreatment-Free Detection of Heavy-Metal Ions in Organic Polluted Water through an Alkyne-Coded SERS Test Kit. ACS Appl. Mater. Interfaces **2016**, 8 (41), 27772–27778.

(154) Fang, G.; Hasi, W.; Sha, X.; Cao, G.; Han, S.; Wu, J.; Lin, X.; Bao, Z. Interfacial Self-Assembly of Surfactant-Free Au Nanoparticles as a Clean Surface-Enhanced Raman Scattering Substrate for Quantitative Detection of As5+ in Combination with Convolutional Neural Networks. J. Phys. Chem. Lett. **2023**, 14 (32), 7290–7298.

(155) Bradl, H. Heavy Metals in the Environment: Origin, Interaction and Remediation; Elsevier, 2005.

(156) Wei, H.; Huang, Y.; Santiago, P. J.; Labachyan, K. E.; Ronaghi, S.; Banda Magana, M. P.; Huang, Y.-H.; Jiang, S. C.; Hochbaum, A. I.; Ragan, R. Decoding the Metabolic Response of Escherichia Coli for

Sensing Trace Heavy Metals in Water. Proc. Natl. Acad. Sci. U. S. A. 2023, DOI: 10.1073/pnas.2210061120.

(157) Xu, G.; Cheng, H.; Jones, R.; Feng, Y.; Gong, K.; Li, K.; Fang, X.; Tahir, M. A.; Valev, V. K.; Zhang, L. Surface-Enhanced Raman Spectroscopy Facilitates the Detection of Microplastics < 1 Mm in the Environment. *Environ. Sci. Technol.* **2020**, *54* (24), 15594–15603.

(158) Tong, H.; Jiang, Q.; Hu, X.; Zhong, X. Occurrence and Identification of Microplastics in Tap Water from China. *Chemosphere* **2020**, *252*, No. 126493.

(159) Klein, S.; Worch, E.; Knepper, T. P. Occurrence and Spatial Distribution of Microplastics in River Shore Sediments of the Rhine-Main Area in Germany. *Environ. Sci. Technol.* **2015**, *49* (10), 6070–6076.

(160) Shan, J.; Zhang, Y.; Wang, J.; Ren, T.; Jin, M.; Wang, X. Microextraction Based on Microplastic Followed by SERS for On-Site Detection of Hydrophobic Organic Contaminants, an Indicator of

Seawater Pollution. J. Hazard. Mater. 2020, 400 (June), No. 123202. (161) Mikac, L.; Rigó, I.; Himics, L.; Tolić, A.; Ivanda, M.; Veres, M. Surface-Enhanced Raman Spectroscopy for the Detection of Microplastics. Appl. Surf. Sci. 2023, 608, 155239.

(162) Hu, R.; Zhang, K.; Wang, W.; Wei, L.; Lai, Y. Quantitative and Sensitive Analysis of Polystyrene Nanoplastics down to 50 Nm by Surface-Enhanced Raman Spectroscopy in Water. *J. Hazard. Mater.* **2022**, 429, No. 128388.

(163) Luo, Y.; Su, W.; Xu, D.; Wang, Z.; Wu, H.; Chen, B.; Wu, J. Component Identification for the SERS Spectra of Microplastics Mixture with Convolutional Neural Network. *Sci. Total Environ.* **2023**, *895* (June), No. 165138.

(164) Peng, X.; Qiao, F.; Zhao, L. Out-of-Domain Generalization from a Single Source: An Uncertainty Quantification Approach. *IEEE Trans. Pattern Anal. Mach. Intell* **2024**, *46*, 1775.