Ultra-Sensitive Detection of PFASs using Surface Enhanced Raman Scattering and Machine Learning: A Promising Approach for Environmental Analysis

Joshua C. Rothstein[1], Jiaheng Cui[2], Yanjun Yang[2], Xianyan Chen[3], and Yiping Zhao[1]

1. Department of Physics and Astronomy, University of Georgia, Athens, GA 30602, USA
2. School of Electrical and Computer Engineering, University of Georgia, Athens, GA 30602, USA
3. Department of Epidemiology & Biostatistics , University of Georgia, Athens, GA 30602, USA

Introduction

The per- and polyfluoroalkyl substances (PFASs) are a diverse class of over 3000 chemicals in use since the 1950s in industrial and consumer products. They came under regulatory and scientific scrutiny due to high bioaccumulation and persistence in the environment, as well as toxicity. Current EPA proposed standard requires a detection limit of PFAS close to 4 ppt, which imposes a significant challenge for current detection methods. Surface-enhanced-Raman scattering (SERS) spectroscopy is a very promising technology to address the challenges of PFAS detection. However, there are three challenges associated with SERS-based PFAS detection. First, high high-enhancement SERS substrates are required to provide adequately strong signals for the desired limits of detection. Second, the affinity of PFAS molecules to the designed SERS substrates shall be strong enough to demonstrate good SERS signals. Different substrates may have varying affinities with different analytes depending on their interactions. Finally, the SERS spectra from different PFAS molecules must be distinguishable. Many PFAS molecules have remarkably similar molecular structures, which can result in similar SERS or Raman spectra.

Objectives

- Show that the Raman spectra of different PFAS molecules, even with the same functional groups but different carbon chain numbers, are able to be used to differentiate the PFAS in solution.
- Integrate SERS with machine learning (ML) to differentiate and quantify various PFAS compounds in water and methanol.
- Use thiol-modified SERS substrates to improve the differentiation and quantification capabilities of the SERS-ML method.

Fabrication of AgNR substrates

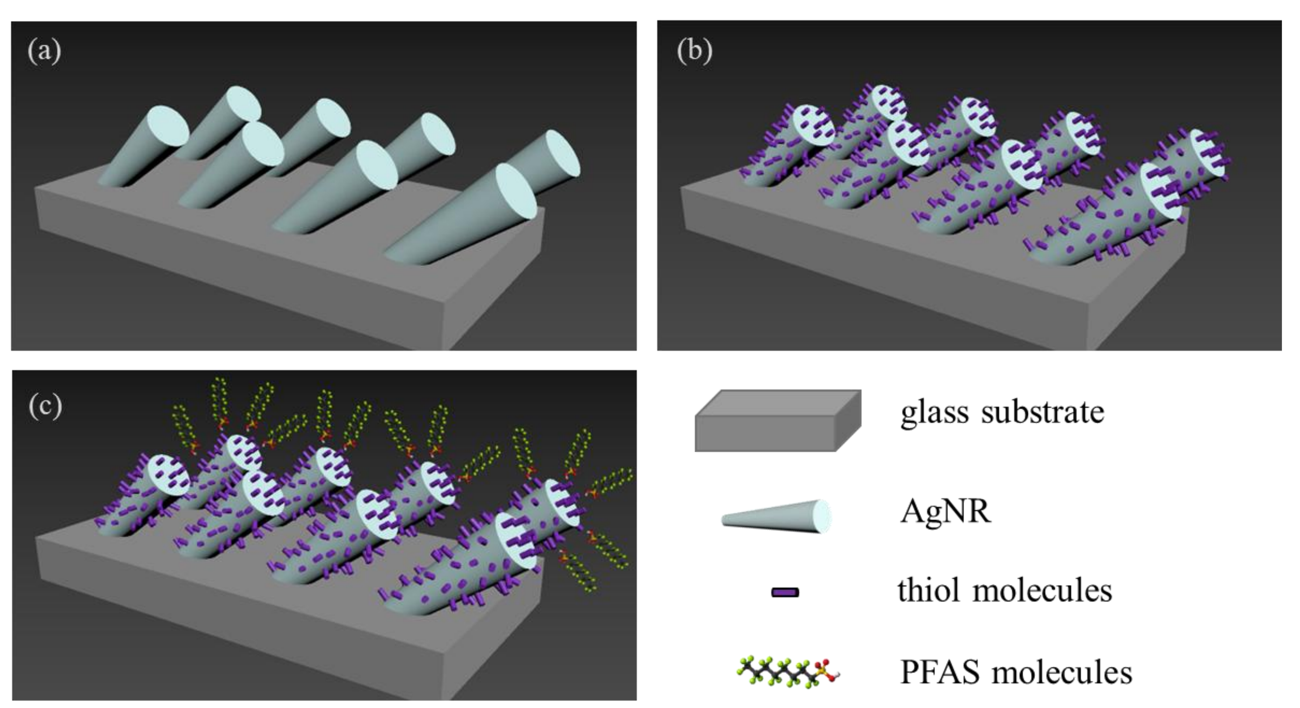

Fig. 1 (a) Illustration of Ag nanorods on a glass substrate. (b) A self-assembled monolayer of thiol molecules coats the nanorods. (c) The charged PFAS molecules are electrostatically attracted to the thiol molecule monolayer.

Glass microscope slides were cut into 0.5-inch × 0.5-inch square pieces. Followed by a standard cleaning procedure. Before SERS measurements, the AgNR substrates were cleaned by argon plasma for 90 seconds. **Figure 1a** illustrates the geometry of the AgNR array SERS substrate.

To improve the affinity of PFAS molecules to AgNR substrate, it may be advantageous to put a self-assembled thiol molecule layer on the surface to change the surface charge as shown in **Figure 1b**.

It is envisioned that when PFAS molecules are dispensed on the thiol-modified surface, they will be aligned in a certain way according to electrostatic interaction, as shown in **Figure 1c**.

SERS spectra of PFASs on AgNR substrates

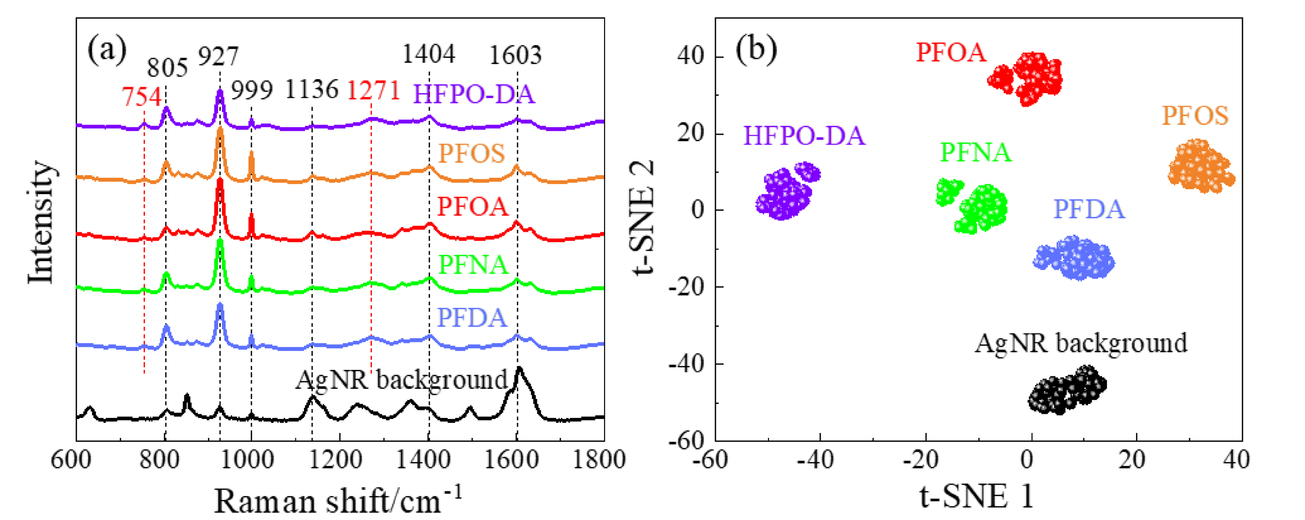

Fig. 2. (a) The average SERS spectra of selected PFAS solutions (in methanol) at a fixed 10^3 ppt concentration. The average SERS spectrum of AgNR substrate is shown as a reference. (b) The t-SNE plot of the SERS.

Figure 2a shows the average SERS spectra of 10^3 ppt PFOS, PFOA, PFNA, PFDA, and HFPO-DA in methanol.

t-SNE analysis was implemented is shown in **Figure 2b.** The SERS spectra can form well-separated clusters, which demonstrates a clear differentiation capability of SERS.

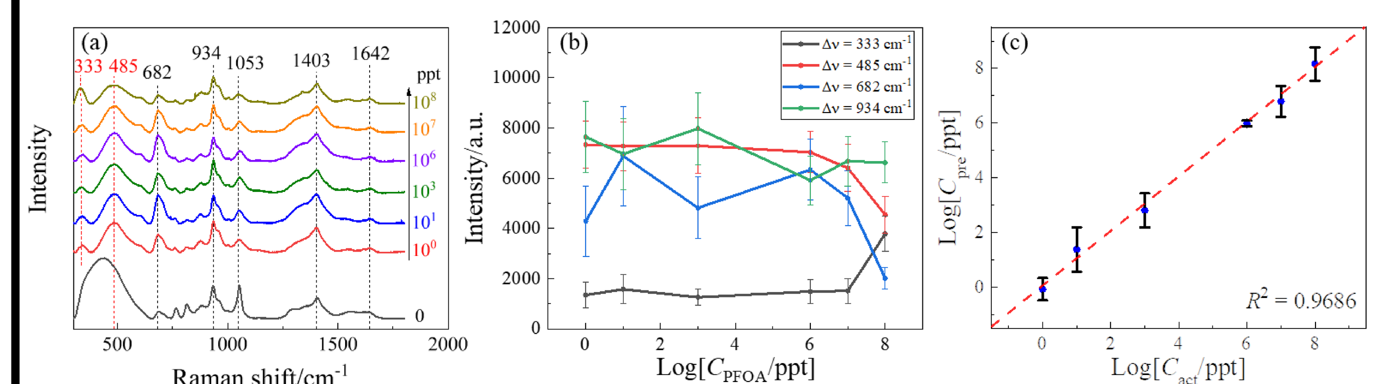

Fig. 3 (a) The representative average SERS spectra of PFOA in methanol at the concentrations of 10^0, 10^1, 10^3, 10^6, 10^7, and 10^8 ppt, respectively. (b) The semi-log plot of the peak intensities at $\Delta\nu$= 333, 485, 682, 934 cm^{-1} versus PFOA concentration C_{PFOA}. (c) The log-log plot of C_{pre} versus C_{act} of PFOA via an optimal SVR model.

In order to demonstrate the quantification capability of SERS, concentration-dependent SERS spectra of PFOA have been measured, shown in **Figure 3a**.

Traditional method (**Figure 3b**): plot the SERS intensities of characteristic peaks as a function of concentration. The relationship exhibits significant variations, making it difficult to differentiate the concentration based on these peaks.

To circumvent this problem, we can apply ML-based regression models. A support vector regression (SVR) model was used to predict the concentration. **Figure 3c** shows a log-log plot of the predicted concentration versus the actual concentration. The average R^2 value resulting from 10 independent trials was 0.95 ± 0.01, denoting an excellent fit.

SERS spectra of PFASs on cysteine-functionalized substrates

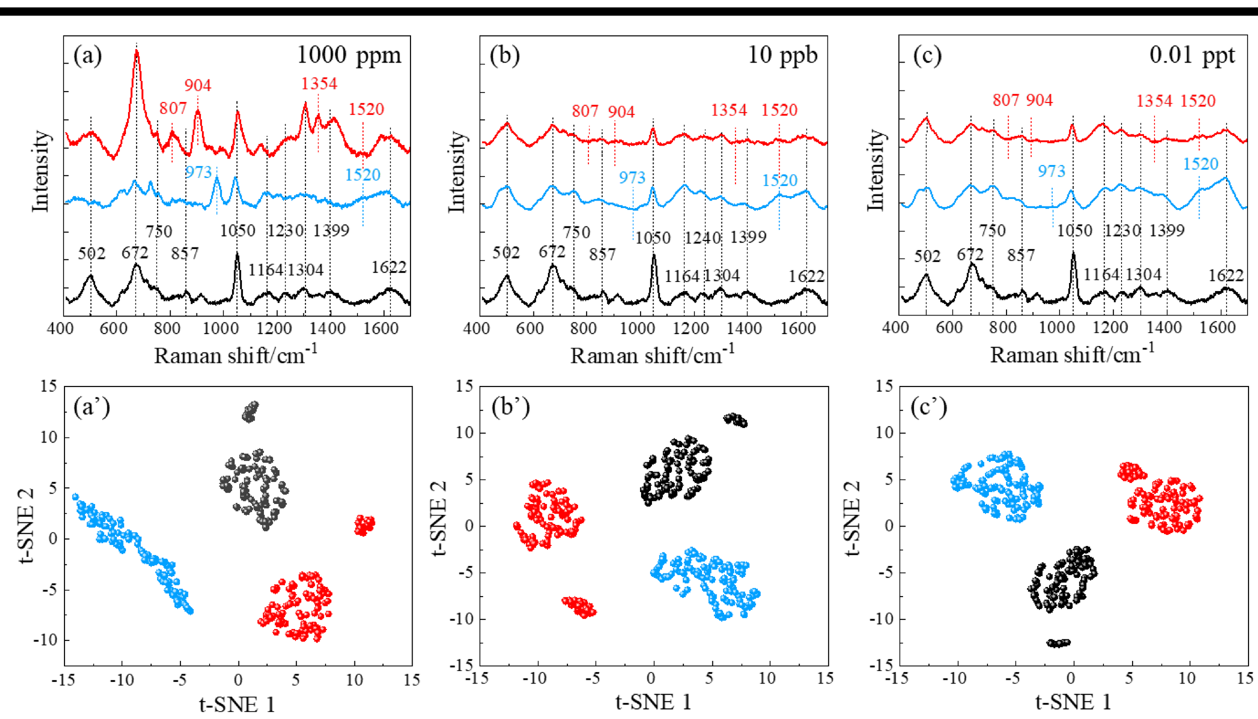

Fig. 4. The SERS spectra of (a) concentrations of 1000 ppm (10^9 ppt), (b) 10 ppb (10^4 ppt), and (c) 0.01 ppt of PFOA and PFOS on cysteine-modified AgNR substrates and the corresponding (a', b', c') t-SNE plots. In all figures, the colors red, blue, and black represent PFOA, PFOS, and water on cysteine-modified AgNR substrates.

SERS spectra of PFASs on MCH-functionalized substrates

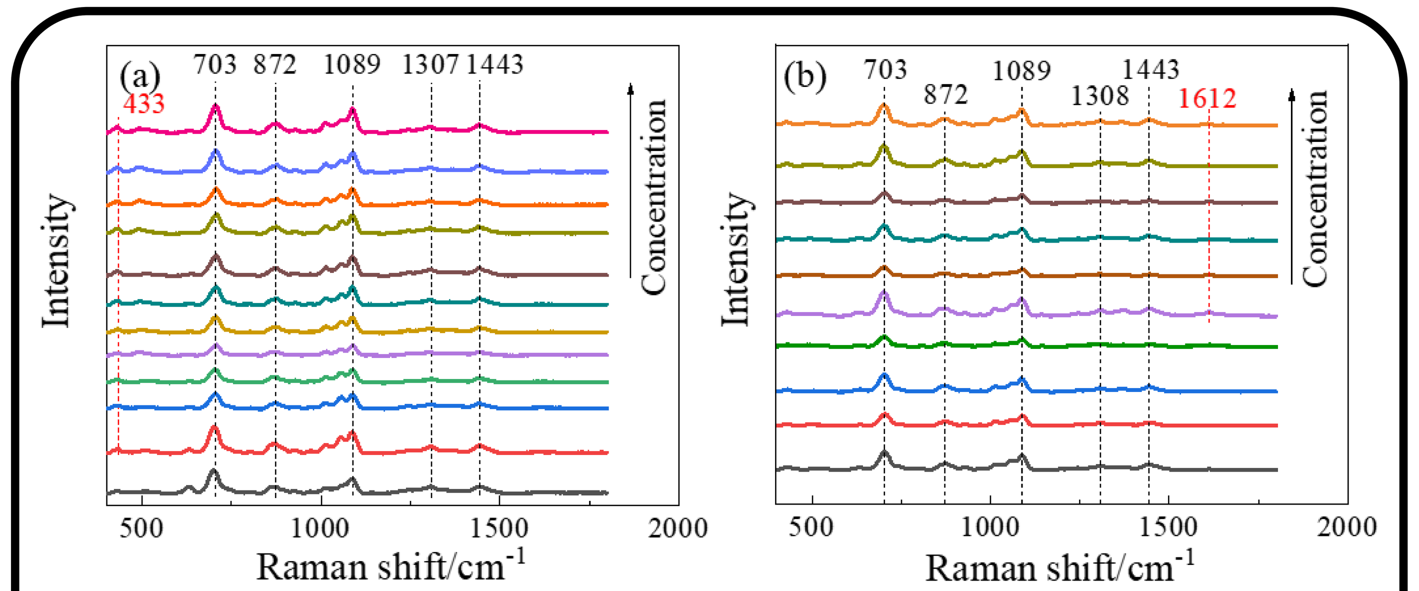

Fig. 5. The average SERS spectra from MCH-modified AgNR substrates: (a) PFOA of concentrations of 0, 10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, and 10^9 ppt (from bottom to top); and (b) PFOS with concentrations of 0, 4.28×10^{-2}, 4.28×10^{-1}, 4.28, 4.28×10^1, 4.28×10^2, 4.28×10^3, 4.28×10^4, 4.28×10^5, 4.28×10^6 ppt (from bottom to top).

MCH is a short-chain Alkanethiol molecule with an end group of -OH. When the AgNR surface is functionalized with MCH, the surface will be negatively charged.

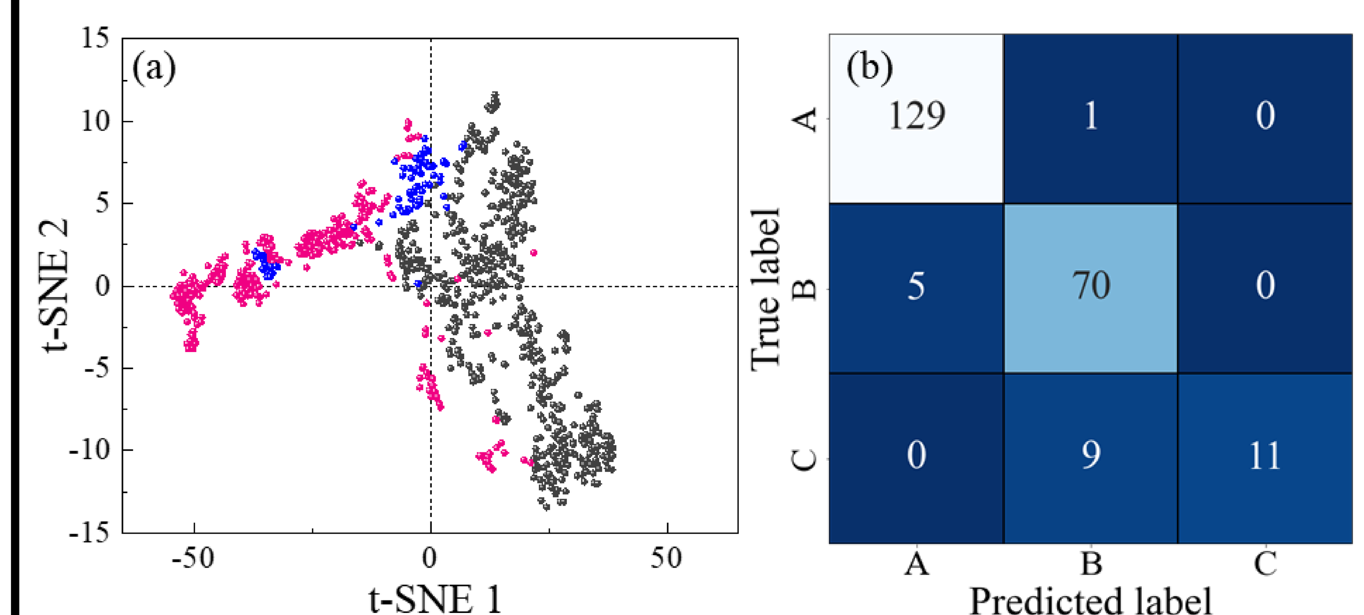

Fig. 6. (a) t-SNE result for PFOS (pink), PFOA (grey), and reference (blue) samples, including spectra from all concentrations. (b) The SVM classification confusion matrix for PFOS, PFOA, and reference samples. A: different concentrations of PFOA; B: different concentrations of PFOS; and C: MCH-modified AgNR substrates.

To demonstrate the capabilities of ML models, a more powerful SVM model with an RBF kernel was employed. Ten independent trials were conducted, resulting in an accuracy of 0.89 ± 0.01. The trial with the highest accuracy (0.93) is demonstrated by the confusion matrix shown in **Figure 6b**.

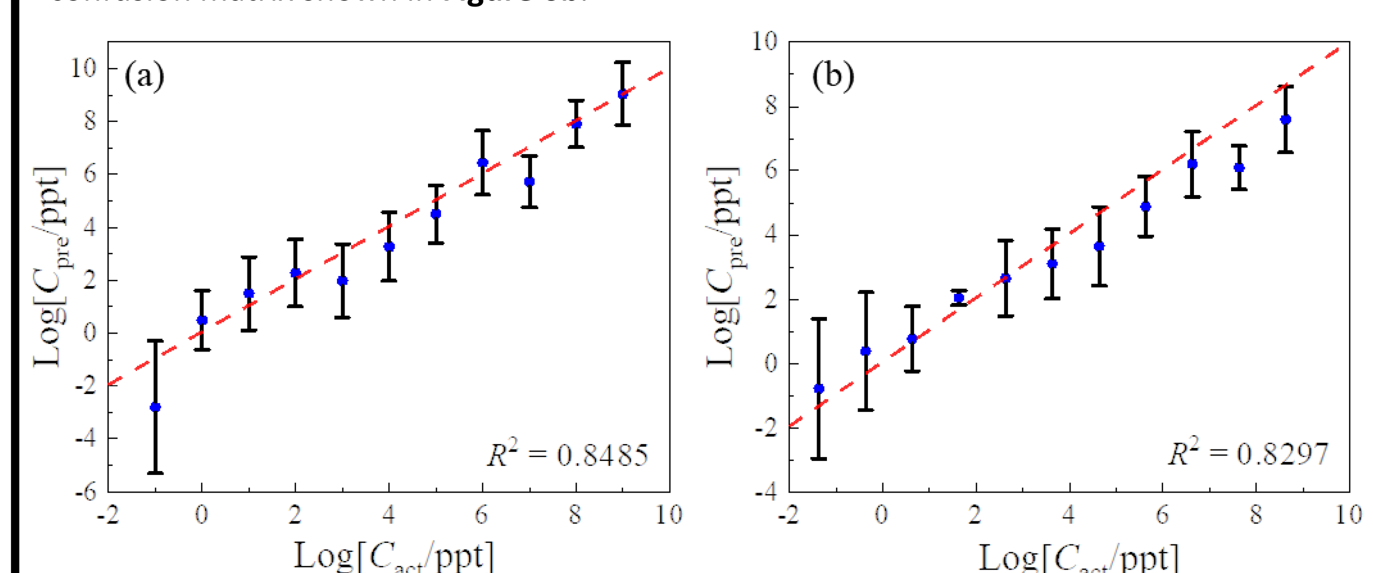

Fig. 7. The log-log plot of C_{pre} versus C_{act} of (a) PFOA and (b) PFOS via two independent SVR models, respectively.

Two separate SVR models were built to quantify concentration-dependent PFOA and PFOS spectra. Ten independent trials were performed for each SVR. The R^2 for PFOS and PFOA are 0.76 ± 0.04 and 0.82 ± 0.01, the best trials are plotted in **Figures 7a and 7b**. By employing one-sample t-tests, the SVR model for PFOA achieved an LOD of 1 ppt, with a p-value of 0.37, while the LOD for PFOS was determined as 4.28 ppt, with a p-value of 0.10.

Results after removing low concentrations

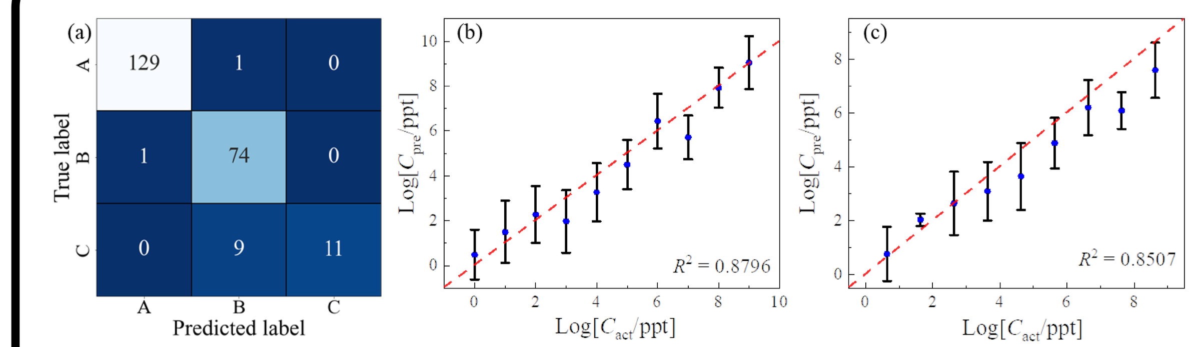

Fig. 8. (a) The SVM classification confusion matrix for PFOS, PFOA, and reference samples after removing the low concentrations. A: different concentrations of PFOA; B: different concentrations of PFOS; and C: MCH-modified AgNR substrates. (b) and (c) the log-log plot of C_{pre} versus C_{act} of PFOA and PFOS via two independent SVR models after removing the low concentrations, respectively.

As illustrated in **Figure 8a**, by excluding concentrations below the LOD, 3 of the 5 misclassified PFOS spectra were eliminated. The overall model accuracy is lifted from 0.93 to 0.95.

The R^2 scores for PFOA and PFOS by the regression models would experience only marginal improvements, rising from 0.85 to 0.88 and from 0.83 to 0.85, as shown in **Figures 8b and 8c**.

Stability of MCH-functionalized AgNR substrates

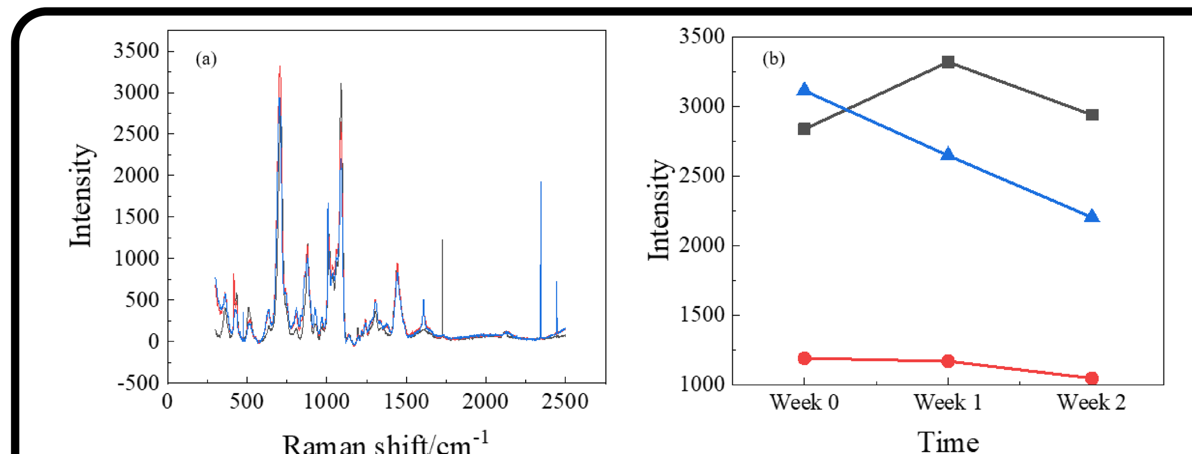

Fig. 9. (a) The averaged spectra of AgNR substrates with an MCH monolayer and PFOA applied. Measurements were taken on the day of creation (black), 1 week after creation (red), and 2 weeks after creation (blue). (b) The intensity of 3 peaks over 2 weeks: $\Delta\nu$= 710 cm^{-1} (black), 878 cm^{-1} (red), and 1089 cm^{-1} (blue).

Conclusions

Our study demonstrates the effectiveness of utilizing SERS spectra and machine learning techniques for the detection and differentiation of various PFAS. We have successfully differentiated and quantified the amounts of PFOA in water, enabling highly sensitive PFAS detection.

To further improve differentiation and quantification, we employed various Alkanethiol molecules to modify the AgNR substrates. Although the spectral features were dominated by the Alkanethiol molecules, discernible changes due to the presence of different concentrations of PFAS and PFOS molecules were observed.

Leveraging an SVM model, we achieved an average accuracy of 93% accuracy in differentiating PFOA, PFOS, and MCH. Once the spectra were accurately classified, further quantification was achieved through an SVR model, capable of predicting concentrations as low as 1 ppt for PFOA and 4.28 ppt for PFOS.

Acknowledgement

Joshua C. Rothstein, Yanjun Yang, and Yiping Zhao were partially supported by the National Science Foundation under the contract #ECCS-1808271. Jiaheng Cui, Xianyan Chen, and Yiping Zhao are funded by USDA NIFA Grant number 2023-67015-39237.