npj | clean water

www.nature.com/npjcleanwater

ARTICLE OPEN

Check for updates

Machine learning assisted dual-functional nanophotonic sensor for organic pollutant detection and degradation in water

Junhu Zhou[1], Ziqian Wu[1], Congran Jin[1] and John X. J. Zhang[1]✉

This study presents a dual-functional thin film, known as Ag nanoparticles decorated, ZnO nanorods coated silica nanofibers (AgNP-ZnONR-SNF), which demonstrates remarkable capabilities in both water purification and organic pollutants sensing. The 3D fibrous structure of ZnONR-SNF provides a large surface-area-to-volume ratio for piezo- and photo-catalytic degradation of organic pollutants under UV irradiation, achieving over 98% efficiency. Ag nanoparticles decorated on ZnONR-SNF form "hot-spot" that significantly enhance the surface-enhanced Raman spectroscopy (SERS) signal, resulting in an enhancement factor of 1056 and an experimental detection limit of 1 pg mL^{-1}. Furthermore, a machine learning algorithm is developed for the qualitative and quantitative detection of multiple contaminants, achieving high accuracy (92.3%) and specificity (89.3%) without the need for preliminary processing of Raman spectra. This work provides a promising nanoengineering solution for water purification and sensing with improved detection accuracy, purification efficiency, and cost-effectiveness.

npj Clean Water (2024)7:3 ; https://doi.org/10.1038/s41545-023-00292-4

INTRODUCTION

Water is an essential resource for all living organisms, and access to clean water is crucial for maintaining good health. Regrettably, recent data from the World Health Organization (WHO) reveals that 2 billion people lack access to safely managed water service as of 2020, including 282 million people with limited services and 122 million drinking surface water[1]. Water pollution has become a serious threat to water safety due to economic development, global population increase, and climate changes. Organic pollutants, including pharmaceuticals, pesticides, organic dyes, detergents, and industrial waste, are of particular concern because they consume oxygen in water, generate toxic residues, and can persist in the water body for a long time due to their chemical stability[2].

Traditional water purification approaches encompass chemical precipitation, filtration, adsorption, ion exchange, chlorination, adsorption, and distillation[3–6]. Although these techniques are widely used in industrial settings, they have limitations such as lengthy processing times, limited removal efficiencies, chemical resistance issues, and risks of re-contamination.

Zinc oxide (ZnO), with its exceptional photo- and piezo-catalytic properties, presents an eco-friendly material for processing contaminated water[7,8]. One major advantage of using ZnO in water treatment is its ability to react with persistent organic contaminants under natural sunlight, which contains UV light. Hence, it offers a more sustainable and eco-friendly approach for removing organic pollutants than traditional methods[9–11]. However, to fully leverage ZnO's properties for wastewater treatment, a well-designed material and system must be in place. ZnO can be fabricated into various nanostructures, such as nanospheres[12], nanorods[13], and nanoflowers[14], but there is still potential for improvement in degradation efficiency, secondary contamination prevention, and repeatability. To tackle these issues, ZnO nanoparticles can be anchored on silica substrates which have robust mechanical strength and chemical stability[15]. Growing ZnO on silica nanofibers (SNFs) can increase the contact area, prevent nano pollution, and enable reusability.

Monitoring water quality is crucial for environmental and human health on both large and small scales. However, detecting organic pollutants in water is challenging due to molecular variability and complexity[16]. Additionally, the necessity for achieving low limits of detection further complicates the monitoring process. Technologies such as photo-luminescence spectroscopy[17], high-resolution mass spectrometry[18], high-performance liquid chromatography[19], and surface-enhanced Raman spectroscopy (SERS)[20] have been used to detect water contaminants. SERS is extensively researched due to its label-free, ultrasensitive, fast, and versatile characteristics. Nevertheless, identifying complex Raman spectra presents challenges, including mixed features, complex datasets, instrument noise interference, and sample property effects. To extract meaningful information from the Raman spectrum, various machine learning (ML) algorithms have been developed, including partial least square (PLS) regression[21,22], support vector machine[23], convolutional neural network[24], recurrent neural network[25] and Deep learning (DL)[26,27]. DL is a type of ML algorithm that employs multiple hidden layers of neural networks to extract complex features from large and diverse datasets. These deep neural networks can efficiently learn and represent nonlinear relationships between input and output data without domain-specific knowledge, allowing accurate classification of Raman spectra for corresponding chemical compositions. However, detecting "out-of-distribution" (OOD) samples (i.e., class of unseen data points), which are significantly different from the training data, remains a common challenge in DL. Despite this challenge, OOD detection is vital for ensuring the reliability of ML algorithms for SERS detection, particularly when training data is limited, and diverse chemicals may be present in the test samples. Current ML-assisted Raman detection techniques also struggle to discern mixtures of molecules and accurately characterize complex mixtures.

This study aims to develop a water purification and detection material by incorporating SNF thin film with ZnO nanorods (ZnONRs) and silver nanoparticles (AgNPs). This innovative material system can efficiently degrade organic pollutants while

[1]Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA. ✉email: John.Zhang@Dartmouth.edu

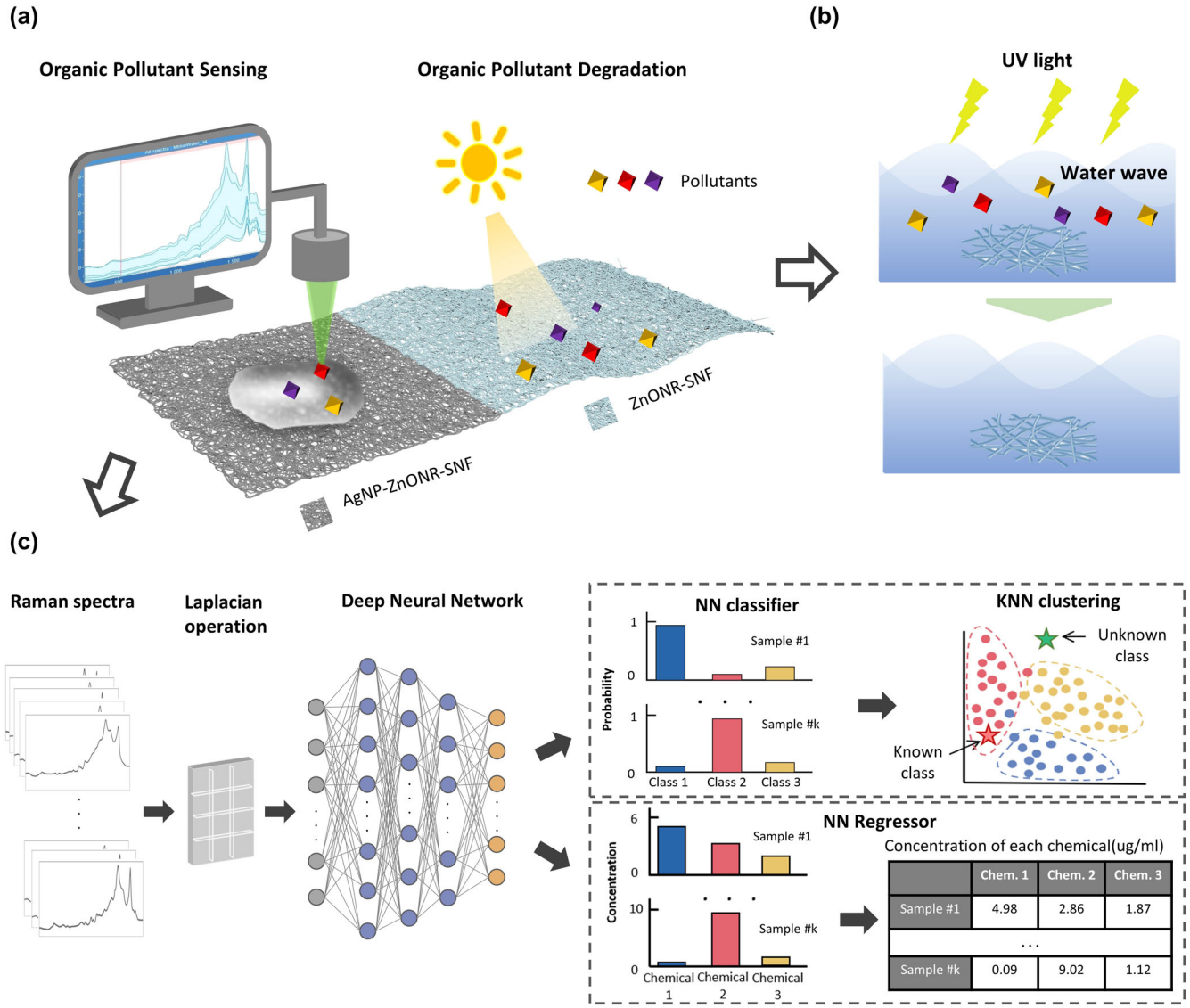


Fig. 1 The schematic diagram of the pollutant degradation, detection, and ML algorithm-assisted data interpretation process of the dual functional device. a Components of the AgNP-ZnONR-SNF system for organic pollutant detection and degradation. **b** The diagrammatic sketch of the photo- and piezo-catalytic degradation process. **c** The schematic demonstration of the ML models.

enabling rapid, quantitative, and label-free detection of these contaminants through SERS (Fig. 1a, b). A DL algorithm is developed to facilitate the qualitative, quantitative, and OOD detection of mixed contaminants without the need of preliminary processing. A Laplacian operation is implemented to effectively extract the Raman peak information from the original Raman spectra. The neural network's output can encompass both regression model, indicating the concentration of multiple analytes, and classification model for identifying analyte components and concentration levels, with an added K-nearest neighbor (KNN) model for detecting unknown classes (Fig. 1c).

RESULTS AND DISCUSSION

Characterization

The electrospun SNF thin film was collected from the aluminum foil and had a diameter of ~12 cm (Supplementary Fig. 1). The Scanning electron microscopy (SEM) analysis was conducted using an FEI Helios 5CX DualBeam scanning electron microscope operating at 5 kV. The energy-dispersive X-ray spectroscopic (EDS) measurements and the chemical mapping were performed with the Oxford Instruments Ultim Max detector attached to the FEI Helios 5CX scanning electron microscope operating at 30 kV. Figure 2a depicts the sequential process of synthesizing electrospun SNF, growing ZnONR, and decorating with AgNP. Figure 2b–e are SEM images demonstrating the structure of SNF, ZnONR-SNF and AgNP-ZnONR-SNF. The AgNPs were distributed three-dimensionally, with some on the top of the ZnONRs and others on their side walls, while pure ZnONRs had a smooth surface (Supplementary Fig. 2). In Fig. 2f-i, the EDS chemical mapping confirms the uniform distribution of Zn, O, and Ag elements in the sample. The average diameter of the SNF, AgNP, and ZnO-SNF are 868 nm, 40 nm, and 3 µm, respectively (Fig. 2j). The characteristic X-ray energy of the key elements, including Zn, Si, O, and Ag, are plotted and labeled in the energy spectrum Fig. 2k. The Ag element has its characteristic energies at 2.98 keV and 3.15 keV, the Zn element at 1.01 keV, the O element at 0.52 keV, and the Si element at 1.74 keV. Additionally, Optical images taken by Keyence 3D Microscope with high magnification reveal the fibrous and 3D structure of the AgNP-ZnONR-SNF thin film (Supplementary Fig. 3). Figure 2l exhibits the normalized UV–Visible optical

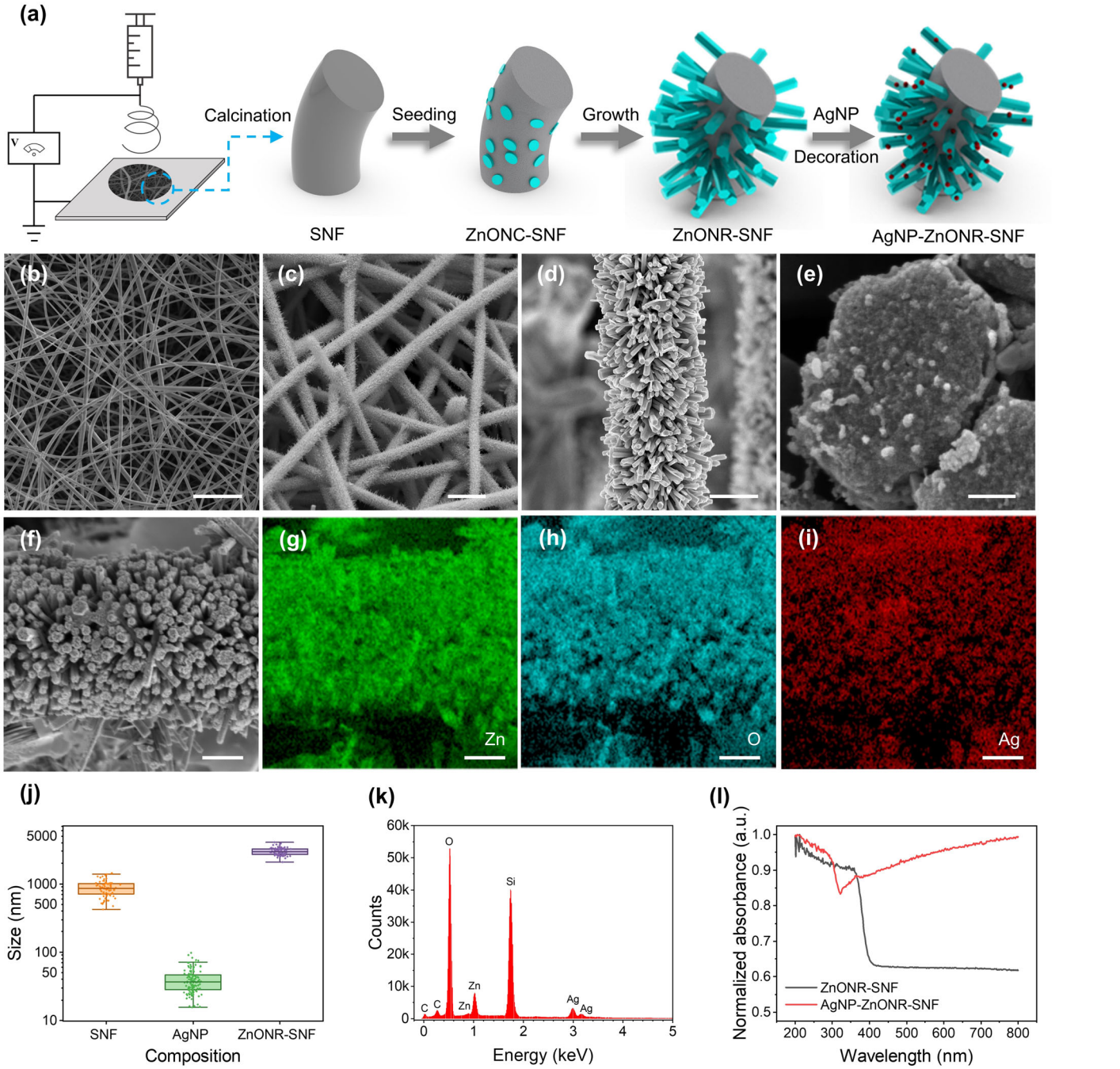


Fig. 2 Fabrication and characterization of the AgNP-ZnONR-SNF thin film. a Schematic diagram describes the synthesis procedure of the ZnONR and AgNP on electrospun SNF thin film. **b** SEM image of the SNF. **c**–**d** SEM images of the ZnONR-SNF. **e** SEM image of the AgNPs on the top of the ZnONR. Scale bars are 20 μm, 10 μm, 2 μm, and 100 nm, respectively. EDS mapping of the AgNP-ZnO-SNF, including **f** SEM image, **g** Zn element), **h** O element, and **i** Ag element. Scale bars are 3 μm. **j** Diameter distribution of the SNF, AgNP, and ZnONR-SNF in semi-logarithmic plot. Within each box, horizontal line denotes median value; boxes extend from the 25th to the 75th percentile of each group's distribution of values; vertical extending lines denote the most extreme values within 1.5 interquartile range of the 25th and 75th percentile of each group; dots denote observations. **k** EDS spectrum of the AgNP-ZnO-SNF. **l** Normalized UV–Visible optical absorption spectrum of the ZnONR-SNF and AgNP-ZnO-SNF.

absorption spectrum of the ZnONR-SNF and AgNP-ZnO-SNF thin films.

Piezo- and photo-catalytic degradation

The photocatalytic activities of ZnONR-SNF thin film in degrading organic dyes, including Methylene Blue (MB), Trypan Blue (TB) and Methyl Orange (MO), at an initial concentration of 10 μg/mL are investigated. The experimental setups are shown in Supplementary Fig. 4. Experimental groups involve ZnONR-SNF material and organic dye solutions, whereas control groups solely contain organic dye solutions, all subjected to UV irradiation and shaking treatment. Control experiments demonstrate that solar UV irradiation and shaking without any photocatalyst is negligible, and the concentration of the dyes remains almost unchanged. When ZnONR-SNF thin films are present, a degradation efficiency greater than 98% is

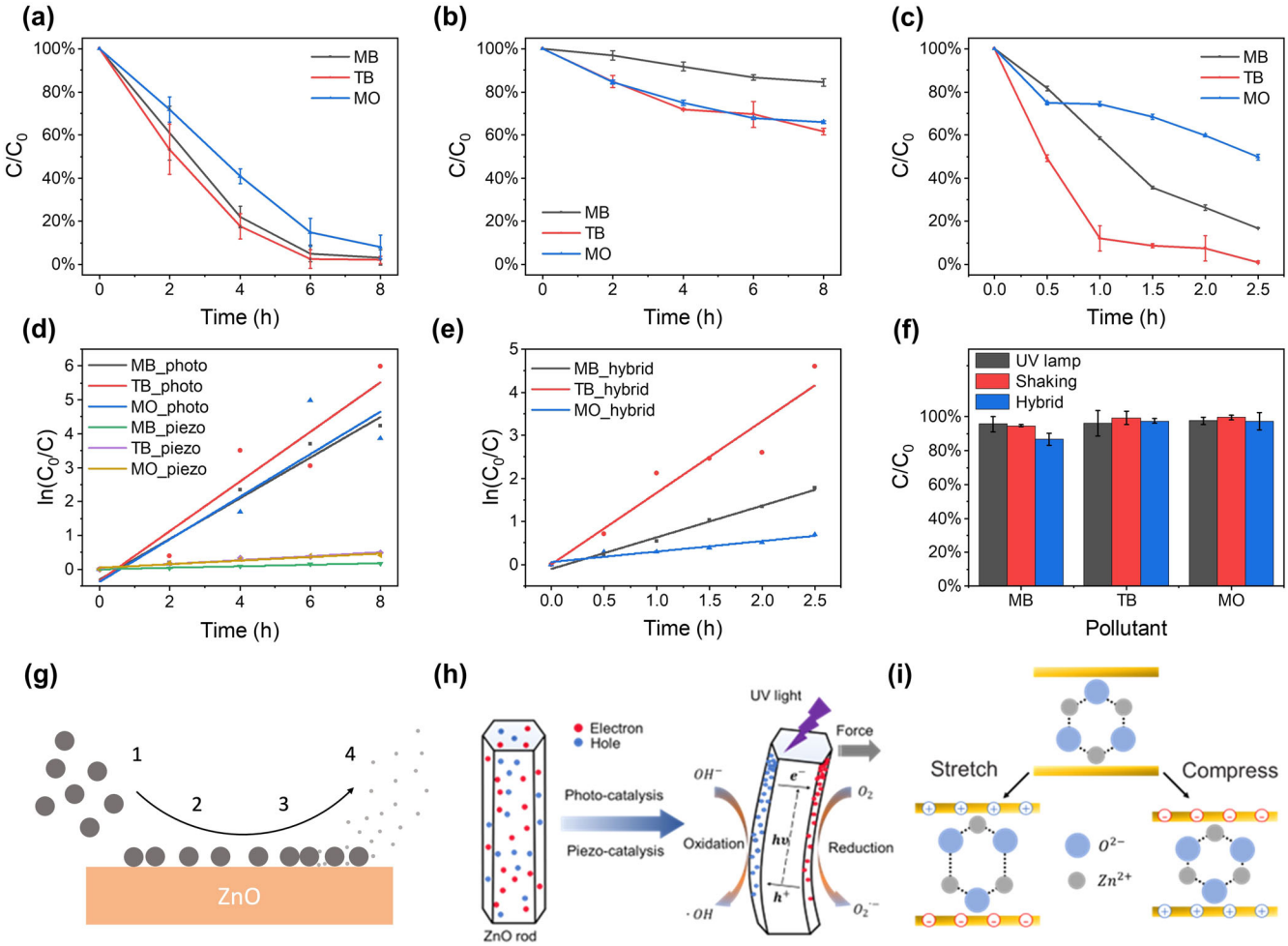


Fig. 3 Degradation results and schematic illustration of the degradation mechanism. a Photocatalytic degradation results. **b** Piezo-catalytic degradation results. **c** Hybrid degradation results. **d** Photo- and piezo-catalytic degradation kinetic curves of dye solutions catalyzed by the ZnONR-SNF. **e** Hybrid degradation kinetic curves of dye solutions catalyzed by the ZnONR-SNF. **f** Degradation results of control groups without the ZnONR-SNF. Error bars represent the standard deviation. **g** Schematic illustration of the degradation process. **h** Schematic graphic showing the piezo- and photo-catalytic properties of the ZnONR. **i** Schematic diagram of the piezoelectric property of ZnO.

achieved against the three organic dyes after 8 h exposure to UV light. Figure 3a–c displays the change in concentration of these organic pollutants over time under different conditions. Notably, the decline of the light absorbance intensity of the organic dyes at their characteristic peak wavelength is due to the cleavage of chromophore groups responsible for dye decoloration. Supplementary Fig. 5 shows the separate decomposition of these organic molecules under UV lamp and sunlight. Photocatalytic degradation efficiency for samples after 2-h sunlight exposure is even higher than under the UV lamp. Figure 3d, e displays the rate constants (k) for photocatalytic degradation, confirming the excellent photocatalytic property of ZnONR-SNF thin films. The degradation outcomes of the control groups lacking the ZnONR-SNF thin films are illustrated in Fig. 3f. In these groups, only a marginal decrease in dye concentrations is observed. In addition to organic dyes, the photocatalysis degradation against antibiotic Ciprofloxacin (Cip) is also studied under both UV light and sunlight (Supplementary Fig. 6). The cost efficiency of the ZnONR-SNF thin film in photocatalytic degradation is closely tied to its reusability. In the process of degrading MB, TB, and MO under UV lamp, the identical ZnONR-SNF samples were employed and subsequently dried in an oven at 50°C for a total of six cycles. The photocatalytic efficiency of ZnONR-SNF thin films remain high for five cycles (Supplementary Fig. 7). The photocatalytic degradation against organic molecules can be explained by the physical and optical properties of ZnO (Fig. 3g-h) as follows[9,28,29]: first organic pollutants diffuse from the liquid phase and are absorbed onto the surface of ZnONR. Then ZnO is irradiated by the UV light with energy larger than its bandgap energy, promoting electrons (e^-) from valence band (VB) to conduction band (CB) and leave holes (h^+) in the VB. Next, the photogenerated electron-hole (e^-/h^+) pairs can migrate to the ZnONR surface, reacting with hydroxide ions (by h^+) and oxygen (by e^-) in water to generate reactive oxygen species (ROS) including hydroxyl radical ($\cdot OH$) and superoxide anion ($O_2^{\cdot -}$). Finally, the ROS can directly oxidize organic pollutant molecules.

The piezo-catalytic activity of ZnONR-SNF thin film on degrading organic dyes is shown in Fig. 3b. The catalyst is stimulated through orbiting shaking using a shaker motor with a rating output of 13.5 W. After 8 h of reaction, the piezo-catalytic efficiency against MB, TB, and MO reaches 15.5%, 38.4%, and 34.1%, respectively. The mechanical force exerted by the flowing water induces deformation in ZnONR, resulting in the formation of a strain field. As a consequence, the outer side of ZnONR experiences stretching, while the inner side of ZnONR undergoes compression. (Fig. 3h). This deformation leads to the generation of an electric field along the ZnONR, which induces a piezoelectric potential and surface charge accumulation on the opposite surface[30]. It allows e^-/h^+ to migrate to the ZnONR surface, triggering subsequent reactions similarly to the photocatalytic

degradation process. Additionally, the piezoelectric potential also contributes to the adsorption of charged organic molecules.

In hybrid experiments, the combination of a UV lamp and a shaker results in the degradation of 96.3% of MB, 66.4% of TB, and 34.8% of MO after 2 h of irradiation. This degradation rate surpasses the individual catalytic activities of the two components. The improved degradation efficiency can be explained by the flowing water and semiconducting nature of ZnONR. The circulating water flow created by the orbiting shaker not only deflects ZnONR to create piezoelectric potential between the opposite surface but also promotes the adsorption of organic molecules. Moreover, the relative displacement of the Zn^{2+} cations regarding the O^{2-} anions in the wurtzite crystal structure (Fig. 3i) generates a piezoelectric potential along the nanorod direction when the ZnONR is bent. This potential difference maintains as long as the strain exists and creates a potential difference between the compressed and the stretched side surface[30]. Correspondingly, the recombination of photo-induced e^-/h^+ pairs inside the ZnONR will be impeded, leading to more efficient carrier separation. Eventually, more redox reaction occurs on the surface of ZnONR, improving the photocatalytic degradation.

SERS detection

Raman spectroscopy provides a label-free and rapid tool for sensing organic molecules. A SERS approach is used in this study to quantitatively measure the concentration of an organic dye and an antibiotic. Before conducting the Raman experiments, the enhancement of localized surface plasmonic resonance was simulated using the finite element method (FEM) by COMSOL. Three scenarios of AgNP placement were observed in the SEM images: on the top side, on the side wall, and at the edge of ZnONR (Supplementary Fig. 8). When AgNPs are in close proximity and interact with incident light, they create a plasmonic "hot-spot" region characterized by a significant enhancement of the localized electromagnetic field[31]. This enhanced electromagnetic field can amplify SERS enhancement factor to the fourth power of the field, and the "hot-spot" sites created by AgNP dimers generate more than 50% of the total SERS signal with only 1% of total surface area[32,33]. Prior research has demonstrated the efficacy of utilizing AgNPs-decorated ZnO mesoporous materials as substrates for SERS-based detection[34]. Nevertheless, achieving precise control over "hot-spot" regions on mesoporous substrates is challenging, potentially leading to less consistent performance in SERS sensing. In our proposed approach, we leverage ZnONR. This multi-functional material functions as a photo- and piezo-catalyst while also providing a stable platform for AgNP decoration and enhancing uniformity in SERS sensing. Additionally, AgNP trimer structure on the top of ZnONR was found to even better enhance the field (Supplementary Fig. 9), which aligns with the experiment observation[35]. According to the simulation results, the diameter and the gap between the AgNPs affect the enhancement, while the locations have limited effect (Supplementary Fig. 10). As most of the peaks of the extinction cross-section area for all AgNP arrangements lay in the range from 475 nm to 550 nm, 532 nm laser is chosen as the excitation source.

In the Raman experiments, MB and Cip, whose chemical structures are shown in Fig. 4a, are used as demonstrations. In the test, 10 µL of the sample solution are dropped on the AgNP-ZnONR-SNF chip which has a size about 5 mm × 5 mm. Figure 4b and Fig. 4c show the raw Raman spectra of the Cip and MB samples, respectively, at concentrations ranging from 100 µg mL^{-1} to 1 pg mL^{-1}. The most distinguishable characteristic peaks of the Cip are at 1382 cm^{-1}, 1465 cm^{-1}, 1605 cm^{-1}, and 1548 cm^{-1}, and MB at 1437 cm^{-1} and 1614 cm^{-1}. It is shown that there is a general trend of concentration of the samples being proportional to the intensity of the Raman signal. Therefore, the intensity of the highest peaks of the MB (1437 cm^{-1}) and Cip (1382 cm^{-1}) dyes are plotted against their concentrations as shown in Fig. 4d and e, on a $\log_{10}$ scale. MB and Cip calibration curves show high linearity of $R^2 = 0.992$ and $R^2 = 0.946$, respectively, demonstrating the device's functionality as a sensor that can quantitatively monitor the concentration of organic dye and antibiotic. To demonstrate the superb signal enhancement from the AgNP-ZnONR-SNF chip, Raman signals on various substrates are measured, including a pristine glass substrate with no specimen, a glass substrate with a drop (10 µL) of MB, ZnONR-SNF with a drop of MB, and AgNP-ZnONR-SNF with a drop of MB solution (Fig. 4f). The intensity at the 1437 cm^{-1} peaks of MB on the AgNP-ZnONR-SNF, ZnONR-SNF, and the glass substrate are 19861.9, 62.1, and 18.8, respectively. The results demonstrate a significant increase in intensity when using the AgNP-ZnONR-SNF substrate compared to the glass substrate. Specifically, the intensity from the ZnONR-SNF substrate is 3.3 times higher than that from the glass substrate, and the intensity from the AgNP-ZnONR-SNF sensor is 1056 times higher than that from the glass substrate. The inset of Fig. 4f shows the spectrum in the 1600–1800 cm^{-1} range, which contains two other characteristic peaks of MB at 1700 cm^{-1} and 1732 cm^{-1}. The intensities of these peaks from the AgNP-ZnONR-SNF are 645.3 and 778.2, respectively, which are about 1.8 times greater than those from the glass substrate. These results indicate that the SERS enhancement mainly occurs in the range of 1050–1650 cm^{-1}, where the enhancement is up to 1056-fold. Enhancement outside this range is less than 2-fold. Several factors contribute to the significant enhancement range of nanosensor, including particle composition (e.g., Ag, Au, or other noble metals) and nanomaterial geometry (e.g., particle size). To achieve a substantial enhancement across a broader spectrum, it is advisable to employ multiple types of nanoparticles and various sizes in decorating the ZnONR. Moreover, the AgNP-ZnONR-SNF chip exhibits relatively consistent and uniform sensing performance on its surface. A color map of the signal intensity of the highest peak (1614 cm^{-1}) of MB dye is plotted (15 × 15 data points on an area of 400 × 400 μm^2) to represent the magnitude of the intensity, as shown in Fig. 4g. Microscale Raman mapping analysis reveals that the magnitude of the signal intensity is generally consistent across this testing area, indicating a uniform sensing ability across the chip. The box chart of Fig. 4h shows the distribution of the Raman signal intensity for the characteristic peak at 1614 cm^{-1} of MB.

ML-assisted detection

Raman spectra obtained by AgNP-ZnONR-SNF chip contain inherent molecular fingerprints for analyte identification (Supplementary Fig. 11a). However, accurately characterizing and explaining the molecular structure of mixtures with overlapping characteristic peaks can be challenging (Supplementary Fig. 11b). To address this issue, a combination of deep neural network is employed. The deep neural network consists of a prior treatment stage, a fully connected neural network structure and KNN to provide qualitative and quantitative detection results. While pre-processing is a standard practice in Raman detection models, the intrinsic complexity of Raman spectra and the presence of background noise frequently lead to the generation of weak signals[36,37]. To enhance the informative content, a prior treatment is implemented leveraging the knowledge of Raman spectrum characteristics. The removal of marginal data points is undertaken, and a Laplacian operation is applied to the raw Raman spectrum data, intensifying the signal gradient and unveiling pertinent Raman peak information. The strengthened data is then fed into a four-layer fully connected neural network structure with sigmoid/ReLU activation between each linear layer, where the size of each layer is 128, 64, 32, and 16. The neural network structure and training parameters are shown in Supplementary Table 1.

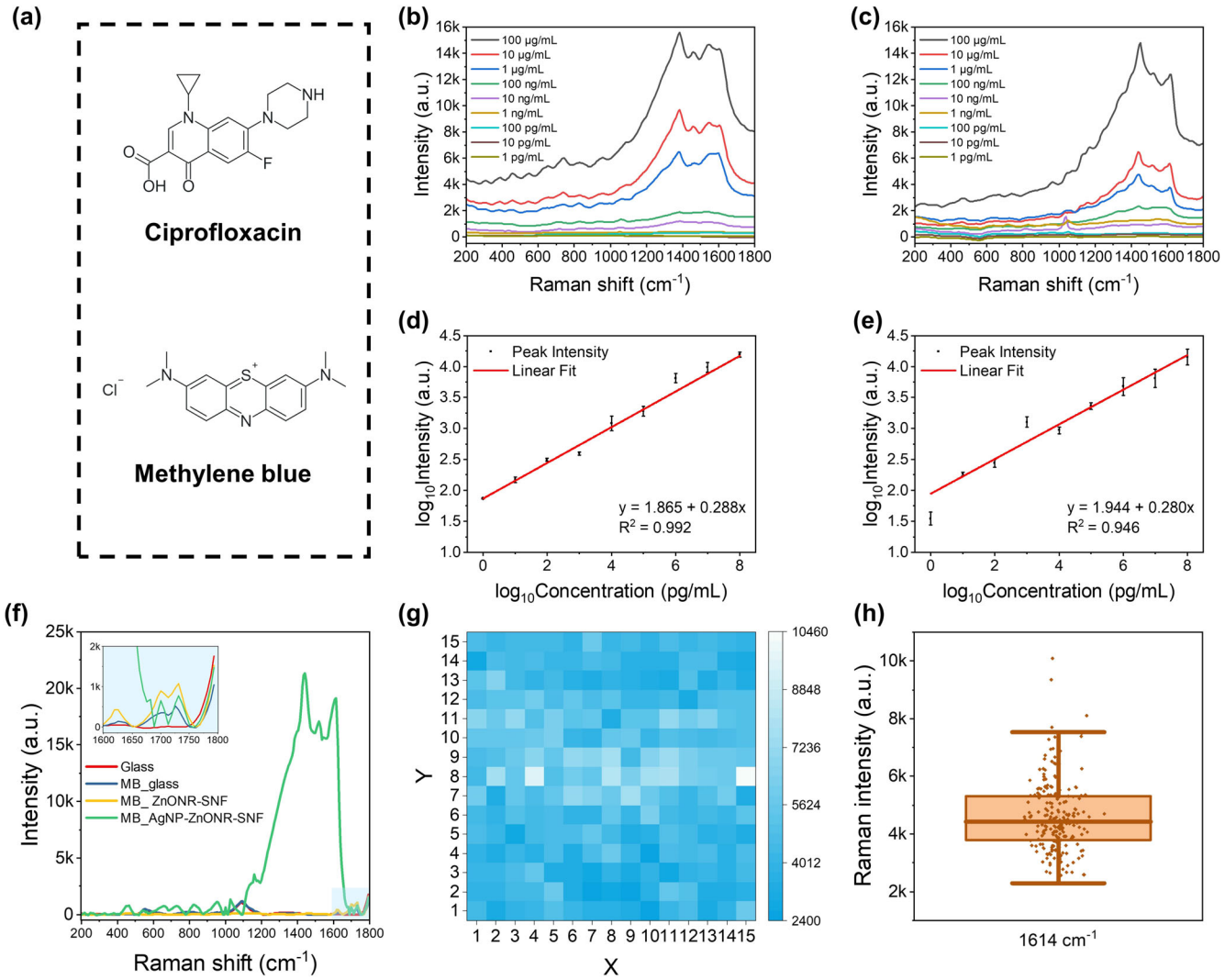


Fig. 4 SERS detection of organic molecules using the AgNP-ZnONR-SNF thin film. a Chemical structure of antibiotic Cip and dye MB. **b** Raman spectra of Cip solution of various concentrations. **c** Raman spectra of MB solution of various concentrations. **d** The Raman intensity vs. concentration curve of Cip (in $\log_{10}$ scale). **e** The Raman intensity vs. concentration curve of MB (in $\log_{10}$ scale). Error bars represent the standard deviation. **f** The SERS enhancement of MB dye (1 μg mL^{-1}) on different substrates. The inserted figure shows the zoomed area between 1600 cm^{-1} and 1800 cm^{-1}. **g** Raman signal mapping (15 × 15 data points on an area of 400 × 400 μm^2) of MB dye's 1614 cm^{-1} characteristic peak showing uniform intensity across the AgNP-ZnONR-SNF chip. Color bar represents Raman signal intensity. **h** The distribution of Raman intensity. Within the box, horizontal line denotes median value; box extend from the 25th to the 75th percentile of the group's distribution of values; vertical extending lines denote the most extreme values within 1.5 interquartile range of the 25th and 75th percentile of the group; dots denote observations.

Neural network classification

The classification is assessed using Raman spectra obtained from a mixture of two types of dyes (MB, TB, MB&TB at different ratios, and water only) across different concentrations, as detected by the AgNP-ZnONR-SNF chip. The output of the neural network is a five-digit tensor. The first four digits of the output tensor correspond to the dye type, providing information about the potential affiliation of the sample component with one of four classes: water, MB only, TB only, and the mixture of TB and MB. The last digit of the output tensor indicates whether the concentration of a specific analyte surpasses a typical cutoff threshold: [1] above the threshold and [0] indicates a smaller concentration. In this study, the concentration threshold was set as 5 μg mL^{-1}. The qualitative detection results are illustrated in Supplementary Fig. 12a, where the positions of the points indicate the predicted sample components, and the color of the points indicates their true components. The statistical results are shown in Supplementary Fig. 12b, with an accuracy of 92.3% in qualitative detection and 90.8% in quantitative detection. A comprehensive summary containing 100 tests is listed in Supplementary Table 2.

Neural network regression

The regression model undergoes additional testing with Raman spectra obtained from mixtures of three components, MB, TB, and Cip, mixed at various concentration ratios. The concentration ratios of the detected analytes are displayed in Fig. 5a, demonstrating a notable concordance with the ground truth curves. The average absolute error, computed within a concentration ratio range of 0 to 10, is 0.327, indicating a high level of agreement. The presence of the particular analyte in the sample can also be determined from the regression output. A notably low concentration ratio in the output suggests the absence of the specific analyte in the sample. For each analyte of the three, we

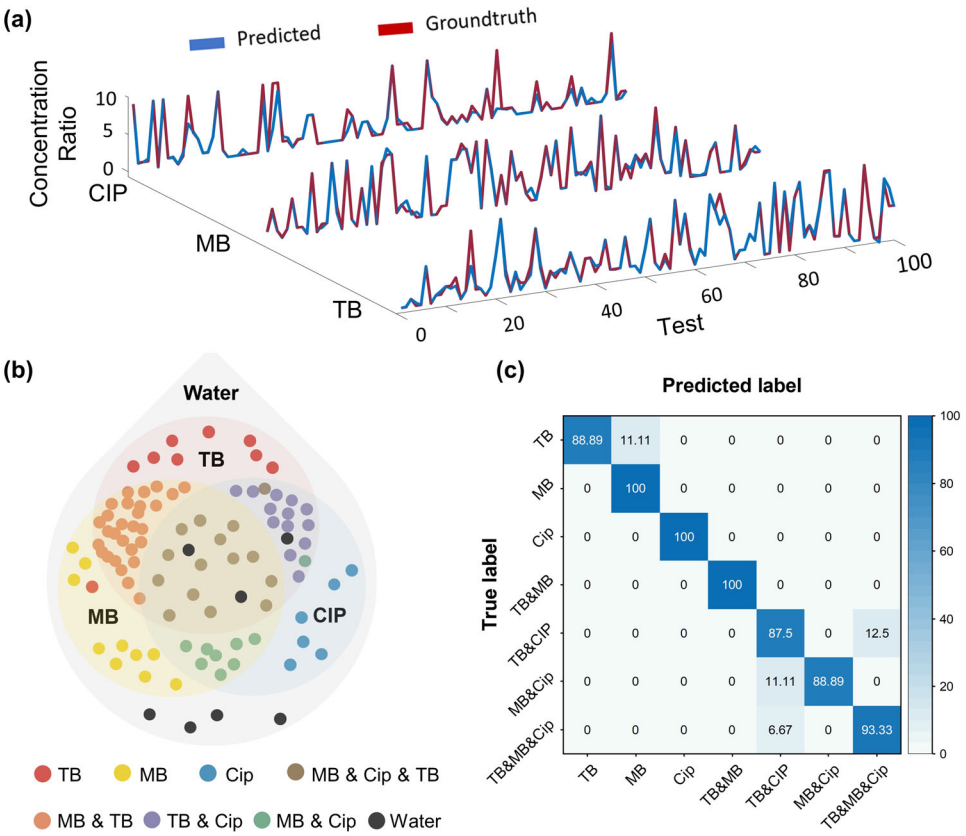


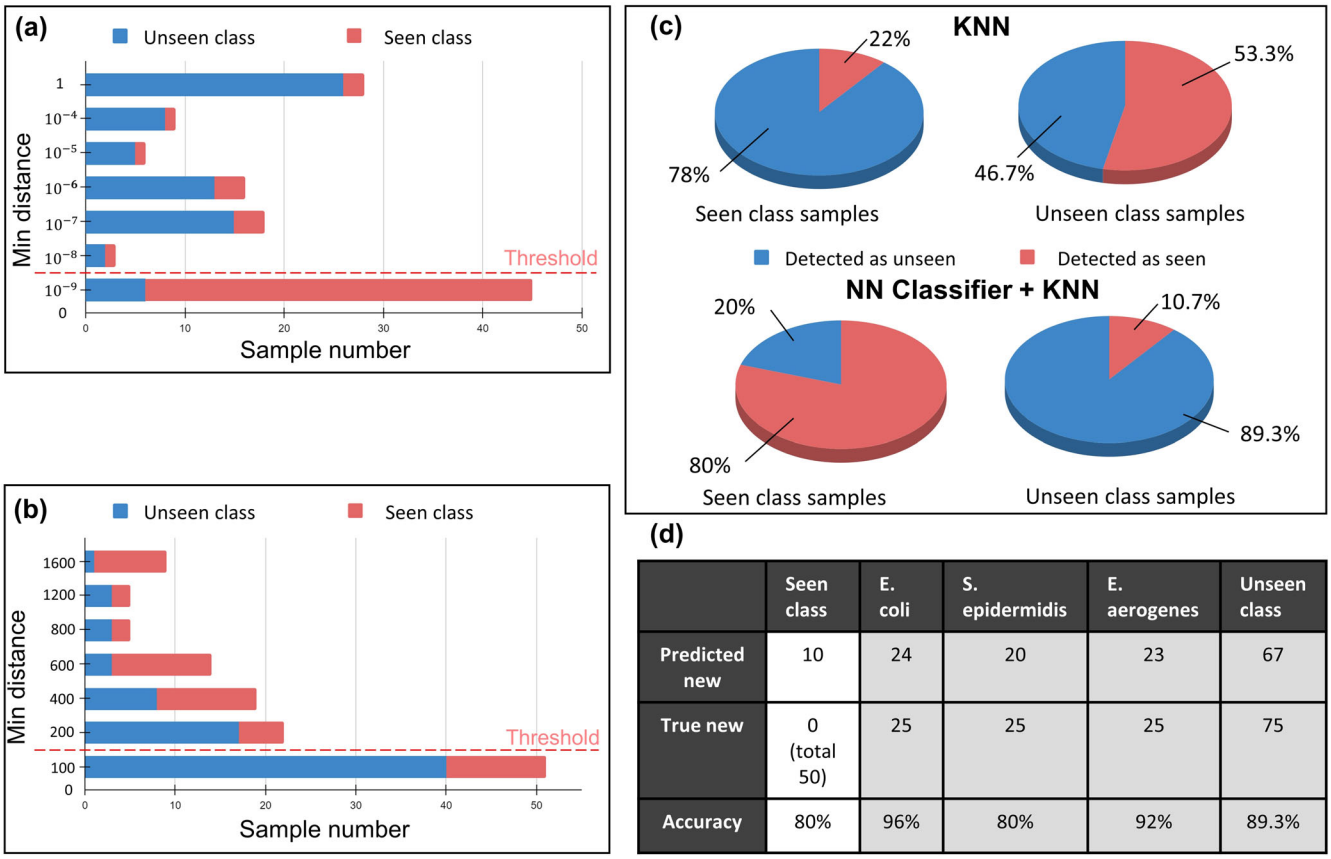
Fig. 5 Neural network regressor for the SERS detection of mixing samples. a Quantitative detection results of three component mixture from neural network regression. **b** Qualitative detection results of neural network regression, where the point colors show the true components, and the point positions show the predicted results. **c** The statistical detection accuracy of each type of sample. The vertical axis denotes the ground truth analytes, while the horizontal axis represents the predicted analytes. The color bar on the right illustrates a colormap ranging from 0% to 100% of predictions.

have existence (1) or non-existence (0) to create eight different combinations in total. Figure 5b illustrates the qualitative detection of neural network regression. Testing points are positioned within the corresponding background area based on the predictions made by the NN regressor. The color of each point represents the actual components, while their positions reflect the corresponding predictions. The results demonstrate a qualitative detection accuracy of 92%. Compared with neural network classification, the neural network regression can directly output the concentration ratio but has a slightly lower accuracy in qualitative detection, making these two models suitable for different scenarios. The statistical detection accuracy of each type of sample is shown in Fig. 5c, from mono-component samples to mixed opponent samples.

Neural network classifier combined with KNN

The KNN model is incorporated with the neural network classifier for detecting unseen dataset. The neural network classifier and the KNN are trained with the same dataset as the neural network regressor, which is the mixtures of three components, MB, TB, and Cip, mixed at different concentration ratios. The output of the neural network classifier is an 8-digit tensor to show the possibility of each class, and 8 clusters are established with the KNN clustering method through supervised training. In the testing scenarios, seen classes points (mixture of MB, TB, and Cip) are used. Additionally, data points from unseen classes, specifically Raman spectra of three bacteria (*E. coli*, *S. epidermidis* and *E. aerogenes*) are incorporated into the evaluation. The minimum distance of each test point to the established clusters is calculated, where seen class points tend to have an extremely small distance to the cluster (in most cases the distance is 0) and the unseen class points typically are far from each established cluster. Therefore, a threshold can be employed to decide if the test point is an unseen class sample. The calculated minimum distances to the established clusters are shown in Fig. 6a, b with thresholds to determine whether the test data points belong to seen classes or unseen classes. A comparison is made between the outcomes of the neural network classifier combined with KNN (Fig. 6a) and that of KNN alone (Fig. 6b). The results show that when employing the neural network classifier and KNN, a significant portion of observed data points is situated within the proximity (threshold) of the clusters, characterized by a small distance. In contrast, only a minimal number of data points from the unseen class exhibit such a proximity. However, outcome of using KNN alone does not provide the apparent dividing line for the two types of samples. The detection results are shown in Fig. 6c, where neural network classifier together with KNN reaches 80% sensitivity (the true seen classes samples are correctly detected as seen) and 89.3% specificity (the true unseen classes samples are correctly detected as unseen). However, applying KNN method alone only yields 22% sensitivity and 46.7% specificity. The detection results together with data distribution are displayed in Fig. 6d.

In comparison, standard analysis technique hierarchical cluster analysis is used for classifying Cip, MB, and TB from 15 different mixtures. As shown in Supplementary Fig. 13, many mixed analytes from different mixtures were inaccurately grouped into one cluster. This suggests that SERS signal combined with DL is a reliable method to identify different organic pollutants in mixtures with high accuracy.



	Seen class	E. coli	S. epidermidis	E. aerogenes	Unseen class
Predicted new	10	24	20	23	67
True new	0 (total 50)	25	25	25	75
Accuracy	80%	96%	80%	92%	89.3%

Fig. 6 Unseen class detection results from the neural network classifier combined with KNN clustering. a The calculated minimum distance toward established clusters using KNN clustering method. **b** The calculated minimum distance toward established clusters using neural network classifier and KNN clustering methods. **c** The accuracy of unseen class points detection from KNN alone, and neural network classifier together with KNN. **d** The statistics of unseen class points detection using neural network classifier and KNN.

Laplacian operation

The detection of different analytes is critically based on the Raman peaks. The model implements the Laplacian operation to strengthen the useful information before the deep neural network without additional processing of Raman spectra. To validate the improvement from the Laplacian operator, the learning curves of neural network regression are compared with and without Laplacian operation are shown in Supplementary Fig. 14, which confirms that incorporating the Laplacian operation leads to a significant decrease in loss over the training.

CONCLUSIONS

The presented research has developed a dual-functional AgNP-ZnONR-SNF nanophotonic sensor by combining electrospinning, hydrothermal, and wet-chemical synthesis techniques. The ZnONR-SNF thin films exhibit remarkable efficacy in decomposing organic dyes (MB, TB, and MO) when exposed to low-intensity UV and mechanical radiation, and they can be reused. Even after undergoing five cycles, the film maintains its impressive degradation efficiency. The results demonstrate the potential of this low-cost, eco-friendly technology for water decontamination against organic pollutants. Additionally, the vertically aligned ZnONR provide suitable geometry for the incorporation of Ag nanoparticles. This configuration creates high-density "hot-spots" for SERS enhancement. Consequently, the AgNP-ZnONR-SNF thin films serve as an effective SERS substrate, facilitating the detection of exceptionally low concentrations of both organic dye (MB) and antibiotics (Cip). A machine learning algorithm is developed for both qualitative and quantitative pollutants detection based on the Raman spectra obtained from AgNP-ZnONR-SNF nanosensor. It has an accuracy of 92.3% in qualitative detection and 90.8% in quantitative classification. This label-free dual-functional nanophotonic sensor, assisted by a machine learning algorithm, is sensitive, fast, and portable, holding great promise towards environmental monitoring and sustainability studies.

METHODS

Chemical and reagents

Tetraethyl orthosilicate (TEOS, 98%), formic acid, ethanol, Polyvinylpyrrolidone (PVP, $M_w = 1{,}300{,}000\ g\ mol^{-1}$), Zinc acetate ($(CH_3CO_2)_2Zn$, 99.99%), zinc nitrate hexahydrate ($Zn(NO_3)_2{\cdot}6H_2O$, 98.0%), hexamethylenetetramine ($C_6H_{12}N_4$, HMTA, ≥99.0%), silver nitrate powder ($AgNO_3$, ≥99.0%), Methylene blue (MB, $C_{16}H_{18}ClN_3S{\cdot}xH_2O$, M_w=319.85 g mol^{-1}), Trypan Blue (TB, $C_{34}H_{24}N_6O_{14}S_4Na_4$, $M_w = 960.81\ g\ mol^{-1}$), Methyl Orange (MO, $C_{14}H_{14}N_3NaO_3S$, $M_w = 327.33\ g\ mol^{-1}$) and Ciprofloxacin (Cip, $C_{17}H_{18}FN_3O_3$, $M_w = 331.34\ g\ mol^{-1}$) are purchased from Sigma-Aldrich Inc. Deionized (DI) water is from a Milli-Q water ultrapure water purification system.

Silica nanofiber fabrication

The silica nanofiber thin film was prepared through electrospinning according to a reported protocol with modifications[38]. In a typical run, 1.9 g of TEOS, 3.15 g of ethanol, 2.0 g of water and 0.04 g of formic acid were mixed with 0.9 g of PVP. The mixture was stirred for 1 h at room temperature until a transparent solution was formed. The solution was electrospun at a feed rate

of 0.9 mL h^{-1} through a 22G stainless needle under a high voltage (16 kV). The ejected silica nanofibers were collected by a flat aluminum plate collector 10 cm away from the needle tip. The silica nanofiber thin film was peeled off from the aluminum foil gently and was then subjected to the calcination for 3 h at 800 °C to remove PVP and other solvent residues.

ZnO nanorods growth

The ZnO nanorods were created based on a seeding-growth method we reported previously[39]. First, the silica nanofiber thin film was immersed in 5 mM zinc acetate solution (in DI water) and then vacuumized to remove air bubbles. It was then transferred into an oven at 180 °C for 20 min for thermal decomposition of the zinc acetate to create ZnO seeds. This process was repeated three times. Then, zinc nitrate hexahydrate (35 mM) and HMTA (25 mM) were added to 45 mL of DI water to provide the hydrothermal growth solution. Next, the seeded thin film was placed in the growth solution in a beaker which was covered by aluminum foil and placed in a water bath for 3 h at a temperature of 90 °C. This growth cycle was repeated twice to form the ZnO nanorods. Finally, the thin film was rinsed with DI water to remove excessive ZnO residuals and dried in an oven at 50 °C.

Silver nanoparticles decoration

Ag nanoparticles were fabricated on top of the ZnO nanorods by UV irradiation. The fabricated thin film was immersed in a 5 mM $AgNO_3$ solution (in DI water) and irradiated under a UV lamp (365 nm, 30 W) for 30 min. Then, the film was washed with DI water and dried in an oven at 50 °C.

SEM and EDS characterization

The morphologies of SNFs, ZnONR-SNF and AgNP-ZnONR-SNF were studied by SEM performing on a FEI Helios 5CX DualBeam scanning electron microscope operating at 5 kV. The EDS measurements and the chemical mapping were performed with the SDD X-ray detector (Oxford®) attached to the SEM microscope operating at 30 kV.

Photo- and piezo-catalytic degradation

All organic molecules were dissolved in distilled water at a concentration of 10 μg mL^{-1} under sonication at room temperature. Calibration curves were obtained by considering the characteristic UV-Vis absorbance values (MB at 664 nm, TB at 590 nm, MO at 463 nm), obtained from a series of diluted solutions at prefixed concentration values. ZnONR-SNF films were cut into 0.8×0.8 cm^2 pieces and employed in glass vials with 3 mL of pollutant solutions for each degradation experiment. The control experiments were carried out in the absence of ZnONR-SNF films. Photocatalytic degradation experiments were carried out by applying a 30 W UV lamp (irradiated from upper side of the vials) at a fixed distance of around 5 cm. In piezo-catalytic degradation experiments, vials were fixed on the MTS 2/4 digital shaker orbiting at 300 rpm. In hybrid experiments, vials were fixed at the shaker and irradiated by the UV lamp simultaneously. All experiments were performed in the dark room at room temperature.

The degradation efficiency was measured by means of light absorbance. First, 30 μL solution was withdrawn from the vials at a 2-h time interval (0 h, 2 h, 4 h, 6 h, 8 h) and placed in a 96-well UV-Star microplate. Then the light absorbance was read by a microplate reader (TECAN SPARK 10 M) at each characteristic peak wavelength. According to the light absorbance readout and recorded calibration curve obtained previously, the concentration of the organic pollutants after degradation was finalized. The degradation efficiency can be calculated by the following equation:

$$Degradation\ efficiency = \frac{C_0 - C}{C_0} \times 100\% \quad (1)$$

Where C_0 is the initial concentration and C is the measured concentration at different time intervals.

SERS detection

The SERS measurements were processed in a liquid system. The AgNP-ZnONR-SNF film was first cut into a square piece (0.5 cm × 0.5 cm). Then the film was focused by a 10× objective lens. Next, the sample aqueous solution (10 μL) was added onto the film. Finally, SERS signals were obtained point-by-point from the mapping grid using a 532 nm laser as the excitation source.

Machine learning algorithm development

The machine learning algorithm consists of a Laplacian operator, a deep neural network with two output modes to show the qualitative and quantitative detection results, and a KNN cluster method which is combined with the classification mode output to enable detection of unseen classes. Every Raman spectrum collected displays the Raman signals of chemical structures in the samples across a 238-point Raman shift axis. The Laplacian operation was directly implemented on the Raman spectra without additional need of preliminary processing. After the Laplacian operation, the treated data was taken into the deep neural network to output the detection results through the two modes. In the first mode, the neural network is a classification (fully connected layers with sigmoid as activation layer) to show the analyte components and concentration level. The output of this mode is a tensor with part of the tensor to show the possibility that the sample belongs to a specific class and part of the tensor either 0 or 1 to indicate if the specific chemical is above a typical cutoff concentration. In the second mode, the neural network is a regression (fully connected with ReLU as activation layer) to directly show the concentration of multiple chemicals that are included in the sample. The output of this mode is a tensor with arbitrary numbers showing the concentration of each chemical, where an extremely small number indicates that the specific analyte does not exist. Both two modes can output the qualitative and quantitative detection of mixed analytes to satisfy different application requests. The model is implemented with python. To enhance OOD detection, specifically for identifying unseen classes, a KNN clustering method is incorporated with the neural network classifier. Clusters aligned with the known classes from the neural network classifier's output can be established through the KNN clustering model. By calculating the distance of a test point to the established clusters after clustering, the test point can be identified as seen class data or unseen class data.

PyTorch, along with its associated toolsets such as Scikit-learn, NumPy, and Skorch, was employed to implement the machine learning models. The model was executed on a standard workstation equipped with an Intel(R) Core(R) i9-9980 XE CPU running at 3.00 GHz, featuring 18 CPU cores, and an 8GB NVIDIA GeForce RTX 2080Ti.

During each training process, the training and testing data were randomly divided. To ensure reliable outcomes, the process was repeated ten times, with the final error and prediction accuracy calculated across these ten training iterations. For instance, in the experiments involving the regressor model, we employed a tenfold cross-validation approach using a total of 1095 data samples. This dataset was further divided into 995 training samples and 100 testing samples. Additional details regarding the sizes of the training and testing datasets can be found in Supplementary Table 1. Furthermore, Supplementary Figs. 15 and 16 present the confusion matrices pertaining to both the training and testing samples.

DATA AVAILABILITY

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

CODE AVAILABILITY

For access to detailed code implementations, please contact the authors directly.

Received: 1 May 2023; Accepted: 17 November 2023;
Published online: 16 January 2024

REFERENCES

1. World Health Organization. Progress on household drinking water, sanitation and hygiene 2000–2020: five years into the SDGs. Joint Water Supply, & Sanitation Monitoring Programme http://apps.who.int/bookorders. (2021).
2. Madhav, S. et al. Water pollutants: sources and impact on the environment and human health. In *Sensors Water Pollutants Monitoring: Role of Material* (eds Devi, P., Kumar, P., Singh, P. & Patil, S.) 43–62 (Springer Singapore, 2019).
3. Capodaglio, A. G., Hlavínek, P. & Raboni, M. Physico-chemical technologies for nitrogen removal from wastewaters: a review. *Ambient. e Agua - Interdiscip. J. Appl. Sci.* **10**, 445–458 (2015).
4. Vermeer, A. W. P., McCulloch, J. K., Van Riemsdijk, W. H. & Koopal, L. K. Metal ion adsorption to complexes of humic acid and metal oxides: deviations from the additivity rule. *Environ. Sci. Technol.* **33**, 3892–3897 (1999).
5. Perić, J., Trgo, M., & Vukojević Medvidović, N. Removal of zinc, copper and lead by natural zeolite—a comparison of adsorption isotherms. *Water Res.* **38**, 1893–1899 (2004).
6. Franco, A. et al. Photocatalytic decolorization of methylene blue in the presence of TiO2/ZnS nanocomposites. *J. Hazard. Mater.* **161**, 545–550 (2009).
7. Raizada, P., Sudhaik, A. & Singh, P. Photocatalytic water decontamination using graphene and ZnO coupled photocatalysts: a review. *Mater. Sci. Energy Technol.* **2**, 509–525 (2019).
8. Weldegebrieal, G. K. Synthesis method, antibacterial and photocatalytic activity of ZnO nanoparticles for azo dyes in wastewater treatment: a review. *Inorg. Chem. Commun.* **120**, 108140 (2020).
9. Ong, C. B., Ng, L. Y. & Mohammad, A. W. A review of ZnO nanoparticles as solar photocatalysts: synthesis, mechanisms and applications. *Renew. Sustain. Energy Rev.* **81**, 536–551 (2018).
10. Sá, A. S. et al. A brief photocatalytic study of zno containing cerium towards ibuprofen degradation. *Materials* **14**, 5891 (2021).
11. Qu, Y. et al. Enhanced photocatalytic degradation of antibiotics in water over functionalized N,S-doped carbon quantum dots embedded ZnO nanoflowers under sunlight irradiation. *Chem. Eng. J.* **382**, 123016 (2020).
12. Saleh, S. M. ZnO nanospheres based simple hydrothermal route for photocatalytic degradation of azo dye. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **211**, 141–147 (2019).
13. Liu, X. & Chen, C. Mxene enhanced the photocatalytic activity of ZnO nanorods under visible light. *Mater. Lett.* **261**, 127127 (2020).
14. Xu, Y. et al. Simple synthesis of ZnO nanoflowers and its photocatalytic performances toward the photodegradation of metamitron. *Mater. Res. Bull.* **76**, 235–239 (2016).
15. Zhou, J., Nie, Y., Jin, C. & Zhang, J. X. J. Engineering biomimetic extracellular matrix with silica nanofibers: from 1D material to 3D network. *ACS Biomater. Sci. Eng.* **8**, 2258–2280 (2022).
16. Vikesland, P. J. Nanosensors for water quality monitoring. *Nat. Nanotechnol.* **13**, 651–660 (2018).
17. Pooja, & Chowdhury, P. Functionalized CdTe fluorescence nanosensor for the sensitive detection of water borne environmentally hazardous metal ions. *Opt. Mater.* **111**, 110584 (2021).
18. Peter, K. T. et al. Using high-resolution mass spectrometry to identify organic contaminants linked to urban stormwater mortality syndrome in Coho Salmon. *Environ. Sci. Technol.* **52**, 10317–10327 (2018).
19. De Paepe, E. et al. Ultra-high-performance liquid chromatography coupled to quadrupole orbitrap high-resolution mass spectrometry for multi-residue screening of pesticides, (veterinary)drugs and mycotoxins in edible insects. *Food Chem.* **293**, 187–196 (2019).
20. Liu, S., Cheng, R., Chen, Y., Shi, H. & Zhao, G. A simple one-step pretreatment, highly sensitive and selective sensing of 17B-estradiol in environmental water samples using surface-enhanced Raman spectroscopy. *Sens. Actuators B Chem.* **254**, 1157–1164 (2018).
21. Goetz, M. J., Coté, G. L., Erckens, R., March, W. & Motamedi, M. Application of a multivariate technique to Raman spectra for quantification of body chemicals. *IEEE Trans. Biomed. Eng.* **42**, 728–731 (1995).
22. Hedegaard, M. et al. Discriminating isogenic cancer cells and identifying altered unsaturated fatty acid content as associated with metastasis status, using K-means clustering and partial least squares-discriminant analysis of Raman maps. *Anal. Chem.* **82**, 2797–2802 (2010).
23. Widjaja, E., Zheng, W. & Huang, Z. Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines. *Int. J. Oncol.* **32**, 653–662 (2008).
24. Ho, C. S. et al. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* **10**, 4927 (2019).
25. Wang, P. et al. Discrimination of blood species using Raman spectroscopy combined with a recurrent neural network. *OSA Contin.* **4**, 672 (2021).
26. 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. T*rends Anal. Chem*. **124**, 115796 (2020).
27. Fan, X., Ming, W., Zeng, H., Zhang, Z. & Lu, H. Deep learning-based component identification for the Raman spectra of mixtures. *Analyst* **144**, 1789–1798 (2019).
28. Herrmann, J.-M. Heterogeneous photocatalysis: fundamentals and applications to the removal of various types of aqueous pollutants. *Catal. Today* **53**, 115–129 (1999).
29. Qi, K., Cheng, B., Yu, J. & Ho, W. Review on the improvement of the photocatalytic and antibacterial activities of ZnO. *J. Alloy. Compd.* **727**, 792–820 (2017).
30. Wang, Z. L. & Song, J. Piezoelectric nanogenerators based on zinc oxide nanowire arrays. *Science* **312**, 242–246 (2006).
31. Lee, H. K. et al. Designing surface-enhanced Raman scattering (SERS) platforms beyond hotspot engineering: emerging opportunities in analyte manipulations and hybrid materials. *Chem. Soc. Rev.* **48**, 731–756 (2019).
32. Ding, S. Y., You, E. M., Tian, Z. Q. & Moskovits, M. Electromagnetic theories of surface-enhanced Raman spectroscopy. *Chem. Soc. Rev.* **46**, 4042–4076 (2017).
33. Fang, Y., Seong, N.-H. & Dlott, D. D. Measurement of the distribution of site enhancements in surface-enhanced Raman scattering. *Science* **321**, 388–392 (2008).
34. Wang, Z., Meng, G., Huang, Z., Li, Z. & Zhou, Q. Ag-nanoparticle-decorated porous ZnO-nanosheets grafted on a carbon fiber cloth as effective SERS substrates. *Nanoscale* **6**, 15280–15285 (2014).
35. Chen, G. et al. Measuring ensemble-averaged surface-enhanced Raman scattering in the hotspots of colloidal nanoparticle dimers and trimers. *J. Am. Chem. Soc.* **132**, 3644–3645 (2010).
36. Gautam, R., Vanga, S., Ariese, F. & Umapathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2**, 11798–11818 (2015).
37. Bocklitz, T., Walter, A., Hartmann, K., Rösch, P. & Popp, J. How to pre-process Raman spectra for reliable and stable models? *Anal. Chim. Acta* **704**, 47–56 (2011).
38. Chen, W. S. et al. Chemically modified electrospun silica nanofibers for promoting growth and differentiation of neural stem cells. *J. Mater. Chem. B* **2**, 1205–1215 (2014).
39. Zhang, M. et al. Silver nanoparticle on zinc oxide array for label-free detection of opioids through surface-enhanced Raman spectroscopy. *RSC Adv.* **11**, 11329–11337 (2021).

ACKNOWLEDGEMENTS

Grateful acknowledgment is extended to the Electron Microscope Facility at Dartmouth College for their invaluable support. Special thanks go to Paul A. Defino, Instrumentation Lab Manager at Dartmouth College, for his assistance with the operations of the Raman spectroscope.

AUTHOR CONTRIBUTIONS

J.Z. conceived and designed the experiments, performed the experiments, analyzed the data and wrote the paper; Z.W. analyzed the data, contributed analysis tools and wrote the paper; C.J. conceived and designed the experiments, performed the experiments and wrote the paper. The three authors share co-first authorship for this work. J.Z. (corresponding author): supervised the study, reviewed and edited the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41545-023-00292-4.

Correspondence and requests for materials should be addressed to John X. J. Zhang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.