



*Citation for published version:*

Xie, L, Luo, S, Liu, Y, Ruan, X, Gong, K, Ge, Q, Li, K, Valev, VK, Liu, G & Zhang, L 2023, 'Automatic Identification of Individual Nanoplastics by Raman Spectroscopy Based on Machine Learning', *Environmental Science and Technology*, vol. 57, no. 46, pp. 18203-18214. <https://doi.org/10.1021/acs.est.3c03210>

*DOI:*

[10.1021/acs.est.3c03210](https://doi.org/10.1021/acs.est.3c03210)

*Publication date:*

2023

*Document Version*

Peer reviewed version

[Link to publication](#)

This document is the Accepted Manuscript version of a Published Work that appeared in final form in *Environmental Science and Technology*, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see <https://pubs.acs.org/doi/10.1021/acs.est.3c03210>

**University of Bath**

## **Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Automatic Identification of Individual Nanoplastics by Raman spectroscopy based on Machine Learning

Lifang Xie<sup>1,2,3</sup>, Siheng Luo<sup>4,5</sup>, Yangyang Liu<sup>1,2,3</sup>, Xuejun Ruan<sup>1,2,3</sup>, Kedong Gong<sup>1,2,3</sup>, Kejian Li<sup>1,2,3</sup>, Qiuyue Ge<sup>1,2,3</sup>,  
Ventsislav Kolev Valev<sup>6</sup>, Guokun Liu<sup>4,5</sup>, Liwu Zhang<sup>1,2,3\*</sup>

<sup>1</sup>Department of Environmental Science & Engineering, Fudan University, Shanghai, 200433, Peoples' Republic of China.

<sup>2</sup>Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention, Fudan University, Shanghai, 200433, Peoples' Republic of China.

<sup>3</sup>Shanghai Institute of Pollution Control and Ecological Security, Shanghai, 200092, Peoples' Republic of China.

<sup>4</sup>State Key Laboratory of Marine Environmental Science, College of the Environment and Ecology, Xiamen University, Xiamen 361005, and P. R. China.

<sup>5</sup>Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, Center for Marine Environmental Chemistry & Toxicology, Xiamen University, Xiamen 361102, China.

<sup>6</sup>Centre for Photonics and Photonic Materials and Centre for Nanoscience and Nanotechnology, Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, U.K.

## ABSTRACT

The increasing prevalence of nanoplastics in the environment underscores the need for effective detection and monitoring techniques. Current methods mainly focus on microplastics, while accurate identification of nanoplastics is challenging due to their small size and complex composition. In this work, we combined highly reflective substrates and machine learning to accurately identify nanoplastics using Raman spectroscopy. Our approach established Raman spectroscopy datasets of nanoplastics, incorporated peak extraction and retention data processing, and constructed a random forest model that achieved an average accuracy of 98.8% in identifying nanoplastics. We validated our method with tap spiked water samples, achieving over 97% identification accuracy, and demonstrated the applicability of our algorithm to real-world environmental samples through experiments on rainwater, detecting nanoscale Polystyrene (PS) and

28 Polyvinyl chloride (PVC). Despite the challenges of processing low-quality nanoplastic Raman  
29 spectra and complex environmental samples, our study demonstrated the potential of using random  
30 forests to identify and distinguish nanoplastics from other environmental particles. Our results  
31 suggest that the combination of Raman spectroscopy and machine learning holds promise for  
32 developing effective nanoplastic particle detection and monitoring strategies.

33

34 **KEYWORDS:** Raman Spectroscopy; Machine Learning; Nanoplastics; Microplastics; Random  
35 Forest

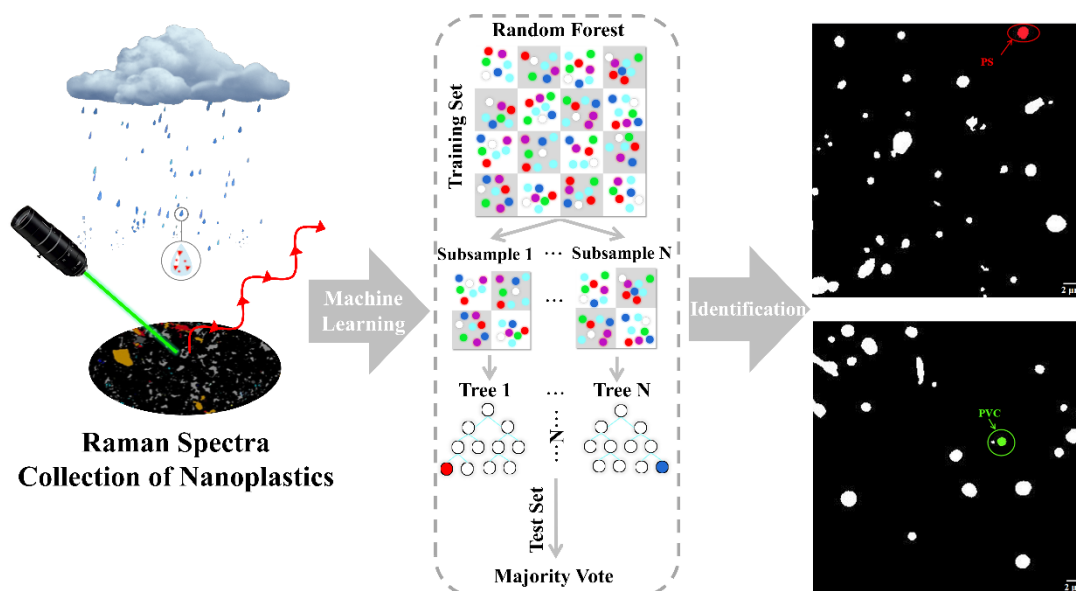
36

### 37 **SYNOPSIS**

38 This study combined Raman spectroscopy and machine learning to accurately identify nanoplastics,  
39 with successful application to complex environmental samples.

40

### 41 **TOC**



42

43

## 44 1. INTRODUCTION

45 Due to the massive manufacturing of plastic products and the unregulated disposal of plastic waste,  
46 plastic particles are released into the environment and circulated worldwide in large quantities<sup>1-4</sup>.  
47 Researchers have demonstrated the widespread presence of microplastics (<5 mm), e.g. in lakes<sup>5,6</sup>,  
48 oceans<sup>7, 8</sup>, and sediments<sup>9</sup>. In addition to microplastics, recently, researchers have started to  
49 recognize the importance of nanoplastics (<1  $\mu\text{m}$ ), which have been detected in samples of the  
50 atmosphere<sup>10,11</sup>, seawater<sup>12</sup>, snow<sup>13</sup>, and soil<sup>14</sup>. In contrast to microplastics, nanoplastics are small  
51 enough to penetrate biological barriers and transfer between tissues<sup>15,16</sup>, their presence in blood<sup>17</sup>  
52 and placenta<sup>18</sup> has been confirmed, which poses a significant risk to human health<sup>19</sup>. Therefore, the  
53 development of effective methods to detect nanoplastics in the environment is of direct relevance  
54 for both human health and environmental science.

55 To assess environmental impact, both the composition and the concentration of nanoplastics  
56 are of obvious interest. Raman spectroscopy is a widely used and effective technique for identifying  
57 microplastics, including those found in soil (150-200  $\mu\text{m}$ )<sup>20</sup> and marine environments (>10  $\mu\text{m}$ )<sup>21</sup>.  
58 Raman Spectral Imaging has been utilized to identify plastic microbeads larger than 4  $\mu\text{m}$ <sup>22</sup>. Micro-  
59 Raman spectroscopy allows for spectral analysis with spatial resolution below 1  $\mu\text{m}$ <sup>23</sup>, which could  
60 aid in the detection of plastic particles of submicron and nano sizes in environmental samples.  
61 Recent research has indicated that Raman mapping can be employed to visualize, image, and  
62 identify nanoplastics down to 100 nm in size<sup>24</sup>. However, due to the extremely weak Raman signal  
63 usually present in nanoplastics, accurate identification is problematic. The Raman signal (peak  
64 intensity) can be significantly enlarged by structural methods, such as optimizing the optical  
65 structure<sup>25</sup> or designing surface-enhanced Raman substrates<sup>10</sup>. However, robust and reliable Raman  
66 identification requires processing large amounts of data, whose analysis by highly skilled scientists  
67 does not scale up.

68 Machine learning has received increasing attention from researchers for its ability to automate  
69 the analysis of complex spectral data. The combination of machine learning and Raman  
70 spectroscopy can improve the efficiency and accuracy of micro(nano)plastic identification, and the  
71 introduction of algorithms related to image recognition can be applied to the analysis of spectral  
72 matrices<sup>26</sup>. Principal component analysis (PCA) is capable of extracting key features of data from

73 multiple dimensions of datasets, and has been applied to analyze a variety of microplastics as well  
74 as to optimize Raman imaging of plastics<sup>27-31</sup>. However, the composition analysis of unknown  
75 particles also requires supervised learning, and existing methods include PCA-linear discriminant  
76 analysis (PCA-LDA)<sup>32</sup>, Support vector machine (SVM)<sup>32</sup>, Linear Partial Least Squares Regression  
77 by Intervals (iPLS-R)<sup>33</sup> and Competitive Adaptive Weighted Sampling (CARS/PLS-R)<sup>33</sup>, to  
78 improve and quantify the accuracy. Current publications mainly report on the spectral classification  
79 of microplastics, targeting plastic particles larger than 15  $\mu\text{m}$ , and the collected databases consist of  
80 high-quality Raman spectra<sup>32-35</sup>. However, these databases are insufficient to meet the existing  
81 needs of nanoplastics detection. To date, only a few studies have attempted to use chemometric tools  
82 for spectral analysis of submicron and nanoplastics. As the size decreases, complications arise, such  
83 as changes in plastic crystallinity, and some characteristic Raman peaks become too weak or  
84 shifted<sup>36</sup>. The weak signal features of the Raman spectra of nanoplastics in the atmosphere can be  
85 easily covered by noise<sup>10</sup>. Therefore, there is a need to establish methods for obtaining weak signals  
86 of nanoplastics and specialized Raman databases for nanoplastics. Additionally, it is essential to  
87 establish effective qualitative analysis methods to facilitate the identification of nanoplastics in the  
88 environment.

89 The low-quality Raman spectra of nanoplastics make it impractical to be identified by the  
90 traditional dataset of traditional high-quality spectra for matching<sup>23</sup>. To address the low-quality  
91 Raman spectra, random forests have a higher tolerance for noise over other algorithms<sup>37,38</sup>. Random  
92 forest is an ensemble learning model that uses bagging and random feature selection to construct  
93 multiple independent decision tree structures, and then integrates the voting results of each decision  
94 tree for prediction<sup>39</sup>. The ability of random forests to handle noisy and high-dimensional data,  
95 achieve high accuracy, and provide interpretable results makes it a valuable tool for scientific  
96 research in many fields<sup>40</sup>. In this work, we applied machine learning approach in automatic  
97 processing of Raman spectra to identify nanoplastics. In terms of detection methods, the highly  
98 reflective aluminum sheets (95% reflectivity) facilitate automatic image identification of  
99 nanoplastics as well as signal collection. For data processing, the peak extraction and retention  
100 (PEER) algorithm served to extract meaningful features from weak Raman signals (for technical  
101 details see ref.<sup>41</sup>). To identify unknown plastic samples, we constructed a random forest model  
102 enabling successful discrimination between five common plastic contaminants, namely

103 Polyethylene (PE), polytetrafluoroethylene (PTFE), Polystyrene (PS), polymethyl methacrylate  
104 (PMMA) and Polyvinyl chloride (PVC), with an average accuracy of 98.8%, an average sensitivity  
105 of 98.5%, and an average specificity of 100%. By combining machine learning with Raman  
106 spectroscopy, our method significantly improves the accuracy and efficacy of nanoplastic  
107 identification in complex environments. We demonstrated the effectiveness of our method through  
108 successful lab tests (utilizing tap water mixed with plastic particles) with identification accuracy  
109 exceeding 97%. Furthermore, the applicability of our method to real-world environmental samples  
110 was demonstrated through experiments on natural rainwater, where nanoscale PS and PVC were  
111 detected. Our work successfully addresses the challenges of identifying low-quality nanoplastic  
112 Raman spectra and handling complex environmental samples, demonstrating the value of machine  
113 learning in developing effective nanoplastic particle detection and monitoring strategies.

114

## 115 **2. Experimental Section**

### 116 **2.1 Chemicals and Materials**

117 Polyethylene (PE) and polytetrafluoroethylene (PTFE) pellets (with diameters  $\leq 1 \mu\text{m}$ ) were  
118 purchased from Qihong Plastic Material Co, China. Polystyrene (PS) and polymethyl methacrylate  
119 (PMMA) spheres (with diameters of 360 nm, 500 nm, and  $1 \mu\text{m}$ ) were purchased from Shanghai  
120 Huge Biotech Co, China. Polyvinyl chloride (PVC) pellets (with diameters of 300 nm  $\sim 1 \mu\text{m}$ ) used  
121 in the experiments was provided by Shanghai Guanbu Electromechanical Technology Co, China.  
122 Amorphous recycled PTFE material was provided by Dongguan Teplas Long Chemical Raw  
123 Material Co, China. The nanoplastics were dispersed with anhydrous ethanol or deionized water  
124 and diluted to 50  $\mu\text{g/mL}$  for subsequent analysis of individual plastic particles. Following the  
125 dispersion and dilution process, the nanoplastic solution was filtered through cellulose filter with  
126 pore size of  $1 \mu\text{m}$ . This filtration step was performed to eliminate large plastic particles and to  
127 guarantee that only nano-sized particles were present in the sample for subsequent analysis. The  
128 aluminum sheets (95% reflectivity) were purchased from Shenzhen Dixuan Metal Co., Ltd. in China,  
129 and silicon wafers were purchased from Suzhou Research Material Micro-Nano Technology Co,  
130 China. All substrates were thoroughly cleaned by ultrasonication in ethanol and subsequently rinsed

131 with ultrapure water to remove any possible contaminants. Additionally, great care was taken in  
132 handling the substrates to prevent any potential scratches or contamination during the experiments.  
133

## 134 **2.2 Sample preparation**

135 Samples extracted from the environment are usually complex and pretreatment is required to remove  
136 impurities that could cause signal interference. Biological contamination requires special attention  
137 as it can lead to strong fluorescence that interferes with the Raman signals from nanoplastics. Tap  
138 water mixed with a variety of plastics was first filtered through cellulose filter with pore size of 1  
139  $\mu\text{m}$  to remove large particulate matter. Next, it was filtered through cellulose filter with pore size of  
140 0.25  $\mu\text{m}$  to capture nanoscale particulate matter. To reduce the interference of organic matter and  
141 biological substrates, hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) solution (30%) was added to the above filters and  
142 the samples were digested at room temperature for 24 h. Afterward, we applied ultrasonic water  
143 bath to release particles potentially trapped on the filter membrane surface, allowing them to fall  
144 into the solution and be collected. Finally, concentration was performed by constant temperature  
145 evaporation at 50 °C. Throughout the process, interference from equipment and laboratories was  
146 minimized by using clean glass containers, wearing cotton white coats, and covering the sampler  
147 with aluminum foil.

148 When it rains, the micro(nano)plastic particles present in the atmosphere fall under the wash  
149 and are carried by the rain. Rainwater was collected in clean glass containers on the roof of the  
150 Department of Environmental Science building, at Fudan University. Rainwater was pretreated  
151 using the same method as we used for tap water. For sample preparation, the pretreated sample was  
152 dropped onto the surface of aluminum plates (95% reflectance) to have it dry, and subsequently  
153 examined using Raman spectroscopy. The specific experimental flow chart is shown in Figure 1 (a).  
154 We collected rainwater for a total of 3 days in March 2023, and the samples were measured in  
155 triplicate to ensure accuracy and reproducibility.

156

## 157 **2.3 Characterization**

158 Raman spectroscopy was performed with an XploRA Plus confocal Raman spectrometer (Jobin  
159 Yvon, HORIBA Gr, France) equipped with a  $\times 100$  Olympus microscope objective. The external-  
160 cavity diode laser (532 nm) used in our study has a total power output of 100 mW. In order to  
161 minimize potential damage to the plastic samples while maintaining high sensitivity and accuracy  
162 in our measurements, we utilized only 10% of the laser's energy for nanoplastic detection, as  
163 demonstrated in Figure S1. No noticeable melting or damage to the plastics was observed, and no  
164 carbon peaks were detected in the Raman spectroscopy. Spectra were collected using a multichannel  
165 EMCCD device with confocal imaging of  $0.5\ \mu\text{m XY}$  and the diffraction grating density had 1200  
166 lines per mm. The spectra were collected from 200 to  $2000\ \text{cm}^{-1}$ , with 2 spectra accumulations for  
167 sample at a 30 s acquisition time per spectrum. Particle identification and automatic acquisition of  
168 spectra are implemented by Particle Finder in LabSpec 6. The optical microscope images were  
169 processed using Image Pro software.

170

## 171 **2.4 Data Processing**

172 (1) Establishment of nanoplastic datasets. The spectra included in the nanoplastic database  
173 were obtained directly from the plastic samples. To establish the internal Raman spectral dataset, a  
174 total of 1000 individual nanoparticles were examined, covering five different plastic types  
175 (including PE, PTFE, PS, PVC, and PMMA). For each specific plastic category, 200 nanoparticles  
176 were selected for subsequent analysis (raw data of nanoplastic datasets were provided in SI).

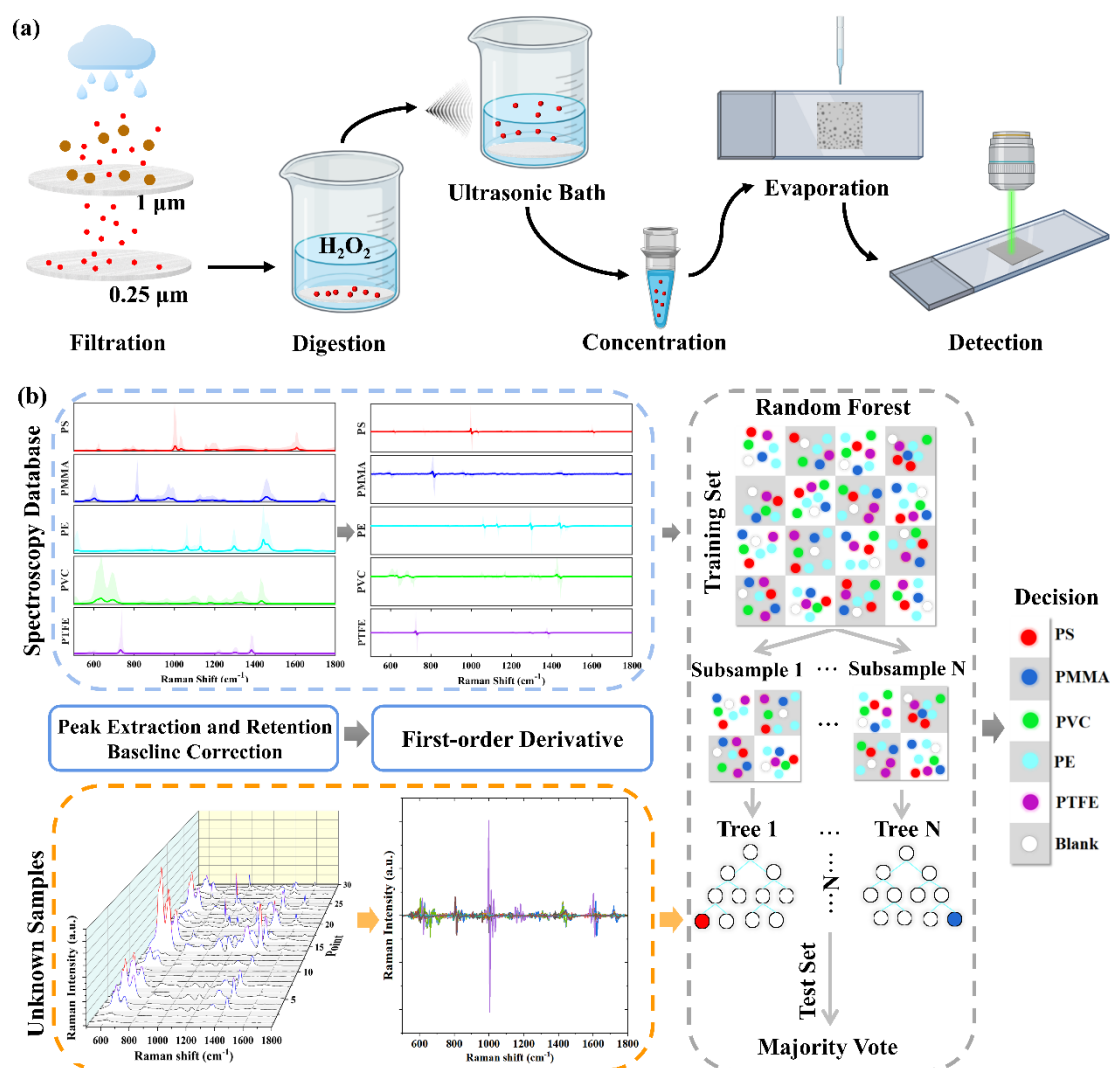
177 (2) Data pre-processing. The obtained Raman spectra were cut to the range of  $500\text{-}1800\ \text{cm}^{-1}$   
178 and later processed by polynomial equations for baseline correction. Weak Raman signal extraction  
179 and denoising are required because the inherent noise of the system interferes with the identification  
180 of plastic feature peaks. The PEER algorithm has good signal extraction and retention, as described  
181 earlier<sup>41</sup>. To reduce the influence of overlapping peaks and the background on the model  
182 construction, the subsequent steps use the first-order derivatives of the Raman spectra.

183 (3) PCA analysis. The pre-processed data were analyzed by Origin (2021) Principal  
184 Component Analysis for Spectroscopy software. In this study, the spectral wave number variables

185 were compressed into 10 principal components (PCs) for explaining the variance (>95%). The PCs  
186 that accounted for the top three were selected for 3D mapping to achieve separation of the Raman  
187 database.

188 (4) Random forest analysis. A random forest model was developed using MATLAB R2021b  
189 (The MathWorks) software to analyze the pre-processed data. The training database consisted of  
190 1,200 spectra, including five types of plastics with 200 spectra each, plus 200 blank spectra. The  
191 Random Forest classifier was constructed by selecting a random subset from the training set  
192 (randomly choosing half of the dataset), creating 100 uncorrelated decision trees, and aggregating  
193 the outputs of the decision trees for the final prediction, as depicted in Figure 1(b). The importance  
194 of features within the random forest was determined by counting the features used in the decision  
195 tree nodes. Upon completion of the model construction, the model's performance was evaluated by  
196 calculating the out-of-bag classification error on the test dataset (the other half of the dataset).  
197 Specific reference metrics included accuracy, sensitivity, and specificity<sup>32, 42, 43</sup> (specific formulas  
198 are shown in SI). To ensure logical and scientific research coherence, the entire process, including  
199 dataset partitioning, model training, and validation, was repeated 100 times. Each iteration involved  
200 randomly selecting a new set of training data, leading to slightly different training and, consequently,  
201 potentially different results. The displayed outcome represents the average value, calculated from  
202 these 100 tests. Support Vector Machine (SVM) and Back Propagation (BP) Neural Networks were  
203 developed using MATLAB R2021b (The MathWorks) software to compare the performance of  
204 random forests. To address the needs of real-world detection and evaluate the reliability of the  
205 constructed model, the test set employed actual nanoplastic samples (spiked tap water samples,  
206 rainwater samples) to assess the random forest model.

207



208

209 **Figure 1.** (a) Pre-processing workflow for the environmental samples and (b) schematic diagram of  
 210 the random forest model. The raw data need to be pre-processed (the PEER algorithm and baseline  
 211 correction) and later processed by first-order derivatives to build the random forest model. A  
 212 randomly selected training set creates 100 uncorrelated decision trees to build the random forest  
 213 classifier, and then the outputs of the decision trees are aggregated for the final prediction.

214

### 215 3. Results and Discussion

#### 216 3.1 Raman detection and spectral database establishment for nanoplastics

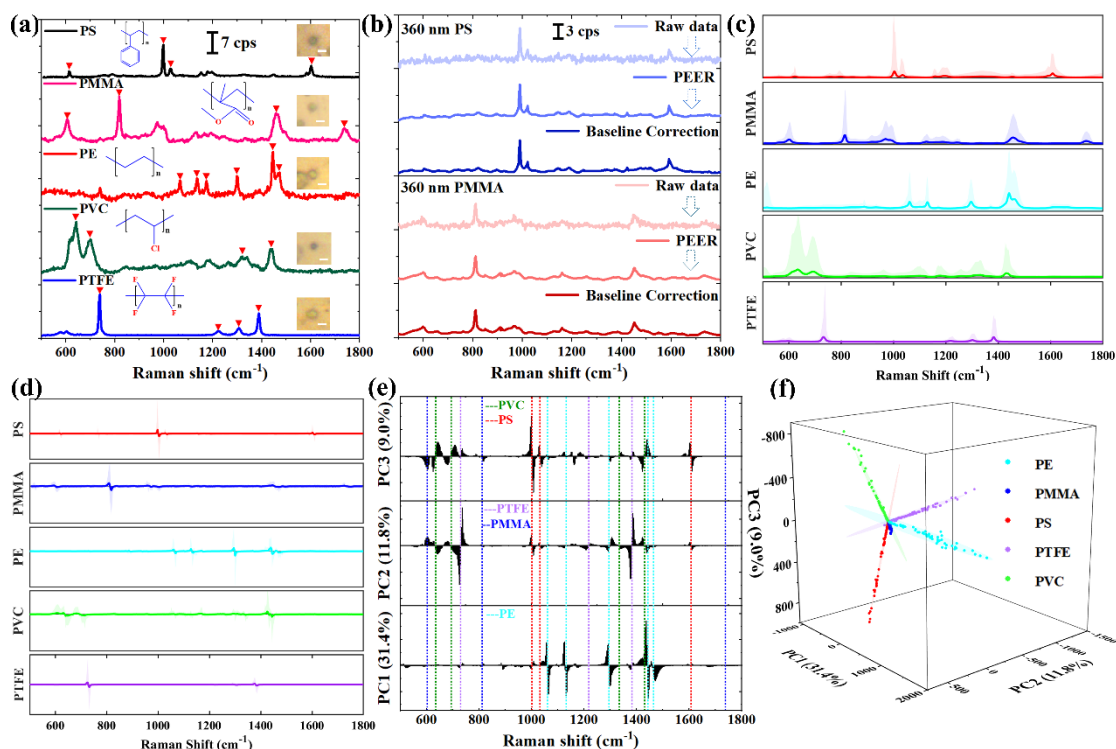
217 To effectively detect nanoplastics, it is crucial to develop a reliable detection method, with the  
 218 selection of the substrate playing a vital role. Aluminum, silicon, and glass slides were first

219 compared for their suitability as substrates for Raman spectroscopy acquisition, as shown in Figure  
220 S2(a). PS nanoparticles with a diameter of 360 nm served as the model. The characteristic peak of  
221 PS ( $\sim 1000\text{ cm}^{-1}$ ) experienced significant interference from the silicon secondary peak ( $\sim 936\text{ cm}^{-1}$ )  
222 and the fluorescence of the glass slide within the  $800\text{-}1200\text{ cm}^{-1}$  range. In contrast, aluminum  
223 presented a clean signal background (aluminum foil, pressed aluminum plate, and polished  
224 aluminum plate all exhibited no background signal, as depicted in Figure S2(b)). In addition, the  
225 smooth properties of polished aluminum plates (95% reflectance) allow nanoplastics to be observed  
226 under optical microscopy ( $\times 100$ ) with clear imaging and facilitated the automatic identification of  
227 particles. Polished aluminum plates, with their low background signal, high reflectance, and low  
228 cost, offer a promising option for detecting nanoplastics using Raman spectroscopy. In this study,  
229 we focused on a selection of the most common and representative plastics found in the environment.  
230 These types of plastics have been detected in various environmental samples<sup>10-14</sup>. They served to  
231 establish our nanoplastic standard spectral datasets. Raman data were collected from individual  
232 plastic particles  $\leq 1\text{ }\mu\text{m}$  (as shown in Figure 2(a)). (1) PE, with C-C as the main chain, has a  
233 relatively simple structure. Its Raman peaks are dominated by C-C stretching ( $\sim 1064, 1131\text{ cm}^{-1}$ ),  
234  $\text{CH}_2$  rocking and bending ( $\sim 1173, 1298, 1442, 1460\text{ cm}^{-1}$ )<sup>44</sup>. (2) PS has branched chains substituted  
235 by benzene rings, so its characteristic peaks are composed of ring breathing ( $\sim 1000\text{ cm}^{-1}$ ), C-H  
236 rocking ( $\sim 1030\text{ cm}^{-1}$ ) and C-C stretching ( $\sim 1604\text{ cm}^{-1}$ )<sup>45</sup>. (3) PMMA, with heteroatoms (O) on the  
237 branched chains, whose structure are mainly linked by C-C and (C=O)-O-C, with C=O stretching  
238 ( $\sim 1729\text{ cm}^{-1}$ ), C-COO stretching ( $\sim 599\text{ cm}^{-1}$ ), and C-O-C symmetric stretching ( $\sim 808\text{ cm}^{-1}$ ) as  
239 Raman features<sup>46</sup>. (4) PVC and PTFE, substituted by halogen atoms (Cl, F) on the branched chain.  
240 PVC can be specifically distinguished by C-Cl stretching ( $\sim 635, 694\text{ cm}^{-1}$ )<sup>47</sup>. PTFE has C-C  
241 vibrations ( $1215\text{ cm}^{-1}$ ) and is qualitatively distinguished by C-F vibrational stretching ( $\sim 729, 1298,$   
242  $1379\text{ cm}^{-1}$ )<sup>48</sup>. The composition of nanoplastics was determined based on the location of Raman  
243 characteristic peaks as well as on their relative intensity (the main characteristic peaks and band  
244 assignments of different plastics are shown in Table S1). Nanoplastics also provide information  
245 about the C-H peaks in the Raman wavenumbers near  $3000\text{ cm}^{-1}$ . However, due to the high similarity  
246 of peak shapes and positions between PTFE and PE, as well as between PVC and PMMA, they  
247 cannot be well distinguished from each other (as shown in Figure S3). Moreover, the Raman spectra  
248 of plastics provide limited useful information in the range of  $1800\text{-}3200\text{ cm}^{-1}$  and focusing on a

249 narrower spectral range reduces computational complexity. Therefore, we chose to concentrate on  
250 the spectral features in the range of 500-1800  $\text{cm}^{-1}$ , as these features are more robust and reliable  
251 for identifying and differentiating various types of nanoplastics.

252 Next, to check the dependence of Raman peaks on particle size, we compared the Raman  
253 spectral signal intensities of nanoplastics (PS, PMMA) with different sizes (1  $\mu\text{m}$ , 500 nm, 360 nm),  
254 as shown in Figure S4(a). As the size decreases, the Raman signal intensity of individual  
255 nanoplastics significantly declines, and there is a clear proportional relationship between the  
256 characteristic peak area and the particle area, as shown in Figure S4(b). The smallest detectable  
257 sizes for PS and PMMA were found to be 360 nm. For other plastic materials, the minimum  
258 detectable particle size was determined to be  $\sim 500$  nm (PTFE),  $\sim 500$  nm (PVC), and  $\sim 600$  nm (PE)  
259 at the current level of techniques, as it is not feasible to produce fixed-size particles for these  
260 materials. Notably, although the characteristic bands of the 360 nm PS particles are still identifiable,  
261 the signal-to-noise ratio (SNR) has deteriorated. However, for nanoplastics, clear identification in  
262 such low-quality spectra is evidently necessary. The random noise in our experiments originates  
263 from the charge-coupled devices (dark current and readout noise) and from the signal itself  
264 (scattered particle noise)<sup>49</sup>. The PEER algorithm consists of peak identification, retention, and  
265 denoising, and is capable of removing the interference of noise on spectra, where overlapping peaks  
266 and spikes coexist. The Raman spectra of 360 nm PS and 360 nm PMMA were used as an example  
267 to compare the denoising effect of the PEER algorithm executed before and after on the spectra.  
268 Figure 2(b) shows that the original Raman spectrum of PS has both sharp single peaks ( $\sim 1000$   $\text{cm}^{-1}$ ),  
269 and peaks with poor SNR ( $\sim 1604$   $\text{cm}^{-1}$ ). The PMMA feature peak at  $\sim 808$   $\text{cm}^{-1}$  is sharp, while  
270  $\sim 1452$  and  $599$   $\text{cm}^{-1}$  are strongly affected by the background noise. Upon comparison, the Raman  
271 peak positions, shapes and intensities of the smoothed spectra correspond well to those of the  
272 original spectra. As illustrated in Figure S5, a comparison of the Raman characteristic peaks of  
273 PMMA particles with different sizes before and after PEER processing demonstrates this  
274 correspondence. We observe that the method of smoothing using the PEER algorithm can extract  
275 the Raman signal peaks from the background noise. Next, the background of the denoised PS  
276 nanoplastic Raman spectral data was removed, in order to reduce the background fluctuations in the  
277 Raman signal. Asymmetric least squares were used for baseline correction. The smoothing and  
278 baseline removal preprocessing steps change neither the wave number nor the intensity of the peaks.

279           Plastics have multiple vibrational characteristics, including some distinct features and many  
280 overlapping characteristic bands. Conventional visual identification methods are time-consuming  
281 and not suitable for processing large amounts of data. Moreover, subtle differences in low-quality  
282 spectra can be overlooked. In Figure 2(c), the Raman spectral datasets of plastic particles, including  
283 PS, PMMA, PTFE, PVC, and PE, are displayed. PCA serves as an example of unsupervised learning,  
284 operating without prior knowledge of the components. By projecting the spectral data onto a set of  
285 linearly uncorrelated directions called principal components (PCs), 85.3% of the variance in the  
286 original data can be captured with the three PCs that contributed the most, as shown in Figure S6(a).  
287 However, the degree of aggregation among different libraries is relatively poor, as illustrated in  
288 Figure S6(b). To make the database more clearly separated, first-order derivative processing was  
289 subsequently applied (as shown in Figure 2(d)). The first-order derivative processing represents the  
290 fluctuations of the characteristic peaks more intuitively and makes the PCA analysis free from the  
291 interference introduced by the absolute intensity of the Raman signal. The PC loading plot (Figure  
292 2(e)) calculates the importance of each band and shows the various prominent features (indicated  
293 by dashed lines) corresponding to the nanoplastics. The PE feature peak ( $\sim 1064, 1173, 1298, 1442,$   
294  $1460 \text{ cm}^{-1}$ ) in PC1 (Variance  $\approx 31.4\%$ ) has the largest contribution. PC2 (Variance  $\approx 11.8\%$ )  
295 represents characteristic bands of PMMA ( $\sim 599, 808, 1729 \text{ cm}^{-1}$ ), and PTFE ( $\sim 729, 1215, 1379 \text{ cm}^{-1}$ ).  
296 PC3 (Variance  $\approx 9.0\%$ ) contains more information, including the characteristic bands of PS  
297 ( $\sim 1000, 1030, 1604 \text{ cm}^{-1}$ ) and of PVC ( $\sim 635, 694, 1333, 1431 \text{ cm}^{-1}$ ). To visualize the data structure  
298 of various plastic spectra, PC fraction visualization plot was generated for the spectral data of  
299 different types of plastics (Figure 2(f)). The five types of plastics were assembled into distinct  
300 clusters, attributed to the substantial differences in their spectra. This facilitated classification using  
301 machine learning algorithms. The first-order derivative processing allowed sharp spectral peaks in  
302 the spectra to occupy a larger proportion and reduced the range and values of the PC scores, filtering  
303 out invalid false peak information, which benefits the differentiation of weak spectra. To ensure the  
304 validity of the dataset inputted into subsequent models, all subsequent steps employed first-order  
305 derivative processing.  
306



307

308 **Figure 2.** (a) Raman spectra of nanoplastics with different compositions (PS, PMMA, PE, PVC,  
 309 and PTFE) are shown, with a white scale bar of 1  $\mu\text{m}$ . (b) Raman spectra processed using the Peak  
 310 Extraction and Retention (PEER) algorithm and baseline correction are shown, with 360 nm PS and  
 311 360 nm PMMA spectra as examples. The Raman spectral datasets are displayed after the  
 312 preprocessing step (c) and after being processed by first-order derivatives (d), with the average of  
 313 all spectra shown as a solid line. (e) The Principal Component (PC) loading plot for the first-order  
 314 derivative dataset is shown. Highlighted features are marked with dashed lines, indicating the most  
 315 significant bands contributing to the variance in the data. (f) 3D plot of PC scores from different  
 316 samples is displayed, with each point corresponding to a Raman spectrum.

317

### 318 3.2 Establishment of Random Forest

319 PCA is primarily employed to simplify high-dimensional Raman spectral data by reducing its  
 320 dimensionality. This process renders the data more manageable and visually interpretable for  
 321 nanoplastic datasets. However, PCA may struggle to adequately distinguish weak signal spectra of  
 322 nanoplastics. The projections of PC1&PC2 fractions (Figure 3(a)) and PC1&PC3 fractions (Figure  
 323 S7(a)) for different nanoplastic particles illustrate this problem. This issue is particularly

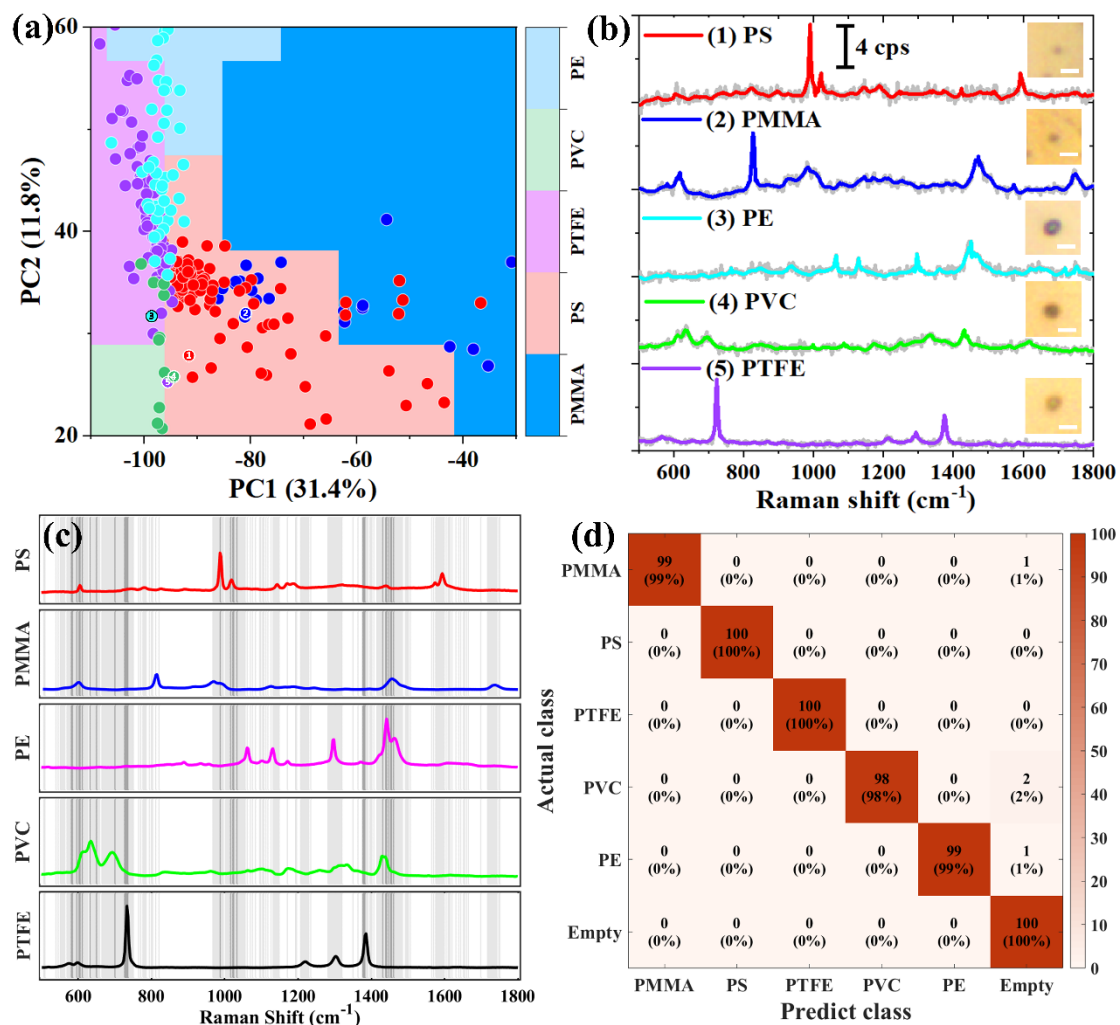
324 pronounced for nanoparticles of very small sizes, which exhibit low SNR, resulting in extremely  
325 high similarity between spectra. Figure 3(b) displays the low-intensity Raman spectra of various  
326 types of nanoplastics corresponding to points 1, 2, 3, 4, and 5 in Figure 3(a), emphasizing the  
327 challenge of identifying nanoplastics at low SNR.

328 In contrast, the random forest algorithm is capable of effectively classifying overlapping  
329 spectra by utilizing multiple decision trees and feature selection. Consequently, we employed the  
330 random forest algorithm to classify the Raman spectral datasets into multiple classes, demonstrating  
331 its effectiveness in distinguishing various types of nanoplastics. 100 uncorrelated decision trees  
332 were created from a randomly selected subset of the training set (randomly selected half of the  
333 dataset) to build a random forest classifier, and then the output of the decision trees was pooled, so  
334 that the model could make the final prediction, with the representative architecture shown in Figure  
335 1(b). The performance of the model was tested by calculating the out-of-bag classification error on  
336 the test dataset (the other half of the dataset). As the tree increases, the error decreases significantly.  
337 When the tree increases to 100, the random forest achieves a good fit (Figure S8(a)). The  
338 performance of the classifier can be represented by the confusion matrix (Figure 3(d)) that calculates  
339 the percentage assigned to the test data compared to the real labels. To avoid potential overfitting,  
340 the decision tree was grown by ranking the importance of sample features and subsequently using  
341 the main features to explicitly classify the training samples. Figure 3 (c) illustrates that the model's  
342 feature utilization was primarily concentrated on the plastic feature peaks (the focused part is  
343 marked with shading). Within the wave number range, the darker the color for each peak, the greater  
344 its contribution to the recognition judgment in the model. The random forest demonstrated superior  
345 feature extraction for  $\sim 729, 1379 \text{ cm}^{-1}$  (PTFE),  $1000 \text{ cm}^{-1}$  (PS),  $635 \text{ cm}^{-1}$  (PVC),  $1442, 1460 \text{ cm}^{-1}$   
346 (PE), and  $599 \text{ cm}^{-1}$  (PMMA). The random forest method exhibits significant advantages over PCA  
347 in effectively classifying overlapping spectra and improving the classification accuracy of nano-  
348 plastic particles, especially when dealing with low SNR spectra. The confusion matrix in Figure  
349 S7(b) demonstrates the classification accuracy of the random forest model for the low-quality  
350 spectra in Figure 3(b), highlighting the suitability of the random forest method for identifying and  
351 differentiating various types of nano-plastics in scientific research.

352 To underscore the merits of the Random Forest algorithm for nanoplastic Raman spectra  
353 classification, we compared its performance with that of Support Vector Machine (SVM) and Back

354 Propagation (BP) Neural Network using identical training and test sets (refer to Figure S8). Each of  
355 the three models underwent 100 tests. The Random Forest algorithm outshined the other methods,  
356 achieving an average accuracy of 98.8%, an average sensitivity of 98.5%, and an average specificity  
357 of 100%. In comparison, the overall recognition accuracy of the BP Neural Network stood at 88.5%,  
358 lower than that of the Random Forest. The SVM's overall recognition accuracy was even more  
359 modest at 68.0%, significantly lagging behind both the Random Forest and BP Neural Network.  
360 This exceptional performance establishes the Random Forest algorithm as the preferred choice for  
361 Raman spectral classification of nanoplastics. Recycled plastics may undergo various treatments  
362 that will alter their spectra, potentially introducing new peaks or causing a significant decrease in  
363 spectral quality that affects the fingerprint of the plastic sample (e.g., recycled PTFE material, as  
364 illustrated in Figure S9(a)). To evaluate the model's ability to handle such variations, we introduced  
365 Raman spectra of recycled PTFE materials into our model for classification. We collected Raman  
366 spectra from 20 recycled PTFE nanoparticle samples and observed that the Random Forest model  
367 effectively classified them into PTFE categories with 100% accuracy (as depicted in Figure S9(b)).  
368 This finding demonstrates that our model exhibits robust resistance to interference and can adapt to  
369 changes in the spectral data of recovered materials.

370



371

372 **Figure 3.** (a) Scatterplot of PC1 and PC2 Scores for Various Nanoplastic Particles. Each point  
 373 corresponds to a Raman spectrum. (b) Low-Quality Raman Spectra of Diverse Nanoplastic Types,  
 374 corresponding to points 1, 2, 3, 4, and 5 in scatterplot (a). (c) Visualization of the Random Forest  
 375 Recognition Model. Gray fill indicates the features extracted by the model, representing the  
 376 classification boundaries differentiating distinct nanoplastic particle categories. Darker shades of  
 377 gray signify more important features. (d) Raman Spectra Identification of Micro- and Nanoplastics  
 378 using the Random Forest Model. The confusion matrix displays the classification percentages for  
 379 the first-order derivative-processed spectral test set across multiple categories.

380

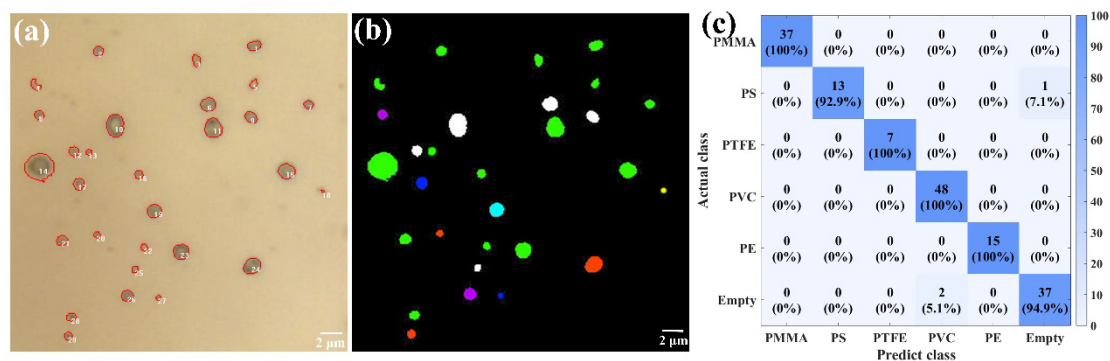
### 381 3.3 Identification of nanoplastics in spiked tap water

382 To demonstrate the practical applicability of the random forest model, the first challenge is to  
 383 distinguish plastics in complex mixtures. For this purpose, we performed a lab test: tap water was

384 spiked with a variety of nanoplastics, and this sample was digested, filtered, concentrated and then  
385 dropped on smooth aluminum sheets for Raman detection (as shown in Figure 1(a)). To achieve a  
386 large range of particle identification, we used automatic image recognition to frame out the location  
387 of the particles under optical microscopy (as shown in Figure 4(a)). Here, the highly reflective  
388 aluminum sheet facilitated large-scale processing and automatic image identification. The framed  
389 particles underwent Raman scanning, and the collected Raman spectra were subsequently analyzed  
390 using our random forest model. The corresponding identification results were subsequently mapped  
391 onto the original image, with different plastics assigned distinct colors, and unknown or "silent"  
392 (exhibiting no signal) particles classified as empty sets. In the example presented in Figure 4(b), a  
393 total of 29 particles were observed within this region (the raw Raman spectra in Figure S10). The  
394 random forest model successfully labeled 2 PMMA particles, 3 PS particles, 15 PVC particles, 2  
395 PTFE particles, 1 PE particle, and 5 unknown particles (including spectra of impurities and spectra  
396 without signal in tap water), along with 1 misclassified particle (had a true classification as an empty  
397 set but was misclassified as PVC). This method can clearly identify a large range of particles and is  
398 applicable to real environmental samples. In this tap water spiked sample, 6 areas were examined  
399 under the optical microscope (Figure 4 (a), Figure S11(a-d)) and Figure S12(a)), each measuring  
400  $35 \times 35 \mu\text{m}$ . A total of 160 particles were observed, as depicted in the particle size distribution in  
401 Figure S12(a), with the majority of the particles being smaller than  $1 \mu\text{m}$ . The presence of particles  
402 with an area larger than  $1 \mu\text{m}^2$  can be attributed to the potential agglomeration of smaller particles  
403 during the concentration process. The type of plastic particles was determined through manual  
404 comparison to establish the true labels. Subsequently, the random forest assigned predicted labels,  
405 and the confusion matrix was utilized to evaluate the performance of the random forest (as depicted  
406 in Figure 4(c)). Based on the confusion matrix calculations, three particles were misclassified (with  
407 the Raman raw data of misclassified particles shown in Figure S13(b-d)). The presence of mixed  
408 signals in Raman spectra due to the adherence of small particles of different substances poses a  
409 challenge for accurate particle classification. In our study, we observed the misclassified particle in  
410 Region No. 4 (as shown in Figure S13(c)) where Raman spectra exhibited a mixture of plastic  
411 particles of PS and PMMA. The particle was incorrectly identified as empty sets by the random  
412 forest model, primarily due to the low respective scores of PS and PMMA. However, we later  
413 confirmed the false negatives using Pearson correlation analysis. The correlation coefficient matrix

414 of the particle compared to the plastic standard sample showed highly significant correlation  
 415 between this sample and the standard samples of PS and PVC (as shown in Figure S13(c)). The  
 416 particles identified as yellow in Region No. 1 and No. 5 exhibited clear bulges at 500-1000  $\text{cm}^{-1}$ ,  
 417 leading to misclassification as PVC (as shown in Figure S13(b) and S13(d)). However, upon closer  
 418 examination, these particles did not display the characteristic features of PVC overall, as evidenced  
 419 by p-values  $> 0.0001$ . Consequently, they were classified as empty sets. In this particular case, our  
 420 random forest model achieved high classification accuracy of approximately 98.1%, with a  
 421 sensitivity of 99.1% and specificity of 94.8%. We also conducted SEM analysis to further validate  
 422 the accuracy of our random forest model. Figure S12 displays the optical microscopy image (a) and  
 423 SEM image (b), as well as the particle classification (c) and corresponding element distribution (d-  
 424 g) of Region No. 6 of tap water nanoplastic spiked samples. The results confirmed that the elemental  
 425 distribution of F matched the position of PTFE identified by the random forest model, and that the  
 426 elemental distribution of C matched the position of PS identified by the random forest model, due  
 427 to the higher C density of PS compared to PTFE.

428



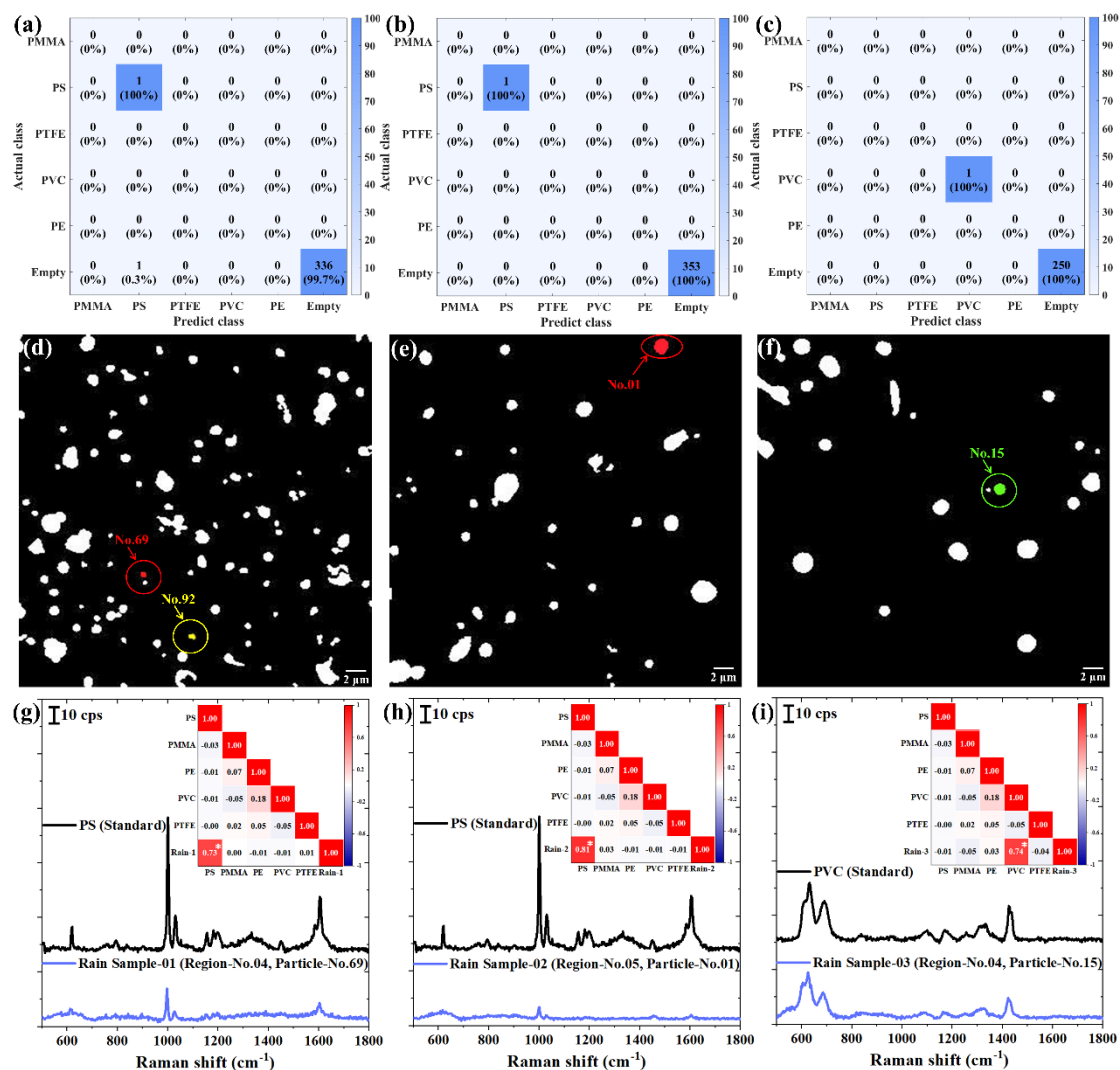
430 **Figure 4.** (a) Automated particle identification under optical microscopy for Region No.1 of tap  
 431 water sample spiked with nanoplastics. (b) Particle classification in (a) using the random forest  
 432 model (blue represents PMMA particles, red represents PS particles, green represents PVC particles,  
 433 purple represents PTFE particles, cyan represents PE particles, white represents unknown particles,  
 434 and yellow represents misclassified particles). (c) Confusion matrix illustrating the classification of  
 435 Raman spectra for particles in the six regions of spiked tap water samples, grouped into different  
 436 percentage classes.

437

### 438 3.4 Identification of nanoplastics in natural rainwater

439 Rainwater has been identified as a potential indicator of the abundance and distribution of  
440 microplastics and nanoplastics in the atmospheric environment, with several studies investigating  
441 the presence of microplastics in rainwater<sup>50-54</sup>. In Wuhan, China, the abundance of microplastics in  
442 rainwater ranged from ~2-19 particles/L, with identified polymer types including PE, PP, PVC, PS,  
443 and other plastic particles smaller than 1 mm<sup>52</sup>. However, samples collected through 37 µm pore  
444 size stainless steel mesh resulted in the loss of submicron and nanoplastics. Similarly, in Seoul,  
445 Korea, rainwater was found to contain 48.1% of plastic particles ranging from 20-50 µm and 34.7%  
446 of plastic particles ranging from 50-100 µm, but the assessment of nanoplastics was neglected due  
447 to the minimum pore size of 25 µm for the employed retention filter membrane<sup>53</sup>. Hence, it is  
448 essential to conduct further investigations into nanoplastics in the environment to gain a  
449 comprehensive understanding of plastic pollution. In our study, we collected environmental  
450 rainwater samples to detect and identify submicron plastics and nanoplastics. We collected rainwater  
451 samples over three consecutive days and followed the pre-processing steps to extract the  
452 nanoplastics, which were dispersed onto the substrate. For each sample, we analyzed an area of 5 ×  
453 35 µm × 35 µm, using image recognition software to identify the particulate matter and performed  
454 Raman detection automatically. Figure S14(a-c) shows that the particle size distribution in the  
455 rainwater is wide, with the vast majority of particles smaller than 1 µm, and approximately 60-70%  
456 of the particles are in the nanoscale range. The collected Raman spectra were analyzed using a  
457 random forest model, which mapped different types of plastics to different colors, while unknown  
458 and no-signal particles were grouped into the empty set category. Figure 5(a-c) presents the results  
459 of the first day of rainwater detection, where 338 particles were detected in rainwater sample 01,  
460 and one was identified as PS (size area of ~0.30 µm<sup>2</sup>), accounting for 0.29%. In rainwater sample  
461 02, 354 particles were detected, and one was identified as PS particles (size area of ~1.50 µm<sup>2</sup>),  
462 accounting for 0.28%. In rainwater sample 03, 251 particles were detected, and one was identified  
463 as PVC (size area of ~0.90 µm<sup>2</sup>), accounting for 0.40%. The regions where the nanoplastics  
464 identified by the random forest are located, are shown in Figure 5(d-f). As illustrated in Figure 5(g-  
465 i), the three particles identified as plastic exhibited characteristic peaks in their Raman spectra that  
466 matched those of the corresponding standard plastic samples. Moreover, the corresponding

467 correlation coefficients were all greater than 0.7 (P-values < 0.0001), which is consistent with the  
468 random forest model's judgment. However, no nanoplastic particles were detected in the rainwater  
469 samples collected on the subsequent two days, primarily due to the first flushing effect of rainwater<sup>53</sup>.  
470 As the duration of rainfall increased, more atmospheric pollutants were deposited. Other particulate  
471 matter in rainwater was predominantly composed of carbon-based soot-like particles and salt  
472 particles. We collected Raman spectra of common salts in rainwater, including NaCl, NaSO<sub>4</sub>, and  
473 CaCO<sub>3</sub>, and used the random forest model for classification. All these spectra were accurately  
474 classified as "non-plastic" and added to the empty set, demonstrating the ability of our algorithm to  
475 distinguish plastic particles from salt crystals (as shown in Figure S15). During the analysis of  
476 rainwater sample 01, a non-plastic particle in region No.04 was misclassified and marked in yellow  
477 in Figure 5(d). To eliminate false positives, we utilized Pearson correlation analysis. The raw Raman  
478 spectrum of the misclassified particle is shown in Figure S16, exhibiting a clear bulge dominated  
479 by a carbon peak at 1000-1800 cm<sup>-1</sup>. Further inspection revealed that it did not exhibit the overall  
480 characteristics of PS (p-value > 0.0001) and was therefore classified as empty set. All data were  
481 manually identified, and the random forest model achieved a high classification accuracy of over  
482 99% in the analysis of rainwater samples, as confirmed by the confusion matrix calculations. Our  
483 results demonstrate that machine learning techniques are promising and applicable for identifying  
484 environmental nanoplastics. The model can efficiently identify relevant plastic data from large  
485 datasets and reduce manual processing time. Further refinement and expansion of the spectral  
486 datasets may improve the accuracy of the model in identifying a wider range of plastic particles.  
487  
488  
489



490

491 **Figure 5.** Environmental tests for identification in natural rainwater. The confusion matrix shows  
 492 that the Raman spectra of all particles in the selected areas of rainwater samples 01 (a), 02 (b), and  
 493 03 (c) are classified into different percentage classes. The particles in rainwater sample 01-Region  
 494 04 (d), rainwater sample 02-Region 05 (e), and rainwater sample 03-Region 04 (f) were identified  
 495 by the random forest model (blue for PMMA particles, red for PS particles, green for PVC particles,  
 496 purple for PTFE particles, cyan for PE particles, white for unknown particles, and yellow for  
 497 misclassified particles). (g) Raman spectrum of particle No. 69 in Region 04 of rainwater sample  
 498 01 (classified as PS), and Pearson correlation coefficient matrix of the particles compared to the  
 499 plastic standard sample. (h) Raman spectrum of particle No. 01 in Region 05 of rainwater sample  
 500 02 (classified as PS) and Pearson correlation coefficient matrix of the particles compared to the  
 501 plastic standard sample. (i) Raman spectrum of particle No. 15 in Region 04 of rainwater sample 03  
 502 (classified as PVC) and Pearson correlation coefficient matrix of particles compared to the plastic

503 standard samples. (P-values < 0.0001 are indicated by \*, indicating a highly significant correlation  
504 between this sample and the plastic standard sample).

505

#### 506 **4. Environmental Implication**

507 The increasing prevalence of nanoplastic particles in the environment highlights the urgent need for  
508 effective detection and monitoring strategies<sup>55-60</sup>. However, current sampling methods and Raman  
509 detection techniques mainly focus on plastic particles in the micrometer size range<sup>22, 23, 50-54, 61</sup>,  
510 limiting the detection of nanoplastic particles. Therefore, further investigation of nanoplastics in the  
511 environment is necessary to gain a comprehensive understanding of plastic pollution. Although  
512 Machine learning has been applied to many studies on microplastic identification<sup>32-35</sup>, existing data  
513 systems typically use high-quality Raman spectra of plastic samples, with strong signals and clear  
514 single peaks, which are not suitable for identifying nanoplastic particles. The significant signal loss  
515 caused by the dramatic reduction in nanoplastic particle size<sup>10, 36</sup> makes it challenging to match their  
516 spectra to those of macroscopic plastic samples. As a result, relying on commercial data systems to  
517 identify nanoplastic particles may result in poor matching accuracy and require manual evaluation  
518 by researchers. Further research is needed to develop and optimize Raman detection techniques for  
519 identifying nanoplastic particles in various environmental samples.

520 In summary, our study provides a promising solution to this problem by demonstrating the  
521 potential of Raman spectroscopy and machine learning algorithms in identifying and characterizing  
522 nanoplastic particles. The combination of Raman spectroscopy and multivariate analysis provides  
523 an information-rich and robust method that overcomes the limitations of conventional spectral  
524 analysis for individual peak determination of micro- and nanoplastics. In our study, we developed  
525 five Raman spectral datasets of nanoplastic particles, used advanced data pre-processing techniques,  
526 and applied the random forest algorithm to achieve high classification accuracy. Our proposed  
527 random forest algorithm outperformed other methods such as BP, SVM, and PCA, achieving an  
528 average accuracy of 98.8%, an average sensitivity of 98.5%, and an average specificity of 100%.  
529 As a laboratory test, our method achieved 97.9% recognition accuracy when analyzing tap water  
530 samples with added plastic particles. Furthermore, the applicability of our algorithm to real-world

531 environmental samples was demonstrated through experiments on rainwater, where nanoscale PS  
532 and PVC were detected.

533 As a novel contribution, our work successfully addresses the challenge of identifying low-  
534 quality nanoplastic Raman spectra and dealing with complex environmental samples. By  
535 incorporating spectra with weaker signals, we improve the classification accuracy of nanoplastic  
536 particles. We validate our proposed random forest algorithm by including reference standards for  
537 recycled plastics and salts. The comprehensiveness of the plastics and additives included in the  
538 spectral database determines the success of qualitative matching<sup>62</sup>. Therefore, further research is  
539 necessary to develop a comprehensive nanoplastic database that considers different compositions  
540 of plastic mixtures with additives and chemical structure changes during aging<sup>63, 64</sup>. Machine  
541 learning algorithms applied to nanoplastic identification offer a promising strategy to improve the  
542 accuracy and efficiency of plastic pollution monitoring in the environment. Our results demonstrate  
543 the value of machine learning in developing effective detection and monitoring strategies for  
544 nanoplastic particles, which are crucial for understanding the extent of plastic pollution in the  
545 environment.

546

#### 547 **Declaration of Competing Interest**

548 The authors declare no conflict of interest.

549

#### 550 **Acknowledgment**

551 The authors gratefully acknowledge financial support from National Natural Science Foundation of  
552 China (No. 22176036, No. 21976030 and No. 22006020), the Natural Science Foundation of  
553 Shanghai (No. 19ZR1471200). V.K.V. acknowledges support from the Royal Society through the  
554 University Research Fellowships and the Royal Society grants PEF1\170015 and RGF\EA\180228,  
555 as well as the EPSRC grant EP/T001046/1. V.K.V and L.Z. acknowledge the International  
556 Collaboration Awards 2020 of the Royal Society (No. ICA\R1\201088).

557

558 **Additional information**

559 **Supplementary Data.** Links to supporting information for the remaining data will be provided upon  
560 receipt.

561 **Correspondence and requests for materials** should be addressed to L.Z.

562

563 **References**

- 564 1. Allen, S.; Allen, D.; Phoenix, V. R.; Le Roux, G.; Durántez Jiménez, P.; Simonneau, A.; Binet, S.;  
565 Galop, D., Atmospheric transport and deposition of microplastics in a remote mountain catchment.  
566 *Nature Geoscience* **2019**, *12*, (5), 339-344.
- 567 2. Feng, S.; Lu, H.; Liu, Y., The occurrence of microplastics in farmland and grassland soils in the  
568 Qinghai-Tibet plateau: Different land use and mulching time in facility agriculture. *Environ. Pollut.*  
569 **2021**, *279*, 116939.
- 570 3. Bessa, F.; Ratcliffe, N.; Otero, V.; Sobral, P.; Marques, J. C.; Waluda, C. M.; Trathan, P. N.; Xavier,  
571 J. C., Microplastics in gentoo penguins from the Antarctic region. *Sci. Rep.* **2019**, *9*, (1), 14191.
- 572 4. Mitrano, D. M.; Wick, P.; Nowack, B., Placing nanoplastics in the context of global plastic pollution.  
573 *Nat. Nanotechnol.* **2021**, *16*, (5), 491-500.
- 574 5. Baldwin, A. K.; Corsi, S. R.; Mason, S. A., Plastic Debris in 29 Great Lakes Tributaries: Relations  
575 to Watershed Attributes and Hydrology. *Environ. Sci. Technol.* **2016**, *50*, (19), 10377-10385.
- 576 6. Hu, L. L.; Chernick, M.; Hinton, D. E.; Shi, H. H., Microplastics in Small Waterbodies and Tadpoles  
577 from Yangtze River Delta, China. *Environ. Sci. Technol.* **2018**, *52*, (15), 8885-8893.
- 578 7. Schmidt, C.; Krauth, T.; Wagner, S., Export of Plastic Debris by Rivers into the Sea. *Environ. Sci.*  
579 *Technol.* **2017**, *51*, (21), 12246-12253.
- 580 8. Ding, J. F.; Jiang, F. H.; Li, J. X.; Wang, Z. X.; Sun, C. J.; Wang, Z. Y.; Fu, L.; Ding, N. X.; He, C.  
581 F., Microplastics in the Coral Reef Systems from Xisha Islands of South China Sea. *Environ. Sci.*  
582 *Technol.* **2019**, *53*, (14), 8036-8046.
- 583 9. Zheng, Y. F.; Li, J. X.; Cao, W.; Liu, X. H.; Jiang, F. H.; Ding, J. F.; Yin, X. F.; Sun, C. J., Distribution  
584 characteristics of microplastics in the seawater and sediment: A case study in Jiaozhou Bay, China.  
585 *Sci. Total. Environ.* **2019**, *674*, 27-35.
- 586 10. Xu, G.; Cheng, H.; Jones, R.; Feng, Y.; Gong, K.; Li, K.; Fang, X.; Tahir, M. A.; Valev, V. K.; Zhang,  
587 L., Surface-Enhanced Raman Spectroscopy Facilitates the Detection of Microplastics <1 µm in the  
588 Environment. *Environ. Sci. Technol.* **2020**, *54*, (24), 15594-15603.
- 589 11. Morales, A. C.; Tomlin, J. M.; West, C. P.; Rivera-Adorno, F. A.; Peterson, B. N.; Sharpe, S. A. L.;  
590 Noh, Y.; Sendesi, S. M. T.; Boor, B. E.; Howarter, J. A.; Moffet, R. C.; China, S.; O'Callahan, B. T.;  
591 El-Khoury, P. Z.; Whelton, A. J.; Laskin, A., Atmospheric emission of nanoplastics from sewer pipe  
592 repairs. *Nat. Nanotechnol.* **2022**, *17*, (11), 1171-1177.
- 593 12. Ter Halle, A.; Jeanneau, L.; Martignac, M.; Jardé, E.; Pedrono, B.; Brach, L.; Gigault, J.,  
594 Nanoplastic in the North Atlantic Subtropical Gyre. *Environ. Sci. Technol.* **2017**, *51*, (23), 13689-  
595 13697.

- 596 13. Materic, D.; Kasper-Giebl, A.; Kau, D.; Anten, M.; Greilinger, M.; Ludewig, E.; van Sebille, E.;  
597 Rockmann, T.; Holzinger, R., Micro- and Nanoplastics in Alpine Snow: A New Method for  
598 Chemical Identification and (Semi)Quantification in the Nanogram Range. *Environ. Sci. Technol.*  
599 **2020**, *54*, (4), 2353-2359.
- 600 14. Wahl, A.; Le Juge, C.; Davranche, M.; El Hadri, H.; Grassl, B.; Reynaud, S.; Gigault, J., Nanoplastic  
601 occurrence in a soil amended with plastic debris. *Chemosphere* **2021**, *262*, 127784.
- 602 15. Peng, L.; Fu, D.; Qi, H.; Lan, C. Q.; Yu, H.; Ge, C., Micro- and nano-plastics in marine environment:  
603 Source, distribution and threats - A review. *Sci. Total. Environ.* **2020**, *698*, 134254.
- 604 16. Ribeiro, F.; O'Brien, J. W.; Galloway, T.; Thomas, K. V., Accumulation and fate of nano- and micro-  
605 plastics and associated contaminants in organisms. *TRAC-TREND anal.chem.* **2019**, *111*, 139-147.
- 606 17. Leslie, H. A.; van Velzen, M. J. M.; Brandsma, S. H.; Vethaak, A. D.; Garcia-Vallejo, J. J.; Lamoree,  
607 M. H., Discovery and quantification of plastic particle pollution in human blood. *Environ. Int.* **2022**.
- 608 18. Schmidt, S., Breach of Security? Placental Uptake of Micro- and Nanoplastic Particles. *Environ*  
609 *Health Perspect* **2022**, *130*, (11), 114001.
- 610 19. Wagner, S.; Reemtsma, T., Things we know and don't know about nanoplastic in the environment.  
611 *Nat. Nanotechnol.* **2019**, *14*, (4), 300-301.
- 612 20. Käßler, A.; Fischer, D.; Oberbeckmann, S.; Schernewski, G.; Labrenz, M.; Eichhorn, K.-J.; Voit,  
613 B., Analysis of environmental microplastics by vibrational microspectroscopy: FTIR, Raman or  
614 both? *Analytical and Bioanalytical Chemistry* **2016**, *408*, (29), 8377-8391.
- 615 21. Lenz, R.; Enders, K.; Stedmon, C. A.; Mackenzie, D. M. A.; Nielsen, T. G., A critical assessment of  
616 visual identification of marine microplastic using Raman spectroscopy for analysis improvement.  
617 *Mar. Pollut. Bull.* **2015**, *100*, (1), 82-91.
- 618 22. Wright, S. L.; Levermore, J. M.; Kelly, F. J., Raman Spectral Imaging for the Detection of Inhalable  
619 Microplastics in Ambient Particulate Matter Samples. *Environ. Sci. Technol.* **2019**, *53*, (15), 8947-  
620 8956.
- 621 23. Loeder, M. G. J.; Gerdts, G., *Methodology Used for the Detection and Identification of*  
622 *Microplastics-A Critical Appraisal*. 2015; p 201-227.
- 623 24. Sobhani, Z.; Zhang, X.; Gibson, C.; Naidu, R.; Mallavarapu, M.; Fang, C., Identification and  
624 visualisation of microplastics/nanoplastics by Raman imaging(i): down to 100 nm. *Water Res.* **2020**,  
625 *174*.
- 626 25. Fan, X.-g.; Zeng, Y.; Zhi, Y.-L.; Nie, T.; Xu, Y.-j.; Wang, X., Signal-to-noise ratio enhancement for  
627 Raman spectra based on optimized Raman spectrometer and convolutional denoising autoencoder.  
628 *J. Raman. Spectrosc.* **2021**, *52*, (4), 890-900.
- 629 26. Yu, F.; Hu, X., Machine learning may accelerate the recognition and control of microplastic  
630 pollution: Future prospects. *J. Hazard Mater.* **2022**, *432*, 128730.
- 631 27. Luo, Y.; Zhang, X.; Zhang, Z.; Naidu, R.; Fang, C., Dual-Principal Component Analysis of the  
632 Raman Spectrum Matrix to Automatically Identify and Visualize Microplastics and Nanoplastics.  
633 *Analytical Chemistry* **2022**, *94*, (7), 3150-3157.
- 634 28. Fang, C.; Luo, Y.; Zhang, X.; Zhang, H.; Nolan, A.; Naidu, R., Identification and visualisation of  
635 microplastics via PCA to decode Raman spectrum matrix towards imaging. *Chemosphere* **2022**, *286*,  
636 131736.
- 637 29. Luo, Y.; Gibson, C. T.; Chuah, C.; Tang, Y.; Naidu, R.; Fang, C., Raman imaging for the  
638 identification of Teflon microplastics and nanoplastics released from non-stick cookware. *Sci. Total.*  
639 *Environ.* **2022**, *851*, 158293.

- 640 30. Luo, Y.; Zhang, Z.; Naidu, R.; Zhang, X.; Fang, C., Raman imaging of microplastics and  
641 nanoplastics released from the printed toner powders burned by a mimicked bushfire. *Sci. Total.*  
642 *Environ.* **2022**, *849*, 157686.
- 643 31. Luo, Y.; Qi, F.; Gibson, C. T.; Lei, Y.; Fang, C., Investigating kitchen sponge-derived microplastics  
644 and nanoplastics with Raman imaging and multivariate analysis. *Sci. Total. Environ.* **2022**, *824*,  
645 153963.
- 646 32. Jin, N.; Song, Y.; Ma, R.; Li, J.; Li, G.; Zhang, D., Characterization and identification of  
647 microplastics using Raman spectroscopy coupled with multivariate analysis. *Analytica Chimica*  
648 *Acta* **2022**, *1197*, 339519.
- 649 33. da Silva, D. J.; Parra, D. F.; Wiebeck, H., Applying confocal Raman spectroscopy and different  
650 linear multivariate analyses to sort polyethylene residues. *Chem. Eng. J.* **2021**, *426*, 131344.
- 651 34. Lin, J.-y.; Liu, H.-t.; Zhang, J., Recent advances in the application of machine learning methods to  
652 improve identification of the microplastics in environment. *Chemosphere* **2022**, *307*, 136092.
- 653 35. Brandt, J.; Mattsson, K.; Hassellöv, M., Deep Learning for Reconstructing Low-Quality FTIR and  
654 Raman Spectra—A Case Study in Microplastic Analyses. *Analytical Chemistry* **2021**, *93*, (49),  
655 16360-16368.
- 656 36. Winkler, A.; Fumagalli, F.; Cella, C.; Gilliland, D.; Tremolada, P.; Valsesia, A., Detection and  
657 formation mechanisms of secondary nanoplastic released from drinking water bottles. *Water Res.*  
658 **2022**, *222*, 118848.
- 659 37. Agjee, N. e. H.; Mutanga, O.; Peerbhay, K.; Ismail, R., The Impact of Simulated Spectral Noise on  
660 Random Forest and Oblique Random Forest Classification Performance. *Journal of Spectroscopy*  
661 **2018**, *2018*, 8316918.
- 662 38. Breiman, L., Random Forests. *Machine Learning* **2001**, *45*, (1), 5-32.
- 663 39. Oruganti, R. K.; Biji, A. P.; Lanuyanger, T.; Show, P. L.; Sriariyanun, M.; Upadhyayula, V. K. K.;  
664 Gadhamshetty, V.; Bhattacharyya, D., Artificial intelligence and machine learning tools for high-  
665 performance microalgal wastewater treatment and algal biorefinery: A critical review. *Sci. Total.*  
666 *Environ.* **2023**, *876*, 162797.
- 667 40. Balakrishnan, V.; Kehrabi, Y.; Ramanathan, G.; Paul, S. A.; Tiong, C. K., Machine learning  
668 approaches in diagnosing tuberculosis through biomarkers - A systematic review. *Progress in*  
669 *Biophysics and Molecular Biology* **2023**, *179*, 16-25.
- 670 41. Luo, S.-h.; Wang, X.; Chen, G.-y.; Xie, Y.; Zhang, W.-h.; Zhou, Z.-f.; Zhang, Z.-m.; Ren, B.; Liu,  
671 G.-k.; Tian, Z.-q., Developing a Peak Extraction and Retention (PEER) Algorithm for Improving  
672 the Temporal Resolution of Raman Spectroscopy. *Analytical Chemistry* **2021**, *93*, (24), 8408-8413.
- 673 42. Morais, C. L. M.; Paraskevaidi, M.; Cui, L.; Fullwood, N. J.; Isabelle, M.; Lima, K. M. G.; Martin-  
674 Hirsch, P. L.; Sreedhar, H.; Trevisan, J.; Walsh, M. J.; Zhang, D.; Zhu, Y.-G.; Martin, F. L.,  
675 Standardization of complex biologically derived spectrochemical datasets. *Nature Protocols* **2019**,  
676 *14*, (5), 1546-1577.
- 677 43. Paraskevaidi, M.; Morais, C. L. M.; Lima, K. M. G.; Snowden, J. S.; Saxon, J. A.; Richardson, A.  
678 M. T.; Jones, M.; Mann, D. M. A.; Allsop, D.; Martin-Hirsch, P. L.; Martin, F. L., Differential  
679 diagnosis of Alzheimer's disease using spectrochemical analysis of blood. *Proceedings of the*  
680 *National Academy of Sciences* **2017**, *114*, (38), E7929-E7938.
- 681 44. Sato, H.; Shimoyama, M.; Kamiya, T.; Amari, T.; Šašić, S.; Ninomiya, T.; Siesler, H. W.; Ozaki, Y.,  
682 Raman spectra of high-density, low-density, and linear low-density polyethylene pellets and  
683 prediction of their physical properties by multivariate data analysis. *Journal of Applied Polymer*

- 684 *Science* **2002**, *86*, (2), 443-448.
- 685 45. Menezes, D. B.; Reyer, A.; Marletta, A.; Musso, M., Glass transition of polystyrene (PS) studied by  
686 Raman spectroscopic investigation of its phenyl functional groups. *Materials Research Express*  
687 **2017**, *4*, (1), 015303.
- 688 46. Xu, X.; Ming, H.; Zhang, Q.; Zhang, Y., Properties of Raman spectra and laser-induced  
689 birefringence in polymethyl methacrylate optical fibres. *Journal of Optics A: Pure and Applied*  
690 *Optics* **2002**, *4*, (3), 237.
- 691 47. Rusen, E.; Marculescu, B.; Butac, L.; Preda, N.; Mihut, L., The Synthesis and Characterization of  
692 Poly Vinyl Chloride Chemically Modified with C60. *Fullerenes, Nanotubes and Carbon*  
693 *Nanostructures* **2008**, *16*, (3), 178-185.
- 694 48. Mihaly, J.; Sterkel, S.; Ortner, H.; Kocsis, L.; Hajba, L.; Furdyga, É.; Mink, J., FTIR and FT-Raman  
695 Spectroscopic Study on Polymer Based High Pressure Digestion Vessels. *Croatica Chemica Acta*  
696 *(CCA@chem.pmf.hr)*; Vol.79 No.3 **2006**, 79.
- 697 49. Barton, S. J.; Ward, T. E.; Hennelly, Bryan M., Algorithm for optimal denoising of Raman spectra.  
698 *Analytical Methods* **2018**, *10*, (30), 3759-3769.
- 699 50. Yuan, Z.; Pei, C.; Li, H.; Lin, L.; Liu, S.; Hou, R.; Liao, R.; Xu, X., Atmospheric microplastics at a  
700 southern China metropolis: Occurrence, deposition flux, exposure risk and washout effect of rainfall.  
701 *Sci. Total. Environ.* **2023**, *869*, 161839.
- 702 51. Wu, J.; Ye, Q.; Sun, L.; Liu, J.; Huang, M.; Wang, T.; Wu, P.; Zhu, N., Impact of persistent rain on  
703 microplastics distribution and plastisphere community: A field study in the Pearl River, China. *Sci.*  
704 *Total. Environ.* **2023**, *879*, 163066.
- 705 52. Sang, W.; Chen, Z.; Mei, L.; Hao, S.; Zhan, C.; Zhang, W. b.; Li, M.; Liu, J., The abundance and  
706 characteristics of microplastics in rainwater pipelines in Wuhan, China. *Sci. Total. Environ.* **2021**,  
707 *755*, 142606.
- 708 53. Do, T.; Park, Y.; Lim, B.; Kim, S.; Chae, M.-Y.; Chun, C.-H., Effect of the first-flush phenomenon  
709 on the quantification of microplastics in rainwater. *Mar. Pollut. Bull.* **2023**, *187*, 114559.
- 710 54. Liu, J.; Xu, G.; Ruan, X.; Li, K.; Zhang, L., V-shaped substrate for surface and volume enhanced  
711 Raman spectroscopic analysis of microplastics. *Frontiers of Environmental Science & Engineering*  
712 **2022**, *16*, (11), 143.
- 713 55. Ferreira, I.; Venâncio, C.; Lopes, I.; Oliveira, M., Nanoplastics and marine organisms: What has  
714 been studied? *Environ. Toxicol. Pharmacol.* **2019**, *67*, 1-7.
- 715 56. Gigault, J.; Halle, A. t.; Baudrimont, M.; Pascal, P.-Y.; Gauffre, F.; Phi, T.-L.; El Hadri, H.; Grassl,  
716 B.; Reynaud, S., Current opinion: What is a nanoplastic? *Environ. Pollut.* **2018**, *235*, 1030-1034.
- 717 57. Lambert, S.; Wagner, M., Characterisation of nanoplastics during the degradation of polystyrene.  
718 *Chemosphere* **2016**, *145*, 265-8.
- 719 58. Gigault, J.; Pedrono, B.; Maxit, B.; Ter Halle, A., Marine plastic litter: the unanalyzed nano-fraction.  
720 *Environ. Sci. Nano* **2016**, *3*, (2), 346-350.
- 721 59. Jambeck Jenna, R.; Geyer, R.; Wilcox, C.; Siegler Theodore, R.; Perryman, M.; Andrady, A.;  
722 Narayan, R.; Law Kara, L., Plastic waste inputs from land into the ocean. *Science* **2015**, *347*, (6223),  
723 768-771.
- 724 60. Xie, L.; Gong, K.; Liu, Y.; Zhang, L., Strategies and Challenges of Identifying Nanoplastics in  
725 Environment by Surface-Enhanced Raman Spectroscopy. *Environ. Sci. Technol.* **2023**, *57*, (1), 25-  
726 43.
- 727 61. Koelmans, A. A.; Mohamed Nor, N. H.; Hermsen, E.; Kooi, M.; Mintenig, S. M.; De France, J.,

- 728 Microplastics in freshwaters and drinking water: Critical review and assessment of data quality.  
729 *Water Res.* **2019**, *155*, 410-422.
- 730 62. Munno, K.; De Frond, H.; O'Donnell, B.; Rochman, C. M., Increasing the Accessibility for  
731 Characterizing Microplastics: Introducing New Application-Based and Spectral Libraries of Plastic  
732 Particles (SLoPP and SLoPP-E). *Analytical Chemistry* **2020**, *92*, (3), 2443-2451.
- 733 63. Li, J.; Wang, L.; Xu, Z.; Zhang, J.; Li, J.; Lu, X.; Yan, R.; Tang, Y., A new point to correlate the  
734 multi-dimensional assessment for the aging process of microfibers. *Water Res.* **2023**, *235*, 119933.
- 735 64. Song, Y. K.; Hong, S. H.; Eo, S.; Shim, W. J., Fragmentation of nano- and microplastics from virgin-  
736 and additive-containing polypropylene by accelerated photooxidation. *Environ. Pollut.* **2023**, *327*,  
737 121590.
- 738