analytical chemistry

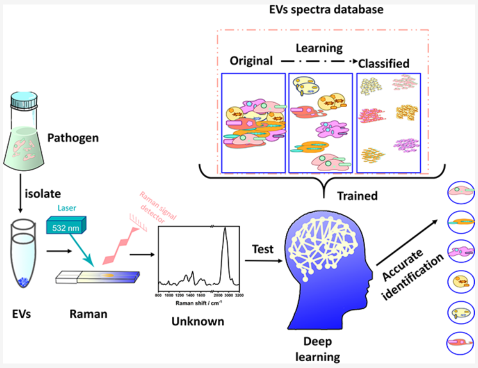
pubs.acs.org/ac

Article

Deep Learning-Enabled Raman Spectroscopic Identification of Pathogen-Derived Extracellular Vesicles and the Biogenesis Process

Yi-Fei Qin,[⊥] Xin-Yu Lu,[⊥] Zheng Shi, Qian-Sheng Huang, Xiang Wang, Bin Ren,* and Li Cui*

Cite This: https://doi.org/10.1021/acs.analchem.2c02226



ABSTRACT: Pathogenic bacterial infections, exacerbated by increasing antimicrobial resistance, pose a major threat to human health worldwide. Extracellular vesicles (EVs), secreted by bacteria and acting as their "long-distance weapons", play an important role in the occurrence and development of infectious diseases. However, no efficient methods to rapidly detect and identify EVs of different bacterial origins are available. Here, label-free Raman spectroscopy in combination with a new deep learning model of the attentional neural network (aNN) was developed to identify pathogen-derived EVs at Gram$^{\pm}$, species, strain, and even down to physiological levels. By training the aNN model with a large Raman data set from six typical pathogen-derived EVs, we achieved the identification of EVs with high accuracies at all levels: exceeding 96% at the Gram and species levels, 93% at the antibiotic-resistant and sensitive strain levels, and still above 87% at the physiological level. aNN enabled Raman spectroscopy to interrogate the bacterial origin of EVs to a much higher level than previous methods. Moreover, spectral markers underpinning EV discrimination were uncovered from subtly different EV spectra via an interpretation algorithm of the integrated gradient. A further comparative analysis of the rich Raman biochemical signatures of EVs and parental pathogens clearly revealed the biogenesis process of EVs, including the selective encapsulation of biocomponents and distinct membrane compositions from the original bacteria. This developed platform provides an accurate and versatile means to identify pathogen-derived EVs, spectral markers, and the biogenesis process. It will promote rapid diagnosis and allow the timely treatment of bacterial infections.

■ INTRODUCTION

Bacterial infections, especially antibiotic-resistant (AMR) pathogen infections, are listed as the top 10 threats to global health by the World Health Organization.[1,2] In 2019, the global burden of infections with drug-resistance-related pathogens was estimated at 4.95 million deaths, of which 1.27 million deaths were directly attributable to AMR.[3] During infections, pathogenic bacteria-derived extracellular vesicles (EVs) have been regarded as bacterial "long-distance weapons" in initiating pathological events.[4,5] EVs are nanosized lipid bilayer spheres with a size ranging from 30 to 400 nm.[6] They are naturally secreted from cells in all domains of life, including bacteria, mammalian cells, archaea, and fungi.[7] Numerous studies have shown that EVs are abundantly released from pathogenic bacteria and harbor the components from the parental pathogens, such as proteins, nucleic acids, carbohydrates, lipids, and virulence factors.[8,9] Moreover, these nanoscale particles are small enough to cross the blood−brain, placental, and air−blood barriers that bacteria cannot reach to deliver bacterial toxins to host cells, eliciting the host immune system and potentially accelerating bacterial infections.[5,10,11] For example, pathogen-derived EVs have been found to permeate the blood−brain barrier of infected mice[12] and to be present in body fluids and tissue biopsies, such as in the cerebrospinal fluid of patients with meningococcal disease.[10] In addition, EVs derived from pulmonary commensal microbes were reported to deliver lipopolysaccharide (LPS) and lipoprotein contents to macrophages, inducing interleukin-17β expression and promoting pulmonary fibrosis.[13] In another study, EVs from pathogenic *Neisseria gonorrheae*, *Escherichia coli*, and *Pseudomonas aeruginosa* caused mitochondrial dysfunction and immune responses in macrophages.[14] Because pathogen-derived EVs play a leading role in pathological initiation and are widely present in body fluids, they have great potential as a diagnostic test for bacterial infections. Therefore, rapid and accurate identification of EVs of different bacterial origins is very important for the early diagnosis of infection, timely intervention, and treatment guidance.

However, unlike mammalian cell-derived EVs with specific membrane protein markers such as ALIX, TSG101, CD9, and CD63,[15] specific recognition components in pathogen-

Received: May 24, 2022
Accepted: August 16, 2022

originated EVs are largely lacking or obscure, hindering the employment and development of efficient identification strategies. Currently, conventional methods used to detect and identify cancer cell- or tissue-derived EVs, such as Western blot and enzyme-linked immunosorbent assays, cannot work on bacterial EVs.[16] In some cases, Gram classification of bacterial EVs is possible on the basis of the unique surface component of lipoteichoic acid and LPS in Gram-positive and Gram-negative bacteria,[17] but a high taxonomic classification is not allowed. In addition, although 16S rRNA-based genomic sequencing can be used to identify bacteria at the genus and even species level,[18] it is not applicable for bacteria-derived EVs because there are no gene databases available for comparison with bacteria EVs. Moreover, EVs do not always harbor enough genetic material for identification. Accordingly, there remains a critical technical hurdle in identifying EVs from different pathogen sources.

Raman spectroscopy is emerging as a valuable tool for the label-free, rapid, and sensitive identification of biological samples.[19] Raman spectroscopy can provide an intrinsic biochemical profile of cells, including nucleic acids, proteins, lipids, carbohydrates, pigments, and metabolites,[20] by exciting the samples with a laser to obtain their fingerprint vibrational information in an inelastic way. This information provides an exquisite molecular basis for resolving the cellular identity. Through the combination with state-of-the-art machine learning and deep learning algorithms, subtle spectral difference can be captured from complex Raman spectra with improved efficiency, enabling the differentiation of different biological samples by Raman spectroscopy. Currently, Raman spectroscopy was assisted by a support vector machine (SVM) to identify pathogenic *Legionella* at the genus, species, and down to subspecies level.[21] Recently, Raman spectroscopy was coupled with a convolutional neural network (CNN)-based deep learning approach to accurately identify 30 bacterial pathogens, including methicillin-resistant and susceptible strains, and validate the results on clinical bacteria isolated from patients.[22] A combination of Raman spectroscopy and residual neural network (ResNet)-based deep learning was able to achieve a binary classification of EVs derived from normal and lung cancer cells.[23] However, Raman spectroscopy has never been explored for discerning bacterial pathogen-derived EVs, whose compositions are distinct from those of human cells. In addition, the diagnosis and treatment of bacteria-induced infections require the identification of pathogens at the species, strain (e.g., drug-resistant and sensitive), and even down to physiological levels. The discrimination of bacteria with increased similarity is becoming more challenging, and it is still unclear about the power of Raman spectroscopy for discriminating bacteria-derived EVs at various taxonomic or physiological levels. Moreover, although EV biogenesis is generally regarded as the process that involves the shuttling of cellular contents from donor bacteria to EVs,[5,24] a detailed understanding of the substances loaded in EVs is still lacking. The rich chemical information on Raman spectra may provide a clue to interrogating the biogenesis process of EVs.

In this work, a total of six typical clinical bacterial pathogens and their derived EVs were collected and extracted. A large Raman database including 4335 spectra from 3 independent measurements of EVs was obtained. To explore the power of Raman spectroscopy in resolving EVs to different levels, a new deep learning model of the attentional neural network (aNN) and the feature extraction algorithm of the integrated gradient (IG) were developed and employed to improve the identification accuracy and interpretation capability, respectively. Their performance was compared to that of five other machine/deep learning algorithms previously reported for bacteria identification. By using these techniques, we aim to (1) explore the capacity of Raman spectroscopy in identifying EVs from different pathogen origins at different hierarchies from Gram-positive (G^+) and Gram-negative (G^-), species, strains, and down to physiological states; (2) explore the potential spectral markers of different EVs with the newly developed interpretation algorithm; and (3) reveal the biogenesis process of EVs by systematically comparing the spectral signatures of EVs with their host pathogens. This work provides a new and accurate means for identifying pathogen-derived EVs and their generation process. Combined with the single-particle analysis platform, our method is expected to be applied to clinical samples (blood, urine, and sputum). It holds great promise for the early diagnosis of bacterial source of infection, guiding timely treatment and answering the fundamental question of EV-triggered infection.

■ METHODS

Bacterial Species and Growth Conditions. The information on bacterial species is listed in Table S1. Bacteria were cultured at 37 °C and 150 rpm overnight. Unless explicitly stated, all chemicals used in the cultivation process were purchased from Sinopharm Chemical Reagent Co., China.

EV Isolation. To isolate EVs from bacteria, the culture medium (500 mL) was centrifuged first at 8000*g* for 20 min to remove detached cells. The resulting supernatant was then concentrated with an ultrafiltration centrifugal tube (100-KDa, Millipore, USA) to reach a volume of 10 mL. Then the medium was centrifuged at 12 000*g* for 45 min, processed by 0.22 μm membrane filtration (Millipore, USA), and ultracentrifuged at 120 000*g* for 90 min. Pellets were resuspended, pooled in 12 mL of sterile water instead, and reultracentrifuged at 120 000*g* for 90 min at 4 °C. The EVs were obtained by suspending the pellet in 200 μL of sterile water.

TEM and nFCM Analysis. The characterization of EVs with transmission electron microscopy (TEM) was performed according to our previous report with some modifications.[25] The EVs' concentration and size distribution were quantified using the Nanoflow cytometer (N30E, NanoFCM).[26] A NanoFCM silica nanospheres cocktail containing a mixture of silica nanospheres (diameters: 68 to 155 nm) was used as a size reference. For the EVs' concentration calculation, a NanoFCM silica nanosphere cocktail at different concentrations was first measured to create a standard working curve. EVs diluted with PBS were then measured, and the particle concentration was calculated using the standard working curve.

Raman Spectra Acquisition and Data Processing. All bacterial samples were washed with sterile water three times to remove residual media. Isolated EV samples can be directly used for detection. An aliquot of 3 μL of sample was dropped onto the aluminum (Al) foil to ensure a low and featureless background signal, which was air-dried at room temperature prior to Raman measurements. Raman spectra were measured with a LabRAM Aramis confocal Raman microscope (HORIBA Jobin Yvon, Japan) equipped with a 300 g/mm grating. A 532 nm Nd:YAG laser (Laser Quantum, USA) with a power of 5 mW was focused onto the sample spot through a 100× objective (Olympus, NA = 0.90). To acquire consistent

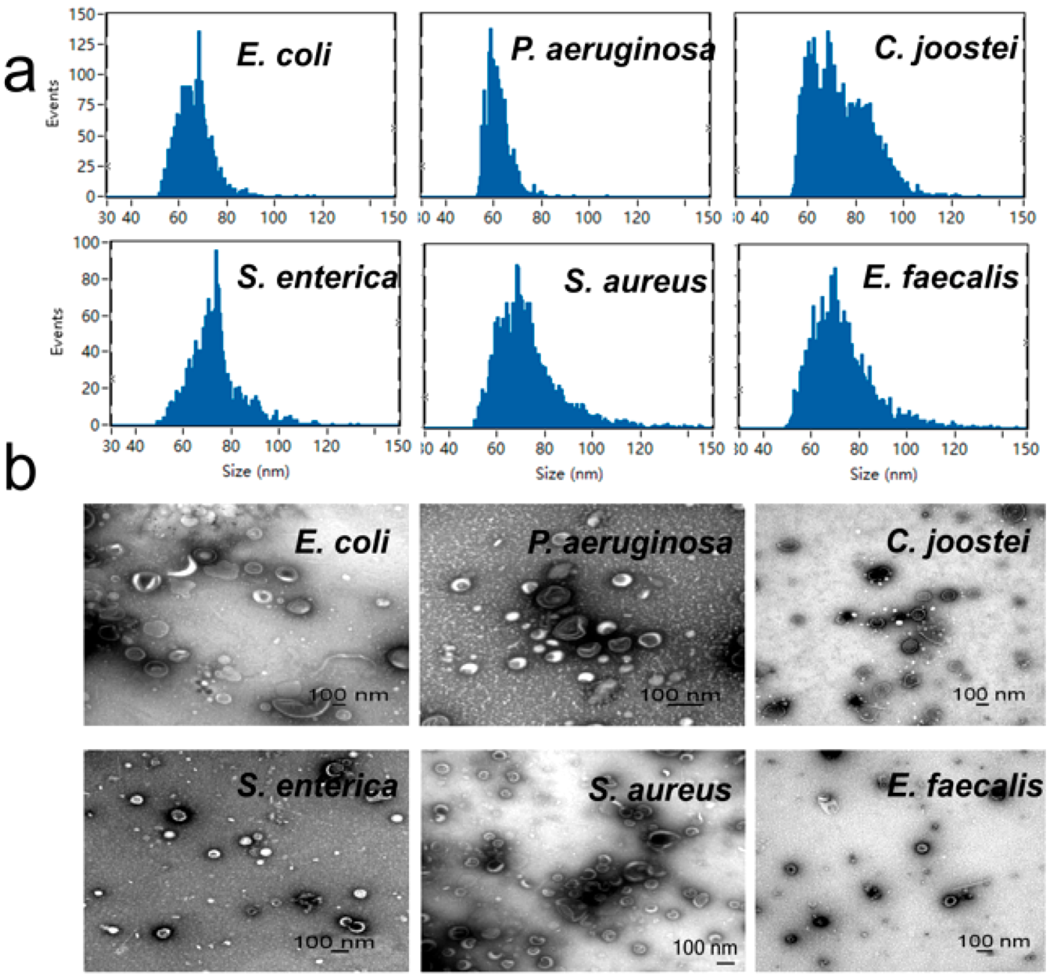


Figure 1. Characterization of EVs obtained from six different pathogens. (a) Particle size distribution histogram of EVs derived from *E. coli*, *P. aeruginosa*, *C. joostei*, *S. enterica*, *S. aureus*, and *E. faecalis* using an nFCM measurement. (b) Wide-field TEM image of EVs. The representative image is chosen from three independent repeats. Scale bar, 100 nm.

Raman spectra, the Raman spectrometer was calibrated with the Raman band of a silicon wafer at 520.6 cm^{-1}. The measurement was performed at an acquisition time of 9 s. A total of three independent biological replicates and two technical replicates were used for each sample. Biological replicates mean that each bacterial species was cultured for a total of three times, followed by EV extraction and Raman detection at each time. Technical replicates refer to when preparing EVs for Raman analysis at each time, two copies were prepared simultaneously to eliminate human or machine error. Savitzky−Golay filter with a window size of 7 and an order of 3 was used to denoise. After the subtraction of the biofluorescence background, spectra were individually normalized with the maximum peak value.

Statistical Analysis. All statistical comparisons of Raman spectra were performed with SPSS 25.0 (SPSS, Chicago, IL, USA) software. The results were presented as the mean ± standard deviation (SD). After verifying the normality and equality, we conducted one-way analysis of variance (ANOVA) for the differences among multiple groups followed by a post hoc Tukey's test. Individual data that did not meet the above requirements were analyzed with Kruskal−Wallis H (K). A p value of <0.05 was defined as statistically significant. The intensity of Raman bands from hundreds of spectra was displayed by points, and the quartiles distribution was shown as a box plot.

Machine Learning Model. In this study, principal component analysis (PCA) and PCA combined with linear discriminant analysis (PCA-LDA) analysis were implemented by MATLAB 2017a.[27] SVM with different kernels was implemented by Python API of sci-kit learn. The architecture of ResNet is similar to that of Ho.[22]

The aNN module used in this article is inspired by the convolutional block attention module.[28] The attention module is divided into channel attention and wavenumber attention. Channel attention is used to filter characteristics of different signal frequencies, and wavenumber attention is used to filter characteristics of different Raman shifts. Details of the architecture and mechanism of attention module are shown in Text S1.

During training, we found that the training difficulty for different samples is not the same, and easily identified species often occupied the main gradient of back-propagation, making it difficult to correct the misclassification in time. We used weighted cross entropy as the loss function to solve this problem. According to the training results, we adjusted the weights of one or several categories in time to achieve better classification performance.

The formula for weighted cross entropy is as follows

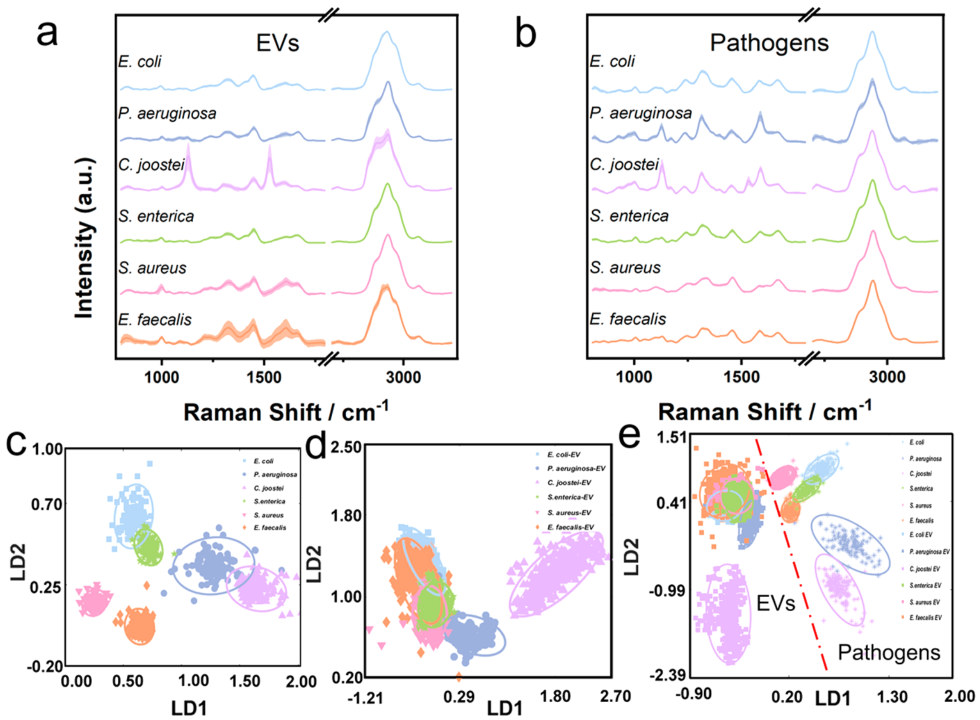


Figure 2. Raman signatures and classification of pathogenic bacteria and their derived EVs. Average Raman spectra of (a) EVs and (b) their host pathogenic bacteria. The thick solid line and shadow area represent the mean and standard deviations of spectra, respectively. PCA-LDA analysis of the Raman spectra obtained from (c) pathogens, (d) EVs, and (e) both.

$$\mathrm{loss}(x,\ \mathrm{class}) = -w_{\mathrm{class}} \log\left(\frac{\exp(x_{\mathrm{class}})}{\sum_{i=1}^{n} \exp(x_i)}\right) \tag{1}$$

where x is the output of the model, class is the category involved in calculating the loss, n is the total number of categories, and w is the weight of a category. The weight distribution of the four classification levels is shown inTable S2.

Unless otherwise specified, traditional CNN mentioned below refers to the aNN without the attention module, which is convenient for comparing the effect of the attention module.

IG-Based Raman Spectral Feature Extraction. After training, an integrated gradient (IG)[29] was used to uncover distinct spectral markers. An integrated gradient is a back-propagation-based feature-attribution algorithm. Unlike a traditional saliency map, a baseline needs to be selected first. Then the integral from the baseline to the input spectrum was calculated to represent the relative difference between them. The gradient value of the ith Raman shift was calculated as follows

$$I_i = (x_i - x_i') \int_0^1 \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} \, d\alpha \tag{2}$$

$$I_i = x_i \int_0^1 \frac{\partial F(\alpha x)}{\partial x_i} \, d\alpha \tag{3}$$

where x is the input spectrum, x' is the baseline spectrum, x_i is the intensity of the ith Raman shift of the input spectrum, F is the model that need to be interpreted, and α is a scale factor.

We choose the zero vector as a baseline in this article. As a result, eq 2 can be simplified to eq 3. Details about the integrated gradient are shown in Text S2.

■ RESULTS AND DISCUSSION

Isolation and Characterization of Pathogen-Derived EVs. EVs were extracted from six typical pathogenic bacterial species, including *E. coli*, *P. aeruginosa*, *Chryseobacterium joostei*, *Salmonella enterica*, *Staphylococcus aureus*, and *Enterococcus faecalis* (Table S3). Among them, *S. aureus* and *E. faecalis* are Gram-positive and *E. coli*, *P. aeruginosa*, *C. joostei*, and *S. enterica* are Gram-negative. These bacteria can cause numerous clinical infections including bacteremia, urinary tract infections, bacterial endocarditis, meningitis, and acute pneumonias.[30,31] Each pathogen was cultured in triplicate and harvested in the stationary phase.

EVs were successfully isolated using an ultracentrifugation protocol and characterized following the guidelines of the Information for Studies of Extracellular Vesicles 2018.[16] NanoFCM was employed to determine the size and quantity of the collected EVs. Figure 1a shows that most EVs fall in the size range of 40−150 nm, which is the typical size of EVs reported in previous works.[5,24] The average size of EVs derived from the six pathogens did not differ significantly: 66.3 ± 7.0 nm (*E. coli*), 61.8 ± 4.9 nm (*P. aeruginosa*), 73.6 ± 11.8 nm (*C. joostei*), 72.6 ± 10.5 nm (*S. enterica*), 72.7 ± 12.7 nm (*S. aureus*), and 72.9 ± 12.7 nm (*E. faecalis*). TEM was also used to characterize the morphology of EVs, and a classical "cup-shaped" morphology was observed (Figure 1b). As such, both the size analysis and TEM images demonstrated the successful extraction of pathogen-derived EVs.

Raman Profiling and Classification of EVs and Their Host Pathogens. By using the as-prepared EVs, a large Raman database containing 4335 spectra from 3 independent batches of EVs was established. To minimize the potential deviation, EVs were all isolated from pathogens in the same growth stage of the stationary phase. Spectra from three independent biological replicates are consistent. Figure 2a,b

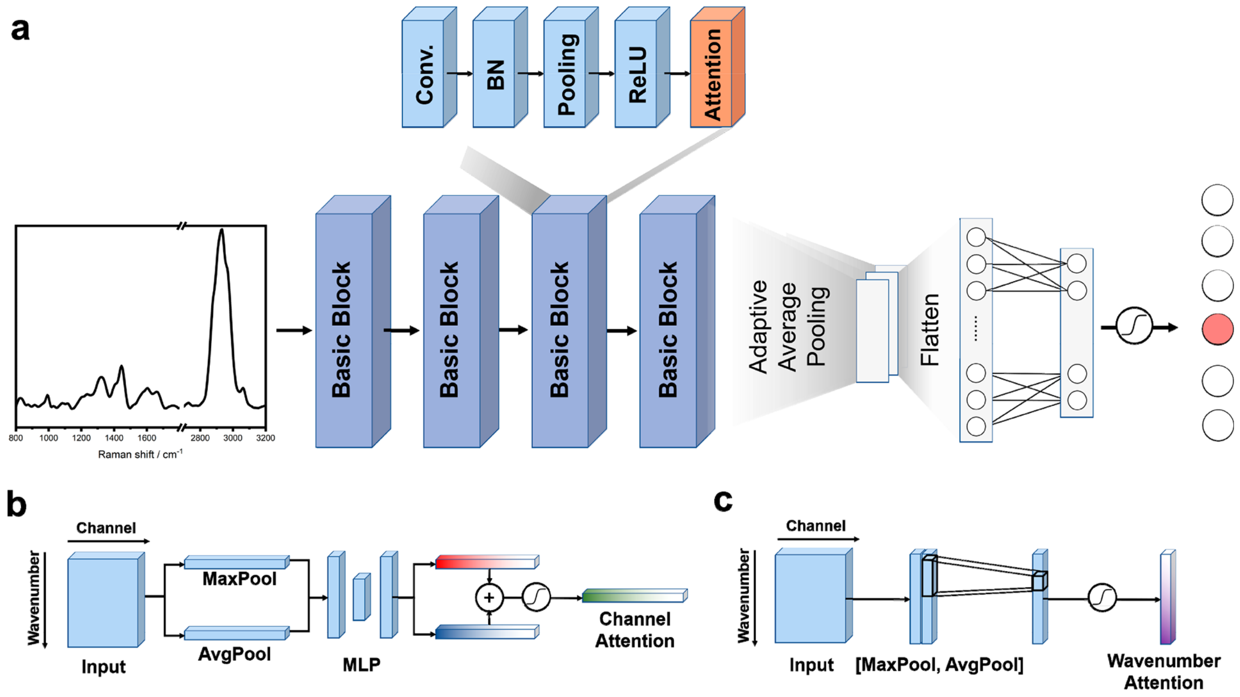


Figure 3. Construction of the aNN model. (a) aNN structure diagram. aNN includes four convolution modules, four attention modules, and a classifier based on a fully connected layer. The convolution block consists of a convolution layer, a batch regularization layer, a pooling layer, and an activation layer. (b) Structural diagram of channel attention. The channel attention first compresses the wavenumber direction information through the parallel maximum pooling layer and the average pooling layer and then uses the multilayer perceptron (MLP) to filter the features before adding the parallel compressed information. After activating the sigmoid function, the attention distribution of the channel direction is generated. (c) Structure diagram of wavenumber attention. Wavenumber attention goes through parallel maximum pooling and average pooling operations. Different from the channel attention, the compressed information needs to use a convolutional layer with a 1 × 1 kernel to reduce the dimensionality and then be activated by the sigmoid function to obtain the wavenumber attention distribution.

shows the average Raman spectra with corresponding standard deviations of EVs and their parent pathogens in the spectral ranges of 800−1800 and 2700−3200 cm^{-1}.

Prominent Raman peaks related to the biological constituents of EVs can be observed at 826 cm^{-1} (O−P−O stretch DNA), 1003 cm^{-1} (ring breathing of phenylalanine), 1130 cm^{-1} (C−C stretching of carotenoid-like substance), 1205 cm^{-1} (C−N stretching and N−H bending of amide III), 1320 cm^{-1} (ring breathing of DNA/RNA bases), 1450 cm^{-1} (CH_2 bending of lipid), 1525 cm^{-1} (C−C and conjugated C=C stretching of carotenoid), 1600 cm^{-1} (C−C in-plane bending mode of phenylalanine and tyrosine), and 2938 cm^{-1} (C−H stretching of lipids and proteins). The presence of these peaks indicated that EVs contained nucleic acids, proteins, lipids, and pigment. A direct visual inspection of EV spectra revealed that except for *C. joostei*-derived EVs, all other EVs exhibited highly similar spectral features, indicating their similar compositions. By comparison, relatively obvious differences were observed between EVs and parental pathogens, revealing that the chemical compositions of EVs were distinct from bacteria.

To highlight and visualize the spectral difference, all of the obtained high-dimension spectral data were subjected to multivariate dimension-reducing processing via PCA-LDA. The derived two-dimensional scores plotted at the 0.95 confidence level are shown in Figure 2c−e. It is interesting to find that the PCA-LDA clusters of six pathogens are all well separated (Figure 2c) whereas those of EVs overlapped to a certain degree (Figure 2d), indicating a smaller spectral difference among EVs than among parental bacteria. It is therefore more challenging to classify and identify EVs than pathogens. Moreover, when pooling all of the spectral data from pathogens and EVs, PCA-LDA plots of EVs and pathogens were clearly divided into two categories, indicating their large spectral difference (Figure 2e). This finding indicates that although nearly all of the biological materials carried by EVs are from the parent cells, some biomaterials are preferentially carried by EVs or at least in different portions from parent cells. Recent studies on the proteomics and transcriptome of bacteria and archaea reported the similar phenomenon that some specific protein and RNA cargoes were preferably carried by EVs.[32−34]

Deep Learning-Enabled EV Identification at Multihierarchies via the Attentional Neural Network. Because the PCA-LDA scores plot did not provide clear cluster boundaries for the various pathogen-derived EVs, we employed a new deep learning algorithm of aNN to construct classification models to identify EVs of different pathogen origins based on their Raman spectra. aNN is constructed by incorporating a convolutional neural network with an attention module in order to better extract the differences in the spectra. It includes four convolution modules, four attention modules, and a classifier based on a fully connected layer (Figure 3a). Attention is a mechanism to reasonably distribute computing resources by weighting features. To apply this mechanism to Raman spectroscopy, we designed channel attention and wavenumber attention (Figure 3b,c), which can emphasize the key points and weaken the unimportant features in both dimensions.

In order to determine the capabilities of this aNN-based deep learning model to meet the above classification

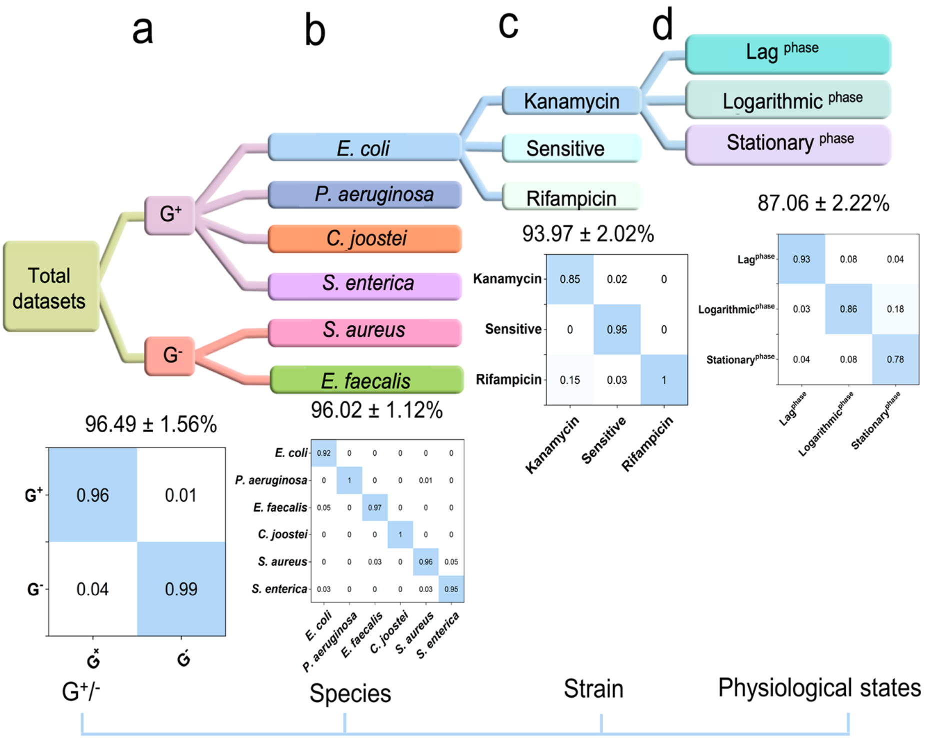


Figure 4. Discrimination performance of the aNN model at different levels. The horizontal and vertical coordinates of confusion matrices are true labels and predicted labels, respectively. Diagonal values exhibited the accuracy at (a) G^+/G^-, (b) species, (c) strain, and (d) down to physiological levels. *Test sets were from independent experimental results not included in the training sets.

requirements, a four-level classification data set of EVs was introduced, including G^+ and G^-, species, strains, and physiological states. The top level contained EVs from two species of G^+ pathogens and four species of G^- pathogens. The second level contained EVs from six different pathogenic species. The third level contained EVs from one antibiotic-sensitive strain and two resistant strains of *E. coli*. (The resistance genes were located on the plasmid and chromosome and were resistant to kanamycin and rifampicin, respectively.) The last level focused on EVs from different physiological states of the kanamycin-resistant *E. coli* strain, including the lag phase, logarithmic phase, and stationary phase. Such a four-level classification is clinically relevant. In the process of bacterial infection, pathogens with antibiotic resistance tend to cause more therapeutic difficulty.[35] Moreover, even for the same bacterial strain, different physiological states such as growth phases may result in different infection risks. EVs can be produced during all bacterial growth phases. The function of EVs is closely related to the contents derived from their parent bacteria in the context of pathogen infections.[36] Subtle differences in pathogens may lead to great function variation in their secreted vesicles. Therefore, apart from identifying EVs from different species, it is also crucial to identify EVs derived from the same species with different antibiotic resistance as well as the same strain in different physiological states. Considering that the spectra produced by bacteria along the taxonomic level become increasingly similar, it is anticipated to be more challenging to distinguish the associated EVs.

To perform aNN, the training and testing sets used here account for about 80 and 20% of all spectra, respectively. The specific numbers of spectra allocated to the training set and testing set at different levels are listed in Table S4. To make sure that the test set was independent of the training data sets, the testing data set was gathered from independently cultured samples that did not participate in the training process. As shown in Figure 4, the aNN model manifested satisfactory EV classification results, with average accuracies of 96.49 ± 1.56, 96.02 ± 1.12, 93.97 ± 2.02, and 87.06 ± 2.22% at the G^{+}/G^-, species, strain, and physiological levels, respectively. Of note, the accuracy was more than 95% at the species level, more than 90% at the strain level, and still more than 85% at the physiological level. This result demonstrates that aNN enabled the Raman method to pinpoint the origin of EVs to a level far beyond other methods, such as Gram classification.[17] EVs that merely differ in growth stages still meet a high classification accuracy, supporting the view that EV components reflect the physiological state of original bacteria.[7,37]

To compare the performance of aNN, five other machine and deep learning models were employed to identify EVs with the same four-level classification data set, including SVM with a linear kernel, SVM with a radial basis function (RBF) kernel, LDA, and CNN and ResNet with architectures similar to that of Ho.[22] The classification results are shown in Table S5. The accuracy of LDA and SVM was just around 70% at the strain level and even less than 35% at the physiological level. For ResNet and CNN, although these two models also achieved more than 95% accuracy when the classification task was not very difficult (at Gram and species levels), the aNN model performed more accurately at the more challenging strain and physiological levels. Clearly, compared with traditional algorithms, aNN achieved the best performance in terms of the accuracy. Apart from the high accuracy, the advantage of aNN also lies in that it is more portable and efficient. As shown in Table S6, aNN requires only 17% of the number of parameters and 6% of the floating point operations (FLOPs)

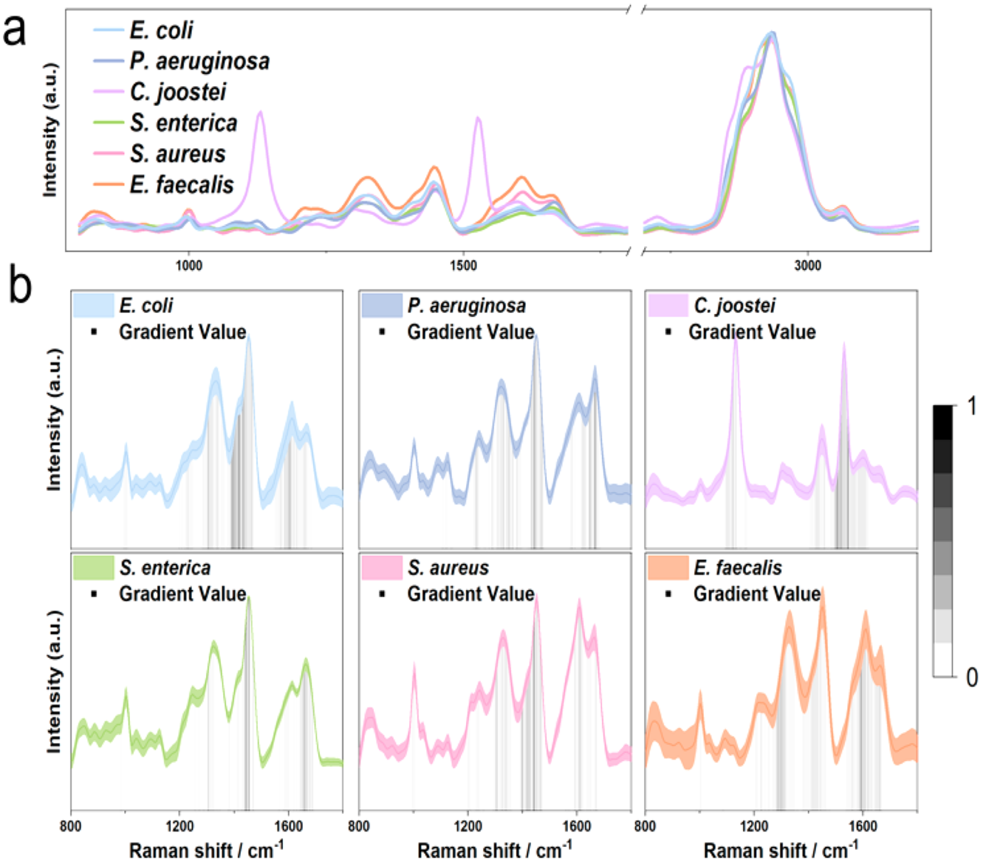


Figure 5. IG uncovers distinct spectral markers of different EVs. (a) Average Raman spectra acquired from six pathogen-derived EVs. (b) Visualization of the decisive bands of specific EVs based on their Raman spectra. The gradient value, represented as a bar, is an indicator to measure the importance of the spectral bands. The darker the bar, the larger the gradient value and the larger the contribution of the spectral band to the EV classification.

required by ResNet. Compared with CNN, aNN improves the performance by using a small amount of computing resources.

Therefore, Raman spectroscopy combined with aNN is a reliable and efficient method of identifying EVs of bacterial origins at a high classification level from Gram, species, strains (antibiotic-resistant or sensitive), and down to physiological states. It will appear as an advanced identification method for EVs and will facilitate the application of EVs in infectious disease diagnosis.

IG-Enabled Uncovering of Distinct Spectral Markers for Pathogenic EVs. From the above results, deep learning has enabled the accurate classification of bacterial EVs but provided no knowledge of what makes the difference. The underlying Raman spectral features that contribute to EV discrimination may be of great value as a potential biomarker of EV.[38] However, except for the unique bands of the carotenoid pigment (1130 and 1525 cm^{-1}) that can directly identify *C. luteola* EVs from other EVs (Figure 5a), it is challenging to decipher the decisive spectral features to distinguish EVs at the species level.

Here, IG was used to extract the decisive bands of EVs. Previously, a saliency map has been widely used in the artificial intelligence-related Raman field. However, a saliency map cannot solve the problem of gradient saturation, which makes the gradient independent of the feature important. An integrated gradient is the evolution of the saliency map. Using the path integration from the baseline to the input spectrum can well avoid the above problems and make the measure of feature importance more accurate.

We use the spectra from species-level data to test the ability of IG to extract distinct spectral features. As shown in Figure 5b, the order of contribution of Raman peaks to EV classification is ranked as the gradient distribution. Most of the gradient distributions were concentrated in the range of 1400−1600 cm^{-1}. Of note, *C. joostei* EVs displayed strong gradient signals at 1130 and 1525 cm^{-1}, consistent with our previous results via direct spectral observations (Figure 5a), demonstrating the reliability of IG. All of the Raman bands that contribute most to the identification of one specific EV at the species level and the corresponding assignments are listed in Table S7. We found that the primary contributors were species-dependent. The key spectral markers were present at around 1413 cm^{-1} (*E. coli*), 1168 cm^{-1} (*P. aeruginosa*), 1552 cm^{-1} (*E. faecalis*), 1130 and 1525 cm^{-1} (*C. joostei*), 1024 cm^{-1} (*S. aureus*), and 1456 cm^{-1} (*S. enterica*), respectively. These results indicated that the interpretation method enabled the extraction of spectral markers even from the subtle spectral differences among EVs. These spectral markers may facilitate the rapid diagnosis of parental pathogenic bacteria and thus timely treatment.

Comparison of Raman Features between EVs and Bacteria. The production of EVs is a conserved process that carries contents from their parent bacteria. However, there is still a limited understanding of the content shuttling during EV biogenesis. Whether cargoes are selectively packaged into EVs

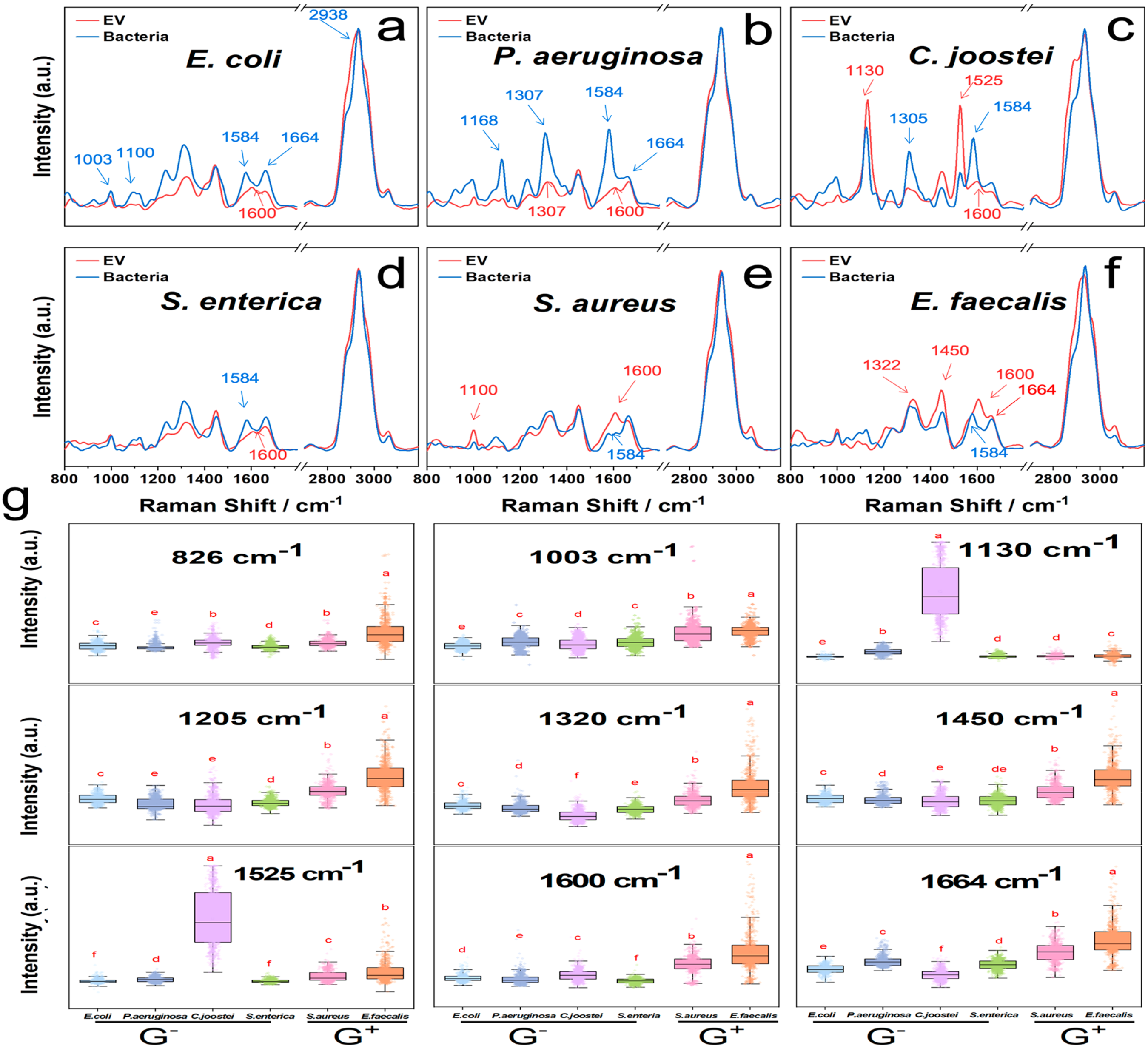


Figure 6. Comparison of Raman spectral features of six pathogens and their EVs. (a−f) Average Raman spectra of pathogens (blue solid line) and their EVs (red solid line). (g) Quantitative comparison of the Raman band intensities among different EVs. The results represent the means ± SD. Different letters (e.g., a and b) were defined as statistically significant ($P < 0.05$).

or are merely appendages of EV biogenesis is still under debate.[33,39] In the previous sections, we indicated that Raman spectra of EVs were obviously separated from the parental pathogens (Figure 3e). The underpinning biomolecular composition may confer crucial biological information related to the specific contents enclosed in EVs and the biogenesis process.

Here, we compared the Raman spectra from the EVs and their parental cells after normalization with the highest peak at 2938 cm^{-1} (C−H stretch of lipids and proteins) (Figure 6a−f). Most Raman peaks attributed to nucleic acids, proteins, and lipids showed a relatively lower intensity in EVs than that of bacteria. For example, the band at 1100 cm^{-1} (O−P−O backbone stretch of DNA) detected in *E. coli* was almost absent in its EVs (Figure 6a). Moreover, bands at 1003 cm^{-1} (phenylalanine; Figure 6a), 1307 cm^{-1} (CH_3/CH_2 twisting or bending mode of lipid/collagen; Figure 6b), 1584 cm^{-1} (ring breathing of guanine and adenine) showed similar trends (Figure 6b). Because EVs are just nanometer lipid bilayer particles with limited encapsulated contents, it becomes apparent that the relative abundance of these contents in EVs is less than that in the originating bacteria.

We also found that some Raman peaks with a low relative intensity in bacteria exhibited a higher relative intensity in EVs. For example, the relative intensities of 1130 and 1525 cm^{-1} (carotenoid-like) in *C. joostei* EVs were higher than those in the counterpart pathogens (Figure 6c), as were the peaks at 1322 cm^{-1} (nucleic acids) and 1450 cm^{-1} (lipid) in *E. faecal* EVs (Figure 6e). Moreover, the 1584 cm^{-1} peak (ring breathing of guanine and adenine) apparent in all bacterial spectra decreased and became less obvious in bacterial EVs. Instead, a neighboring peak at 1600 cm^{-1} (the C═C in-plane bending mode of phenylalanine) increased and showed up in all EVs. These results indicated that some biocomponents from parental bacteria were selectively encapsulated and used to enrich their EVs. These results are in agreement with our previous work that some DNA at low concentrations in the natural environment may be enriched in EVs isolated from the environment.[40] Previous studies using microscopy and a multiomics approach also showed that there existed the selective enrichment of specific DNA fragments, miRNA, and protein in EVs.[15,33,41]

More interestingly, during the selective packaging process, different types of EVs were found to use different package choices. For example, the amide I band at 1664 cm^{-1} displayed a similar high intensity in both *E. faecal* and the EVs (Figure 6f). In contrast, this band detected in *E. coli* and *S. enterica* decreased in their EVs. Likewise, previous work compared the relative levels of each miRNA in the cell body and EVs and found that some miRNAs were enriched in EVs from all types of cells, whereas other miRNAs were enriched only in one or two types of EVs.[15] In addition, we found that the average spectra of EVs from G^+ and G^- bacteria were visibly distinguishable. To semiquantitively illustrate such a variation, we compared the relative intensity of Raman peaks normalized by C−H peaks (Figure 6g). It can be found that the bands at 826, 1003, 1205, 1320, 1450, 1600, and 1664 cm^{-1} in G^+ bacteria-derived EVs (G^+ EV) were stronger than those of G^- bacteria-derived EVs (G^- EV), indicating that the corresponding contents of G^+ EVs were more abundant than those of G^- EVs. The reason might be related to the different secretion modes of EVs from G^+ and G^- bacteria. The classic production route of G^- EVs is through blebbing from the outer membrane; therefore, EVs get enriched with outer-membrane and periplasmic cargoes but do not have direct access to cytoplasmic compounds.[5] By comparison, EVs from G^+ bacteria originated from the inner membrane and can therefore package rich contents from the cytoplasm.[24] Together, these findings indicate that there may be a unique retention and depletion pattern of EV contents.

Furthermore, a noticeable difference was found in the C−H vibrational region from 2800 to 3050 cm^{-1}. The C−H bond-rich membrane lipid showed a prominent contrast between EVs and their corresponding bacteria. Such changes reflected the difference in the composition of EVs and bacterial membranes. As we know, the formation of the EVs' membrane structure is by entrapment of the originating cell membrane contents.[42] Nevertheless, carbohydrates and proteins are randomly distributed on the cell membrane, and the blebbing region of EV is random, resulting in the variation of the EV membrane compared to the overall cell membrane. The progradation of EVs has been previously reported to cause the outer members' remodeling,[7] supporting our Raman observation of the difference between EVs and the bacterial outer membrane. To conclude, multiple biocomponents were found to be selectively encapsulated by EVs during the biogenesis process from the rich Raman chemical signatures, and this encapsulation was bacteria-dependent. Moreover, EVs exhibited different membrane compositions from those of parental bacteria. Our work provides some clues to the characteristics of EV biogenesis. These data support the hypothesis proposed by many review or comment articles.[5,7] In the future, the further coupling with proteomics and metabolomics will provide a more detailed understanding of the molecular events involved in the EV biogenesis.[43]

■ CONCLUSIONS

A novel approach coupling label-free Raman spectroscopy with a deep learning model of aNN was developed to enable the rapid and accurate identification of pathogenic bacteria-derived EVs, a major precursor to the induction of bacterial infectious diseases. A large Raman data set from EVs derived from six typical pathogens at different taxonomic and physiological levels was established to explore the discriminatory power of the method. By training and testing the aNN using the Raman spectra of independently extracted and measured EVs samples, we realized the accurate identification of EVs with high accuracies at all levels from G^+/G^- (96.49 ± 1.56%), species (96.02 ± 1.12%), antibiotic-resistant and sensitive strains (93.97 ± 2.02%,), and even down to the physiological level (87.06 ± 2.22%). Clearly, Raman spectroscopy coupled with aNN achieves a much higher resolution of the bacterial origin of EVs than previous methods. Moreover, aNN outperformed five other deep/machine learning algorithms previously used for bacterial identification by showing higher accuracy and occupying less computational resources. Furthermore, employing the interpretation algorithm of IG to extract the decisive bands from subtly different EV spectra, spectral markers that contribute to the identification of pathogen-derived EVs at the species level, was used. Finally, a detailed comparative Raman feature study of EVs and parental pathogens clearly revealed the specific biocomponents packaged in EVs during EV biogenesis and the associated selective encapsulation and distinct membrane compositions.

The developed deep learning-enabled Raman identification strategy is easily extendable to EVs secreted by other pathogenic bacteria. It advances EV identification methods. However, direct clinical implementation is still challenging with the current Raman spectrometer, with a spatial resolution that is not high enough for individual EV at the nanoscale. Clinical samples often contain a blend of EVs of different origins including bacteria, human cells, and tissues. The discrimination of the origin of EVs and the characterization of their abundance are required by clinical diagnosis. Recently, a single-particle automated Raman trapping system capable of capturing and analyzing individual nanosized EVs has been reported.[44] Together with the deep learning-enabled Raman identification strategy developed here, they will promote early EV-based infectious disease diagnosis and the timely treatment of bacterial infections.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c02226.

> Key mechanism of the attention module; mechanism and advantage of the integrated gradient; list of pathogens' phylogeny; data sets from the training set and testing set; performance of different machine and deep learning models; number of parameters and computational complexity of ResNet, CNN, and aNN; summary of spectral markers to distinguish different EVs; and hyperparameters of aNN (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Bin Ren − *State Key Laboratory of Physical Chemistry of Solid Surfaces, Collaborative Innovation Center of Chemistry for*

Energy Materials (i-ChEM), Department of Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China; orcid.org/0000-0002-9821-5864; Email: bren@xmu.edu.cn

Li Cui − *Xiamen Key Laboratory of Indoor Air and Health, Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China;* orcid.org/0000-0002-0708-8899; Email: lcui@iue.ac.cn; Fax: +86-592-6190780

Authors

Yi-Fei Qin − *Xiamen Key Laboratory of Indoor Air and Health, Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China; College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China*

Xin-Yu Lu − *State Key Laboratory of Physical Chemistry of Solid Surfaces, Collaborative Innovation Center of Chemistry for Energy Materials (i-ChEM), Department of Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China*

Zheng Shi − *College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China; State Environment Protection Key Laboratory of Satellite Remote Sensing, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China*

Qian-Sheng Huang − *Xiamen Key Laboratory of Indoor Air and Health, Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China;* orcid.org/0000-0002-3788-3164

Xiang Wang − *State Key Laboratory of Physical Chemistry of Solid Surfaces, Collaborative Innovation Center of Chemistry for Energy Materials (i-ChEM), Department of Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China;* orcid.org/0000-0003-1937-0725

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c02226

Author Contributions

[⊥]Y.-F.Q. and X.-Y.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (21922608 and 22176186) and the Chinese Academy of Sciences (ZDBS-LY-DQC027).

■ REFERENCES

(1) DeAntonio, R.; Yarzabal, J. P.; Cruz, J. P.; Schmidt, J. E.; Kleijnen, J. *Hum Vaccin Immunother* **2016**, *12* (9), 2422−40.

(2) Fleischmann, C.; Scherag, A.; Adhikari, N. K.; Hartog, C. S.; Tsaganos, T.; Schlattmann, P.; Angus, D. C.; K. Reinhart, T. *Am. J. Respir Crit Care Med.* **2016**, *193* (3), 259−272.

(3) Murray, C. J. L.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Aguilar, G. R.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; et al. *Lancet* **2022**, *399* (10325), 629−655.

(4) Kaparakis-Liaskos, M.; Ferrero, R. L. *Nat. Rev. Immunol* **2015**, *15* (6), 375−87.

(5) Schwechheimer, C.; Kuehn, M. J. *Nat. Rev. Microbiol* **2015**, *13* (10), 605−19.

(6) Shao, H.; Im, H.; Castro, C. M.; Breakefield, X.; Weissleder, R.; Lee, H. *Chem. Rev.* **2018**, *118* (4), 1917−1950.

(7) Toyofuku, M.; Nomura, N.; Eberl, L. *Nat. Rev. Microbiol* **2019**, *17* (1), 13−24.

(8) Colombo, M.; Raposo, G.; Thery, C. *Annu. Rev. Cell Dev Biol.* **2014**, *30*, 255−89.

(9) EL Andaloussi, S.; Mager, I.; Breakefield, X. O.; Wood, M. J. A. *Nat. Rev. Drug Discov* **2013**, *12* (5), 347−357.

(10) Thery, C.; Zitvogel, L.; Amigorena, S. *Nat. Rev. Immunol* **2002**, *2* (8), 569−79.

(11) Vanaja, S. K.; Russo, A. J.; Behl, B.; Banerjee, I.; Yankova, M.; Deshmukh, S. D.; Rathinam, V. A. K. *Cell* **2016**, *165* (5), 1106−1119.

(12) Bittel, M.; Reichert, P.; Sarfati, I.; Dressel, A.; Leikam, S.; Uderhardt, S.; Stolzer, I.; Phu, T. A.; Ng, M.; Vu, N. K.; et al. *J. Extracell Vesicles* **2021**, *10* (12), No. e12159.

(13) Yang, D.; Chen, X.; Wang, J.; Lou, Q.; Lou, Y.; Li, L.; Wang, H.; Chen, J.; Wu, M.; Song, X.; et al. *Immunity* **2019**, *50* (3), 692−706.

(14) Deo, P.; Chow, S. H.; Han, M. L.; Speir, M.; Huang, C.; Schittenhelm, R. B.; Dhital, S.; Emery, J.; Li, J.; Kile, B. T.; et al. *Nat. Microbiol* **2020**, *5* (11), 1418−1427.

(15) Garcia-Martin, R.; Wang, G.; Brandao, B. B.; Zanotto, T. M.; Shah, S.; Kumar Patel, S.; Schilling, B.; Kahn, C. R. *Nature* **2022**, *601* (7893), 446−451.

(16) Thery, C.; Witwer, K. W.; Aikawa, E.; Alcaraz, M. J.; Anderson, J. D.; Andriantsitohaina, R.; Antoniou, A.; Arab, T.; Archer, F.; Atkin-Smith, G. K.; et al. *J. Extracell Vesicles* **2018**, *7* (1), 1535750.

(17) Hendrix, A.; De Wever, O. *Trends Microbiol* **2022**, *30* (3), 213−216.

(18) Wang, Y.; Zhang, R.; Li, J.; Wu, Z.; Yin, W.; Schwarz, S.; Tyrrell, J. M.; Zheng, Y.; Wang, S.; Shen, Z.; et al. *Nat. Microbiol* **2017**, *2*, 16260.

(19) Stockel, S.; Kirchhoff, J.; Neugebauer, U.; Rosch, P.; Popp, J. *J. Raman Spectrosc.* **2016**, *47* (1), 89−109.

(20) Butler, H. J.; Ashton, L.; Bird, B.; Cinque, G.; Curtis, K.; Dorney, J.; Esmonde-White, K.; Fullwood, N. J.; Gardner, B.; Martin-Hirsch, P. L.; et al. *Nat. Protoc* **2016**, *11* (4), 664−87.

(21) Kusic, D.; Kampe, B.; Rosch, P.; Popp, J. *Water Res.* **2014**, *48*, 179−89.

(22) Ho, C. S.; Jean, N.; Hogan, C. A.; Blackmon, L.; Jeffrey, S. S.; Holodniy, M.; Banaei, N.; Saleh, A. A. E.; Ermon, S.; Dionne, J. *Nat. Commun.* **2019**, *10* (1), 4927.

(23) Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y. G.; Choi, B. H.; Kang, K. W.; Jeong, H.; Park, Y.; et al. *ACS Nano* **2020**, *14* (5), 5435−5444.

(24) Brown, L.; Wolf, J. M.; Prados-Rosales, R.; Casadevall, A. *Nat. Rev. Microbiol* **2015**, *13* (10), 620−30.

(25) Qin, Y.; Zhang, J.; Avellan-Llaguno, R. D.; Zhang, X.; Huang, Q. *Environ. Pollut.* **2021**, *272*, 116005.

(26) Tian, Y.; Gong, M.; Hu, Y.; Liu, H.; Zhang, W.; Zhang, M.; Hu, X.; Aubert, D.; Zhu, S.; Wu, L.; et al. *J. Extracell Vesicles* **2020**, *9* (1), 1697028.

(27) Martin, F. L.; Kelly, J. G.; Llabjani, V.; Martin-Hirsch, P. L.; Patel, I. I.; Trevisan, J.; Fullwood, N. J.; Walsh, M. J. *Nat. Protoc* **2010**, *5* (11), 1748−60.

(28) Woo, S. H.; Park, J.; Lee, J. Y.; Kweon, I. S. *Lect Notes Comput. Sc* **2018**, *11211*, 3−19.

(29) Sundararajan, M.; Taly, A.; Yan, Q. Q. *Pr. Mach. Learn. Res.* **2017**, 70.

(30) Tong, S. Y.; Davis, J. S.; Eichenberger, E.; Holland, T. L.; Fowler, V. G., Jr. *Clin Microbiol Rev.* **2015**, *28* (3), 603−61.

(31) Curran, C. S.; Bolig, T.; Torabi-Parizi, P. *Am. J. Respir Crit Care Med.* **2018**, *197* (6), 708−727.

(32) Lee, E. Y.; Choi, D. Y.; Kim, D. K.; Kim, J. W.; Park, J. O.; Kim, S.; Kim, S. H.; Desiderio, D. M.; Kim, Y. K.; Kim, K. P.; et al. *Proteomics* **2009**, *9* (24), 5425−36.

(33) Liu, J.; Cvirkaite-Krupovic, V.; Commere, P. H.; Yang, Y.; Zhou, F.; Forterre, P.; Shen, Y.; Krupovic, M. *ISME J.* **2021**, *15* (10), 2892−2905.

(34) Resch, U.; Tsatsaronis, J. A.; Le Rhun, A.; Stubiger, G.; Rohde, M.; Kasvandik, S.; Holzmeister, S.; Tinnefeld, P.; Wai, S. N.; Charpentier, E. *mBio* **2016**, *7* (6), No. e00207-16.

(35) Walsh, C. *Nature* **2000**, *406* (6797), 775−81.

(36) Wang, X.; Eagen, W. J.; Lee, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (6), 3174−3184.

(37) Briaud, P.; Frey, A.; Marino, E. C.; Bastock, R. A.; Zielinski, R. E.; Wiemels, R. E.; Keogh, R. A.; Murphy, E. R.; Shaw, L. N.; Carroll, R. K. *mSphere* **2021**, *6* (5), No. e00676-21.

(38) Hoshino, A.; Kim, H. S.; Bojmar, L.; Gyan, K. E.; Cioffi, M.; Hernandez, J.; Zambirinis, C. P.; Rodrigues, G.; Molina, H.; Heissel, S.; et al. *Cell* **2020**, *182* (4), 1044−1061.

(39) Bonnington, K. E.; Kuehn, M. J. *Biochim. Biophys. Acta* **2014**, *1843* (8), 1612−9.

(40) Qin, Y.; Guo, Z.; Huang, H.; Zhu, L.; Dong, S.; Zhu, Y. G.; Cui, L.; Huang, Q. *Environ. Sci. Technol.* **2022**, *56* (9), 5653−5663.

(41) Wang, H.; Beier, N.; Boedeker, C.; Sztajer, H.; Henke, P.; Neumann-Schaal, M.; Mansky, J.; Rohde, M.; Overmann, J.; Petersen, J.; et al. *mSystems* **2021**, *6* (1), No. e00693-20.

(42) Deatherage, B. L.; Lara, J. C.; Bergsbaken, T.; Rassoulian Barrett, S. L.; Lara, S.; Cookson, B. T. *Mol. Microbiol.* **2009**, *72* (6), 1395−1407.

(43) Xu, J.; Preciado-Llanes, L.; Aulicino, A.; Decker, C. M.; Depke, M.; Gesell Salazar, M.; Schmidt, F.; Simmons, A.; Huang, W. E. *Anal. Chem.* **2019**, *91* (12), 7729−7737.

(44) Penders, J.; Nagelkerke, A.; Cunnane, E. M.; Pedersen, S. V.; Pence, I. J.; Coombes, R. C.; Stevens, M. M. *ACS Nano* **2021**, *15* (11), 18192−18205.

https://doi.org/10.1021/acs.analchem.2c02226