Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 326 (2025) 125207

Contents lists available at ScienceDirect

Spectrochimica Acta Part A:
Molecular and Biomolecular Spectroscopy

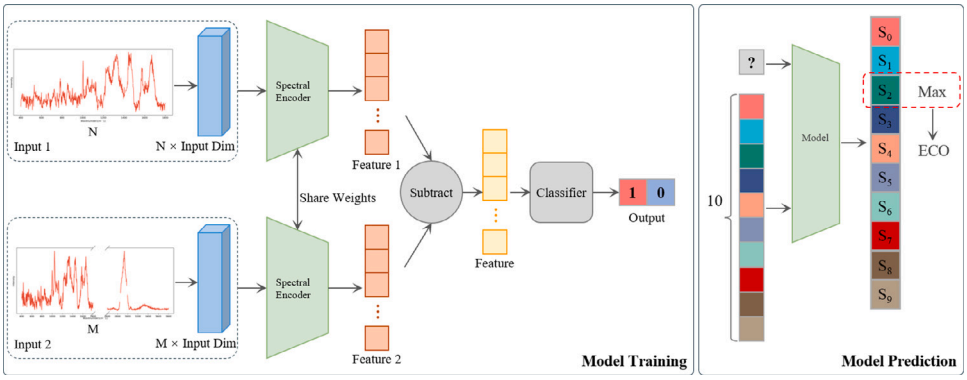
journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

Siamese network for classification of Raman spectroscopy with inter-instrument variation for biological applications

Xiaodong Bao, Lindong Shang, Fuyuan Chen, Hao Peng, Yu Wang, Xusheng Tang, Yan Ge, Bei Li *

Key Laboratory of Optical System Advanced Manufacturing Technology, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, 130033, PR China
University of Chinese Academy of Sciences, Beijing, 100049, PR China

GRAPHICAL ABSTRACT



HIGHLIGHTS

- Proposing a modular Siamese neural network model with freely pluggable spectral encoders.
- This model solves the Raman spectral classification problem in biological applications with inter-instrument variation.
- This model can be used for fusion training and prediction of Raman spectra of different resolutions (different data lengths).
- The model can be trained using only a small amount of data and achieves high classification accuracy.

ARTICLE INFO

Keywords:
Siamese network
Inter-instrument variation
Raman spectroscopy
Classification

ABSTRACT

Raman spectroscopy has emerged as a highly sensitive, rapid, and label-free detection method, extensively utilized in biological research. Presently, it is frequently paired with artificial intelligence (AI) algorithms to facilitate identification and classification tasks. However, variations in the settings across different Raman spectrometers, along with the sensitive and continuous nature of biological Raman signals, can subtly alter the acquisition of these signals. This can potentially impact the classification outcomes of the spectra. Moreover, Raman spectra with disparate resolutions pose challenges for effective model training. In this study, we introduce a modularized Siamese neural network, equipped with multiple projection layers to segregate the model components. This design allows our model to support the core module spectral encoder's pluggability.

* Corresponding author at: Key Laboratory of Optical System Advanced Manufacturing Technology, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, 130033, PR China.
E-mail address: beili@ciomp.ac.cn (B. Li).

https://doi.org/10.1016/j.saa.2024.125207
Received 15 July 2024; Received in revised form 19 September 2024; Accepted 22 September 2024
Available online 26 September 2024

The model determines the classification results by extracting the features of Raman spectra with inter-instrument variation, mapping these feature distances into spectral similarities, and finally, comparing a set of similarities. Our experimental results demonstrate the feasibility of training the model with only 10 spectra per category, using bacterial datasets we created. We compared the classification outcomes of three distinct spectral encoders, with the most effective model achieving a classification accuracy exceeding 90%. Furthermore, we successfully implemented the fusion training and prediction of Raman spectra with different resolutions. In conclusion, our model enhances the validity and comparability of Raman spectral acquisition for biological applications and diversifies the methods of Raman spectral acquisition.

1. Introduction

Raman spectroscopy has emerged as a frequently employed spectral analysis technique across various research fields, including biological analysis [1–4], real-time monitoring of chemical reactions [5,6], and material identification [7–9], among others. The primary components of a Raman spectrometer, namely a laser, a spectrometer, and a detector, play pivotal roles in determining the accuracy of Raman spectral measurements. The laser's stability [10,11], the spectrometer's sensitivity [12], and the CCD detector's stability [13] directly influence the measurement precision. Environmental parameters, such as temperature, humidity, and pressure, during the Raman detection process can also impact the measurement outcomes [10,13,14]. In a long-term stability study of the Raman spectrometer, Raj et al. [15] observed fluctuations up to $\pm 0.3\ cm^{-1}$ the band positions and up to 2% in the corresponding Raman intensities over extended periods. To enhance the Raman spectrometer's measurement accuracy, researchers and manufacturers have used various standard materials' Raman spectra, such as neon emission lines, broadband white light lamps [16], NIST standards [17], and standard organic compounds [18], for wavenumber and intensity calibration. From a data processing perspective, techniques such as aligning with reference spectra [19] or modeling approaches like ANOVA-simultaneous component analysis (ASCA) [20] and extended multiplicative signal correction (EMSC) [21,22] can significantly reduce variations in Raman spectra. However, these methods are insufficient to address the issues arising from signal nonlinearity. To further explore the inter-instrument variation of Raman spectrometers, Guo et al. [23] proposed a large-scale cross-laboratory study to evaluate the setup-dependence of Raman spectra. Their findings indicated that the routinely applied spectrometer calibration does not fully eliminate all setup-related effects on Raman spectra. Indeed, variations are inevitable in Raman spectroscopy measurements due to numerous factors, contributing to inter-instrument variation.

In an effort to effectively address the inter-instrument variation in Raman spectra, several standard Raman spectral databases have been established by researchers and institutions for reference and comparison. Examples include the RRUFF database [24], which contains Raman spectra for minerals; Sadtler, featuring a Raman spectra database for organic and inorganic compounds; SDBS [25], which houses a Raman spectra database for organic compounds; and PNNL [26], offering a near-infrared Raman spectra database for chemicals and common substances. Several of these Raman spectra databases, when combined with deep learning algorithms, have been successfully applied to material identification. For instance, Liu et al. [7] achieved an 88.4% classification accuracy using deep convolutional neural networks on the RRUFF dataset, outperforming non-deep learning methods. Zhang et al. [8] employed transfer learning methods based on Deep Neural Network (DNN) and Convolutional Neural Network (CNN), which respectively improved the classification accuracy by 5.0% and 4.1% compared to non-transfer models. Contrastive representation learning has also shown promising performance in classification experiments conducted on multiple publicly available Raman spectral datasets [27]. In the context of linear regression models, Park et al. [28] proposed the use of the minimum-variance network (MVNet) based on self-supervised learning to reduce inter-laboratory variation in surface-enhanced Raman scattering spectroscopy (SERS) quantitative analysis. However, it is important to note that most of these databases and applications primarily target materials, organic and inorganic compounds, and the like. These substances typically exhibit high signal-to-noise ratios and clear Raman characteristic peaks, and are thus less affected by variations. More specifically, these algorithms tolerate minor wavenumber and intensity drifts.

However, in biological applications, the Raman spectra collected often exhibit relatively low signal-to-noise ratios. For instance, the Raman spectra of certain bacteria may bear significant similarities. Consequently, the continuous signal performance in the fingerprint region may be quite subtle, making it easier for the aforementioned differences to obscure or confound some Raman signals. This low repeatability also complicates the design of standard Raman databases for reference in biological applications [23]. Despite these challenges, researchers have not been deterred from exploring algorithms for Raman spectroscopy in biological applications. For example, Ho et al. [1] constructed a Raman spectral dataset of 30 common bacterial pathogens, representative of the majority of infections in intensive care units worldwide, and achieved a classification accuracy of 82.2% using the Residual Network (ResNet). Liu et al. [2] employed the progressive growing of Generative Adversarial Net (PGGAN) to generate high-resolution synthetic Raman spectra for data augmentation. This approach yielded a 99% accuracy in detecting deep-sea microorganisms using the ResNet in Raman spectroscopy. In another study, Liu et al. [3] utilized Long Short-Term Memory (LSTM) to rapidly identify biological Raman spectra with low signal-to-noise ratios. The Transformer [29], renowned for its excellent performance in attention mechanisms in the fields of Natural Language Processing and image processing, has also been applied to the identification of biological tissues [30] and microorganisms [4] using Raman spectroscopy. While these different architectural models have achieved commendable classification results in Raman spectroscopy tasks, they have not addressed the issue of inter-instrument variation. The Raman spectra they utilized were either collected by a single Raman spectrometer, or the dataset was shuffled during model training.

In summary, the challenge of inter-instrument variation in Raman spectroscopy for biological applications is significant. To address this, we propose a classification model based on a modular design incorporating multiple projection layers in the Siamese network. Siamese networks have found applications in various fields of analytical chemistry, including Raman spectroscopy [9,27,31], mass spectroscopy [32,33], and near-infrared spectroscopy [34]. However, they have been predominantly applied to single datasets and do not consider inter-instrument variation. Our model diverges from the direct training of a traditional multi-classification network. Instead, it compares whether Raman spectra collected by two instruments belong to the same class, thereby outputting the feature distance between the two spectra, which is then mapped into a similarity value. By comparing the similarity values with the reference spectra of each known class, the class of the unknown spectrum can be determined. Specifically, our research work has made the following contributions:

1. We assembled three Raman spectral datasets, representing data collected from different instruments and at varying resolutions, using 10 common pathogenic bacteria. We performed an inter-instrument variation analysis on these datasets using clustering and classification algorithms.

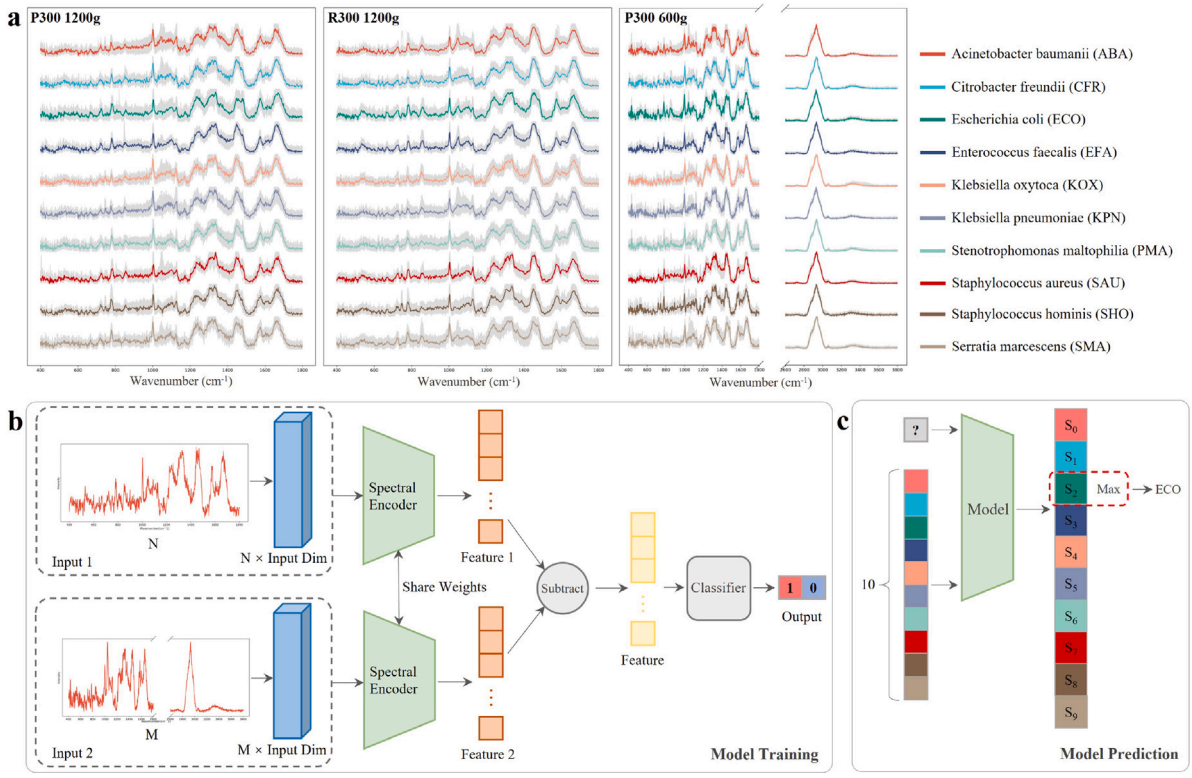


Fig. 1. Visualization of datasets and the architecture of our Siamese network. **a**: Visualization of preprocessed spectra of the three datasets. Each bacterium has 400 spectra, and solid lines represent the average spectrum for each bacterium in each dataset. **b**: Schematic representation of the Siamese network trained with spectra inputs of different lengths. The same class spectra training label is 1, and different class spectra training label is 0. **c**: Schematic representation of the model using reference spectra to predict unknown spectra, where the class corresponding to the reference spectrum with the highest similarity value is assigned as the prediction class.

2. We adopted a modular design approach, which facilitates the easy integration and removal of Raman spectroscopy encoders. We also compared the classification effects of three different architectural models — Transformer, ResNet, and LSTM — as core spectral encoders.
3. Our model does not require deliberate data augmentation and can be trained effectively with only a small amount of data, achieving superior classification accuracy compared to the baseline algorithm.
4. Thanks to the design of multiple projection layers, the dimensional projection transformation of the input layer can maximize the deconstruction of the data input and model coupling relationship, enabling the fusion of Raman data of different resolutions (different data lengths) for training.
5. We have also verified the usability of our model on CPUs.

In addition, the implementation code of Python for this research is available on GitHub page (https://github.com/lonseajim/Siamese-Network).

2. Materials and methods

2.1. Sample preparation

We selected 10 bacteria for our study: Acinetobacter baumannii (ABA), Citrobacter freundii (CFR), Escherichia coli (ECO), Enterococcus faecalis (EFA), Klebsiella oxytoca (KOX), Klebsiella pneumoniae (KPN), Stenotrophomonas maltophilia (PMA), Staphylococcus aureus (SAU), Staphylococcus hominis (SHO), and Serratia marcescens (SMA). These are common pathogenic bacteria, and constructing datasets using their Raman spectra provides a more universal approach.

We cultured and prepared samples of these 10 bacterial strains using a uniform method. Each strain was inoculated at a 1% seeding rate into 2 mL of Lysogeny broth (LB) and placed in a shaking incubator overnight (incubation conditions: 37 °C, 180 rpm). After incubation, 1 mL of each bacterial culture was transferred to clean, sterile 1.5 mL EP tubes. The samples were vortexed and then centrifuged at 9000 rpm for 2 min to collect 100 μL of bacterial sediment. The sediment was resuspended in sterile water, and this process was repeated three times. After the final wash, 100 μL of bacterial sediment was retained and resuspended in 900 μL of sterile water. Each bacterial sample was diluted to an appropriate concentration, vortexed for 2 min, and 2 μL of bacterial suspension was pipetted onto a chip and air-dried for further use.

2.2. Raman spectral datasets

To construct datasets suitable for inter-instrument variation analysis, we collected three sets of Raman spectra from each of the 10 pathogenic bacterial samples, as illustrated in Fig. 1a:

First dataset: Using the P300 Raman spectrometer (Hooke Instruments, Changchun, China. Objective: Olympus, 100×, NA=0.8) with a 600 g/mm grating, we gathered 400 Raman spectra for each bacterium. The spectral range of the data spans from 400 cm^{-1} to 3800 cm^{-1}, with a raw data length of 1701. We have named this dataset "P300 600 g".

Second dataset: Using the P300 Raman spectrometer with a 1200 g/mm grating, we collected 400 Raman spectra for each bacterium. The spectral range of the data spans from 400 cm^{-1} to 1800 cm^{-1}, with a raw data length of 701. We have named this dataset "P300 1200 g".

Third dataset: Using the R300 Raman spectrometer (Hooke Instruments, Changchun, China. Objective: Olympus, 100×, NA=0.8) with a 1200 g/mm grating, we collected 400 Raman spectra for each bacterium. The spectral range of the data spans from 400 cm^{-1} to 1800 cm^{-1}, with a raw data length of 701. We have named this dataset "R300 1200 g".

Table 1
Number of base layers used for different models.

Model	Description of base layers
Classification-ResNet	Residual Layer × 6
Classification-Transformer	Transformer Encoder Layer × 3
Classification-LSTM	LSTM Layer × 2
Siamese-ResNet	Residual Layer × 1
Siamese-Transformer	Transformer Encoder Layer × 1
Siamese-LSTM	LSTM Layer × 1

These three datasets collectively encompass 12,000 Raman spectra. Each spectrum was exposed for 5 s, with an excitation wavelength of 532 nm and a laser power of 5 mW.

For clustering analysis, classification tasks, and Siamese network classification tasks involving spectra of the same resolution, we utilized datasets with the same data length, namely the P300 1200 g and R300 1200 g datasets. For Siamese network classification tasks involving spectra of different resolutions, we employed datasets with varying data lengths, specifically the P300 600 g and R300 1200 g datasets.

In the subsequent content, the dataset name will be used as a prefix, and the bacteria abbreviation will be used as a suffix to denote specific bacterial categories within a dataset. For instance, "P300 600g-ABA" refers to the Raman spectral data of Acinetobacter baumannii in the P300 600 g dataset.

The datasets can be obtained by contacting the corresponding author.

2.3. Data preprocessing

We applied conventional preprocessing steps to all Raman spectra, including spike removal, baseline subtraction, and normalization [3]. Baseline subtraction was executed using the adaptive iteratively reweighted penalized least squares (airPLS) algorithm, and normalization was conducted using the min–max normalization method. The preprocessing algorithm has been submitted to GitHub and was executed with default parameters. For the data obtained with the 1200 grating, the spectral range and data length remained unchanged after preprocessing, spanning 400 cm^{-1} to 1800 cm^{-1} with a data length of 701. Using the P300 with the 600 g/mm grating results in a reduced resolution compared to the 1200 g/mm grating, but it allows for the capture of spectral features in the high wavenumber region. However, as the peaks of bacterial Raman spectra in the high wavenumber region are strong, normalization based on this region may result in small signal values in the fingerprint region. First, the raw spectra were subjected to the aforementioned spike removal and baseline subtraction. Then, to maintain the signal intensity in the fingerprint region, the spectra were cropped into the fingerprint region (400 cm^{-1} to 1800 cm^{-1}) and the high wavenumber region (2600 cm^{-1} to 3800 cm^{-1}). Each region was individually normalized using a min–max normalization method, and finally, the two regions were concatenated. The data length was reduced from the original 1701 to 1302. The discarded wavenumber range of 1800 cm^{-1} to 2600 cm^{-1} is the signal-silent region of biological Raman spectra. This cropping operation reduces the data length and conserves model training resources.

These preprocessing algorithms are implemented in Python 3.10.

2.4. Classification task

Models: We utilized three models with different architectures: ResNet [35], a variant based on CNN; Transformer [29], based on a self-attention mechanism; and LSTM [36], a variant based on Recurrent Neural Network (RNN). All three models are representative of their respective architectures. Fig. 2a illustrates the basic structure of the ResNet. It can learn spectral features of varying depths through convolutional blocks and addresses the degradation problem in deep networks by utilizing residual connections, allowing for more effective feature extraction. Fig. 2b shows the basic structure of the Transformer. Its attention mechanism enables parallel processing of sequence information, improving training and inference speed, while also capturing long-range dependencies more effectively. Fig. 2 c presents the basic structure of the LSTM. It excels at handling sequential information by utilizing gated units that efficiently retain and leverage both long-term and short-term memory. The use of the base layer of each classification model is illustrated in Table 1, where we aimed to align as closely as possible with our referenced research work [1–4] in order to validate the spectral encoding capabilities of the models in our dataset.

Training: The classification task was trained using a conventional dataset splitting ratio of 8:2 for the training and validation set data volumes. The loss function used was CrossEntropyLoss, and the batch size was set to 512. The highest accuracy in the validation set was saved as the best model. We conducted five cross-validations to verify the stability of the model.

Cross-dataset testing: We tested using all data from another dataset. For instance, if you train on the P300 1200 g dataset, use the R300 1200 g dataset for cross-dataset testing.

2.5. Siamese network

Model: Drawing inspiration from the Clip model [37], we introduced multiple projection layers into the concatenated network to realize the modular design of the model, as depicted in Fig. 1b. Initially, we connected a linear projection layer after the Raman spectral inputs. This enabled us to project two different inputs to the same dimension, or even different data lengths. It also made it possible to decouple the inputs from the model. This arrangement allows a set of data inputs and the corresponding layer linear projection layer to form a learnable unit. For the same classification task, model fine-tuning can be accomplished by freezing the other layers and learning only the linear projection layer corresponding to the new set of data. This design significantly conserves training resources.

We used the base layer of the three models as a spectral encoder and passed the output of the encoder through a linear projection layer that realizes the projection of the output feature values to a fixed dimension. This design decouples the spectral encoding layer from the classifier, thereby realizing the pluggability of the spectral encoder. Due to the small absolute quantity of training data, to prevent the model from becoming overly large and leading to overfitting, each Siamese network model, as outlined in Table 1, employs just one base layer as a spectral encoder.

As shown in Fig. 2d, two linear layers and one ReLU layer are used to realistically form a binary classifier. This is finally normalized with Sigmoid to obtain a value representing the spectral similarity.

Training: Two inputs to the model are treated as a positive pair if they belong to the same class, and as a negative pair if they belong to different classes. If the two datasets each have N classes with M spectra per class, then the absolute number of training data is 2NM, while the number of input pairs is $(NM)^2$. This expands the training set and naturally facilitates data enhancement. However, the positive–negative contrast is $\frac{1}{N-1}$. When $N = 10$ and $M = 10$, the ratio of positive to negative sample size is $\frac{1}{9}$. Due to the imbalance in the ratio of positive to negative samples, we maintain the ratio of positive and negative pairs in each training batch at 0.5 by randomly sampling the positive and negative pairs [27]. The batch size is set to 1024 so that during each training process, more non-duplicate data can be extracted each time to improve the coverage of the training data.

The loss function is calculated such that if the input is a positive pair, the similarity value approaches 1; otherwise, the similarity value approaches 0. The loss function uses Binary Cross Entropy (BCE) Loss. Due to the small training set and the large test set, the model is trained to save the best model with a loss value less than 0.01. We extracted

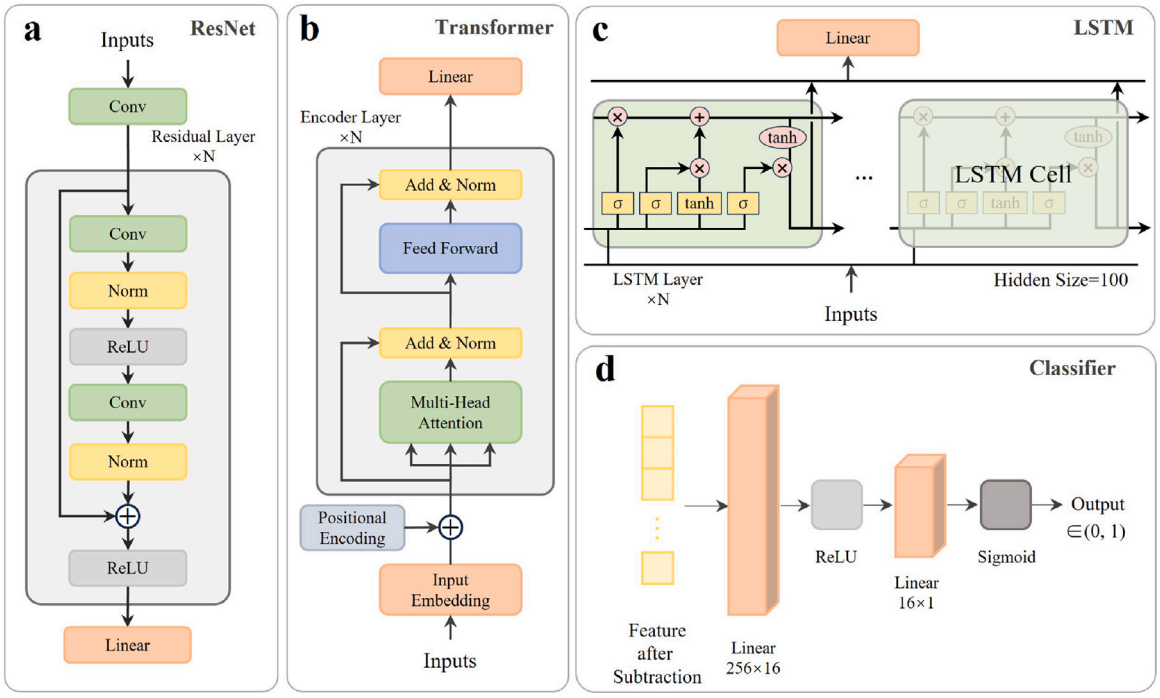


Fig. 2. Structures of various models. **a**: The ResNet model consists of 1 convolutional layer, *N* residual layers, and 1 linear layer. **b**: The Transformer model consists of 1 embedding layer, *N* encoder layers, and 1 linear layer. **c**: The LSTM model consists of *N* LSTM layers and 1 linear layer. **d**: Design of the classifier module in the Siamese network.

the training set five times without repetition to verify the stability of the model.

Prediction: One spectrum from each of the 10 classes in the training set of one dataset is selected as the reference spectral group, as illustrated in Fig. 1 c. The 3900 spectra in the other dataset, which are not involved in the training, are each modeled with the reference spectral group for prediction. Each time, 10 similarity values are output, and the class with the highest value is considered as the predicted class.

Both the classification tasks and the Siamese network classification tasks employed the Adam optimizer with a learning rate of 0.0001 and betas set to (0.5, 0.999) [1]. All deep learning models were constructed using PyTorch2.0. Training was carried out on a GPU (NVIDIA Tesla V100 32G).

2.6. Model testing on CPU

To validate the deployability and usability of our model, we directly assessed the training and prediction time of the Siamese network on the CPU of the computer (CPU: Intel i7-11700, RAM: 32 GB, GPU: Not available) that controls the Raman spectrometer. We utilized two datasets with the same resolution: P300 1200 g and R300 1200 g.

3. Results and discussion

3.1. Cluster analysis

Before conducting the clustering analysis, we first calculated the signal-to-noise ratio (SNR) for the Raman spectra of the two datasets, P300 1200 g and R300 1200 g. The signal-to-noise ratio serves as a crucial index for evaluating the spectral output capability of Raman spectrometers, and it also reflects the expression capability of different spectrometers for the same substance. A large difference in the signal-to-noise ratio indicates a significant disparity between the two spectrometers, necessitating adjustments to the instrument settings.

As per Fig. 3a, the average signal-to-noise ratio for the Raman spectra of these 10 bacteria generally remained between 4–7 for the 1200 g/mm grating settings of the two spectrometers. The signal-to-noise difference between the Raman spectra of the same bacteria on both spectrometers did not exceed 2. Combined with the spectral curves in Fig. 1a, it is evident that the spectral expressivity of the two Raman spectrometers is closely matched, and that the selected 10 bacterial samples performed stably on the two Raman spectrometers. However, we currently cannot predict, let alone completely eliminate, the more subtle differences in spectrometer settings. This introduces inter-instrument variation in Raman spectroscopy for biological applications that can affect the clustering and classification results of the data.

Next, we performed a t-Distributed Stochastic Neighbor Embedding (t-SNE) 2D clustering analysis on both datasets. In the t-SNE 2D clustering visualization results shown in Fig. 3 c, the distribution of instrument categories versus bacterial categories is clear. The blue dashed box represents the R300 1200 g dataset and the red dashed box represents the P300 1200 g dataset. Except for P300 1200g-EFA and R300 1200g-SHO, the overall distribution of the P300 1200 g data is above the black dotted line and the distribution of bacterial categories is more dispersed. Conversely, the R300 1200 g data has an overall distribution below the black dashed line and a relatively concentrated distribution of six bacterial categories, ABA, KPN, CFR, PMA, KOX, and SMA. Notably, both bacteria, ABA and KPN, in both datasets showed significant distributional confusion. However, in an ideal clustering situation without inter-instrument variation, signs of the same color representing the same bacterial category should be distributed closer together while maintaining a distance from the distribution of other bacterial categories. It is thus clear that the inter-instrument variation even exceed the category differences of some bacteria.

3.2. Results of the classification task

In order to evaluate the feature extraction capability and the generalization ability of the three models - Transformer, ResNet, and LSTM - for the application of Raman spectroscopy in cross-dataset classification, we conducted classification tests using each of the three models. As shown in Table 2, all three models achieved over 98% classification

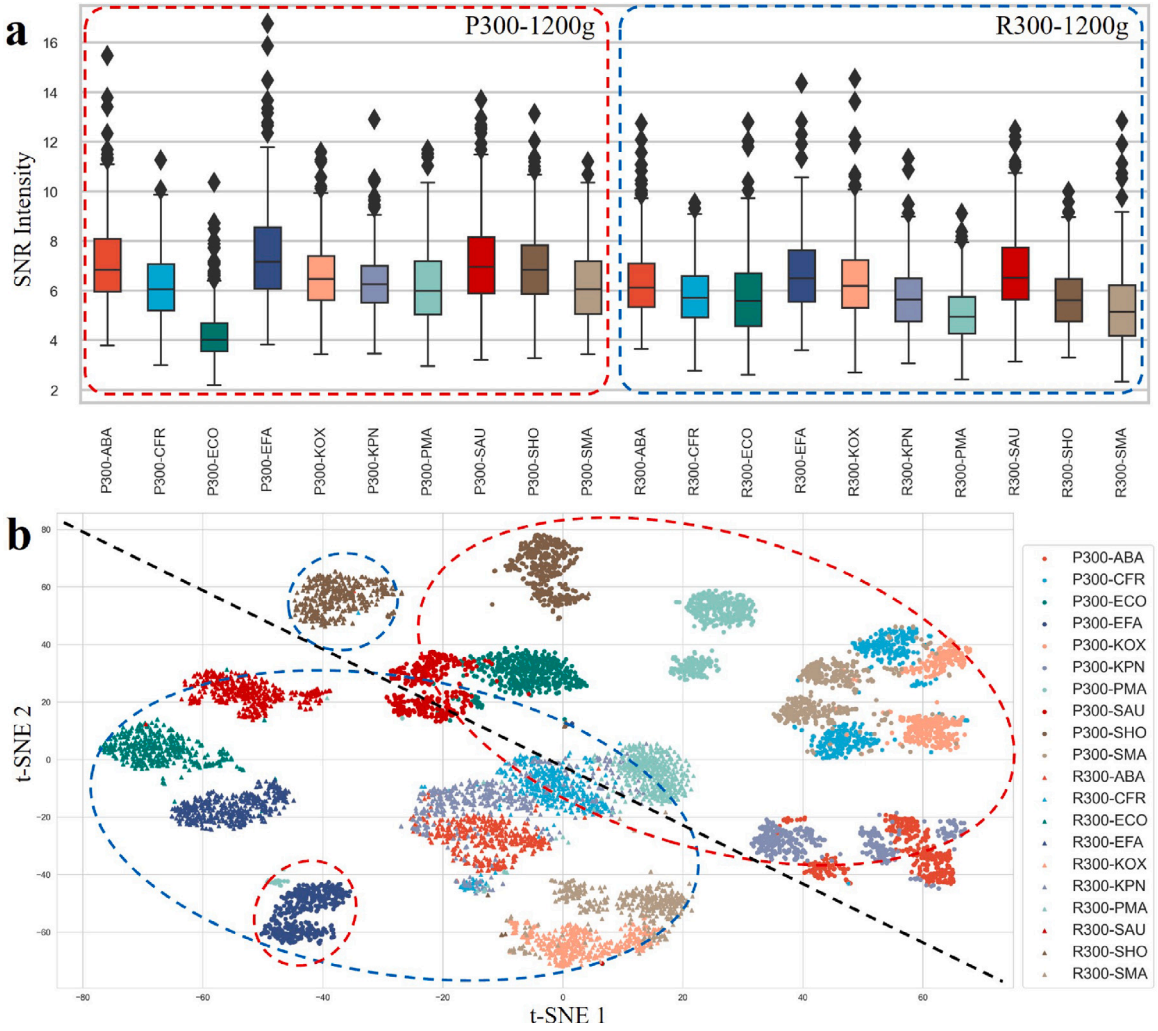


Fig. 3. Signal-to-noise ratio (SNR) and cluster analysis results for the P300 1200 g and R300 1200 g datasets. **a**: SNR of two datasets based on bacterial classes. **b**: t-SNE 2D clustering analysis based on bacterial classes.

Table 2
Model training on P300 1200 g, validation accuracy on P300 1200 g, and test accuracy on R300 1200 g.

Dataset	Classification-Transformer	Classification-ResNet	Classification-LSTM
Validation on P300 1200 g	98.6 ± 0.9%	99.2 ± 0.4%	**99.9 ± 0.1%**
Test on cross-datasets R300 1200 g	19.1 ± 1.0%	20.8 ± 2.7%	**39.1 ± 1.6%**

Table 3
Model training on R300 1200 g, validation accuracy on R300 1200 g, and test accuracy on P300 1200 g.

Dataset	Classification-Transformer	Classification-ResNet	Classification-LSTM
Validation on R300 1200 g	96.4 ± 1.1%	96.5 ± 0.6%	**98.8 ± 0.5%**
Test on cross-datasets P300 1200 g	19.4 ± 0.9%	28.9 ± 4.4%	**45.1 ± 4.5%**

accuracy on the P300 1200 g training-validation set. However, the bset model, Classification-LSTM, only achieved an accuracy of 39.1 ± 1.6% on the cross-dataset R300 1200 g. Similarly, as per Table 3, all three models also achieved over 96% classification accuracy on the R300 1200 g training-validation set, but the Classification-LSTM model only achieved an accuracy of 45.1 ± 4.5% on the cross-dataset P300 1200 g.

From a dataset perspective, the classification models trained on the R300 1200 g dataset generally have lower accuracy than the models trained on the P300 1200 g dataset, but they perform better in terms of generalization on the other dataset. This can be explained by referring to Fig. 3 c: the distribution of each category in the R300 1200 g dataset is more spread out, meaning the features learned by the classification model will also be more spread out, reducing the weight of salient features. Such a model would increase the likelihood of misidentification. However, Raman spectra from the same category between instruments are inherently different, and dispersed features are more likely to cover spectra with differences, resulting in better generalization.

From a modeling perspective, the validation accuracies of the three models are close, and all perform well, demonstrating excellent feature extraction capabilities for Raman spectra. However, none of the models show good cross-dataset classification generalization ability. Even for the best model, Classification-LSTM, the test on the cross-dataset accuracy drops more than 50% compared to the validation accuracy. This further validates that for biological applications, the presence of inter-instrument variation has a significant impact on classification.

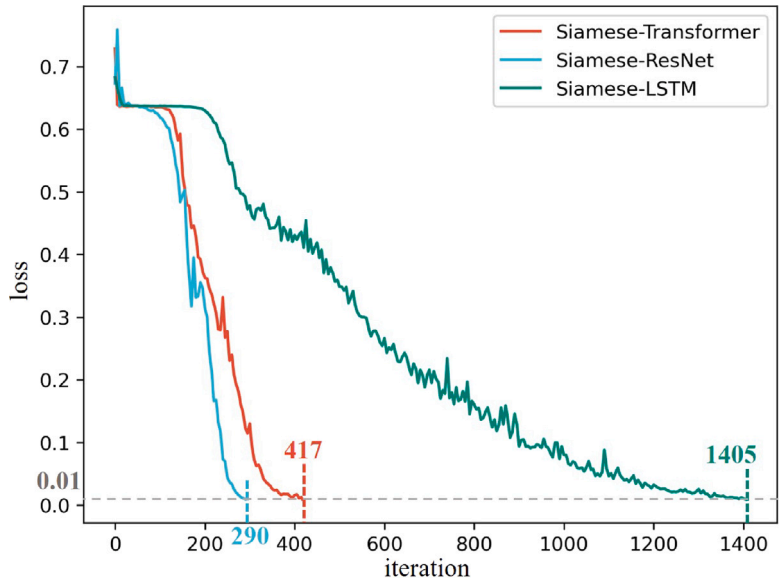


Fig. 4. Loss curves for three Siamese network models and training iterations, with the training objective set loss to 0.01.

Table 4
Classification accuracy when predicting P300 1200 g using R300 1200 g as the reference set.

Model	Accuracy
1NN Euclidean	24.9 ± 2.4%
1NN Cosine	25.5 ± 3.5%
1NN Manhattan	23.7 ± 2.3%
Siamese-Transformer	87.0 ± 3.8%
Siamese-ResNet	**92.6 ± 1.2%**
Siamese-LSTM	73.5 ± 2.0%

Table 5
Classification accuracy when predicting R300 1200 g using P300 1200 g as the reference set.

Model	Accuracy
1NN Euclidean	29.0 ± 3.4%
1NN Cosine	31.9 ± 3.1%
1NN Manhattan	27.8 ± 3.6%
Siamese-Transformer	83.2 ± 1.7%
Siamese-ResNet	**84.4 ± 2.5%**
Siamese-LSTM	61.9 ± 2.5%

If a classification model is to be used across datasets, the current conventional method is to use model fine-tuning [8,38], which involves retraining an already trained classification model on a new dataset with a relatively small amount of data from the new dataset. While this can yield good results on the new dataset, there are two issues:

1. The accuracy of the fine-tuned model on the original dataset may decrease.
2. The process of pre-training the model still requires a large amount of data.

3.3. Siamese network classification task with the same resolution spectra

In the classification experiments, it has been shown that all three models – Transformer, ResNet, and LSTM – have the ability to extract excellent features from spectra. In this experiment, since the training set only uses 10 spectra from each type of bacterial Raman spectra, it is challenging to train with traditional classification models without data augmentation. Therefore, in the comparison experiments, we did not train traditional classification models for comparison. Instead, we chose 1NN Euclidean, 1NN Cosine, and 1NN Manhattan – three spectral matching algorithms commonly found in commercial software [31] as the baseline for comparison. The comparison was made with three Siamese networks with different spectral encoders, namely Siamese-Transformer, Siamese-ResNet, and Siamese-LSTM.

As demonstrated in Tables 4 and 5, among the three baseline methods, the highest classification accuracy for the two datasets using similarity measures was achieved by the 1NN Cosine algorithm, with respective accuracies of 25.5 ± 3.5% and 31.9 ± 3.1%. Even the worst-performing model among the three different spectral encoders, Siamese-LSTM, has accuracies of 73.5 ± 2.0% and 61.9 ± 2.5%. It is evident that the accuracy of the three Siamese networks is significantly higher than that of the three baseline algorithms, and also higher than that of the cross-dataset test in the classification task.

From the classification results of the concatenated networks with three different spectral encoders, Siamese-ResNet has the best classification accuracy on both datasets, with respective accuracies of 92.6 ± 1.2% and 84.4 ± 2.5%. Moreover, as shown in Fig. 4, Siamese-ResNet reduces the loss value to 0.01 with the least number of iterations. Under the same experimental conditions, the encoder based on the ResNet architecture demonstrated superior capability in extracting spectral features. This reflects the characteristic strength of CNN models in capturing both shallow and deep features, leading to a more effective representation of spectral features. Therefore, among the three models, the Siamese-ResNet is more suitable for classification tasks in a Siamese network configuration.

The Siamese network is characterized by its ability to map the feature distributions of each class of spectra in two datasets to each other. Features of similar spectra are mapped closer, while features of different spectra are mapped further apart. This is why we can achieve good classification results even when we use only a small amount of spectra in the training set, as long as this small amount of spectra can roughly represent the feature distribution of each class. From the classification results, the accuracy of predicting the P300 1200 g dataset using the R300 1200 g dataset was generally higher than predicting the R300 1200 g dataset using the P300 1200 g dataset. This indicates that the distribution of each class in the P300 1200 g dataset was more significant than that in the R300 1200 g dataset, a conclusion also corroborated by the clustering results in Fig. 3b.

In Fig. 5a, the prediction results of the P300 1200 g dataset show that 21% of KPN spectra were predicted as ABA, and 15% of SMA

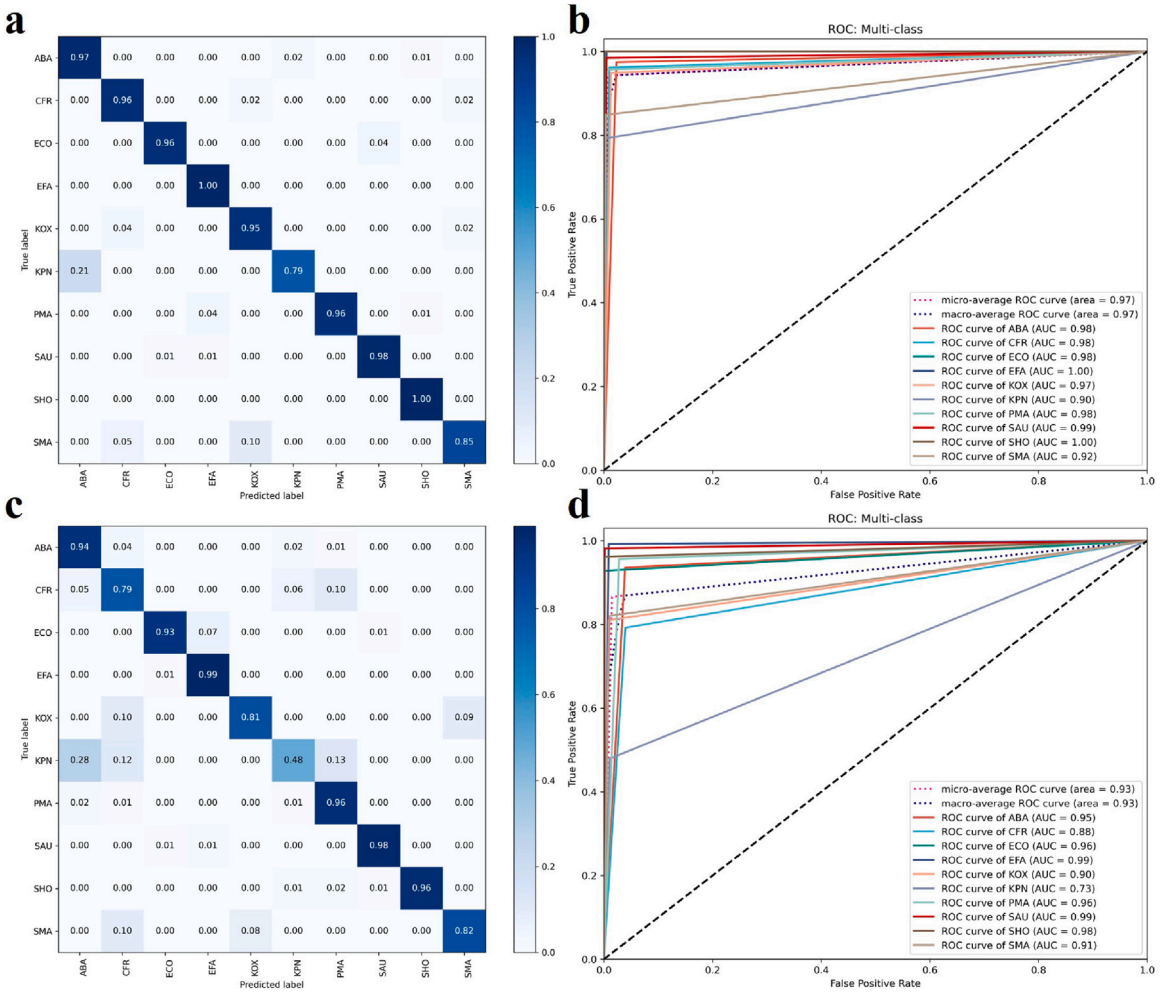


Fig. 5. Results of mutual prediction of datasets with the same resolution using Siamese-ResNet. Confusion matrix (**a**) and ROC curve (**b**) for predicting the P300 1200 g using the R300 1200 g as the reference set. Confusion matrix (**c**) and ROC curve (**d**) for predicting the R300 1200 g using the P300 1200 g as the reference set.

Table 6
Classification accuracy when predicting P300 600 g using R300 1200 g as the reference set.

Model	Accuracy
Siamese-Transformer	93.2 ± 0.8%
Siamese-ResNet	**94.1 ± 1.9%**

Table 7
Classification accuracy when predicting R300 1200 g using P300 600 g as the reference set.

Model	Accuracy
Siamese-Transformer	82.6 ± 1.3%
Siamese-ResNet	**84.5 ± 2.4%**

spectra were predicted as CFR and KOX. These results align with the clustering analysis, where the distribution of KPN spectra in the P300 1200 g dataset clustered with ABA spectra, and the distribution of SMA, CFR, and KOX spectra partially clustered together. In Fig. 5c, the prediction results of the R300 1200 g dataset show that the classification accuracy of CFR, KOX, KPN, and SMA was significantly lower, while the categories most often predicted were ABA, CFR, and PMA. The classification of these six classes of bacteria is consistent with the distribution observed in the clustering analysis.

The receiver operating characteristic (ROC) curves in Figs. 5b and d show that the Siamese network classifiers exhibited high specificity, even for KPN, which performed poorly in both datasets. This may be due to the large ratio difference between positive and negative pairs during training.

3.4. Siamese network classification task with the different resolution spectra

We selected the Siamese-ResNet model, which performed the best in the classification experiments with spectra of the same resolution, and the Siamese-Transformer model, which exhibited a more balanced performance, for the classification experiments with different resolutions. Due to the multiple projection layers and modular structure design, we only needed to modify the input dimension of one linear projection layer to apply the model to spectra with different resolutions.

As displayed in Table 6, the accuracy of predicting the P300 600 g dataset using the R300 1200 g dataset exceeds 93%. Remarkably, for the Siamese-Transformer, the accuracy was improved by 6.2% compared to predicting the P300 1200 g dataset, almost reaching the same level as Siamese-ResNet. As shown in Fig. 6, after adding the signal from the high wavenumber region of the Raman spectra, the clustering results of the P300 600 g dataset were more dispersed compared to the P300 1200 g dataset. There were no apparent instances of mixed distributions among different spectral classes. The distributions of KPN and ABA were notably separated, leading to a significant improvement in the classification accuracy of KPN, as shown in Fig. 7a. This was a key factor contributing to the overall improvement in accuracy.

As per Table 7 and Fig. 7c, the performance of predicting the R300 1200 g dataset using the P300 600 g dataset was consistent with that of

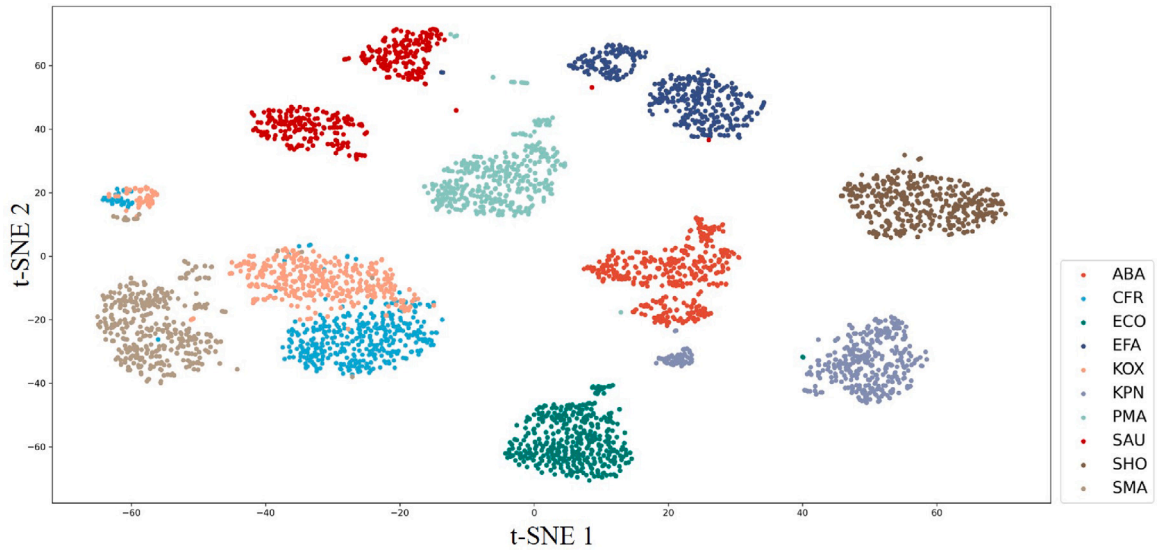


Fig. 6. t-SNE 2D clustering analysis based on bacterial classes of P300 600 g.

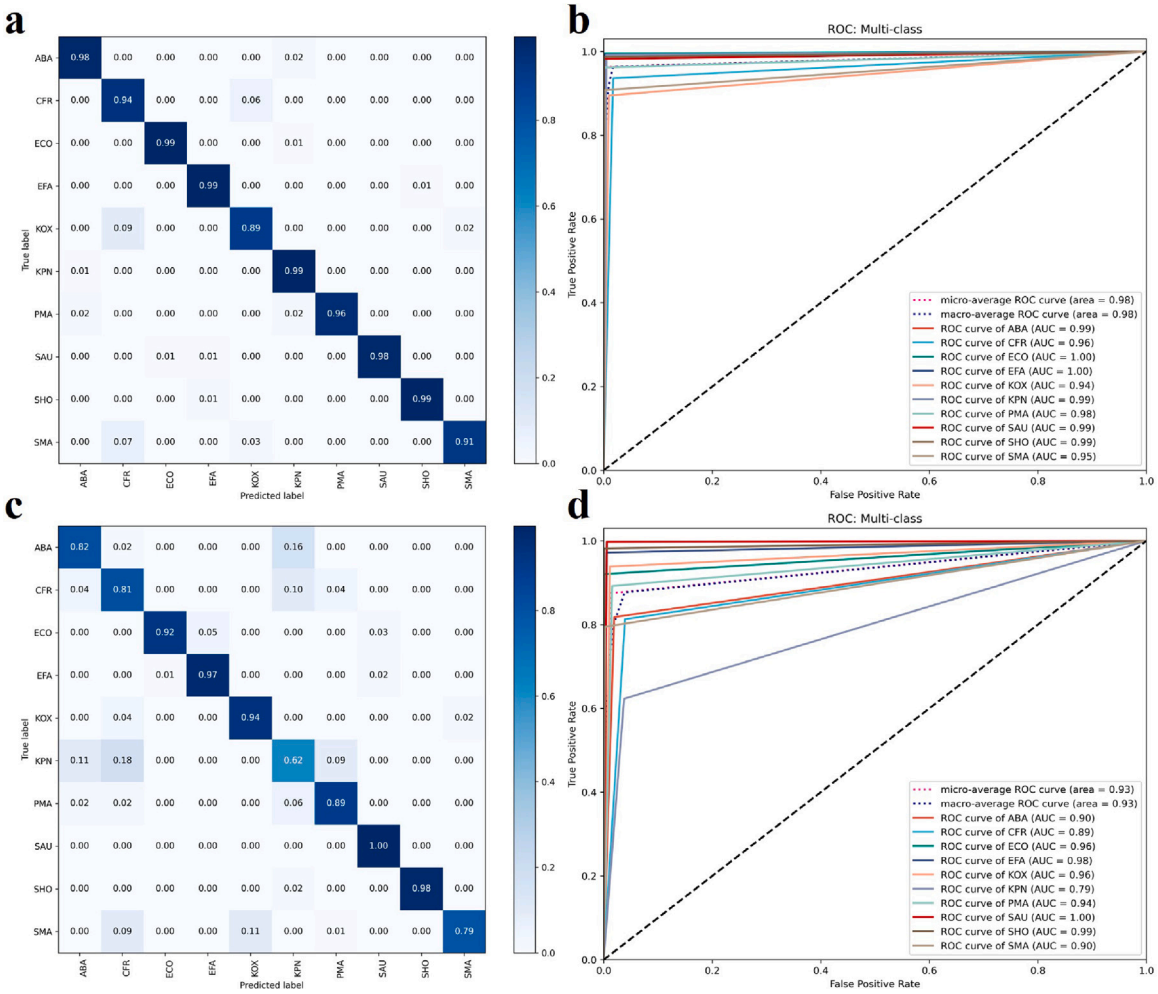


Fig. 7. Results of mutual prediction of datasets with the different resolution using Siamese-ResNet. Confusion matrix(**a**) and ROC curve(**b**) for predicting the P300 600 g using the R300 1200 g as the reference set. Confusion matrix(**c**) and ROC curve(**d**) for predicting the R300 1200 g using the P300 600 g as the reference set.

Table 8
Training and prediction duration of Siamese networks.

Task	Model	On GPU(NVIDIA Tesla V100 32G)	On CPU(Intel i7-11700)
Training one iteration	Siamese-ResNet	1.045 s	9.454 s
	Siamese-Transformer	**1.021 s**	**2.152 s**
Predict one spectrum	Siamese-ResNet	0.023 s	0.090s
	Siamese-Transformer	**0.018 s**	**0.036 s**

predicting it using the P300 1200 g dataset. Therefore, the classification accuracy of our model appeared to be more closely related to the spectral feature distribution of the dataset itself. The ROC curves in Figs. 7b and d still demonstrated the high specificity of the model, even at low sensitivity.

3.5. Siamese network testing on CPU

Taking into account the high cost of GPU hardware, we carried out performance testing of our model on a CPU to increase its universality. Given that the Siamese-LSTM did not perform well in the same resolution spectral classification task, it was not included for comparison in this experiment.

As shown in Table 8, it is apparent that the time required to train one iteration of the Siamese-Transformer model on the CPU is approximately $\frac{1}{4}$ of that of the Siamese-ResNet. However, in Fig. 4, the iterations of Siamese-Transformer completes one training only about 100 more than Siamese-ResNet. When taken together, completing one model training on the CPU is 1800s (30 min) less with the Siamese-Transformer encoder than with the Siamese-ResNet. This is attributed to the ability of the Transformer encoder to process sequential data in parallel through its attention mechanism, thereby reducing training time. It is more challenging for the CPU to perform convolutional computations in CNN networks, which results in a significantly higher training time on the CPU for Siamese-ResNet. Although the time of predicting a spectrum using the Siamese-Transformer is also advantageous, it is not noticeable in the experiments because the absolute times are all relatively small.

Therefore, in a laboratory equipped with high-performance GPUs, selecting the ResNet spectral encoder might result in higher accuracy. However, if only CPUs are available and there is a necessity to reduce the training duration, the Transformer spectral encoder is also an excellent choice.

4. Conclusion

In this paper, we introduce a modular Siamese network classification model with multiple projection layers to address the issue of classifying Raman spectra with inter-instrument variation in biological applications. We explain the working principle of the model in conjunction with the clustering distribution results of Raman spectra. The model extracts features from spectra collected by various Raman spectrometers and performs a distance-to-similarity mapping between the features, which still maintains a high classification accuracy with minimal data for training. The design of multiple projection layers is also utilized to execute the fusion training and mutual prediction of Raman spectra of different lengths. The design of one input corresponding to one linear projection layer simplifies the fine-tuning of the new input. The decoupled modular design makes it easier to plug and unplug the core spectral encoder module. Our model can also be seamlessly deployed to CPU platforms.

In few-shot learning, it can be challenging to encompass all features of a category when constructing a training set. This limitation may result in local features of the spectral data replacing global features, leading to a decrease in classification accuracy. Therefore, in our future research, we aim to explore a data augmentation algorithm that can simulate the distribution of Raman spectral features. This will allow us to train models with higher classification accuracy using less data. Additionally, we are dedicated to constructing larger and more comprehensive biological Raman spectral datasets. This would enable the training of a more robust Raman spectroscopy encoder as a pretraining model, facilitating broader applications through fine-tuning for downstream tasks and integrating various output modules.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank Hooke Instruments Ltd. for its Raman spectrometers.

References

[1] Chi-Sing Ho, Neal Jean, Catherine A. Hogan, Lena Blackmon, Stefanie S. Jeffrey, Mark Holodniy, Niaz Banaei, Amr A.E. Saleh, Stefano Ermon, Jennifer Dionne, Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning, Nat Commun. 10 (1) (2019) 4927.
[2] Bo Liu, Kunxiang Liu, Nan Wang, Kaiwen Ta, Peng Liang, Huabing Yin, Bei Li, Laser tweezers Raman spectroscopy combined with deep learning to classify marine bacteria, Talanta 244 (2022) 123383.
[3] Kunxiang Liu, Fuyuan Chen, Lindong Shang, Yuntong Wang, Hao Peng, Bo Liu, Bei Li, Deep learning-based ultra-fast identification of Raman spectra with low signal-to-noise ratio, J. Biophotonics 17 (1) (2024) e202300270.
[4] Bo Liu, Kunxiang Liu, Xiaoqing Qi, Weijia Zhang, Bei Li, Classification of deep-sea cold seep bacteria by transformer combined with Raman spectroscopy, Sci. Rep. 13 (1) (2023) 3240.
[5] Stipe Lukin, Krunoslav Užarević, Ivan Halasz, Raman spectroscopy for real-time and in situ monitoring of mechanochemical milling reactions, Nat. Protoc. 16 (7) (2021) 3492–3521.
[6] Ermanno Miele, Wesley M. Dose, Ilya Manyakin, Michael H. Frosz, Zachary Ruff, Michael F.L. De Volder, Clare P. Grey, Jeremy J. Baumberg, Tijmen G. Euser, Hollow-core optical fibre sensors for operando Raman spectroscopy investigation of Li-ion battery liquid electrolytes, Nature Commun. 13 (1) (2022) 1651.
[7] Jinchao Liu, Margarita Osadchy, Lorna Ashton, Michael Foster, Christopher J. Solomon, Stuart J. Gibson, Deep convolutional neural networks for Raman spectrum recognition: a unified solution, Analyst 142 (21) (2017) 4067–4074.
[8] Rui Zhang, Huimin Xie, Shuning Cai, Yong Hu, Guo-kun Liu, Wenjing Hong, Zhong-qun Tian, Transfer-learning-based Raman spectra identification, J. Raman Spectrosc. 51 (1) (2020) 176–186.
[9] Jae-Hyeon Park, Hyeong-Geun Yu, Dong-Jo Park, Hyunwoo Nam, Dong Eui Chang, Dynamic one-shot target detection and classification using a pseudo-siamese network and its application to Raman spectroscopy, Analyst 146 (22) (2021) 6997–7004.
[10] Akifumi Takamizawa, Shinya Yanagimachi, Takeshi Ikegami, External cavity diode laser with very-low frequency drift, Appl. Phys. Express 9 (3) (2016) 032704.
[11] Akifumi Takamizawa, Shinya Yanagimachi, Takeshi Ikegami, Ryuzo Kawabata, External cavity diode laser with frequency drift following natural variation in air pressure, Appl. Opt. 54 (18) (2015) 5777–5781.

[12] Frank-Thomas Lentes, Marc K. Th Clement, Norbert Neuroth, Hans-Jürgen Hoffmann, Yuiko T. Hayden, Joseph S. Hayden, Uwe Kolberg, Silke Wolff, Optical properties, in: The Properties of Optical Glass, Springer, 1998, pp. 19–164.
[13] Satoshi Fukura, Tomoyuki Mizukami, Shoko Odake, Hiroyuki Kagi, Factors determining the stability, resolution, and precision of a conventional Raman spectrometer, Appl. Spectrosc. 60 (8) (2006) 946–950.
[14] Abdelkader Mestari, Robert Gaufrès, Patrice Huguet, Behaviour of the calibration of a Raman spectrometer with temperature changes, J. Raman Spectrosc. 28 (10) (1997) 785–789.
[15] Ankit Raj, Henryk A. Witek, Hiro-o Hamaguchi, Evaluating stability of a Raman spectrometer for long-time experiments, J. Raman Spectrosc. 52 (5) (2021) 1032–1047.
[16] Didier Hutsebaut, Peter Vandenabeele, Luc Moens, Evaluation of an accurate calibration and spectral standardization procedure for Raman spectroscopy, Analyst 130 (8) (2005) 1204–1214.
[17] Steven J. Choquette, Edgar S. Etz, Wilbur S. Hurst, Douglas H. Blackburn, Stefan D. Leigh, Relative intensity correction of Raman spectrometers: NIST SRMs 2241 through 2243 for 785 nm, 532 nm, and 488 nm/514.5 nm excitation, Appl. Spectrosc. 61 (2) (2007) 117–129.
[18] Jason D. Rodriguez, Benjamin J. Westenberger, Lucinda F. Buhse, John F. Kauffman, Standardization of Raman spectra for transfer of spectral libraries across different instruments, Analyst 136 (20) (2011) 4232–4240.
[19] Jason D. Rodriguez, Benjamin J. Westenberger, Lucinda F. Buhse, John F. Kauffman, Quantitative evaluation of the sensitivity of library-based Raman spectral correlation methods, Anal. Chem. 83 (11) (2011) 4061–4067.
[20] Age K. Smilde, Jeroen J. Jansen, Huub C.J. Hoefsloot, Robert-Jan A.N. Lamers, Jan Van Der Greef, Marieke E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data, Bioinformatics 21 (13) (2005) 3043–3048.
[21] Nils Kristian Afseth, Achim Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, Chemometr. Intell. Lab. Syst. 117 (2012) 92–99.
[22] Shuxia Guo, Achim Kohler, Boris Zimmermann, Ralf Heinke, Stephan Stockel, Petra Rosch, Jurgen Popp, Thomas Bocklitz, Extended multiplicative signal correction based model transfer for Raman spectroscopy in biological applications, Anal. Chem. 90 (16) (2018) 9787–9795.
[23] Shuxia Guo, Claudia Beleites, Ute Neugebauer, Sara Abalde-Cela, Nils Kristian Afseth, Fatima Alsamad, Suresh Anand, Cuauhtemoc Araujo-Andrade, Sonja Askrabic, Ertug Avci, et al., Comparability of Raman spectroscopic configurations: a large scale cross-laboratory study, Anal. Chem. 92 (24) (2020) 15745–15756.
[24] Barbara Lafuente, Robert T. Downs, Hexiong Yang, Nate Stone, Thomas Armbruster, Rosa Micaela Danisi, et al., The power of databases: the RRUFF project, Highlights Mineral. Crystallogr. 1 (2015) 25.
[25] Takeshi Saito, Shinichi Kinugasa, Development and release of a spectral database for organic compounds-key to the continual services and success of a large-scale database, Synth. Engl. Ed. 4 (1) (2011) 35–44.
[26] Brenda M. Kunkel, Yin-Fong Su, Russell G. Tonkyn, Eric G. Stephan, Alan G. Joly, Jerome C. Birnbaum, Kristin H. Jarman, Timothy J. Johnson, Raman database considerations for near-infrared systems, in: Optics and Photonics for Counterterrorism and Crime Fighting VII; Optical Materials in Defence Systems Technology VIII; and Quantum-Physics-Based Information Security, vol. 8189, SPIE, 2011, pp. 42–50.
[27] Bo Li, Mikkel N. Schmidt, Tommy S. Alstrøm, Raman spectrum matching with contrastive representation learning, Analyst 147 (10) (2022) 2238–2246.
[28] Seongyong Park, Abdul Wahab, Minseok Kim, Shujaat Khan, Self-supervised learning for inter-laboratory variation minimization in surface-enhanced Raman scattering spectroscopy, Analyst 148 (7) (2023) 1473–1482.
[29] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[30] Min Chang, Chen He, Yi Du, Yemin Qiu, Luyao Wang, Hui Chen, RaT: Raman transformer for highly accurate melanoma detection with critical features visualization, Spectrochim. Acta A 305 (2024) 123475.
[31] Jinchao Liu, Stuart J. Gibson, James Mills, Margarita Osadchy, Dynamic spectrum matching with one-shot learning, Chemometr. Intell. Lab. Syst. 184 (2019) 175–181.
[32] Florian Huber, Lars Ridder, Stefan Verhoeven, Jurriaan H. Spaaks, Faruk Diblen, Simon Rogers, Justin J.J. Van Der Hooft, Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships, PLoS Comput. Biol. 17 (2) (2021) e1008724.
[33] Florian Huber, Sven van der Burg, Justin J.J. van der Hooft, Lars Ridder, MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra, J. Cheminformatics 13 (1) (2021) 84.
[34] Zheng An-Bing, Yang Hui-Hua, Pan Xi-Peng, Yin Li-Hui, Feng Yan-Chun, On-site identification of counterfeit drugs based on near-infrared spectroscopy Siamese-network modeling, IEEE Access 9 (2020) 3195–3206.
[35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[36] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[37] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al., Learning transferable visual models from natural language supervision, in: International Conference on Machine Learning, PMLR, 2021, pp. 8748–8763.
[38] Saumya Singh, Dipak Kumbhar, Dhanya Reghu, Shwetha J. Venugopal, P.T. Rekha, Silpa Mohandas, Shruti Rao, Ambica Rangaiah, Sneha K. Chunchanur, Deepak Kumar Saini, et al., Culture-independent Raman spectroscopic identification of bacterial pathogens from clinical samples using deep transfer learning, Anal. Chem. 94 (42) (2022) 14745–14754.