

Article

# Siamese Networks for Clinically Relevant Bacteria Classification Based on Raman Spectroscopy

Jhonatan Contreras <sup>1,2,†</sup> , Sara Mostafapour <sup>1,†</sup> , Jürgen Popp <sup>1,2</sup> and Thomas Bocklitz <sup>1,2,3,\*</sup> 

<sup>1</sup> Institute of Physical Chemistry (IPC) and Abbe Center of Photonics (ACP), Friedrich Schiller University Jena, Member of the Leibniz Centre for Photonics in Infection Research (LPI), Helmholtzweg 4, 07743 Jena, Germany; jhonatan.contreras@uni-jena.de (J.C.); sara.mostafapourghasrodashti@uni-jena.de (S.M.); juergen.popp@uni-jena.de (J.P.)

<sup>2</sup> Leibniz Institute of Photonic Technology, Member of Leibniz Health Technologies, Member of the Leibniz Centre for Photonics in Infection Research (LPI), Albert Einstein Straße 9, 07745 Jena, Germany

<sup>3</sup> Institute of Computer Science, Faculty of Mathematics, Physics & Computer Science, University Bayreuth Universitaetsstraße 30, 95447 Bayreuth, Germany

\* Correspondence: thomas.bocklitz@uni-jena.de

† These authors contributed equally to this work.

**Abstract:** Identifying bacterial strains is essential in microbiology for various practical applications, such as disease diagnosis and quality monitoring of food and water. Classical machine learning algorithms have been utilized to identify bacteria based on their Raman spectra. However, convolutional neural networks (CNNs) offer higher classification accuracy, but they require extensive training sets and retraining of previous untrained class targets can be costly and time-consuming. Siamese networks have emerged as a promising solution. They are composed of two CNNs with the same structure and a final network that acts as a distance metric, converting the classification problem into a similarity problem. Classical machine learning approaches, shallow and deep CNNs, and two Siamese network variants were tailored and tested on Raman spectral datasets of bacteria. The methods were evaluated based on mean sensitivity, training time, prediction time, and the number of parameters. In this comparison, Siamese-model2 achieved the highest mean sensitivity of  $83.61 \pm 4.73$  and demonstrated remarkable performance in handling unbalanced and limited data scenarios, achieving a prediction accuracy of 73%. Therefore, the choice of model depends on the specific trade-off between accuracy, (prediction/training) time, and resources for the particular application. Classical machine learning models and shallow CNN models may be more suitable if time and computational resources are a concern. Siamese networks are a good choice for small datasets and CNN for extensive data.

**Keywords:** Siamese networks; machine learning; bacteria classification; Raman spectroscopy



**Citation:** Contreras, J.; Mostafapour, S.; Popp, J.; Bocklitz, T. Siamese Networks for Clinically Relevant Bacteria Classification Based on Raman Spectroscopy. *Molecules* **2024**, *29*, 1061. <https://doi.org/10.3390/molecules29051061>

Academic Editor: Xuefeng Wang

Received: 3 January 2024

Revised: 7 February 2024

Accepted: 27 February 2024

Published: 28 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Bacteria analysis including bacteria detection, identification, discrimination, and antibiotic resistance is important in food safety control, infectious disease prevention, and environmental monitoring [1,2]. Culture-based methods are the most used methods in bacteria detection and identification [3]. However, these methods are slow, labor-intensive, and do not meet requirements for the rapid detection of bacteria. Improved analytical methods such as polymerase chain reaction (PCR) [4,5] and immunological assays [6,7] have been developed to reduce overall testing time with high specificity but have limitations for field testing and require expensive reagents. Therefore, alternative methods to improve testing efficiency are needed.

Raman spectroscopy is a powerful technique for the rapid, sensitive, and non-destructive detection of bacteria. Raman spectroscopy is a powerful technique that uses scattered light to provide a chemical fingerprint of the vibrational modes within the sample's composing

molecules [8,9]. The technique has been applied in numerous fields where identifying unknown substances is crucial, including food quality control [10,11], pharmaceutical drug characterization [12], forensic investigations [13], and bacterial detection [14–16]. However, interpreting untargeted spectral data from Raman spectroscopy can be challenging due to the complexity of the data and the properties of biological samples. To address this, chemometric techniques are commonly used to analyze Raman data in the fields of chemistry, biology, and biochemistry [17,18].

Within chemometrics, classical machine learning methods have been widely used for data modeling such as unsupervised clustering, classification, and regression tasks. Dimension reduction, often the initial step in data modeling is crucial in extracting valuable features from the data. Principal component analysis (PCA) and partial least squares (PLS) are two dimension reduction techniques utilized in data modeling [19]. Selecting an appropriate number of components is critical in this area as too few components might not adequately capture the underlying trends in the data. Conversely, an excessive number of components might introduce noise, reducing the effectiveness of the model [20,21]. Following dimension reduction, the processed data can be used as input for various classification models, including linear discriminant analysis (LDA), support vector machine (SVM), and random forest (RF).

The LDA method is a widely utilized classification method that aims to maximize the ratio of between-class variance to within-class variance, thereby enhancing class separability. However, it encounters challenges with small sample sizes, where the number of samples is less than the number of variables and is limited to modeling only linearly separable data. The small sample size issue can be mitigated by applying PCA before LDA, a technique known as PCA-LDA. To address non-linearity, where classes are not linearly separable, kernel functions can be employed to extend LDA's applicability [22,23].

SVM [24] uses kernel functions to handle nonlinear classification problems by maximizing the interval between samples and decision boundary. However, this method is sensitive to missing data, which can affect its accuracy. Random forests (RFs) [25] are particularly adept at processing large datasets due to their ensemble approach, which combines multiple decision trees to improve classification accuracy and handle noise. However, the algorithm's efficiency can be compromised by the increased training time associated with a large number of decision trees, and it is possible that the model may overfit when dealing with highly noisy data [26].

A vital step in these classical machine learning methods is spectral pre-processing, which varies depending on the spectrometer and its configuration, which could produce different noise characteristics and artifacts. At the same time, the pre-processing of the signal has the potential risk of introducing further errors and variability [27–30]. Therefore, deep learning techniques in vibrational spectroscopy have been investigated in recent years. Deep neural networks have many advantages compared to classical machine learning, such as handling large datasets, both pre-process and without pre-processing and achieving superior performance [30–32]. Application of different deep learning methods like artificial neural network (ANN), convolutional neural network (CNN), auto-encoder, generative adversarial network (GAN), recurrent neural network (RNN), etc. in Raman spectroscopy data analysis can be found in previous studies [33,34].

The predecessor of deep learning is artificial neural networks (ANNs); however, among deep learning techniques, convolutional neural networks (CNNs) have been commonly used for spectral matching since they can extract fingerprint characteristics effectively, which leads to higher classification accuracy [33]. For example, a CNN is utilized to rapidly identify *Salmonella serovars* [35], distinguish between live and dead *Salmonella* [36], discrimination of clinically significant pathogens [37], discriminate between Carbapenem-resistant and Carbapenem-sensitive *Klebsiella pneumoniae* strains [38], and identify antibiotic resistance and virulence encoding factors in *Klebsiella pneumoniae* [39].

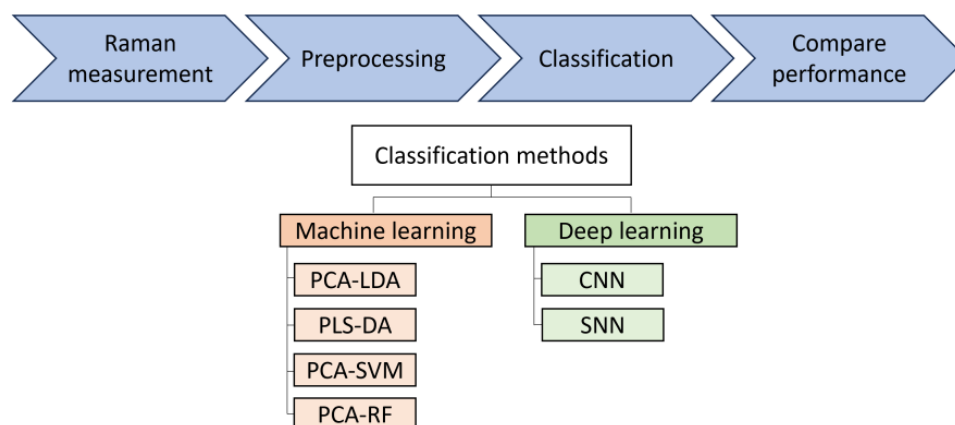
CNNs are trained in a supervised manner and optimized directly for each reference substance or class in the training database, and their performance strongly depends on the amount of training data. Additionally, retraining is needed when the reference database or the training set are modified, which induces impractical computational costs, i.e., the addition of a new class or the availability of more training data [34,40,41]. To address these limitations, a Siamese network has been proposed, which converts the classification problem into a similarity problem and solves the issue of limited available data for training CNNs [42–47], where, in many practical cases, only a few spectra are available per substance or class.

Siamese networks aim to determine similarity or dissimilarity between pairs of data points. Unlike CNN classification methods that rely on large datasets for training, Siamese networks work differently. In the training process the model is presented with pairs of examples with indications of whether they belong to the same class (similar) or different class (dissimilar), and the model learns to generate a similarity score or distance measure for each input pairs, providing a better understanding of how closely or distantly related they are. This approach differs from classification strategies that focus on assigning singular labels to individual instances. Siamese networks are more flexible and efficient when data is limited.

Our Siamese network architecture consists of two identical convolutional neural networks (CNNs) that share the same weights. The extracted features are fed into a final dense layer that calculates the similarity metric between the two spectra. This function can determine how similar or dissimilar two spectra are. In order to introduce CNNs, the model description section aims to provide a comparison between classical machine learning methods and deep learning techniques in Raman spectroscopy analysis. The comparison includes the accuracy metrics and the computational costs of these methods. The paper will also discuss potential future directions in Raman spectroscopy, mainly using Siamese networks over conventional CNNs in tasks where data availability is limited.

## 2. Results and Discussion

In this study, different classical machine learning and deep learning methods for analysis of Raman spectra from bacteria datasets are investigated. A workflow of this study is shown in Figure 1.



**Figure 1.** Bacteria Raman data analysis workflow. Different classical machine learning and deep learning method performance is investigated. PCA-LDA: principal component analysis–linear discriminant analysis; PLS-DA: partial least squares–discriminant analysis; PCA-SVM: PCA–support vector machine; RF: random forest; CNN: convolutional neural network; SNN: Siamese neural network.

### 2.1. Classification on a Six-Bacterial-Species Dataset

This Raman spectral dataset contains 5420 single bacteria spectra, which include six bacterial species and around 900 spectra per class showing a balanced dataset. The samples were cultivated in nine independent biological replicates. We use two batches of cross-validation to evaluate the stability of the results. In every cross-validation fold, the test set contains two batches, and the rest is used as a training and validation set (70% training and 30% validation). Table 1 presents the performance metrics of five different classical machine learning models: PCA-LDA, PCA-SVM, PLS-DA, and PCA-RF as classical machine learning models, and Shallow CNN, Deeper CNN, Siamese-model1, and Siamese-model2 as deep learning models. The models are evaluated based on sensitivity, specificity, training time, and the number of parameters, where sensitivity is the ability of the model to identify positive instances correctly and specificity indicates the false positive rate of the model. The number of parameters in PCA-LDA, PCA-SVM, and PCA-RF shows the number of principal components for dimensional reduction and in PLS-DA is the number of latent variables in PLS decomposition.

**Table 1.** Mean sensitivity across 36 cross-validated models on a six-bacterial-species dataset, including training time, prediction time for one sample, and the number of parameters.

	Sensitivity (%)	Specificity (%)	Training Time (s)	Prediction Time (s) One Sample		Number Parameters
PCA-LDA	79.85 ± 4.01	95.97 ± 0.80	1.79	0.0002		21
PCA-SVM	80.51 ± 4.75	96.10 ± 0.95	8.65	0.0002		21
PLS-DA	78.58 ± 3.81	95.72 ± 0.76	5.27	0.0002		21
PCA-RF	79.15 ± 4.80	95.82 ± 0.95	98.57	0.0003		21
Shallow CNN	82.80 ± 13.54	96.52 ± 0.89	800	0.040		14.7 K
Deeper CNN	84.13 ± 12.30	96.90 ± 0.83	800	0.047		19.6 M
				<i>k</i> = 10	<i>k</i> = 50	
Siamese model1	82.65 ± 4.39	96.62 ± 0.82	2000	0.070	0.105	19.6 M
Siamese model2	83.61 ± 4.73	96.75 ± 0.92	2000	0.072	0.185	19.6 M

The fitting time for the classical machine learning models cannot be compared directly with the CNN models. Although we set the maximum number of epochs for all networks to 200 with early stopping based on validation data and patience of 20 epochs, the actual number of epochs required for convergence varies depending on several factors. Random initialization, learning rate, batch normalization, and model complexity are among the factors that can influence the number of epochs needed for convergence. Therefore, we report the duration of an entire training model (200 epochs), it is essential to consider that the actual number of epochs required may vary depending on the specific model and its parameters.

The prediction time reported in Table 1 corresponds to the time needed to predict a single bacteria spectrum. Classical machine learning model prediction is straightforward and recommended when time is the most critical aspect as the prediction of a single spectrum took only around 0.0002 s. Table 1 displays high specificity values for all models, indicating correct identification of negative results. However, sensitivity is better with any of the CNN methods as the mean sensitivity of classical machine learning methods is approximately 80%.

Prediction times for the shallow and deeper CNN models are constant. However, for the Siamese networks, during testing, we averaged *N* times over *k*-shots. This means that for each test spectrum, *k* samples per class are randomly selected, this is repeated *N* times, and the average is calculated. Table 1 presents the prediction time for two values of *k*-shots for *N* = 1. Predicting 50 shots takes a significant amount of time as it involves comparing our input spectrum 300 times with 50 samples for each of the six classes, which becomes even more problematic when the number of classes is higher. However, we have

found experimentally that the values of  $k$  can range between 10 and 30, with  $N$  equal to 1. During our experiments, we observed that excluding the distance predictions (output of the Siamese networks) falling below the 10th percentile and above the 90th percentile instead of a simple mean value yields a more accurate prediction.

The number of parameters indicates the complexity of the models and should be compared as well. Classical machine learning methods have only 20 principal components, thus only 21 parameters are fitted, with 20 weights and the classification threshold, whereas all other models have millions of parameters, except for the shallow CNN, which has only 14.7 K parameters. The two highest sensitivity values are for the deeper CNN and Siamese-model2. Although the CNN sensitivity is 0.52% higher than that of the Siamese network, the standard deviation provides insight into the more general performance of the model across the 36 different cross-validated models, indicating that Siamese-model2 has a more stable behavior. In summary, the CNN-based approaches perform similarly, with  $82.80 \pm 13.54$  (shallow CNN),  $84.13 \pm 12.30$  (deeper CNN),  $82.65 \pm 4.39$  (Siamese-model1), and  $83.61 \pm 4.73$  (Siamese-model2) mean sensitivity.

## 2.2. Pre-Training and Fine-Tuning on 15 Bacterial Strains

In this section, we utilized 50% of the classes from the public Raman spectra dataset [14] to improve the visualization embeddings created by the Siamese network. Next, we tested the effectiveness of our methodology by analyzing all 30 available in a balance and imbalance scenario.

Siamese-model2 was employed due to its superior performance as determined by our evaluation metrics and considerations, and to evaluate the effectiveness of our approach. We pre-trained the Siamese network using only 50% of the available classes of the reference dataset, which allowed us to emulate a more realistic scenario.

Training first the sub-network of the Siamese network separately speeds up the process, but it is also possible to train the network end-to-end. This subnetwork was trained using the triplet semi-hard loss function, which utilizes triplets, where the negative example is farther from the anchor than the positive one, to create an embedding vector separating the classes. To identify these triplets efficiently, we used Semi-Hard online learning in each batch, allowing us to optimize the subnetwork's performance.

Figure S1a shows the learned embedding projections of three PCA components that capture only 52% of the variance in the testing data, comprising 15 classes. Each point in the plot is color-coded based on the ground truth label, and the corresponding number is also shown. We further highlight some of the clusters that are well separated from the other classes in Figure S1b,c. While it is possible to differentiate at least half of the classes based on the embeddings, the variance between the other classes is low, and the intra-class variance is high. Thus, even for similar Raman data trained on a comparable spectrum, fine-tuning is necessary to achieve optimal performance.

After fine-tuning the subnetwork using half of the classes of the fine-tuning dataset, we obtained a second embedding vector. Figure S2 displays the embedding projections of three PCA components that capture 47% of the variance in the testing data. Each point in the plot is color-coded based on the ground truth label, which may result in some colors being repeated. We observe that the majority of the classes can be distinguished from each other based on the embedding, indicating that the fine-tuning process has enhanced the subnetwork. Figure 2 displays the confusion matrix derived from the testing dataset, where the model was tested with ten shots. Analysis of the confusion matrix reveals an average sensitivity of 72.0%, an average precision of 76.1%, and an F1 score of 0.71. The F1 score integrates precision and sensitivity into a single metric to gain a better understanding of model performance. The F1 score is calculated by:

$$\text{F1 score} = \frac{2 \times \text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

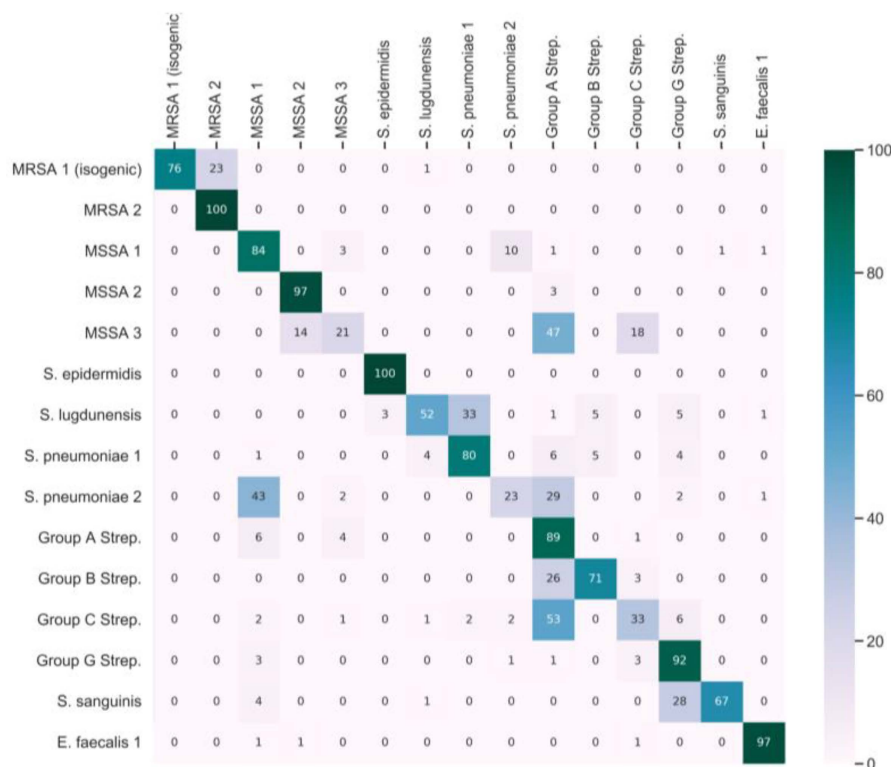


Figure 2. Confusion matrix for the Siamese network. Testing data across the first 15 bacteria strains of a 30-bacterial-strain dataset.

### 2.3. Classification on a 30-Bacterial-Strain Dataset

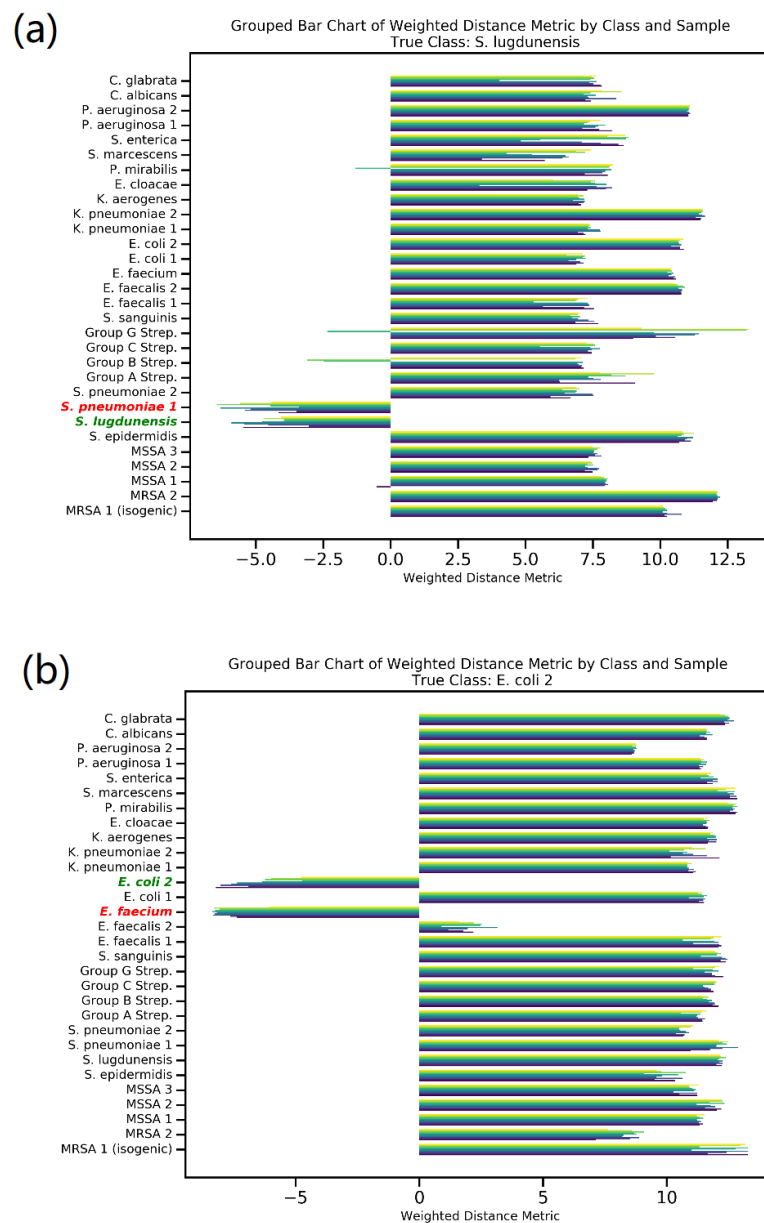
In this section, Siamese model2 is pre-trained using the reference data dataset [14], and subsequently we examine two distinct fine-tuning scenarios. The initial scenario entails pre-training on 30 classes, followed by training the Siamese network with a balanced dataset, consisting of 100 spectra for each class. In contrast, the second scenario adopts an unbalanced dataset approach, where half of the classes are represented by 100 spectra each, and the remaining half by only 30 spectra each. Moreover, the pre-training process incorporates only 50% of the classes, a strategy specifically designed to assess the model’s proficiency in managing scenarios characterized by limited data availability.

Figure 3 showcases the confusion matrices for the two scenarios under investigation. This detailed representation highlights the classification outcomes when training the model under balanced versus unbalanced conditions. For the testing phase, ten shots were employed.

In the balanced scenario, utilizing the full dataset yielded a notable average sensitivity of 80.0%, an average precision of 82.0%, and an F1 score of 80.0%. For the unbalanced scenario, the model achieved an average sensitivity of 73.0%, an average precision of 78.1%, and an F1 score of 72.9%. These results are particularly significant considering the challenging conditions: the absence of certain classes during pre-training and the limited sample size in the fine-tuning stage. Despite these constraints, the Siamese model attained a commendable prediction accuracy of 73.0%. The Siamese model shows potential in complex scenarios, indicating an ability to address new data classes and manage limited data, which suggests its adaptability and robustness for practical applications.



distance metric between the input spectra; specifically, (a) *S. lugdunensis* and (b) *E. coli*2, in relation to all reference bacteria. In this chart, a greater distance signifies reduced similarity between the spectra, whereas a smaller or negative distance denotes higher similarity. Figure 4 uses green to denote the true class and red to indicate the incorrectly predicted class. The model focuses exclusively on these two specific classes in the scenarios presented, effectively disregarding the other 28. It is important to note that this model's primary function is not to classify data in the conventional sense but rather to compare input data against a reference spectrum, as reflected in its selective identification process. Furthermore, Figure S4 graphically presents the distribution of the weighted distance metric, contrasting correct versus incorrect predictions. This visualization clearly demonstrates that samples with incorrect predictions tend to cluster at higher distances, while those correctly classified generally align with lower distance values.



**Figure 4.** Comparison of misclassified spectra (a) *S. lugdunensis* and (b) *E. coli* 2. Using weighted distance metrics and grouped bar chart illustration of similarity measures in the 30-class Siamese model2.

In the previous section, we described a model trained on a balanced dataset that achieved a notable classification accuracy of 80%. However, given the complex and diverse



nature of bacteria strains, we incorporated a refined metric called “Rank-2 accuracy”. It considers the accuracy of the second-highest prediction along with the top choice. This metric is relevant when using a Siamese network, which solves a similarity problem, especially in scenarios with 30 different bacterial classes, because it better evaluates the model’s performance. As summarized in Table 2 which also includes Rank-3 accuracy, Rank-1 accuracy stands at 80.26% with 2408 correct classifications, Rank-2 accuracy enhances the model’s precision to 90.26% for an additional 300 spectra, and Rank-3 accuracy further elevates this metric to 93.46% for 96 spectra.

**Table 2.** Hierarchical accuracy metrics, Siamese model2 performance across Ranks 1–3. Shows accuracy percentages and counts of classified instances. Testing data across the 30 bacteria strains.

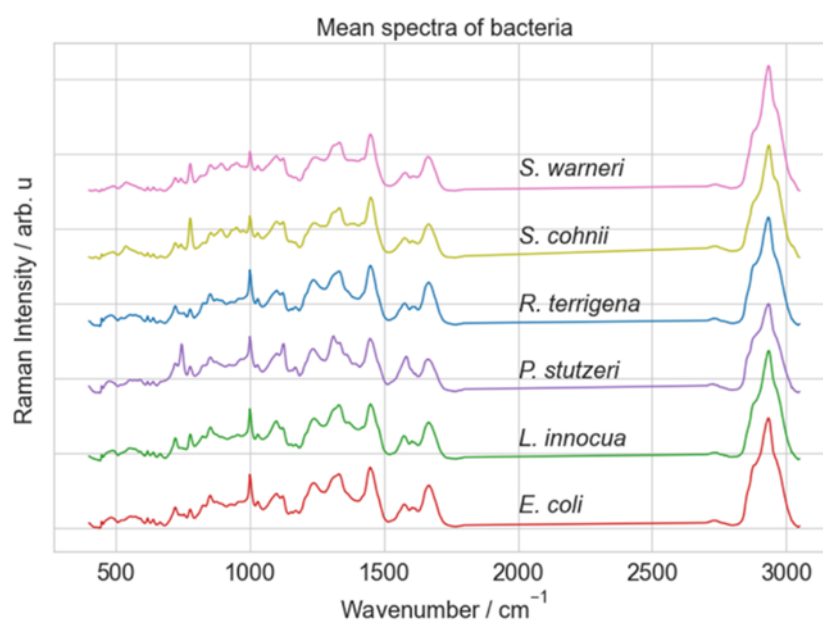
	Overall Accuracy	Count
Rank-1	80.26	2408
Rank-2	90.26	300
Rank-3	93.46	96

### 3. Materials and Methods

#### 3.1. Data Description

The dataset utilized in this study comprised Raman spectra obtained from six distinct bacterial species: *Escherichia coli* DSM 423, *Klebsiella terrigena* DSM 2687, *Pseudomonas stutzeri* DSM 5190, *Listeria innocua* DSM 20649, *Staphylococcus warneri* DSM 20316, and *Staphylococcus cohnii* DSM 20261. All bacterial species were obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) and were cultivated in nine independent biological replicates.

The Raman spectra were acquired using a Raman microscope (Bio Particle Explorer, rap.ID Particle Systems GmbH). The spectral data were collected over a range of 240 to 3190  $\text{cm}^{-1}$ . The dataset contains 5420 preprocessed spectra with 584 wavenumbers. The number of spectra per class is around 900 spectra showing a balanced dataset. Figure 5 shows the mean spectra for each class. More details about samples and the Raman spectrometer can be found in ref. [48].

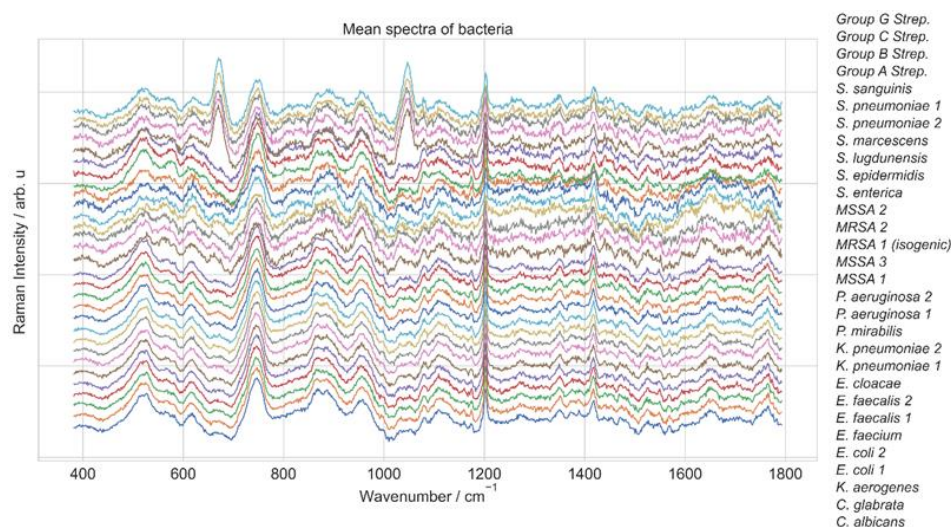


**Figure 5.** Mean spectra of Raman spectra dataset [48] obtained from six distinct bacterial species.

Prior to analysis, the Raman spectra underwent preprocessing steps, cosmic spikes removal, wavenumber calibration, spectra alignment, and baseline correction. The preprocessed spectra were then used as inputs for the classification algorithms. The efficacy of various classification algorithms was evaluated utilizing cross-validated weighted accuracy.

We analyzed a second publicly available Raman spectra dataset [14], including 30 common bacterial pathogens. Table S1 describes the species name, label, and isolate code. These pathogens were chosen as they are responsible for most infections observed in intensive care units worldwide. The researchers deposited bacterial cells onto gold-coated silica substrates to create the reference database. They collected spectra from monolayer regions of each strain, ensuring high-quality Raman spectra with minimal interference.

The dataset comprises three subsets: a reference dataset, a reference-finetune dataset, and a test dataset. The reference dataset contains 60,000 evenly distributed Raman spectra from 30 bacterial strains, with 2000 spectra per class. These spectra were used to pre-train machine learning models [14,45]. The reference-finetune dataset was employed to fine-tune the models, while the test dataset was used to evaluate the performance of the models. The reference-finetune dataset and a test dataset contain 3000 spectra representing all 30 bacterial strains with 100 spectra per class. Figure 6 shows the mean spectra for each of the 30 bacterial strains. The spectral data were collected over 381.98 and 1792.4  $\text{cm}^{-1}$ , distributed uniformly over 1000 wavenumbers. As preprocessing, the spectra intensity was individually normalized between 0 and 1.



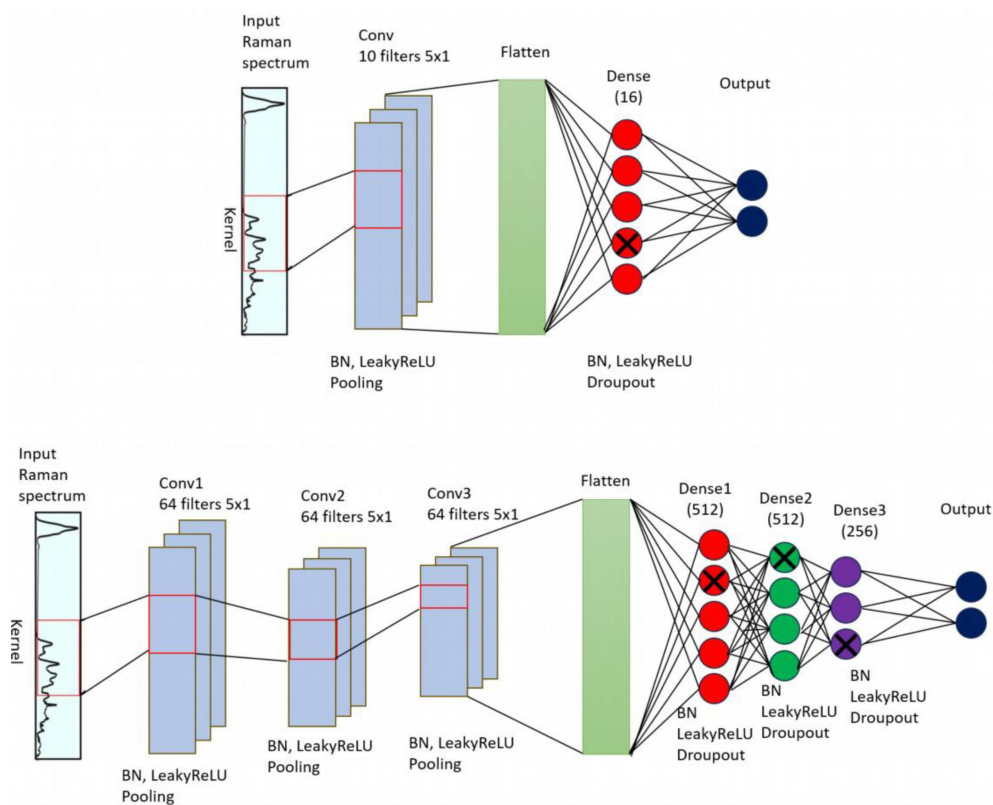
**Figure 6.** Mean spectra of Raman spectra dataset [14] obtained from 30 distinct bacterial species.

### 3.2. One-Dimensional Convolutional Neural Network Model Description

One-dimensional convolutional neural networks (1D-CNNs) have become a popular tool for processing and analysis of sequential data in various domains, such as speech recognition [49], music analysis [50], and financial forecasting [51], for their ability to extract relevant features from sequential data. In a 1D-CNN, the input data is represented as a sequence of values.

Figure 7 illustrates the network architecture for the two 1D-CNN models that are used in this study. In the case of Raman data, the input contains the entire spectrum, where each point represents an intensity in a specific wavenumber. The network then applies a set of filters (kernels) to the input data, with each filter moving across a small window of adjacent data points at a time. The filters perform convolutions, which extract important features or patterns. These features are then passed through additional convolutional layers to create increasingly complex representations of the original input sequence. Additionally, a non-linear activation function (Leaky ReLU) introduces nonlinearity to the model after each layer. Finally, the output of the last convolutional layer is fed through a set of fully connected layers, which perform the final classification followed by SoftMax activation that

normalizes the  $k$ -dimensional output vector of real values with different dynamic ranges to real values in the range  $[0,1]$ , that sum to 1 and can be viewed as a probability output.



**Figure 7.** CNN architecture for Raman spectrum classification. It is composed of a set of convolutional layers that extract features, followed by fully connected layers that perform the classification. Above: shallow CNN with one convolutional and one dense layer, below: deeper CNN with three convolutional and three dense layers.

Figure 7 presents the general concept of a 1D CNN. However, the selection of the optimal values of the hyperparameters, such as the number of layers, kernels, kernel size, dropout value, activation function, batch normalization decay, and momentum, depends on the dataset. Typically, those values are selected through trial and error. We report the results for two configurations, a shallow CNN model (Figure 5 above) composed of a convolutional layer and a dense layer, and a deeper CNN formed by three convolutional layers and three dense layers (Figure 5 below). A more detailed review of these CNN architectures which were built based on Tensorflow framework in Python is shown in Table 3. The shallow CNN model includes a convolutional, a pooling, and a dense layer. The Raman spectrum as input imported into the 1D-CNN model. The convolutional kernel with size  $10 \times 1$  and step size 5 are used for feature extraction. This procedure is followed by a batch normalization (BN) layer. The Leaky ReLU function is used as the activation function, which can help the network learn complex data, improve nonlinear modeling capabilities of the network, and provide more accurate predictions. The pooling layer reduces the dimensionality of the feature vector, enhances the robustness of the network, and obtains lower resolution feature data. The data from pooling layer are input to a fully connected (Dense) layer for classification. The difference between shallow CNN and deeper CNN models is in the number of convolutional, pooling, dense layers, and kernel size.

**Table 3.** Comparison of the architectures of the deep learning models used within this article.

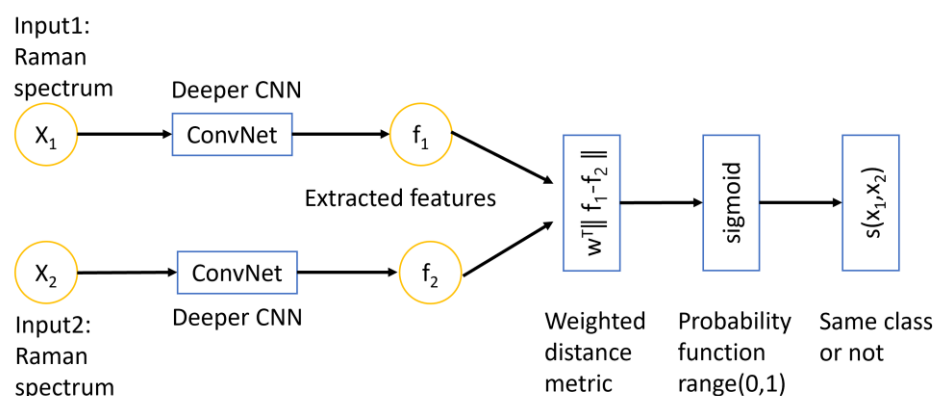
	Convolutional Layers	Fully Connected Layers
Shallow CNN	Conv (10,5) BatchNorm + LeakyReLU + pooling	Dense (16) BatchNorm + LeakyReLU + Dropout
Deeper CNN	3 × Conv (64,5) BatchNorm + LeakyReLU + pooling	2 × Dense (512) + Dense (256) BatchNorm + LeakyReLU + Dropout
	Feature vector embedding	Learnable distance metric
Siamese model1	Deeper CNN	Dense (1) Sigmoid
Siamese model2	Deeper CNN	Dense (64) + Dense (16) + Dense (1) Sigmoid

### 3.3. One-Dimensional Siamese Network Model Description

The Siamese network is a type of deep learning architecture specifically designed for solving problems related to similarity and distance. Figure 8 shows that the architecture of this network consists of two identical sub-networks with the same weights and architecture. Each sub-network takes in a Raman spectrum as input and produces feature vector embedding that captures the essential information in a higher dimensional space. The two feature vectors  $f_1$  and  $f_2$  are then compared using a learnable weighted distance metric:

$$d = w^T \cdot \|f_1 - f_2\|$$

where the weights of the metric are learned from data during the training process. In our network, it is implemented using a fully connected layer followed by a sigmoid activation. In other words, the metric is not fixed but can adapt to the specific characteristics of the data being analyzed. It measures the distance or similarity between two spectra, the output of the Siamese network is a binary classification, which reveals whether the two input spectra belong to the same class or not.



**Figure 8.** The conceptual architecture of a Siamese network. The inputs are two Raman spectra that analyzed by two identical CNN models. The convolution network is the deeper CNN model that was introduced in previous section. The extracted features are compared using a learnable weighted distance metric and followed by a sigmoid function which predicts the probability between the range of 0 and 1. The output determines whether the two inputs belong to the same class or not (the two inputs are similar or not).

### 3.4. One-Dimensional Siamese Network Model Training

Since the dataset size is small, we decided not to create a paired spectra dataset and instead generated them randomly during training, which allowed us to reduce overfitting. Given spectrum  $x_i$ , another reference spectrum  $x_j$  is randomly selected from the training

data set with its corresponding class label. The paired label  $y_i$  is generated as follows: if the input spectra belong to the same category,  $y_i$  is set to zero, indicating poor distance and similarity. On the other hand, if the input spectra belong to different classes,  $y_i$  is set to one, indicating significant distance and negative similarity. In this way, the problem is transformed into a binary classification in which we use the binary cross entropy as the loss function.

$$L = -\frac{1}{N} \sum_{i=1}^N y_i \log p_i + (1 - y_i) \log(1 - p_i)$$

where  $N$  is the number of paired spectra,  $p_i$  is the outcome of the Siamese network and  $y_i$  is the paired label. The models were trained with the Adam optimizer with a learning rate of  $6 \times 10^{-5}$  and batch size of 64. We report the results for two configurations. For both cases, the feature extraction part is identical. The difference lies in the learnable distance function. In the first case, we use only one dense layer, and in the second configuration, we use a more complex function that uses three dense layers, as shown in Table 3.

#### 4. Conclusions

As discussed earlier, to utilize Raman spectroscopy in real world applications like diagnostics, chemometrics and machine learning are needed. In this contribution different classical machine learning concepts are compared, which feature different properties. Table 1 presents the results of several classical machine learning models trained and evaluated on a six-bacterial-species dataset, including PCA-LDA, PCA-SVM, PLS-DA, PCA-RF, shallow CNN, deeper CNN, and two variants of the Siamese model. The models were compared based on mean sensitivity, training time, prediction time, and number of parameters. Based on the application scenario, the restrictions of the application and the available data different models perform best.

If the goal is to maximize model performance, the best choice would be the Siamese model 2, which achieved the highest mean sensitivity of  $83.61 \pm 4.73$ . However, if time and computational resources are a concern, the classical machine learning and shallow CNN models would be more suitable, as they required significantly less training and prediction time and contained fewer parameters than the deeper CNN models. Additionally, if the dataset is small, the shallow CNN model may be a better option, as it achieved reasonable accuracy while requiring less training time and having fewer parameters than the deeper CNN and Siamese models. Overall, the choice of model should depend on the specific trade-off between model performance, training/prediction time, and resources that are most important for a particular application [52].

In addition, Siamese-model2 was utilized to classify 30 bacterial strains. The sub-network was pre-trained separately using the triplet semi-hard loss function to generate an efficient embedding vector to separate the classes. We learned the learnable weighted distance metric using the fine-tuning dataset, which improved the model's ability to differentiate between bacterial strains converting the classification problem into a similarity problem. To better emulate real-world conditions, only half of the available classes from the reference dataset were utilized to pre-train the model. We fine-tuned Siamese-model2 using a limited and highly unbalanced dataset. This presented a challenging task, as 15 classes had 100 samples each, while the remaining 15 classes only had 30 samples each. Nevertheless, such situations are often encountered in medical diagnostics. Despite these difficulties, Siamese-model2 demonstrated remarkable performance, achieving a prediction accuracy of 73%. Our findings highlight the model's ability to generalize to new classes and handle limited data scenarios effectively. These properties make Siamese-model2 a versatile and robust tool, with immense practical applications. Furthermore, using all available data, we obtained a significantly higher accuracy of 80.04% for 30 classes, further underscoring the potential of Siamese-model2 in bacterial strain identification. With further improvements and optimizations, this approach could lead to more accurate and efficient bacterial classification, aiding in the field of microbiology and disease diagnosis.

**Supplementary Materials:** The supplementary material can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules29051061/s1>. Figure S1. Pre-trained Embedding projections of three PCA components that capture only 52% of the variance in the testing data. (a) 15 classes clusters, (b,c) some of the clusters that are well separated from the other classes. Figure S2. Fine-tuned Embedding projections of three PCA components that capture only 47% of the variance in the testing data. (a) 15 classes clusters, (b,c) some of the clusters that are well separated from the other classes. Figure S3. Mean spectra and standard deviation from training data for three bacterial strains. Figure S4. Distribution of Weighted Distance Metric for Correct vs Incorrect Predictions. Testing Data Across the 30 Bacteria Strains. Table S1. The species name, figure label, isolate code, and empiric antibiotic treatment. Data sourced from “Rapid identification of pathogenic bacteria using spectroscopy and deep learning”, HO, Chi-Sing, et al. Nature communications, 2019 [14]. Table S2. Confusion matrix of different classification methods for six bacteria species dataset.

**Author Contributions:** Conceptualization, T.B.; code generation, S.M. and J.C.; discussion, J.P., T.B., S.M. and J.C.; draft preparation, S.M. and J.C.; writing—reviewing, S.M., J.C., J.P. and T.B.; funding acquisition J.P. and T.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is part of the Collaborative Research Centre AquaDiva of the Friedrich Schiller University Jena, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB 1076—Project Number 218627073. This work is additionally supported by the BMBF, funding program Photonics Research Germany (FKZ: 13N15710 (LPI-BT3), FKZ: 13N15466 (LPI-BT1)) and is integrated into the Leibniz Center for Photonics in Infection Research (LPI). The LPI initiated by Leibniz-IPHT, Leibniz-HKI, Friedrich Schiller University Jena and Jena University Hospital is part of the BMBF national roadmap for research infrastructures.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in this study is public and available in the cited references.

**Acknowledgments:** We would like to thank the authors of reference 14 and 48 for their work.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Váradi, L.; Luo, J.L.; Hibbs, D.E.; Perry, J.D.; Anderson, R.J.; Orenga, S.; Groundwater, P.W. Methods for the detection and identification of pathogenic bacteria: Past, present, and future. *Chem. Soc. Rev.* **2017**, *46*, 4818–4832. [CrossRef] [PubMed]
2. Law, J.W.-F.; Ab Mutalib, N.-S.; Chan, K.-G.; Lee, L.-H. Rapid methods for the detection of foodborne bacterial pathogens: Principles, applications, advantages and limitations. *Front. Microbiol.* **2015**, *5*, 770. [CrossRef]
3. Gracias, K.S.; McKillip, J.L. A review of conventional detection and enumeration methods for pathogenic bacteria in food. *Can. J. Microbiol.* **2004**, *50*, 883–890. [CrossRef]
4. Vinayaka, A.C.; Ngo, T.A.; Kant, K.; Engelsmann, P.; Dave, V.P.; Shahbazi, M.-A.; Wolff, A.; Bang, D.D. Rapid detection of *Salmonella enterica* in food samples by a novel approach with combination of sample concentration and direct PCR. *Biosens. Bioelectron.* **2019**, *129*, 224–230. [CrossRef] [PubMed]
5. Jian, C.; Luukkonen, P.; Yki-Järvinen, H.; Salonen, A.; Korpela, K. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS ONE* **2020**, *15*, e0227285. [CrossRef]
6. Seo, S.-H.; Lee, Y.-R.; Jeon, J.H.; Hwang, Y.-R.; Park, P.-G.; Ahn, D.-R.; Han, K.-C.; Rhie, G.-E.; Hong, K.-J. Highly sensitive detection of a bio-threat pathogen by gold nanoparticle-based oligonucleotide-linked immunosorbent assay. *Biosens. Bioelectron.* **2015**, *64*, 69–73. [CrossRef]
7. Wu, W.; Li, J.; Pan, D.; Li, J.; Song, S.; Rong, M.; Li, Z.; Gao, J.; Lu, J. Gold nanoparticle-based enzyme-linked antibody-aptamer sandwich assay for detection of *Salmonella Typhimurium*. *ACS Appl. Mater. Interfaces* **2014**, *6*, 16974–16981. [CrossRef]
8. Raman, C.V.; Krishnan, K.S. A new type of secondary radiation. *Nature* **1928**, *121*, 501–502. [CrossRef]
9. Popp, J.; Tuchin, V.V.; Chiou, A.; Heinemann, S.H. *Handbook of Biophotonics, Volume 3: Photonics in Pharmaceuticals, Bioanalysis and Environmental Research*; Wiely-VCH Verlag & Co. KGaA: Weinheim, Germany, 2012; Volume 3.
10. Amjad, A.; Ullah, R.; Khan, S.; Bilal, M.; Khan, A. Raman spectroscopy based analysis of milk using random forest classification. *Vib. Spectrosc.* **2018**, *99*, 124–129. [CrossRef]
11. Sun, Y.; Tang, H.; Zou, X.; Meng, G.; Wu, N. Raman spectroscopy for food quality assurance and safety monitoring: A review. *Curr. Opin. Food Sci.* **2022**, *47*, 100910. [CrossRef]
12. Lussier, F.; Thibault, V.; Charron, B.; Wallace, G.Q.; Masson, J.-F. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *TrAC Trends Anal. Chem.* **2020**, *124*, 115796. [CrossRef]

13. Huang, T.-Y.; Yu, J.C.C. Development of crime scene intelligence using a hand-held Raman spectrometer and transfer learning. *Anal. Chem.* **2021**, *93*, 8889–8896. [[CrossRef](#)] [[PubMed](#)]
14. Ho, C.-S.; Jean, N.; Hogan, C.A.; Blackmon, L.; Jeffrey, S.S.; Holodniy, M.; Banaei, N.; Saleh, A.A.; Ermon, S.; Dionne, J. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* **2019**, *10*, 4927. [[CrossRef](#)] [[PubMed](#)]
15. Kukula, K.; Farmer, D.; Duran, J.; Majid, N.; Chatterley, C.; Jessing, J.; Li, Y. Rapid detection of bacteria using raman spectroscopy and deep learning. In Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 27–30 January 2021; pp. 796–799.
16. Liu, B.; Liu, K.; Wang, N.; Ta, K.; Liang, P.; Yin, H.; Li, B. Laser tweezers Raman spectroscopy combined with deep learning to classify marine bacteria. *Talanta* **2022**, *244*, 123383. [[CrossRef](#)] [[PubMed](#)]
17. Rodriguez, L.; Zhang, Z.; Wang, D. Recent advances of Raman spectroscopy for the analysis of bacteria. *Anal. Sci. Adv.* **2023**, *4*, 81–95. [[CrossRef](#)]
18. Mukherjee, A.; Su, A.; Rajan, K. Deep learning model for identifying critical structural motifs in potential endocrine disruptors. *J. Chem. Inf. Model.* **2021**, *61*, 2187–2197. [[CrossRef](#)]
19. Guo, S.; Rösch, P.; Popp, J.; Bocklitz, T. Modified PCA and PLS: Towards a better classification in Raman spectroscopy-based biological applications. *J. Chemom.* **2020**, *34*, e3202. [[CrossRef](#)]
20. Gracia, A.; González, S.; Robles, V.; Menasalvas, E. A methodology to compare dimensionality reduction algorithms in terms of loss of quality. *Inf. Sci.* **2014**, *270*, 1–27. [[CrossRef](#)]
21. Salem, N.; Hussein, S. Data dimensional reduction and principal components analysis. *Procedia Comput. Sci.* **2019**, *163*, 292–299. [[CrossRef](#)]
22. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear discriminant analysis: A detailed tutorial. *AI Commun.* **2017**, *30*, 169–190. [[CrossRef](#)]
23. Lasalvia, M.; Capozzi, V.; Perna, G. A comparison of PCA-LDA and PLS-DA techniques for classification of vibrational spectra. *Appl. Sci.* **2022**, *12*, 5345. [[CrossRef](#)]
24. Tewes, T.J.; Kerst, M.; Platte, F.; Bockmühl, D.P. Raman microscopic identification of microorganisms on metal surfaces via support vector machines. *Microorganisms* **2022**, *10*, 556. [[CrossRef](#)] [[PubMed](#)]
25. Seifert, S. Application of random forest based approaches to surface-enhanced Raman scattering data. *Sci. Rep.* **2020**, *10*, 5436. [[CrossRef](#)]
26. Jiang, Y.; Luo, J.; Huang, D.; Liu, Y.; Li, D.-D. Machine learning advances in microbiology: A review of methods and applications. *Front. Microbiol.* **2022**, *13*, 925454. [[CrossRef](#)]
27. Bocklitz, T.; Putsche, M.; Stüber, C.; Käs, J.; Niendorf, A.; Rösch, P.; Popp, J. A comprehensive study of classification methods for medical diagnosis. *J. Raman Spectrosc.* **2009**, *40*, 1759–1765. [[CrossRef](#)]
28. Bocklitz, T.; Walter, A.; Hartmann, K.; Rösch, P.; Popp, J. How to pre-process Raman spectra for reliable and stable models? *Anal. Chim. Acta* **2011**, *704*, 47–56. [[CrossRef](#)]
29. Ryabchykov, O.; Guo, S.; Bocklitz, T. Analyzing Raman spectroscopic data. *Phys. Sci. Rev.* **2018**, *4*, 20170043.
30. Guo, S.; Popp, J.; Bocklitz, T. Chemometric analysis in Raman spectroscopy from experimental design to machine learning-based modeling. *Nat. Protoc.* **2021**, *16*, 5426–5459. [[CrossRef](#)]
31. Ma, D.; Shang, L.; Tang, J.; Bao, Y.; Fu, J.; Yin, J. Classifying breast cancer tissue by Raman spectroscopy with one-dimensional convolutional neural network. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *256*, 119732. [[CrossRef](#)] [[PubMed](#)]
32. Liu, J.; Osadchy, M.; Ashton, L.; Foster, M.; Solomon, C.J.; Gibson, S.J. Deep convolutional neural networks for Raman spectrum recognition: A unified solution. *Analyst* **2017**, *142*, 4067–4074. [[CrossRef](#)]
33. Pradhan, P.; Guo, S.; Ryabchykov, O.; Popp, J.; Bocklitz, T.W. Deep learning a boon for biophotonics? *J. Biophotonics* **2020**, *13*, e201960186. [[CrossRef](#)]
34. Luo, R.; Popp, J.; Bocklitz, T. Deep learning for Raman spectroscopy: A review. *Analytica* **2022**, *3*, 287–301. [[CrossRef](#)]
35. Sun, J.; Xu, X.; Feng, S.; Zhang, H.; Xu, L.; Jiang, H.; Sun, B.; Meng, Y.; Chen, W. Rapid identification of salmonella serovars by using Raman spectroscopy and machine learning algorithm. *Talanta* **2023**, *253*, 123807. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, J.; Zhang, J.; Ding, J.; Lin, Q.; Young, G.M.; Jiang, C. Rapid identification of live and dead Salmonella by surface-enhanced Raman spectroscopy combined with convolutional neural network. *Vib. Spectrosc.* **2022**, *118*, 103332. [[CrossRef](#)]
37. Tang, J.-W.; Li, J.-Q.; Yin, X.-C.; Xu, W.-W.; Pan, Y.-C.; Liu, Q.-H.; Gu, B.; Zhang, X.; Wang, L. Rapid discrimination of clinically important pathogens through machine learning analysis of surface enhanced Raman spectra. *Front. Microbiol.* **2022**, *13*, 843417. [[CrossRef](#)]
38. Liu, W.; Tang, J.-W.; Lyu, J.-W.; Wang, J.-J.; Pan, Y.-C.; Shi, X.-Y.; Liu, Q.-H.; Zhang, X.; Gu, B.; Wang, L. Discrimination between carbapenem-resistant and carbapenem-sensitive *Klebsiella pneumoniae* strains through computational analysis of surface-enhanced Raman spectra: A pilot study. *Microbiol. Spectr.* **2022**, *10*, e02409-21. [[CrossRef](#)] [[PubMed](#)]
39. Lu, J.; Chen, J.; Liu, C.; Zeng, Y.; Sun, Q.; Li, J.; Shen, Z.; Chen, S.; Zhang, R. Identification of antibiotic resistance and virulence-encoding factors in *Klebsiella pneumoniae* by Raman spectroscopy and deep learning. *Microb. Biotechnol.* **2022**, *15*, 1270–1280. [[CrossRef](#)]
40. Kazemzadeh, M.; Martinez-Calderon, M.; Xu, W.; Chamley, L.W.; Hisey, C.L.; Broderick, N.G. Cascaded deep convolutional neural networks as improved methods of preprocessing raman spectroscopy data. *Anal. Chem.* **2022**, *94*, 12907–12918. [[CrossRef](#)]

41. Wu, M.; Wang, S.; Pan, S.; Terentis, A.C.; Strasswimmer, J.; Zhu, X. Deep learning data augmentation for Raman spectroscopy cancer tissue classification. *Sci. Rep.* **2021**, *11*, 23842. [[CrossRef](#)]
42. Dong, X.; Shen, J. Triplet loss in siamese network for object tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 459–474.
43. Melekhov, I.; Kannala, J.; Rahtu, E. Siamese network features for image matching. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 378–383.
44. Park, J.-H.; Yu, H.-G.; Park, D.-J.; Nam, H.; Chang, D.E. Dynamic one-shot target detection and classification using a pseudo-Siamese network and its application to Raman spectroscopy. *Analyst* **2021**, *146*, 6997–7004. [[CrossRef](#)]
45. Li, B.; Schmidt, M.N.; Alstrøm, T.S. Raman spectrum matching with contrastive representation learning. *Analyst* **2022**, *147*, 2238–2246. [[CrossRef](#)]
46. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Proceedings of the Computer Vision—ECCV 2016 Workshops, Amsterdam, The Netherlands, 8–10 and 15–16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 850–865.
47. Tian, X.; Wang, P.; Tian, Y.; Zhang, R.; Jiang, Z.; Gao, J. Classification method based on Siamese-like neural network for inter-species blood Raman spectra similarity measure. *J. Biophotonics* **2023**, *16*, e202200377. [[CrossRef](#)] [[PubMed](#)]
48. Ali, N.; Girmus, S.; Rösch, P.; Popp, J.; Bocklitz, T. Sample-size planning for multivariate data: A Raman-spectroscopy-based example. *Anal. Chem.* **2018**, *90*, 12485–12492. [[CrossRef](#)] [[PubMed](#)]
49. Huang, J.-T.; Li, J.; Gong, Y. An analysis of convolutional neural networks for speech recognition. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, Australia, 19–24 April 2015; pp. 4989–4993.
50. Costa, Y.M.; Oliveira, L.S.; Silla, C.N., Jr. An evaluation of convolutional neural networks for music classification using spectrograms. *Appl. Soft Comput.* **2017**, *52*, 28–38. [[CrossRef](#)]
51. Cao, J.; Wang, J. Stock price forecasting model based on modified convolution neural network and financial time series analysis. *Int. J. Commun. Syst.* **2019**, *32*, e3987. [[CrossRef](#)]
52. Althnian, A.; AlSaeed, D.; Al-Baity, H.; Samha, A.; Dris, A.B.; Alzakari, N.; Abou Elwafa, A.; Kurdi, H. Impact of dataset size on classification performance: An empirical evaluation in the medical domain. *Appl. Sci.* **2021**, *11*, 796. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.