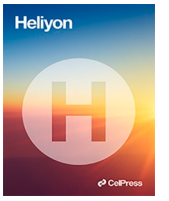
Heliyon 10 (2024) e27824

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Unveiling the efficacy of a bulk Raman spectra-based model in predicting single cell Raman spectra of microorganisms

Thomas J. Tewes [a], Mario Kerst [a], Svyatoslav Pavlov [a], Miriam A. Huth [a], Ute Hansen [b], Dirk P. Bockmühl [a,*]

[a] Faculty of Life Sciences, Rhine-Waal University of Applied Sciences, Marie-Curie-Straße 1, 47533, Kleve, Germany
[b] Faculty of Communication and Environment, Rhine-Waal University of Applied Sciences, Friedrich-Heinrich-Allee, 47475, Kamp-Lintfort, Germany

ARTICLE INFO

Keywords:
Raman
Single cell
Bulk
Microorganisms
Convolutional neural network
CNN

ABSTRACT

In a previous publication, we trained predictive models based on Raman bulk spectra of microorganisms placed on a silicon dioxide protected silver mirror slide to make predictions for new Raman spectra, unknown to the models, of microorganisms placed on a different substrate, namely stainless steel. Now we have combined large sections of this data and trained a convolutional neural network (CNN) to make predictions for single cell Raman spectra. We show that a database based on microbial bulk material is conditionally suited to make predictions for the same species in terms of single cells. Data of 13 different microorganisms (bacteria and yeasts) were used. Two of the 13 species could be identified 90% correctly and five other species 71%–88%. The six remaining species were correctly predicted by only 0%–49%. Especially stronger fluorescence in bulk material compared to single cells but also photodegradation of carotenoids are some effects that can complicate predictions for single cells based on bulk data. The results could be helpful in assessing universal Raman tools or databases.

1. Introduction

Raman spectroscopy is a vibrational spectroscopic technology based on the inelastic scattering of light. Emitted photons interact with a given molecule, resulting in a change of energy and a wavelength of the backscattered light being associated with molecular vibration [1,2]. Raman spectra contain information of chemical components in specimens and permit the characterization of various samples, whether solid, liquid or gaseous [3]. Differences in the biochemical composition of microorganisms lead to different Raman spectra, which makes the differentiation feasible [4,5]. Since Raman spectra of microorganisms are often very similar, especially in closely related species, and there are external environmental influences on spectra, whether sample-, instrument- or e.g., fluorescence-related, mathematical pre-treatment methods such as baseline correction, smoothing and normalization, are often necessary [6] to extract the maximum information from a spectrum or a large data set for the corresponding prediction to be made. Reducing dimensions via principal component analysis (PCA) and the associated elimination of features that are unnecessary or disturbing for accurate predictions comprises another data pretreatment [7]. In many cases, differentiation is not enabled by clear species-specific signatures, but by subtle nuances within large datasets [8]. These can be exploited by various classification methods such as linear discriminant analysis (LDA) [9,10], support vector machines (SVMs) [11,12] or convolutional neural networks (CNNs)

* Corresponding author.
E-mail address: dirk.bockmuehl@hsrw.eu (D.P. Bockmühl).

https://doi.org/10.1016/j.heliyon.2024.e27824
Received 16 October 2023; Received in revised form 29 February 2024; Accepted 7 March 2024
Available online 9 March 2024

[13,14].

Numerous studies have shown that robust predictive models for microorganisms based on Raman microspectroscopy (RMS) in combination with chemometric methods are feasible, and often very accurate [4,15–17]. One major advantage of RMS is that complex signals from individual bacterial cells can be detected [18,19]. Single bacterial cells can be differentiated using surface-enhanced Raman microscopy (SERS) or specific metal substrates [20–22], avoiding time-consuming pre-enrichment and cultivation. This is particularly interesting for medical diagnostics [23], where examination results must be available quickly in order to treat patients correctly. For example, Wang et al. recently demonstrated an interesting SERS approach in combination with different machine learning methods for the rapid identification of *Mycobacterium tuberculosis* in sputum samples [24]. In general, there have been major recent advances in clinical SERS applications for the rapid and accurate detection of bacterial pathogens [25–27], whereby the studies confirm the great potential of integrating SERS techniques and machine learning algorithms for the rapid and reliable identification of various bacterial pathogens, but the authors always point out the still existing wide gap between basic research and the clinical application of SERS technology. Likewise, in the food industry [28–30], time is crucial for product safety and a long shelf life. It might be important to mention that there are also culture-related influences on a microbial Raman spectrum, e.g. depending on the nutrition media [31,32] or the incubation time [33]. Therefore, usually standardized culture methods are used for specific applications, or the variation occurring in the application field (including a possible environmental matrix) is considered within the data of the predictive models. Likewise, the substrate itself also matters in RMS [34], being particularly significant for single cell measurements. Basically, the classification of microorganisms by Raman spectroscopy always involves the development of a database containing Raman spectra of the species to be identified. Already at this stage, a choice of the wavenumber region, the microscopic lens, excitation wavelength, spectral resolution and measurement settings such as exposure time, laser power and accumulations must be made, based on the possible options, since all these parameters can also affect Raman spectra.

After data collection, data pre-treatment is performed and may vary depending on the initial data and the task at hand but is usually similar in essence. Influences on the baseline, e.g. caused by fluorescence or differences in the overall intensity of a Raman spectrum, can be removed by baseline correction and normalization [35]. Smoothing algorithms can prevent misinterpretation from non-specific noise. The preprocessed data is then used to train models based on various classification methods, often using cross-validation to counteract overfitting [36]. In the end, an independent validation of the models [37,38] should always be performed with new data unknown to the model. There is increasing Raman spectroscopic data and growing biological knowledge due to the rapid development of single cell Raman spectroscopic applications with simultaneous progress in machine learning [39].

In a recent review, Rodriguez et al. [40] conclude that the use of advanced data processing methods will allow for a better interpretation of Raman spectra and the detection of minor differences in biological samples. Since it might be a problem to compare the results obtained from different data processing programs, future work should address this problem by developing a universal tool or database for more reliable results [40]. In our recent studies we showed that differentiation of bacteria and yeasts after standardized growth conditions using microbial bulk material for spectral recording was possible [11]. In this context, we were able to get predictions with accuracies of over 98% using SVMs and have also shown that models based on Raman spectra acquired from bulk material on a silicon dioxide (SiO_2) protected silver mirror slide are largely suitable for making predictions of Raman spectra obtained from microorganisms on stainless steel substrates [11]. Although the substrate does influence Raman spectra considerably [34], these results prove the general option to make predictions for spectra collected under other conditions. This fact is interesting, especially when considering the idea of large, universal predictive models. We have also shown that predictions for dry surface biofilms (DSBs) can be made using our models, which were originally not designed for this purpose at all, albeit with larger limitations [41].

In a different approach [42], we used Raman microscopy to train predictive models in order to identify different isolates and species of fungal spores using models that contained both bulk and single cell spectra [42] with the fungal spores studied being many times larger than, for example, bacterial cells. The question whether Raman spectra of bulk microbial material can be used to make predictions for Raman spectra obtained from single cells of the same species was subject to the investigations presented here. The extensive collection of data from bulk material is easier and more practical than collecting Raman spectra from single cells. Single cells must first be located and targeted, which could be done automatically by image recognition [43]. Considering practical applications of Raman single cell analysis, it becomes clear that single cells require more reference data for robust identifications, as the origin of individual bacteria has not been standardized and large, mainly growth-related, variations in cell composition may occur [44]. Should it be possible to make predictions from bulk data for single cells, the preceding data collection for a particular single cell application could potentially be facilitated. In addition, the results of our work could be helpful in assessing universal tools or databases. The collected bulk Raman spectra from our previous work [11], were used to train a NN without prior PCA. This model only includes bulk microbial spectra recorded on silver mirror slides as well as bulk spectra recorded on stainless-steel slides. We additionally recorded Raman spectra of bacterial single cells and yeasts on stainless-steel slides and tested the predictive capabilities of the model. Thus, our main goal was to assess the suitability of microbial bulk Raman spectra as a basis for single cell predictions. The subordinate goal of this work, namely the assessment of more universal Raman spectroscopic tools, was deepened by training and evaluating additional models that include both bulk and single cell spectra.

2. Materials and methods

2.1. Growth conditions and sample preparation

The detailed growth conditions and sample preparation for bulk material are described in Ref. [11]. Glycerol stocks of the microorganisms obtained from the Leibniz-Institut DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH,

Braunschweig, Germany), were stored at −80 °C, thawed, streaked on universal tryptic soy agar (TSA) plates (Merck, Darmstadt, Germany) or malt extract agar (MEA) (Merck, Darmstadt, Germany) using sterile inoculating loops and incubated at 30 °C for 24 h depending on species (see Table 1). From these cultures, a subculture was obtained by transferring material to other agar plates and incubated for 24 h. These cultures were then used for the investigation. The microorganisms used and the growth parameters, including variations, are listed in Table 1. Although the microorganisms used are important in many areas such as hygiene, medicine or the food sector, the initial aim in selecting the microorganisms was to collect as broad a spectrum of data as possible with microorganisms that were suspected of having both great Raman spectroscopic similarities and great differences.

2.1.1. Bulk material

One ml 0.9 % sterile NaCl solution was pipetted in 1.5 ml sterile reaction tubes. Cell material was collected with an inoculation loop and transferred into the NaCl solution. After mixing with a vortex mixer the suspension was centrifuged (Heraeus Fresco 17, Thermo Scientific, Dreieich, Germany) for 3 min at 5000 rpm. The supernatant was discarded and the cell pellet with some residual moisture was homogenized with a sterile pipette tip and then 1 µl was transferred to a SiO2 protected silver mirror slide (PFR14-P02, Thorlabs, Bergkirchen, Germany) and to a highly polished stainless-steel slide (Renishaw, Pliezhausen, Germany) and then dried for 15 min at 20 °C.

2.1.2. Single cells

One ml of sterile 0.9% saline was placed in 1.5 ml reaction tubes. As with the bulk material samples, material was transferred but then a serial dilution was prepared. For this, 900 µl of sterile NaCl solution was transferred into 1.5 ml reaction tubes and 100 µl of the previous solution was transferred, mixed via vortexer, and transferred in the next tube with 900 µl NaCl solution. For the analysis of the single cells, the 1:1000 dilutions were used since the distribution of the cells there proved to be appropriate. At this point it must be mentioned that the salt solution crystallizes on the slides, which can make the localization of single cells more difficult. Experiments using deionized water instead of NaCl solution simplify the localization of single cells but show a change in the Raman spectra (Supplementary data S1), which is why saline solution was used despite the rather more difficult conditions in localizing the single cells. Droplets of one µl were placed on the slides using sterile pipette tips and dried at 20 °C for about 15 min.

2.1.3. Spectral recording

A confocal Raman microscope (inVia, Renishaw, Gloucestershire, UK) was used with an excitation wavelength of 633 nm (Helium–Neon (HeNe)) and a 100× magnification lens. The detected spectral region was between 606 cm^{-1}–1736 cm^{-1} with a resolution of about 1.1 cm^{-1} (1800 l/mm grating). All spectra were recorded with about 3.5 mW laser power on sample (50 % of the total power of the laser model used). Laser diameter was about 7–8 µm. In total, Raman spectra of 13 species were collected in this work, of which 11771 bulk spectra were recorded on silver mirrors, 7599 bulk spectra were recorded on stainless steel slides, and 551 single cell spectra were recorded on stainless steel slides.

Table 1
List of species included in the predictive model with their DSMZ-No. and information on the number of spectra per substrate, number of independently grown cultures, used exposure time, accumulations per spectrum and growth conditions; all spectra have been recorded with an excitation wavelength of 633 nm at about 3.5 mW laser power on bulk and single cell samples.

Microorganism	Abbreviation	DSMZ-No.	Spectra on silver slide	Spectra on stainless steel slide	independent cultures	Single cell spectra on stainless steel slide	Exposure time; accumulations	Nutrition media	Cultivation time
Acinetobacter radioresistens	*Ara*	6976	725	581	2	41	1.5; 20	TSA	24 h
Brevundimonas diminuta	*Bdi*	7234	897	585	4	40	1.5; 20	TSA	24 h
Candida albicans	*Cal*	1386	400	584	2	42	1.5; 20	MEA	24 h
Candida boidinii	*Cbo*	70034	686	588	2	41	1.5; 20	MEA	24 h
Chryseobacterium indologenes	*Cin*	16777	684	616	2	40	1.5; 15	TSA	24 h
Enterococcus faecium	*Efa*	2146	680	602	3	42	1.5; 20	TSA	24 h
Enterococcus hirae	*Ehi*	3320	1423	593	3	50	1.5; 20	TSA	24 h
Escherichia coli	*Eco*	423	1190	593	4	45	1.5; 20	TSA	24 h
Micrococcus luteus	*Mlu*	1790	1842	616	7	41	1.5; 15	TSA	24 h
Pseudomonas aeruginosa	*Pae*	939	385	589	2	42	1.5; 20	TSA	24 h
Pseudomonas fluorescens	*Pfl*	50090	654	583	2	41	1.5; 20	TSA	24 h
Staphylococcus aureus	*Sau*	799	1094	616	4	43	1.5; 15	TSA	24 h
Staphylococcus epidermidis	*Sep*	1798	1111	453	3	43	1.5; 20	TSA	24 h

2.1.4. Bulk material

The spectra were recorded in a spiral shape pattern from the inside to the outside with a distance of about 2–4 μm between the measuring spots. This was done to bleach out possible fluorescence interferences. Similar spirals were recorded at several locations on the surface of the samples, each comprising 50 to 300 spectra. Fig. 1 (a) depicts the spirally distributed measuring points on an exemplary microbial bulk surface, which start in the center and then move outwards. Although the laser spot diameter (Fig. 1 (b)) is larger than the distance between the measuring points, the effective focus range changes with each measuring point, so that a certain Raman spectroscopic signal variation occurs due to different sample positions, as well as a bleaching effect. Fig. 1 (c) illustrates an example of *Pfl* spectra of the first 25 measuring points of such a spiral measurement and the associated fluorescence reduction.

Information about all species used for modelling, DSM-numbers, and additional information about used exposure time, accumulations per spectrum and growth conditions can be found in Table 1. For all species that visibly contained color pigments, only 15 accumulations were used, and the spectra were not recorded in a spiral pattern but side by side with a spacing of 4 μm. On the one hand, this procedure should lead to the carotenoids not photodegrading, but on the other hand, the measuring points should still be close enough to each other for a photobleaching effect to occur [11].

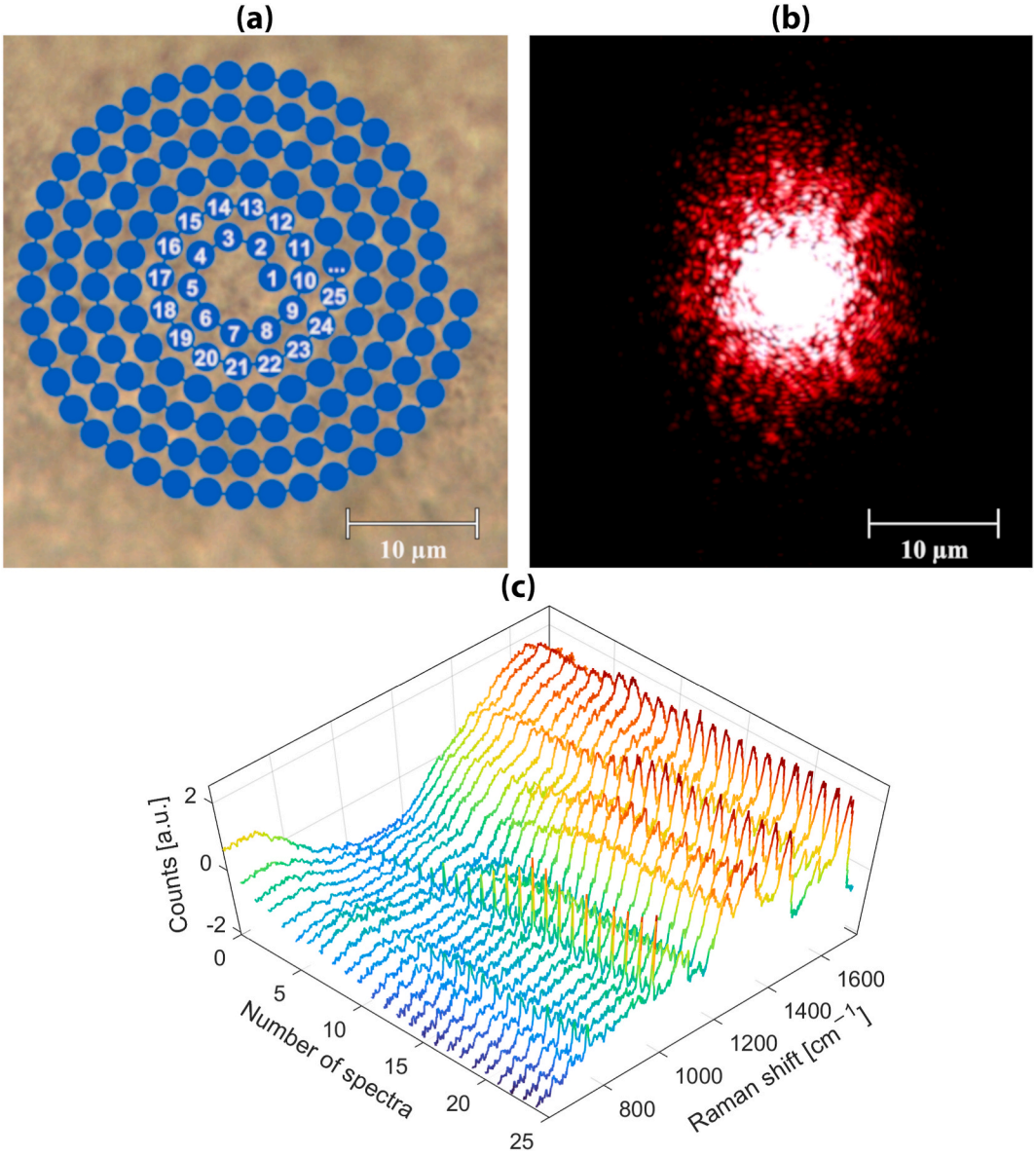


Fig. 1. Spiral-shaped measuring points on a bacterial bulk surface (a), laser focus on this sample (b) and exemplary untreated but normalized *Pfl* spectra of the first 25 measuring points of a spiral-shaped measurement (c).

2.1.5. Singe cells

Approximately 20 individual cells per species were targeted and two spectra of the same cell were recorded in direct succession, using the same measurement settings as for the bulk measurements (Table 1).

2.2. Data preprocessing

For data preprocessing MATLAB R2022b was used (MathWorks, Massachusetts, USA). After the spectra were interpolated, baseline correction ("msbackadj") and smoothing (Savitzky-Golay filtering) took place using MATLAB code below:

yOut = msbackadj(X, Intensities, 'WindowSize',50, 'StepSize',50)

y = sgolayfilt(yOut,3,13);

yOut = Baseline corrected Raman spectra X = Raman shift axis Intensities = Raman spectra (without Raman shift axis) y = baseline corrected and smoothed Raman spectra.

All spectra were normalized (z-score) after baseline subtraction and smoothing. The entire spectra, i.e., all 1015 data points per spectrum, were used to train the model without applying any dimension reduction methods. Only the bulk spectra were used for modeling (19370 spectra).

2.3. Model development

The Deep Network Designer App of MATLAB R2023a (MathWorks, Massachusetts, USA) was used for model development. The basic structure of the CNN, shown in Fig. 2, is: Input Layer - > Hidden Layers - > Output Layer with the input layer being the previously mentioned bulk spectral data. The hidden layer is comprised of a flatten layer for restructuring the input matrix into a vector, followed by three convolutional blocks combining three layers. Each convolutional block consists of a convolutional layer with the parameters (S = filter size, N = number of filters) $S_1 = 78$, $S_2 = 289$, $S_3 = 135$ and $N_1 = 27$, $N_2 = 100$, $N_3 = 77$ respectively. The convolutional layer is followed by a batch normalization layer and a rectified linear activation layer. After the convolutional blocks follows a fully connected layer with a 13-feature output. In the output layer data is sorted with a softmax layer for probability distribution for each input class and fully classified with a classification layer according to the species classes specified in the input layer. The parameters for the convolutional layers were optimized using the Experimental Manager App of MATLAB R2023a (MathWorks, Massachusetts, USA) with a Bayesian optimization algorithm that aims to maximize cross-validation accuracy. A total of 75 trials were trained with a possible range of 1–300 for filter size and number of filters for the convolutional layers. No time limit was set for the training of each trial. For each training the learning rate was set at 0.02, the momentum at 0.9 and a maximum of 60 epochs. When building the CNN based on microbial bulk Raman spectra, the focus was not on accurate predictions for Raman spectra of single cells, but on high accuracy within cross-validation. However, the basic architecture and modification of the network layers of the CNN was inspired by various publications [17,45,46].

In addition, models were trained that contained a certain percentage of single cell Raman spectra. The architecture of the CNNs were not changed. The used single cell spectra were shuffled and 10%, 20%, 25% and 30% of all single cell spectra were added to the bulk training data and new models were trained. It is therefore possible that due to the random selection of certain spectra, slightly more or fewer single cell spectra are included for individual species.

3. Results and discussion

3.1. Comparison of bulk and single cell Raman spectra

With the measurement settings and instrument parameters used, strong Raman spectra from single cells could be obtained. Interfering fluorescence did not play a notable role when single cells were present, whereas fluorescence was an issue in bulk measurements for many species when using a 633 nm excitation wavelength [11]. From various publications it can be concluded that 633 nm excitation wavelength is a viable compromise between still strong Raman signals and still manageable fluorescence disturbances [11,41,47–49].

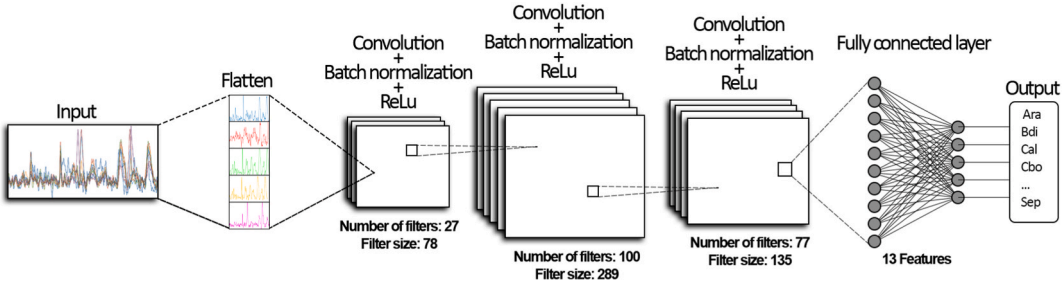


Fig. 2. Visual representation of CNN architecture used in this study.

Fig. 3 (a) illustrates all bulk Raman spectra of the 13 species studied with highlighted mean spectra and the corresponding single cell counterpart in Fig. 3 (b). To highlight the elementary differences directly, the single cell mean spectra of a respective species were subtracted from the corresponding bulk Raman spectra in Fig. 3 (c). Most strikingly, differences in the spectra of microorganisms containing carotenoids (*Cin*, *Mlu*, *Sau* – for abbreviations, see Table 1) could be observed. For *Cin* and *Sau*, almost no detectable signals were present in the mean spectrum at about 1530 cm^{-1} and 1132 cm^{-1} (*Cin*) resp. 1522 cm^{-1} and 1159 cm^{-1} (*Sau*). The effect was smaller for *Mlu*, where the characteristic peaks triggered by carotenoids [50] were still largely visible, albeit noticeably weaker. Thus, although carotenoids seem ideal features for the detection and differentiation of microbial cells by Raman microscopy [50], they are a potential confounding factor for classification when their signals are present in the bulk data but not in the single cell spectra. On the one hand it became obvious especially with *Mlu*, that the carotenoid signals were still relatively strong in the first recorded spectrum of a cell, but almost disappeared in the second spectrum. With some cells, however, there were no longer any noticeable carotenoid signals even with the first acquisition, or a significant photodegradation already occurred through the accumulation of the 15 individual spectra. In *Cin* there were a few spectra still showing visible carotenoid signals, while in *Sau* there were none. The species-specific differential decrease in carotenoid signaling could be related to the type of carotenoids such as Staphyloxanthin in *Sau* [51], Sarcinaxanthin in *Mlu* [52], or combinations of carotenoid and Flexirubin in *Cin* [53]. It may also be due to interactions of carotenoids with other cellular components [50].

Comparing the bulk and single cell spectra, *Pae* and *Pfl* showed particularly little variation (Fig. 3 (c)), which seems counterintuitive at first due to the fact that these organisms should exhibit a strong fluorescence in bulk material due to the production of pyoverdins [54,55]. The fact that this was not the case might be explained by an interplay of several factors. First, due to the specific measurement process when analyzing the bulk material (as described in Ref. [11]), fluorescence might have largely bleached out. In addition, after the relatively short growth period of only 24 h, the number of interfering fluorophores produced might still have been relatively low. Using saline could also act as a kind of washing step, which could also have reduced fluorescence [56]. As far as yeasts are concerned, further noticeable features were observed. *Cal* showed many peculiarities in the difference spectrum (Fig. 3 (c)) and at the bulk spectra of *Cal* from Fig. 3 (a), the partially weak expressions of the signals and a "pulling up" of the spectra in the front region at 600 cm^{-1} can be seen. This is probably due to strong fluorescence, the influence of which was still visible after baseline correction. Other baseline filter methods could prove to be more efficient. In contrast, the fluorescence of the other yeast species, *Cbo, was* lower. However, an additional signature at about 1154 cm^{-1} was noticeable in *Cbo*, which was no longer visible in the mean spectrum. Bednárová et al. reported that yeast vacuoles can be reliably identified by this signal, triggered by polyphosphates, which is absent in other yeast compartments [57]. Thus, the fact that the band appears only sometimes may reflect different physiological states of the cells [57], given the fact that the cells are relatively large (about 10 μm), and the vacuoles are not always sufficiently in focus of the laser. Finally, it was possible to obtain spatially resolved specific signals of individual cell compartments of the investigated yeasts by Raman microscopy [57]. In the single cell spectra, the band at about 1154 cm^{-1} did not appear or was extremely weak (Fig. 3 (b)).

After data pretreatment and especially in the representation in Fig. 3, differences in the spectra depending on different positions or the substrate are not recognizable. For this reason, Fig. 4 shows an example of the untreated but normalized Raman spectra of *Ehi* from

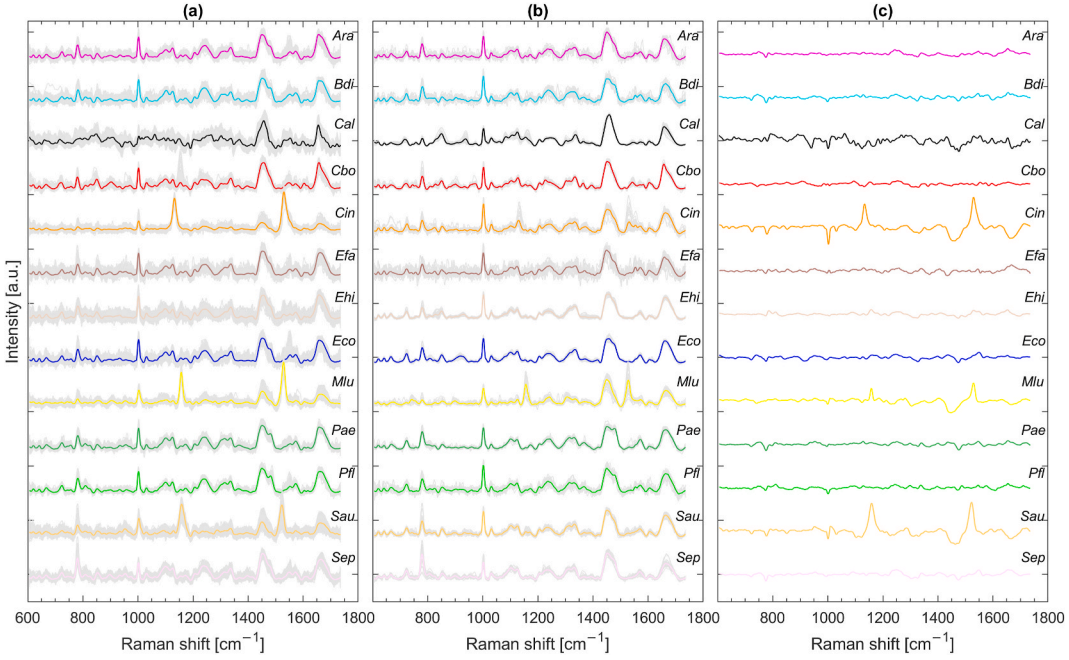


Fig. 3. All Raman spectra (grey curves) and the respective mean spectrum highlighted in color obtained from either bulk material (a) or single cells (b). The difference spectrum of each bulk mean spectrum to the corresponding single cell mean spectrum is shown in (c).

four different positions on the silver slide (Fig. 4 (a)) and the stainless-steel slide (Fig. 4 (c)). The variation is smaller on the stainless-steel slide, but this is not solely due to the substrate, but can also be related to individual differences in the samples or minor deviations in the execution [11]. Due to the SiO_2 protective layer on the silver surface, roughness or damage is easily noticeable. For this reason and due to the bulk material covering the substrate, any SERS effects are unlikely and have not been observed, but generally cannot be completely ruled out. What becomes clear is that there are systematic differences in the Raman spectra depending on the measurement areas, which can be particularly illustrated by performing a PCA (Fig. 4 (b)). It is also striking that the spectra of the different areas in the PCA do not necessarily cluster, but that the clusters also mix (Fig. 4 (d)), which confirms that variations also occur in a single spiral area despite the proximity of the individual measurement points to each other.

3.2. Predictive models and cross-validation

The structure of the CNN was derived empirically, starting with a singular convolutional block. Increasing the number of convolutional blocks led to higher accuracy of the predictions, however after three blocks no further improvement was observed. Commonly, in a CNN the convolutional layer is followed by a pooling layer to reduce the features extracted by the convolutional filter. The CNN used in this work does not use pooling layers as they lead to reduction in prediction accuracy. This could be due to the removal of features with low significance, which might still contribute to the overall classification [58].

The confusion matrix generated using five-fold cross-validation (Fig. 5) shows the percent prediction accuracies and the corresponding number of incorrectly predicted microorganisms from other groups. Overall, the model suggests a high prediction accuracy in the cross-validation with an average of 99.0%. The least accurate predictions within the cross-validation were for *Eco* at 97.1% and *Efa* at 97.5%. It should be mentioned that good prediction accuracies from cross-validation within large data sets with high similarities within the data must be viewed with caution, as overfitting still might influence the results. Because Raman spectral changes in microorganisms are often small and can be affected by batch-to-batch variations or instrumental/methodical impacts multiple

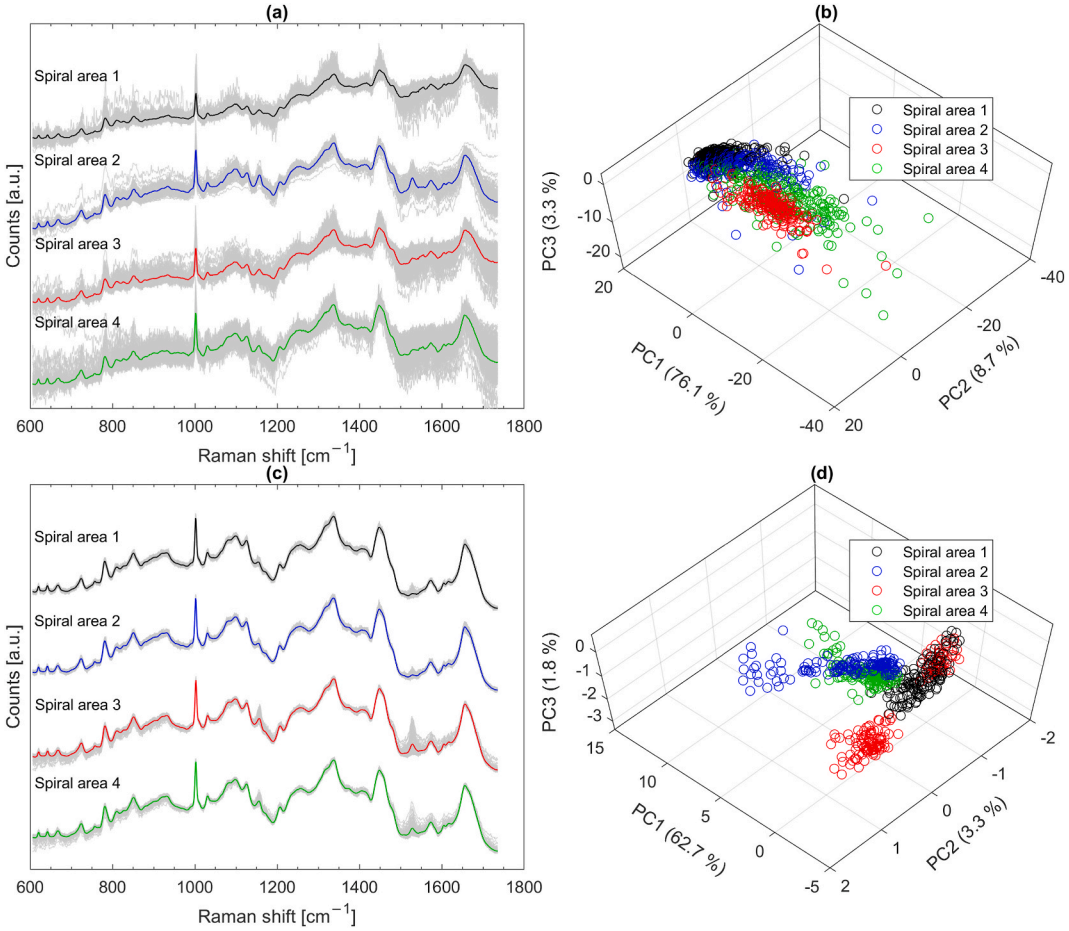


Fig. 4. Exemplary variability in the untreated but normalized bulk Raman spectra of the bacterium *Ehi* depending on different spirally arranged sample measurement areas on the silver slides (a) and the stainless steel slides (c), as well as the first three PCs of a PCA with the spectra of the silver slide (b) and stainless steel slide (d) with indication of the explained variance in %.

independent batches are recommended in Raman cell studies [59] which is partly the case here (Table 1) but could nevertheless be further expanded. Especially, however, in our application of the model to completely different data, namely single cell spectra, deviations are expected anyway and a good cross-validation result within the bulk model does not necessarily refer to the ability to make predictions for single cells. Nevertheless, the general applicability of the model has been proven here.

Fig. 6 shows the confusion matrix describing the predictions of the bulk CNN model for single cell spectra. The average prediction accuracy achieved across all species is 60.1%. *Efa* and *Sep* were predicted 90.5% and 86.0% correctly, respectively. The yeast *Cbo* was classified 90.2% correctly and the other yeast (*Cal*) 88.1% where the 11.9% incorrectly predicted species was *Cbo*. Within the investigated *Pseudomonas* genera *Pae* was relatively accurate identified with about 81.0%, while *Pfl* was predicted less accurate with about 48.8%. Most of the incorrect predictions for *Pfl* were *Bdi* and *Eco* (about 24% each) and only 2.4% within the same genus of *Pseudomonas* (*Pfl*). *Ehi* was identified 74.0% correctly, but most of the incorrect predictions were within the same genus of *Enterococcus* (10.0% *Efa*), however also 10.0% of the *Ehi* spectra were predicted as *Cbo*, a completely different cell type. Predictions for *Cin*, *Mlu*, and *Sau* were particularly unprecise, which might be explained by the influences of photodegradation of carotenoids described above. *Sau* is never correctly identified, which is not surprising given the total absence of carotenoid signals in single cells. The Raman spectra of the single cells of *Cin* still partially contained carotenoid signals, but these were barely detectable on average (Fig. 3 (b)), which is reflected in the prediction accuracy of 20.0%. The *Mlu* carotenoids were least affected by photodegradation; its specific signatures at about 1156 and 1528 cm^{-1} were still well discernible on average (Fig. 3 (b)). *Mlu* single cells were predicted 70.7% correctly. Fig. 7 demonstrates the principal problem of photodegradation for classification models. As an example, 100 additional spectra were recorded one after the other from a bulk sample from the exact same location with the measurement settings described in the methodology. As this is a bulk sample, the process to complete degradation takes longer than for single cells. Fig. 7 (a) shows the degradation of the two bands typical for carotenoids at about 1159 cm^{-1} and 1522 cm^{-1} of a *Sau* sample (not included in the model, but additional 100 spectra). *Sau*, like *Efa*, are gram-positive cocci. The more carotenoids are degraded, the more the *Sau* spectra resemble the *Efa* spectra. The first two principal components of a PCA with bulk spectra from *Efa*, *Sau* and the photodegraded *Sau* (Fig. 7 (b)) illustrate this problem. If, for example, a classification method uses the first two principal components as predictors, *Sau* bulk spectra with bleached carotenoid signals could be misinterpreted as *Efa*. Conversely, this illustrates the problem of making predictions from bulk spectra for single cell spectra, since single cell Raman spectra from *Sau* do not show any carotenoid signals, but the *Sau* bulk spectra do.

The measurement settings seem to be suitable for all microorganisms without carotenoids, since strong signatures could be detected

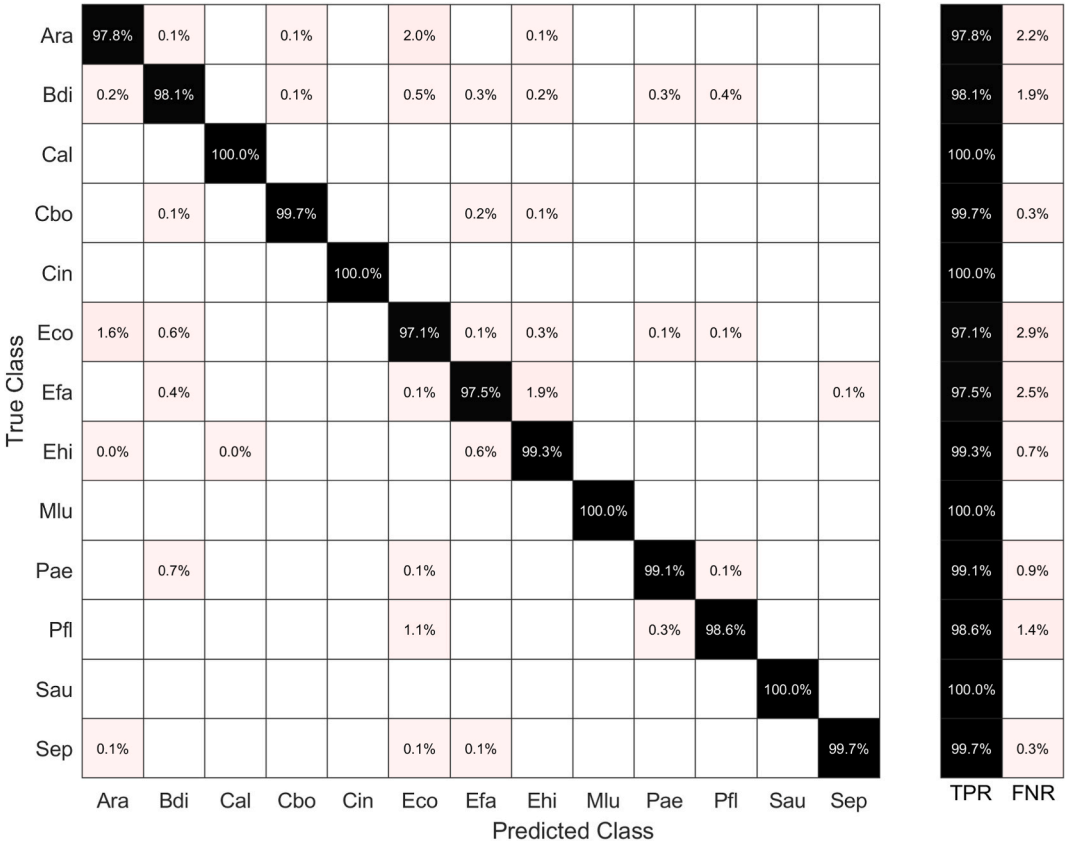


Fig. 5. Confusion matrix of CNN with bulk spectra using 5-fold cross validation, with true positive rate (TPR) and false negative rate (FNR).

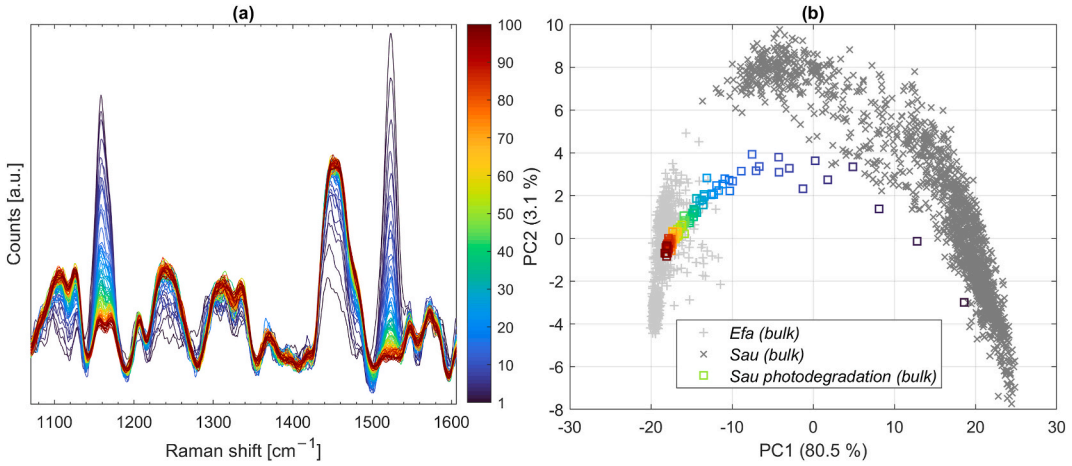


Fig. 6. Decrease in carotenoid-related bands at approximately 1159 cm^{-1} and 1522 cm^{-1} by 100 measurements taken consecutively on the same spot of a *Sau* bulk sample, where the color bar indicates the number of spectra from the first spectrum (deep purple) to the hundredth spectrum (deep red) (a); First two principal components of a PCA with bulk spectra of *Efa* and *Sau* and spectra of the intentionally photodegraded *Sau* sample (b) with indication of the explained variance in %.

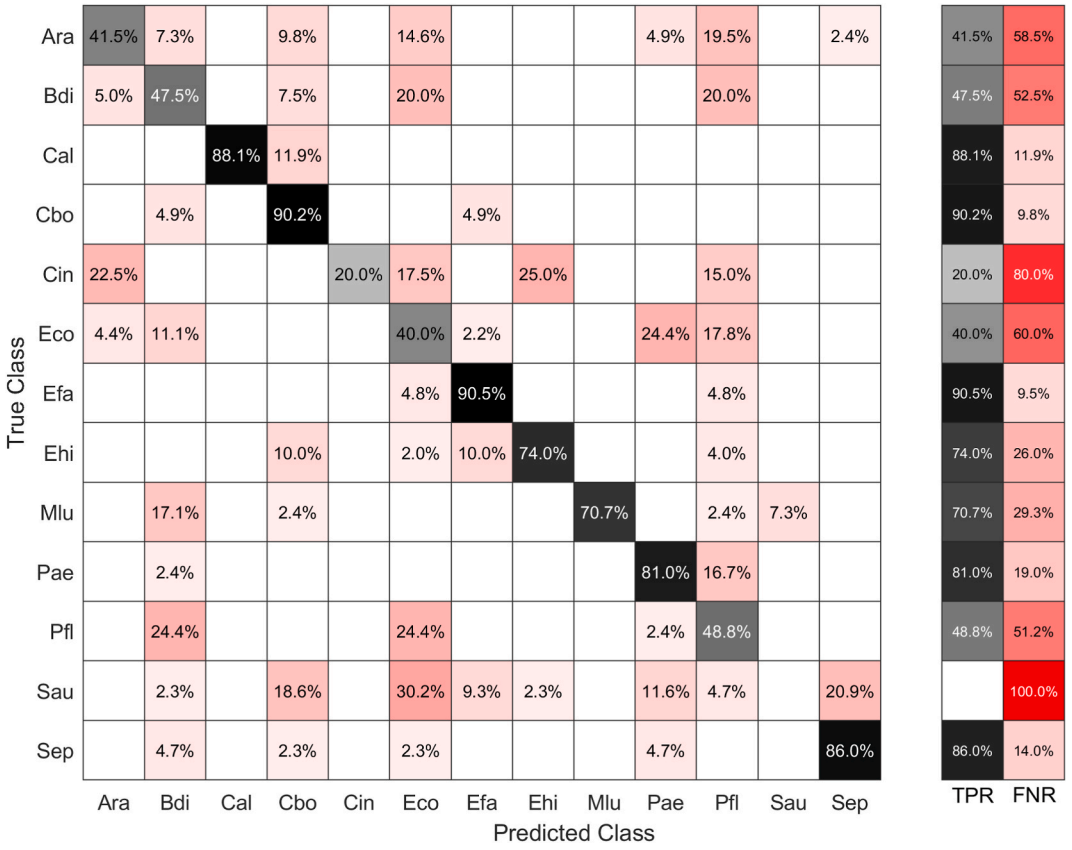


Fig. 7. Confusion matrix for predictions of single cell Raman spectra via CNN based on bulk spectra, with true positive rate (TPR) and false negative rate (FNR).

(Fig. 3 (b)) and no burning of the cells took place (no characteristic bands as found in burnt microorganisms [60]). However, lower laser powers or shorter measurement times could yield more consistent spectra of the carotenoid-containing microorganisms, whereby other, possibly important signals may be less pronounced. Changing the excitation wavelength may also be useful. In our study, an excitation wavelength of 633 nm was used. Frequently 532 nm is used for single cell applications [59], which in particular detects

cytochrome in cells due to resonance enhancement [59,61]. Excitation in the 514.5 nm range was also shown to cause photodamage in living cells, but not 660 nm [59,62]. An excitation of 785 nm had been shown to be suitable for studies in living cells due to lower phototoxic effects [59,63]. Fourier transform (FT) Raman spectroscopy using a near-infrared (NIR) excitation wavelength of 1064 nm could also be advantageous. For example, Naumann et al. report that essentially fluorescence-free spectra were obtained even from highly stained bacterial samples [64]. The work of Prucek et al. and Kairyte et al. also confirms the avoidance of fluorescence background through NIR excitation and they achieved meaningful Raman spectra [65,66].

If a model is based on data recorded using the same conditions, photodegradation is not a major problem in differentiation, since this effect would be then included in the data contained in the model. Although predictions were particularly inaccurate for *Bdi* (47.5%) the 20.0% assignment to *Pfl* can be considered interesting, after all *Brevundimonas* formerly belonged to the genus *Pseudomonas* due to certain taxonomic similarities [67]. However, some other inaccurate predictions should be considered more critical, with *Eco* being correctly predicted by about 40.0% but incorrectly predicted by 24.4% as *Pae* and 17.8% as *Pfl*. Considering the problem with carotenoids and incorrect predictions for the same genus, our study shows that bulk Raman spectra data are suitable to make predictions for single cells to a not inconsiderable extent. Five of the 13 species included were over 80% correctly predicted. These include the yeasts *Cbo* and *Cal* which were correctly classified at 90.2% and 88.1%, respectively. *Efa*, an important fecal indicator [68] and test germ in analytical microbiology [69], with nearly 90.5% correct identification. As well as *Sep*, a ubiquitous skin colonizer and biofilm former [70] with 86.0% and the typical nosocomial germ *Pae* [71] with 81.0%. *Ehi* and *Mlu* are correctly predicted at over 70%. The most inaccurately predicted microorganisms besides the carotenoid-containing bacteria *Sau* and *Cin* are *Eco* (40.0%), *Ara* (41.5%), *Bdi* (47.5%), and *Pfl* (48.8%).

Since one purpose of our work was to assess possible universal tools or databases, we considered the predictive performance of other models based on the same architecture but containing both bulk spectra and certain proportions of single cell spectra. For this purpose, 10%, 20%, 25% and 30% of the single cell spectra were added to the bulk spectra, the models were retrained and predictions were made for the remaining single cell spectra. Table 2 shows that the cross-validation accuracies changed only marginally and is between 99.0% (no single cell spectra in the model) and 98.6% (30% single cell spectra in the model). However, the prediction accuracy for the remaining single cell spectra that were not included in the models improves remarkably and increases from 60.1% (no single cell spectra at all) to up to 75.1% (30% single cell spectra). This not only speaks in favor of our basic idea of the development of more universal tools, but also shows that Raman spectra of single cells, which are somewhat more time-consuming to collect, could possibly be saved and reliable predictions could still be made. Nevertheless, it must be considered that the single cell Raman spectra came from the same batch in each case and possibly even a Raman spectrum from one and the same cell is included in the model as a reference. More independent data are needed to strengthen these observations.

In addition to looking at the overall prediction accuracy (Table 2), Fig. 8 shows the individual predictions for single cells from the CNN, to which 30% of the single cell spectra were added for training. The improvement in the predictions of *Sau* is particularly evident, increasing from 0% via the bulk model to 86.2%. The carotenoid-containing *Cin* was also predicted correctly at 86.7% instead of 20%. The predictions for *Ehi* improved from 74.0% to 94.7% and for *Eco* from 40.0% to 62.5%. *Ara* was also correctly classified with 57.1% instead of only 41.5%. The predictions of many other species also improved, but only slightly, such as *Pae* from 81.0% to 84.6%, *Cbo* (from 90.2% to 92.0%) and *Sep* from 86.0% to 87.5%. The predictions for *Efa* and *Mlu* remain about the same. The prediction accuracies have only worsened for *Bdi*, *Cal* and *Pfl*.

4. Conclusion

Our data suggest that bulk Raman spectra can be useful for identifying single cell spectra to some extent. Overall, five of the 13 included species are correctly predicted with more than 80%. Seven out of 13 species are predicted correctly with more than 70% and six of 13 species are correctly predicted at less than 50%. Most importantly, it could be shown that higher fluorescence in bulk material and stronger photodegradation in single cells are major factors for less accurate prediction performance. Thus, there might be a potential for larger universal applications, for example, when including both bulk and single cell data. The addition of single cell spectra to the bulk models, without changing the model architecture, already led to a drastic improvement in prediction accuracy. With the complexity of neural networks, the model can be adjusted and improved further to increase the prediction rate for the current selection of species or to adjust for future species that may be added to the database. The adaptability of CNNs can elevate problems arising from difficult measurements and can be optimized to tackle specific issues such as the degradation of carotenoids. The development of a fully optimized neural network in terms of structure and hyperparameters however, is outside the scope of this paper. Many

Table 2
Prediction accuracy of the five-fold cross-validation of models that contain a certain percentage of single cell spectra, as well as the overall prediction performance for the single cell spectra not contained in the model.

Percentage of randomly selected single cell spectra added to model training	Overall predictive accuracy of the five-fold cross-validation of the model.	Overall predictive accuracy for remaining (not included in the model) single cells.
0%	99.0%	60.1%
10%	98.9%	69.6%
20%	98.6%	72.1%
25%	98.8%	73.7%
30%	98.6%	75.1%

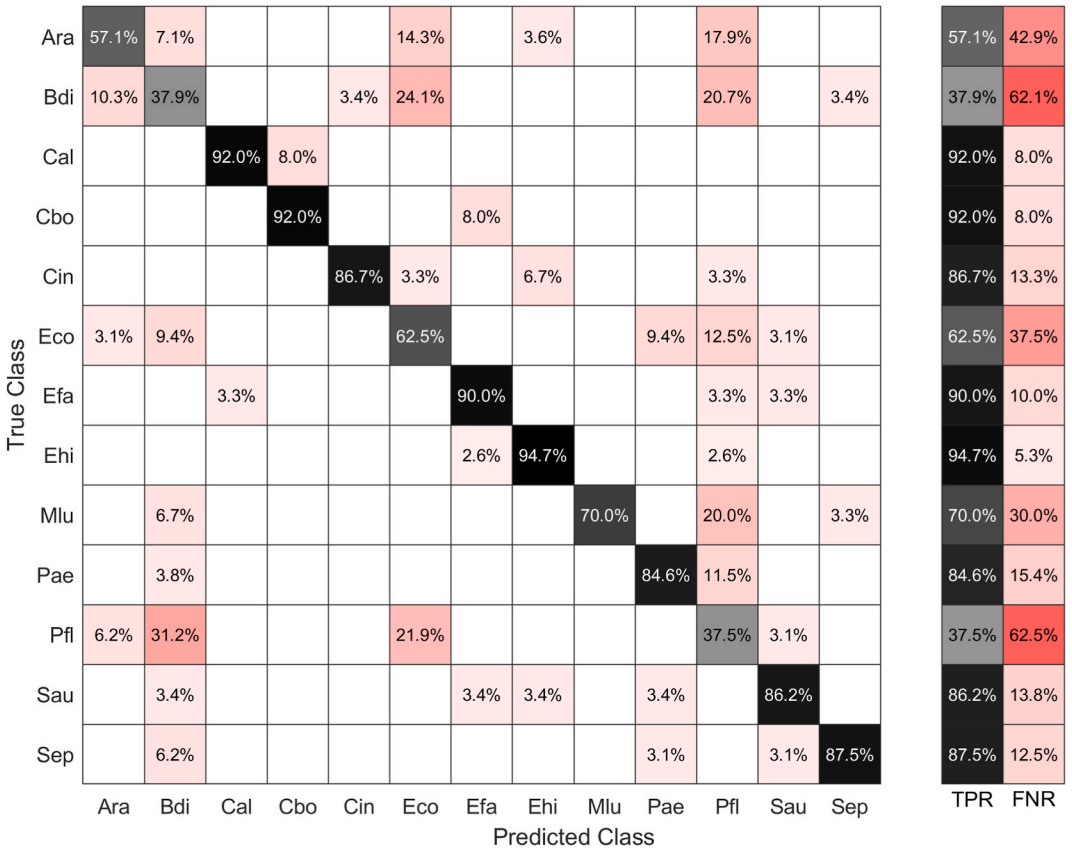


Fig. 8. Confusion matrix for predictions of single cell Raman spectra via CNN based on bulk and 30% of the single cell spectra, with true positive rate (TPR) and false negative rate (FNR).

publications address highly specialized applications such as detection of *Brucella* in milk [20], differentiation of *Staphylococcus* species for possible clinical diagnostics [72] or the differentiation between specific antibiotic resistant bacteria [73]. Likewise, there are studies using numerous species included in a predictive model [30] but much less frequently Raman is considered as a kind of universal spectroscopic approach, which, for example matrix-assisted laser desorption/ionization mass spectrometry MALDI-TOF MS allows for the identification of a wide range of species [74], whereby cultivation is necessary for this and single cell measurements are not possible there. Despite many advantages, such as speed, no or low invasive measurements and single cell analysis, Raman spectroscopy is not regularly used for the analysis of e.g. clinical samples, maybe because complex models are often required and the spectra alone usually cannot be directly interpreted [75]. Further reasons may include different data pretreatment methods or the fact that chemometric methods are required in the first place to convert the spectral data into easily interpretable information [76]. In this regard, projects like RAMANMETRIX [75] are interesting, where users can apply relatively simple chemometric methods to their Raman data, and even some data sets are freely available. Thus, more intensive exchange and collaboration could lead to biological Raman spectroscopic applications that make their way out of the lab.

Data availability

Sections of the data will be made available on request.

CRediT authorship contribution statement

Thomas J. Tewes: Writing – original draft, Investigation, Conceptualization. **Mario Kerst:** Investigation. **Svyatoslav Pavlov:** Investigation. **Miriam A. Huth:** Investigation. **Ute Hansen:** Validation, Conceptualization. **Dirk P. Bockmühl:** Writing – review & editing, Supervision.

Declaration of competing interest

No funding was received to assist with the preparation of this manuscript.

The authors have no financial or proprietary interests in any material discussed in this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27824.

References

[1] P. Carey, Biochemical Applications of Raman and Resonance Raman Spectroscopes, Elsevier Science, 2012.
[2] D.A. Long, Survey of light-scattering Phenomena, in: The Raman Effect, 2002, pp. 3–18.
[3] S. Ewen, D. Geoffrey, The Raman experiment – Raman instrumentation, sample presentation, data handling and practical aspects of interpretation, in: Modern Raman Spectroscopy, 2019, pp. 21–75.
[4] S. Stöckel, J. Kirchhoff, U. Neugebauer, P. Rösch, J. Popp, The application of Raman spectroscopy for the detection and identification of microorganisms, J. Raman Spectrosc. 47 (2016) 89–109, https://doi.org/10.1002/jrs.4844.
[5] M. Harz, P. Rösch, J. Popp, Vibrational spectroscopy-A powerful tool for the rapid identification of microbial cells at the single-cell level, Cytometry 75 (2009) 104–113, https://doi.org/10.1002/cyto.a.20682.
[6] T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, J. Popp, How to pre-process Raman spectra for reliable and stable models? Anal. Chim. Acta 704 (2011) 47–56, https://doi.org/10.1016/j.aca.2011.06.043.
[7] W. Schumacher, S. Stöckel, P. Rösch, J. Popp, Improving chemometric results by optimizing the dimension reduction for Raman spectral data sets, J. Raman Spectrosc. 45 (2014) 930–940, https://doi.org/10.1002/jrs.4568.
[8] T.J. Tewes, M.C. Welle, B.T. Hetjens, K.S. Tipatet, S. Pavlov, F. Platte, D.P. Bockmühl, Understanding Raman Spectral Based Classifications with Convolutional Neural Networks Using Practical Examples of Fungal Spores and Carotenoid-Pigmented Microorganisms. AI, vol. 4, 2023, pp. 114–127, https://doi.org/10.3390/ai4010006.
[9] Y. Zhu, W. Liu, S. Liu, M. Li, L. Zhao, L. Xu, N. Wang, G. Zhao, Q. Yu, Preparation of AgNPs self-assembled solid-phase substrate via seed-mediated growth for rapid identification of different bacterial spores based on SERS, Food Res. Int. 160 (2022) 111426, https://doi.org/10.1016/j.foodres.2022.111426.
[10] S.D. Dryden, S. Anastasova, G. Satta, A.J. Thompson, D.R. Leff, A. Darzi, Rapid uropathogen identification using surface enhanced Raman spectroscopy active filters, Sci. Rep. 11 (2021) 8802, https://doi.org/10.1038/s41598-021-88026-9.
[11] T.J. Tewes, M. Kerst, F. Platte, D.P. Bockmühl, Raman microscopic identification of microorganisms on metal surfaces via support vector machines, Microorganisms 10 (2022) 556, https://doi.org/10.3390/microorganisms10030556.
[12] K. Rebrošová, S. Bernatová, M. Šiler, M. Uhlirova, O. Samek, J. Ježek, V. Holá, F. Růžička, P. Zemanek, Raman spectroscopy—a tool for rapid differentiation among microbes causing urinary tract infections, Anal. Chim. Acta 1191 (2022) 339292, https://doi.org/10.1016/j.aca.2021.339292.
[13] J. Zhang, J. Zhang, J. Ding, Q. Lin, G.M. Young, C. Jiang, Rapid identification of live and dead Salmonella by surface-enhanced Raman spectroscopy combined with convolutional neural network, Vib. Spectrosc. 118 (2022) 103332, https://doi.org/10.1016/j.vibspec.2021.103332.
[14] J. Sun, X. Xu, S. Feng, H. Zhang, L. Xu, H. Jiang, B. Sun, Y. Meng, W. Chen, Rapid identification of salmonella serovars by using Raman spectroscopy and machine learning algorithm, Talanta 253 (2023) 123807, https://doi.org/10.1016/j.talanta.2022.123807.
[15] P. Rösch, M. Harz, M. Krause, J. Popp, Fast and reliable identification of microorganisms by means of Raman spectroscopy, in: Biophotonics 2007: Optics in Life Science, Optical Society of America, 2007, p. 6633_45.
[16] S. Pahlow, S. Meisel, D. Cialla-May, K. Weber, P. Rösch, J. Popp, Isolation and identification of bacteria by means of Raman spectroscopy, Adv. Drug Deliv. Rev. 89 (2015) 105–120, https://doi.org/10.1016/j.addr.2015.04.006.
[17] C.S. Ho, N. Jean, C.A. Hogan, L. Blackmon, S.S. Jeffrey, M. Holodniy, N. Banaei, A.A.E. Saleh, S. Ermon, J. Dionne, Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning, Nat. Commun. 10 (2019), https://doi.org/10.1038/s41467-019-12898-9.
[18] K. Maquelin, Confocal Raman Microspectroscopy : a Novel Diagnostic Tool in Medical Microbiology, Erasmus University Rotterdam, 2002.
[19] K.C. Schuster, E. Urlaub, J.R. Gapes, Single-cell analysis of bacteria by Raman microscopy: spectral information on the chemical composition of cells and on the heterogeneity in a culture, J. Microbiol. Methods 42 (2000) 29–38, https://doi.org/10.1016/S0167-7012(00)00169-X.
[20] S. Meisel, S. Stöckel, M. Elschner, F. Melzer, P. Rösch, J. Popp, Raman spectroscopy as a potential tool for detection of Brucella spp. in milk, Appl. Environ. Microbiol. 78 (2012) 5575–5583, https://doi.org/10.1128/AEM.00637-12.
[21] P. Rösch, M. Harz, M. Schmitt, K.D. Peschke, O. Ronneberger, H. Burkhardt, H.W. Motzkus, M. Lankers, S. Hofer, H. Thiele, J. Popp, Chemotaxonomic identification of single bacteria by micro-Raman spectroscopy: application to clean-room-relevant biological contaminations, Appl. Environ. Microbiol. 71 (2005) 1626–1637, https://doi.org/10.1128/AEM.71.3.1626-1637.2005.
[22] S.A. Strola, J.-C. Baritaux, E. Schultz, A.C. Simon, C. Allier, I. Espagnon, D. Jary, J.-M. Dinten, Single bacteria identification by Raman spectroscopy, J. Biomed. Opt. 19 (2014) 111610, https://doi.org/10.1117/1.jbo.19.11.111610.
[23] K. Maquelin, C. Kirschner, L.-P. Choo-Smith, N. van den Braak, H.P. Endtz, D. Naumann, G.J. Puppels, Identification of medically relevant microorganisms by vibrational spectroscopy, J. Microbiol. Methods 51 (2002) 255–271, https://doi.org/10.1016/S0167-7012(02)00127-6.
[24] L. Wang, X.-D. Zhang, J.-W. Tang, Z.-W. Ma, M. Usman, Q.-H. Liu, C.-Y. Wu, F. Li, Z.-B. Zhu, B. Gu, Machine learning analysis of SERS fingerprinting for the rapid determination of Mycobacterium tuberculosis infection and drug resistance, Comput. Struct. Biotechnol. J. 20 (2022) 5364–5377, https://doi.org/10.1016/j.csbj.2022.09.031.
[25] M. Usman, J.-W. Tang, F. Li, J.-X. Lai, Q.-H. Liu, W. Liu, L. Wang, Recent advances in surface enhanced Raman spectroscopy for bacterial pathogen identifications, J. Adv. Res. 51 (2023) 91–107, https://doi.org/10.1016/j.jare.2022.11.010.
[26] J.-W. Tang, J.-Q. Li, X.-C. Yin, W.-W. Xu, Y.-C. Pan, Q.-H. Liu, B. Gu, X. Zhang, L. Wang, Rapid discrimination of clinically important pathogens through machine learning analysis of surface enhanced Raman spectra, Front. Microbiol. 13 (2022).
[27] W. Liang, T. Jia-Wei, L. Fen, U. Muhammad, W. Chang-Yu, L. Qing-Hua, K. Hai-Quan, L. Wei, G. Bing, Identification of bacterial pathogens at genus and species levels through combination of Raman spectrometry and deep-learning algorithms, Microbiol. Spectr. 10 (22) (2022) e02580, https://doi.org/10.1128/spectrum.02580-22.
[28] A. Zhu, S. Ali, Y. Xu, Q. Ouyang, Z. Wang, Q. Chen, SERS-based Au@Ag NPs Solid-phase substrate combined with chemometrics for rapid discrimination of multiple foodborne pathogens, Spectrochim. Acta Part A Mol Biomol Spectrosc 270 (2022) 120814, https://doi.org/10.1016/j.saa.2021.120814.
[29] Y. Du, D. Han, S. Liu, X. Sun, B. Ning, T. Han, J. Wang, Z. Gao, Raman spectroscopy-based adversarial network combined with SVM for detection of foodborne pathogenic bacteria, Talanta 237 (2022) 122901, https://doi.org/10.1016/j.talanta.2021.122901.
[30] O.H. Dib, A. Assaf, E. Grangé, J.F. Morin, C.B.Y. Cordella, G. Thouand, Automatic recognition of food bacteria using Raman spectroscopy and chemometrics: a comparative study of multivariate models, Vib. Spectrosc. 126 (2023) 103535, https://doi.org/10.1016/j.vibspec.2023.103535.
[31] K. Mlynáriková, O. Samek, S. Bernatová, F. Růžička, J. Ježek, A. Hároniková, M. Šiler, P. Zemánek, V. Holá, Influence of culture media on microbial fingerprints using Raman spectroscopy, Sensors 15 (2015) 29635–29647, https://doi.org/10.3390/s151129635.
[32] M. Harz, P. Rösch, K.D. Peschke, O. Ronneberger, H. Burkhardt, J. Popp, Micro-Raman spectroscopic identification of bacterial cells of the genus Staphylococcus and dependence on their cultivation conditions, Analyst 130 (2005) 1543–1550, https://doi.org/10.1039/b507715j.

[33] D. Hutsebaut, K. Maquelin, P. De Vos, P. Vandenabeele, L. Moens, G.J. Puppels, Effect of culture conditions on the achievable taxonomic resolution of Raman spectroscopy disclosed by three Bacillus species, Anal. Chem. 76 (2004) 6274–6281, https://doi.org/10.1021/ac049228l.

[34] L. Mikoliunaite, R.D. Rodriguez, E. Sheremet, V. Kolchuzhin, J. Mehner, A. Ramanavicius, D.R.T. Zahn, The substrate matters in the Raman spectroscopy analysis of cells, Sci. Rep. 5 (2015) 1–10, https://doi.org/10.1038/srep13150.

[35] N. Kanno, S. Kato, M. Ohkuma, M. Matsui, W. Iwasaki, S. Shigeto, Machine learning-assisted single-cell Raman fingerprinting for in situ and nondestructive classification of prokaryotes, iScience 24 (2021) 102975, https://doi.org/10.1016/j.isci.2021.102975.

[36] S. Guo, P. Rösch, J. Popp, T. Bocklitz, Modified PCA and PLS: towards a better classification in Raman spectroscopy–based biological applications, J. Chemom. 34 (2020) e3202, https://doi.org/10.1002/cem.3202.

[37] S. Stöckel, S. Meisel, B. Lorenz, S. Kloß, S. Henk, S. Dees, E. Richter, S. Andres, M. Merker, I. Labugger, P. Rösch, J. Popp, Raman spectroscopic identification of Mycobacterium tuberculosis, J. Biophot. 10 (2017) 727–734, https://doi.org/10.1002/jbio.201600174.

[38] H. Shen, P. Rösch, J. Popp, Fiber probe-based Raman spectroscopic identification of pathogenic infection microorganisms on agar plates, Anal. Chem. 94 (2022) 4635–4642, https://doi.org/10.1021/acs.analchem.1c04507.

[39] D. Wang, P. He, Z. Wang, G. Li, N. Majed, A.Z. Gu, Advances in single cell Raman spectroscopy technologies for biological and environmental applications, Curr. Opin. Biotechnol. 64 (2020) 218–229, https://doi.org/10.1016/j.copbio.2020.06.011.

[40] L. Rodriguez, Z. Zhang, D. Wang, Recent advances of Raman spectroscopy for the analysis of bacteria, Anal Sci Adv 4 (2023) 81–95, https://doi.org/10.1002/ansa.202200066.

[41] T.J. Tewes, I. Centeleghe, J.-Y. Maillard, F. Platte, D.P. Bockmühl, Raman microscopic analysis of dry-surface biofilms on clinically relevant materials, Microorganisms 10 (2022).

[42] B.T. Hetjens, T.J. Tewes, F. Platte, F. Wichern, The application of Raman spectroscopy in identifying Metarhizium brunneum, Metarhizium pemphigi and Beauveria bassiana, Biocontrol Sci. Technol. (2021) 1–12, https://doi.org/10.1080/09583157.2021.2007851.

[43] R. Zhang, S. Zhao, Z. Jin, N. Yang, H. Kang, Application of SVM in the food bacteria image recognition and count, in: 2010 3rd International Congress on Image and Signal Processing, 2010, pp. 1819–1823.

[44] U. Schmid, P. Rösch, M. Krause, M. Harz, J. Popp, K. Baumann, Gaussian mixture discriminant analysis for the single-cell differentiation of bacteria using micro-Raman spectroscopy, Chemometr. Intell. Lab. Syst. 96 (2009) 159–171, https://doi.org/10.1016/j.chemolab.2009.01.008.

[45] Y. Liu, J. Xu, Y. Tao, T. Fang, W. Du, A. Ye, Rapid and accurate identification of marine microbes with single-cell Raman spectroscopy, Analyst 145 (2020) 3297–3305, https://doi.org/10.1039/C9AN02069A.

[46] J. Liu, M. Osadchy, L. Ashton, M. Foster, C.J. Solomon, S.J. Gibson, Deep convolutional neural networks for Raman spectrum recognition: a unified solution, Analyst 142 (2017) 4067–4074, https://doi.org/10.1039/C7AN01371J.

[47] N. Tarcea, M. Harz, P. Rösch, T. Frosch, M. Schmitt, H. Thiele, R. Hochleitner, J. Popp, UV Raman spectroscopy—a technique for biological and mineralogical in situ planetary studies, Spectrochim. Acta Part A Mol Biomol Spectrosc 68 (2007) 1029–1035, https://doi.org/10.1016/j.saa.2007.06.051.

[48] L. Zeiri, S. Efrima, Surface-enhanced Raman scattering (SERS) of microorganisms, Isr. J. Chem. 46 (2006) 337–346, https://doi.org/10.1560/IJC_46_3_337.

[49] R.M. Jarvis, N. Law, I.T. Shadi, P. O'Brien, J.R. Lloyd, R. Goodacre, Surface-enhanced Raman scattering from intracellular and extracellular bacterial locations, Anal. Chem. 80 (2008) 6741–6746, https://doi.org/10.1021/ac800838v.

[50] B.N.V. Kumar, B. Kampe, P. Rösch, J. Popp, Characterization of carotenoids in soil bacteria and investigation of their photodegradation by UVA radiation via resonance Raman spectroscopy, Analyst 140 (2015) 4584–4593, https://doi.org/10.1039/C5AN00438A.

[51] J.H. Marshall, G.J. Wilmoth, Pigments of Staphylococcus aureus, a series of triterpenoid carotenoids, J. Bacteriol. 147 (1981) 900–913, https://doi.org/10.1128/jb.147.3.900-913.1981.

[52] N. Roman, H. Sm, A. Trygve, G. Audun, B. Per, B. Trygve, Biosynthetic pathway for γ-cyclic Sarcinaxanthin in micrococcus luteus: heterologous expression and evidence for diverse and multiple catalytic functions of C50 carotenoid cyclases, J. Bacteriol. 192 (2010) 5688–5699, https://doi.org/10.1128/jb.00724-10.

[53] M. del C. Montero-Calasanz, M. Göker, M. Rohde, C. Spröer, P. Schumann, H.-J. Busse, M. Schmid, B.J. Tindall, H.-P. Klenk, M. Camacho, Chryseobacterium hispalense sp. nov., a plant-growth-promoting bacterium isolated from a rainwater pond in an olive plant nursery, and emended descriptions of Chryseobacterium defluvii, Chryseobacterium indologenes, Chryseobacterium wanjuense and Chryseob, Int. J. Syst. Evol. Microbiol. 63 (2013) 4386–4395, https://doi.org/10.1099/ijs.0.052456-0.

[54] R. Kahlon, Pseudomonas Molecular and Applied Biology, NV-. Springer, Switzerland, 2016.

[55] N. Folschweiller, J. Gallay, M. Vincent, M.A. Abdallah, F. Pattus, I.J. Schalk, The interaction between pyoverdin and its outer membrane receptor in Pseudomonas aeruginosa leads to different conformers: a time-resolved fluorescence study, Biochemistry 41 (2002) 14591–14601, https://doi.org/10.1021/bi0259711.

[56] D.F.M. Willemse-Erix, J. Jachtenberg, T. Bakker Schut, W. van Leeuwen, A. van Belkum, G. Puppels, K. Maquelin, Towards Raman-based epidemiological typing of Pseudomonas aeruginosa, J. Biophot. 3 (2010) 506–511, https://doi.org/10.1002/jbio.201000026.

[57] L. Bednárová, J. Palacký, V. Bauerová, O. Hrušková-Heidingsfeldová, I. Pichová, P. Mojzeš, Raman microspectroscopy of the yeast vacuoles, Spectros. Int. J. 27 (2012) 746597, https://doi.org/10.1155/2012/746597.

[58] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J Big Data 8 (2021) 53, https://doi.org/10.1186/s40537-021-00444-8.

[59] C. Krafft, J. Popp, The many facets of Raman spectroscopy for biomedical analysis, Anal. Bioanal. Chem. 407 (2015) 699–717, https://doi.org/10.1007/s00216-014-8311-9.

[60] B. Mirvaux, N. Recham, J. Miot, M. Courty, S. Bernard, O. Beyssac, C. Davoisne, M. Sougrati, A. Demortière, F. Guyot, J.M. Tarascon, D. Larcher, Iron phosphate/bacteria composites as precursors for textured electrode materials with enhanced electrochemical properties, J. Electrochem. Soc. 163 (2016) A2139, https://doi.org/10.1149/2.0101610jes.

[61] M. Kakita, M. Okuno, H. Hamaguchi, Quantitative analysis of the redox states of cytochromes in a living L929 (NCTC) cell by resonance Raman microspectroscopy, J. Biophot. 6 (2013) 256–259, https://doi.org/10.1002/jbio.201200042.

[62] G.J. Puppels, J.H.F. Olminkhof, G.M.J. Segers-Nolten, C. Otto, F.F.M. de Mul, J. Greve, Laser irradiation and Raman spectroscopy of single living cells and chromosomes: sample degradation occurs with 514.5 nm but not with 660 nm laser light, Exp. Cell Res. 195 (1991) 361–367, https://doi.org/10.1016/0014-4827(91)90385-8.

[63] I. Notingher, S. Verrier, H. Romanska, A.E. Bishop, J.M. Polak, L.L. Hench, *In situ* characterisation of living cells by Raman spectroscopy, Spectroscopy 16 (2002) 408381, https://doi.org/10.1155/2002/408381.

[64] D. Naumann, S. Keller, D. Helm, C. Schultz, B. Schrader, FT-IR spectroscopy and FT-Raman spectroscopy are powerful analytical tools for the non-invasive characterization of intact microbial cells, J. Mol. Struct. 347 (1995) 399–405, https://doi.org/10.1016/0022-2860(95)08562-A.

[65] R. Prucek, V. Ranc, L. Kvítek, A. Panáček, R. Zbořil, M. Kolář, Reproducible discrimination between Gram-positive and Gram-negative bacteria using surface enhanced Raman spectroscopy with infrared excitation, Analyst 137 (2012) 2866–2870, https://doi.org/10.1039/C2AN16310A.

[66] K. Kairyte, Z. Luksiene, M. Pucetaite, V. Sablinskas, Differentiation of bacterial strains by means of surface enhanced FT-Raman spectroscopy, Lith J Phys Tech Sci 52 (2012) 276–283, https://doi.org/10.3952/lithjphys.52309.

[67] P. Segers, M. Vancanneyt, B. Pot, U. Torck, B. Hoste, D. Dewettinck, E. Falsen, K. Kersters, P. De Vos, Classification of Pseudomonas diminuta Leifson and Hugh 1954 and Pseudomonas vesicularis Büsing, Döll, and Freytag 1953 in Brevundimonas gen. nov. as Brevundimonas diminuta comb. nov. and Brevundimonas vesicularis comb, nov., Respectively, Int. J. Syst. Evol. Microbiol. 44 (1994) 499–510, https://doi.org/10.1099/00207713-44-3-499.

[68] A.B. Boehm, L.M. Sassoubre, Enterococci as Indicators of Environmental Fecal Contamination, 2014.

[69] M. Suchomel, A. Lenhardt, G. Kampf, A. Grisold, Enterococcus hirae, Enterococcus faecium and Enterococcus faecalis show different sensitivities to typical biocidal agents used for disinfection, J. Hosp. Infect. 103 (2019) 435–440, https://doi.org/10.1016/j.jhin.2019.08.014.

[70] M. Otto, Staphylococcus epidermidis — the "accidental" pathogen, Nat. Rev. Microbiol. 7 (2009) 555–567, https://doi.org/10.1038/nrmicro2182.

[71] A. Cross, J.R. Allen, J. Burke, G. Ducel, A. Harris, J. John, D. Johnson, M. Lew, B. MacMillan, P. Meers, R. Skalova, R. Wenzel, J. Tenney, Nosocomial infections due to Pseudomonas aeruginosa: review of recent trends, Rev. Infect. Dis. 5 (1983) S837–S845, https://doi.org/10.1093/clinids/5.Supplement_5.S837.
[72] K. Rebrošová, M. Šiler, O. Samek, F. Růžička, S. Bernatová, J. Ježek, P. Zemánek, V. Holá, Differentiation between Staphylococcus aureus and Staphylococcus epidermidis strains using Raman spectroscopy, Future Microbiol. 12 (2017) 881–890, https://doi.org/10.2217/fmb-2016-0224.
[73] Z. Al-Shaebi, F. Uysal Ciloglu, M. Nasser, O. Aydin, Highly accurate identification of bacteria's antibiotic resistance based on Raman spectroscopy and U-net deep learning algorithms, ACS Omega 7 (2022) 29443–29451, https://doi.org/10.1021/acsomega.2c03856.
[74] A. Wieser, L. Schneider, J. Jung, S. Schubert, MALDI-TOF MS in microbiological diagnostics—identification of microorganisms and beyond (mini review), Appl. Microbiol. Biotechnol. 93 (2012) 965–974, https://doi.org/10.1007/s00253-011-3783-4.
[75] D. Storozhuk, O. Ryabchykov, J. Popp, T. Bocklitz, RAMANMETRIX: a delightful way to analyze Raman spectra, arXiv Prepr arXiv220107586 (2022).
[76] S. Guo, O. Ryabchykov, N. Ali, R. Houhou, T. Bocklitz, 4.13 - comprehensive chemometrics, in: S. Brown, R. Tauler (Eds.), Walczak BBT-CC, Second E, Elsevier, Oxford, 2020, pp. 333–359.