ACS Partner Journal

CHEMICAL & BIOMEDICAL IMAGING

Open Access

pubs.acs.org/ChemBioImaging

Article

Enhancing Open-World Bacterial Raman Spectra Identification by Feature Regularization for Improved Resilience against Unknown Classes

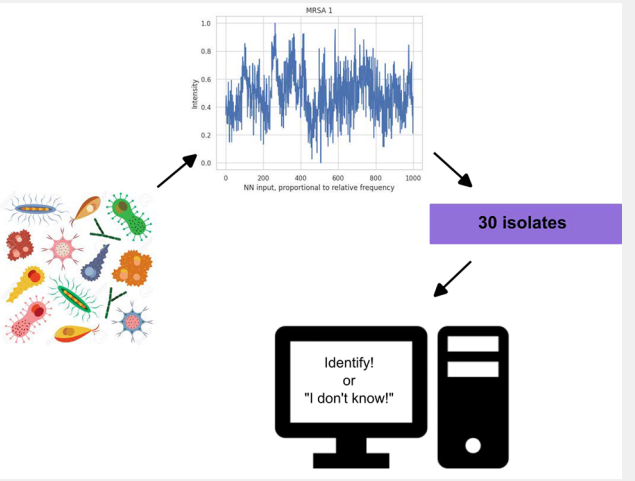
Yaroslav Balytskyi,* Nataliia Kalashnyk, Inna Hubenko, Alina Balytska, and Kelly McNear

Cite This: *Chem. Biomed. Imaging* 2024, 2, 442−452

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The combination of deep learning techniques and Raman spectroscopy shows great potential offering precise and prompt identification of pathogenic bacteria in clinical settings. However, the traditional closed-set classification approaches assume that all test samples belong to one of the known pathogens, and their applicability is limited since the clinical environment is inherently unpredictable and dynamic, unknown, or emerging pathogens may not be included in the available catalogs. We demonstrate that the current state-of-the-art neural networks identifying pathogens through Raman spectra are vulnerable to unknown inputs, resulting in an uncontrollable false positive rate. To address this issue, first we developed an ensemble of ResNet architectures combined with the attention mechanism that achieves a 30-isolate accuracy of 87.8 ± 0.1%. Second, through the integration of feature regularization by the Objectosphere loss function, our model both achieves high accuracy in identifying known pathogens from the catalog and effectively separates unknown samples drastically reducing the false positive rate. Finally, the proposed feature regularization method during training significantly enhances the performance of out-of-distribution detectors during the inference phase improving the reliability of the detection of unknown classes. Our algorithm for Raman spectroscopy empowers the identification of previously unknown, uncataloged, and emerging pathogens ensuring adaptability to future pathogens that may surface. Moreover, it can be extended to enhance open-set medical image classification, bolstering its reliability in dynamic operational settings.



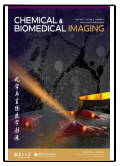
KEYWORDS: *Raman spectroscopy, machine learning, ResNet, pathogen identification, Open-Set learning, Objectosphere*

I. INTRODUCTION AND PROBLEM STATEMENT

Raman spectroscopy involves the scattering of light and its interaction with the chemical bonds present in the material under investigation. This interaction produces a unique spectrum, akin to a fingerprint, that characterizes the material's chemical composition and molecular structure.[1] It was independently discovered in 1928 by Raman[2] and Landsberg,[3] and the appearance of laser spectrometers[4,5] further expanded its capabilities and applications. Raman spectroscopy is a reliable, sensitive, nondestructive, and versatile analytical technique to determine the chemical composition and molecular structure of complex substances,[6] where it is already used in a number of applications,[7] while its portability makes it valuable for both laboratory and field applications.[1] In addition, its unique properties make it a promising tool for biomedical science,[8] including disease diagnosis.[9−12]

One of the crucial applications of Raman spectroscopy is the identification of bacterial infections, which are responsible for approximately 7 million deaths worldwide each year.[13,14] While there are effective methods for detecting pathogenic bacteria, such as enzyme-linked immunosorbent assay (ELISA), polymerase chain reaction (PCR), and sequencing-based approaches, these methods often involve significant time requirements to produce results.[15−18] Even methods that use clinical tools such as BioFire are limited by the testing panel, such as the BCID2, which identifies 33 species.[19] Furthermore, clinical diagnostic procedures for identifying specific pathogens often involve time-consuming microbiological culture (up to

Received: January 22, 2024
Revised: April 21, 2024
Accepted: April 23, 2024
Published: May 6, 2024



48 h) and antibiotic susceptibility testing (up to 24 h).[20,21] During this waiting period, broad-spectrum antibiotics (BSAbx) are commonly prescribed as a precautionary measure.[20,21] However, it is important to note that while BSAbx can be life-saving, they should be used judiciously due to their potential side effects and contribution to antibiotic resistance. Excessive use of BSAbx can disrupt the healthy gut microbiome, leading to the overgrowth of pathogens like *Candida albicans* and *Clostridium difficile*.[20,21] Disturbingly, the Centers for Disease Control and Prevention (CDC) has reported that over 30% of patients are treated with antibiotics unnecessarily.[22] The delay in accurately detecting pathogens leads to extended hospital stays, escalated medical expenses, heightened antibiotic resistance, and ultimately increased mortality rates.[23] To address this issue, Raman spectroscopy offers immense potential as a highly sensitive, culture-free, cost-effective, and rapid identification method. By employing Raman spectroscopy, targeted antibiotics can be administered, thus mitigating the development of antimicrobial resistance.[24] This approach allows for timely and effective treatment decisions, minimizing the negative impacts associated with delayed pathogen identification.

The application of Raman spectroscopy extends beyond the identification of pathogenic bacteria and encompasses diverse areas such as the diagnosis of COVID-19,[25] food safety,[26] identification of contaminants in pharmaceuticals,[27] and homeland security.[28−30] However, in this Article, our primary focus is on the identification of pathogenic bacteria using the bacteria-ID data set.[24]

To extract meaningful information from Raman spectra, data analysis and processing are necessary. While manual approaches, such as the "Ramanome" concept utilizing 31 specific Raman peaks, have been employed,[31,32] they are not sufficiently reliable. This is because spectral information encompasses more complex characteristics beyond these 31 peaks, and interclass differences pose challenges for manual classification.[33,34] Moreover, due to the low probability of Raman scattering, meaningful spectral information can be easily obscured by background noise.[35] Additionally, the large volume of spectral data can be challenging to handle in practical applications, necessitating reliable and efficient quantitative methods facilitated by machine learning-driven tools.[36]

We follow the definition of machine Learning (ML) by Francois Chollet as "the effort to automate intellectual tasks normally performed by humans".[37] The ML framework aims to find a suitable representation of the data, allowing classification rules to be automatically derived rather than hard-coded. "Deep learning (DL) is a specific subfield of machine learning: a new take on learning representations from data that puts an emphasis on learning successive layers of increasingly meaningful representations".[37] In our paper, the above-mentioned layers of data representations are implemented using deep neural networks (DNNs). According to ref 37, unlike DL, shallow learning approaches use only one or two consecutive data representation layers.

Shallow learning models, in particular, principal component analysis (PCA) combined with linear discriminant analysis (LDA), are often used to analyze the Raman spectra.[38−42] However, these methods only work in a "closed-world" environment, and our previous work, discussed later in the text, has shown improvements. DL models have also been successfully applied to classify molecular spectra[43−48] and have shown better performance compared to shallow ones.[49] The vanishing gradient problem[50] prevents a further boost of the model's performance by a naive approach of adding extra layers. ResNet architecture[51] fixes this problem by introducing skip connections. ResNet and its modifications have been successfully applied to classify Raman spectra, outperforming shallow models by a large margin, as shown by other authors[24,52,53] and in our previous work.[54]

While DNNs excel at identifying classes encountered during the training phase, their behavior becomes unpredictable when confronted with spectra belonging to unknown classes that were not part of the training data, known as out-of-distribution (OOD) samples.[55,56] Typically, the SoftMax layer[57] is used to interpret DNN outputs as probabilities, and the classification result corresponds to the output with the highest SoftMax score. However, as outlined in ref 58, even a slight difference in logit values between the winning and runner-up classes can lead to vastly different probability values from the SoftMax layer. Moreover, the SoftMax procedure involves logit value normalization, rendering it inherently closed-world[55] and thus unable to reliably identify OOD samples. Consequently, DNNs often produce incorrect and overly confident predictions when faced with OOD samples. For example, as shown in refs 59 and 60, DNNs encounter "foolish" and "rubbish" images visually far from the class from the training catalog but still produce high confidence scores. Another example is the incorrect and confident classification of a *crab* image as the *clapping* class, even though no crab-related items were present during training.[61] This necessitates the use of specialized ML techniques capable of identifying OOD samples, as the false positive (FP) rate estimated on large-scale data sets exceeded 70% and, in some cases, was close to 100%.[62,63]

The biggest concern in terms of clinical use is that a classifier trained on known species of bacteria would identify a new type of bacteria as belonging to a known class with high confidence.[64] This issue is challenging to mitigate in practice, as it is difficult to anticipate and account for all the potential classes that a classifier might encounter in an unpredictable environment. Some ML systems have been developed to handle this problem, particularly in areas like medical image classification,[65] safety-critical applications,[66] and environmental monitoring.[67] To tackle this challenge, in our research, we introduce new features to our previous ML algorithm in ref 54 which are designed to accurately identify pathogenic bacteria using Raman spectroscopy in real-world scenarios. This improved algorithm can now handle noisy data while maintaining relatively high accuracy. It not only successfully classifies the pathogens already listed in its catalog but also reliably distinguishes and rejects pathogens that are not included in the catalog thereby enhancing patient care and treatment outcomes.

Our Article is structured as follows. In Section II, we describe the data set and its division into in-catalog and out-of-catalog samples. In Section III, we present our custom ResNet architecture that leverages the strength of the attention mechanism to achieve enhanced performance compared to existing DNN architectures in closed-world scenarios. Furthermore, we highlight the limitations associated with typical closed-world approaches. In Section IV, we combine our backbone ResNet architecture with Entropic Open-Set and Objectosphere loss functions, demonstrating drastic improvement over the naive closed-world approaches in handling the

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452

Table 1. List of Pathogen Names and Their Division into Categories $\mathcal{K}$, $\mathcal{I}$, and $\mathcal{N}$

class	pathogen name
p_1, used as $\mathcal{K}$	Group A Strep, Group B Strep, Group C Strep, Group G Strep, *Escherichia coli* 1, *E. coli* 2, *Enterobacter cloacae*, *Proteus mirabilis*, *Serratia marcescens*, *Candida albicans*
p_2, used as $\mathcal{K}/\mathcal{I}$	*Enterococcus faecalis* 1, *E. faecalis* 2, *Enterococcus faecium*, *Pseudomonas aeruginosa* 1, *P. aeruginosa* 2
p_3, used as $\mathcal{I}$	*Staphylococcus epidermidis*, *Staphylococcus lugdunensis*, *Streptococcus sanguinis*, *Klebsiella aerogenes*, *Candida glabrata*
p_4, used as $\mathcal{N}$	MRSA 1, MRSA 2, MSSA 1, MSSA 2, MSSA 3, *Streptococcus pneumoniae* 1, *S. pneumoniae* 2, *Klebsiella pneumoniae* 1, *K. pneumoniae* 2, *Salmonella enterica*

unknowns. To minimize the occurrence of inconclusive outcomes for in-catalog samples, we augment our combined deep neural network with the one-vs-rest classifier in Section V. In Section VI, we demonstrate a substantial enhancement in the performance of novelty detectors following the implementation of our proposed feature regularization, as opposed to using naive approaches. Therefore, we demonstrate that our integrated DNN architecture outperforms currently available techniques for both closed- and open-world applications. We present our conclusions in Section VII. Our previous proof-of-concept work was published in ref 54.

II. OVERVIEW OF OPEN-SET LEARNING STRATEGY AND SPLITTING THE DATA INTO $\mathcal{K}/\mathcal{I}/\mathcal{N}$ CATEGORIES

In general, current Open-Set learning approaches belong to two broad classes: generative and discriminative.[55] Generative methods model the joint distribution of input features and labels to estimate the probability that a given sample is OOD, while discriminative methods directly learn the decision boundary between classes to classify the input data based on their features. However, generative methods have been shown to be less efficient than discriminative methods with a well-chosen background class for any but simple data sets,[68,69] so we will focus on discriminative methods in our further considerations.

In this discriminative modeling, we classify the data into three categories: $\mathcal{K}$, $\mathcal{I}$, and $\mathcal{N}$, similarly to our previous work.[54] $\mathcal{K}$ corresponds to the known category, the classes of interest that the DNN prioritizes to identify. $\mathcal{I}$ consists of classes belonging to the background category that the DNN "ignores" in order to identify $\mathcal{K}$ with greater confidence in a procedure called feature regularization. Finally, $\mathcal{N}$ corresponds to samples corresponding to classes not seen during the training phase of the DNN and which the DNN seeks to distinguish from the $\mathcal{K}$ classes. Only $\mathcal{K}$ and $\mathcal{I}$ are used during the training of the DNN, and the DNN is completely unaware of $\mathcal{N}$ until the testing stage.

For our purposes, we use the bacteria-ID data set,[24] which contains 30 pathogen classes shown in Table 1, with 2000 spectra per class for training, 100 spectra per class for fine-tuning, and 100 spectra per class for testing. To test our ML algorithm in open-world learning settings, we split our data set into four parts, p_1, p_2, p_3, and p_4. We assign the pathogen group p_1 to $\mathcal{K}$ or "known classes of interest" since those are extremely common and contagious. Antibiotic-resistant or susceptible pathogens corresponding to the p_4 group are particularly harmful to patients and a burden on healthcare systems. Misclassification of these pathogens is extremely problematic, especially if any errors are made in the classification between a susceptible strain of the pathogen and a resistant strain (such as MSSA vs MRSA). Therefore, we classified the p_4 pathogens group as $\mathcal{N}$, to highlight the ability of our algorithm to identify "never before seen" samples while keeping high accuracy on the known ones. The p_2 and p_3 groups are often antibiotics susceptible but typically found in the body. In our experiments, we tested both options, assigning them as both $\mathcal{K}$ and $\mathcal{I}$ in different runs of the experiments for this proof of concept work.

However, as mentioned before, it is necessary to carefully assign the pathogen classes to the background category $\mathcal{I}$ in order for the DNN to be efficient. Note that p_1 and p_2 are closer in their characteristics than p_1 and p_3. Both groups p_1 and p_2 consist mainly of streptococcal species and have streptococcal species associated with respiratory and invasive infections. Although groups p_1 and p_3 also share some common features, such as the presence of Staphylococcus species, in general, p_1 and p_2 are much closer to each other in their characteristics. Due to the low signal-to-noise ratio of this data set, it is necessary to keep $\mathcal{K}$ and $\mathcal{I}$ *sufficiently distinct* from each other to avoid false positives or misclassifications.

As we will later demonstrate, if this condition is not met, the DNN is forced to focus on highly similar samples and "ignore" them simultaneously, which is an inconsistent task and leads to significant performance degradation. The case when $\mathcal{K} = p_1$ and $\mathcal{I} = p_2$ has significantly worse performance compared to all other data partitions, which highlights the importance of choosing $\mathcal{I}$ correctly. It is much more efficient to assign $\mathcal{K} = p_1$ and $\mathcal{I} = p_3$ or $\mathcal{K} = p_1 + p_2$ and $\mathcal{I} = p_3$ as we show further in the text.

III. BACKBONE NEURAL NETWORK ARCHITECTURE AND LIMITATIONS OF CLOSED-WORLD APPROACHES

We construct our DNN architecture based on custom ResNet architecture, similar to our previous work.[54] However, we implement several adjustments to mitigate spectral noise and effectively address the inherent low signal-to-noise ratio present in the bacteria-ID data set. Through experiments, we found that employing the mainstream SE-ResNet architecture,[70] which incorporates the ResNet architecture alongside the squeeze-and-excitation (SE) attention mechanism in all residual connection layers, did not result in improved accuracy compared to a standard ResNet architecture for this data set. Instead, it resulted in overfitting, prompting us to only augment the last residual block of our custom ResNet architecture with the SE attention mechanism. Surprisingly, this adjustment led to significantly improved performance compared to both SE-ResNet and standard ResNet architectures. Both during training and fine-tuning, we utilized the Adam optimizer provided by TensorFlow,[71] with a batch size of 32 and a validation set comprising 20% of the total data set. However, for training, we set the learning rate to 10^{-5}, whereas for fine-tuning, we employed a smaller learning rate of 10^{-6}. In

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452

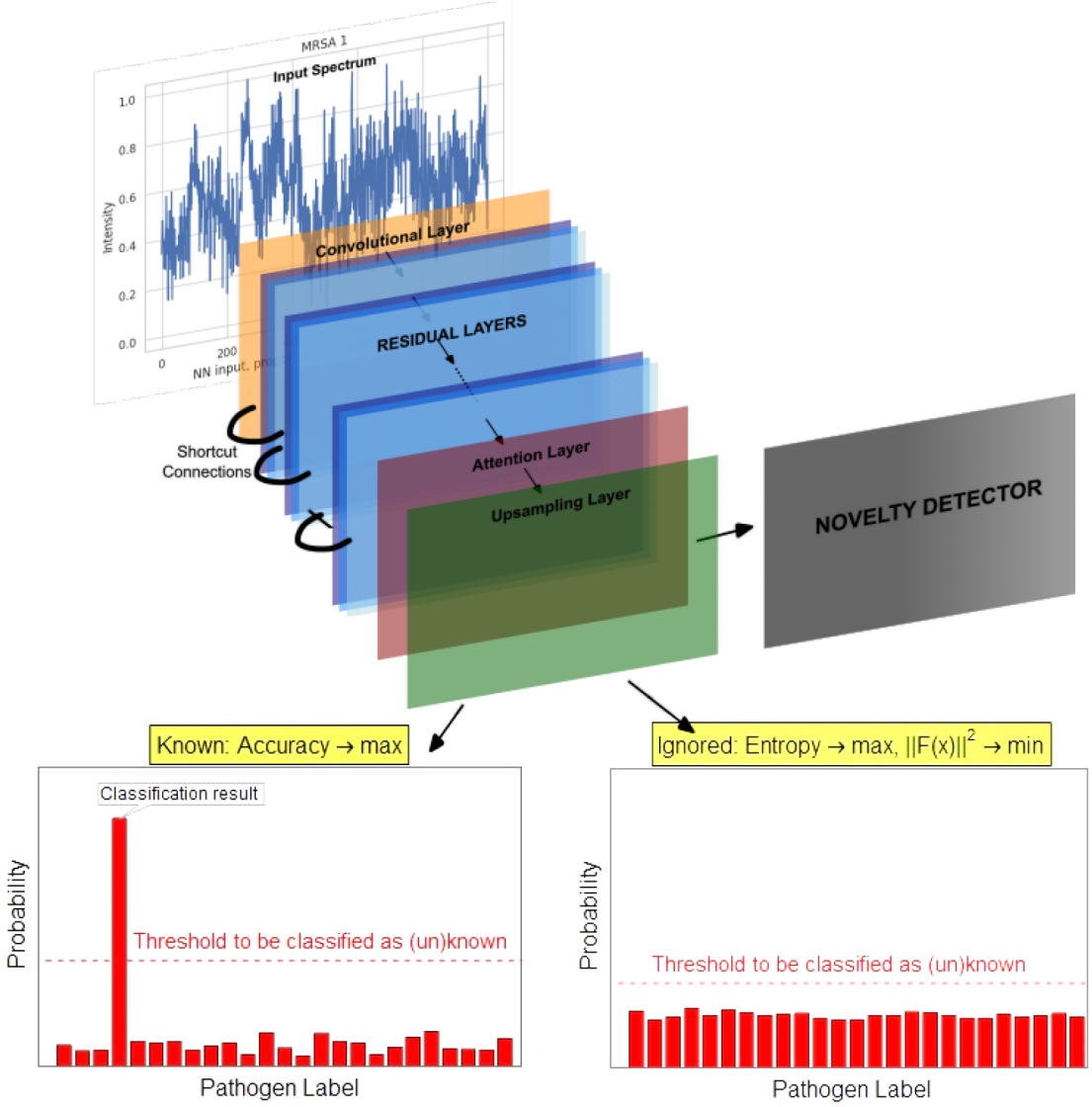


Figure 1. Schematic representation of our custom ResNet architecture. The DNN processes input pathogen Raman spectra into classification probabilities while simultaneously detecting OOD samples. Section IV describes feature regularization using the $\mathcal{I}$ class, while an additional OOD detector is implemented in Section VI. For a pathogen to be classified as belonging to the known catalog, it must exceed a threshold and pass the novelty detector test.

addition, during the fine-tuning stage, we only fine-tuned the last three layers of our custom DNN to mitigate the risk of overfitting. Importantly, a similar architecture utilizing only one small SE-ResNet module was demonstrated to be highly efficient for breast cancer histopathological image classification.[72] Overall, our experiments revealed that integrating a single attention block enhances the performance of ResNet. However, excessive attention within the DNN can result in overfitting and negatively affect performance. A schematic representation of our DNN is shown in Figure 1. Finally, similarly to our previous work, we assembled our ResNets into an ensemble:

$$\text{Prediction Ensemble} = \frac{1}{5}\sum_{i=1}^{5}\text{Prediction}[i] \tag{1}$$

To demonstrate the stability of our model's performance, first we conducted 20 runs of our model and assessed the accuracy of a single run using all 30 pathogen classes, $\mathcal{K} = p_1 + p_2 + p_3 + p_4$, $\mathcal{I} = \varnothing$. We subsequently grouped these 20 models into 4 ensembles, each consisting of 5 models. As illustrated in Table 2, the average 30-isolate accuracy of an individual model run stands at 87.5% ± 0.4%, whereas the accuracy of the ensemble is 87.8% ± 0.1%. Thus, using a model ensemble results in a marked increase in accuracy and a reduction in variance. The corresponding correlation table is shown in Figure 2. The architecture we propose surpasses the

Table 2. Single Runs and Ensemble Accuracies on All 30 Classes, $\mathcal{K} = p_1 + p_2 + p_3 + p_4$, $\mathcal{I} = \varnothing$

run no.	accuracy of a run	ensemble accuracy
1–5	87.6%, 87.2%, 88.2%, 87.7%, 87.4%	87.8%
6–10	88.3%, 87.9%, 87.4%, 87.8%, 87.7%	87.9%
11–15	87.7%, 87.5%, 86.9%, 87.3%, 87.1%	87.9%
16–20	86.7%, 87.5%, 87.2%, 87.4%, 87.1%	87.6%
accuracy	87.5 ± 0.4%	87.8 ± 0.1%

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452

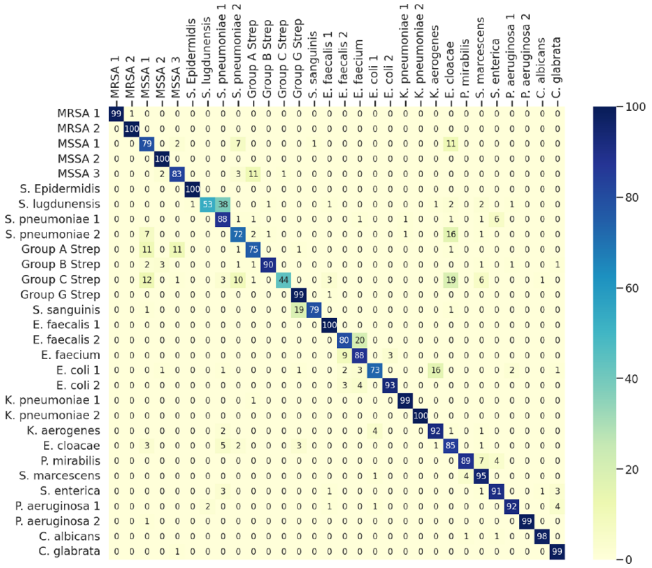


Figure 2. Correlation table for all 30 pathogens in "closed-world" settings when all pathogen classes are known, $\mathcal{K} = p_1 + p_2 + p_3 + p_4$; $\mathcal{I}$, $\mathcal{N} = \varnothing$. Average accuracy = 87.8 ± 0.1%.

current state-of-the-art closed-world DNNs in terms of 30-isolate accuracy on the bacteria-ID data set. Existing models achieve accuracies of 82.2 ± 0.3%,[24] 84.7 ± 0.3%,[53] 86.3%,[73] and 86.7 ± 0.4%,[52] respectively. Meanwhile, our model achieves an 8-treatment accuracy of 97.7% ± 0.2%, which is comparable to the available models.[24,52,53,73] However, in scenarios where more targeted antibiotics are used, our proposed DNN architecture may outperform its counterparts due to its higher 30-isolate accuracy. Therefore, we not only utilize our model in the closed-world setting but also adapt it for the open-world settings considered in the following Sections, leveraging its remarkable performance.

Before we get into the open-world settings, we need to establish the required definitions. The input of the DNN x corresponds to the intensity values of the input Raman scattering spectra, and the corresponding output represents the probability of the spectrum belonging to a specific class of pathogens p, given by the logit values $l_p(x)$. The logit values are obtained by multiplying the output from the second to last layer of the DNN, called the deep features $F(x)$, by the weights W, and the probability of a spectrum belonging to a particular class of pathogens p is obtained from "softmaxing" procedure, defined as

$$l_p(x) = W \cdot F(x),\ S_p(x) = \frac{e^{l_p(x)}}{\sum_p e^{l_p(x)}} \tag{2}$$

The resulting value is in the interval $S_p(x) \in [0, 1]$ with $\sum_p S_p(x) = 1$, and thus $S_p(x)$ can be interpreted as a probability measure.

In the case where the input belongs to the category $\mathcal{K}$, it is classified as the pathogen that has the highest softmax score in eq 2. In the perfect case scenario, the DNN's output from the sample belonging to ith class in the pathogen catalog should return the following:

$$(\text{Pathogen class} \in \mathcal{K})[i] \rightarrow \underbrace{[0, ..., \underbrace{1}_{i\text{th position}}, ..., 0]}_{\text{Length}=|\mathcal{K}|} \tag{3}$$

where $|\mathcal{K}|$ represents the number of pathogen classes in the $\mathcal{K}$ catalog.

When unknowns are present in the test set, it may be tempting to allocate a dedicated output node for the "unknown" category. However, as demonstrated by previous research in ref 74, this approach proves ineffective except for simple academic data sets. Similarly, our prior investigation in ref 54 revealed that such an approach is ill-suited for Raman spectroscopy applications in open-world scenarios. In essence, since there may be multiple spectra associated with the unknown category, assigning a single node to all of them is inefficient.

A more efficient strategy for managing the $\mathcal{N}$ category involves thresholding the softmax score.[58,75,76] This technique operates under the assumption that there is ample distinction between the categories $\mathcal{K}$ and $\mathcal{N}$ within the feature space, resulting in the DNN's output on $\mathcal{N}$ approaching the following:

$$\mathcal{N} \rightarrow \underbrace{\left[\frac{1}{|\mathcal{K}|}, ..., \frac{1}{|\mathcal{K}|}\right]}_{\text{Length}=|\mathcal{K}|} \tag{4}$$

with Shannon entropy[77] reaching its maximum value $\log_2(|\mathcal{K}|)$. Therefore, there is no dedicated node for the unknown category further in the text, since it is assumed that $\mathcal{N}$ is distributed among the nodes that correspond to the $\mathcal{K}$ category.

In the case in which the condition in eq 4 is fulfilled, it is possible to introduce the threshold Λ such that $\mathcal{K}$ and $\mathcal{N}$ are separated as $\max(\mathcal{N}) < \Lambda$ while $\max(\mathcal{K}) > \Lambda$. In practice, if the maximum value of the softmax score is less than Λ, this is classified as an inconclusive "I don't know what it is" result, which may mean that the sample belongs to category $\mathcal{N}$ outside the $\mathcal{K}$ catalog. Another possibility is that the sample belongs to $\mathcal{K}$ but with low confidence. Thus, our goal is to separate samples outside the catalog $\mathcal{N}$ while minimizing the number of inconclusive results for samples in the catalog.

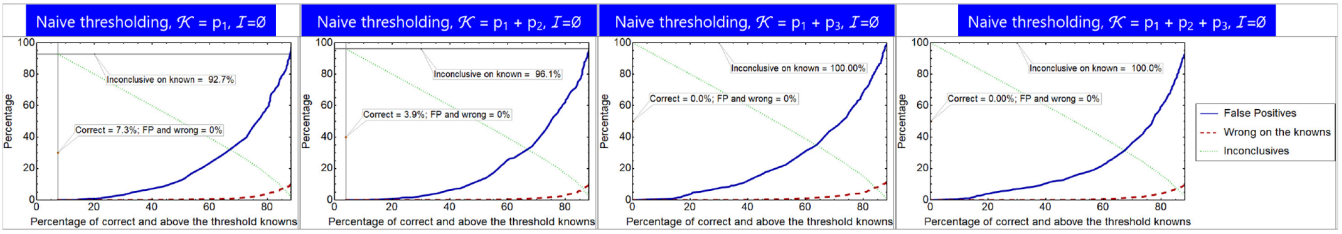


Figure 3. False positive rate on $\mathcal{N}$, error, and inconclusive rates on $\mathcal{K}$ as a function of correct classification rate for naive approaches.

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452

Since the correct classification, error, and inconclusive rates on $\mathcal{K}$ as well as the FP rate on $\mathcal{N}$ are all functions of the global threshold Λ, it is convenient to represent the FP, error, and inconclusive rates as a function of the correct classification rate, and we plot the results corresponding to the naive threshold approaches in Figure 3. A striking feature can be observed: the FP rate for unknowns is much higher than the error probability for knowns and can be close to 100%, so special methods are needed to solve this problem, which will be implemented in the following Sections. In practice, as Figure 3 demonstrates, $\mathcal{K}$ and $\mathcal{N}$ are not sufficiently separated, resulting in a high FP rate, and therefore the assumption in eq 4 is false. In the next Section, we show that by introducing an additional "ignored" category $\mathcal{I}$ and Open-Set methods, this problem can be mitigated, as illustrated in Figure 4. The solid lines

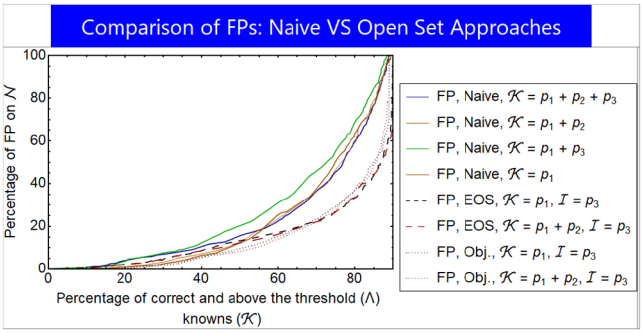


Figure 4. Comparison of false positive rates by naive vs Open-Set methods.

corresponding to the FP rates of the naive approaches are much higher than the dotted and dashed lines corresponding to the Open-Set methods labeled "EOS" and "Obj." and discussed further.

IV. FEATURE REGULARIZATION BY ENTROPIC OPEN-SET AND OBJECTOSPHERE METHODS

In order to improve separation between $\mathcal{K}$ and $\mathcal{N}$, we introduce the "ignored" category, $\mathcal{I}$. This approach was originally proposed in ref 74 and has been shown to be highly efficient for open-world Raman spectroscopy purposes.[54]

This method consists of modifying the loss function during the training and consists of two parts. First, the Entropic Open-Set (EOS) loss function[74] is defined as

$$V_E(x) = \begin{cases} -\log(S_p(x)), \text{ if } x \in \mathcal{K} \\ -\frac{1}{|\mathcal{K}|}\sum_{p=1}^{|\mathcal{K}|}\log(S_p(x)), \text{ if } x \in \mathcal{I} \end{cases} \quad (5)$$

Thus, for category $\mathcal{K}$, it reduces to the usual categorical cross-entropy loss function, while for the case where $x \in \mathcal{I}$, $V_E(x)$ aims to maximize the Shannon entropy and uniformly distribute the output of the DNN over the knowns.

Second, in general, classes belonging to $\mathcal{K}$ tend to have higher absolute values of deep features $\|F(x)\|$ than classes belonging to $\mathcal{N}$. Thus, the Objectosphere loss function aims to increase this separation by using the deep feature $F(x)$ parameter as

$$V_O(x) = V_E(x) + \alpha\begin{cases} \max(\beta - \|F(x)\|^2, 0), \text{ if } x \in \mathcal{K} \\ \|F(x)\|^2, \text{ if } x \in \mathcal{I} \end{cases} \quad (6)$$

where the values of α and β are adjusted to minimize the inconclusive rate on the $\mathcal{K}$ category by the model cross-validation, and $\|\cdot\|$ is a regular Euclidean norm. This leads to a minimization of the FP rate on $\mathcal{I}$, and this property generalizes to the $\mathcal{N}$ category, even though DNN is unaware of $\mathcal{N}$ until the testing phase.

The corresponding FP rates on $\mathcal{N}$, error, and inconclusive rates on $\mathcal{K}$ as a function of correct classification rate for the Entropic-Open Set (EOS) and Objectosphere (Obj.) approaches with different choices of $\mathcal{K}$ and $\mathcal{I}$ are shown in Figure 6. As mentioned earlier in Section II, it is crucial to choose the right data set for the $\mathcal{I}$ category. One can increase the global threshold Λ and reduce the FP frequency as well as the error rate to zero at the cost of increasing the frequency of inconclusive results on $\mathcal{K}$. The FP rates of all Open-Set learning experiments are plotted in Figure 5, and one can observe a noticeably higher FP rate when $\mathcal{I}$ includes p_2, showed by solid lines and marked by the $\surd$ sign.

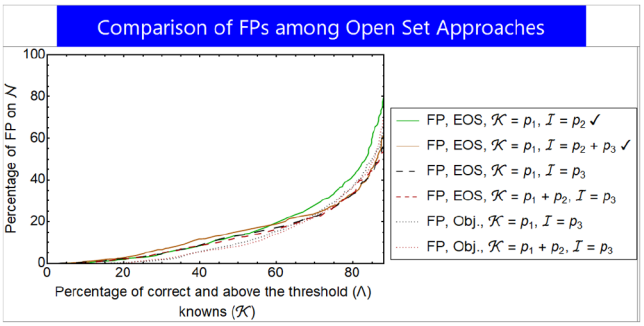


Figure 5. Comparison of FP rates of all Entropic Open-Set (EOS) and Objectosphere (Obj.) experiments. A noticeably higher FP rate when the $\mathcal{I}$ data set includes p_2 can be observed, shown by the $\surd$ mark.

Although the Open-Set learning methods described above significantly improve the DNN's performance in the open world as shown in Figure 4, there are a significant number of inconclusive results on $\mathcal{K}$. For example, as shown in Figure 6, if the threshold is increased so that both the FP and the error rates are zero, the highest rate of conclusive result achieved is around 18%. Therefore, in the next Section, we combine the Open-Set approaches implemented here with a one-versus-the-rest classifier to increase the number of conclusive results on $\mathcal{K}$ while keeping the FP and error rate zero.

V. COMBINATION WITH THE ONE-VS-REST CLASSIFIER

To reduce the number of inconclusive results, instead of the global (Λ), we introduce the per-class threshold Λ' to classify the DNN's output $S(x)$:

$$S(x) = \underbrace{[s_1, \ldots, s_{|\mathcal{K}|}]}_{\text{Length}=|\mathcal{K}|}, \Lambda = \underbrace{[\lambda_1, \ldots, \underbrace{\lambda_i}_{i\text{th position}}, \ldots, \lambda_{|\mathcal{K}|}]}_{\text{Length}=|\mathcal{K}|}$$

If the maximum value of the output exceeds the threshold value for the class, $\max_{k\in\mathcal{K}}(s_k) = s_i > \lambda_i$, the spectrum is classified

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452

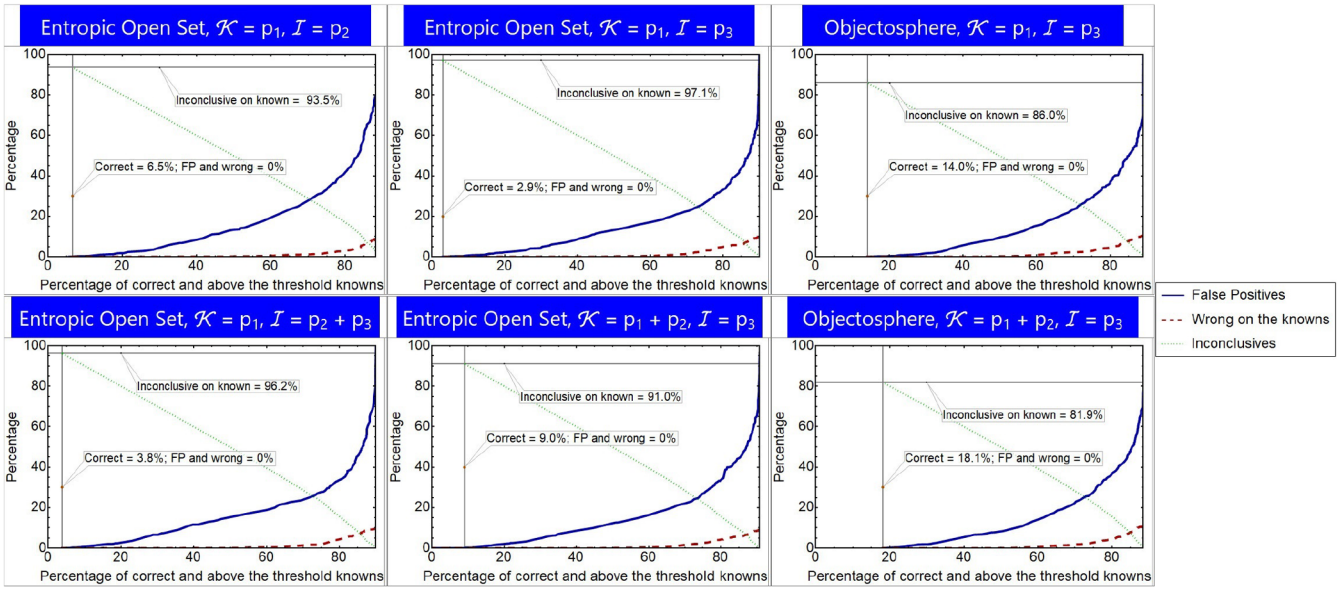


Figure 6. FP rate on $\mathcal{N}$, error, and inconclusive rates on $\mathcal{K}$ as a function of correct classification rate for Entropic Open-Set (EOS) and Objectosphere (Obj.) approaches.

as belonging to (Pathogen class $\in \mathcal{K}$)[i]. Otherwise, in the case $s_i < \lambda_i$, $i \in \mathcal{K}$, the result is inconclusive.

Since different classes have different rates of confidence represented by the softmax score, the class-adaptive threshold leads to a higher *average* rate of conclusive outcomes. Increasing the per-class threshold reduces both the FP and error rates while increasing the number of inconclusive outcomes, and we compute $\langle\lambda\rangle_{\mathcal{K}}^{\mathrm{FP/Err.=0\%}}$ providing both FP = 0% and error rate = 0%. A similar approach was successfully used for open-world text classification as a part of the DOC model.[78]

The corresponding results are provided in Table 3, and one can observe that when the category $\mathcal{I}$ is chosen appropriately, the Entropic Open-Set and Objectosphere approaches consistently outperform the naive thresholding. For example, naive thresholding with $\mathcal{K} = p_1 + p_2 + p_3$ and $\mathcal{I} = \varnothing$ have $\langle\lambda\rangle_{p_1+p_2}^{\mathrm{FP/Err.=0\%}} = 38.1 \pm 2.1\%$, naive thresholding with $\mathcal{K} = p_1 + p_2$ and $\mathcal{I} = \varnothing$ have $\langle\lambda\rangle_{p_1+p_2}^{\mathrm{FP/Err.=0\%}} = 45.1 \pm 0.4\%$, while EOS and Obj. with $\mathcal{K} = p_1 + p_2$ and $\mathcal{I} = p_3$ have 55.5 ± 1.4% and 55.4 ± 1.8%, respectively. However, as mentioned before, when p_2 is included in the $\mathcal{I}$ data set, this leads to degradation of performance, as marked by the $\surd$ sign in Table 3.

Table 3. Comparison of the Average Rate of Conclusive Results over the $\mathcal{K}$ Category by Naive and Open-Set Methods

	$\langle\lambda\rangle_{p_1}^{\mathrm{FP/Err.=0\%}}$	$\langle\lambda\rangle_{p_1+p_2}^{\mathrm{FP/Err.=0\%}}$	$\langle\lambda\rangle_{p_1+p_2+p_3}^{\mathrm{FP/Err.=0\%}}$
naive: $\mathcal{K} = p_1$, $\mathcal{I} = \varnothing$	40.3 ± 1.0%	N/A	N/A
naive: $\mathcal{K} = p_1 + p_2$, $\mathcal{I} = \varnothing$	46.7 ± 0.8%	45.1 ± 0.4%	N/A
EOS: $\mathcal{K} = p_1$, $\mathcal{I} = p_2$ $\surd$	42.3 ± 1.0%	N/A	N/A
naive: $\mathcal{K} = p_1 + p_3$, $\mathcal{I} = \varnothing$	30.9 ± 1.2%	N/A	N/A
EOS: $\mathcal{K} = p_1$, $\mathcal{I} = p_3$	46.7 ± 1.4%	N/A	N/A
Obj.: $\mathcal{K} = p_1$, $\mathcal{I} = p_3$	51.8 ± 1.8%	N/A	N/A
naive: $\mathcal{K} = p_1 + p_2 + p_3$, $\mathcal{I} = \varnothing$	37.0 ± 2.6%	38.1 ± 2.1%	44.6 ± 1.4%
EOS: $\mathcal{K} = p_1$, $\mathcal{I} = p_2 + p_3$ $\surd$	48.1 ± 1.2%	N/A	N/A
EOS: $\mathcal{K} = p_1 + p_2$, $\mathcal{I} = p_3$	52.1 ± 1.0%	55.5 ± 1.4%	N/A
Obj.: $\mathcal{K} = p_1 + p_2$, $\mathcal{I} = p_3$	51.4 ± 1.5%	55.4 ± 1.8%	N/A

A comparison of the conclusive results for different classes of pathogens is presented in Figures 7 and 8. One can observe

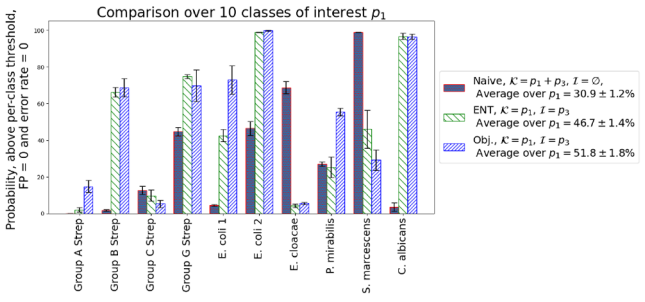


Figure 7. Comparison of conclusive outcomes over classes $\mathcal{K} = p_1$ for naive and Open-Set methods. Error bars represent one standard deviation over four ensembles.

that while Entropic Open-Set and Objectosphere approaches have a higher rate of conclusive results *on average* over the classes of interest, there are pathogen classes on which naive thresholding has a higher rate of conclusive results. This observation can be attributed to both the inherent properties of the pathogens under study and potential artifacts associated with the DNN. For example, *Serratia marcescens* is a Gram-negative bacterium associated with nosocomial infections exhibiting unique biochemical characteristics resulting in distinct spectral signatures compared to other bacterial

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452

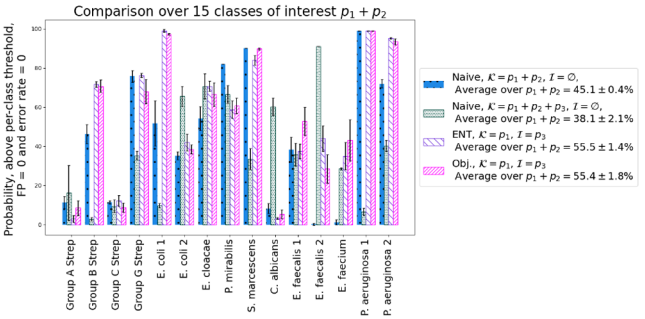


Figure 8. Comparison of conclusive outcomes over classes $\mathcal{K} = p_1 + p_2$ for naive and Open-Set methods. Error bars represent one standard deviation over four ensembles.

pathogens which may have contributed to its distinguishability from the OOD samples. Additionally, although efforts were made to select the $\mathcal{I}$ category to be distinct from $\mathcal{K}$, there may be relevant spectral features present that are suppressed during feature regularization. Therefore, in future studies, we plan to design the $\mathcal{I}$ category explicitly rather than selecting it from the available set of pathogens.

VI. SUPPLEMENT OF OOD DETECTORS AND THEIR EVALUATION

Finally, we implement and test OOD detectors that aim to separate $\mathcal{N}$ samples from $\mathcal{K}$ *after* training, namely, Mahalanobis,[79] OpenMax,[80] and ODIN[81] detectors.

As shown in Figure 9 and Figure 10, the ODIN detector based on input perturbations and temperature scaling performs much better than the other two OOD detectors and separates $\mathcal{K}$ and $\mathcal{I}$ with a significant margin while the other two detectors have a significant $\mathcal{K}/\mathcal{I}$ overlap, similarly to the findings of ref 62.

At the same time, as shown in Figure 11, the feature regularization by the Objectosphere loss function during the training boosts the ODIN performance even more and leads to a significantly larger margin separating $\mathcal{K}$ and $\mathcal{I}$ scores. For the Objectosphere with $\mathcal{K} = p_1$ and $\mathcal{I} = p_3$, the maximum value of the ODIN score for known classes is $\max(\text{ODIN}(\mathcal{K})) = 4.7 \times 10^{-6}$, and the minimum value of the ODIN score for never seen before classes is $\max(\text{ODIN}(\mathcal{N})) = 3.5 \times 10^{-2}$ while for naive thresholding with $\mathcal{K} = p_1 + p_3$ the corresponding values are $\max(\text{ODIN}(\mathcal{K})) = 3.4 \times 10^{-8}$ and $\max(\text{ODIN}(\mathcal{N})) = 1.6 \times 10^{-2}$. For the case of Objectosphere with $\mathcal{K} = p_1 + p_2$ and $\mathcal{I} = p_3$, the $\mathcal{K}/\mathcal{I}$ margin is even larger; the maximum value of the ODIN score for known classes is $\max(\text{ODIN}(\mathcal{K})) = 4.5 \times 10^{-6}$, and the minimum value of the ODIN score for never seen before classes is $\max(\text{ODIN}(\mathcal{N})) = 6.7 \times 10^{-2}$, in comparison with naive thresholding for $\mathcal{K} = p_1 + p_2 + p_3$ with the corresponding values being $\max(\text{ODIN}(\mathcal{K})) = 5.4 \times 10^{-8}$ and $\max(\text{ODIN}(\mathcal{N})) = 1.8 \times 10^{-2}$. Additionally, as one can observe from Figure 11, the histogram corresponding to Objectosphere is noticeably shifted toward larger values of the ODIN scores. Similarly, even though the OpenMax detector performs worse than ODIN, the histograms corresponding to Objectosphere are shifted toward larger values as well.

If during the training stage, in addition to focusing on the knowns, the DNN has its features regularized by means of the Objectosphere loss function in eq 6, it leads to a significantly improved separation between knowns and unknowns in comparison with the application of the OOD detector with the naive approaches alone leading to improved reliability of inference.

VII. CONCLUSIONS AND FUTURE WORK

Machine learning-enabled Raman spectroscopy holds significant promise as a label-free, accurate, and rapid method for identifying pathogens and hazardous contaminants, contributing to the preservation of human lives. However, the reliability and robustness of ML models used in such applications pose limitations, particularly in critical scenarios where complete knowledge of all possible classes cannot be assumed and when there are substantial disparities between test and training data. To address this gap, we developed a unified approach that addresses the problem of reliable and robust classification of open-world Raman spectra by leveraging the capabilities of ResNet combined with the SE attention mechanism and Objectosphere loss function. We evaluated the proposed method on the bacteria-ID database and demonstrated that it not only provides better or comparable performance to state-of-the-art methods under closed-world conditions but also provides robust identification of out-of-catalog samples. Combination with the one-vs-rest classifier significantly improves the number of inconclusive outcomes while keeping the FP and error rate zero. Additionally, we showed that the conjunction of OOD detectors with our architecture boosts their performance and found that the ODIN detector performs significantly better than the Mahalanobis and OpenMax detectors, making it a valuable supplement for OOD detection.

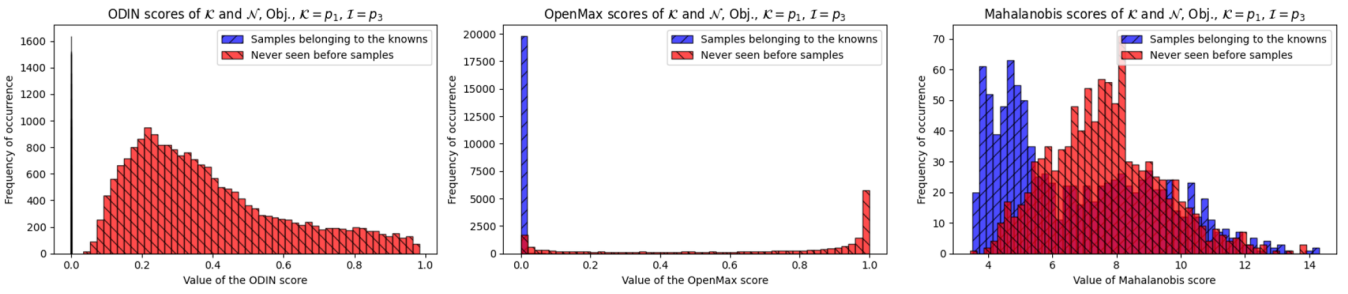


Figure 9. Comparison of OOD detectors after the model training with Objectosphere loss function with $\mathcal{K} = p_1$ and $\mathcal{I} = p_3$. A significantly better $\mathcal{K}/\mathcal{N}$ separation can be observed for the ODIN compared to the OpenMax and Mahalanobis.

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452

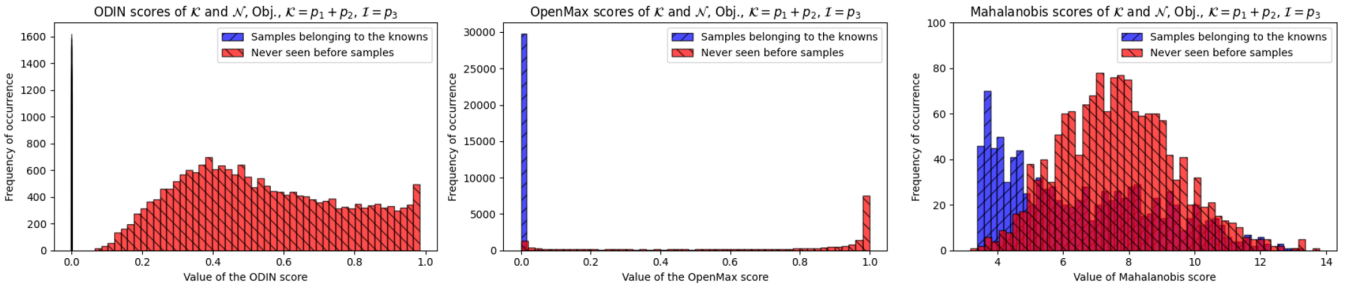


Figure 10. Comparison of OOD detectors after the model training with Objectosphere loss function with $\mathcal{K} = p_1 + p_2$ and $\mathcal{I} = p_3$. Again, significantly better $\mathcal{K}/\mathcal{N}$ separation can be observed for the ODIN detector.

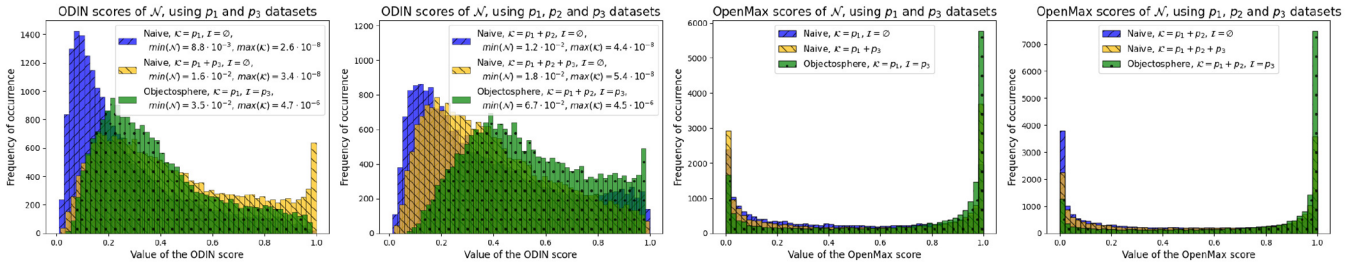


Figure 11. Comparison of ODIN and OpenMax detectors in combination with naive approaches and Objectosphere. One can observe a significantly better $\mathcal{K}/\mathcal{N}$ for the case of Objectosphere.

In the future, we aim to adapt our ML algorithm to cater to other critical applications such as public safety and environmental monitoring, benefiting from the adaptability of our proposed model to analyze Raman spectra in diverse contexts. Furthermore, given the versatility and adaptability of our approach, our future goal is to adapt it for medical image classification in open-world scenarios.

AUTHOR INFORMATION

Corresponding Author

Yaroslav Balytskyi − *Department of Physics and Astronomy, Wayne State University, Detroit, Michigan 48201, United States*; orcid.org/0000-0002-7582-6232; Email: hr6998@wayne.edu

Authors

Nataliia Kalashnyk − *National University of Civil Protection of Ukraine, Cherkasy 18034, Ukraine*

Inna Hubenko − *Cherkasy Medical Academy, Cherkasy 18000, Ukraine*; orcid.org/0000-0002-3798-6489

Alina Balytska − *Cherkasy Medical Academy, Cherkasy 18000, Ukraine*

Kelly McNear − *UCCS BioFrontiers Center, University of Colorado, Colorado Springs, Colorado 80918, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/cbmi.4c00007

Notes

The authors declare no competing financial interest.

Our model is publicly available: https://github.com/BalytskyiJaroslaw/PathogensRamanOpenSet.git.

ACKNOWLEDGMENTS

This work was supported by funding from the UCCS Cyber Seed Grant as well as the UCCS BioFrontiers Center. Y.B. appreciates fruitful discussions with Drs. Tristan Paul, Guy Hagen, and Sang-Yoon Chang.

REFERENCES

(1) Pelletier, M. J. *Analytical applications of Raman spectroscopy*; Blackwell Science: Oxford, 1999; Vol. *427*.

(2) Raman, C. V. A new radiation. *Indian J. Phys.* **1928**, *2*, 387−398.

(3) Landsberg, G. Eine neue erscheinung bei der lichtzerstreuung in krystallen. *Naturwissenschaften* **1928**, *16*, 557−558.

(4) Platonenko, V.; Khokhlov, R. On the mechanism of operation of a raman laser. *Sov. Phys. JETP* **1964**, *19*, 378−381.

(5) Hendra, P. J.; Stratton, P. Laser-raman spectroscopy. *Chem. Rev.* **1969**, *69* (3), 325−344.

(6) Lafuente, B.; Downs, R. T.; Yang, H.; Stone, N. 1. The power of databases: The rruff project. In *Highlights in mineralogical crystallography*; De Gruyter: Berlin, Germany, 2015; pp 1−30.

(7) Orlando, A.; Franceschini, F.; Muscas, C.; Pidkova, S.; Bartoli, M.; Rovere, M.; Tagliaferro, A. A comprehensive review on raman spectroscopy applications. *Chemosensors* **2021**, *9* (9), 262.

(8) Krafft, C.; Popp, J. The many facets of raman spectroscopy for biomedical analysis. *Anal. Bioanal. Chem.* **2015**, *407*, 699−717.

(9) Li, S.; Chen, G.; Zhang, Y.; Guo, Z.; Liu, Z.; Xu, J.; Li, X.; Lin, L. Identification and characterization of colorectal cancer using raman spectroscopy and feature selection techniques. *Opt. Express* **2014**, *22* (21), 25895−25908.

(10) Sigurdsson, S.; Philipsen, P. A.; Hansen, L. K.; Larsen, J.; Gniadecka, M.; Wulf, H.-C. Detection of skin cancer by classification of raman spectra. *IEEE transactions on biomedical engineering* **2004**, *51* (10), 1784−1793.

(11) Ellis, D. I.; Cowcher, D. P.; Ashton, L.; O'Hagan, S.; Goodacre, R. Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool. *Analyst* **2013**, *138* (14), 3871−3884.

(12) Guevara, E.; Torres-Galván, J. C.; Ramírez-Elías, M. G.; Luevano-Contreras, C.; González, F. J. Use of raman spectroscopy to screen diabetes mellitus with machine learning tools. *Biomedical Optics Express* **2018**, *9* (10), 4998−5010.

(13) Fleischmann, C.; Scherag, A.; Adhikari, N. K.; Hartog, C. S.; Tsaganos, T.; Schlattmann, P.; Angus, D. C.; Reinhart, K. Assessment

of global incidence and mortality of hospital-treated sepsis. current estimates and limitations. *American Journal of Respiratory and Critical Care Medicine* **2016**, *193* (3), 259−272.
(14) DeAntonio, R.; Yarzabal, J.-P.; Cruz, J. P.; Schmidt, J. E.; Kleijnen, J. Epidemiology of community-acquired pneumonia and implications for vaccination of children living in developing and newly industrialized countries: A systematic literature review. *Human Vaccines & Immunotherapeutics* **2016**, *12* (9), 2422−2440.
(15) Howes, P. D.; Rana, S.; Stevens, M. M. Plasmonic nanomaterials for biodiagnostics. *Chem. Soc. Rev.* **2014**, *43* (11), 3835−3853.
(16) Chung, H. J.; Castro, C. M.; Im, H.; Lee, H.; Weissleder, R. A magneto-DNA nanoparticle system for rapid detection and phenotyping of bacteria. *Nat. Nanotechnol.* **2013**, *8* (5), 369−375.
(17) Levy, M. M.; Evans, L. E.; Rhodes, A. The surviving sepsis campaign bundle: 2018 update. *Intensive Care Medicine* **2018**, *44* (6), 925−928.
(18) Chaudhuri, A.; Martin, P.; Kennedy, P.; Andrew Seaton, R.; Portegies, P.; Bojar, M.; Steiner, I. Efns guideline on the management of community-acquired bacterial meningitis: report of an efns task force on acute bacterial meningitis in older children and adults. *Eur. J. Neurol.* **2008**, *15* (7), 649−659.
(19) Berinson, B.; Both, A.; Berneking, L.; Christner, M.; Lütgehetmann, M.; Aepfelbacher, M.; Rohde, H. Usefulness of BioFire FilmArray BCID2 for blood culture processing in clinical practice. *J. Clin. Microbiol.* **2021**, *59* (8), 10−1128.
(20) Brusselaers, N.; Vogelaers, D.; Blot, S. The rising problem of antimicrobial resistance in the intensive care unit. *Ann. Intensive Care* **2011**, *1* (1), 1−7.
(21) Sullivan, Å.; Edlund, C.; Nord, C. E. Effect of antimicrobial agents on the ecological balance of human microflora. *Lancet Infectious Diseases* **2001**, *1* (2), 101−114.
(22) Fleming-Dutra, K. E.; Hersh, A. L.; Shapiro, D. J.; Bartoces, M.; Enns, E. A.; File, T. M.; Finkelstein, J. A.; Gerber, J. S.; Hyun, D. Y.; Linder, J. A.; et al. Prevalence of inappropriate antibiotic prescriptions among us ambulatory care visits, 2010−2011. *JAMA* **2016**, *315* (17), 1864−1873.
(23) Davies, J.; Davies, D. Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews* **2010**, *74* (3), 417−433.
(24) 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* (1), 1−8.
(25) Jadhav, S. A.; Biji, P.; Panthalingal, M. K.; Krishna, C. M.; Rajkumar, S.; Joshi, D. S.; Sundaram, N. Development of integrated microfluidic platform coupled with surface-enhanced raman spectroscopy for diagnosis of covid-19. *Med. Hypoth.* **2021**, *146*, 110356.
(26) Neng, J.; Zhang, Q.; Sun, P. Application of surface-enhanced raman spectroscopy in fast detection of toxic and harmful substances in food. *Biosens. Bioelectron.* **2020**, *167*, 112480.
(27) Valet, O.; Cherris, J.; Lee, K.; Lankers, M. Raman spectroscopy for identification of contaminant materials in pharmaceuticals. *MJH Life Sci.* **2016**, 8−17.
(28) Izake, E. L. Forensic and homeland security applications of modern portable raman spectroscopy. *Forensic science international* **2010**, *202* (1−3), 1−8.
(29) Choi, J.; Kim, J.-H.; Oh, J.-W.; Nam, J.-M. Surface-enhanced raman scattering-based detection of hazardous chemicals in various phases and matrices with plasmonic nanostructures. *Nanoscale* **2019**, *11* (43), 20379−20391.
(30) Mogilevsky, G.; Borland, L.; Brickhouse, M.; Fountain, A. W. Raman spectroscopy for homeland security applications. *Int. J. Spectrosc.* **2012**, *2012*, 1.
(31) Teng, L.; Wang, X.; Wang, X.; Gou, H.; Ren, L.; Wang, T.; Wang, Y.; Ji, Y.; Huang, W. E.; Xu, J. Label-free, rapid and quantitative phenotyping of stress response in e. coli via ramanome. *Sci. Rep.* **2016**, *6* (1), 34359.
(32) Tao, Y.; Wang, Y.; Huang, S.; Zhu, P.; Huang, W. E.; Ling, J.; Xu, J. Metabolic-activity-based assessment of antimicrobial effects by d2o-labeled single-cell raman microspectroscopy. *Analytical chemistry* **2017**, *89* (7), 4108−4115.
(33) Klein, D.; Breuch, R.; von der Mark, S.; Wickleder, C.; Kaul, P. Detection of spoilage associated bacteria using raman-microspectroscopy combined with multivariate statistical analysis. *Talanta* **2019**, *196*, 325−328.
(34) Yu, S.; Li, H.; Li, X.; Fu, Y. V.; Liu, F. Classification of pathogens by raman spectroscopy combined with generative adversarial networks. *Science of The Total Environment* **2020**, *726*, 138477.
(35) Nie, S.; Emory, S. R. Probing single molecules and single nanoparticles by surface-enhanced raman scattering. *science* **1997**, *275* (5303), 1102−1106.
(36) Rohleder, D. R.; Kocherscheidt, G.; Gerber, K.; Kiefer, W.; Köhler, W.; Möcks, J.; Petrich, W. H. Comparison of mid-infrared and raman spectroscopy in the quantitative analysis of serum. *Journal of Biomedical Optics* **2005**, *10* (3), 031108.
(37) Chollet, F. *Deep learning with Python*; Simon and Schuster: New York, 2021.
(38) Zheng, X.; Yin, L.; Lv, G.; Lv, X.; Chen, C.; Wu, G. Serum log-transformed raman spectroscopy combined with multivariate analysis for the detection of echinococcosis. *Optik* **2021**, *226*, 165687.
(39) Liu, W.; Sun, Z.; Chen, J.; Jing, C. Raman spectroscopy in colorectal cancer diagnostics: Comparison of pca-lda and pls-da models. *J. Spectrosc.* **2016**, *2016*, 1.
(40) Houhou, R.; Rösch, P.; Popp, J.; Bocklitz, T. Comparison of functional and discrete data analysis regimes for raman spectra. *Anal. Bioanal. Chem.* **2021**, *413* (22), 5633−5644.
(41) Song, D.; Yu, F.; Chen, S.; Chen, Y.; He, Q.; Zhang, Z.; Zhang, J.; Wang, S. Raman spectroscopy combined with multivariate analysis to study the biochemical mechanism of lung cancer microwave ablation. *Biomedical Optics Express* **2020**, *11* (2), 1061−1072.
(42) Kongklad, G.; Chitaree, R.; Taechalertpaisarn, T.; Panvisavas, N.; Nuntawong, N. Discriminant analysis pca-lda assisted surface-enhanced raman spectroscopy for direct identification of malaria-infected red blood cells. *Methods and protocols* **2022**, *5* (3), 49.
(43) Liu, Y.; Upadhyaya, B. R.; Naghedolfeizi, M. Chemometric data analysis using artificial neural networks. *Appl. Spectrosc.* **1993**, *47* (1), 12−23.
(44) Maruthamuthu, M. K.; Raffiee, A. H.; De Oliveira, D. M.; Ardekani, A. M.; Verma, M. S. Raman spectra-based deep learning: A tool to identify microbial contamination. *MicrobiologyOpen* **2020**, *9* (11), No. e1122.
(45) Thrift, W. J.; Ronaghi, S.; Samad, M.; Wei, H.; Nguyen, D. G.; Cabuslay, A. S.; Groome, C. E.; Santiago, P. J.; Baldi, P.; Hochbaum, A. I.; et al. Deep learning analysis of vibrational spectra of bacterial lysate for rapid antimicrobial susceptibility testing. *ACS Nano* **2020**, *14* (11), 15336−15348.
(46) 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 in Analytical Chemistry* **2020**, *124*, 115796.
(47) Peiffer-Smadja, N.; Dellière, S.; Rodriguez, C.; Birgand, G.; Lescure, F.-X.; Fourati, S.; Ruppé, E. Machine learning in the clinical microbiology laboratory: has the time come for routine practice? *Clinical Microbiology and Infection* **2020**, *26* (10), 1300−1309.
(48) Lu, W.; Chen, X.; Wang, L.; Li, H.; Fu, Y. V. Combination of an artificial intelligence approach and laser tweezers raman spectroscopy for microbial identification. *Anal. Chem.* **2020**, *92* (9), 6288−6296.
(49) Gniadecka, M.; Wulf, H.; Mortensen, N. N.; Nielsen, O. F.; Christensen, D. Diagnosis of basal cell carcinoma by raman spectroscopy. *J. Raman Spectrosc.* **1997**, *28* (2−3), 125−129.
(50) Hochreiter, S.; Bengio, Y.; Frasconi, P.; Schmidhuber, J.; et al. *Gradient flow in recurrent nets: The difficulty of learning long-term dependencies*; IEEE Press: New York, 2001.
(51) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer*

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452

Vision and Pattern Recognition, Washington, DC, November, 2016; pp 770−778.
(52) Deng, L.; Zhong, Y.; Wang, M.; Zheng, X.; Zhang, J. Scale-adaptive deep model for bacterial raman spectra identification. *IEEE Journal of Biomedical and Health Informatics* **2022**, *26* (1), 369−378.
(53) Zhou, B.; Tong, Y.-K.; Zhang, R.; Ye, A. Ramannet: a lightweight convolutional neural network for bacterial identification based on raman spectra. *RSC Adv.* **2022**, *12* (40), 26463−26469.
(54) Balytskyi, Y.; Bendesky, J.; Paul, T.; Hagen, G. M.; McNear, K. Raman spectroscopy in open-world learning settings using the objectosphere approach. *Analytical chemistry* **2022**, *94* (44), 15297−15306.
(55) Geng, C.; Huang, S.-j.; Chen, S. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence* **2021**, *43* (10), 3614−3631.
(56) Parmar, J.; Chouhan, S.; Raychoudhury, V.; Rathore, S. Open-world machine learning: applications, challenges, and opportunities. *ACM Computing Surveys* **2023**, *55* (10), 1−37.
(57) Bridle, J. S. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neurocomputing: Algorithms, architectures and applications*; Springer: New York, 1990; pp 227−236.
(58) Matan, O.; Kiang, R.; Stenard, C.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; Jackel, L.; Le Cun, Y. Handwritten character recognition using neural network architectures. *4th USPS Advanced Technology Conference*; 1990; Vol. *2*, pp 1003-1011.
(59) Goodfellow, I.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations*, December 19, 2014.
(60) Nguyen, A.; Yosinski, J.; Clune, J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2015; pp 427−436.
(61) Drummond, N.; Shearer, R. The open world assumption. *eSI Workshop: The Closed World of Databases meets the Open World of the Semantic Web*; 2006; Vol. *15*, p 1.
(62) Roady, R.; Hayes, T. L.; Kemker, R.; Gonzales, A.; Kanan, C. Are open set classification methods effective on large-scale datasets? *PLoS One* **2020**, *15* (9), No. e0238302.
(63) Sehwag, V.; Sitawarin, C.; Bhagoji, A. N.; Cullina, D.; Mittal, P.; Song, L.; Chiang, M. Analyzing the robustness of open-world machine learning. In *AISec 2019-Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, Proceedings of the ACM Conference on Computer and Communications Security*; Association for Computing Machinery: London, U.K., November 11, 2019; pp 105−116.
(64) Ren, J.; Liu, P. J.; Fertig, E.; Snoek, J.; Poplin, R.; Depristo, M.; Dillon, J.; Lakshminarayanan, B. Likelihood ratios for out-of-distribution detection. *Adv. Neur. Inf. Process. Syst.* **2019**, *32*.
(65) Guo, X.; Gichoya, J. W.; Purkayastha, S.; Banerjee, I. Margin-aware intraclass novelty identification for medical images. *Journal of Medical Imaging* **2022**, *9* (1), 014004−014004.
(66) Zhou, T.; Zhang, L.; Han, T.; Droguett, E. L.; Mosleh, A.; Chan, F. T. An uncertainty-informed framework for trustworthy fault diagnosis in safety-critical applications. *Reliability Engineering & System Safety* **2023**, *229*, 108865.
(67) Morgan, M. M.; Braasch, J. Open set classification strategies for long-term environmental field recordings for bird species recognition. *J. Acoust. Soc. Am.* **2022**, *151* (6), 4028−4038.
(68) Lee, K.; Lee, H.; Lee, K.; Shin, J. Training confidence-calibrated classifiers for detecting out-of-distribution samples. In *6th International Conference on Learning Representations, ICLR 2018*, Vancouver, Canada, April 30−May 3, 2018.
(69) Yu, Y.; Qu, W.-Y.; Li, N.; Guo, Z. Open category classification by adversarial sample generation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*; 2017; pp 3357−3363, .
(70) Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2018; pp 7132−7141.
(71) Pang, B.; Nijkamp, E.; Wu, Y. N. Deep learning with tensorflow: A review. *Journal of Educational and Behavioral Statistics* **2020**, *45* (2), 227−248.
(72) Jiang, Y.; Chen, L.; Zhang, H.; Xiao, X. Breast cancer histopathological image classification using convolutional neural networks with small se-resnet module. *PloS one* **2019**, *14* (3), No. e0214587.
(73) Al-Shaebi, Z.; Uysal Ciloglu, F.; Nasser, M.; Aydin, O. Highly accurate identification of bacteria's antibiotic resistance based on raman spectroscopy and u-net deep learning algorithms. *ACS omega* **2022**, *7* (33), 29443−29451.
(74) Dhamija, A. R.; Günther, M.; Boult, T. Reducing network agnostophobia. *Adv. Neur. Inf. Process. Syst.* **2018**, *31*.
(75) De Stefano, C.; Sansone, C.; Vento, M. To reject or not to reject: that is the question-an answer in case of neural classifiers. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2000**, *30* (1), 84−94.
(76) Fumera, G.; Roli, F. Support vector machines with embedded reject option. In *International Workshop on Support Vector Machines*; Springer: New York, 2002; pp 68−82.
(77) Shannon, C. E. A mathematical theory of communication. *Bell system technical journal* **1948**, *27* (3), 379−423.
(78) Shu, L.; Xu, H.; Liu, B. Doc: Deep open classification of text documents. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Copenhagen, Denmark; September, 2017; pp 2911−2916.
(79) Lee, K.; Lee, K.; Lee, H.; Shin, J. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Adv. Neur. Inf. Process. Syst.* **2018**, *31*.
(80) Bendale, A.; Boult, T. E. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016; pp 1563−1572.
(81) Liang, S.; Li, Y.; Srikant, R. Enhancing the reliability of out-of-distribution image detection in neural networks. In *6th International Conference on Learning Representations*; ICLR: Vancouver, Canada, 2018.

https://doi.org/10.1021/cbmi.4c00007
Chem. Biomed. Imaging 2024, 2, 442−452