





Identifying Bacteria Species on Microscopic Polyculture Images Using Deep Learning

Adriana Borowa , Dawid Rymarczyk , Dorota Ochońska, Agnieszka Sroka-Oleksiak ,
Monika Brzychczy-Włoch, and Bartosz Zieliński 

Abstract—Preliminary microbiological diagnosis usually relies on microscopic examination and, due to the routine culture and bacteriological examination, lasts up to 11 days. Hence, many deep learning methods based on microscopic images were recently introduced to replace the time-consuming bacteriological examination. They shorten the diagnosis by 1–2 days but still require iterative culture to obtain monoculture samples. In this work, we present a feasibility study for further shortening the diagnosis time by analyzing polyculture images. It is possible with multi-MIL, a novel multi-label classification method based on multiple instance learning. To evaluate our approach, we introduce a dataset containing microscopic images for all combinations of four considered bacteria species. We obtain ROC AUC above 0.9, proving the feasibility of the method and opening the path for future experiments with a larger number of species.

Index Terms—Deep learning, medical imaging, microorganisms, microscopy, neural networks.

I. INTRODUCTION

RECOGNIZING genera and species of bacteria is crucial in many domains, such as medicine, veterinary, biochemistry, food industry, and farming, because they can cause many diseases, including infectious ones. Therefore, it is necessary to automate the recognition process and, in consequence, accelerate bacteriological diagnosis. This would allow for implementation of appropriate medical prophylaxis and enable target treatment.

Manuscript received 12 April 2022; revised 25 August 2022 and 20 September 2022; accepted 20 September 2022. Date of publication 26 September 2022; date of current version 5 January 2023. This work was supported in part by the Foundation for Polish Science through Team-Net Program under Grant POIR.04.04.00-00-14DE/18-00, and in part by the European Union under the European Regional Development Fund and Priority Research Area DigiWorld through Excellence Initiative Program, Research University at the Jagiellonian University, Kraków, Poland. The BX63 microscope was purchased as part of the Initiative of Excellence (Research University IDUB Project) supported by the Ministry of Science and Higher Education. (Corresponding author: *Adriana Borowa*.)

Adriana Borowa, Dawid Rymarczyk, and Bartosz Zieliński are with the Faculty of Mathematics and Computer Science, Jagiellonian University, 30-348 Kraków, Poland (e-mail: ada.borowa@student.uj.edu.pl; dawid.rymarczyk@student.uj.edu.pl; bartosz.zielinski@uj.edu.pl).

Dorota Ochońska, Agnieszka Sroka-Oleksiak, and Monika Brzychczy-Włoch are with the Department of Molecular Medical Microbiology, Chair of Microbiology, Faculty of Medicine, Jagiellonian University Medical College, 31-121 Kraków, Poland (e-mail: dorota.ochonska@uj.edu.pl; agnieszka.sroka@uj.edu.pl; m.brzychczy-wloch@uj.edu.pl).

Digital Object Identifier 10.1109/JBHI.2022.3209551

The standard bacterial diagnostics procedure [1], presented in the upper part of Fig. 1, starts with collecting various types of test materials, such as swabs, scraps of skin lesions, urine, blood, or cerebrospinal fluid. Then, the clinical material is directly cultured on special media under specific temperature conditions (usually for 1–2 days, blood and cerebrospinal fluid samples require prior cultivation in automated closed systems for additional 1–5 days). Often bacteria colonies are too close to each other and it is not feasible to obtain monoculture colonies after the first inoculation on the culture medium.

To obtain samples with single species, they need to be separated in an iterative process (1–2 days). The initial identification of bacteria is based on microscopic observation, which takes into account the growth rate, type, shape, and color. Such analysis allows only approximate identification due to species similarity, in consequence, a bacteriological examination is required. It is a set of pre-laboratory and laboratory procedures aimed at identifying microorganisms and determining their drug sensitivity. Diagnostic diagrams in bacteriology consist of the following laboratory procedures:

- 1) microscopic examination of the direct preparation;
- 2) inoculating the material on an appropriate medium to obtain pure bacterial cultures;
- 3) morphological macro and microscopic observations of the obtained cultures;
- 4) testing the physiological properties of pure cultures;
- 5) immunological research; and
- 6) determination of sensitivity to antimicrobial substances, including drugs.

Conventional bacteriological examination may take up to 11 days.

Due to the long time required for the standard process of species identification and its high costs [2], it is beneficial to use methods that do not rely on conventional methods. Existing solutions can automatically distinguish between bacteria and fungi species [3], [4] or even bacteria clones [5] using microscopic images and deep learning methods. However, it is only possible for monoculture images, which requires multiple culture iterations.

In this work, we present a feasibility study for a further acceleration of diagnosis by reducing the number of culture iterations. For this purpose, we introduce a multi-label classification method based on multiple instance learning [6] to address a shortage of GPU memory when training the model on high resolution images. Firstly, we split each image into

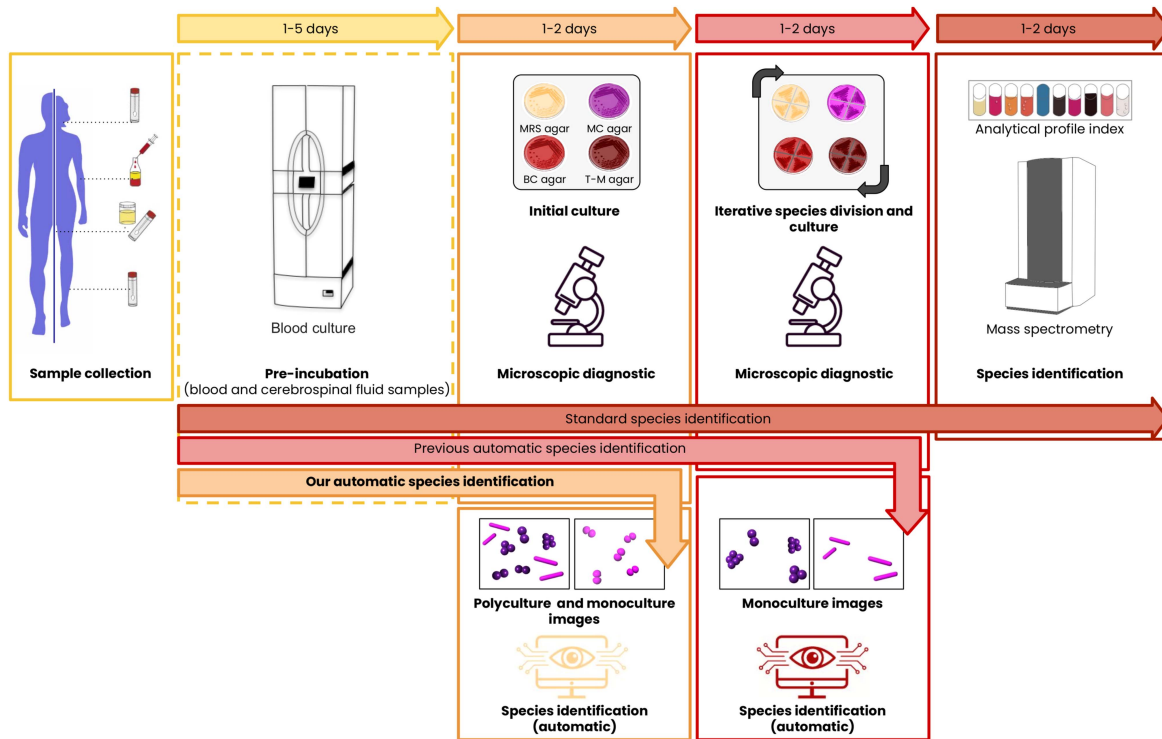


Fig. 1. Standard microbiological diagnosis requires iterative species division culture and biochemical tests, extending the diagnosis process up to 11 days. While the existing methods for automatic species identification do not require biochemical tests, they still need iterative species division and culture as they require monoculture images. In contrast, our method works on polyculture images. Hence, diagnosis shortens to 2–7 days.

smaller patches to which we assign the image labels. Because each patch is associated with a label, we can train a patch classifier. The output of its penultimate layer serves to generate a patch representation. Finally, we aggregate representations of all patches belonging to the analyzed image and pass the cumulative representation to the classifier.

To evaluate our approach, we introduce a dataset containing microscopic images for all combinations of four considered bacteria species. Moreover, we provide results for different variants of our method and compare them with existing state-of-the-art approaches. Additionally, we perform extensive ablation studies: on set of species of high resemblance as well as how image magnification and amount of training data influence the performance. Our contributions can be summarized as follows:

- Shortening the time of bacteria identification with methods classifying polyculture images.
- Introducing multi-MIL, a multi-label classification method based on multiple instance learning, with increased interpretability compared to existing methods.
- Providing a methodology for creating controlled datasets of polyculture images of exactly known species, similar to real-life images.

II. RELATED WORKS

For various medical purposes, e.g. an epidemiological investigation [5] and an infection diagnosis [3], the classification of microbiological organisms, especially bacteria, is essential.

Traditional methods of microorganism identification and classification are expensive and labour-intensive [7]. Therefore, researchers have been developing machine learning techniques to improve or even automate recognition of non-living infectious agents (e.g. viruses [8]) and microorganisms such as algae [9], bacteria [10], fungi [11], and protozoa [12]. However, according to our best knowledge, existing methods focus on identifying a single microbe per microscopy image.

Identification of microbes can be described in the context of types of imaging, taxonomy, and computational methods such as deep learning. However, due to the variety of approaches, we decided to present the related works chronologically, emphasizing computer vision methods and bacteria species identification.

One of the first works [13] clustered dinoflagellate cyst with self-organized maps (SOMs) on microscope-mounted camera images. Later, in [14] an artificial neural network was trained using contour invariant moment and morphological features extracted from microscopy images to identify wastewater bacteria. Just a year later, a probabilistic neural network [15] was used to classify five microorganisms, stained with fluorescent dyes and captured with a light microscope. The authors used nine morphological features to describe microbes in images of single bacteria extracted from the original microscopy image. At the same time, in [16], decision trees were used to identify *Mycobacterium tuberculosis* from ZN-stained sputum smear images. Hiremath and Bannigidad of [17] exploited information about cocci bacteria geometry and extracted morphological features, such as sphericity, to train a 3σ method, kNN, and ANN classifiers.

In successive years, researchers explored the classification of micro-organisms using methods such as random forest for classification of tuberculosis bacteria [18], minimal sequential optimization for an algae image classification [19], genetic programming for representing an image and optimum-path forest classifier [20]. The last work examined bright-field microscopy images of the 15 most common species of protozoan cysts, helminth eggs, and larvae with fecal impurities. Priya and Srinivasan of [21] also delve into tuberculosis research by extraction of fifteen Fourier descriptors passed to a multi-layer perceptron with activations classified via support vector machines. Meanwhile, five species of *Staphylococcus* bacteria were identified in hyperspectral microscopic images [22], and their classification was conducted with SVM and Partial Least Square Discriminant Analysis (PLS-DA).

Then, [3] classified bacteria colony using deep learning approach. Similarly, in [23] authors used texture features extracted from CNNs to identify gut bacteria in larval zebrafish using 3D light-sheet fluorescence microscopy images. Lakshmi and Sivakumar of [24] compared a multitude of methods, i.e. kNN, SVM, RF, ANN, and CNN which achieved the highest accuracy. In [25], a system for environmental microorganism classification on microscopic images was presented, and the authors used Conditional Random Fields (CRF) and Deep Convolutional Neural Networks (DCNN).

In recent years, CNN with Raman spectroscopy was used in [26], which used the database of thirty yeast and bacterial isolates of five species. Arredondo-Santoyo et al. of [27] investigated standard features, expert features, and features extracted using deep neural networks. This approach involved various machine learning algorithms, i.e. logistic regression, KNN, SVM, and random forest for classification. They also present the problem of dye decolorization in fungal strains. Fungus classification was also focus of [4] with the use of Fisher Vector and Random Forest on features extracted from AlexNet neural network [28]. Seven food-borne pathogens, captured with hyperspectral imaging, were classified using pixel features and a classifier based on SVM, competitive adaptive weighted sampling, and particle swarm optimization [29]. A novel approach based on coherent time-lapse images was used in [30] to detect live bacteria, even mixes of two species.

In the latest research, [5] used attention-based multiple instance learning pooling to classify clones of *Klebsiella pneumoniae* as well as persistence homology to obtain explanations of the model and description of each clone. Yu et al. of [31] created a hierarchical classification model for taxonomy purposes with PCA, LDA, and random forest, using gold nanoparticles measurements. Then, transfer learning was used in [32] with ResNet-18 [33] to detect longitudinal bacterial fission and in [34] with atrous convolution. Finally, [35] used various convolutional architectures to generate image representation which were then concatenated and classified with xgboost [36]. More detailed insights about microbe classification can be found in reviews [37], [38], [39].

According to our best knowledge, none of the aforementioned works consider a dataset of mixed bacteria species captured

on microscopy images as an alternative approach for iterative species division and culture in microbiological diagnosis. Therefore, in this work, we describe the results of a feasibility study on this problem.

III. MATERIALS

For this feasibility study, we selected four representative species of Gram-positive (*Lactobacillus plantarum*, *Staphylococcus aureus*) and Gram-negative bacteria (*Escherichia coli*, *Neisseria gonorrhoeae*) with significantly different morphology [40], [41], [42] (see Fig. 2). All microorganisms came from the American Type Culture Collection (ATCC).

L. plantarum belongs to the genus *Lactobacillus* called Lactic Acid Bacteria (LAB) and is facultatively anaerobic or strictly anaerobic rods. These microorganisms are a component of microbiota of the mouth, vagina, stomach, intestines, and genitourinary tract, especially in breastfed infants. Also, they are found in water, sewage, plants, food products, human body, and warm-blooded animals. The LAB bacteria are most commonly isolated in urine specimens and blood cultures due to transient bacteremia, endocarditis, or opportunistic septicemia. *Lactobacillus* strains are also very widely used as probiotics [43]. *S. aureus* is the best-known, highly virulent member of the genus *Staphylococcus* which are important pathogens in humans, causing a wide spectrum of life-threatening systematic diseases, including infections of the bones, skin, soft tissue, urinary tract, and opportunistic infections. They also cause sepsis and septic shock [44]. *E. coli* is an important member of the family *Enterobacteriaceae* and the most common aerobic, Gram-negative rods in the gastrointestinal tract. This bacteria is associated with various diseases, including gastroenteritis and extraintestinal infections such as urinary tract infections, meningitis, sepsis, and hemorrhagic colitis. Moreover, the presence of *E. coli* in the human intestine is an important indicator of fecal contamination of water, food, and medicines [45]. *N. gonorrhoeae* is the etiological factor of gonorrhea, one of the most widespread sexually transmitted diseases. These bacteria are strictly human pathogens [46]. Pure cultures of *E. coli* (strain ATCC 25922) were grown overnight at 37°C on MacConkey agar (MAC agar, Merck Germany), *L. plantarum* (strain ATCC14431) was isolated from MRS medium (De Man, Rogosa and Sharpe agar, Oxoid, UK), *N. gonorrhoeae* was selected from Theyer-Martin medium (T-M medium, Graso, Poland), and *S. aureus* was cultivated on Columbia CNA Agar with 5% Sheep Blood (CNA agar, Becton Dickinson, Germany). Then, samples of each bacteria were isolated from single bacterial colonies using a 1 μ l calibrated loop (Bionovo, Poland) and mixed on the surface of a basic microscope slide in a drop of saline. Bacteria species were mixed in all possible combinations to create samples containing up to 4 different species. Additional replicate was made for each mix, resulting in two microscopic preparation on two different slides. After fixing the slides over the flame of a hot burner for 10 seconds, they were Gram-stained using a commercially available kit (Merck, Poland) according to the manufacturer's instructions [41], [47]. Finally, microscopic images of samples were taken from 10 different locations per

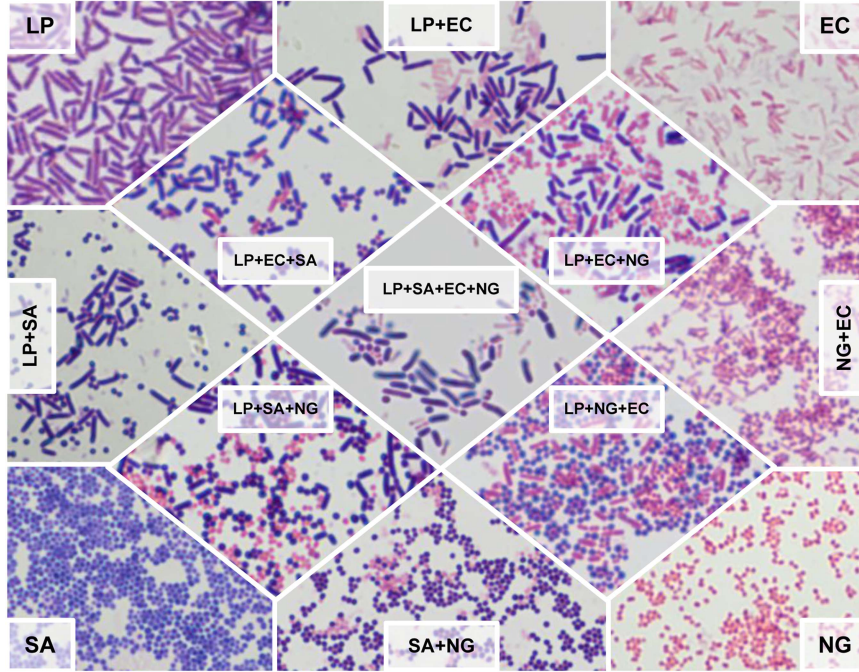


Fig. 2. In our database, bacteria species were mixed in all possible combinations to create samples containing 1, 2, 3, and 4 different species. The abbreviation EC refers to *Escherichia coli*, LP to *Lactobacillus plantarum*, NG to *Neisseria gonorrhoeae*, and SA to *Staphylococcus aureus*.

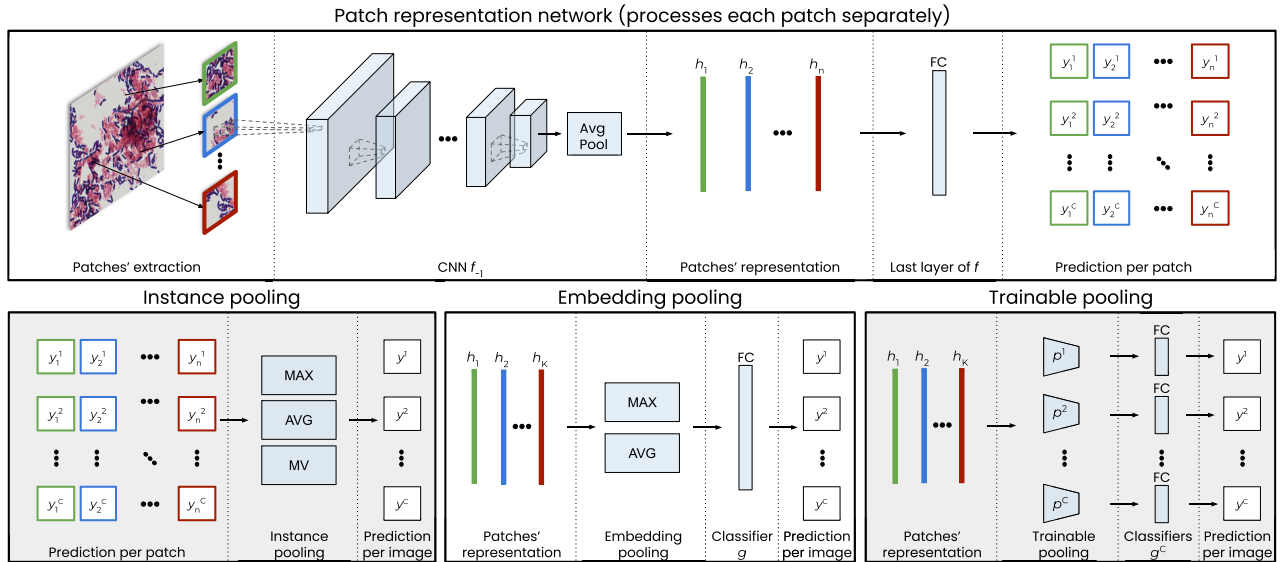


Fig. 3. Our method identifies bacteria species based on a ninput image. Its following steps divide the image into patches, create their representations, and aggregate and classify them.

slide. The resolution of obtained images was 4912×3684 pixels. Images were taken using an Olympus BX63 microscope with $100\times$ super-apochromatic objective under oil-immersion. The photographic documentation was then produced with an Olympus Hamamatsu camera ORC and CellSense software (Olympus).

IV. METHODS

To identify bacteria species, we develop a pipeline (see Fig. 3) which for a given image returns labels $y^c \in \{0, 1\}$ for $c = 1, \dots, C$ corresponding to each of C bacteria species. The

pipeline starts with an image preprocessing and extracting its patches $X = \{x_1, \dots, x_n\}$. Then, it generates representations of patches $\{h_1, \dots, h_n\}$ using a representation network f without the last layer (denoted f_{-1}). Patches' representations are then aggregated into an image representation h using various types of pooling p . Finally, a multi-label classifier g obtains C predictions.

A. Preprocessing Images and Extracting Patches

First, we decrease each image size by two in each dimension (magnification: $1/4x$) and we divide images into patches of

resolution 250×250 pixels using a sliding window mechanism with stride 125. This introduces some redundancy in information but allows us to include each bacteria cell. Some patches may not contain any material or bacteria overlapping so much that is impossible to classify them. This is due to bacterial cells being characterized by low density, they refract and absorb light poorly, which makes it difficult to distinguish them from the background, therefore they are clearly visible in the microscope only after staining. The microscopic preparation is prepared on a degreased, cooled glass slide by applying and spreading (smear) drops of the bacterial suspension using a loop. Although there are loops with a strictly defined mesh diameter, eg $1 \mu\text{l}$, $10 \mu\text{l}$ (so-called calibrated loops), while making a smear, even a calibrated loop cannot be controlled in any way by the random pattern of cells obtained on the slide. To reduce use of such uncontrolled patterns, we calculate the standard deviation σ_p of the pixel intensities and remove patches with $\sigma_p \notin [2, 15]$. The interval value for standard deviation σ_p was obtained experimentally using a training dataset to maximize the number of patches with clearly visible bacteria cells. Finally, following a good practice [48] and previous research [4], [5], [49], we normalize the remaining patches by subtracting the mean and dividing by the standard deviation. Both values are again derived from training patches. On average, we've obtained 160 patches per image resulting in total in 47149 patches obtained from 293 images. Detailed information about each experiment are presented in Table II.

B. Generating Patches' Representations

To derive a meaningful patch representation, we use a transfer learning technique. We pretrain ResNet-18 [33] neural network on ImageNet [50]. Then, we replace the last layer of the pretrained neurons with four neurons corresponding to four bacteria species and finetune the model (denoted f) with previously extracted patches. The resulted model without the final layer (denoted f_{-1}) is used to generate patches' representation h_i .

C. Aggregation and Classification

Here, we first recall the multiple instance learning definition and then provide its specific implementations, including instance and embedding-based methods, recurrent neural network, attention-based methods, and our novel multi-label Multiple Instance Learning (multi-MIL).

a) Definition: A typical supervised problem assumes that a single input x corresponds to a single output y of the model. However, in Multiple Instance Learning (MIL) [51], each input is represented by a bag of instances $X = \{x_i\}_{i=1}^n$ of variable size n , which also corresponds to a single output y . Moreover, in the standard MIL assumption there is binary $y \in \{0, 1\}$ and hidden binary labels $y_i \in \{0, 1\}$ of each instance (unavailable during training), where $y = 1$ if at least one $y_i = 1$. However, this assumption does not fit multi-label classification of bacteria species with C binary outputs.

b) Instance and embedding-based methods: The simplest MIL approaches, called instance-based methods, aggregate the predictions for bag instances using maximum

(MAX), average (AVG) or majority voting (MV) operator: $y = p(\{f(x_1), \dots, f(x_n)\})$, where $p \in \{MAX, AVG, MV\}$. On the other hand, more complicated embedding-based methods apply maximum or average pooling to embeddings of bag instances and then apply a classifier to derive an output $y = g(h)$, where $h = p(\{h_1, \dots, h_n\})$, and $h_i = f_{-1}(x_i)$.

c) Recurrent networks: Embeddings $\{h_1, \dots, h_n\}$ can also be considered as a sequence [52] and passed to Recurrent Neural Network (RNN) that jointly aggregates and classifies bags of various sizes. We employ this strategy with LSTM [53] and GRU [54] models to extend the number of baseline approaches.

d) Attention-based MIL: Embedding-based methods are imperfect because they apply pooling operations to all embedding without considering the importance of particular instances. As a result, a classifier can obtain irrelevant features. Hence, weighted average poolings were introduced based on the attention mechanism: Attention-based Multiple Instance Learning Pooling (AbMILP) [6] and Loss-based Attention (LA) [55].

In the case of AbMILP, pooling p is defined as

$$p(\{h_1, \dots, h_n\}) = \sum_{i=1}^n a_i h_i, \quad (1)$$

where weight a_i is described by

$$a_i = \frac{\exp(w^T \tanh(Vh_i))}{\sum_j^N \exp(w^T \tanh(Vh_j))}, \quad (2)$$

with trainable parameters w and V . Notice that the sum of all weights within the bag equals 1. Hence, the model works for various sizes of a bag.

In comparison, LA model simplifies the computation of weights to

$$a_i = \frac{\exp(w^T h_i)}{\sum_j^N \exp(w^T h_j)}, \quad (3)$$

with trainable parameter w . Moreover, w is reused as the parameter of classifier g to model hidden labels of instances and increase the interpretability. It is possible thanks to the simplified a_i computations and the same dimension of h and h_i .

Both AbMILP and LA return an aggregated bag representation, which is passed to the classifier g to obtain a prediction.

e) Multi-MIL: When there is a single output label, LA can be used to directly link the weights of instances with their influence on a prediction. However, LA cannot be directly used in a multi-label setup. At the same time, AbMILP can be used, but the correspondence between weights and influence is difficult to observe. Therefore, we introduce a multi-label version of those models, called multi-AbMILP and multi-LA. For this purpose, we provide separate weighted average pooling and classifier for each class

$$y = [y^c], \text{ where } h^c = g^c(h^c) \text{ and } h^t = p^t(h_i). \quad (4)$$

Hence, there are four different pairs of poolings and classifiers for four considered bacteria species. As a result, we obtain a direct correspondence of weights and influence, which improves the interpretability of the methods.

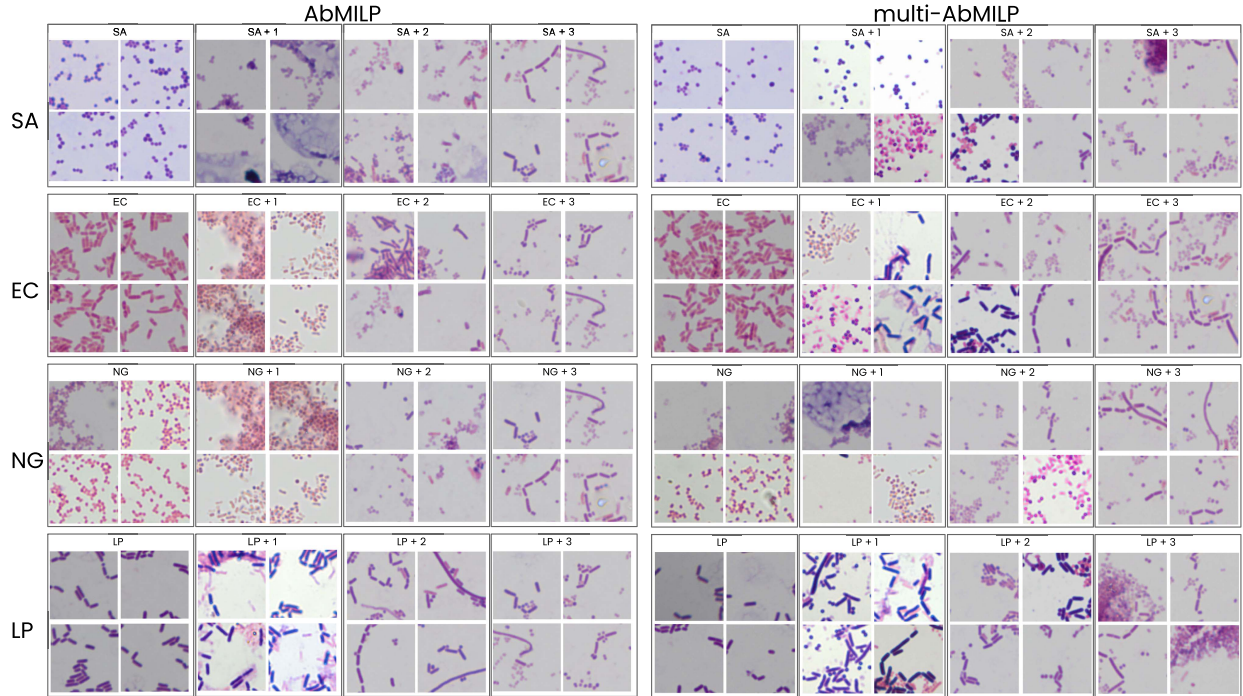


Fig. 4. The most important patches according to AbMILP and multi-AbMILP models. Description over each set presents how many additional species are present in the image except the primary species.

V. EXPERIMENTAL SETUP

We repeat all experiments five times. Each time, for each mix, we randomly assign one of its two slides to the training set and the second one to the testing set. This way, we eliminate the possible environmental bias. Moreover, all models are trained in three scenarios:

- *poly-poly*: f , p , and g are trained both on monoculture and polyculture images,
- *mono-poly*: f is trained on only monoculture images, but p and g are trained on both types of images,
- *mono-mono*: f , p , and g are trained only on monoculture images.

However, they are always tested using all images. We decided to use three training scenarios to estimate the importance of species combinations for the model training. It is essential because the number of combinations grows exponentially with the number of species. From this perspective, the poly-poly scenario obtains the highest accuracy but requires polyculture images. At the same time, the mono-mono scenario obtains the lowest accuracy but requires only monoculture images. That is why we test a third scenario, where the representation network f is trained on monoculture images while pooling p and classifier g use both monoculture and polyculture images. Because p and g have much fewer parameters than f , the last scenario can be used in future research to limit the number of polycultures images.

To finetune the representation network, we use batch size 64, an initial learning rate of 10^{-4} , which decreases 10 times every 1000 iteration. Hyperparameters were obtained from the preliminary experiments with learning rates from the range

[0.00000001, 0.001] and batch size values from {16, 32, 64}. In the preliminary experiments, we operated only the training set. All network layers are finetuned with binary cross-entropy loss and trained for 5000 iterations until the loss function reaches the plateau. Training images were augmented using color jittering, random rotation (90° , 180° , 270°), and random flip with an augmentation probability of 0.5. During the testing phase of the model, images are not augmented.

We performed the hyperparameters search for the classification network using grid search over learning rate from the range [0.000005, 0.001] and weight decay from the range [0.00001, 0.05]. We use a standard number of three attention heads [6] and batch size 1 due to the variability of the bag length.

We perform all the experiments on a workstation with four 12 GB GPU and 64 GB RAM. On average, it takes 10 hours to train the representation network and 2 hours to generate patch representations for the classification step. Training pooling and classifier lasts up to 4 hours. Both networks were implemented using PyTorch and Adam optimizer [56] with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

VI. RESULTS AND DISCUSSION

Table I presents the overall accuracy and ROC AUC in three considered scenarios (described in Section V) for ten different methods. In bold, we mark the best method and methods that are not significantly worse. We obtain them by comparing the best method to all others using the Wilcoxon signed-rank test. Results are significantly different if the p-value is smaller than 0.05.

TABLE I

RESULTS OF PATCH-BASED (CNN) AND IMAGE-BASED METHODS (REMAINING ROWS) FOR ALL THREE SCENARIOS PRESENTED IN SECTION V. RESULTS IN BOLD ARE STATISTICALLY BETTER ACCORDING TO THE WILCOXON SIGNED-RANK TEST

Method	Accuracy (%)			ROC AUC		
	poly-poly	mono-poly	mono-mono	poly-poly	mono-poly	mono-mono
CNN	84.8 ± 1.7	62.3 ± 2.6		0.935 ± 0.014	0.821 ± 0.022	
Instance + MV	87.7 ± 1.9	60.6 ± 1.4		n/a	n/a	
Instance + mean	87.7 ± 1.8	60.7 ± 1.5		0.943 ± 0.012	0.835 ± 0.038	
Instance + max	74.1 ± 3.9	72.7 ± 5.4	72.7 ± 5.4	0.954 ± 0.010	0.827 ± 0.055	
Embedding + mean	90.4 ± 0.9	81.8 ± 0.8	61.8 ± 1.5	0.960 ± 0.007	0.887 ± 0.014	0.754 ± 0.115
Embedding + max	89.9 ± 1.0	82.6 ± 1.6	60.5 ± 2.1	0.961 ± 0.015	0.887 ± 0.012	0.735 ± 0.137
RNN	86.7 ± 2.2	79.0 ± 2.4	61.6 ± 1.1	0.912 ± 0.014	0.852 ± 0.016	0.698 ± 0.017
AbMILP	90.1 ± 1.6	85.2 ± 1.1	62.3 ± 2.2	0.972 ± 0.009	0.899 ± 0.012	0.846 ± 0.037
multi-AbMILP	89.0 ± 5.9	95.4 ± 2.4	64.1 ± 1.5	0.944 ± 0.041	0.950 ± 0.029	0.860 ± 0.062
multi-LA	91.6 ± 0.7	96.0 ± 2.5	71.1 ± 4.2	0.954 ± 0.014	0.986 ± 0.014	0.882 ± 0.037

TABLE II

NUMBER OF IMAGES AND PATCHES IN TRAINING DATASETS FOR EACH EXPERIMENT AVERAGED OVER 5 FOLDS

SA + EC + NG + LP						
Dataset	Magnification	% of dataset	# of images	# of patches	Section	
poly	1x	100	139	91 340	VII-B	
mono	1x	100	36	23 537	VII-B	
poly	1/4x	10	13	2 093	VII-C	
poly	1/4x	50	69	11 124	VII-C	
poly	1/4x	100	138	21 727	IV, VII-C	
mono	1/4x	100	35	5 627	IV, VII-C	
poly	1/16x	100	130	4 673	VII-B	
mono	1/16x	100	31	1 199	VII-B	
SA + SH + SAP						
Dataset	Magnification	% of dataset	# of images	# of patches	Section	
poly	1/4x	100	70	13 236	VII-A	
mono	1/4x	100	30	6 534	VII-A	

A. Polyculture Images in All Training Steps (Poly-Poly)

To estimate the upper bound of problem performance, we train the models in the first scenario with monoculture and polyculture images. Almost all of the methods obtain ROC AUC over 0.9. Nevertheless, the highest ROC AUC is obtained with embedding-based methods, AbMILP, and multi-AbMILP (0.961, 0.944, and 0.972 respectively). This is expected because distributions of training and testing sets are similar and the information about polyculture images is propagated throughout the entire pipeline. However, this solution does not scale up for the growing number of recognized bacteria species because it is impractical to create polyculture images of all possible mixes.

B. Polyculture Images in the Pooling and the Classifier (Mono-Poly)

In the second scenario, the representation network is trained only on monoculture images, but the pooling and classifier are also trained on polyculture images. In this case, the CNN and instance-based methods work the same as in the mono-mono scenario. One can observe that results for multi-MIL methods are better than in the poly-poly scenario. Both multi-AbMILP and multi-LA give ROC AUC over 0.95. It indicates that multi-MIL methods do not require polyculture images when training the representation network. Therefore, they should behave satisfactorily when the number of recognized species grows.

C. Only Monoculture Images in All Training Steps (Mono-Mono)

In the third scenario, all steps are trained only with monoculture images containing single bacteria species. We observe a big decrease in accuracy for this scenario across all methods. It indicates that the polyculture information is crucial when training the pooling and classifier because when trained on single bacteria images, the model becomes confused seeing an image of polyculture. However, polyculture images are not necessary to train the representation network. Moreover, ROC AUC of attention-based methods AbMILP, multi-AbMILP, and multi-LA, is relatively high, again confirming their relevance.

D. Interpretability Aspects of Methods

Figure 4 presents the most important patches, i.e. patches with the largest weights in a pooling method. AbMILP model focuses on images with mixed bacteria species, while multi-AbMILP prefers images focused on one species. A similar trend is observed in Fig. 5, where we additionally present the least important patches, i.e. patches with the smallest weight in a pooling method. The figure shows that the AbMILP does not capture the nature of each species, while multi-AbMILP focuses on the most important patches with characteristic features of a given species. For example, in NG, we observe that the least important are patches with purple rods, while the most important ones are round and pink, which corresponds to the nature of NG that is a Gram-negative (pink) cocci (round). Therefore, we conclude that current MIL approaches cannot explain the results for each task, like the AbMILP model, which always weighs the patches similarly, no matter which species it predicts. In contrast, the multi-MIL models provide individual prediction interpretations for each task, making them more interpretable.

VII. ABLATION

In this section, we provide additional results on bacteria species identification using polyculture images. Firstly, we study how the deep learning models perform on bacteria species that are much more similar. Then, we analyze how the image magnification influences the model effectiveness, as well as how many training examples are required to obtain a meaningful model.

TABLE III
RESULTS FOR THE SECOND DATABASE CONTAINING 3 SPECIES FROM *STAPHYLOCOCCUS* GROUP

Method	Accuracy (%)			ROC AUC		
	poly-poly	mono-poly	mono-mono	poly-poly	mono-poly	mono-mono
CNN	81.4 ± 1.8	80.3 ± 1.5		0.961 ± 0.003	0.969 ± 0.004	
Instance + MV	84.9 ± 2.5	79.7 ± 2.1		n/a	n/a	
Instance + mean	84.9 ± 2.5	80.0 ± 1.9		0.964 ± 0.001	0.962 ± 0.003	
Instance + max	60.3 ± 4.6	62.9 ± 3.4		0.985 ± 0.003	0.991 ± 0.002	
Embedding + mean	82.0 ± 1.7	92.6 ± 2.0	41.8 ± 5.9	0.973 ± 0.001	0.979 ± 0.002	0.885 ± 0.064
Embedding + max	83.4 ± 4.1	85.7 ± 4.5	35.7 ± 4.2	0.971 ± 0.006	0.975 ± 0.002	0.853 ± 0.065
RNN	60.3 ± 3.4	58.6 ± 12.2	84.7 ± 3.3	0.911 ± 0.024	0.927 ± 0.027	0.963 ± 0.025
AbMILP	83.4 ± 3.8	84.9 ± 3.8	35.4 ± 2.6	0.981 ± 0.004	0.987 ± 0.003	0.962 ± 0.009
multi-AbMILP	91.7 ± 2.8	90.8 ± 1.2	70.7 ± 3.2	0.983 ± 0.004	0.959 ± 0.005	0.915 ± 0.020
multi-LA	95.3 ± 2.0	91.3 ± 2.9	63.6 ± 6.4	0.987 ± 0.003	0.975 ± 0.011	0.904 ± 0.018

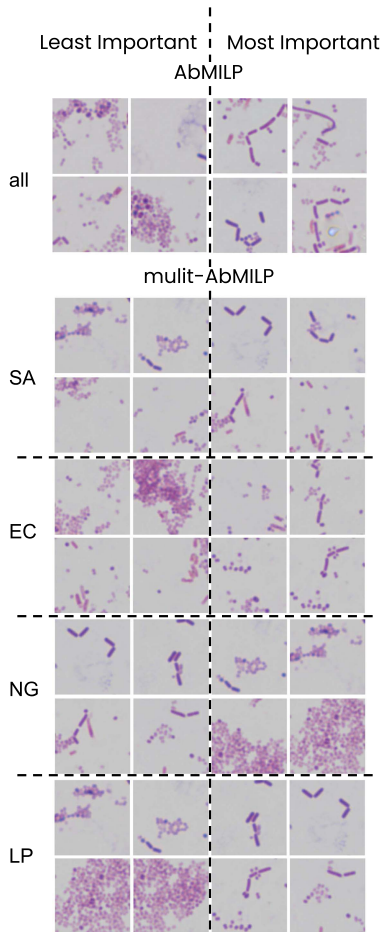


Fig. 5. The least and the most important patches according to AbMILP and multi-AbMILP models.

A. SA+SH+SAP

In this experiment, we check the performance of deep learning algorithms on polyculture bacteria images of species of a high resemblance. We use Gram-stained images of bacteria from the *Staphylococcus* group. They cause a wide spectrum of life-threatening systemic diseases and can be found on the skin, in the nostrils, urinary tract, and female reproductive tract. Those species can be commonly found in the human population and even 30% of humans can carry *Staphylococcus aureus*. In

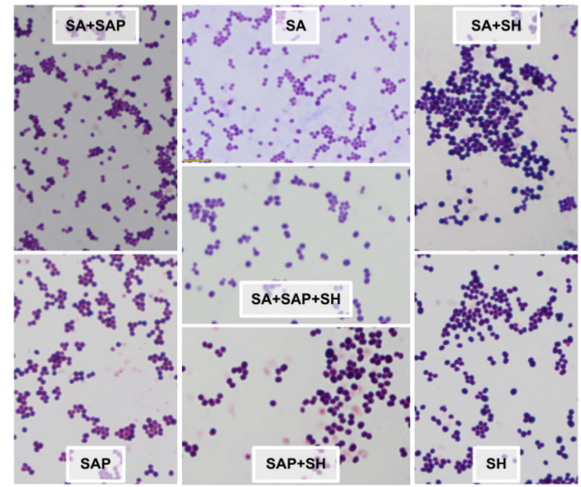


Fig. 6. Examples of images in the SA+SH+SAP database. The abbreviation SA refers to *Staphylococcus aureus*., SH to *Staphylococcus haemolyticus*, and SAP to *Staphylococcus saprophyticus*.

our subjective opinion, they are very similar to each other and the difference (mostly in the cell size) is barely perceptible in microscopic slides by the human eye. We are studying the following 3 species: *Staphylococcus haemolyticus* (SH), *Staphylococcus saprophyticus* (SAP), and *Staphylococcus aureus* (SA). Examples of those species are presented in Fig. 6. It is worth noting that even though those species are very similar and present a challenge to a deep learning model, our database contains only 3 of them which makes the classification problem slightly easier.

In Table III, we present the results on the datasets consisting of similar bacteria species. One can observe that the multi-MIL approach once again surpasses all the other methods, especially in a poly-poly scenario. Also, we observe that the accuracy of the models is poor in the mono-mono scenario. This is strictly related to the high resemblance of the staphylococci species to each other and the overfitting of the model. Indicating, that it is important to use polyculture images in the training phase to obtain a meaningful model.

B. Image Magnification

In this ablation study, we test 3 magnification of patches. We followed the same procedure for patch generation as in Section

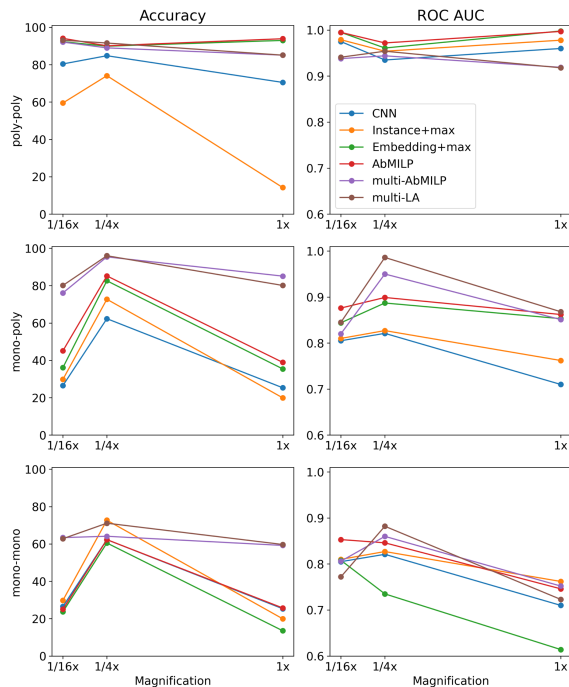


Fig. 7. Results of models trained with different magnification of patches.

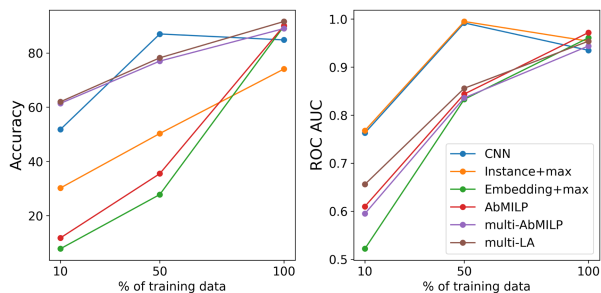


Fig. 8. Results of models trained with different percent of training data in the poly-poly scenario.

IV-A (1/4x) and introduced images in original size (1x) and in size decreased by 4 in each dimension (1/16x). Fig. 7 presents that using 1/4x magnification, in almost all cases, results in the best performance.

C. Percent of Training Data

We trained models with 10%, 50% and 100% of training data to study the amount of images needed for satisfactory results. Testing was performed on entire testing set in each case. Fig. 8 shows majority of methods have the best performance when trained on 100% of data but CNN-based methods can be also used with only 50% of training data.

VIII. CONCLUSION

This work introduces multi-label classification methods based on multiple instance learning to identify bacteria species on polyculture images. Our method takes advantage of the fact

that the multiple instance learning methods automatically assign interpretable weights to instances. Moreover, it introduces a mechanism that allows for multi-label classification without a decrease in the aforementioned interpretability. Experiments conducted on the specially created bacteria mixes database resulted in high ROC AUC values of up to 0.972, which supports the success of this feasibility study. In the future, we plan to expand the database to new bacteria species and other microorganisms, thus creating a tool for a fast and reliable microbiological diagnosis and, in consequence, a faster treatment. Additionally, we plan to analyze how different imaging techniques for capturing the bacteria species, such as novel microscopes that operate in nanoscale resolutions, influence the performance of artificial intelligence methods. However, those novel solutions are in early adaptation stages and it is challenging to create a substantial dataset for deep learning methods.

REFERENCES

- [1] J.-C. Lagier, S. Edouard, I. Pagnier, O. Mediannikov, M. Drancourt, and D. Raoult, "Current and past strategies for bacterial culture in clinical microbiology," *Clin. Microbiol. Rev.*, vol. 28, no. 1, pp. 208–236, 2015.
- [2] K. E. Tan, B. C. Ellis, R. Lee, P. D. Stamper, S. X. Zhang, and K. C. Carroll, "Prospective evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: A bench-by-bench study for assessing the impact on time to identification and cost-effectiveness," *J. Clin. Microbiol.*, vol. 50, no. 10, pp. 3301–3308, 2012.
- [3] B. Zieliński, A. Plichta, K. Misztal, P. Spurek, M. Brzywczy-Włoch, and D. Ochońska, "Deep learning approach to bacterial colony classification," *Plos One*, vol. 12, no. 9, 2017, Art. no. e0184554.
- [4] B. Zieliński, A. Sroka-Oleksiak, D. Rymarczyk, A. Piekarczyk, and M. Brzywczy-Włoch, "Deep learning approach to describe and classify fungi microscopic images," *Plos One*, vol. 15, no. 6, pp. 1–16, 2020. [Online]. Available: <https://doi.org/10.1371/journal.pone.0234806>
- [5] A. Borowa, D. Rymarczyk, D. Ochońska, M. Brzywczy-Włoch, and B. Zieliński, "Deep learning classification of bacteria clones explained by persistence homology," in *Proc. Int. Joint Conf. Neural Netw.*, 2021, pp. 1–8.
- [6] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *Proc. Int. Conf. Mach. Learn. PMLR*, 2018, pp. 2127–2136.
- [7] R. Franco-Duarte et al., "Advances in chemical and biological methods to identify microorganisms-from past to present," *Microorganisms*, vol. 7, no. 5, 2019, Art. no. 130.
- [8] L. Nanni, E. De Luca, M. L. Facin, and G. Maguolo, "Deep learning and handcrafted features for virus image classification," *J. Imag.*, vol. 6, no. 12, 2020, Art. no. 143.
- [9] E. Medina, M. R. Petraglia, and J. G. R. C. Gomes, "Neural-network based algorithm for algae detection in automatic inspection of underwater pipelines," in *Proc. Mex. Int. Conf. Artif. Intell.*, Springer, 2016, pp. 141–148.
- [10] T. Ahmed, M. F. Wahid, and M. J. Hasan, "Combining deep convolutional neural network with support vector machine to classify microscopic bacteria images," in *Proc. IEEE Int. Conf. Elect., Comput. Commun. Eng.*, 2019, pp. 1–5.
- [11] M. W. Tahir, N. A. Zaidi, A. A. Rao, R. Blank, M. J. Vellekoop, and W. Lang, "A fungus spores dataset and a convolutional neural network based approach for fungus detection," *IEEE Trans. Nanobiosci.*, vol. 17, no. 3, pp. 281–290, Jul. 2018.
- [12] Z. Liang et al., "Cnn-based image analysis for malaria diagnosis," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, 2016, pp. 493–496.
- [13] A. F. Weller, A. J. Harris, and J. A. Ware, "Artificial neural networks as potential classification tools for dinoflagellate cyst images: A case using the self-organizing map clustering algorithm," *Rev. Palaeobotany Palynology*, vol. 141, no. 3-4, pp. 287–302, 2006.

- [14] L. Xiaojuan and C. Cunshe, "An improved BP neural network for wastewater bacteria recognition based on microscopic image analysis," *WSEAS Trans. Comput.*, vol. 8, no. 2, pp. 237–247, 2009.
- [15] S. Kumar and G. S. Mittal, "Rapid detection of microorganisms using image processing parameters and neural network," *Food and Bioprocess Technol.*, vol. 3, no. 5, pp. 741–751, 2010.
- [16] Y. Zhai, Y. Liu, D. Zhou, and S. Liu, "Automatic identification of Mycobacterium tuberculosis from ZN-stained sputum smear: Algorithm and system design," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2010, pp. 41–46.
- [17] P. Hiremath and P. Bannigidad, "Identification and classification of cocci bacterial cells in digital microscopic images," *Int. J. Comput. Biol. drug Des.*, vol. 4, no. 3, pp. 262–273, 2011.
- [18] S. Ayas and M. Ekinci, "Random forest-based tuberculosis bacteria classification in images of ZN-stained sputum smear samples," *Signal, Image Video Process.*, vol. 8, no. 1, pp. 49–61, 2014.
- [19] S. Promdaen, P. Wattuya, and N. Sanevas, "Automated microalgae image classification," *Procedia Int. Conf. Comput. Sci.*, vol. 29, pp. 1981–1992, 2014.
- [20] C. T. Suzuki, J. F. Gomes, A. X. Falcao, J. P. Papa, and S. Hoshino-Shimizu, "Automatic segmentation and classification of human intestinal parasites from microscopy images," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 3, pp. 803–812, Mar. 2013.
- [21] E. Priya and S. Srinivasan, "Automated object and image level classification of TB images using support vector neural network classifier," *Biocybernetics Biomed. Eng.*, vol. 36, no. 4, pp. 670–678, 2016.
- [22] Y. Seo, B. Park, A. Hinton, S.-C. Yoon, and K. C. Lawrence, "Identification of Staphylococcus species with hyperspectral microscope imaging and classification algorithms," *J. Food Meas. Characterization*, vol. 10, no. 2, pp. 253–263, 2016.
- [23] E. A. Hay and R. Parthasarathy, "Performance of convolutional neural networks for identification of bacteria in 3D microscopy datasets," *PLoS Comput. Biol.*, vol. 14, no. 12, 2018, Art. no. e1006628.
- [24] S. Lakshmi and R. Sivakumar, "Chlorella algae image analysis using artificial neural network and deep learning," in *Biologically Rationalized Computing Techniques For Image Processing Applications*. Berlin, Germany: Springer, 2018, pp. 215–248.
- [25] S. Kosov, K. Shirahama, C. Li, and M. Grzegorzec, "Environmental microorganism classification using conditional random fields and deep convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 248–261, 2018.
- [26] C.-S. Ho et al., "Rapid identification of pathogenic bacteria using raman spectroscopy and deep learning," *Nature Commun.*, vol. 10, no. 1, pp. 1–8, 2019.
- [27] M. Arredondo-Santoyo et al., "Automatic characterisation of dye decolourisation in fungal strains using expert, traditional, and deep features," *Soft Comput.*, vol. 23, no. 23, pp. 12 799–12 812, 2019.
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [29] E. Bonah, X. Huang, R. Yi, J. H. Aheto, and S. Yu, "Vis-NIR hyperspectral imaging for the classification of bacterial foodborne pathogens based on pixel-wise analysis and a novel CARS-PSO-SVM model," *Infrared Phys. Technol.*, vol. 105, 2020, Art. no. 103220.
- [30] H. Wang et al., "Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning," *Light: Sci. Appl.*, vol. 9, no. 1, pp. 1–17, 2020.
- [31] T. Yu, S. Su, J. Hu, J. Zhang, and Y. Xianyu, "A new strategy for microbial taxonomic identification through micro-biosynthetic gold nanoparticles and machine learning," *Advanced Materials*, 2022, Art. no. 2109365.
- [32] C. Garcia-Perez, K. Ito, J. Geijo, R. Feldbauer, N. Schreiber, and W. Zu Castell, "Efficient detection of longitudinal bacteria fission using transfer learning in deep neural networks," *Front. Microbiol.*, vol. 12, 2021.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [34] S. Patel, "Bacterial colony classification using atrous convolution with transfer learning," *Ann. Romanian Soc. Cell Biol.*, pp. 1428–1441, 2021.
- [35] F. Ahmad, A. Farooq, and M. U. Khan, "Deep learning model for pathogen classification using feature fusion and data augmentation," *Curr. Bioinf.*, vol. 16, no. 3, pp. 466–483, 2021.
- [36] T. Chen et al., "Xgboost: Extreme gradient boosting," R package version 0.4-2, vol. 1, no. 4, pp. 1–4, 2015.
- [37] C. Li et al., "A brief review for content-based microorganism image analysis using classical and deep neural networks," in *Proc. Int. Conf. Inf. Technol. Biomed.*, Springer, 2018, pp. 3–14.
- [38] A. Locke, S. Fitzgerald, and A. Mahadevan-Jansen, "Advances in optical detection of human-associated pathogenic bacteria," *Molecules*, vol. 25, no. 22, 2020, Art. no. 5256.
- [39] Y. Zhang, H. Jiang, T. Ye, and M. Juhas, "Deep learning for imaging and detection of microorganisms," *Trends Microbiol.*, vol. 29, no. 7, pp. 569–572, 2021.
- [40] A. Popescu and R. Doyle, "The Gram stain after more than a century," *Biotechnic Histochemistry*, vol. 71, no. 3, pp. 145–151, 1996.
- [41] C. Gram, "Ueber die isolirte farbung der schizomyceten in schnitt- und trockenpreparaten," *Fortschritte der Medicin*, vol. 2, pp. 185–189, 1884.
- [42] C. Friedlander, "Die mikrokokken der pneumonie," *Fortschritte der Medicin*, vol. 1, pp. 715–733, 1883.
- [43] M. Bodaszewska-Lubas, M. Brzywczy-Włoch, T. Gosiewski, and P. Heczko, "Antibacterial activity of selected standard strains of lactic acid bacteria producing bacteriocins: Pilot study," *Postepy Higieny i Medycyny Doswiadczalnej*, vol. 66, pp. 787–794, 2012.
- [44] J. A. Lindsay and M. T. Holden, "Staphylococcus aureus: Superbug, super genome?," *Trends Microbiol.*, vol. 12, no. 8, pp. 378–385, 2004.
- [45] Z. D. Blount, "The natural history of model organisms: The unexhausted potential of *E. coli*," *eLife*, vol. 4, 2015, Art. no. e05826, doi: [10.7554/eLife.05826](https://doi.org/10.7554/eLife.05826).
- [46] A. P. R. d. Costa-Lourenço, K. T. B. d. Santos, B. M. Moreira, S. E. L. Fracalanza, and R. R. Bonelli, "Antimicrobial resistance in *Neisseria gonorrhoeae*: History, molecular mechanisms and epidemiological aspects of an emerging global threat," *Braz. J. Microbiol.*, vol. 48, pp. 617–628, 2017.
- [47] U. Srinivasan et al., "Gram stains: A resource for retrospective analysis of bacterial pathogens in clinical studies," *Plos One*, vol. 7, no. 10, 2012, Art. no. e4289.
- [48] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT press, 2016.
- [49] A. Borowa, D. Rymarczyk, D. Ochońska, M. Brzywczy-Włoch, and B. Zieliński, "Classifying bacteria clones using attention-based deep multiple instance learning interpreted by persistence homology," 2020.
- [50] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [51] J. Foulds and E. Frank, "A review of multi-instance learning assumptions," *Knowl. Eng. Rev.*, vol. 25, no. 1, pp. 1–25, 2010.
- [52] D. Bychkov et al., "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.
- [53] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [54] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," in *Proc. 8th Workshop Syn., Semantics Struct. Stat. Transl.*, 2014, pp. 103–111.
- [55] X. Shi, F. Xing, Y. Xie, Z. Zhang, L. Cui, and L. Yang, "Loss-based attention for deep multiple instance learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 04, 2020, pp. 5742–5749.
- [56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.