

Efficient Deep Learning Architectures for Fast Identification of Bacterial Strains in Resource-Constrained Devices

Rafael Gallardo García · Sofía Jarquín Rodríguez · Beatriz Beltrán Martínez ·
Carlos Hernández Gracidas · Rodolfo Martínez Torres

Received: date / Accepted: date

Abstract This work presents twelve fine-tuned deep learning architectures to solve the bacterial classification problem over the Digital Image of Bacterial Species Dataset. The base architectures were mainly published as mobile or efficient solutions to the ImageNet challenge, and all experiments presented in this work consisted in making several modifications to the original designs, in order to make them able to solve the bacterial classification problem by using fine-tuning and transfer learning techniques. This work also proposes a novel data augmentation technique for this dataset, which is based on the idea of artificial zooming, strongly increasing the performance of every tested architecture, even doubling it in some cases. In order to get robust

and complete evaluations, all experiments were performed with 10-fold cross-validation and evaluated with five different metrics: top-1 and top-5 accuracy, precision, recall, and F1 score. This paper presents a complete comparison of the twelve different architectures, cross-validated with the original and the augmented version of the dataset, the results are also compared with several literature methods. Overall, eight of the eleven architectures surpassed the 0.95 score in top-1 accuracy with our data augmentation method, being 0.9738 the highest top-1 accuracy. The impact of the data augmentation technique is reported with relative improvement scores.

Keywords Bacterial Identification · Data Augmentation · Deep Learning · DIBaS Dataset · MobileNets

Rafael Gallardo García
Language and Knowledge Engineering Laboratory, Benemérita Universidad Autónoma de Puebla, México
E-mail: rafael.gallardo@alumno.buap.mx
ORCID: 0000-0001-5085-3501

Sofía Jarquín Rodríguez
Faculty of Chemical Sciences, Benemérita Universidad Autónoma de Puebla, México
E-mail: ana.jarquin@alumno.buap.mx
ORCID: 0000-0002-5552-3699

Beatriz Beltrán Martínez
Language and Knowledge Engineering Laboratory, Benemérita Universidad Autónoma de Puebla, México
E-mail: bbeltran@cs.buap.mx
ORCID: 0000-0003-4528-4222

Carlos Hernández Gracidas
Faculty of Physical and Mathematical Sciences, CONACYT-Benemérita Universidad Autónoma de Puebla, México
E-mail: cahernandezgr@conacyt.mx
ORCID: 0000-0003-0267-6306

Rodolfo Martínez Torres
Language and Knowledge Engineering Laboratory, Benemérita Universidad Autónoma de Puebla, México
E-mail: beetho@cs.buap.mx

1 Introduction

Bacteria are small, unicellular microorganisms which are found almost everywhere on Earth. Most bacteria are not harmful to humans, because less than 1% of the different species of bacteria make people sick [18].

Due to the impact of bacteria in human life, the recognition of bacterial genera or species is a very common and important task in many areas such as medicine, veterinary science, biochemistry, food industry, or farming [38]. Traditional laboratory methods for the identification of bacterial strains commonly require an expert with knowledge and experience in the field. On the other hand, most techniques designed for rapid and automated identification of microbiological samples are based on biochemical or modular biology technologies [11, 33]. These traditional approaches are well established, however they are expensive and time consuming since they require complex sample preparation [2].

Automatizing the process of bacteria identification is very promising in the field of bioimage informatics. This field has yielded powerful solutions for specific image analysis tasks such as object detection, motion analysis or morphometric features [32]. Even so, most image analysis algorithms have been developed for very specific tasks or biological assays. Faster, more precise and less expensive computational approaches to bacterial strain classification are necessary. This paper presents a comparison among several deep learning architectures when solving the task of bacterial strain classification. All experiments were performed with architectures that have been proposed as efficient or mobile solutions to image classification, with the main advantage of low computational cost in terms of memory and number of calculations. In the experiments, the architectures were initialized with a pre-trained model (also called transfer-learning) over the ImageNet dataset [4], and then fine-tuned to our specific problem. The networks were trained from scratch in the cases where pre-trained models were unavailable. The evaluation of each proposal was performed using 10-fold cross validation over two different datasets: the original one, and a new one, generated by using data augmentation techniques. We present comparative Tables of all architectures, trained and evaluated over both datasets, reporting cross-validation averages of top-1 accuracy, top-5 accuracy, weighted precision, weighted recall, and weighted F1 score.

The main contributions of this work are summarized as follows:

- We present proposals that reach and exceed several methods available in the literature when measuring top-1 classification accuracy.
- We carry out a complete and robust evaluation of each method, training and testing with 10-fold cross validation in both datasets, and measuring performance with five different metrics.
- This work indicates that near-state-of-the-art performance can be achieved without using extremely expensive networks: all of our proposals have less than 10 million parameters.
- The efficient and low-resource nature of the networks we used, makes it easy to adapt and scale them to new problems and datasets, also making implementation and deployment on mobile or embedded devices easy.
- Derived from our experiments, this work contains comparative tables of relative improvement, which demonstrate that our data augmentation technique considerably increases the overall performance of several architectures, sometimes with increments of more than 100% when comparing against experiments with the original dataset.

2 State of the art

Holmberg et al. [8] published the first attempt to classify bacterial strains in an automated way. They used classification trees to extract the most important features and then classified them with artificial neural networks, achieving an accuracy of 76% in a dataset with 5 species of bacteria.

Later, in 2001, Liu et al. [15] presented a computer-aided system to extract size and shape measurements of digital images of microorganisms to classify them into their appropriate morphotype. Their work aimed to classify automatically each cell into one of 11 predominant bacterial morphotypes. This classifier had an accuracy of 96% on a set of 1,471 cells and 97% on a set of 4,270 cells.

Trattner et al. [36] proposed an automatic tool to identify microbiological data types with computer vision and statistical modeling techniques. Their methodology provided an objective and robust analysis of visual data, to automatize bacteriophage typing methods.

Some works consist in the identification of just one species of bacteria. Forero et al. [5] presented a method for automatic identification of *Mycobacterium tuberculosis* by using Gaussian mixture models. The authors used geometrical and color features of the images.

In 2013, Ahmed et al. [2] published a light scatter-based approach to bacterial classification using distributed computing (due to the computational complexity of their feature extractors). They used Zernike and Chebyshev moments and Haralick texture features, then, the best features were selected using Fisher's discriminant. The classifier was a support vector machine with a linear kernel. The classification accuracy varied from 90% to 99% depending on the species.

The Digital Image of Bacterial Species (DIBaS) dataset was presented in 2017 [38], Zieliński et al. published a deep learning approach to bacterial colony classification [38]. In this work, the authors used Convolutional Neural Networks as feature extractors and support vector machines or random forests as classifiers. Their overall highest accuracy over the DIBaS dataset was 97.24%, achieved by Fisher vector [25] encoders combined with VGG-M [30] network.

In 2018, Faruk Nasip and Zengin Kenan [20] published another deep learning approach to solve the DIBaS classification task. The authors used two different architectures, which mainly consist in bacteria classification using a VGG Net-based [30] approach and an AlexNet-based [14] approach, achieving 98.25% and 97.53% of classification accuracy, respectively.

More recently, in 2019, three works over the DIBaS dataset were published. The first one, by Khalifa et al.

[12], reached a 98.22% of classification accuracy with a deep convolutional neural network (CNN) and data augmentation techniques, which they called “Deep bacteria”. The second one, by Muhammed Talo [34], outperformed “Deep bacteria”, achieving a 99.2% of classification accuracy by fine-tuning a pre-trained model of ResNet-50. The third work [26] reached a 95.09% of classification accuracy by fine-tuning a pre-trained version of the MobileNet V2 architecture [27], and using some data augmentation techniques.

To the best of our knowledge, just three papers about bacteria classification over the DIBaS dataset were published in 2020. Elaziz et al. [1] achieved a 98.68% of classification accuracy, by using fractional-order orthogonal moments to extract features; also, they proposed a new feature selection method, which they called SSATLBO (Salp Swarm Algorithm + teaching-based learning optimization). In [7], authors achieved a 94.22% of classification accuracy, by using a fine-tuned version of a pre-trained MobileNet V2 architecture and data augmentation techniques. The last method, proposed by Satoto et al. [28], reached a 98.59% (over a subset of the DIBaS dataset, with just four classes) of classification accuracy by using a custom CNN topology and data augmentation techniques.

3 Traditional methods for bacterial identification

In clinical microbiology laboratories, manual and automated methods are used for bacterial identification. Classification of microorganisms to the species level is one of the most important tasks to microbiologists and other scientists involved in many areas. Phenotypic methods are the standard to identify them. However, they have some limitations for certain types of microorganisms. Molecular methods and proteomics-based methods are complementary to phenotypic methods and allow to overcome some limitations, although their implementation is less usual in some laboratories due to the higher cost and experience required for their application.

3.1 Phenotypic methods

Traditional bacterial phenotypic identification includes some macroscopic features and biochemical properties such as morphology, hemolysis, culture medium, production of certain metabolites, type of growth in relation to atmosphere, temperature and nutrition. Preliminary identification of most cultivable bacteria is based on such morphological characteristics [6]. Manual testing to identify bacteria requires from hours to several

days and a large amount of biological material, so that most laboratories use commercial biochemical test panels. The results of these panels are obtained in a short time-period using automated systems [37].

The following list contains some of the most common techniques that belong to this category:

- Culture medium.
- Multi-test galleries.
- Automated systems.

3.1.1 Advantages

These methods are inexpensive and very useful in the clinical laboratory because there are many different culture media for bacterial strains. Currently, biochemical test panels can be read with automated systems and provide results in short time-periods, these methods also allow many microorganisms to be analyzed simultaneously.

3.1.2 Disadvantages

Most phenotypic methods are laborious, require 36–48 hours and are limited to bacteria that are difficult to grow, also these methods consume many reagents and are not always useful to identify the microorganism to the species level.

3.2 Microscopy techniques

Optical microscopy is a simple and fast method that often guides to clinical diagnosis. However, this method is not successful by itself, usually because the identification of microorganisms is limited by factors such as small cells, variation in size or resolution of the optical microscope. Visualization under the microscope is more specific and easier to perform with the use of fluorescent markers. [19].

The following list contains some of the most common techniques that belong to this category:

- Transmission Electron Microscopy (TEM).
- Scanning Electron Microscopy (SEM).
- Confocal Laser Scanning Microscopy (CLSM).

3.2.1 Advantages

Microscopic observation is useful to reveal morphological features of isolated microorganisms and their diagnosis, mainly in the identification of microorganisms in biofilms. In microscopic techniques, fluorescent dyes are commonly used to get detailed visualization and higher-quality micrographs.

3.2.2 Disadvantages

Most microscopes are expensive and personnel must be trained to use them correctly, living cells are difficult to differentiate from dead cells in samples and a poorly prepared or contaminated sample reduces contrast and resolution. For these reasons, microscopy alone is not very accurate for microorganism identification.

3.3 Proteomics-based methods

Proteomics is the study and characterization of the set of proteins expressed by a genome (proteome) [3]. Methods based on mass spectrometry (MS) are important tools for microbial typing, they detect the mass-charge ratio (m/z) of the analyte. Particularly, the method known as Matrix Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) works by generating a peptide mass fingerprint (PMF) in the form of a spectral profile that is characteristic of each species, the result is compared with existing databases for its identification [31, 29].

3.3.1 Advantages

MALDI-TOF gets results in a few minutes, the protein spectrum that is characteristic of each microorganism provides an identification at the species level, the cost of the reagents is low and allows to analyze a large number of isolates simultaneously.

3.3.2 Disadvantages

The MALDI-TOF equipment is expensive and requires a strong initial invest. The detection is not direct from clinical samples, it also requires frequent calibrations and quality controls, the commercial databases used for the comparison and identification of microorganisms are extensive but still limited.

3.4 Molecular methods

Molecular methods are complementary or alternative procedures, which provide high-throughput, more sensitive and discriminatory analysis of microorganisms or genetic polymorphisms through the amplification and detection of specific nucleic acid targets. Some strains do not have a specific feature, the same strain can generate different results in repeated tests, most of the time molecular methods can solve these problems [37, 6].

The following list contains some of the most common techniques that belong to this category:

- Real-time PCR.
- Pulse-Field Gel Electrophoresis (PFGE).
- 16S-RNA.

3.4.1 Advantages

Most of these methods are fast and sensitive, they can identify bacteria that are difficult to grow under laboratory conditions, and are able to detect many pathogens simultaneously.

3.4.2 Disadvantages

Molecular methods are expensive and not widely used in routine clinical tests. In some cases, identification at the genus and species level could be confusing or incorrect when the obtained sequence is compared with the sequences of the databases.

4 Materials and methods

This section describes the methodologies we used to generate the dataset (the class distribution of the samples is also presented). Details about implementation, fine-tuning, training, and evaluation of all the deep learning architectures are provided as well.

4.1 Digital Images of Bacteria Species dataset

The original version of DIBaS dataset contains a total of 33 species of microorganisms, approximately 20 RGB images (2048×1532 pixels) per each species. It was published as the main dataset of the paper *Deep learning approach to bacterial colony classification* [38]. We removed the set of images of *Candida albicans* colonies as it is considered fungi [21]. The dataset was collected by the Chair of Microbiology of the Jagiellonian University in Krakow. The samples were stained using the Gramm's method. All images were taken with an Olympus CX31 Upright Biological Microscope and an SC30 camera with a 100 times objective under oil-immersion [38]. The DIBaS dataset is publicly available ¹.

4.2 Data augmentation

Our data augmentation strategy is aimed to provide at least 35 different images per each sample in the original dataset. The proposed approach tries to simulate different levels of zoom for the same bacterial sample,

¹ <http://misztal.edu.pl/software/databases/dibas/>

which is achieved by cropping different regions of different sizes from the full image. Specifically, we obtained five crops per image, with seven different sizes: from 100 to 700, with steps of 100, as we can see in Algorithm 1. All of the proposed architectures are designed to process inputs of 224x224 pixels; hence, after data augmentation, we resized every sample to the input size using Lanczos interpolation. At the end, our augmented version of DIBaS had a total of 24,073 samples (including the original images). Detailed information about sample distribution across classes is available in Table 1. We provide some graphical examples of the samples produced by our data augmentation method: Figure 1 shows some sample images generated with our data augmentation method and Fig. 2 presents a detailed visualization of the augmentation process for one sample of the *Actinomyces israelii* class.

Algorithm 1: Data augmentation process.

```

Result: An augmented version of the DIBaS
dataset.

for Species in DIBaS do
    for Sample in Species do
        for Shape in [100,200,300,400,500,600,700]
            do
                L=Obtain 5 crops of Sample with size
                Shape;
                resize all images in L to 224x224 px;
            end
            add all images in L to augmented folder;
            resize Sample to 224x224px;
            add Sample to augmented folder;
        end
    end

```

4.3 Deep learning architectures

This section contains a short description of each architecture we present in this paper. Overall, this paper presents a total of twelve architectures: all are product of transfer-learning and fine-tuning techniques. To accelerate the training process and increase the chances of a good generalization, we used pre-trained models (when available) on ImageNet to initialize the weights of our networks, after fine-tuning we trained the networks end-to-end. Our fine-tuning phase consisted in the modification of some layers of the networks (commonly, the last or the fully-connected layer), by making it suitable to solve a problem with 32 classes (original architectures were designed to solve a 1000-class problem), details about each implementation are described in the corresponding section.

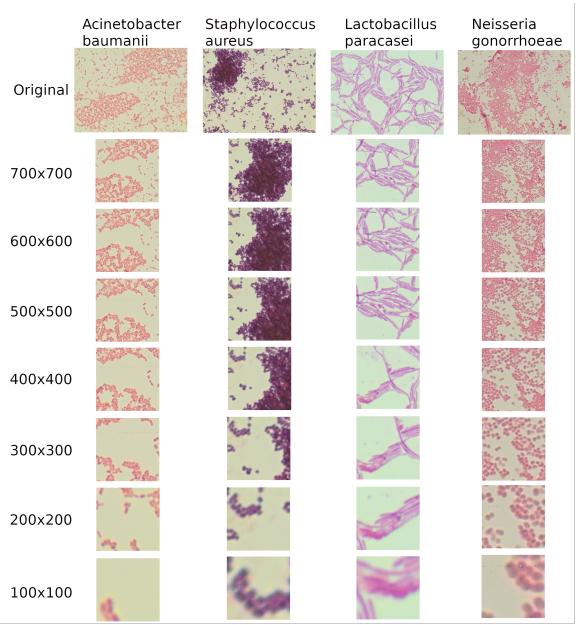


Fig. 1 Augmented samples of 4 of the 32 classes in the dataset. This figure shows the original image and one sample of each amplification.

Table 1 Samples per species, both in original and augmented versions of the dataset.

Genera	Species	# Original	# Augmented
Acinetobacter	baumannii	20	709
Actinomyces	israelii	23	828
Bacteroides	fragilis	23	828
Bifidobacterium	spp.	23	828
Clostridium	perfringens	23	828
Enterococcus	faecalis	20	720
	faecium	20	720
Escherichia	coli	20	720
Fusobacterium	spp.	23	828
Lactobacillus	casei	20	720
	crispatus	23	720
	delbrueckii	23	720
	gasseri	23	720
	jehnsenii	23	720
	johnsonii	23	720
	paracasei	23	720
	plantarum	23	720
	reuteri	23	720
	rhamnosus	23	720
	salivarius	23	720
Listeria	monocytogenes	22	792
Micrococcus	spp.	21	756
Neisseria	gonorrhoeae	23	828
Porphyromonas	gingivalis	23	828
Propionibacterium	acnes	23	828
Proteus	spp.	23	720
Pseudomonas	aeruginosa	23	720
Staphylococcus	aureus	23	720
	epidermidis	23	720
	saprophyticus	23	720
Streptococcus	agalactiae	23	720
Veionella	spp.	22	792

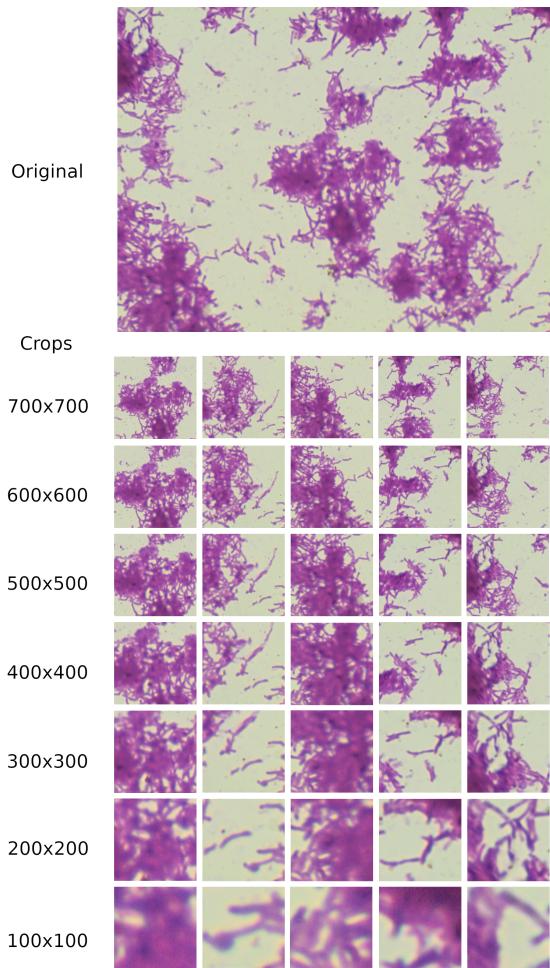


Fig. 2 Illustration of each sample generated by our data augmentation method (see Algorithm 1), applied to one image of the *Actinomyces israelii* class.

4.3.1 EfficientNet

It was published under the name *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks* [35]. One of the main contributions of the paper was a scaling method that uniformly scales all dimensions of a network (depth, width, and resolution), by using a *compound coefficient*. The compound scaling method uses a compound coefficient ϕ to uniformly scale the network dimensions. Depth (d), width (w), and resolution (r) of the network can be calculated with ϕ as follows:

$$\begin{aligned} d &= \alpha^\phi, \\ w &= \beta^\phi, \\ r &= \gamma^\phi, \end{aligned} \tag{1}$$

where α , β , and γ are constants that can be determined by a small grid search [35].

The authors proposed a family of models, called *EfficientNets*. The largest model achieved a state-of-the-art score of 84.3% of top-1 accuracy on ImageNet classification while being 8.4x smaller and 6.1x faster on inference than the best existing ConvNet at that moment [35]. Their base architecture is named EfficientNet-B0 and contains a total of 5.3 million parameters. EfficientNet-B1 to B7 were obtained by using different ϕ values on equation 1.

We used the implementation and pre-trained models of EfficientNet in PyTorch, provided in the EfficientNet-PyTorch repository ². The experiments included the three lighter versions of the EfficientNet, i.e., all the versions with less than 10 million parameters. The architectures are named as follows:

- efficientnet-b0: EfficientNet-B0, initialized with a pre-trained model on ImageNet.
- efficientnet-b1: EfficientNet-B1, initialized with a pre-trained model on ImageNet.
- efficientnet-b2: EfficientNet-B2, initialized with a pre-trained model on ImageNet.

All the EfficientNet-based architectures were trained using transfer learning. The original B0, B1, and B2 versions had a total of 5.3, 7.8, and 9.2 million parameters. We performed fine-tuning by modifying the fully connected layer of the network to have 32 output neurons. The final architectures have a total of 4, 6.6, and 7.7 million parameters, respectively.

4.3.2 MobileNet V2

MobileNet V2 was first reported in the paper *MobileNetV2: Inverted Residuals and Linear Bottlenecks* [27], where the authors proposed a new mobile architecture, based on their main contribution: a novel layer, called inverted residual with linear bottleneck. This novel layer takes a low-dimensional compressed representation as input, and then it is expanded to a high dimension, where it is filtered with a lightweight depth-wise convolution. The base architecture had a total of 3.4 million parameters and achieved a top-1 accuracy score of 72.0% over the ImageNet classification.

In our experiments, we used the official PyTorch implementation of the MobileNetV2, as well as its pre-trained model on ImageNet. We present just one version of this architecture, called mobilenet_v2, which is the result of different experiments and combinations of hyperparameters (detailed information available in [7]), from where we selected the best-performing architecture.

² <https://github.com/lukemelas/EfficientNet-PyTorch>

We used transfer learning to initialize our architecture; for fine-tuning, we changed the network topology by modifying the last fully connected layer of the original MobileNetV2. In the end, the network had a total of 32 output neurons. After fine-tuning, the total parameter number of the network was reduced to 2.2 millions.

4.3.3 MobileNet V3

Andrew Howard et al, presented the next generation of MobileNets (direct successor of MobileNetV2) in the paper *Searching for MobileNetV3* in 2019 [9]. They tuned MobileNetV3 to be efficient in mobile CPUs with a combination of hardware-aware network architecture search (NAS) and complementary approaches. The authors released two versions: MobileNetV3-large and MobileNetV3-small, with 5.4 and 2.5 million parameters, respectively, and claimed that, compared to MobileNetV2, the largest model is 3.2% more accurate on ImageNet classification while reducing latency by 15%. In top-1 classification accuracy of ImageNet, the “-large” model achieved 73.8% and the “-small” model achieved a 64.9%.

Our experiments were performed with a GitHub implementation of MobileNetV3³. Code and pre-trained models come from the same source. We used both the large and small version of the network to generate our solutions, which are named as follows:

- mobilenet_v3_small: MobileNet V3 Small, implemented and trained as described in the original paper, initialized with a pre-trained model on ImageNet.
- mobilenet_v3_large: MobileNet V3 Large, implemented and trained as described in the original paper, initialized with a pre-trained model on ImageNet.

As with the other architectures, we used transfer-learning to initialize the networks, then, we fine-tuned each version by modifying the last fully connected layer, in a way that the output is composed by 32 neurons. After fine-tuning, the number of parameters dropped from 5.4 and 2.5 millions, to 4.2 and 1.5 millions, for the large and small models, respectively.

4.3.4 ShuffleNet V2

First presented in the paper *ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design* [17]. Their best proposal achieved a top-1 accuracy score of 73.3% on ImageNet classification.

³ <https://github.com/d-li14/mobilenetv3.pytorch>

All the experiments were carried out using the official PyTorch [22] implementation of this architecture. We used pre-trained models when available on the PyTorch model zoo. The main difference among the four implementations lies in the output channel multiplier, which can be 0.5, 1, 1.5, or 2, which leads the number of parameters of the network to be 1.4, 2.3, 3.5, and 7.4 millions, respectively.

In total, we present experiments with 4 different ShuffleNet V2-based architectures, which are named as follows:

- shufflenet_v2_x0_5: ShuffleNet V2 with a 0.5 output channel multiplier. This version was initialized with a pre-trained model on ImageNet.
- shufflenet_v2_x1_0: ShuffleNet V2 with a 1.0 output channel multiplier. This version was initialized with a pre-trained model on ImageNet.
- shufflenet_v2_x1_5: ShuffleNet V2 with a 1.5 output channel multiplier. This version was trained from scratch because no pre-trained models were available.
- shufflenet_v2_x2_0: ShuffleNet V2 with a 2.0 output channel multiplier. This version was trained from scratch because no pre-trained models were available.

As we said before, we used transfer-learning in shufflenet_v2_x0_5 and shufflenet_v2_x1_0; the remaining architectures were trained from scratch. For fine-tuning, we changed the network topology by modifying the last fully connected layer of all the architectures in a way that all of them had a total of 32 output neurons (which is the number of classes in our problem). After fine-tuning, the number of parameters of the networks with output channel multipliers of 0.5, 1, 1.5, and 2 was reduced to 0.3, 1.2, 2.5 and 5.5 millions, respectively.

4.3.5 SqueezeNet

It appeared for the first time in the paper *SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters and <0.5 MB Model Size* [10], as a proposal to provide some of the advantages of small CNNs, such as: less communication across servers during distributed training, less bandwidth to export models from the cloud to autonomous cars and to provide CNNs feasible to deploy on field-programmable gate arrays (FPGAs) and other hardware with limited memory. Their base architecture achieved a top-1 accuracy score of 57.5% on the ImageNet classification task. Their main contribution consisted in a model size reduction up to 510x, when compared to AlexNet [14]; also, SqueezeNet models can be saved to disk with a size between 0.47 and 4.8 MB. The vanilla version had a total of 1.248 million parameters.

Similar to MobileNetV2 and ShuffleNetV2, we used the official PyTorch implementation and the pre-trained models on ImageNet to experiment with the SqueezeNet architecture, all available on the PyTorch model zoo. We present two SqueezeNet-based architectures to solve our problem, which are called as follows:

- squeezenet1_0: Vanilla version of SqueezeNet, also called SqueezeNet 1.0. This version was initialized with a pre-trained model on ImageNet.
- squeezenet1_1: A slightly modified version of the vanilla architecture, also called SqueezeNet 1.1. This version was initialized with a pre-trained model on ImageNet. SqueezeNet 1.1 has 2.4x less computation and 12,000 less parameters ⁴.

After transfer-learning, we fine-tuned both networks by modifying the last convolutional layer, to have a total of 32 output filters therein. After fine-tuning, the total parameters of the network reduced to 0.751 and 0.738 millions in squeezenet1_0 and squeezenet1_1, respectively.

4.4 Experiments

Above we have just described our architectures and some details about the implementation, in this section we are providing details about the process of training and evaluation with cross-validation.

4.4.1 Loss function, optimizer and hyper-parameters

The following list contains relevant information about our experimental setup:

- All architectures were trained by using the cross-entropy loss function (a combination between log softmax and the negative log likelihood loss).
- All architectures use the Adam optimizer [13], with the L2 penalty implemented as proposed in the paper *Decoupled Weight Decay Regularization* [16].
- All architectures are trained end-to-end, even when we used transfer-learning to initialize the weights. This allows the models to learn more problem-specific details.
- Learning rate, weight decay and the number of epochs are adjusted according to the performance of each architecture when training.
- Our cross-validation setup uses 10 folds in every experiment. The dataset is randomized but it uses the same random state in all experiments (to ensure uniform and fair evaluations).

⁴ <https://github.com/forresti/SqueezeNet/>

4.4.2 Software

The following list contains relevant information about the language, libraries and software:

- Dataset scripts, training and evaluation code are fully written in Python 3.
- We used PyTorch [23] as our main framework.
- The training and evaluation process of each architecture is a Jupyter Notebook ⁵. Hence, we have two notebooks for each architecture, one for the original dataset and one for the augmented version.
- All models are evaluated with five metrics: top-1 and top-5 accuracy, weighted precision, weighted recall and weighted F1 score. For the last three metrics, we used the Scikit-Learn implementation [24].

4.4.3 Hardware

Our main workstation during experiments works with an Intel i7-9750H processor, 20 gigabytes of RAM, and a commercial NVIDIA GTX 1660-Ti with 6 gigabytes of VRAM. All experiments, in training and evaluation phases fitted in just 6 gigabytes of VRAM and less than 10 gigabytes of RAM, ensuring the repeatability of our experiments without needing specific or high-tier hardware.

5 Results

To present a complete comparison, we provide four different tables:

- Table 2 contains the results of cross-validation with all our proposals, using the original dataset.
- Table 3 contains the results of cross-validation with all our proposals, using our augmented version of the dataset.
- Table 4 shows a comparison of the number of parameters and Top-1 accuracy among the literature methods and the best of our methods (one for each dataset).
- Table 5 shows the relative difference (or relative improvement) between metrics obtained before and after using our data augmentation method.

All values in the three tables are averages of the 10 folds.

As said above, Table 2 contains the results of all experiments performed over the original version of the dataset, which consists of between 20 and 23 samples per class. The results were consistent across the five

⁵ <https://jupyter.org/>

Table 2 Cross-validation results of our methods when using the original version of the DIBaS dataset. Models with a † were trained from scratch due to a lack of pre-trained models, the remaining ones were initialized with transfer-learning.

Method	# Parameters	Total samples	# Folds	# Epochs per fold	Top-1 Accuracy	Top-5 Accuracy	Precision	Recall	F1 Score
shufflenet_v2_x0_5	0.374 M	669	10	10	0.8893	0.9954	0.9125	0.8893	0.8810
squeezeenet1_1	0.738 M	669	10	15	0.5810	0.9686	0.5675	0.5810	0.5403
squeezeenet1_0	0.751 M	669	10	15	0.4487	0.8895	0.4366	0.4487	0.4047
shufflenet_v2_x1_0	1.286 M	669	10	10	0.9207	0.9940	0.9412	0.9207	0.9170
mobilenet_v3_small	1.505 M	669	10	10	0.9341	0.9954	0.9548	0.9341	0.9342
mobilenet_v2	2.264 M	669	10	10	0.9237	0.9910	0.9488	0.9237	0.9236
† shufflenet_v2_x1_5	2.511 M	669	10	20	0.6307	0.9626	0.6846	0.6307	0.6111
efficientnet-b0	4.048 M	669	10	15	0.8938	0.9895	0.9072	0.8938	0.8873
mobilenet_v3_large	4.243 M	669	10	10	0.9132	0.9925	0.9311	0.9132	0.9077
† shufflenet_v2_x2_0	5.543 M	669	10	20	0.6486	0.9655	0.6827	0.6486	0.6266
efficientnet-b1	6.654 M	669	10	10	0.8354	0.9835	0.8579	0.8354	0.8224
efficientnet-b2	7.746 M	669	10	10	0.8668	0.9940	0.8837	0.8668	0.8558

metrics (i.e. the same model achieved the highest score in all metrics): the mobilenet_v3_small architecture outperformed the other eleven methods in all metrics, only equaled by shufflenet_v2_x0_5 in top-5 accuracy. The second best architecture in top-1 accuracy and F1 score was the mobilenet_v2, only 0.010 points lower than the best model (in both metrics). The shufflenet_v2_x1_0 and mobilenet_v3_large architectures were close to the first one, only exceeded by 0.013 and 0.02 points in top-1 accuracy, respectively. We had models that showed strong under-fitting (reporting low scores in training and testing) to the problem, such as squeezeenet1_0, squeezeenet1_1, shufflenet_v2_x1_5, and shufflenet_v2_x2_0, all close or under the 0.6 of top-1 accuracy, surprisingly, all the trained models achieved top-5 accuracies higher than 0.8. Two architectures were trained from scratch, and both reported a clear underfitting, which we attribute to the lack of data (this set only contains 669 samples in total).

Table 3 shows a completely different scenario. In this set of experiments, only two models scored less than 0.9 in top-1 accuracy and all scored higher than 0.9 in top-5 accuracy. These experiments were conducted with a dataset with a total of 24,073 samples, the original data was augmented by using the method described in section 4.2. In this case, our results remain consistent across all the metrics, the best model (mobilenet_v3_large) achieved the higher accuracy in four of the five metrics: 0.9738 in top-1 accuracy, 0.9745 of precision, 0.0738 in recall, and an F1 score of 0.9738; this model was only outperformed by the mobilenet_v2 architecture in top-5 accuracy (which also scored 0.9745 in precision). The experiments with the same architectures of Table 2 but with our augmented version of the dataset show a strong increment of the scores, which indicates that our data augmentation techniques are highly effective on this problem. From the twelve models, all achieved a score of 0.9 or higher in precision (nine of the 12 exceeded the 0.95), and all models scored 0.99 or

higher in top-5 accuracy, also, eight models surpassed the 0.95 points in top-1 accuracy, recall and F1 score.

The highest top-1 accuracy score reported in the literature is that presented by Muhammed Talo [34], which scored 0.9922 in the original version of the dataset (using 5 fold cross-validation), his method was mainly based on the fine-tuning of a ResNet50 architecture (with an approximate of 23 million parameters). The second best architecture over the original version of the dataset was that presented by Elaziz et al. [1], which consisted in the SSATLBO method, and scored 0.9868 in top-1 accuracy. Table 4 contains five literature methods that performed their experiments over the original DIBaS dataset, the SSATLBO method is not a deep learning-based method, thus, we cannot compare its computational cost by measuring the number of parameters, however, from the four remaining methods, our mobilenet_v3_small architecture scored 0.9341 in top-1 accuracy with just 1.5 million parameters, while the others achieved 0.9724 [38], 0.9825 [20] and 0.9922 [34] scores by using architectures with, approximately, 23, 128, and 128 million parameters.

Table 4 also presents four literature methods that used an augmented version of the DIBaS dataset to perform their experiments. The highest score belongs to the method presented by Khalifa et al., which achieved a top-1 accuracy score of 0.9822 with a custom CNN of, approximately, 0.5 million parameters, their augmented dataset consisted of 6,600 images in training and 5,940 images in testing. The second highest score belongs to our MobileNet v3 Large model (with 4.2 million parameters), which achieved a 0.9738 of top-1 accuracy, trained and evaluated with 10-fold cross-validation over a dataset with 24,073 samples. The third and fourth scores were both achieved by different versions of the MobileNet v2, with top-1 accuracies of 0.9509 [26] and 0.9422 [7], those architectures mainly vary in their hyperparameters and the data, the former used an augmented version of the dataset with approximately 1,200 samples, and the last used a dataset with 26,524 sam-

Table 3 Cross-validation results of our methods when using the augmented version of the DIBaS dataset. Models with a † were trained from scratch due to a lack of pre-trained models, the remaining ones were initialized with transfer-learning.

Method	# Parameters	Total samples	# Folds	# Epochs per fold	Top-1 Accuracy	Top-5 Accuracy	Precision	Recall	F1 Score
shufflenet_v2_x0_5	0.374 M	24073	10	10	0.9557	0.9976	0.9602	0.9570	0.9571
squeezezenet1_1	0.738 M	24073	10	15	0.9136	0.9955	0.9210	0.9136	0.9125
squeezezenet1_0	0.751 M	24073	10	15	0.8882	0.9937	0.9007	0.8882	0.8866
shufflenet_v2_x1_0	1.286 M	24073	10	10	0.9635	0.9978	0.9666	0.9635	0.9636
mobilenet_v3_small	1.505 M	24073	10	10	0.9702	0.9975	0.9711	0.9702	0.9703
mobilenet_v2	2.264 M	24073	10	15	0.9737	0.9980	0.9745	0.9737	0.9737
† shufflenet_v2_x1_5	2.511 M	24073	10	15	0.9001	0.9906	0.9073	0.9001	0.9000
efficientnet-b0	4.048 M	24073	10	10	0.9720	0.9976	0.9728	0.9720	0.9720
mobilenet_v3_large	4.243 M	24073	10	15	0.9738	0.9977	0.9745	0.9738	0.9738
† shufflenet_v2_x2_0	5.543 M	24073	10	15	0.8984	0.9972	0.9706	0.8984	0.8977
efficientnet-b1	6.654 M	24073	10	10	0.9661	0.9965	0.9676	0.9661	0.9661
efficientnet-b2	7.746 M	24073	10	10	0.9715	0.9970	0.9723	0.9715	0.9715

Table 4 Comparative table of the number of parameters and top-1 accuracy of the methods available in the literature and the best of our methods for each version of the dataset. A * in the # Parameters column indicates that the number of parameters was approximated, as the original authors did not report the exact value. Methods with a * indicate one of our proposals.

Method	Description	# Parameters	Data augmentation	Top-1 Test Accuracy
Elaziz et al. [1]	SSATLBO	Not applicable	No	0.9868
Khalifa et al. [12]	Custom CNN	0.541 M *	Yes	0.9822
* Our (original)	MobileNet v3 Small	1.505 M	No	0.9341
Rujichan et al. [26]	MobileNet v2	3.904 M *	Yes	0.9509
Gallardo et al. [7]	MobileNet v2	3.904 M	Yes	0.9422
* Our (augmented)	MobileNet v3 Large	4.243 M	Yes	0.9738
Muhammed Talo [34]	ResNet50	23.573 M *	No	0.9922
Zieliński et al. [38]	Fisher Vectors + VGG-M	128.897 – 139.70 M *	No	0.9724
Nasip and Zengin [20]	VGGNet	128.897 – 139.70 M *	No	0.9825

ples, both proposals used training and test splits instead of cross-validation.

Table 5 contains relative improvements between scores obtained before and after augmenting the DIBaS dataset with Algorithm 1. The relative improvement of each architecture is calculated as follows:

$$r = \left(\frac{M_a - M_o}{M_o} \right) \times 100, \quad (2)$$

where r is the relative improvement, M_a is the score of some architecture in metric M when using the original dataset, and M_o is the score of the same architecture, with the same metric, but when using our data augmentation method.

As we can see in Table 5, all architectures reported a positive relative improvement, which indicates that the data augmentation was effective in all cases. Again, our results are consistent among the five metrics, being squeezezenet1_0 the architecture with the highest relative improvements in all metrics, surpassing the 100% both in Precision and F1 scores, which can be interpreted as doubling the score obtained when only the original data was used. Some other architectures also reported very high relative improvements, such as squeezezenet1_1, shufflenet_v2_x1_5, and shufflenet_v2_x2_0.

6 Discussion

As mentioned before, as well as in other literature works (such as [12]), the low quantity of samples in the original version of the dataset may cause fitting problems in some architectures. Table 2 shows some examples of systems that under-fitted the problem even when using transfer-learning (the case of squeezezenet1_0 and squeezezenet1_0), other architectures like shufflenet_v2_x1_5 and shufflenet_v2_x2_0 were unable to fit the problem from scratch and more tests are needed to see if transfer-learning can help to get better results in the original dataset. From our perspective, the main problem lies in the lack of training data as its clear that the gap between scores in Table 2 and Table 3 for the aforementioned architectures is big. On the other hand, we have those architectures that achieved high scores (0.9 or higher) with the original version of DIBaS (such as mobilenet_v3_small, shufflenet_v2_x1_0, mobilenet_v2 and mobilenet_v3_large), its easy to see that there is no clear relationship between the network capacity (# Parameters column) and the top-1 accuracy score. We speculate that this difference in the network generalization is caused by the variations of the information flow inside the topology of each architecture, but further research and analysis is needed. In Table 2, the three highest top-1 accuracy scores were achieved

Table 5 The relative improvement of all metrics, for each architecture, before and after data augmentation.

Method	Top-1 Accuracy (%)	Top-5 Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
shufflenet_v2_x0_5	7.4	0.2	5.2	7.6	8.6
squeezeenet1_1	57.2	2.7	62.2	57.2	68.8
squeezeenet1_0	97.9	11.7	106.2	97.9	119.0
shufflenet_v2_x1_0	4.6	0.3	2.6	4.6	5.0
mobilenet_v3_small	3.8	0.2	1.7	3.8	3.8
mobilenet_v2	5.4	0.7	2.7	5.4	5.4
shufflenet_v2_x1_5	42.7	2.9	32.5	42.7	47.2
efficientnet-b0	8.7	0.8	7.2	8.7	9.5
mobilenet_v3_large	6.4	0.5	4.6	6.6	7.2
shufflenet_v2_x2_0	38.5	3.2	42.1	38.5	43.2
efficientnet-b1	15.6	1.3	12.7	15.6	17.4
efficientnet-b2	12.0	0.3	10.0	12.0	13.5

by networks that consist of between 1 and 2.5 million parameters, and networks with less than 1 million or higher than 2.5 million parameters scored lower values (with the exempt of mobilenet_v3_large), this could be a clue that this is a good starting point to train good and computationally cheap models with the original dataset without needing strong pre-processing or hyper-parameter tuning to avoid under-or-over fitting in networks with different capacities. Our special cases of networks trained from scratch bring us some interesting data: both models trained from scratch scored <0.7 in top-1 accuracy score, and the remaining ShuffleNet-based architectures achieved higher than 0.8 in the same metric, while having less capacity but making use of pre-trained models to initialize the weights.

The experiments after the data augmentation process are completely different, and we can describe them as successful. Only two models could not score above 0.9 (squeezeenet1_0 and shufflenet_v2_x2_0) in top-1 accuracy, while all models scored higher than 0.99 in top-5 accuracy. The highest top-1 accuracy was achieved by the mobilenet_v3_large model, closely followed by the mobilenet_v2 model (the former is only 0.0001 points better), the latter was the best model in top-5 accuracy score. As we can see in Table 3, the highest scores are from architectures composed of between 1.2 and 7.7 million parameters, which is slightly different to the results of Table 2. We think that the increment in scores is certainly related to the increment of training data, as more data also means stronger and bigger architectures have sufficient information to better solve the problem. The aforementioned importance of the data is also present in the two special cases where we trained the networks from scratch (shufflenet_v2_x1_5 and shufflenet_v2_x2_0), even when ShuffleNet-based models were better than MobileNet V2 in top-1 accuracy scores over ImageNet. Here, those models could not surpass the 0.91 of top-1 accuracy (and MobileNet V2 scored 0.9737), also, the other two ShuffleNet-based architec-

tures with less parameters but initialized with pre-trained models achieved >0.95 of top-1 accuracy, highlighting the importance and impact of transfer-learning in this problem.

Table 5 clearly shows the importance and impact of the data augmentation method. As mentioned before, all architectures reported an improvement, in every metric. We are going to discuss four architectures that showed very interesting results: squeezeenet1_0, squeezeenet1_1, shufflenet_v2_x1_5, and shufflenet_v2_x2_0. The first two reported the highest relative improvement in all metrics, while also being two of the lighter architectures (by # of parameters), both networks have a low number of parameters and thus, they may require more data to generalize well. There is a large difference between the relative improvement reported by these architectures and the remaining ones; we consider that this behavior is due to the low capacity of these two networks. On the other hand, we have the shufflenet_v2_x1_5 and shufflenet_v2_x2_0 architectures, which also reported a high relative improvement, but in this case we have a different situation: both architectures were trained from scratch (i.e. without using transfer learning), so, it is clear that the data augmentation method strongly improved the network generalization and overall performance.

As the task of bacterial classification is not well standardized, it is difficult to present fair comparisons among the literature methods. From the works contained in Table 4, four use the original version of the dataset and five use data augmentation, but every work has different training and test sets; also, some papers used training and test splits and others used n-fold cross-validation; these non-standardized practices can be counterproductive when trying to make fair comparisons of the metrics among various methods available in the literature or when trying to declare a new state-of-the-art method. With the aforementioned, we present Table 4 just as a reference and a general overview of the

literature methods compared against our proposals. We believe that well-standardized practices and evaluations for this task (and dataset) will lead researchers to increasingly better and more useful solutions to the bacterial classification problem, in that sense, we performed all experiments with 10-fold cross validation (a practice that we hope will be replicated in future works), which gives reliability and certainty to our results, and also gives us the opportunity to present robust scores (by avoiding the problem of having easy-or-hard subsets). Another good step to standardize the experiments is to standardize the data, therefore, our augmented version of the data can be easily generated by running our data augmentation script over the original data.

7 Conclusions

This work presented a complete comparison of the performance of different efficient and mobile deep learning architectures to solve the problem of classification of bacterial species. The results reported here provide evidence that near-state-of-the-art performance can be achieved with computationally inexpensive methods. This work also provides empirical evidence that our data augmentation technique can be a game-changer when solving this problem. In total, in this work we proposed 12 architectures, a data augmentation strategy and 24 different models that try to solve the problem, all evaluated with five different metrics. Our results with the original version of the dataset go from 0.4487 to 0.9341 in top-1 accuracy, and from 0.8882 to 0.9738 with the same metric, experimenting with some data augmentation techniques. Overall, this paper presents 5 models that surpassed the 0.95 in top-1 accuracy and 14 that scored above 0.9.

This paper also presented an interesting and strong data-related contribution: our data augmentation method led all architectures to achieve a positive relative improvement. One architecture doubled its performance on two different metrics, while other three architectures reported relative improvements of more than 30% in each metric. The data augmentation technique presented here showed evidence that it can help some architectures to achieve performances that are similar to those networks that used transfer learning or to help networks with low capacity to achieve higher scores (or even doubling them).

Since bacterial identification is a sensitive area, which can involve serious diseases, a standardized version of this task is needed, in order to ensure the reliability of the systems that are product of research works like those presented in this paper. Like we mentioned in Section 6, with the purpose of ensuring the consistence and

repeatability of the experiments, all the proposals presented in this paper were trained and evaluated using 10-fold cross-validation, which also avoids the problem of selecting easy or hard subsets of the dataset. Our experiments are easy to review and analyze because we are publicly sharing our algorithm to obtain the augmented samples as well as our main code and scripts (see Section 8).

This work provides strong evidence that deep but computationally inexpensive networks can achieve near state-of-the-art scores, and thus, be highly effective to solve this problem in the real world, opening up the possibilities to have embedded implementations or mobile applications for automatic bacterial recognition, as it is in this type of devices where maximum benefit can be obtained from systems such as the ones proposed here.

8 Declarations

8.1 Funding

This work did not receive funding of any kind.

8.2 Conflicts of interest/Competing interests

The authors have declared that no conflict of interests or competing interests exist.

8.3 Availability of data and material

In the Data Availability section of their paper [38], authors of the DIBaS dataset provided the following link to access the data:

<http://misztal.edu.pl/software/databases/dibas>

8.4 Code availability

The software requirements, Jupyter notebooks, Python scripts and detailed instructions to repeat our experiments are available on the following GitHub repository: github.com/gallardorafael/EfficientMobileDL_Bacterial

8.5 Authors' contributions

Rafael Gallardo García and Sofía Jarquín Rodríguez contributed to the study conception, design, experimentation and writing of the manuscript. Chemical and biological background of bacterial identification provided in this paper was performed by Sofía Jarquín

Rodríguez. Rafael Gallardo García worked on data processing, system design, programming, training and evaluation related tasks. Beatriz Beltrán Martínez and Carlos Hernández Gracidas provided useful and complete feedback to ensure strong and robust research results, also, they contributed with detailed revisions of the paper structure, content and language-quality. Rodolfo Martínez Torres provided feedback and assistance for some engineering-related tasks. All authors read and approved the final manuscript.

8.6 Ethics approval

As the original data was acquired in [38], none of the authors of this work were exposed to dangerous bacteria, biological material or to dangers of any kind. With the aforementioned, we consider that this section is not applicable to this paper.

8.7 Consent to participate

Not applicable.

8.8 Consent for publication

Not applicable.

References

1. Abd Elaziz, M., Hosny, K.M., Hemedan, A.A., Darwish, M.M.: Improved recognition of bacterial species using novel fractional-order orthogonal descriptors. *Applied Soft Computing* **95**, 106504 (2020)
2. Ahmed, W.M., Bayraktar, B., Bhunia, A.K., Hirleman, E.D., Robinson, J.P., Rajwa, B.: Classification of bacterial contamination using image processing and distributed computing. *IEEE journal of biomedical and health informatics* **17**(1), 232–239 (2012)
3. Aslam, B., Basit, M., Nisar, M.A., Khurshid, M., Rasool, M.H.: Proteomics: technologies and their applications. *Journal of chromatographic science* **55**(2), 182–196 (2017)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255. Ieee (2009)
5. Forero, M.G., Cristóbal, G., Desco, M.: Automatic identification of mycobacterium tuberculosis by gaussian mixture models. *Journal of microscopy* **223**(2), 120–132 (2006)
6. Franco-Duarte, R., Černáková, L., Kadam, S., S Kaushik, K., Salehi, B., Bevilacqua, A., Corbo, M.R., Antolak, H., Dybka-Stępień, K., Leszczewicz, M., et al.: Advances in chemical and biological methods to identify microorganisms—from past to present. *Microorganisms* **7**(5), 130 (2019)
7. Gallardo-García, R., Jarquín-Rodríguez, A., Beltrán-Martínez, B., Martínez, R.: Deep learning for fast identification of bacterial strains in resource constrained devices. In: *Aplicaciones Científicas y Tecnológicas de las Ciencias Computacionales*, pp. 67–78 (2020)
8. Holmberg, M., Gustafsson, F., Hörnsten, E.G., Winquist, F., Nilsson, L.E., Ljung, L., Lundström, I.: Bacteria classification based on feature extraction from sensor data. *Biotechnology techniques* **12**(4), 319–324 (1998)
9. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1314–1324 (2019)
10. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:1602.07360* (2016)
11. Jenkins, S.A., Drucker, D., Keaney, M.G., Ganguli, L.A.: Evaluation of the rapid id 32a system for the identification of bacteroides fragilis and related organisms. *Journal of applied bacteriology* **71**(4), 360–365 (1991)
12. Khalifa, N.E.M., Taha, M.H.N., Hassanien, A.E., Hemedan, A.A.: Deep bacteria: robust deep learning data augmentation design for limited bacterial colony dataset. *International Journal of Reasoning-based Intelligent Systems* **11**(3), 256–264 (2019)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25**, 1097–1105 (2012)
15. Liu, J., Dazzo, F.B., Glagoleva, O., Yu, B., Jain, A.K.: Cmeias: a computer-aided system for the image analysis of bacterial morphotypes in microbial communities. *Microbial Ecology* **41**(3), 173–194 (2001)
16. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017)
17. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 116–131 (2018)
18. MicrobiologySociety: Microbes and disease (2021). URL <https://microbiologysociety.org/why-microbiology-matters/what-is-microbiology/microbes-and-the-human-body/microbes-and-disease.html>
19. Mishra, M., Chauhan, P.: Applications of microscopy in bacteriology. *Microscopy Research* **4**(1), 1–9 (2016)
20. Nasip, Ö.F., Zengin, K.: Deep learning based bacteria classification. In: *2018 2nd international symposium on multidisciplinary studies and innovative technologies (ISMSIT)*, pp. 1–5. IEEE (2018)
21. Nobile, C.J., Johnson, A.D.: Candida albicans biofilms and human disease. *Annual review of microbiology* **69**, 71–92 (2015)
22. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
23. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: *Advances in neural information processing systems*, pp. 8026–8037 (2019)
24. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P.,

- Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011)
25. Perronnin, F., Dance, C.: Fisher kernels on visual vocabularies for image categorization. In: 2007 IEEE conference on computer vision and pattern recognition, pp. 1–8. IEEE (2007)
26. Rujichan, C., Vongserewattana, N., Phasukkit, P.: Bacteria classification using image processing and deep convolutional neural network. In: 2019 12th Biomedical Engineering International Conference (BMEiCON), pp. 1–4. IEEE (2019)
27. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520 (2018)
28. Satoto, B.D., Utomo, M.I., Rulaningtyas, R., Koendhori, E.B.: An auto contrast custom convolutional neural network to identifying gram-negative bacteria. In: 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), pp. 70–75. IEEE (2020)
29. Sauer, S., Kliem, M.: Mass spectrometry tools for the classification and identification of bacteria. *Nature Reviews Microbiology* **8**(1), 74–82 (2010)
30. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
31. Singhal, N., Kumar, M., Kanaujia, P.K., Virdi, J.S.: Maldi-tof mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in microbiology* **6**, 791 (2015)
32. Sommer, C., Gerlich, D.W.: Machine learning in cell biology—teaching computers to recognize phenotypes. *Journal of cell science* **126**(24), 5529–5539 (2013)
33. Spratt, D.A.: Significance of bacterial identification by molecular biology methods. *Endodontic topics* **9**(1), 5–14 (2004)
34. Talo, M.: An automated deep learning approach for bacterial image classification. arXiv preprint arXiv:1912.08765 (2019)
35. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2019)
36. Trattner, S., Greenspan, H., Tepper, G., Abboud, S.: Automatic identification of bacterial types using statistical imaging methods. *IEEE transactions on medical imaging* **23**(7), 807–820 (2004)
37. Váradi, L., Luo, J.L., Hibbs, D.E., Perry, J.D., Anderson, R.J., Orenga, S., Groundwater, P.W.: Methods for the detection and identification of pathogenic bacteria: past, present, and future. *Chemical Society Reviews* **46**(16), 4818–4832 (2017)
38. Zieliński, B., Plichta, A., Misztal, K., Spurek, P., Brzychczy-Włoch, M., Ochońska, D.: Deep learning approach to bacterial colony classification. *PloS one* **12**(9), e0184554 (2017)