Jurnal Elektronika dan Telekomunikasi (JET), Vol. 23, No. 1, August 2023, pp. 47-54
Accredited by KEMDIKBUDRISTEK, Decree No: **158/E/KPT/2021**
doi: 10.55981/jet.533

Bacterial Classification Using Deep Structured Convolutional Neural Network for Low Resource Data

M Faizal Amri[a,*], Asri Rizki Yuliani[b], Dwi Esti Kusumandari[a], Artha Ivonita Simbolon[a], M. Ilham Rizqyawan[a], Ulfah Nadiya[a]

[a]Research Center for Smart Mechatronic,
National Research and Innovation Agency
Jl. Sangkuriang, Dago
Bandung, Indonesia
[b]Research Center for Artificial Intelligence and Cyber Security,
National Research and Innovation Agency
Jl. Sangkuriang, Dago
Bandung, Indonesia

Abstract

Bacterial identification is an essential task in medical disciplines and food hygiene. The characteristics of bacteria can be examined under a microscope using culture techniques. However, traditional clinical laboratory culture methods require considerable work, primarily physical and manual effort. An automated process using deep learning technology has been widely used for increasing accuracy and decreasing working costs. In this paper, our research evaluates different types of existing deep CNN models for bacterial contamination classification when low-resource data are used. They are baseline CNN, GCNN, ResNet, and VGGNet. The performance of CNN models was also compared with the traditional machine learning method, including SIFT+SVM. The performance of the DIBaS dataset and our own collected dataset have been evaluated. The results show that VGGNet achieves the highest accuracy. In addition, data augmentation was performed to inflate the dataset. After fitting the model with augmented data, the results show that the accuracy increases significantly. This improvement is consistent in all models and both datasets.

Keywords: Bacterial classification, Deep learning, Convolutional neural network (CNN), *E-coli*.

I. INTRODUCTION

Bacteria are living microorganisms that are not visible to the naked eye but can be observed under a microscope. While some bacteria benefit our ecosystem, they can also become a carrier of diseases to humans, including infectious ones. It is, therefore, essential to provide accurate identification of living bacteria for a wide range of applications, including clinical diagnosis [1], [2], food production [3], and water quality assessment [4], [5]. The characteristics of bacteria can be examined under a microscope using culture techniques. However, traditional clinical laboratory culture methods require considerable work, primarily physical and manual effort. The recognition and counting of bacteria colonies are mainly performed by the expert. Moreover, some bacteria species belong to the same morphology, making identifying them more difficult. To cope with these problems, an automated identification and classification process can be developed. Machine learning (ML) assisted image processing has been comprehensively used to reduce the workload and improve performance accuracy. Specifically, deep learning (DL) as a subset of ML has achieved incredible success in a number of biomedical applications [6], [7].

Deep learning approaches are able to learn features from large datasets to reach human-level performance. Recently, many studies have been developing an automated system using DL that can assist biologists and related researchers in recognizing microscopic images in large-scale applications, not only bacteria identification but also other living microorganisms. In [8], they provided a systematic review of various ML methods applied for image recognition of four types of microorganisms such as bacteria, algae [9], protozoa [10], and fungi [11], [12]. Other than that, [13] provided a comprehensive survey of digital image processing methods for microorganism counting. The study concluded that there is an urgent need for the automation of bacteria colony detection with high accuracy and acceleration. Most studies usually work with an adequate number of images. However, collecting bacterial data is labor-intensive and error-prone.

To provide a highly accurate system, this paper presents a deep structure convolutional neural network (CNN) for bacteria classification when low resource data have occurred. The utilization of CNN was performed because it has become state-of-the-art for many object detection and classification tasks. The first step in performing our methods was to design a standard CNN architecture with three convolutional layers as a baseline. In the next step, the baseline results were compared with gated CNN and two popular pre-trained models based on

* Corresponding Author.
Email: mfai001@brin.go.id
Received: January 19, 2023 ; Revised: March 31, 2023
Accepted: Apr 26, 2023 ; Published: August 31, 2023

CNN, ResNet, and VGGNet. Not only that, the results also had been compared to SVM classifier to evaluate with a traditional ML method. The remainder of the paper is organized as follows. Section 2 presents related works on bacteria classification and feature extraction. Section 3 explains all the architectures used in this study. Sections 4 and 5 describe our experiment's configuration and dataset. Section 6 shows the results and discussion, the summary of this paper will be concluded in the last part.

II. RELATED WORKS

A. Bacteria Classification

In microbiology, researchers have applied various types of machine learning (ML) techniques to recognize microscopic images from laboratory samples. There have been many works implementing statistical methods [14], artificial neural networks (ANN) [15], and other classifiers [16] to differentiate bacteria species. Among them, deep learning, which is a part of ML methods based on ANN has yielded impressive results on image processing tasks. Several studies have already been done for the recognition of bacteria species using deep neural networks (DNN) [17] and convolution neural networks (CNN) [18], [19]. CNN has recently become one of the most effective yet efficient DL techniques where limited data occurs. Compared to other DL networks, CNN works well on a relatively small amount of data.

In [16], they proposed an automatic detection of tuberculosis bacteria in sputum smear images. The approach involved image segmentation using the threshold method. The segmented regions were used to extract features like shape, granularity, and color. Then, classification was performed using fuzzy membership functions to predict the presence of bacteria. In [20], they classified tuberculosis bacteria into three classes, non-TB, overlapped-TB, and TB, using image segmentation and hybrid multilayer perceptron (MLP). The other work, presented in [17], developed an early detection system of bacterial growth in the form of microscopy images captured inside an agar plate. Then, classification was employed using DNN. The system performance was indicated by the detection of *Escherichia coli* and total coliform bacteria (i.e., *Klebsiella aerogenes* and *Klebsiella pneumoniae* subsp. pneumoniae) in water samples. A study in [18] implemented CNN for the classification and counting of bacteria colonies. The CNN-based classifier was designed for cardinality recognition and outlier rejection. The method was experimentally tested in a real clinical scenario, that is bacteria colonies were extracted from repetitive clinical laboratory culture plates. Moreover, [19] applied a three-dimensional CNN for classifying objects in three-dimensional microscopy images of larval zebrafish intestines as bacteria or non-bacteria. Conventional methods such as random forest and support vector machine (SVM) were also performed as a comparison. The results showed that CNN outperformed both methods.

B. Features

The species of bacteria are identified based on their characteristic features. One of the essential features is the shape of a bacteria cell. Their basic shapes are cylindrical, spherical, and spiral. The second most important features are the shape and size of the bacteria colony. However, different species of bacteria could be in the same morphology. Thus, recognizing bacteria merely on its shape and size would be difficult. Several conventional studies [14], [21] have worked on additional analysis with other microbiological characteristics. In [14], they identified the bacteria species using statistical methods based on geometric features such as circularity, compactness, eccentricity, tortuosity, and length-to-width ratio. In [21], they identified tuberculosis bacteria using additional information on the average color of the images in order to tackle the problem of the bacteria morphology similarities. The other study, presented in [15], applies three ML-based approaches, namely 3σ, K-nearest neighbor (K-NN), and ANN, for bacterial cell classification. The approach worked by applying image segmentation using the global threshold method and extraction of geometric shape features.

The studies mentioned earlier are mostly employed using conventionally trained networks to recognize bacteria for limited species, for instance, only one or five bacteria species. The approach used may identify a specific feature very well. However, it may not be capable of differentiating between species with similar morphology (e.g. shape and size). To recognize more various species in the same morphology, several studies have implemented DL techniques. In [22], the authors proposed using transfer learning-based Inception deep CNN structure to classify bacterial microscopic images. First, deep features were extracted using a pre-trained CNN model. The extracted features were then classified by applying fully-connected layers followed by a softmax layer. Furthermore, a study in [23] employed DL based approach to recognize 33 bacterial species. The approach involved texture feature extraction using scale-invariant feature transform (SIFT) and deep features extraction using three pre-trained CNN models, namely AlexNet, VGG-M, and VGG-VD. In [18], they applied two different ML-based systems, SVM and CNN, for counting bacterial colonies. Handcrafted features were extracted using SVM, while deep features were extracted using CNN. The results presented that the deep learning approach outperformed the handcrafted feature-based one. In a neural network, deep features were extracted relevant to the model's final output by training on thousands of images. Image edge and contour detectors were learned from earlier layers. Meanwhile, the deeper the layers of a network can extract features for more complicated patterns, including textures, shapes, or variations of features learned earlier. DL approach can improve robustness for the recognition of bacteria species.

III. METHODOLOGY

In this study, the performance of some deeply structured convolutional neural network architectures was evaluated. They are standard CNN, gated CNN, ResNet, and VGGNet. These networks are differed based on the depth of the network. To build the methods, it was

started by designing a simple CNN model from scratch to be used as the baseline. Furthermore, to improve the performance, the gating mechanism was added to the architecture of CNN, which became gated CNN. Then two popular pre-trained models (ResNet and VGGNet) were employed. These models were trained on Imagenet and yielded supreme results on a wide range of computer vision tasks. Following the principle of transfer learning, the convolutional layers of the pre-trained models work as a large feature extraction layer. The last fully connected layer, which acted as a classifier, is replaced to suit our problem, that is, to classify bacteria colony data and also apply the fine-tuning approach. Instead of using random initialization, the weights from transferred models as initial weights were used. Some convolutional layers had been frozen to be non-trainable to reduce the model parameters and only update the weights on the unfroze layers. By doing this, it greatly reduces the requirement of large training data and the overfitting problem. In addition, data augmentation had been applied to the dataset in order to improve the neural network training phase. The detailed architecture of each network is described in this section.

A. Baseline Convolutional Neural Network

As a baseline, a simple CNN model with three convolutional layers and two fully-connected layers was applied. The convolutional layer uses 11×11, 5×5, and 3×3 convolution filters with stride 2, respectively. Batch normalization, ReLU activation function, and max pooling were applied after each convolutional layer. The output of the last convolution block was flattened into one feature vector, which became the input to a fully-connected layer consisting of 256 hidden neurons. Dropouts with a rate of 0.5 were applied to the first fully-connected layer to avoid overfitting problems. The output layer consists of neurons whose number corresponds to the bacterial classes. At the last stage, a softmax function was used to obtain the probability of membership for each class. Our CNN model architecture can be seen in Figure 1.

B. Gated Convolutional Neural Network

To improve the baseline model, a gated convolutional neural network (GCNN) was employed, which consists of convolution layers and gating mechanisms. The gate is used to control the information that flows into the succeeding layer. It helps to prevent the vanishing gradient problem during backpropagation [24]. Recently, GCNN has yielded excellent results in an abundance of image-processing tasks [25]. Similar to CNN, GCNN is also one of the most effective algorithms

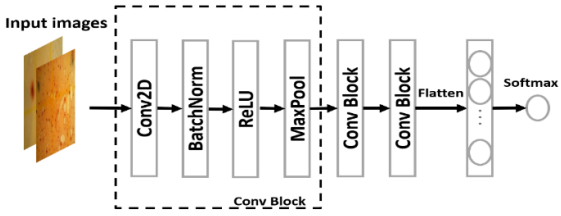


Figure 1. Baseline CNN architecture.

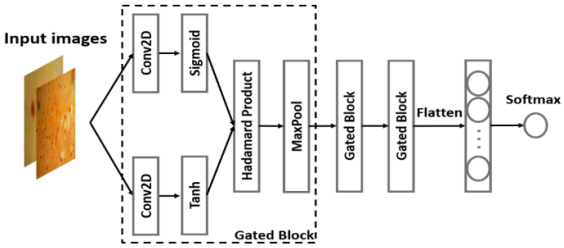


Figure 2. Gated CNN architecture.

in DL techniques because it works on a relatively small amount of data. Compared to DNN, GCNN has a much smaller number of connection weights [26]. Intuitively, the gating mechanism performs convolution operation to the input in two different paths, as shown in Figure 2. Each path might present different information. Following the idea proposed in [27], instead of ReLU activation function, the outputs of convolutional layers are followed by a gated activation unit. The hidden layer after the gated activation unit is expressed in (1):

$$h(X) = tanh(X * W + b) \otimes \sigma(X * V + c) \qquad (1)$$

In this equation, the input image $X \in R^{MxN}$ were put into the network, where M and N represent the dimension of the feature vector of an image at each axis. $Tanh(\cdot)$ function produces the representations, sigmoid function $\sigma(\cdot)$ acts as the gates, and $\otimes$ is the Hadamard product operator. The sigmoid function manages the output of tanh by element-wise multiply operation, $*$ denotes the convolution operator. W and b are the weight matrix and bias respectively. V and c are the kernel weight matrix and bias for the linear gate. The detailed implementation of our GCNN is shown in Figure 2. The network has three gated blocks which consist of three convolutional layers followed by gated structures. Each layer has 11×11, 5×5, and 3×3 convolution filter with stride 2. After the gated activation unit, max pooling over a 3×3 pixel window with stride 2 had been applied. At the final stage, the output of the last gated block was flattened into one feature vector before passing it into softmax classification.

C. ResNet

ResNet, short for Residual Network, is one of popular deep structured convolutional networks that was developed for computer vision tasks [28]. As we know that, deep networks have a notorious problem of vanishing gradient. Model performance often gets saturated when the network goes deeper. Thus, residual networks are introduced as identity shortcut connection

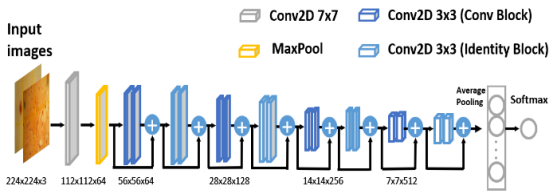


Figure 3. ResNet architecture.

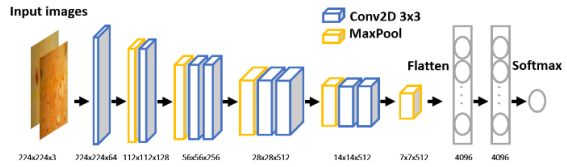


Figure 4. VGGNet architecture.

which skips one or more layers. Using ResNet, an excellent performance could be achieved by training hundreds or thousands of layers. Here, ResNet18 had been implemented. ResNet18 consists of four Conv blocks and four Identity blocks as shown in Figure 3. At last, softmax layer is performed for classification purpose. In this study, the system applied a fine-tuning approach. all convolutional layers had been frozen except the last identity block consisting of two convolutional layers, and the fully-connected layers.

D. VGGNet

VGGNet is another variation of deep-structured convolutional networks proposed by the Visual Geometry Group (VGG) [29]. It can be applied to image classification. VGGNet focuses on the effect of the convolutional neural network depth. It uses smaller filters 3×3 with more depth instead of having large filters (11×11, 5×5) as in AlexNet. Basically, VGGNet is an improvement over AlexNet. There are several VGGNet variants such as VGG11, VGG13, VGG16, and VGG19. Each of VGGNet variants has different configurations in terms of the number of layers and the size of each layer. In our study, VGG11 had been implemented. There are a total of 8 convolutional layers and 3 fully-connected layers in VGG11 architecture. The detailed structure is illustrated in Figure 4. A fine-tuning approach in the VGGNet structure also had been applied. the first seven convolutional layers and the first two fully-connected layers had been frozen. Thus, the model only updates the weights on the last convolutional layer and the last fully-connected layer.

E. Data Augmentation

The bacterial colony dataset used in this study has a limited number of images, compared to the number of trainable parameters of our baseline CNN, GCNN, and pre-trained models. This condition caused a high probability of overfitting. Thus, data augmentation had been implemented to increase the dataset and reduce overfitting. The method focuses on using geometric transformation and Gaussian noise. Geometric transformation is one of the augmentation methods to artificially inflate the dataset by modifying the image position and orientation. Image transformations such as rotation, flipping horizontal and vertical, and the random crop had been applied to add more invariant examples. Data augmentation techniques used in this study are described below:

- Rotation: The rotation scheme rotates an image by a specified angle θ (0 – 180°). Let x and y be the original position at a point (0,0) in an image. Then, $\acute{x}$ and $\acute{y}$ are the new position after θ rotation as shown by the following (2) and (3):

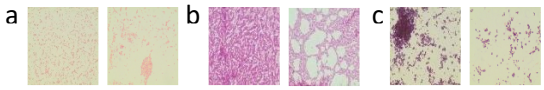


Figure 5. The samples images of three bacteria species from DIBaS dataset: (a) *Escherichia coli*; (b) *Lactobacillus casei*; (c) *Staphylococcus aureus*.

$$\acute{x} = cos\theta * x + sin\theta * y \quad (2)$$

$$\acute{y} = -sin\theta * x + cos\theta * y \quad (3)$$

- Reflection: The reflection scheme flips the image horizontally or vertically. Let x and y be the original image. When the image is flipped horizontally, $\acute{x}$ and $\acute{y}$ be the new position after flipping the image by axis Y as shown in (4). When the image is flipped vertically, x' and y' be the new position after flipping the image by axis X as shown in (5).

$$\begin{bmatrix}\acute{x}\\ \acute{y}\end{bmatrix} = \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} \quad (4)$$

$$\begin{bmatrix}\acute{x}\\ \acute{y}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} \quad (5)$$

- Crop: The crop scheme cut a part of the given image with a desired output size of the crop at a specific location. In this case, a random cropping technique on the original image was used.

IV. Experimental Setup

A. Dataset

In this experiment, the system uses two datasets containing images with bacterial colonies: DIBaS, in which the images contain 33 bacteria species, and our own collected data, in which the images contain *Escherichia coli (E. coli)* detected in water samples taken from different locations.

1) DIBaS

In this paper, a dataset of digital images of bacterial species (DIBaS) published in [5] was used which contains a total of 660 images. Specifically, the data contains images with 33 bacteria species (20 bacteria genera) with 20 images for each class. The dataset consists of colored images that were stained using Gramm's method with the original size of 2048×1532. The sample images of three bacteria species from the dataset are shown in Figure 5. The data was preprocessed by resizing the images into 224×224, the same size used by the pre-trained models, ResNet and VGGNet. For training, the dataset was split into 80% for training, 10% for validation, and 10% for testing.

2) E. coli Dataset

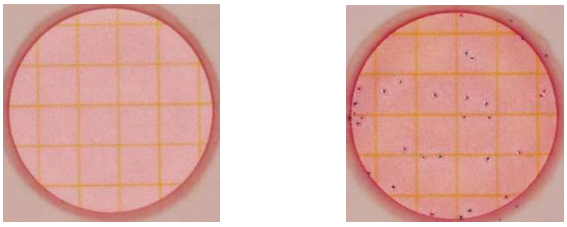
The image data of *Escherichia coli* (*E. coli*) bacteria that were detected in samples from different types of waters had been collected. These waters have three different sources, there are mineral water, sewage, and household water. In total, 262 images of *E. coli* had been collected which comprises three classes: one labeled as a non-bacterial class and two labeled as a bacterial class.

The non-bacterial samples were taken from mineral water. The other two bacterial classes are *E. coli* detected in household water and sewage. To get the datasets, water samples need to be prepared using different treatments before being tested. The preparation was including filtering the samples from unwanted materials and conditioning the water temperature.

The prepared samples were tested using 3M petrifilm in each sample. 3M petrifilm is a medium used to determine the number of colonies of *E. coli* bacteria. The samples should be liquid material or solid material made into a liquid. The use of 3M petrifilm is faster and more practical than detecting *E. coli* bacteria using conventional cultures. Three types of samples were tested on 10 petrifilms for each water, so in the experiments, a total of 30 petrifilms were used. For the next stage, petrifilms were put in a room with good air circulation for 24 to 72 hours to develop bacteria. Figure 6 are examples of petrifilms (a) without *E. coli* bacteria and (b) with *E. coli* bacteria.

The bacteria on petrifilm was captured using our mobile microscope prototype system. The system consists of a Raspberry Pi 4 for device processing, a digital microscope, an SD card, an LCD viewer, and a computer for backup of the data. The data can be stored directly on the SD Card on the Raspberry Pi or computer. The complete system diagram block can be seen in Figure 7.

The data collection was conducted by capturing the image in petrifilm, which in one petrifilm it can take three images from three areas as shown in Figure 8. The captured images were processed and saved into a storage card for further processing step. The total amount of data from the petrifilm of three types of water is shown in Table 1.



(a) (b)

Figure 6. 3M Petrifilm condition (a) without *E. coli* bacteria (b) with *E. coli* bacteria.

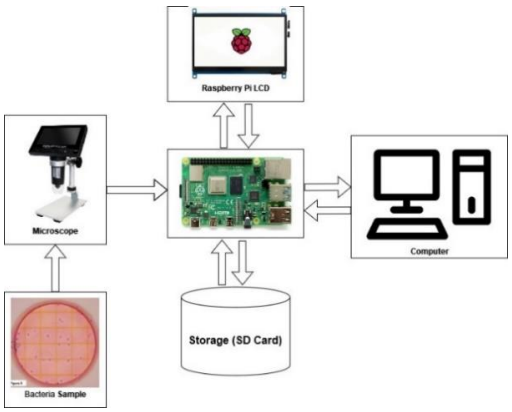


Figure 7. Block diagram system.

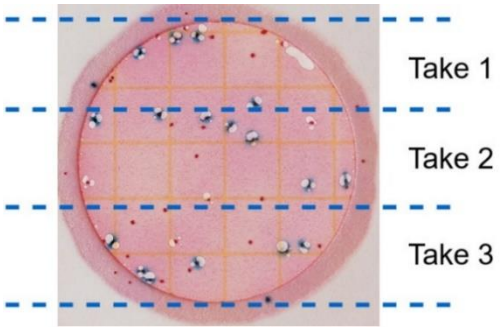
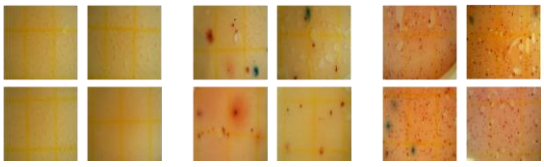


Figure 8. 3M Petrifilm capturing areas.

Figure 9 shows a few samples of our collected bacterial images from different types of water. The augmented methods and the results of all types of water could be seen in Figure 10. Five types of augmented methods had been conducted to the original data to make the amount of each data balanced.

For the experiments, four types of CNN architectures were used to preprocess and train the data. For training, a learning rate of 0.001 and an Adam optimizer were applied. The epochs were set to 50 and the batch size set to 8. Cross-entropy was used as a loss function. Our system trains all the architectures for both DIBaS and our collected *E. coli* dataset. For the pre-trained models, the fine-tuning approach was applied and used pre-trained weights as initialization. Some layers were set to be non-



(a) (b) (c)

Figure 9. Samples of collected images of bacterial (a) mineral water; (b) household water; (c) sewage.

TABLE 1
LIST OF WATER CONTAINING BACTERIAL COLONIES CLASS LABELS WITH THE AMOUNT OF DATA FOR EACH CLASS

Class	Data
Mineral water	30
Household water	123
Sewage	109

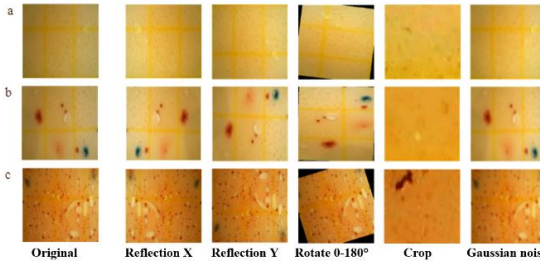


Figure 10. Samples of collected images after various transformations: (a) mineral water; (b) household water; (c) sewage.

trainable to reduce the model parameters and only update the weights to the trainable layers. In addition, SIFT feature extraction was performed together with SVM classifier for comparison. The diagram of our research methods can be seen in Figure 11. The model performance was compared using test accuracy and measured the model size of the training process. All the CNN architectures were implemented in Python using Pytorch packages. The system is trained on Intel i9 10900f, with 32GB memory size and NVidia RTX 3060 12GB GPU.

V. Results and Discussions

1) *DIBaS*

The classification accuracy results from over 33 bacteria species of DIBaS dataset were presented in Table 2. The results showed that data augmentation significantly improves the accuracy of all CNN architecture. The best performance is achieved by VGGNet with an accuracy of 94.44% when training with original data. After fitting the model with augmented data, the accuracy increases significantly to 96.15%. As shown in Table 2, the results with data augmentation (DA) showed consistent improvement over all the CNN architectures. In the beginning, our baseline CNN model performed worse than SIFT+SVM without data augmentation. This might be caused by the small number of images in each class. With DA, the accuracy increases drastically and becomes comparable to GCNN. The number of parameters of ResNet and VGGNet had been reduced to adapt to our task with a small-size dataset. Their parameters have been reduced to 4.73 and 2.49 million respectively. This results in a shorter training time per epoch and reduce overfitting compared to using the original models.

The results of training four types of CNN architectures also were presented, where those were

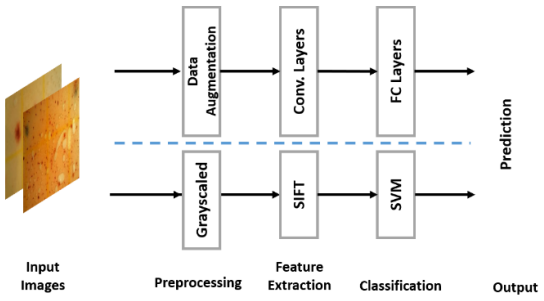


Figure 11. Diagram of research method.

TABLE 2
THE NUMBER OF PARAMETERS (IN MILLIONS) AND THE ACCURACY (%) OF CNN ARCHITECTURES FOR DIBAS

Model	Params	Without DA	With DA
SIFT+SVM	-	81.88	-
CNN	0.79	78.47	92.67
GCNN	1.32	80.56	92.43
ResNet	4.73	91.67	94.11
VGGNet	2.49	**94.44**	**96.15**

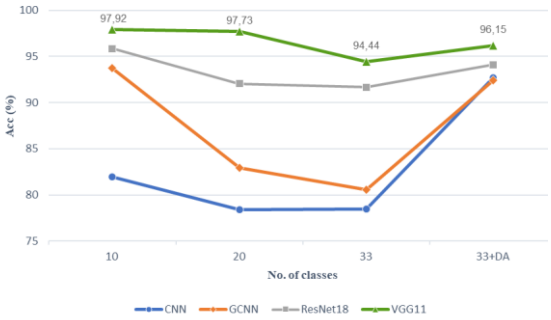


Figure 12. The results of experiments on DIBaS where it was tested on the scalability of CNN architectures.

tested on the scalability of different numbers of classes, as shown in Figure 12. The results showed that accuracy decreases linearly with the number of classes, but it rises again with data augmentation. The decline occurred due to the increasing number of bacterial classes that need to be discriminated against. Since different species of bacteria could be in the same morphology, recognizing more bacterial classes with limited data would be difficult. DL often depends upon large amounts of data to be trained effectively, making it a strong candidate for DA by increasing the number of images for each class.

2) *E. coli Dataset*

Four types of CNN models were evaluated on the three class classifications of *E. coli* dataset, one labeled as a non-bacterial class and two labeled as a bacterial class. The evaluation was conducted by employing the same model architectures as implemented in DIBaS. The number of parameters of pre-trained CNN models had been reduced to adapt to a small size of a dataset. The results show in Table 3. Even without DA, CNN models have achieved high accuracy which is above 90%. This is because the dataset used only contains one bacteria species. In other words, it is not a difficult task to perform a three-class classification. However, as in Table 3, the results showed that the accuracy increases consistently with the increased amount of data. This basically is the premise of data augmentation. The best performance was yielded by VGGNet with an accuracy of 98.12% after performing DA.

Figure 13 shows the confusion matrix of four CNN models on the *E. coli* dataset for the experiments without and with DA. The confusion matrices show that

TABLE 3
THE NUMBER OF PARAMETERS (IN MILLIONS) AND THE ACCURACY (%) OF CNN ARCHITECTURES FOR *E. COLI* DATASET

Model	Params	Without DA	With DA
SIFT+SVM	-	84.91	-
CNN	0.79	93.57	95.62
GCNN	1.32	93.57	95.31
ResNet	4.73	94.64	97.46
VGGNet	2.49	**96.43**	**98.12**

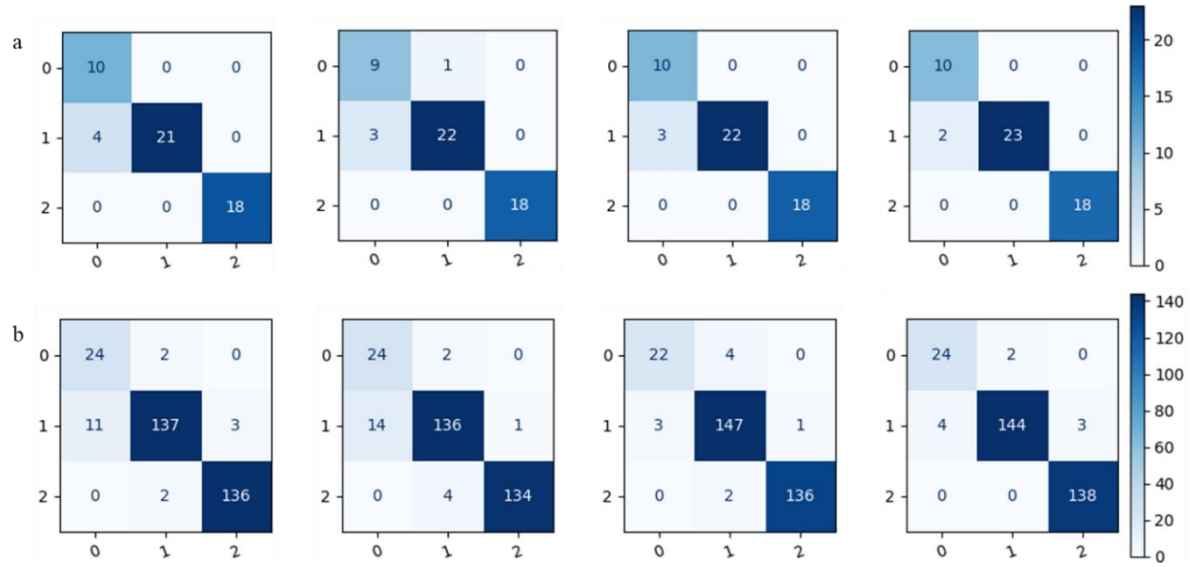


Figure 13. From left to right: confusion matrix for method CNN, GCNN, ResNet, and VGGNet (a) without DA, (b) with DA, where 0: mineral water, 1: household water, 2: sewage.

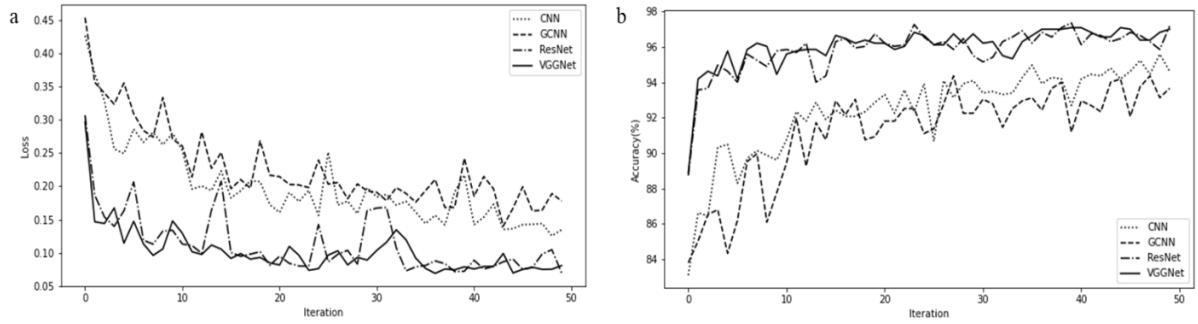


Figure 14. Progression of (a) Training loss, (b) Training accuracy for all CNN architectures.

predicting using pre-trained models improves the prediction result.

Figure 14 shows the progressions of accuracy and loss of training data. The results showed that on *E. coli* dataset, GCNN has the slowest loss to converge and followed by CNN. While ResNet and VGGNet loss convergence are comparable. As expected, GCNN and CNN are the slowest to converge on the progression of accuracy since their losses are also slow to converge.

VI. Conclusion

This paper presents an evaluation of deep learning techniques, especially deep structured convolutional neural networks, for bacterial classification when low-resource data are used. Our research evaluated four types of CNN architectures including baseline CNN, GCNN, ResNet, and VGGNet. Our research also compared the performance of CNN models with traditional machine learning methods such as SIFT+SVM. The models are trained and evaluated using two datasets; DIBaS and our collected *E. coli* data. The results found that VGGNet achieves the highest accuracy. In addition, since the dataset used in our experiment has a small size of data, data augmentation was performed to increase the data. The results show that data augmentation consistently improves the accuracy of all CNN architectures. However, we should say that this study has limitations. First, the methods have not been evaluated with various data. Second, our research have not performed bacterial counting. For future direction, it is substantial to add different types of data collected from different environmental conditions (e.g. river). It is also necessary to count the total number of bacterial colonies grown in water samples for water quality assessment. It is worth noting that automated bacterial classification is not meant to replace the expert diagnosis. However, the implementation could assist the expertise to recognize microscopic images in large-scale applications in a shorter time.

Declarations

Conflict of Interest

The authors declare no conflicts of interest that could potentially influence the research, data analysis, interpretation, or presentation of the findings in this paper..

CRediT Authorship Contribution

MFA: Investigation, Conceptualization, Methodology, Software, Supervision, Visualization, Writing – review & editing; ARY: Software, Investigation, Data curation, Writing-Original draft preparation; DEK, AIS, MIR, UNA: Project administration; Data curation.

Funding

This research is supported by Program Riset Prioritas Nasional (PN) Eks-Deputi Bidang Jasa Ilmiah, LIPI 2021.

REFERENCES

[1] S. Jamal, M. Khubaib, R. Gangwar, S. Grover, A. Grover, and S. E. Hasnain, "Artificial Intelligence and Machine learning based prediction of resistant and susceptible mutations in Mycobacterium tuberculosis," *Sci. Rep.*, vol. 10, no. 1, pp. 1–16, Mar. 2020, doi: 10.1038/s41598-020-62368-2.

[2] A. K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna, and J. J. P. C. Rodrigues, "Identifying pneumonia in chest X-rays: A deep learning approach," *Measurement*, vol. 145, pp. 511–518, Oct. 2019, doi: 10.1016/J.MEASUREMENT.2019.05.076.

[3] M. Deckers *et al.*, "Strategy for the identification of micro-organisms producing food and feed products: Bacteria producing food enzymes as study case," *Food Chem.*, vol. 305, Feb. 2020, doi: 10.1016/J.FOODCHEM.2019.125431.

[4] S. A. Nehal, D. Roy, M. Devi, and T. Srinivas, "Highly sensitive lab-on-chip with deep learning AI for detection of bacteria in water," *International Journal of Information Technology (Singapore)*, vol. 12, no. 2, pp. 495–501, Jun. 2020, doi: 10.1007/S41870-019-00363-1/METRICS.

[5] F. M. Khan, R. Gupta, and S. Sekhri, "Automated Bacteria Colony Counting on Agar Plates Using Machine Learning," *J. Inf. Technol.*, vol. 147, no. 12, p. 04021066, Oct. 2021, doi: 10.1061/(ASCE)EE.1943-7870.0001948.

[6] X. Liu *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *Lancet Digit. Health*, vol. 1, no. 6, pp. e271–e297, Oct. 2019, doi: 10.1016/S2589-7500(19)30123-2.

[7] A. B. Levine, C. Schlosser, J. Grewal, R. Coope, S. J. M. Jones, and S. Yip, "Rise of the Machines: Advances in Deep Learning for Cancer Diagnosis," *Trends Cancer*, vol. 5, no. 3, pp. 157–169, Mar. 2019, doi: 10.1016/J.TRECAN.2019.02.002.

[8] P. Rani, S. Kotwal, J. Manhas, V. Sharma, and S. Sharma, "Machine Learning and Deep Learning Based Computational Approaches in Automatic Microorganisms Image Recognition: Methodologies, Challenges, and Developments," *Arch. Comput. Methods Eng.*, vol. 29, no. 3, pp. 1801–1837, Aug. 2021, doi: 10.1007/S11831-021-09639-X.

[9] J. L. Deglint, C. Jin, A. Chao, and A. Wong, "The Feasibility of Automated Identification of Six Algae Types Using Feed-Forward Neural Networks and Fluorescence-Based Spectral-Morphological Features," *IEEE Access*, vol. 7, pp. 7041–7053, 2019, doi: 10.1109/ACCESS.2018.2889017.

[10] M. A. E. Abdalla and H. Seker, "Recognition of protozoan parasites from microscopic images: Eimeria species in chickens and rabbits as a case study," in *Conf Proc. IEEE Eng. Med. Biol. Soc. (EMBS)*, Jul. 2017, pp. 1517–1520, doi: 10.1109/EMBC.2017.8037124.

[11] B. Zieliski, A. Sroka-Oleksiak, D. Rymarczyk, A. Piekarczyk, and M. Brzychczy-Woch, "Deep learning approach to describe and classify fungi microscopic images," *PLoS One*, vol. 15, no. 6, pp. 1-16, Jun. 2020, doi: 10.1371/JOURNAL.PONE.0234806.

[12] H. Ma *et al.*, "Deep convolutional neural network: a novel approach for the detection of Aspergillus fungi via stereomicroscopy," *J. Microbiol.*, vol. 59, no. 6, pp. 563–572, Jun. 2021, doi: 10.1007/S12275-021-1013-Z/METRICS.

[13] J. Zhang *et al.*, "A comprehensive review of image analysis methods for microorganism counting: from classical image processing to deep learning approaches," *Artif. Intell. Rev.*, vol. 55, no. 4, pp. 2875–2944, Apr. 2022, doi: 10.1007/S10462-021-10082-4/FIGURES/55.

[14] P. Hiremath and P. Bannigidad, "Automated gram-staining characterization of digital bacterial cell images," in *Proc. IEEE Int'l. Conf. on Signal and Image Processing ICSIP*, no. August 2009, 2009.

[15] P. S. 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, doi: 10.1504/IJCBDD.2011.041414.

[16] P. Ghosh, D. Bhattacharjee, and M. Nasipuri, "A hybrid approach to diagnosis of tuberculosis from sputum," in *Int. Conf. Electr. Eng. Inform. Commun. Technol. (ICEEOT) 2016*, pp. 771–776, Nov. 2016, doi: 10.1109/ICEEOT.2016.7754790.

[17] 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, Jul. 2020, doi: 10.1038/s41377-020-00358-9.

[18] A. Ferrari, S. Lombardi, and A. Signoroni, "Bacterial colony counting with Convolutional Neural Networks in Digital Microbiology Imaging," *Pattern Recognit.*, vol. 61, pp. 629–640, Jan. 2017, doi: 10.1016/J.PATCOG.2016.07.016.

[19] 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, p. e1006628, Dec. 2018, doi: 10.1371/JOURNAL.PCBI.1006628.

[20] M. K. Osman, M. Y. Mashor, and H. Jaafar, "Hybrid multilayered perceptron network trained by modified recursive prediction error-extreme learning machine for tuberculosis bacilli detection," in *5th Kuala Lumpur International Conference on Biomedical Engineering 2011, IFMBE Proc.*, vol. 35, Springer, Berlin, Heidelberg, 2011, pp. 667–673, doi: 10.1007/978-3-642-21729-6_163/COVER.

[21] M. G. Forero, G. Cristóbal, and M. Desco, "Automatic identification of Mycobacterium tuberculosis by Gaussian mixture models," *J. Microsc.*, vol. 223, no. 2, pp. 120–132, Aug. 2006, doi: 10.1111/J.1365-2818.2006.01610.X.

[22] M. F. Wahid, T. Ahmed, and M. A. Habib, "Classification of microscopic images of bacteria using deep convolutional neural network," in *The 10th Int. Conf. Electr. Eng. Inform. Commun. Technol. (ICECE) 2018*, pp. 217–220, Feb. 2019, doi: 10.1109/ICECE.2018.8636750.

[23] B. Zieliński, A. Plichta, K. Misztal, P. Spurek, M. Brzychczy-Włoch, and D. Ochońska, "Deep learning approach to bacterial colony classification," *PLoS One*, vol. 12, no. 9, p. e0184554, Sep. 2017, doi: 10.1371/JOURNAL.PONE.0184554.

[24] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language Modeling with Gated Convolutional Networks," in *Proceedings of the 34th International Conference on Machine Learning*, Aug. 2017, pp. 933-941, doi: 10.5555/3305381.3305478.

[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press. [Online]: https://www.scirp.org/(S(czeh2tfqw2orz553k1w0r45))/reference/referencespapers.aspx?referenceid=2859809 (accessed Jan. 14, 2023).

[26] Y. Bengio, *Learning Deep Architectures for AI*. Foundations and Trends® in Machine Learning, vol. 2, no. 1, pp. 1–27, Jan. 2009, doi: 10.1561/2200000006.

[27] P. Lu, X. J. Peng, J. B. Yu, and X. Peng, "Gated CNN for visual quality assessment based on color perception," *Signal Process. Image Commun.*, vol. 72, pp. 105–112, Mar. 2019, doi: 10.1016/J.IMAGE.2018.12.007.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 770–778, Jun. 2016, doi: 10.1109/CVPR.2016.90.

[29] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *3rd International Conference on Learning Representations (ICLR) 2015*, Conference Track Proceedings, May 2015, doi: 10.48550/arxiv.1409.1556.