Bacterial Image Classification

Soni R

Department of Masters Computer Application,
TKM College Of Engineering, Kollam, Kerala, India,
Email: {sonisebastian99, Soni}@gmail.com

ABSTRACT— Numerous uses for bacteria image classification include medical diagnosis, microbiological research, quality control, environmental monitoring, and biodefense. Traditional methods of bacteria classification involve manual observation and analysis by experts, which can be time-consuming and prone to errors. The method of recognising and categorising various species of bacteria using microscopic photographs is known as "bacterial image classification To process the image data and extract pertinent traits that can be used to categorise various species of bacteria, it uses machine learning techniques. In the discipline of deep learning, which is a branch of machine learning, high-level features are extracted from input data, such as photographs, using neural networks with several layers. Convolutional Neural Networks (CNNs), a type of deep learning model, are used to extract pertinent characteristics from images. Convolutional filters are used to the input picture in CNNs in order to extract spatial characteristics from images. Using two CNN models, ResNet-50 and Resnet-34, as well as ANN and LSTM, this application focuses on categorising photos of bacteria. Studies have demonstrated that these methods produce the best outcomes.

I. INTRODUCTION

Millions of people have died as a result of some of the deadliest diseases in recorded human history, which are caused by single-celled germs called bacteria. Bacterial images are grouped into many classes according to their features, such as shape, size, and colour. This process is known as "bacterial image classification." The value of bacterial image categorization rests in its capacity to aid researchers and scientists in comprehending the traits and behaviour of various bacterial species. Modern approaches for identifying bacterial images mainly rely on cutting-edge technology including microscopy, digital imaging, and deep learning algorithms. Bacterial image classification is a relatively young area of study.

Although the old techniques for classifying bacteria were far less accurate than those used today, they still marked significant advancements in the field of microbiology and offered insightful information about the origins of bacterial infections. The accuracy and speed of diagnosis, which are crucial for the prompt treatment and prevention of infectious diseases, could be considerably improved by the use of deep learning algorithms for bacterial image categorization. This work intends to give a summary of the use of deep learning in the classification of bacterial images . The obstacles of classifying bacteria from images, such as changes in bacterial morphology and lighting conditions, are the main topics of this research . Additionally, a number of deep learning models and architectures that have been used to classify bacterial images are presented, along with criteria for measuring their effectiveness . A faster and more accurate identification of bacterial species is one of the possible advantages of developing automated bacterial image categorization systems.

Because bacteria are minute organisms, they cannot be seen with the naked eye. They usually have a diameter of between 0.2 and 10 micrometres, while some species might be larger or smaller. Numerous different forms and structures are possible for bacteria . Deep learning techniques are used as the basis for classification. Image classification and other computer vision tasks use a type of machine learning methods called deep learning algorithms. These artificial neural network-based algorithms are made to analyse and categorise massive volumes of data, such movies or photos. Deep neural networks of the sort known as convolutional neural networks (CNNs) are frequently employed in image categorization applications. Using Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and Artificial Neural Networks (ANNs), this research focuses on the classification of bacterial images.

- In this study, two CNN model variations, ResNet-34 and ResNet-50, are used for the categorization of bacterial images. This study's objectives are:

- using the DiBas dataset to train the two CNN models ResNet-34 and ResNet-50.

- Twenty categories of bacteria have been assigned, and the DiBas dataset has been used to train LSTM and ANN.

II. RELATED WORKS

Single-celled microbes known as bacteria can be found in practically any environment on the planet, including soil, water, and inside other living things . Automated bacterial image classification is the technique of classifying bacterial photographs into several groups based on their visual characteristics using computer vision and machine learning algorithms. A huge dataset of labelled bacterial pictures is needed to train these algorithms. This dataset is used to train the algorithms to correctly identify and categorise the various bacterial pictures . Convolutional neural networks (CNNs), for example, are a common deep learning approach used to obtain high classification accuracy.

In [1], authors employed the Tensorflow machine learning framework with Python programming to achieve training and validation accuracy of more than 75%. Each of the two types of datasets for S. aureus and L. delbruekii has more than 400 sample pictures. Each dataset has been divided into training and test subsets using percentages of 80 and 20, respectively. The purpose of this study was to create a system capable of instantly and automatically classifying bacteria images in standard resolution for mobile and tablet use cases. In paper [2], the author employed a decision tree to cover automatic recognition of 20 bacterial genera and species and got a sensitivity result of approximately 95.94%. The author's original contribution to the work consisted of the choice to conduct the process of identifying bacteria using the simultaneous analysis of the following physical characteristics of bacterial cells: colour, size, shape, number of clusters, cluster shape, as well as density and distribution of the cells. It covers the automatic recognition of twenty bacterial genera and species.

The writers of paper [3] used Deep Convolutional Neural Network (DCNN) and Support Vector Machine (SVM) were combined to create an AI system that performed the task with both train and validation accuracy more than 95%. Eighty percent of the image-dataset, or more than 800 picture samples of seven different bacterial species, have been used to modify and retrain the Inception V3 DCNN architecture.

In work [4], authors perform probabilistic inference using the Naive Bayes classifier based on the input descriptors. The accuracy was roughly 81.08%, according to the authors, who used 64 images for each class of bacteria as the training set and 222 images total, including the three classes of bacteria and other random images like people and aeroplanes , as the test set . Twenty photos for each of the 10 different bacterial species were taken from DIBaS and used in publication [5]'s 200 bacterial images . It achieves accuracy of roughly 97% by using an image preprocessing step that involves histogram equalization , feature extraction using a Bag-of-words model, and classification using a Support Vector Machine (SVM) . To provide the best execution features, the BoW model is applied to the upgraded photos. Effective results show that the proposed algorithm has a great deal of power when employing the BoW model for classifying bacterial picture. In a paper[6], an automated counting system based on image processing was utilised to count the number of bacterial colonies that grow in microbiology lab petri dishes . The pattern detector and counter are the two major components of the system. To create a system that is scalable and reusable, the counter is kept as a separate module. The part of the pattern detector can be changed. In a paper[7] , The study conducted a thorough and critical review of the many Machine Learning approaches used in the field of bacterial classification, along with their shortcomings and potential applications in the future. Additionally, there are many chances and difficulties in applying these methods in the relevant field. are also offered to give academics working in this field a deep understanding.

TABLE I
SOME OF THE BACTERIA SPECIES FROM DIBaS DATASET

Species Name	Number
Veionella	22
Lactobacillus johnsonii	20
Lactobacillus gasseri	20
Proteus	20
Neisseria gonorrhoeae	23
Escherichia coli	20
Staphylococcus epidermidis	20

A. DIBaS dataset

Each of the 33 bacteria species in the Digital photos of Bacteria Species dataset (DIBaS) has an average of 20 photos. The Gramms method was used to stain each and every sample. The photos were shot with a SC30 camera attached to an Olympus CX31 Upright Biological Microscope (Olympus Corporation, Japan). They were assessed using a 100x objective submerged in oil (Japanese Nikon 50). The DIBaS dataset is open to other researchers. Additionally, this dataset underwent blob normalisation preprocessing, which is unique for this kind of image .This change aimed to lessen the weight that colour has while selecting choices in a bias-aware challenge.

Finally, a random flipping, horizontal or vertical translation, or augmentation method has been used. At the conclusion of the pre-processing procedure, the total number of tiny bacterial photos climbed from 400 to 3,43,000, with at least 17,000 images per species. We check for over-fitting, under-fitting, and the ideal dataset split after using Python to divide the data into training and test datasets. The final training and validation data set is then obtained by checking for cross validation using the K-Folds Cross Validation method . After researching the most prevalent bacterial diseases, we selected 20 species from each class to classify that caused the most.

III. METHODOLOGY AND ALGORITHM

The models is created entirely from scratch , which is based on the classification of bacterial species. The dataset is trained and tested using the CNN models ResNet-34, ResNet-50, LSTM, and ANN.
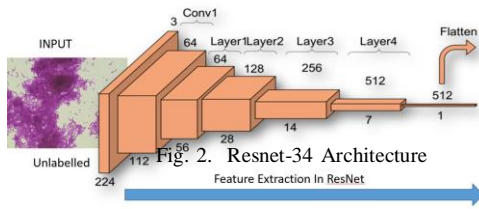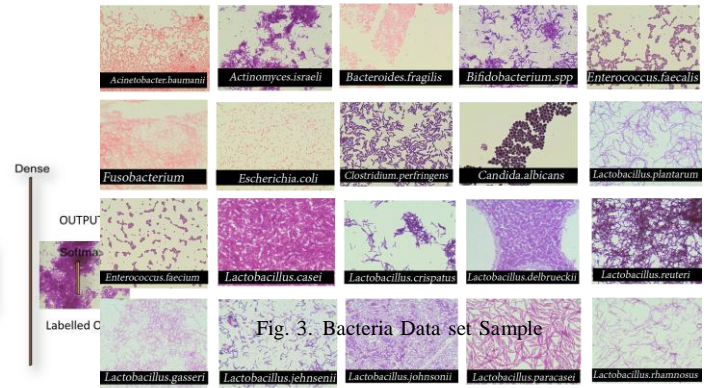

Fig. 2. Resnet-34 Architecture

Fig. 3. Bacteria Data set Sample

I. ResNet-50

ResNet-50 is a convolutional neural network design of 50 layers, including fully connected, convolutional, and pooling layer. It is a deep neural network that can identify features in pictures. Each block in the ResNet-50 design is made up of numerous convolutional layers and a residual connection. A convolutional layer comes initially, followed by a max pooling. The subsequent four blocks each contain two convolutional layers, a residual connection, and a number of residual units. The last block has a fully connected layer that outputs the class probabilities and a global average pooling layer.

II. ResNet-34

Because ResNet-34 has fewer parameters than ResNet-50, it is a more computationally efficient architecture. Nevertheless, it is still capable of accurately classifying images. ResNet-34 has been shown to outperform previous shallow architectures like VGG-19 while using less parameters.

III. LSTM

Long Short-Term Memory, sometimes known as LSTM, is a sort of recurrent neural network architecture created to address the vanishing gradient issue in conventional RNNs. For processing sequential data, such as speech, text, and time series data, LSTMs are highly successful.

IV. ANN

Artificial Neural Networks, often known as ANNs, are a class of machine learning algorithms that take their cues from the structure and operation of the human brain. An artificial neural network (ANN) is made up of interconnected nodes that are arranged into layers. The output layer generates the output prediction, and the input layer receives the input data. The input and output layers are separated by the hidden layers, which are in charge of processing the input data to create the output.

DATA COLLECTION AND PRE-PROCESSING

Each of the 33 bacteria species in the Digital photos of Bacteria Species dataset (DIBaS) has an average of 20 photos. It was gathered by the Jagiellonian University's Microbiology Chair in Krakow, Poland. The DIBaS dataset was used to gather the bacterium photos.

Image resizing, image grey-scaling, and picture normalisation are all included in the pre-processing phases .Since machine learning models can be trained quickly on smaller images, image scaling is an essential stage in the image processing process .The size of the raw photos can change, which could be problematic for many deep learning models.

With OpenCV, the photos are currently grayscaled . The method of gray-scaling involves changing an image from other colour spaces to various shades of grey. It aids in dimension reduction, lessens the complexity of the model, and aids in the detection of edges and corners in the images .The photos have been normalised and are now ready for processing.

IV.TRAINING AND TESTING

The data is divided randomly in the ratio 8:2 after pre-processing. In other words, the model is trained using 80% of the data, then tested using the remaining 20%. After selecting the epochs, the model is now trained.
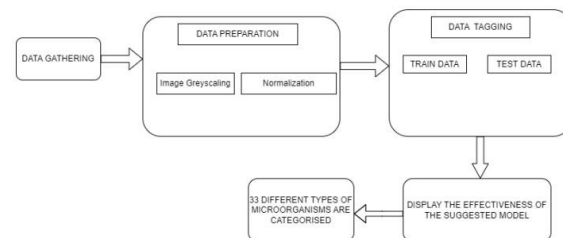
V. RESULTS AND DISCUSSIONS

For the goal of classifying bacteria, two CNN models—ResNet-34 and ResNet-50—have been deployed. There attain 100 percent accuracy for 20 different bacterial species after using the ResNet-50 as our training model. There trained ResNet-50, with the same dataset, which indeed proved to be a better solution for the classification purposes. This paper implements ResNet- 34 with 33 different classes and the accuracy is 82.79 percent which can be treated as remarkable accuracy values. This paper implements LSTM with 33 different classes and the accuracy is 95.83 percent which can be treated as remarkable accuracy values. This paper implements ANN with 33 different classes and the accuracy is 69.23 percent which can be treated as remarkable accuracy values.
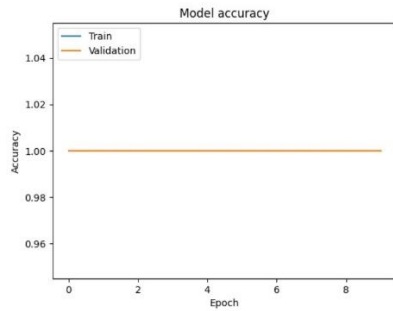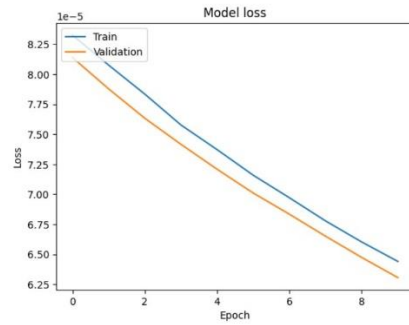
Fig.1 Accuracy of ResNet-50

Fig.5 Final epoch values of ANN

Fig.2 Loss rate of ResNet-50
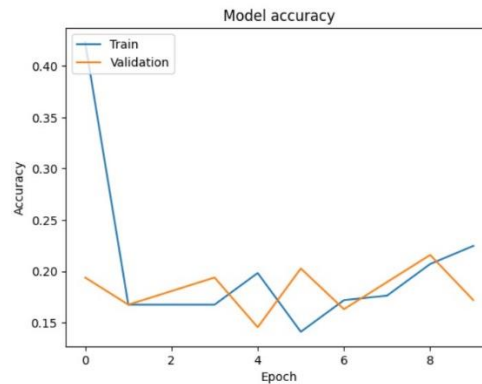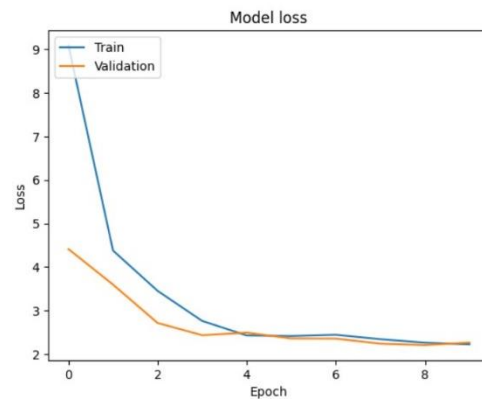
Fig.6 Accuracy of LSTM

Fig.3 final epoch values of ResNet-50

Fig.7 Loss rate of LSTM

Fig.4 Loss rate of ANN

Fig.6 Final epoch values of LSTM

SL NO	MODEL	ACCURACY
1	ResNet-50	100%
2	ResNet-34	82.79%
3	ANN	69%
4	LSTM	95%

VI . CONCLUSION

This article describes a computer-aided method for classifying bacteria to diagnose diseases. For this categorization issue, deep learning neural networks were utilised. Deep residual networks, or different ResNet CNN models as ResNet- 34 and ResNet-50 models, as well as ANN and LSTM, were employed for feature extraction and classification. Using the ResNet-50 CNN model, this study accurately categorised 33 categories of bacteria with a 100% accuracy rate.

REFERENCES

[1] Bacteria Classification Using Image Processing and Deep Learning Treesukon Treebupachatsakul; Suvit Poomrittigul 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC) Year: 2019 | Conference Paper | Publisher: IEEE.

[2] Methods of Classification of the Genera and Species of Bacteria Using Decision Tree January 2020Journal of Telecommunications and Information Technology 4(2019):74-82 DOI:10.26636/jtit.2019.137419.

[3] Combining Deep Convolutional Neural Network with Support Vector Machine to Classify Microscopic Bacteria Images Tasnim Ahmed ; Md. Ferdous Wahid; Md. Jahid Hasan 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)Year: 2019 | Conference Paper | Publisher: IEEE.

[4] Bacteria Identification from Microscopic Morphology using Nave Bayes. Mohamad, Noor Amaleena, et al. "Bacteria identification from microscopic morphology: a survey." International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI) 3.1 (2014): 2319-1015.

[5] Automated classification of Bacterial Images extracted from Digital Microscope via Bag of Words Model 2018 9th Cairo; international biomedical engineering conference (CIBEC). IEEE, 2018.Basma A. Mohamed; Heba M. Afify,2018.

[6] An Image-Processing Based Automated Bacteria Colony Counter Huseyin Ates; Omer Nezih Gerek 2009 24th International Symposium on Computer and Information Sciences Year: 2009 | Conference Paper | Publisher: IEEE.

[7] Automated bacterial classifications using machine learning based computational techniques; Architectures, challenges and open research issues." Archives of Computational Methods in Engineering (2022): 1-22.