Bacteria Classification using Deep Residual Network

A Project Report Submitted

in Partial Fulfilment of the Requirements

for the Degree of

B.Tech (Hon)

in

Computer Science

by

Tumun Shaily
(Roll No. 62)

Indian Institute of Information Technology Kottayam

to

DEPARTMENT OF COMPUTER SCIENCE
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
KOTTAYAM-686635, INDIA

DECLARATION

I, **Tumun Shaily** (**Roll No: 2017BAC0062**), hereby declare that, this report entitled **"Bacteria Classification using Deep Residual Network"** submitted to Indian Institute of Information Technology Kottayam towards partial requirement of **Bachelor of Technology(Hon)** in **Computer Science** is an original work carried out by me under the supervision of **Dr.Kala.S** and has not formed the basis for the award of any degree or diploma, in this or any other institution or university. I have sincerely tried to uphold the academic ethics and honesty. Whenever an external information or statement or result is used then, that have been duly acknowledged and cited.

Kottayam-686635 **Tumun Shaily**

April 2019

CERTIFICATE

This is to certify that the work contained in this project report entitled **"Bacteria Classification using Deep ResidualNetwork"** submitted by **Tumun Shaily** (**Roll No: 2017BAC0062**) to Indian Institute of Information Technology Kottayam towards partial requirement of **Bachelor of Technology(Hon)** in **INDIAN INSTITUTE OF INFORMATION TECHNOLOGY KOTTAYAM** has been carried out by him under my supervision and that it has not been submitted elsewhere for the award of any degree.

Kottayam-686635   (Dr.Kala.S)

April 2019   Project Supervisor

ABSTRACT

Classification of microbe are of great importance in medical science for accurate diagnosis. Many IT industries are using computer vision as their main tool for classification purpose. And the trend continue with the Medical equipment's for achieving better results. In this report We have used Deep Residual Network(ANN) for bacteria classification. For selection of optimum ANN based model we have compared two different ANN model in this project i.e ResNet34 and Resnet50 and later check it efficiency with the previous work of other authors in this same field.

Index Terms—Microbes Classification, Machine learning, Computer vision,Fast.ai, ,Resnet34,ResNet-50,Artificial Neural Networks

Contents

Chapter 1

INTRODUCTION

The pandemic of COVID-19 and horrific Incidents like Influenza (1847-1848),Bubonic plague(18551860),Cholera(1817–1824)and more have showed the need of better medical Equipment's and faster testing. More lives were lost due to lack of proper diagnosis. Many of these plagues pandemic were caused by microbes specifically bacteria's and Virus. After doing a research on these incidents one of the major problem was identification of the microorganism family. In this report we have focused on Classifying bacteria using ANN-model called ResNet34 and Resnet50.1.1General Problem with classifications of microbes and current methods the size of a bacteria fall on the scale from 0.2 to 20 micrometres.Bacteria size are nearly in average so,

Electron microscopes are used to identify or classify them. The microscopic observation and different types of chemical testing is the current approach used in practice for microbe detection and requires expensive equipment.These processes need more time under human observation and hence are slow. When there are lot of samples for testing, this approach gets slower and might be more expensive. In today's scenario, IT sectors are using image processing and many medical labs are also integrating their equipment's with it. In paper[2],authors use Naive Bayes classifier to identify bacteria from microscopic morphology. There are several ways for automatic recognition of bacteria species i.e statistical methods [5], artificial neural networks [6], [7]or other machine learning classifiers [8].In this Report, we present the Resnet34(ANN model) and ResNet50 to solve the microscopic bacteria image classification problem. As our problem contains many different classes of bacteria we are using Resdual Neural Network for more accuracy in classifying different species. ResNet architecture is shown in Fig. 5. After analysing other results from different authors i.e, CNN-KNN (96.4 percentage) and CNN–Naıve Bayes(95.5 percentage) at their best setting we

INPUT
LAYER

HIDDEN
LAYER

OUTPUT
LAYER

X1

X2

X3
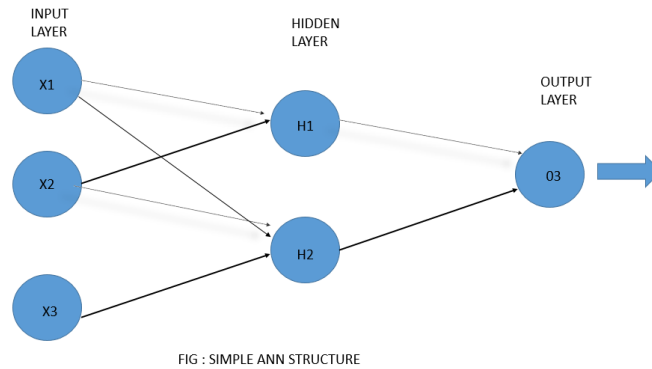
H1

H2

O3

FIG : SIMPLE ANN STRUCTURE

Figure 1.1: General ANN Architecture

wanted to test for the ANN models.

1.0.1 ABOUT RESNET MODEL

Resnet is a type of Artificial Neural network. The basic diagram for ANN can be see Fig (1).Just Like a CNN, ANN were also inspired by the neural architecture of Animal brains. They learn to operate through examples i.e learn features from the dataset(Example) and use it to perform task (Image recognition , Pattern Identification, speech recognition and much more). The key difference between ANN and CNN is that in ANN all the layers are Fully connected whereas in CNN only the last layer is fully connected.

After the evolution of ANN, researchers are using them in

3

task specific purpose. That's why they become more efficient in computer vision.Hence , explain my approach for using ANN model.

One important Feature of ResNet is 'Skip Connection, It mean we can add the output of one layers to other layer without passing from the adjacent layer as can be seen in the Fig 1.2. This results in the ability to train much deeper networks than what was previously possible.

Below is the General Output function:

Y = F(x,Wi+Ws*X)

In the above equation the f(x) and X may have different dimensions Whereas Wi term can be interpreted as 1*1 Convolutions layers.
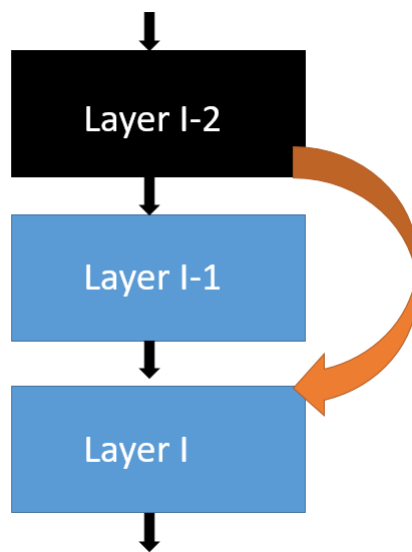
Figure 1.2: General ResNet architecture

Chapter 2

SYSTEM SETUP AND DATASET

2.1 ABOUT DATASET AND PRE-PROCESSING

We use python for cropping the bacteria image (224 pixel size) and then we manually analyse the result images. Finally, an augmentation process has been applied which involves random flipping, horizontal or vertical translation. At the end of the preprocessing step, total number of microscopic bacteria image was increased from 240 to 2200 where each species contains minimum 175 images. After the data is divided into Training and Test dataset using python we check about overfitting , underfitting and just right data-set split.Then we check for cross valida-

Acinetobacter.baumanii   *Actinomyces.israeli*   *Bifidobacterium.*   *Bacteroides.fragilis*
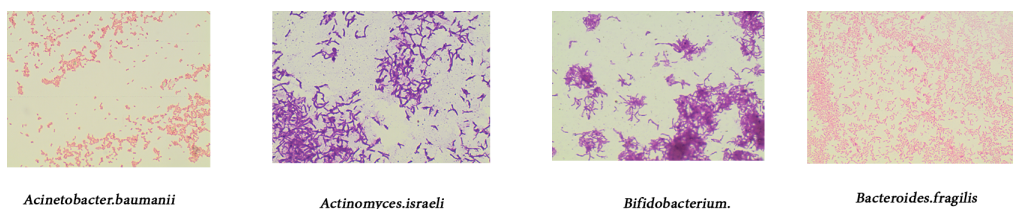
Figure 2.1: Bacteria Data set Sample

tion using K-Folds Cross Validation method which give us the final training and Test data set.

We have chosen 12 such species of Bacteria for classification which caused more disease and were common in most of testing cases.

DIBaS dataset: Bacteria images were collected from DIBaS Data base. Digital Images of Bacteria Species dataset (DIBaS) contains 33 bacteria species with 20 images for each of them. It was collected by the Chair of Microbiology of the Jagiellonian University in Krakow, Poland. All of the samples were stained using the Gramm's method. The images were taken with Olympus CX31 Upright Biological Microscope equipped with a SC30 camera (Olympus Corporation, Japan). They were evaluated using a 100 times objective under oil-immersion (Nikon50, Japan). DIBaS dataset is publicly available to other researchers.

2.2 SYSTEM SET-UP

For ResNet-34 Training Purpose configuration of System : Jupyter notebook with fasti.ai v1 library ResNet-34 model chipset :I5-7th @ 2.71GHz clock speed ram : 16GB Intel inbuilt Graphic card In background we are using Scikit-learn as framework integrated with python modules. For all the data-set analysis and feature relation we have used Matplot (Python). All data set is refine to 224224 pixel size using python script for easy feature extraction. For ResNet-50 training purpose configuration of System : Jupyter notebook with fasti.ai v1 library ResNet-50 model Google TPU.

2.2.1 PROPOSED ARCITECTURE AND SYSTEM SET-UP

We have Used Resnet34 and ResNet-50 as a feature selector and classifier for different scenario. In fig[5], a typical ResNet model is shown. These models bring much better classification performance compared to self made extraction models in image processing. The architecture for feature model is shown in fig[4].
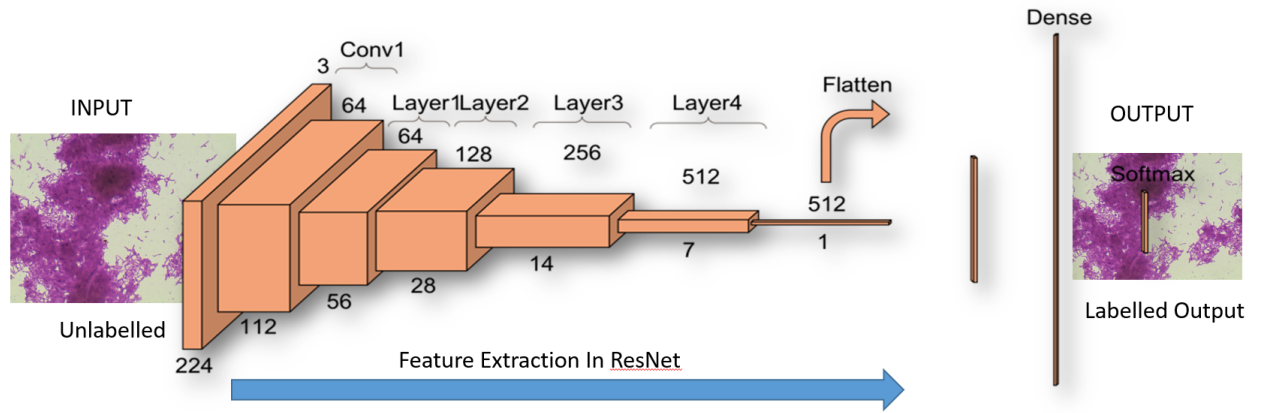
Figure 2.2: Resnet Architecture

2.2.2 CLASSIFICATION AND TRAINING

epoch	train_loss	valid_loss	error_rate	time
0	0.045953	0.016592	0.000000	00:53
1	0.065039	0.004129	0.000000	00:45
2	0.046578	0.001334	0.000000	00:41
3	0.063296	0.000964	0.000000	00:40
4	0.075565	0.001010	0.000000	00:42
5	0.064691	0.001349	0.000000	00:42
6	0.060833	0.001224	0.000000	00:42
7	0.090199	0.001242	0.000000	00:44

Figure 2.3: Final Epoch cycle values of ResNet-34 for 8 classes

```
learn.fit_one_cycle(4)
```

epoch	train_loss	valid_loss	error_rate	time
0	1.929384	1.600818	0.571429	00:47
1	1.172990	0.505385	0.142857	00:49
2	0.808980	0.268321	0.071429	00:48
3	0.632564	0.068968	0.000000	00:47

Figure 2.4: Final Epoch cycle values of ResNet-50 for 12 classes

Chapter 3

Summary

3.1   RESULTS AND DISCUSSIONS

After Using the Resnet34 for our training model we achieve 100 percent accuracy for 8 different bacterial species but their was huge data loss while training and increasing the number of classes hinted toward the bigger ANN model i.e ResNet50 or above. We went with ResNet-50, Which indeed prove to better solution for more number images. We have implemented ResNet-50 with 12 different classes and the accuracy is approx 98 percent which can be treated as remarkable prediction values. At the end of the project our model showed better result then previous work done in this field.
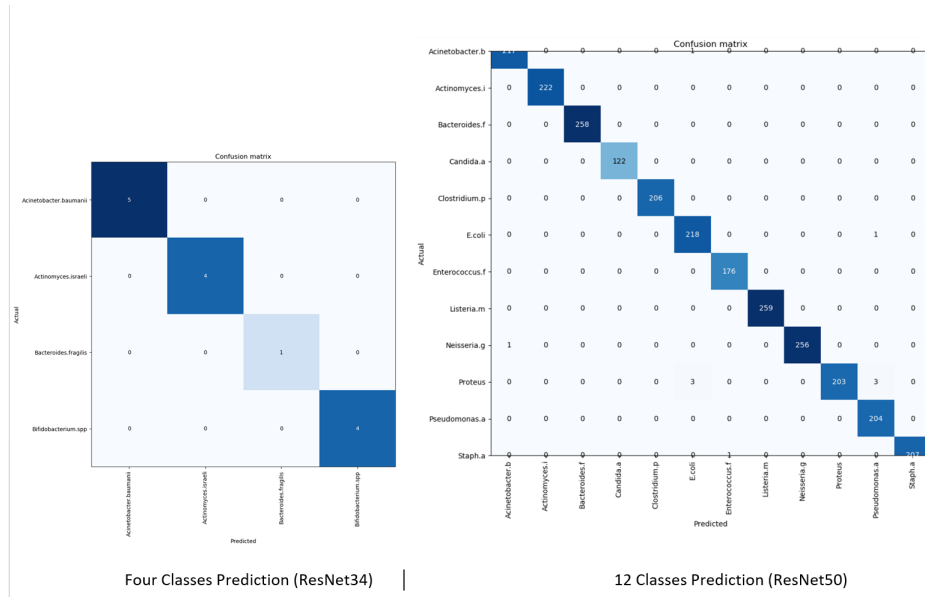
Four Classes Prediction (ResNet34) | **12 Classes Prediction (ResNet50)**

Figure 3.1: a) ResNet34 for Fourclasses b) Resnet50 for 12 classes

MODELS	Accuracy
ResNet-50	0.99
ResNet-34	0.97
CNN-KNN	0.964
CNN-Naïve Bayes	0.955
Decision-Tree	83.77%

Figure 3.2: Comparison from Previous Work.

```
LR Finder is complete, type {learner_name}.recorder.plot() to see the graph.
```
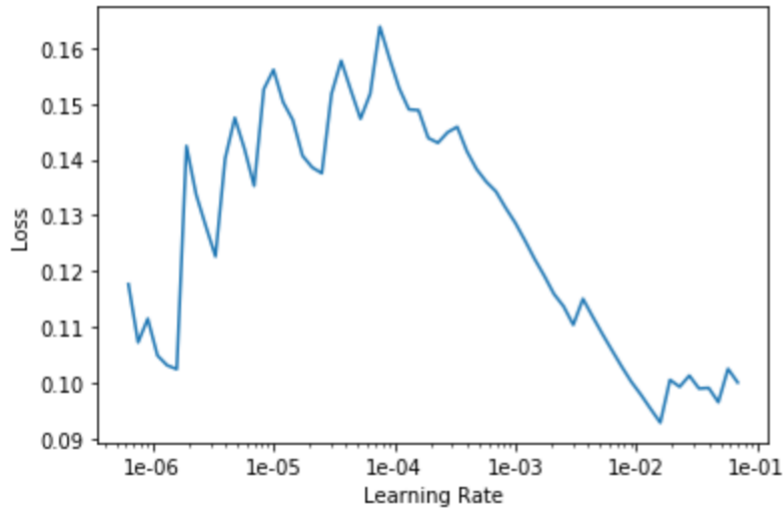

Figure 3.3: Final Epoch cycle values of ResNet-34 for 8 classes

3.2 FUTURE WORK

1. We will add other microbes dataset for classification i.e (Virus,Fungi) in future

2.We will be developing an Advance feature extractor using (VGG models , AlexNet).

3.May use another algorithm for handling different datasets(randomforest).

4.We will try use this Model using Mobile application.

Chapter 4

Bibliography

[1] D. Castro and J. New, The Promise of Articial Intelligence, Center for Data Innovation 2016.

[2] N. A. Mohamad, N. A. Jusoh, Z. Z. Htike, S. L. Win, "Bacteria Identification from Microscopic Morphology using Nave Bayes", International Journal of Computer Science Engineering and Information Technology, April 2014.

[3] Dr. R J. Ramteke and Khachane Monali Y,"Automatic Medical Image Classification and Abnormality Detection Using K Nearest Neighbor ", International Journal of Advanced Computer Research, December 20l2.

[4] Mahbub Hussain, Jordan J. Bird, and Diego R. ,"A Study on CNN Transfer Learning for Image Classification" , Faria

School of Engineering and Applied Science Aston University, January 2012.

[5] Bartosz Zieliński,Anna Plichta,Krzysztof Misztal ,Przemysław Spurek,Monika Brzychczy-Włoch,Dorota Ochońska , Deep learning approach to bacterial colony classification ,September 14, 2017. [6] Trattner S, Greenspan H, Tepper G, Abboud S. Automatic identification of bacterial types using statistical imaging methods. IEEE transactions on medical imaging ,2004.

[7] Anna Plichta , Methods of Classication of the Genera and Species of Bacteria Using Decision Tree, Faculty of Computer Science and Telecommunications, Department of Computer Science, Cracow University of Technology, Krakw, Poland ,PLoS ONE 12(9): e0184554.

[8] Perner P. Classification of HEp-2 cells using fluorescent image analysis anddatamining.In:InternationalSymposiumonMedicalDataAnalysis. Springer; 2001. p. 219224.

[9] De Bruyne K, Slabbinck B, Waegeman W, Vauterin P, De Baets B, Vandamme P. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. Systematic and applied microbiology. 2011; 34(1):2029.

https://doi.org/10.1016/j.syapm.2010.11.003 PMID: 21295428

[10] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2016). "Deep Residual Learning for Image Recognition" (PDF). Proc. Computer Vision and Pattern Recognition (CVPR), IEEE. Retrieved 2020-04-23.

[11] Huang, Gao; Liu, Zhuang; Weinberger, Kilian Q.; van der Maaten, Laurens (2017). "Densely Connected Convolutional Networks" (PDF). Proc. Computer Vision and Pattern Recognition (CVPR), IEEE. Retrieved 2020-04-23.

[12] Kaiming HeXiangyu ZhangShaoqing RenJian Sun , Deep Residual Learning for Image Recognition2016 IEEE Conference On Computer Vision And Pattern Recognitin. , 2016