IMAGE CLASSIFICATION OF CHEST X-RAY DATASET USING CONVOLUTIONAL NEURAL NETWORK (CNN)

By: Gideon Ovuzorie
Sunday Omozusi
Under the supervision of Dr. Yi Sun

Data Science Project Club
School of Physics, Engineering, and Computer Science

Introduction

Pneumonia is a form of acute respiratory infection that affects the lungs. It is the swelling (inflammation) of the tissue in one or both lungs. It is usually caused by a bacterial infection or a virus.

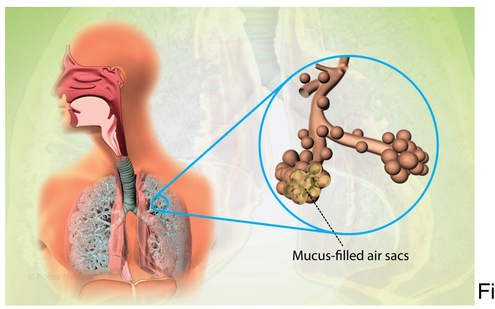

Figure 1

Source: Pneumonia: Symptoms, causes, diagnosis and treatments (msn.com)

Pneumonia is the single largest infectious cause of death in children worldwide. In 2019, it killed 740,180 children under the age of 5 in 2019, accounting for 14% of all deaths of children under five years old.

Project Objectives

This project aims at improving the accuracy of an existing CNN Model which classifies chest x-ray images into normal and pneumonia. This would be achieved by:

- Involving the validation set in the model selection procedure.
- Adding more hidden layers to CNN,
- Tuning hyper parameters, etc.

We will also view the feature maps of each layer to have a better understanding of what the CNN learns from each layer.

The Dataset

We used a Kaggle dataset: Chest X-Ray Images (Pneumonia) which consists of the chest X-ray images of Normal and Pneumonia affected patients, labelled as normal, bacterial pneumonia, and viral pneumonia, respectively. Figure 2 shows three examples from the dataset. It contains 5,840 chest X-ray images divided into the training, validation, and test sets, including 5216, 16, and 624 X-ray images, respectively.

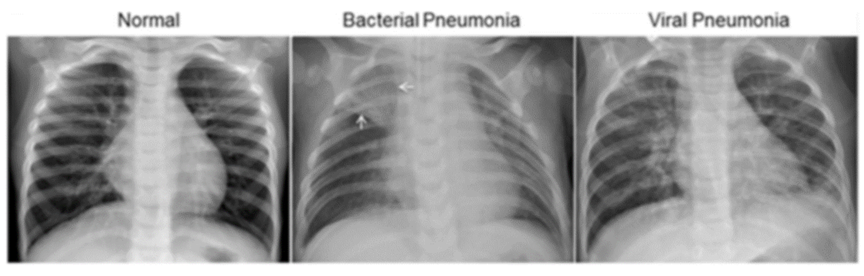

Figure 2: Three examples of X-ray Images in the dataset

The Network Architecture

The model was designed using PyTorch. According to their website, Pytorch is an open-source machine learning framework that accelerates the path from research prototyping to production deployment. It is based on the Torch library, and used mostly in Computer Vision and Natural Language Processing applications.

The Original CNN Model

- The original model was created by FAHAD MEHFOOZ (https://www.kaggle.com/code/fahadmehfoooz/pneumonia-classification-using-pytorch/notebook)
- It consists of three convolutional layers.

The Network Architecture CONT.

- Each layer used a 3x3 filter on the images.
- It does not use the validation set.

Our Improved CNN Model

- A new convolution layer was introduced with a 7x7 filter size. This would make the model learn more images' features to enhance its prediction accuracy.
- After several experiments with different optimizers, Stochastic Gradient Descent (SGD) was adopted as it produced the highest accuracy.
- A validation dataset was introduced in training. This enabled us to guide against overfitting by observing the learning curves (see Figure 3).
- We also introduced some codes that enabled us view the feature maps. This gave us an idea of the features our model was learning in the images (See Figure 4)

Visualisation

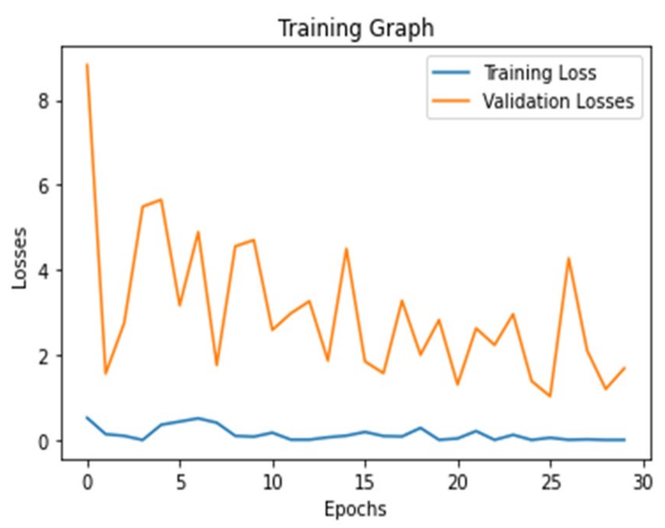

Figure 3 : Learning Curve

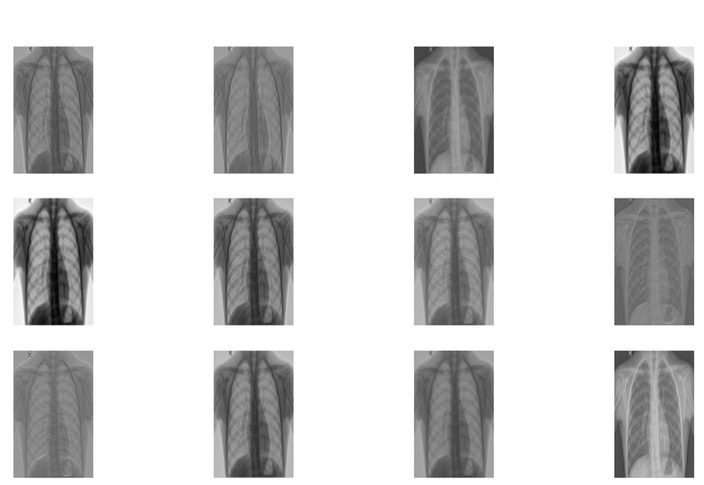

Figure 4: Feature map of the first CNN layer

Performance Measures and Next Steps

The Original CNN Model

Training Parameters	
Optimizer	Adam
Learning Rate	0.01
Loss Function	CrossEntropyLoss
Number of epochs	30

Result	
Loss after the last epoch	0.049
Validation accuracy after the last epoch	N/A
Number of images tested	624
Overall Accuracy	80.13%

Our Improved CNN Model

Training Parameters	
Optimizer	SGD
Learning Rate	0.001
Loss Function	CrossEntropyLoss
Number of epochs	22

Result	
Loss after the last epoch	0.0268
Validation accuracy after the last epoch	1.1315
Number of images tested	624
Overall Accuracy	87.18%

The result shows a 7.05% improvement in our modified model which can be attributed to the introduction of more hidden layers in the neural network. We could also introduce a confusion matrix to show the prediction accuracy of each class in the dataset.

References

WHO, 2021. Pneumonia November, 2021. Available at: https://www.who.int/news-room/fact-sheets/details/pneumonia (Accessed: 10 August 2022).

Pneumonia Classification Using Pytorch. https://www.kaggle.com/code/fahadmehfoooz/pneumonia-classification-using-pytorch/notebook

University of Hertfordshire UH