COSE474-2023F: Final Project Report
"Pneumonia Image Classification on X-Rays using CNN"

Marius Stefan Moldovan

1. Introduction

1.1. Motivation

Pneumonia is an inflammation of the lungs, most often caused by viruses or bacteria. The air sacks may be filled with fluids, causing cough, fever and difficulty breathing. A diagnosis can be difficult because the symptoms are similar to those of a common cold or influenza. Therefore further diagnostic tests such as a blood test and chest X-ray have to be carried out.

1.2. Problem definition

To diagnose pneumonia, a highly trained physician is needed for the Analysis of a chest X-Ray. These physicians have to look at many Images every day, causing exhaustion which results in a higher number of mistakes. By using Deep learning, physicians can be supported in decision making and given hints on what to look at in the picture to not miss any important parts.

The main challenge lies in giving the physician a visual representation of which features (parts of the X-ray) show signs of Pneumonia.

The goal of this paper was to achieve a classification accuracy of 90%.

2. Methods

2.1. The main challenges

The biggest limitation was the limited available data. There were especially limited numbers of x-ray images for healthy people available. This limits the complexity of the model and would result in overfitting for too complex architectures.

2.2. Algorithm

To achieve more flexibility in designing the model, no established architecture was used. The used model consisted of 4 convolutional layers followed by a ReLU activation function and one max pooling layer respectively. After this, there are two dense layers and one layer for the classification. (Fig 1)

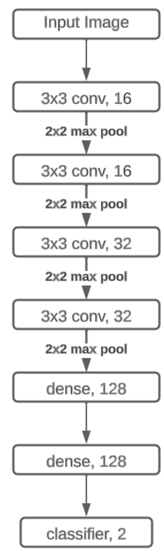

Figure 1. Model architecture

3. Experiments

3.1. computing resources

The model was implemented using the Pytorch Framework.Additionally the cloud based platform "Google Colab" was used to run the code on NVIDIA t4 GPUs.

3.2. Dataset

The dataset is split into two categories: train and test. In total, there are 5,842 chest x-ray images, of which 4,218 show cases of Pneumonia. The images were collected during routine clinical care and went through a quality control which removed any low-quality or unreadable pictures. All of the diagnoses of the images were graded and checked by a total of three expert physicians.

Figure 3.2 shows the Distribution of the two classes in the Dataset. It becomes clear that there are about three times as many pictures showing cases of Pneumonia than there are images of healthy patients, resulting in an imbalanced dataset. This can induce a bias in the model, leading to poor accuracy in the minority class. Additionally, all the images

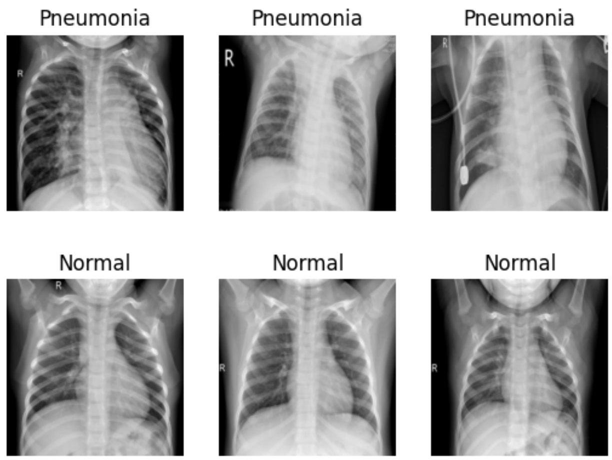

Figure 2. Example images

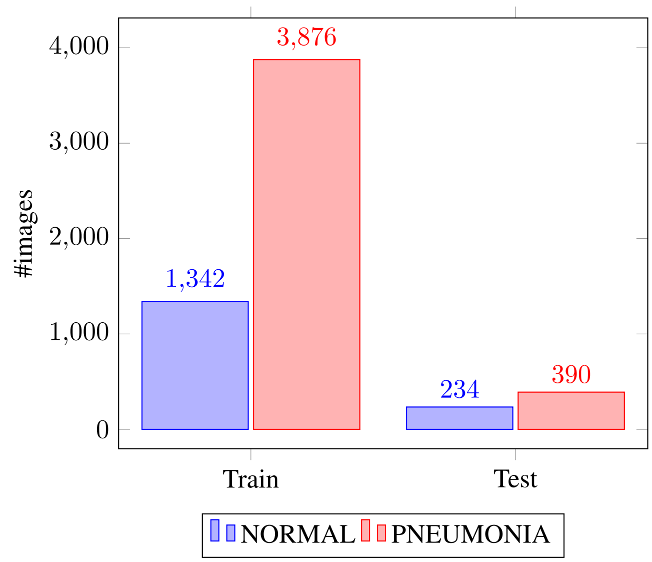

Figure 3. Distribution of Images

have different heights and widths, meaning they have to be cropped to the same size. For this the dimensions of 256x256 were chosen, being the best compromise in the needed GPU-RAM and resolution.

3.3. Experimental design

The used optimizer was Adam with a learning rate of 0.0001. The batch size was set to 32.

The basic model without any regularization added achieved an accuracy of 75% (table 1). To address the problem of the previous mentioned imbalanced dataset, Oversampling of the smaller Class was introduced by using a weighted random sampler. This resulted in an 8% increase in accuracy for the whole model. The main challenge encountered when designing the model was overfitting. To counter this, multiple steps were taken. Firstly the data was augmented by rotating the training images in random directions, increasing the diversity of the trainingset. This resulted in a 3% increase in overall accuracy. Secondly dropout layers, with p=0.5, were added before every dense layer to stop the model from relying on single neurons. In addition to this, L2 regularization was added to the optimizer to prevent gradient explosion. The model also included max pooling layers after every convolution. This further decreased the amount of model parameters. In the end, the model achieved an accuracy of 85%, missing the targeted value by 5%.

The model was trained for 20 Epochs. After an initial big jump in the first epochs, the loss curve starts steadily declining after reaching a value of around 0.5. It should be noted, that there are still noticeable fluctuations in both the loss curve and the accuracy curve. attempts were made to smooth the curves by reducing the learning rate further, but didn't see any success. Training it for further epochs resulted in overfitting, increasing the test loss.

version	Accuracy
CNN	75%
CNN (data balanced dataset)	83%
CNN (balanced dataset, data augmentation)	85%

Table 1. Comparison of accuracy

3.4. Results

Figure 6 shows the Confusion Matrix for the Model. While the predictions for the x-ray images with Pneumonia reach the targeted accuracy of 90%, the model struggles with classifying the cases without Pneumonia and only reaches an accuracy of 73%. The classification therefor produces a lot of type 1 errors resulting in false positives. In a real-life application, this bad accuracy could lead to false patient treatments and is not acceptable. A reason for this might lie in the imbalanced dataset. Even though every batch has the same amount of Pictures of both classes, the Pictures of the Pneumonia cases are more diverse, improving the ability to learn more complex patterns.

3.5. Evaluation

Even after applying multiple generalization techniques, the achieved accuracy of 85% did not reach the target of 90%. This makes it unsuited for use in real-life scenarios. This model still serves as a foundation that can be improved further. When compared with other papers like from Islam, it becomes clear that better results are definetly possible with even more limited data, which achieved accuracies of

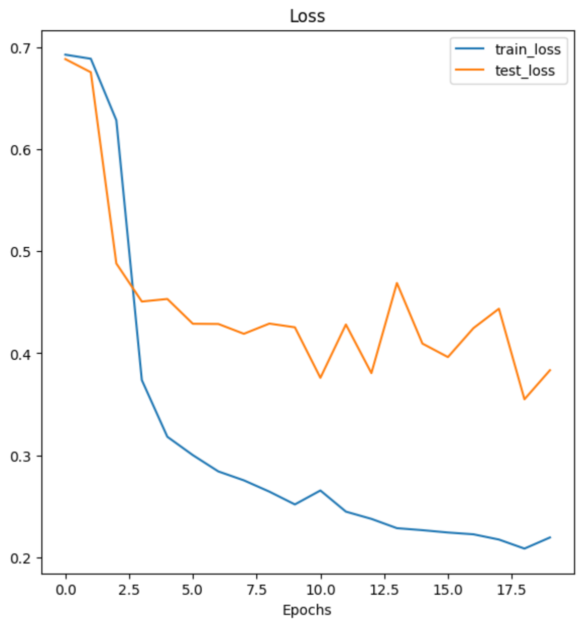

Figure 4. Loss curve

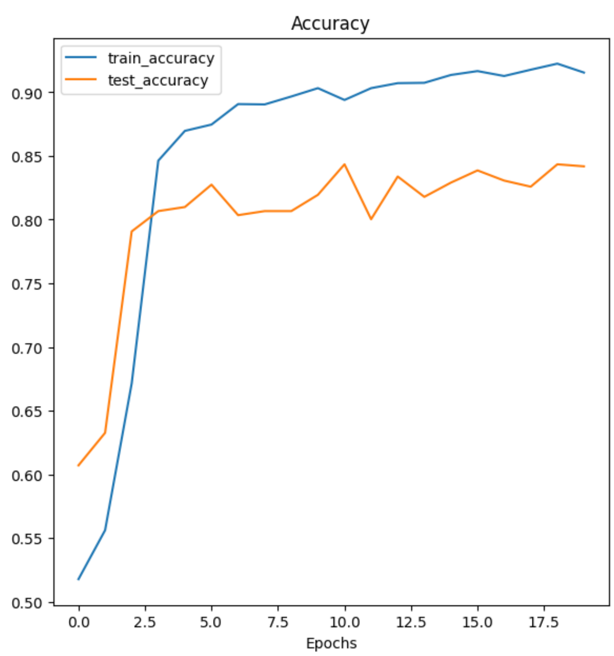

Figure 5. Accuracy curve

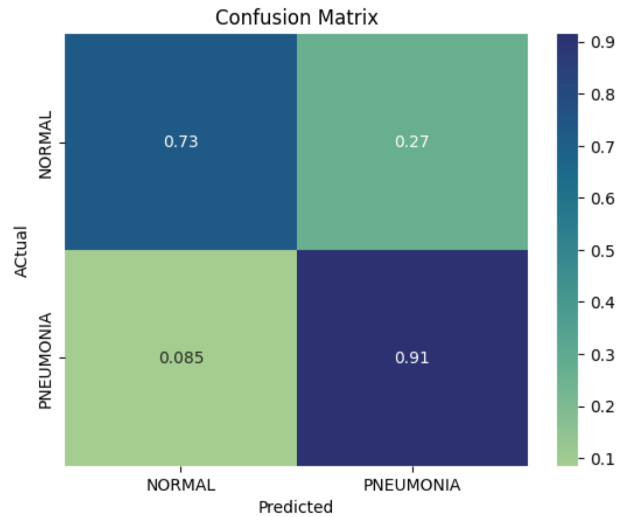

Figure 6. Confusion Matrix

99% by using transfer learning. (Islam, 2022)

4. Future direction

The encountered overfitting is the main reason for the low accuracy. To achieve better results, more data should be collected and used in training. This holds especially true for images that show healthy patients to balance the classes. Furthermore, a pre-trained model could be adapted to this task. This would be the most likely step to take, to improve accuracy.To provide more information to doctors a Class activation map can also be implemented. This would provide visual feedback to where the interesting points on the x-ray image are by overlaying a heatmap, showing the pixels that have the highest weight, on the x-ray image.

References

[1] Islam, M. M., Islam, M. Z., Asraf, A., Al-Rakhami, M. S., Ding, W., Sodhro, A. H. (2022). Diagnosis of COVID-19 from X-rays using combined CNN-RNN architecture with transfer learning. BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 2(4), 100088.