# Pneumonia Detection of Chest X-Ray Images

Alexandre Chapin - achapin@insa-rennes.fr
David Eriksson - daveri-7@student.ltu.se
Oskar Sjögren - osksjg-8@student.ltu.se
https://github.com/AleksandreGit/ADL_Project

## I. Abstract

Convolutional Neural Networks (CNN's) which are used for detecting pneumonia is a highly relevant issue due to the large number of casualties resulting from the disease. The main objective of this report is to investigate the performance of different CNN's, both our own architectures and a pretrained model called EfficientNet which is considered state of the art on binary image classification of lung X-rays. The results of was an obtained accuracy of 94.6% on the test set and a F1-Score of 95.7% using our top performing architecture. Further we did not manage to obtain desired results with the EfficientNet, perhaps due to lack of training capability were we achieve accuracy's ranging from 50-84%.

t-SNE and PCA analysis was performed on the data were global and local structures and similarities between the data was analysed. During classification, heat-maps of the convolutional layers activation's was extracted and a visualization of what the CNN's are "looking" at is analysed. Clearly showing the differences in activation's from the CNN given images of pneumonia and healthy lungs.

## II. Introduction

Pneumonia is one of the major causes for death of children around the world. 2017 had over 800000 registered deaths of children under five years old Dadonaite & Roser (2017). In medicine today the most used technique for detecting pneumonia is by inspection of X-ray images. The images are often unclear and hard to classify for a human eye.

In this paper we will investigate the ability of Convolutional Neural Networks (later referred to as CNN) to detect pneumonia in X-ray images of children lungs. There are a lot of previous work done in this field of study and it's not the scope of this paper to present the new state of the art for image recognition. The goal is to present a good methodology and analysis for binary classification. A task which for a human would be very hard to detect given such abstract features.

## III. Related Work

There has been a lot of research done in the field of computer vision in the recent years and many breakthroughs have been done with Deep neural networks. Computer vision for image classification in medicine is no exception. Marques et al. (2020) used transfer learning on EfficientNetB4 with a few adjustments, such as adding three inner dense layers to classify Covid-19 from lung X-rays. They manage to achieve an average of 99.63% F-1 score, 99.62% accuracy on the binary classification task. For the Covid-19 task this results are considered as state of the art. Ahuja et at. (2020) manage to get 99.4% accuracy with ResNet18 using the same dataset as Marques et al. (2020). Pneumonia classification in X-ray images is like the Covid-19 task a binary classification of X-rays. Rahman et al. (2020) did an extensive research on this with transfer learning on four of the well known CNN's, AlexNet, ResNet, DenseNet, and SqueezeNet. The results showed that DenseNet201 was the model with the best performances on all evaluation indices including accuracy and F-score.

## IV. Experimental Setup

### A. Dataset Used

The data consists of 5,863 X-Ray images of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center. All chest radio graphs were initially screened for quality control by removing all low quality or unreadable scans.

### B. Data augmentation and Preprocessing

Initially the dataset was very unbalanced with a ratio of 2.89 between images labeled pneumonia and healthy. To make the data more balanced and to prevent overfitting towards pneumonia two separate preprocessing functions were used, the initial training set was converted to grayscale then normalized. Healthy lungs were further augmented with random rotation and random flipping later to be concatenated with the original training set. This process evened out the ratio between pneumonia and healthy images. After balancing ratio of pneumonia and healthy images we ended up with a ratio of 1.2.

### C. Model Architectures

The model architectures consisted of three architectures, two self made CNN's and the final one was transferred learning on EfficientNet-b0. The first models architecture, CNN_94, is described in table I due note that after each Conv2d and Linear layer there is an activation function, LeakyReLU, that is used. The second models architecture, CNN_V2, is deeper than the first model and is described in the table II. CNN_V2 also has an activation function, LeakyReLU, present after each of the Conv2d and Linear layers. The final model this paper covered was transferred learning on EfficientNet-b0, the architecture is described in table III

### D. Training

Each of the models were trained on the augmented data set for epochs ranging from thirty to fifty.

TABLE I
ARCHITECTURE OF CNN_94

| Layer | Output |
| --- | --- |
| Conv2d (3x3) | 98 x 98 x 12 |
| MaxPool2d | 49 x 49 x 12 |
| BatchNorm2d | 49 x 49 x 12 |
| Conv2d (3x3) | 47 x 47 x 18 |
| MaxPool2d | 23 x 23 x 18 |
| Conv2d (3x3) | 21 x 21 x 24 |
| Linear | 1000 |
| BatchNorm1d | 1000 |
| Dropout(30%) | 1000 |
| Linear | 100 |
| Dropout(30%) | 100 |
| Linear | 10 |
| Linear | 2 |

TABLE II
ARCHITECTURE OF CNN_V2

| Layer | Output |
| --- | --- |
| Conv2d (5x5) | 96 x 96 x 12 |
| MaxPool2d | 48 x 48 x 12 |
| BatchNorm2d | 48 x 48 x 12 |
| Conv2d (3x3) | 46 x 46 x 12 |
| Conv2d (3x3) | 44 x 44 x 16 |
| Conv2d (3x3) | 42 x 42 x 12 |
| Conv2d (3x3) | 40 x 40 x 12 |
| Dropout (30%) | 40 x 40 x 12 |
| MaxPool2d | 20 x 20 x 12 |
| Linear | 1000 |
| BatchNorm1d | 1000 |
| Linear | 100 |
| Dropout(30%) | 100 |
| Linear | 10 |
| Linear | 2 |

## V. RESULTS

To compare the different architectures that were described in the last section, each one of them were tested on a test set composed of 624 images. The most classic model called CNN_94 obtained an accuracy of 94.6% on the test set with a F1-Score of 95.7%. The second model, CNN_V2, that is a bit deeper obtained worse results with an accuracy of 75.7% on the test set and a F1-Score of 76.0%. Finally some weird results were obtained with the pre-trained EfficientNet, each time it got tested the results were different and the accuracy obtained ranged from 84% accuracy to 50% (that is considered as random with a binary classification). The worse results gained from EfficientNet is possibly caused by EfficientNet being overfitted on ImageNet and for further research one could simply use the architecture instead of applying transfered learning on it.

TABLE III
ARCHITECTURE OF TRANSFERRED LEARNING ON EFFICIENTNET

| Layer | Output |
| --- | --- |
| EfficientNetB0 | 1280 |
| Linear | 500 |
| Dropout(30%) | 500 |
| Linear | 100 |
| Dropout(30%) | 100 |
| Linear | 2 |

## VI. ANALYSIS

In parallel of the test, it has been decided to use some visualisation techniques so as to analyse on a first hand the datas and then see what happens inside of the network during the training.

### A. *Analyse of the dataset*

The techniques used to visualise the dataset were the PCA and t-SNE. t-SNE is great for visulizing high dimensioall data while maintaining local similaities (van der Maaten  Hinton (2008)). We analysed the dataset with and without going through the CNN at hand, this was in an effort to see if any major distinctions could be made before and after going through a trained network. Figure 1 clearly shows that after training the network is capable of distinguish between pneumonia and healthy images both by local and global structure, with the exception of one outlier / missclassification.

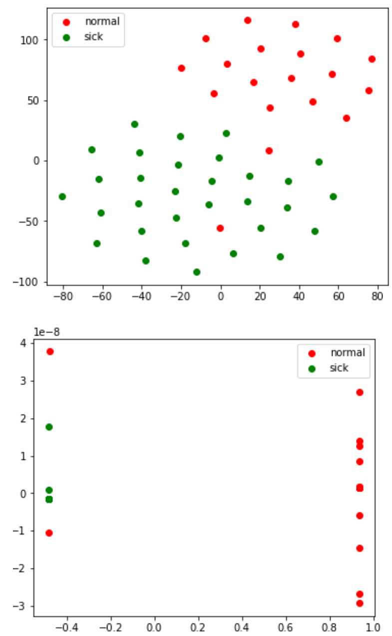


Fig. 1. t-SNE and PCA respectivly after going through CNN_94.

### B. Analyse of the model activation

In order to see what happens inside of the model during the training process, the weights of a certain layer were taken and the activation was compute for a given image to create a Heatmap of it. By doing so, it was easier to see what the networks ”sees” and ”feel” what the network considered most important.

In the heatmaps in figure 2, we see two heatmaps, one for a healthy lung and another one that has pneumonia. It is clearly visible that the healthy one has less activation : it has more ”blue” color inside of the lungs versus the sick one that contains more ”red” color. By comparing those heatmaps to the original images, we could see that the model reacts in priority to the white color. And that is exactly what we want to do : in the subject of this project, it is described that to interpret those X-ray images, radiologist looks for white spots into the lungs (called infiltrates) that identify an infection.

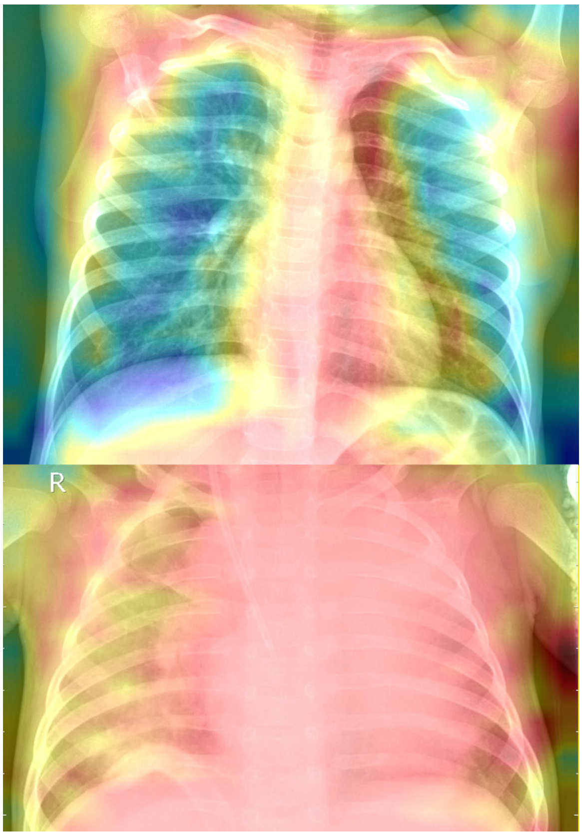


Fig. 2. Heatmaps of a healthy lung (top) and a sick lung (bottom)

### C. *Analyse of the results*

In parallel of the test of our models, some informations were picked up on the classification : which data are well classified and which are misclassified. Those informations were then contained into a confusion matrix. The result obtained for the different models can be observed in the 3. It is interesting to think about the misclassification in the particular case of our project. In medicine we consider it more serious to misclassify someone as healthy but the patient do in fact have pneumonia than to classify a healthy person as sick. The worst case scenario would indeed be that a sick person wouldn't get his treatment. The main goal of our model is to have a 100% recall on the sick datas but in parallel minimizing the miss classification of healthy images.

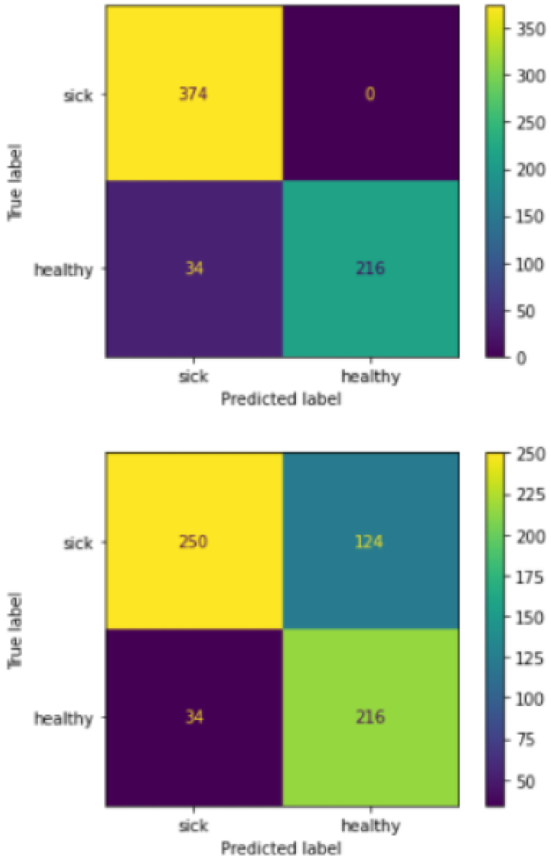


Fig. 3. Confusion matrices obtained on the two CNN models (CNN_94 on top and CNN_v2 on bottom)

## VII. Conclusion & Discussion

Without looking at the accuracy of the CNN one could clearly see that the model would perform well. Given our method of visualising the data before training to see if there are global or local similarities in the data. This could give an indication of how hard the training will be and by correct analysis one could have an indication of how many parameters (with regard to how large the training set is) the network should contain. During development we look at activation heat maps on the training data in a attempt to provide information about whether the neural network is actually "activated" in regions that are relevant to the task or if it is cheating. Finally we visualise the data after passing it through the trained network to see how and if it is capable of separating and clustering similar and unsimilar data. This information is a good indication of whether the network is able to distinguish between images and separate healthy lungs from those with pneumonia. By following this general method developers will have a good framework for analysing different architectures and how well they perform without looking to much at the accuracy.

We also concluded that the ratio of which the dataset contained was really important, when the dataset was evenly balanced the model would misclassify sick and healthy evenly. This could prove to be a big issue due to misclassifying a sick person as healthy is a potential loss of life due to the severity of pneumonia. So by making the dataset more bias towards the pneumonia lungs the model would be a lot better at classifying sick lungs correctly further decreasing the potential fatal mistakes made by the model without lowering the overall accuracy of the model.

## VIII. References


1. Bernadeta Dadonaite and Max Roser (2018) ”Pneumonia”. Published online at OurWorldInData.org. Retrieved from: ’https://ourworldindata.org/pneumonia’

2. Ahuja S., Panigrahi B.K., Dey N., Gandhi T. and Rajinikanth V. (2020). Deep transfer learning - based automated detection of COVID-19 from lung CT scan slices

3. Gonçalo Marques, Deevyankar Agarwal, Isabel de la Torre Díez (2020). Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. DOI: https://doi.org/10.1016/j.asoc.2020.106691

4. Rahman, Tawsifur and Chowdhury, Muhammad E. H. and Khandakar, Amith and Islam, Khandaker R. and Islam, Khandaker F. and Mahbub, Zaid B. and Kadir, Muhammad A. and Kashem, Saad (2020). Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray. DOI: 10.3390/app10093233