Stakeholder report - M3

- A classification case with chest X-rays and pneumonia

by Jonas Røge Jepsen, Michael Bering Olesen and Tobias Maltha Christensen

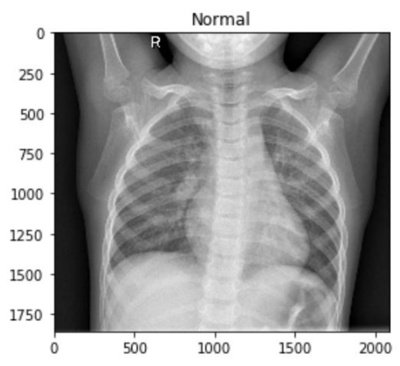
Introduction:

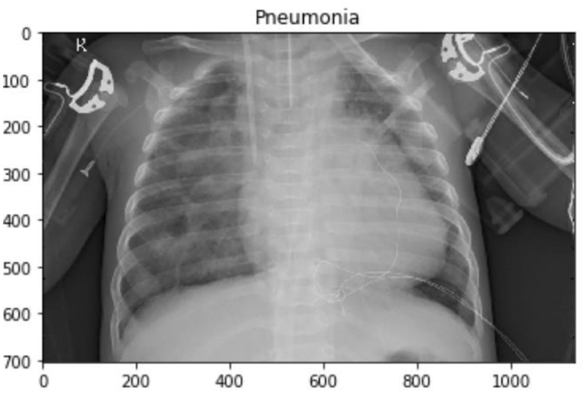
We are working with X-ray images of childrens lungs, and we want to predict whether the children have pneumonia (inflammation of the lung parenchyma) or have "normal" lungs.

The data:

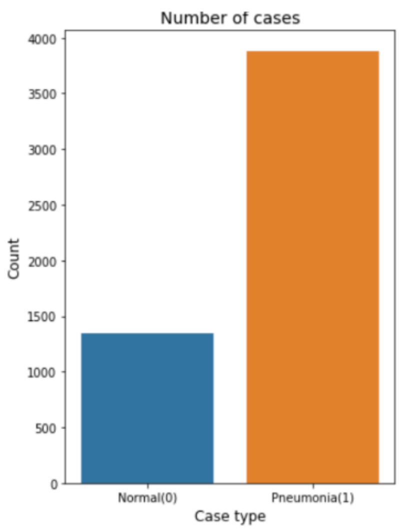
The dataset used in this assignment contains 5863 X-ray images of childrens lungs selected from a women and children's medical center in Guangzhou, China. The data is available from Kaggle and is organized into to three folders: Training, Test and Validation. Each of these folders contain subfolders which categorizes the images into the patient having "normal" lungs or having pneumonia. All the images have been screened for quality where low quality and unreadable scans have been removed. Furthermore, in order to secure the quality several experts have gone through the images to make sure they were labeled correctly [1].

Below is two different examples of images in the dataset. The picture on the left is a image of a "normal" lung, and the one on the right have pneumonia, and as it can be seen it is difficult to spot the difference.

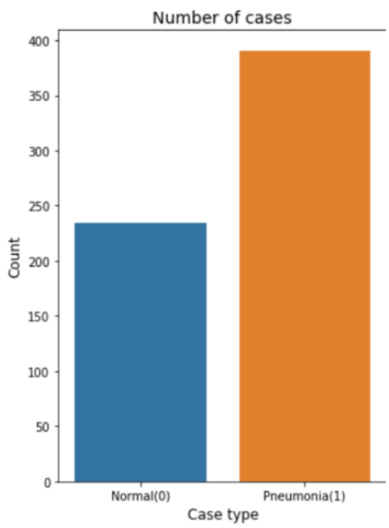




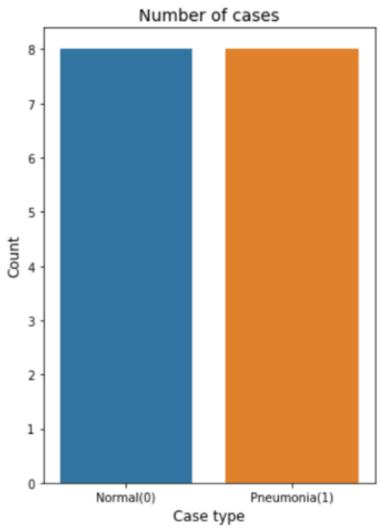
As stated earlier, the data is divided into three subsets. Below is a visualisation of how many of each labels or cases there is in each subset.



Training set



Test set



Validation set

As can be seen above the training part of dataset consists of 1341 images of “normal” lungs and 3875 images of lungs with Pneumonia. The distribution between the two is a little unequal, but there should be enough data for it to work.

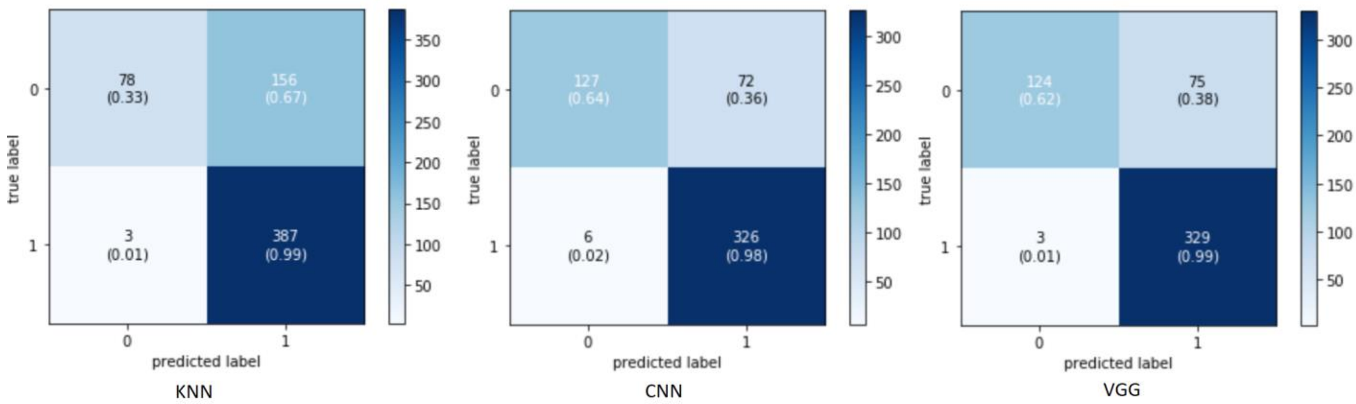
The same seems to be the case with test part, here there is 234 images of ”normal” lungs and 390 images of lungs with Pneumonia. The validation part of the dataset only consist of 16 images 8 of each class, and this seems to be too small to be useful, and for that reason we will not use this part of the dataset, instead we use a small part of the test part of the dataset as the validation set.

Image augmentation

We did some augmentation on the dataset to increase the performance. Which means we manipulated the images by stretching, zooming and flipping them.

Models and results:

We took three different approaches to solve this binary image classification problem, we used a simple K-nearest neighbors (KNN) non-neural model, Convolution Neural network (CNN) and VGG16 which is a pretrained convolutional neural network model, to which to we add some layers at the end. In the model the label 0 is equal to “normal” lungs and 1 is pneumonia.



KNN
As you can see the KNN give as surprising good result. It has a low number of false negatives and high number of true positives, however it takes some extra guesses when it comes to false positives, which isn't great. What we want is both a high number of true negatives and true positives for us to use it in a real situation in a hospital. Hopefully our deep learn model will perform better than the accuracy of KNN, which is 0.75.

To improve the KNN model one could:

- Change the parameters of the model and data
- Use hyperparameter tuning to improve the model
- Perhaps changing the datasets image pixels size

CNN
The confusion matrix for the CNN shows a fine result. However there is still some False negatives and False positives. However, it's better to get True negatives, which makes it slightly better than the KNN model. The accuracy for this model was 85.31%.

VGG
The VGG confusion matrix is slightly better than the CNN. Early results of the VGG had an accuracy of 90%, were this run only gave 85%. Only 3 x-rays got predicted as False negatives which is great and 75 as false positives, which could be better. The accuracy for this model was 85.31%.

Tuning

Furthermore, we did some tuning of our CNN model changing different parameters, to obtain the best result, such as how many pixels were used to train the model and the number neurons. Doing this we were able to obtain a higher accuracy of 89% on our CNN model. This could also have been done on the VGG model, and thereby it might have obtained better results as well.

Conclusion

The best model for detecting pneumonia on x-rays of children's lungs based on this study seems to be the VGG.

Discussion

The accuracy of the deep learning models are in the high-end, being above 85%, but would it be possible to use these models in a real life scenario? The biggest challenge for this are the False negatives and the false positives. It could cause problems for the patients, if they are misdiagnosed by the model. We would not recommend the models being used as an independent decision maker. But, it might be used as a support decision system for the doctor evaluations.

[1]https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia