Binary Image Classification using Chest X-Ray Images

George Paterakis
CSD

Nikolaos Gounakis
CSD

Zubaria Asma
CSD

Abstract

Pneumonia is an inflammatory condition affecting the lungs. Various ways to diagnose , most prominent through Chest X-ray. However, even professional doctors struggle to diagnose pneumonia as it has similar region information with other diseases. To that end, we propose a transfer learning solution for the detection of pneumonia through x-ray images. We rely on pre-trained neural networks in order to extract meaningful information from the images. The information extracted is then used by a classifier that extends the pre-trained networks. Our model achieves 0.89 AUC score.

1 Introduction

Pneumonia causes the death of around 700,000 children every year and affects 7% of the global population. Pneumonia is an infection that inflames the air sacs in one or both lungs. The air sacs may fill with fluid or pus (purulent material), causing cough with phlegm or pus, fever, chills, and difficulty breathing. A variety of organisms, including bacteria, viruses and fungi, can cause pneumonia. The symptoms of pneumonia can develop suddenly over 24 to 48 hours, or they may come on more slowly over several days. Common symptoms of pneumonia include: coughing, difficulty breathing, rapid heartbeat and sweating.

Chest X-rays are primarily used for the diagnosis of this disease however even for experts it is a challenging task to examine chest X-rays. On that end, we propose and automated solution for the detetion of pneumonia through chest x-ray images. Hopefully, this solution could be used alongside expert's diagnosis for a more accurate and fast prediction of pneumonia.

We propose an automated system, using pre-trained neural networks and transfer-learning in order to accurately extract features from images and then using the extracted features, classify the images as healthy or pneumonia patients. In this project, we compare different pre-trained neural networks as well as classifiers in order to find the best configuration (pre-trained neural network and classifier combination). The best configuration achieves 0.89 AUC in a seperate test of images.

In section 2, we present the current solutions to this problem. In section 3, we present the experimental setup and in section 4, we analyse the results of our comparative evaluation. Finally, in section 5, we conclude the paper and discuss future work.

2 Related Work

The dataset was made available by [7]. They propose a transfer learning methodology that utilized a pre-trained neural network (Inception V3) and a classifier (Softmax) while a fine tuning on the pre-trained network is attempted after the training of the classifier layer. Their solution scored 0.96 AUC over predicting healthy and pneumonia individuals. The problem of classifying pneumonia images became viral again as COVID-19 hit the world in 2020. [3] analysed x-ray chest images in the era of coronavirus. They used tailored CNN models to extract meaningful information from the images and classify them intro three categories. Every CNN model used in their study scored over 84% accuracy on the task, which they concluded is promising. [1] proposed an energy efficient deep learning method for pneumonia detection. They used DenseNet-121 in order to extract image features and then a fine-tuned DNN-based feature classification method to detect the type of image. The final classifier had 3 hidden layers, 1024 and 512 layer size for the first and second layer. Their solution achieves 94.4% accuracy and for the detection process, less than a second is needed in order to make an accurate prediction.

3 Experimental Setup

In this section, we provide a detailed overview of the experimental setup. In figure 1, we visualize the pipeline used for the image classification task.

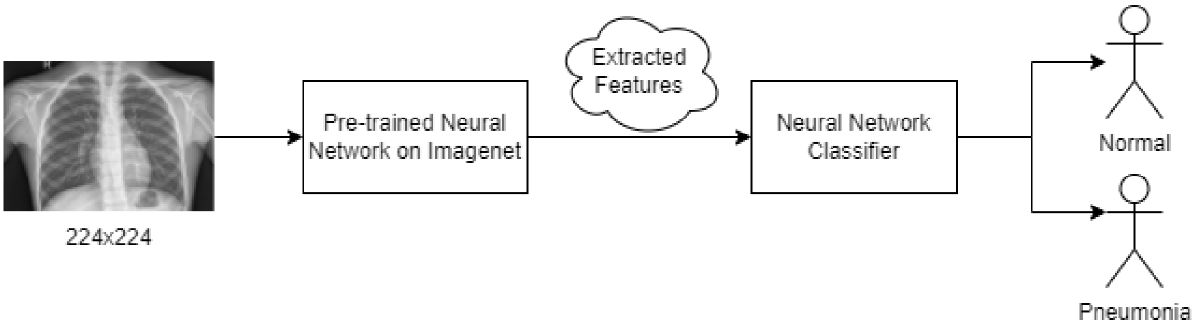


Figure 1: Pipeline of our approach. At first an image is being fed in the pre-trained neural network for feature extraction. Then the extracted features are used by the classifier in order to predict a healthy or a pneumonia individual.

Dataset Split	Samples	IR
Train-Validation	5232	0.35
Test	624	0.6

Table 1: Characteristics of the Dataset presenting number of samples and imbalance ratio. In all cases the pneumonia labeled samples are more than the healthy ones.

3.1 Dataset

We selected a publicly available dataset that contains x-ray chest images from kaggle https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. In table 1 we can see the characteristics of the dataset. Originally, the dataset was split in training, validation and test set. However, the validation set was only 16 samples. We merged the training and the validation set and through the use of python we splitted the new training set into 90% training and 10% validation for a more generalized and less biased validation results.

3.2 Feature Extraction

For the task of feature extraction we used pre-trained neural networks on imagenet dataset. The NN where provided by Keras library in python. An overview of the neural networks used follows.

3.2.1 Resnet 50

ResNet-50 [4] is a convolutional neural network that is 50 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database.

3.2.2 MobileNetV3 small

MobileNetV3 is a convolutional neural network that is tuned to mobile phone CPUs through a combination of hardware-aware network architecture search (NAS) complemented by the NetAdapt algorithm, and then subsequently improved through novel architecture advances. Advances include (1) complementary search techniques. [5]

3.2.3 EfficientNetV2s

A new family of convolutional networks that have faster training speed and better parameter efficiency than previous models. To develop this family of models, they used a combination of training-aware neural architecture search and scaling, to jointly optimize training speed and parameter efficiency. [9]

3.2.4 InceptionV3

Inception v3 is a convolutional neural network for assisting in image analysis and object detection, and got its start as a module for Googlenet. It is the third edition of Google's Inception Convolutional Neural Network, originally introduced during the ImageNet Recognition Challenge. The design of Inceptionv3 was intended to allow deeper networks while also keeping the number of parameters from growing too large: it has "under 25 million parameters", compared against 60 million for AlexNet. [8]

3.2.5 Xception

Xception by Google, stands for Extreme version of Inception, is reviewed. With a modified depthwise separable convolution, it is even better than Inception-v3. [2]

3.2.6 DenseNet121

DenseNet (Dense Convolutional Network) is an architecture that focuses on making the deep learning networks go even deeper, but at the same time making them more efficient to train, by using shorter connections between the layers. [6]

3.3 Classification

For the image classification task, we provide the extracted features to two DNNs with 1 hidden layer that is composed of a Dense Relu layer of 1024 size or an Average2DPooling layer. Finally the output layer consists of a SoftMax activation function that is appropriate for binary classification.

3.4 Metrics and Evaluation process

For the evaluation we used a simple train-validation and test set split. We do not perform any preprocessing to the images except for resizing them to 224x244 pixels.

We used 10 epochs for training, binary crossentropy as loss function and adam optimizer with default learning rate (0.001).

The metric used is AUC. AUC shows the ability of the classifier to accurately distinguish a negative from a positive class making it robust to class imbalances.

4 Results

The best configuration is Resnet50 with average2dpooling (Global Average Pooling) and softmax as classifier. The AUC is 0.89 in the test set. As we can see from figure 2, average2dpooling is better than using a Dense,Relu network as hidden layer (Fully Connected Layers). This can be explained by figure 3. We can see that the training performance of the Dense layer is 1 on 2-3 epoch which means that the validation performance can't improve and the classifier overfits. On the other hand, we can see that this is not the case for average2dpooling where as epochs increase both training and validation set AUC increases.

5 Conclusions and Future work

In this project, we shown how pre-trained neural networks can be effectively used for the pneumonia image classification task, without fine-tuning them, or the classifier layer hyperparameters and by using a few epochs for training like 10, instead of 100 [7]. The pre-trained networks can extract meaningful features that will later be fed to a classifier for a very accurate prediction of Pneumonia that will help experts to make a decision. We want to note that in future work we will extend this project to more classifiers and as well fine-tune the pre-trained neural networks on the train data and also tune the parameters of the classifiers. We believe our results are promising and hope to improve on the state-of-the art performance achieved by the authors in [7].

Acknowledgments

We would like to thank the professor, Christos Kozanitis, and the teaching assistant, Aggelos Sinogeorgos, for their feedback and comments as well as for the course material.

Availability

Our code will be publicly available on github https://github.com/NicolaiGoon/HY-543-Project

References

[1] Bandar Almaslukh. A lightweight deep learning-based pneumonia detection approach for energy-efficient medical systems. *Wireless Communications and Mobile Computing*, 2021:5556635, Apr 2021.

[2] François Chollet. Xception: Deep learning with depthwise separable convolutions, 2016.

[3] Karim Hammoudi, Halim Benhabiles, Mahmoud Melkemi, Fadi Dornaika, Ignacio Arganda-Carreras, Dominique Collard, and Arnaud Scherpereel. Deep learning on chest x-ray images to detect and evaluate pneumonia cases at the era of covid-19, 2020.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[5] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. *CoRR*, abs/1905.02244, 2019.

[6] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2016.

[7] Daniel S Kermany, Michael Goldbaum, and Cai. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, page 1122—1131.e9, 2018.

[8] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.

[9] Mingxing Tan and Quoc V. Le. Efficientnetv2: Smaller models and faster training. 2021.

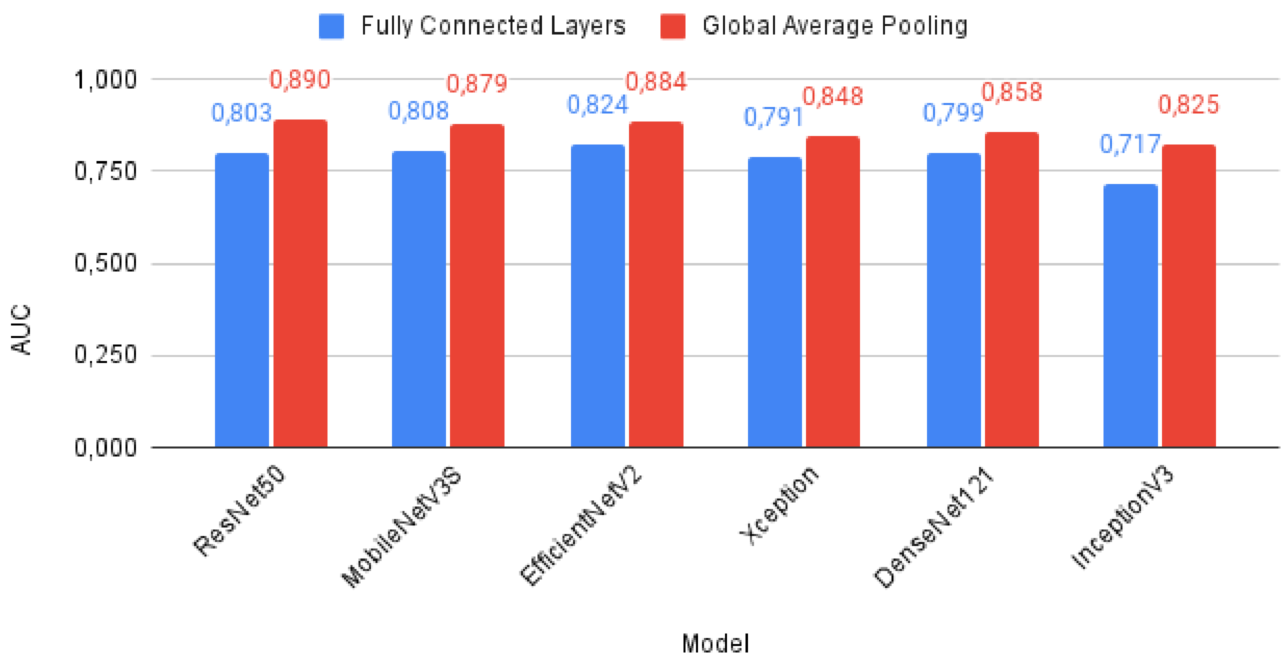
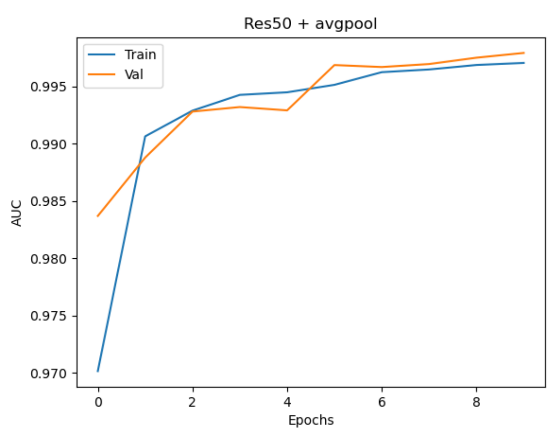
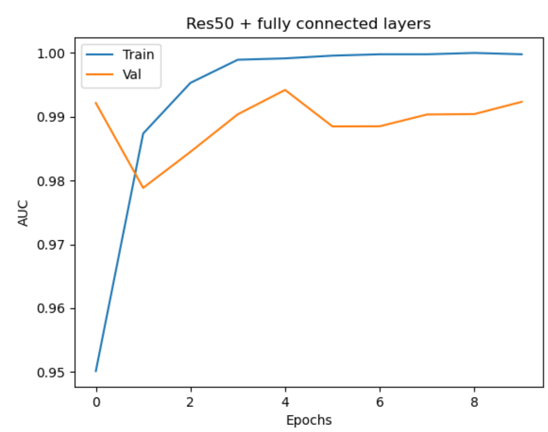


Figure 2: The results over the test split of the dataset. We can see each combination of pre-trained NN and classifier. ResNet50 takes the first place by using Global Average Pooling.



(a) ResNet50 + Global Average Pooling



(b) ResNet50 + Fully Connected Layers

Figure 3: Train vs Validation AUC among Epochs , for ResNet50