Transfer learning for COVID-19 detection

Theodoros Panagiotakopoulos
thepan@kth.se

Adrian Campoy Rodriguez
adriancr@kth.se

Gustavo Teodoro Döhler Beck
gtdb@kth.se

Nimara Doumitrou-Daniil
nimara@kth.se

Abstract

SARS-COV-2 constitutes a novel virus with a relatively high infection rate [2]. As such, early and effective detection is needed in order to track and impede the spread of the pandemic. We propose a Deep Learning approach for detecting COVID-19 based on X-ray images. Aside from detection, our network provides insight about the affected area, the core locations of infection and how similar diseases are when represented by networks. Our findings illustrate promising results both in localisation as well as AUC, reaching a ROC curve performance of 0.88 on the COVID-19 X-ray Dataset. Our report also analyses the importance of Transfer Learning, Model Architecture and Data Pre-processing.

1 Introduction

COVID-19s alarming infection rate requires early and robust detection. While the diagnosis is confirmed using polymerase chain reaction (PCR), infected patients with pneumonia may appear on chest X-ray and computed tomography (CT) images with a pattern that is difficult to spot with the human eye [7]. We propose utilising X-rays, as they not only prove a promising alternative that is already widely used in medicine for detection (e.g. for pneumonia), but also offer the possibility of visualising and localising the disease, giving further insight to its severity.

Our model builds upon well established architectures for robustness and efficacy. It consists of a core deep Convolutional Neural Network which is augmented with a fully connected layer for classification. Inspired by [11], we implemented a ResNet architecture and adopted its ResNet-50 variant. We also utilised a pretrained (on ImageNet) ResNet-50 and a DenseNet-121, as described in [9].

All of the aforementioned architectures were initially trained on the ChestX-ray14 Dataset [8] which constitutes a **multilabel problem** where patients can exhibit one or more diseases, as described in section 3. It also contains multiple patients who do not exhibit any of the diseases included in this data set. Its complexity, which derives from the high degree of class overlap (multiple combination of diseases that coexist), and quantity (approximately 112 thousand X-rays) will help our model learn a latent representation of X-ray Images, bolstering our performance on the rather limited COVID dataset.

We visualize the effect of the learnt latent representation by comparing the performance of Self Organizing Maps and t-distributed Stochastic Neighbour Embedding (t-SNE) before and after applying our Neural Network on our images. This improvement is expected as our latent representation acts as a dimensionality reduction (similar to PCA), projecting our data into a dense, smaller and information rich space.

In the end, our DenseNet network achieves an average ROC score of 0.81 on the ChestX-ray14 Dataset and a ROC score of 0.83 on the COVID-19 dataset. We also consider a 15-classification

Project report for DD2424 - https://github.com/gustavo-beck/DD2424-Deep_Learning-COVID-Project/

task by sampling from both datasets, achieving a 0.89 ROC for COVID-19. Finally, we provide heatmaps for disease localisation (CAMs).

1.1 Outline

The remainder of this report is structured in the following way: In Section 2, we will introduce and familiarise the reader with related work in this field so that he will be able to better conceptualise our approach and reasoning. Next, section 3 will succinctly describe our data and its particularities. In Section 4, we will present in more detail the methods that were utilised to produce our results. After obtaining a firm understanding of our approach, in Section 5, we will illustrate and reason our obtained results. Finally, in Section 6, the key findings and takeaways of the report will be summarized.

2 Related work

Our work was inspired by the findings of Xiaosong Wang et al., 2017, [11] and Pranav Rajpurkar et al., 2017, [9] which utilised ResNet-50 and DenseNet-121 respectively on the ChestX-ray8 and ChestX-ray14 (the former only inspects 8 of the 14 classes). The strategy of both papers consists on using pre-trained networks and employ transfer learning to primarily identify thoracic diseases represented in X-ray images. After comparing both studies, our approach focused on reverse engineering them and build our own model.

2.1 ResNet for ChestX-ray8

As the main application of ChestX-ray8 dataset, Xiaosong Wang et al. [11] presented a unified weakly-supervised multi-label image classification and pathology localization framework. In details, they tailored different pre-trained Deep Convolutional Neural Network (DCNN) architectures such as AlexNet, GoogLeNet, VGGNet-16 and ResNet-50. In addition, they removed the fully-connected layers and the final classification layers of the original architectures and substituted them with a transition layer, a global pooling layer and a prediction layer.

After exhaustive training, Xiaosong Wang et al. concluded that their best model was their variation of the ResNet-50 trained with the following splitting: training set 70%, validation set 10% and testing set 20% of the 108,948 frontal-view X-ray images.

2.2 DenseNet for ChestX-ray14

Similarly to the ResNet model presented previously, this model (CheXNet), proposed by Pranav Rajpurkar et al. [9], aims to accomplish similar goals: output the probability of pneumonia along with a heatmap localizing the areas of the X-ray most indicative of this pathology. However, in this case the network is trained on a data set with 14 different diseases.

The model's architecture is comprised of a Dense Convolutional Network with 121 layers composed by convolutional, pooling, transition, and classification layers. They replaced the final fully connected layer with one that has fourteen outputs (one for each disease), after which they apply a sigmoid nonlinearity. Furthermore, they employ transfer learning alike Xiaosong Wang et al., initializing the weights from a model pre-trained on ImageNet.

This architecture appeared to show very good results by randomly splitting the dataset into training (28744 patients, 98637 images), validation (1672 patients, 6351 images), and test (389 patients, 420 images). We believe that the 420 images in testing may not be representative enough in a dataset with more than 100,000 images.

3 Data

Two data sets were used. On one hand, a data set containing 112,120 chest X-ray exams from the NIH [8] (also known as ChestX-ray14). The main goal of this data set was pre-training our network to be able to compensate for the small data set from COVID-19 as well as applying a multilabel

classification to identify a set of 14 thoracic diseases. On the other hand, the COVID-19 X-ray open data collection [4] was used to obtain images of patients with pneumonia caused by COVID-19.

3.1 ChestX-ray14

The NIH Chest X-ray data set is composed of 112,120 X-ray images with disease labels from 30,805 unique patients. In order to create these labels, the authors used Natural Language Processing to text-mine disease classifications from the corresponding radiological reports. Thus, labels are expected to be >90% accurate.

Each X-ray image may have between 0 and 14 labels. Each of the 14 labels corresponds to a disease: Atelactasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Plural Thickening and Hernia. Images having 0 labels were classified as "No Findings", meaning none of the aforementioned diseases were found. Since each patient could have a number of diseases, we treated our data as for a multi-label problem, where each sample had a 14 dimensional label y where $y_i = 1 \iff$ the patient exhibited the i-th disease.

3.2 COVID-19 image open data collection

Due to the global efforts against the spread of the novel coronavirus COVID-19, there are many initiatives to gather data regarding this disease. As a result, J. Cohen and L. Dao [4] created an X-Ray and CT open dataset of patients which are positive or suspected to be positive on COVID-19 or other viral and bacterial pneumonias (e.g. MERS, SARS, and ARDS). The main goal of this data set is to develop AI based approaches to predict and understand the infection. These approaches could help improving prognostic/diagnostic predictions to triage and manage patient care.

The data set continues being updated. Nonetheless, by the time this project took place, the data set contained 358 images. CT images and X-rays taken from sagittal plane were discarded, keeping only X-ray images from the coronal plane. Finally, our data set consisted of 232 COVID-19 images and 63 non-COVID images.

3.3 Processing X-ray images

Data pre-processing is an essential step in a model's formulation and can be divided into some distinctive steps such as data cleaning, data transformation and feature selection [3]. Deep neural networks have shown impressive results in discriminative tasks, nevertheless improving the generalizability of these models is one of the most difficult challenges [5]. Data augmentation techniques have shown to be most effective in increasing the network's robustness and enhancing generalization. Moreover, the images in both datasets showed varying, often too large, sizes for effective usage in our neural network. Therefore, we resized the images in 224×224, before passing them through a series of transformations.

First and foremost, due to lung symmetry, a random horizontal flip was applied. Augmentation continued with random rotation, jittering and affine transformation. By utilizing jittering, key image characteristics such as contrast and brightness are randomly altered while using the affine transformation, a projection of the image is created. By exploiting those methods overfitting is decreased. Furthermore, the network is prevented from learning particular image characteristics, which often occur in this type of images with objects on the patient's body (i.e. necklaces, earrings, medical implants, etc) or certain features meant to aid clinicians such as letters or labels.

4 Methods

Our methodology can be dissected into three key parts: Data Visualisation, Disease Prediction and finally disease localisation.

4.1 Data Visualisation using Self Organising Maps and t-distributed Stochastic Neighbour Embedding

Self Organising Maps [6] (SOM) are unsupervised topology preserving mappings of data which are often employed for visualisation. They are topology preserving, in the sense that data x_n which lie

close (and describe similar concepts) together in the original high dimensional space maintain this similarity in their lower dimensional mapping y_n. The output space is typically one, two or three dimensional, since it can be plotted, facilitating its descriptive qualities.

SOMs map a high N-dimensional data onto d neurons-clusters in a low dimensional space. These neurons share notions of neighbourhoods, being strongly linked with others that lie near them. For instance, for two dimensions, one can arrange them in a grid. Each output neuron is characterized by a weight vector w_i which is used to compare it with each sample. These vectors constitute the learnable parameters of this model, which are updated using competitive learning, where, given a sample x_j only the closest weight w_i and its neighbourhood gets updated (pulled towards x_j). The size of the neighbourhood usually varies, starting off large, before shrinking over time (e.g. at each epoch). For instance, our approach consisted on deciding the neighbours to be updated based on a heuristic function:

$$h_t(w) = \exp\left(-\frac{d(w, w_h)^2}{2\sigma_t^2}\right)$$

Where, d is the distance of each weight vector (w) with the closest (winner) to the sample weight(w_h) and σ the radius that is used to compute distance in the output space:

$$\sigma_t = \sigma_0 \exp(-\frac{t^2 \log(4\sigma_0)}{T^2})$$

Where t the current epoch and T the total number of epochs. Finally, the weights are updated with η as learning rate: $\Delta \boldsymbol{w} = \eta h(\boldsymbol{x} - \boldsymbol{w})$. Therefore, after some iterations the value of $h(t)$ becomes smaller and only a few neighbours (some times only one) are updated. At this point the algorithm transits from a cooperative behaviour to a competitive in order to become more specific to each class. Finally, after training it's possible to represent the images into a smaller dimension by assessing the trained weights. Similarly, t-distributed Stochastic Neighbour Embedding (t-SNE) is a more novel approach for the same problem. It first induces a distribution p in the original N-dimensinal space based on their mutual Mahalanobis (Gaussian based) distance. That is, $p_{i|j} = \eta \cdot \mathcal{N}(x_i - x_j, \sigma_j^2)$ (η is normalisation term) and $p_{i,j} = \frac{p_{i|j} + p_{j|i}}{2 \cdot |samples|}$. It then induces a distribution q in the output space, based on the Cauchy distribution $q_{i,j} = \eta \cdot \frac{1}{1+||y_i - y_j||^2}$. Since we wish to maintain the topology of our data, we can simply aim to minimise the Kullback-Leibler divergence between our p and q distributions using gradient descent. Note that σ_i is a tunable hyperparameter which regulates the neighbourhood of x_i, as the distribution $p_{i|j}$ can be seen as a gaussian centered around x_j. A large σ_i induces a large neighbourhood and leads to mappings y_i where most data lie close together. On the other hand, small σ_i creates sharp small neighbourhoods which often lead to mappings y_i organised in clusters.

We will use the two aforementioned methods to visualize our data and see the effect of our networks on them. Namely, we will input both our original images and the output of the final convolution (before the classification) to compare our original space with our latent representation.

4.2 Disease Prediction

For prediction, we will utilise Deep Neural Networks. For this matter, we implemented the ResNet architecture and imported (from pytorch) the pretrained models of ResNet and DenseNet. We aim to compare our own non-pretrained model with the latter two to measure the effect of network pretraining as well as architecture.

In all cases, the same classification fully connected layer was used, followed by a 14 dimensional sigmoidal layer (we use sigmoid since our problem is multilabel). The same loss function as in [9] was utilised, that is:

$$L(X, y) = \sum_{c=1}^{14} [-y_c \log p(Y_c = 1|X) - (1 - y_c) \log p(Y_c = 0|X)]$$

which will be optimised using the Adam with standard parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We will also experiment with the initial learning rate, weight decay and data transformations (see Section 5). Note that, as explained in the introduction, each sample X has a label $y \in \mathbb{R}^{14}$ such that $y_i = 1 \iff$ the patient exhibits the i-th disease. Thus, our loss aims to minimise total (for each class) binary cross entropy.

We will then compare the results of non-pretrained ResNet-50, pretrained ResNet-50 and DenseNet by measuring their per class and average Area Under Curve of our ROC (Receiver operating characteristic) curves. The ROC is generated by plotting the true positive rate with the false positive rate. A high AUC indicates that we are able to achieve high true positive rates with low false positive rates. We prefer basing our mode evaluation on ROC scores, rather than accuracy, as our data set is highly imbalanced.

After selecting the best model, we will apply it for transfer learning on the COVID-19 data set, where we will once more measure ROC scores. For this task, we keep the underlying deep convolutional network intact, only swapping out its classification layer. Finally, we will compare the performance of our pretrained models (on chestX-ray14) with non pretrained ones (training directly on the COVID data set). Note that for COVID-19, we will simply consider the binary classification (absence or existence of COVID-19) as there are far too few samples from the other diseases for any meaningful multi label classification.

4.3 Disease localisation

An equally important task is that of disease localisation. For this task, we utilise Class Activation Mappings (class), which generate a heatmap based on a specific class c and the activations of our final convolutional layer. It showcases the part of the image that the network focuses on when it aims to predict disease c. For instance, for Cardiomegaly, we would expect our network to focus on the heart region (figure 3a).

Let $S \times S \times K$ be the dimensions of the final convolution (K: channels-filters). In our case, all models have S=7. Let f_k be the activation (after applying ReLU) of the k-th 7×7 convolution and let $w_{c,k}$ be the corresponding weight which affects this convolution when inspecting class c. Then, we can generate a CAM $M_c = \sum_k f_k \cdot w_{c,k}$.

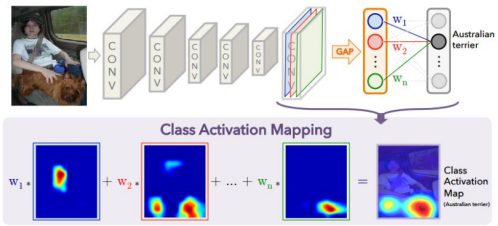


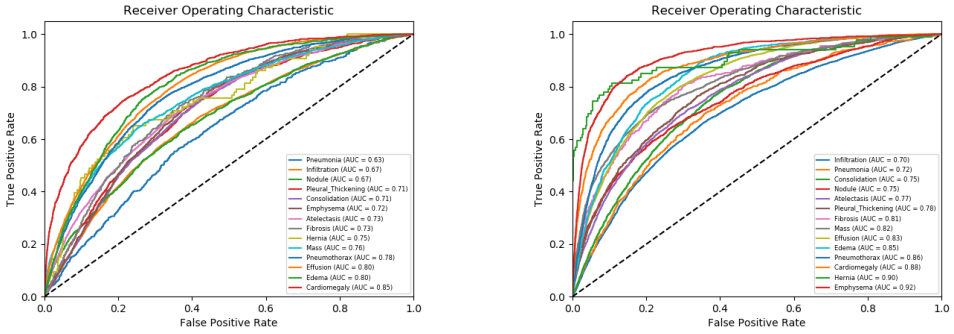
Figure 1: CAM illustration [1]

5 Experiments

5.1 Multi-label disease classification on ChestX-ray14

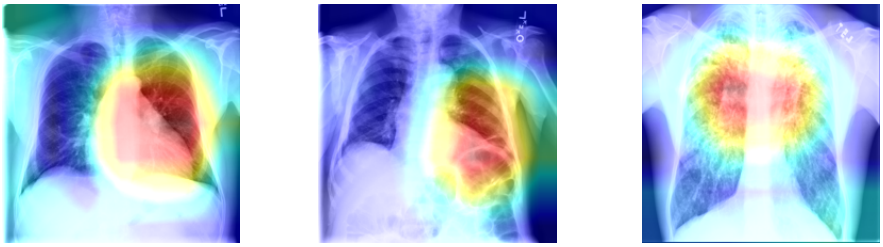
To employ and evaluate the models, the dataset was split into sets according to the included guidelines: 57% train, 20% validation and 23% test (of a total 112120 x-rays). The results (Figure 2a and 2b) bolster our prior hypothesis that the network benefits from being pre-trained. Our non pre-trained ResNet-50 managed to learn certain patterns (see Figure 4a) achieving an average ROC score of 0.74. On the other hand, the pre-trained DenseNet-121 managed to reach an average score of 0.81. Note that our pre-trained ResNet-50 performance lay in between the two shown models. The DenseNet was trained with an initial learning rate of 0.0001, a weight decay of $5e-5$ and reducing the learning rate after 5 batches without improvement of validation loss. The Resnet was trained with an initial learning rate of 0.001, a weight decay of $1e-5$ and reducing the learning rate after 1 batch without improvement of validation loss. Batches of 16 images were used in both models.

The heatmaps illustrate our network's perception of the diseaseses. All heat maps are generated with DenseNet-121 which exhibited the best ROC curves. We see, for instance, that it successfully focuses on the heart region (center-left) for Cardiomegaly (enlarged heart). We also see how it looks on different locations, depending on the disease, which are always, however, within the lung area.



(a) Non pretrained ResNet-50 (b) Pretrained DenseNet-121

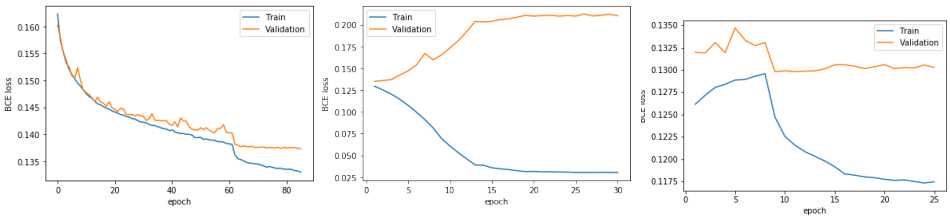
Figure 2: ROC curves of different model architectures pretrained and not pretrained on the Test set.



(a) (b) (c)

Figure 3: CAM for best performing model in 3 different diseases: (a) Cardiomegaly, (b) Atelectasis and (c) Fibrosis

In addition, our loss curves showcase the learning process. For our own non pre-trained ResNet-50 the loss curves for both validation and train decrease smoothly. On the other hand, utilising DenseNet-121 and applying only horizontal flipping (with a probability of 0.5) lead to early overfit, despite the use of weight decay, due to the model's increased capacity (for reference, looking at our training set, we managed a ROC score of 1 in almost every disease). To battle this issue, we applied Color Jitter, Random Rotations (from -20 to 20 degrees), Affine Transformations, data reshuffling and increased the weight decay. These techniques helped our model avoid overfitting (especially reshuffling) and managed to reach a final binary cross-entropy validation loss of 0.128.



(a) (b) (c)

Figure 4: Loss learning curves for models: (a) ResNet, (b) DenseNet121 applying horizontal flipping and (c) DensNet121 with additional augmentation (Reshuffling, Color Jitter and Affine Transformation)

Finally, our results were compared to other studies. It worth mentioning that this task is very difficult, due to different splittings of the data set. Therefore, we decided to include as a fair comparison a model created by Joseph, R. [10] since it splitted data set as the same way that we did. Table 1 can be interpreted as: Atelectasis (A), Cardiomegaly (B), Effusion (c), Infiltration (D), Mass (E), Nodule (F), Pneumonia (G), Pneumothorax (H), Consolidation (I), Edema (J), Emphysema (K), Fibrosis (L), Pleural Thickening (M) and Hernia (N).

Architechture	A	B	C	D	E	F	G	H	I	J	K	L	M	N
ResNet-50	0.7069	0.8141	0.7362	0.6128	0.5609	0.7164	0.6333	0.7891	-	-	-	-	-	-
CheXNet-121	0.8094	0.9248	0.8638	0.7345	0.8676	0.7802	0.7680	0.8887	0.7901	0.8878	0.9371	0.8047	0.8062	0.9164
Joseph, R. model	0.7355	0.8940	0.7727	0.6714	0.7841	0.6851	0.6832	0.8077	0.6932	0.7777	0.8279	0.7289	0.7250	0.9261
Our model (DenseNet-121)	0.7705	0.8823	0.8291	0.6992	0.8172	0.7536	0.7194	0.8602	0.7507	0.8467	0.9157	0.8150	0.7804	0.8996

Table 1: Comparison of ROC curves performance for each pathology [11]. We outperformed Joseph's model and produced similar results with CheXNet on a much larger test set.

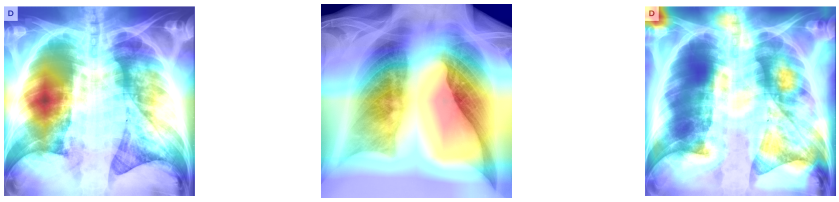
5.2 COVID-19 Classification and Localisation

Our next task is to apply our model for COVID-19 classification. For this task, we considered the binary classification of COVID-19 versus no COVID-19, as all the other diseases were far too few for any meaningful classification (the second most frequent illness had only 16 images).

After training our pretrained DenseNet-121 (on chestXray14) for 100 epochs and picking the one with the best ROC validation curve, our model was able to reach a ROC score of 0.88 on the test set. Note that when training with the non-pretrained DenseNet-121, we only achieved a ROC score of 0.73. To deal with the class imbalance, we oversampled our images in the training set in order to have an even number of representatives from each class. Finally, training with a model pretrained only on ImageNet gave similar results (ROC score 0.87), however, the heatmaps indicate that it may have found shortcuts and overfitted, focusing on data particularities irrelevant to the disease (figure 5c). This highlights that CAMs, in conjunction with traditional metrics (AUC), can be used for examining model robustness.

	Pretrained on ChestXray14	Pretrained only on ImageNet	Not Pretrained
ROC score on Test	0.88	0.87	0.73

Table 2: Roc score on COVID-19 dataset. Pretraining increased our ROC score by 0.1.



(a) (b) (c)

Figure 5: Models' generated heatmaps on COVID 19's x-rays, (a) and (b) correspond to the chestXnet pretrained model, while (c) to the model only pretrained on Imagenet.

5.3 Classification of ChestX-ray14 and COVID-19

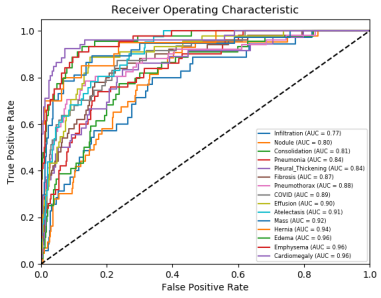


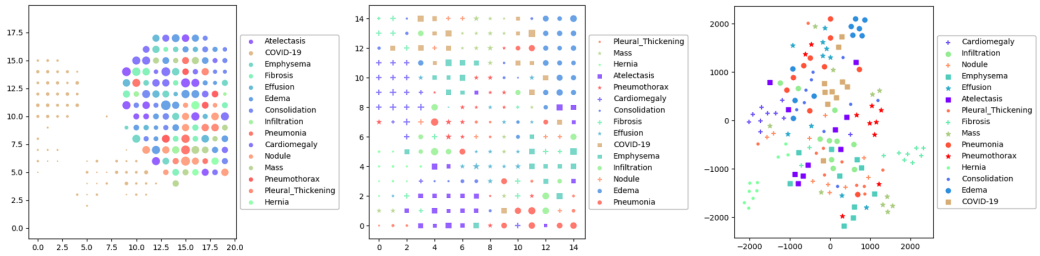
Figure 6: Test ROC of DenseNet121 classifying 15 diseases

It was considered an interesting experiment trying to apply our model similarly to how it would be used in a more realistic situation: As a useful clinical decision support tool, a machine learning model should be able to guide clinicians not only providing a binary classification of a specific disease, but also being able to detect a range of pathologies as broad as possible. For this task, we created a balanced dataset with over 200 samples of each of the 15 diseases (14 from ChestX-ray14 and COVID-19) and trained our pretrained DenseNet121 model. This was done for 100 epochs and, as in previous experiments, the model with the best ROC validation curve was picked. With this model, the network was able to reach a ROC score of 0.89 for COVID-19 on the test set while maintaining similar ROC scores for the remaining 14 diseases compared to the results obtained in section 5.1. As it can be observed, the ROC score for COVID-19 obtained in this experiment was slightly higher than that of section 5.2 (0.88). The

reason could be that the data set is balanced in this case, which helps the network to distinguish better among classes.

5.4 Network's Pathologies Landscape

After analysing our model classifications upon all the pathologies (including COVID-19), we decided to reason some of our results based on a visualization of all the pathologies spread on a 2-D grid. For such task, two methods were implemented: Self-Organizing Maps (figure 7b) and T-distributed Stochastic Neighbor Embedding (figure 7c). Figure 7a refers to the representation before training the model on the X-ray data set, while Figures 7b and 7c on the latent representation.



(a) Normal SOMs. (b) SOMs on the CNN (c) t-SNE on the CNN

Figure 7: Low dimension representation of the ChestX-ray14 and COVID-19 images using different approaches.

Some takeaways that can be taken from both charts it's the similarities that COVID-19 has with some particular classes. These similarities can be inferred based on the proximity between COVID-19 and other pathologies. For instance, in both scenarios, COVID-19 turned to be close to Edema, Pneumonia and Emphysema. Relying on this assumption, one way to boost our model in order to increase the performance to predict COVID-19 correctly, would be pre-train models upon data sets composed with more examples of these pathologies. This kind of technique can be beneficial when there is not many samples from one particular class, like COVID-19. Also, we see that the latent representation is more structured, separating the different diseases into clusters (compare 7a with 7b), while initially the model was focused on differentiating two data sets. However, after learning thoracic images the representations became more alike and the model managed to better cluster the diseases.

6 Conclusions

Our Deep Neural Network approach produced noteworthy results, showcasing great potential and robustness in diagnostic tasks. Its latent representation was able to properly separate the classes and created distinct clusters, by producing a new vector space that can model sample characteristics more accurately. Furthermore, we showed that the network can be utilized for other tasks, not just discriminative, by applying SOMs to uncover class relationships and using heatmaps (CAM) to highlight the affected region. Using the latter, the network's decision can be investigated to avoid overfitting due to data set particularities (e.g. small and unbalanced data sets). Our experiments show that an effective way of avoiding this is using pre-trained models, as they initialize the networks in better regions of the learning space, avoiding such undesirable overfit solutions.

Even though many studies have been conducted on top of X-rays thoracic disease detection, it's hard to infer that the performance of one method it's significantly better than the others, especially due to training and testing sets with different ratios between studies having a huge impact. Nevertheless, our approach reached state of the art performance and constitutes a robust diagnostic and disease localisation framework. Nonetheless, in order to achieve even better results we propose: (1) exploit further features provided by many of the medical data sets (e.g. age and gender); (2) discretize pneumonia into its origins (e.g. bacterial or viral); (3) collect more COVID-19 samples.

References

[1] Learning deep features for discriminative localization. http://cnnlocalization.csail.mit.edu/.

[2] Sars-cov-2 transmission. https://www.worldometers.info/coronavirus/transmission/.

[3] A. Chakrabarty, S. Mannan, and T. Cagin. Chapter 8 - inherently safer design. In A. Chakrabarty, S. Mannan, and T. Cagin, editors, *Multiscale Modeling for Process Safety Applications*, pages 339 – 396. Butterworth-Heinemann, Boston, 2016.

[4] J. P. Cohen, P. Morrison, and L. Dao. Covid-19 image data collection. *arXiv 2003.11597*, 2020.

[5] C. S. . T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 2019.

[6] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[7] M.-Y. Ng, E. Lee, J. Yang, F. Yang, X. Li, H. Wang, M. Lui, C. Lo, B. S. T. Leung, P. Khong, C. Hui, K.-y. Yuen, and M. Kuo. Imaging profile of the covid-19 infection: Radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, 2:e200034, 02 2020.

[8] N. I. of Health Chest. National institutes of health chest x-ray dataset. https://www.kaggle.com/nih-chest-xrays/data#Data_Entry_2017.csv, Feb 2018.

[9] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Y. Ding, A. Bagul, C. Langlotz, K. S. Shpanskaya, M. P. Lungren, and A. Y. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *CoRR*, abs/1711.05225, 2017.

[10] ryanpjoseph17. Chexnet. https://www.kaggle.com/ryanpjoseph17/chexnet?scriptVersionId=33836308, May 2020.

[11] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *CoRR*, abs/1705.02315, 2017.