Chest X-ray Disease Diagnosis

Anush Ananthakumar, Amol Singbal, Arindam Duttagupta
903272750, 903291649, 903327355
Georgia Institute of Technology

Abstract

X-rays are one of the most cost-effective and common medical diagnosis technique available in recent times. Since their discovery in 1895 by Rontgen, X rays are widely used in medical tests due to many desiring properties, mainly because they can penetrate less dense matter like skin but not bones which enables to obtain a clear image on a photographic paper. Accurate classification of X-ray images is a vital task as it is a precursor for future treatment. However, clinical diagnosis of X-ray images often require teams of trained radiologists to accurately estimate the type of disease which often consumes time and is expensive. Deep Learning Models have traditionally performed very well on classification problems and it has been shown that these models perform better with more training data. One major hurdle in creating large X-ray image datasets is the lack of resources for labeling so many images. The release of Openi dataset having 4143 images provided the data needed for training deep networks. In this research, the dataset to be used is the NIH Chest X-Ray Dataset, which is comprised of 112,120 X Ray Images from 30,805 unique patients. We present a CNN based approach along with a Faster Region proposal Convolutional Neural Network for accurately classifying a given X-ray Image in one of the 15 classes along with localization on the area of attention. The accuracy of the classifier and the area over the union are promising for this research with the classifier having an accuracy of over 54% over all the classes and a Jaccard Index of 0.5913 for the localizer which are comparable with the state of the art.

Youtube Link:

https://youtu.be/tPka_wppwUE

Introduction

Identifying abnormalities in the chest is frequently accomplished using chest radiographs as they are the most cost-effective medical examinations in practice. A large amount of chest X-ray (CXR) images are accumulated in many hospitals and this database can facilitate to build learning paradigms which can classify and identify the abnormalities with high precision. Traditional manual classification of CXR images is severely impacted by confusion, misdiagnosis and communication gaps with experts which can severely impact patient's health. Moreover localization of attention to identify the region of anomaly in these CXR images is a highly daunting task and requires an expert with lot of experience and training. Increasing research is conducted in the domain of disease classification and attention model automation algorithms which can improve computation speed and accuracy.

Recent advances in this domain of X-ray image classification increasing utilize deep learning techniques such as Long Short Term Memory Networks (LSTMs) [1], Convolutional Neural Networks (CNNs), recurrent attention models and reinforcement learning. The effectiveness of utilizing a deep convolutional neural network for pneumonia detection from chest x-ray has been demonstrated in [2] using their 121 layer CheXNet. This network has been developed for classification of 14 diseases and trained using the ChestX-ray14 dataset. Another recent research [3] focussed on classifying the CXR as either corresponding to frontal or lateral view based on features extracted such as body size ration, pyramid of histograms of orientation gradient, shape descriptor based on contour detection and image profile. This algorithm utilized a Support Vector Machines (SVM) [4] based classifier for using these features so as to classify the CXR into frontal or lateral categories. Further improvement on such classification has been proposed in [5] where an ensemble of deep convolutional networks is utilized for classification. Also this research proposes a method of spatial abnormality detection wherein occlusion based classification as described in [6] can localize the required pathology in the CXR. Another specific classification of disease is implemented in [7] which implements an attention guided CNN for classification of the Thorax disease. Chest X-Ray images are utilized to learn a three-branch attention guided convolutional neural network which integrates global information with learning from disease specific regions to generate a heat map and detect the anomalous region from the CXR. Such an approach highly improves the performance with average AUC greater than 0.85 and is compared with various deep learning frameworks and local learning networks.

Multi-Label classification is an important domain of research in radiology as a single radiograph can be determined to contain multiple pathological outcomes. Research in [8] aims to tackle the issue of multi-label classification by leveraging interdependencies among target labels using LSTMs to predict different pathological patterns from CXR. Such a method is effective in dealing with scares data by extracting relevant features from available samples. Further localization techniques such as CNN based classification followed by patch based segmentation passed through a fully connected network is utilized in [9] for improving classification and localization of thoracic diseases using CXR images. Such a model predicts the probability score of each patch which can be combined together to produce a non-rectangular localization of region in the input image.

Our research is similarly based on deep learning framework for image classification and detection of pathological region in the CXR images. This research utilizes the concepts of Convolutional Neural Networks as they have proved to be highly effective and provide accurate results as evidenced by the literature. Localization has been performed by a unique technique of Faster Region Proposal Convolutional Neural Network (Faster RCNN) which has scarcely been utilized in the literature for localization of pathological regions. This research builds on these concepts to build effective deep learning based framework for efficient classification which yields high accuracy and reduced computation for localization of the region of interest. This would greatly assist medical practitioners to effectively identify and locate pathogens from chest radiographs. Although this research utilizes only a subset of available dataset of images and highly scarce localization ground truth to develop the proposed algorithm, it serves as a useful algorithm for representation learning, feature extraction, classification and localization in radiographs.

Section 2 consists of the proposed deep learning framework followed by the technical implementation and detailed analysis of the algorithms utilized in this research in Section 3. Further Section 4 highlights the observed results and performance of the classification and localization algorithms for various pathogens. Finally we initiate a discussion on such a research and present the conclusion to this research in Section 5.

Methodology

This research aims to automate the classification of CXR data into different categories using a convolutional neural network. This is followed by using a localization algorithm such as Region Proposal Convolutional Neural Network (R-CNN) which essentially creates a bounding box around the region which requires attention and contributes towards improving the confidence of the classified disease. Such a framework greatly reduces computation time due to the high speed of both the classifier and the localization algorithm which are based on deep learning. Initially the CXR image is histogram equalized to adjust image intensities to enhance contrast. This is essential in X-ray images so as to delineate the regions effectively and develop a more robust learning framework.

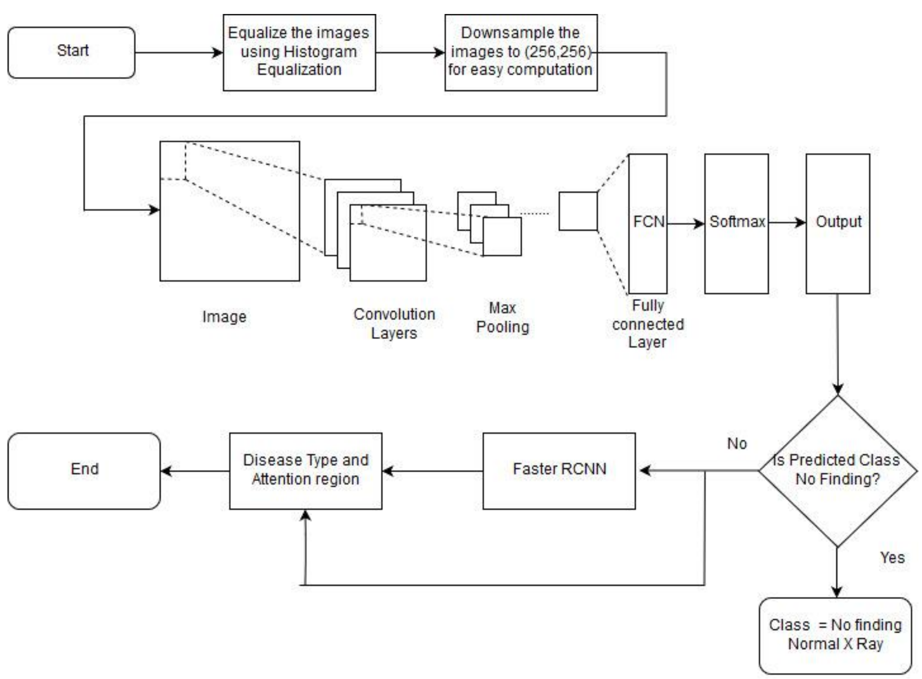

Figure 1: Working of Proposed Algorithm

The region proposal based bounding box generation is trained on dataset of scarcely labelled bounding box regions. Hence this algorithm is efficient in generating the attention regions for only certain disease classes on which it is trained. This research is unique as it utilizes a Faster R-CNN network which has not been researched previously in the literature. Such a method further augments the performance of the classification as the detected region is found only if the image is predicted to contain a specific disease learned during the training phase. Using deep networks similar to VGG16 greatly improve the performance of the classification which can perform improved feature extraction and learns the spatial and higher dimension similarity for classifying the CXR images into different pathology classes. The overall workflow of the proposed algorithm is detailed in Fig 1. The test image is initially histogram equalized to improve contrast followed by down-sampling so as to reduce the computation required by the classification framework. Further as the classifier is trained on images of size 256x256, the test image is resized to fit the input size of the classifier framework. Classified results are then utilized to understand if a localization is required. The attention region is highlighted by the use of Faster RCNN which can efficiently generate bounding boxes which can be used along with the classification result to understand the disease in the CXR along with the region of attention.

Detailed explanation of the different components of the research are discussed successively.

A. Histogram equalization:

Histogram equalization is a technique used for adjusting image intensities to maximize global contrast. It is based on the principle that maximum contrast is achieved when the image has a uniform histogram viz, all pixel intensities are equally probable. Hence, given an image, this technique aims to convert the image histogram as close as possible to the 'ideal' histogram. Transformation is perfect when the image is continuous function which is practically never the case. An advantage of Histogram Equalization is that it is a fairly simple technique and can be easily inverted to get the original histogram back [13]. This technique is especially useful while dealing with medical images as it helps areas of lower contrast gain a higher contrast which makes them distinguishable from the image. The equalized histogram is generally estimated by initially computing the cumulative distribution function and then mapping the function back to the transformed gray levels, which is a computationally inexpensive procedure.

B. Convolutional Neural Networks:

Convolutional Neural Networks (CNNs) find numerous applications in image/video classification, semantic segmentation etc. As the name suggests, the hidden layers in a CNN would be convolutional layers to extract a set of more complex features as compared to the previous layer. Apart from convolutional layers, max pool, fully connected layers are also present. Training a CNN often involves selection of a variety of hyper parameters such as filter size, convolution stride, pooling stride etc. The below figure illustrates the in depth architecture of a CNN [14]. Every layer produces a set of feature maps characterizing a particular feature in the image. For example, for the first convolutional layer might generate a map which shows all image edges. Or the map could give an estimate of the gradients present. The successive layers generate more complex feature maps from these simple features which are then used for accurate classification of an image. Max pooling is usually deployed to control the number of parameters to train on.

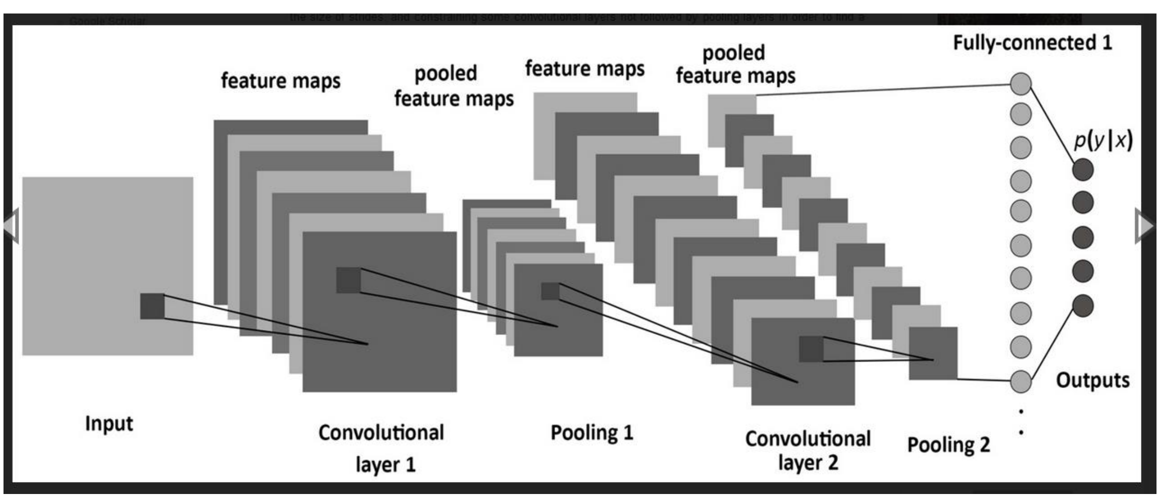

Figure 2: CNN Architecture

C. Faster Region Proposal Convolutional Neural Network (Faster R-CNN):

Faster R-CNN [10] has two networks: region proposal network (RPN) for generating region proposals and a network using these proposals to detect objects.The faster R-CNN based detector is essentially derived from R-CNN and fast R-CNN. R-CNNs begin with a region proposal framework followed by a CNN based feature extractor and then classification of the features. However the region proposal framework utilized in R-CNN was computationally expensive and slowed down the detection process. Faster R-CNN improves the performance by reducing the time required to compute the region proposals by changing the algorithm for the Region Proposal Networks (RPNs). The intuition here is that the computed region proposals depend on features calculated in the forward propagation stage of the CNN, and thus, during real time testing, the results of the convolution layers can be shared, thereby reducing the cost of computation to a marginally small amount. RPN predicts the possibility of an anchor/bounding box being background or foreground, and refine the anchor. The overall loss of the RPN comprises of the classification loss and the regression loss.

$$L_{loc}(t^u, v) = \sum_{i \in \{x,y,w,h\}} smooth_{L_1}(t_i^u - v_i)$$

$$where$$

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases}$$

The working of the Faster RCNN network is explained in Figure 2. Using such a technique to localize the position of the anomalies or attention region has not been previously experimented in the literature to the best of our knowledge. Such a localization provides dual benefit of localizing the required region and reducing computation time. Further this object detector framework also gives a confidence score for each bounding box which can be utilized to remove any regions which have been predicted to contain anomalies with lower confidence. The output of this network is then combined with the classification label to generate the final output.

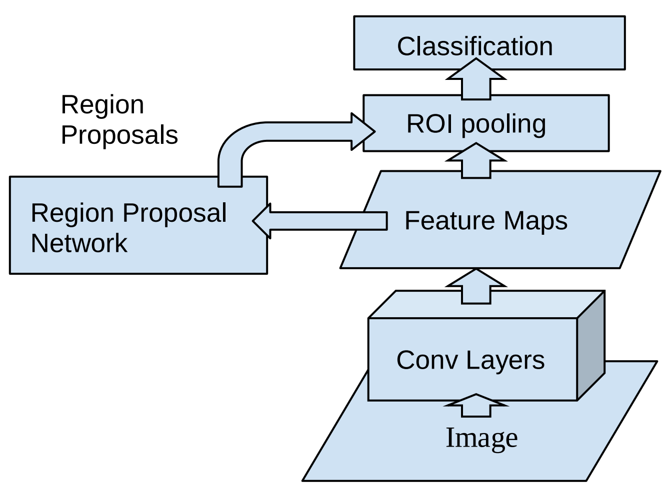

Figure 3: Faster R-CNN Architecture

Data and Experimental Setup

The dataset used in this research is the publicly available NIH Chest X-Ray [11] dataset which is comprised of 112,120 X-Ray Images with disease labels from 30,805 unique patients.

Each Image is a 1024x1024 image having 8-bit precision. The data also consists of:

1. A csv file listing the Image Index, the label and the bounding box specifications (height, width and the position in the image)

2. A patient data csv file which consists of patient parameters such as Age, Gender, ID, Disease Labels etc.

The NIH Dataset after reducing to single label classes consists a total of 15 classes each of which denote a specific kind of chest ailment. For the purpose of this paper, we would not be using the full NIH Dataset, but rather a 5% random sample of the NIH. This reduced dataset contains 5,606 Images spread across 15 classes as described in [12]. As part of the preprocessing step, we first compute the histogram equalized version of the training Images to enable better contrast and then down sample them to 256x256 before feeding them into our network.

For training the Faster-RCNN framework however, we consider the csv file containing the bounding box positions for 984 different images belonging to 9 different classes. This is because we could not find the bounding box labels associated with the reduced dataset of NIH that was used for training the convolutional neural network. This research utilizes Python3 for training the classifiers.

Results

The proposed algorithm has been tested on the National Institutes of Health Chest X-Ray Dataset. Initial pre-processing is essential as it improves the contrast and helps in extracting relevant features by the classifier network. Figure 4 depicts the result of the equalization process on 2 different CXR images.

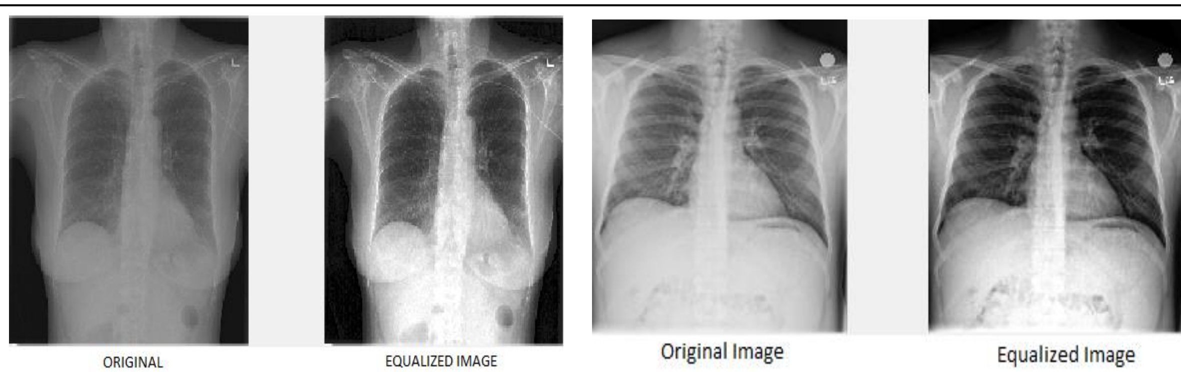

Figure 4: Histogram Equalization

The performance of the Faster R-CNN is evaluated using the Jaccard Index which computes the Intersection over Union so as to identify the performance of the localization algorithm. The performance of the Faster RCNN yields a Jaccard Index of 0.5913 on 196 testing images. The output of the localization algorithm along with the classification result has been depicted in Figure 5. Such a localization would enable to build attention models regarding which regions need to be focussed so as to understand the anomaly in the CXR image.

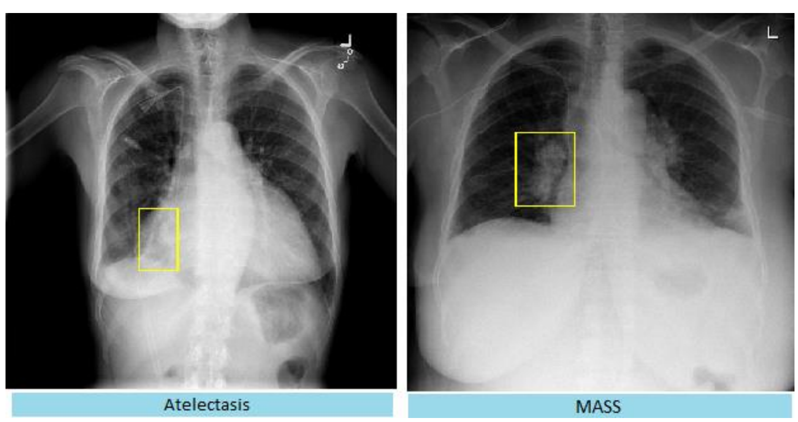

Figure 5: Result of Algorithm

The performance of the Convolutional Neural Network on the reduced 5% dataset has been highlighted in Figure 6. Reduction of the dataset for single label classification yields 15 distinct disease classes. The confusion matrix presented in Figure 6, yields the performance of the multiclass classifier for each disease class and also shows the percentage of misclassifications so as to understand which classes are considered similar by the classifier. Although the overall performance of the classifier is approximately 0.5481, the individual classification is good and the lower overall accuracy can be attributed to poor performance in just a few classes.

The various class labels are: 1. Atelectasis, 2. Consolidation, 3. Infiltration, 4. Pneumothorax, 5. Edema, 6. Emphysema, 7. Fibrosis, 8. Effusion, 9. Pneumonia, 10. Pleural thickening, 11. Cardiomegaly, 12. Nodule Mass, 13. Hernia and 14. No Labels respectively. It can be observed that many of the images of Consolidation, Infiltration, Pneumothorax have been misclassified to be normal images with 'No Finding'. Also the performance of the classifier is low for CXR images of Pneumonia whereas it performs the best when evaluating Effusion in CXR images. Such a confusion matrix provides an insight into the ability and performance of the classifier.

Original Labels

Predicted Labels

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.67	0.12	0.03	0	0	0	0.03	0.03	0	0	0	0.07	0	0	0.07
2	0.22	0.42	0	0	0.02	0	0	0	0	0	0	0	0	0	0.28
3	0	0	0.41	0	0	0	0	0	0	0	0	0	0	0	0.19
4	0	0	0	0.43	0	0	0.24	0	0.11	0.03	0	0.03	0	0	0.26
5	0.03	0	0	0	0.64	0	0	0.11	0	0	0.03	0	0.19	0	0
6	0	0	0	0.12	0	0.52	0	0	0.21	0	0	0.12	0	0.03	0
7	0	0.2	0	0	0	0.09	0.71	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0.1	0	0.88	0	0	0.02	0	0	0	0
9	0.1	0	0	0	0.12	0.03	0.08	0	0.23	0.1	0	0	0.23	0.11	0
10	0	0	0	0.23	0	0	0	0	0	0.32	0.32	0	0	0	0.13
11	0	0	0.07	0.03	0	0	0.1	0	0	0	0.61	0	0	0.1	0.09
12	0	0.12	0	0.03	0	0	0	0.21	0.12	0	0	0.43	0.09	0	0
13	0.12	0.03	0	0.12	0	0.09	0	0	0	0	0	0.1	0.54	0	0
14	0	0	0	0.1	0	0.03	0.03	0.03	0.03	0.09	0	0	0	0.69	0
15	0	0.14	0.03	0.03	0	0	0	0	0	0	0	0	0	0	0.8

Figure 6: Confusion matrix of CNN

Conclusion

The proposed research develops a single channel framework for classifying different types of CXR images along with building a localization framework for building attention regions around anomalies in the X-Rays. This is accomplished by utilizing deep learning models comprising of Convolutional Neural Networks and Faster Region Proposal CNN. CXR images after contrast enhancement are passed through the classifier and localizer thereby providing the disease class and region of interest. The performance of the proposed CNN gives an overall accuracy of 54.81% for classification into 15 different classes. Further deeper analysis of the CNN yields the diseases in which the performance is superior and the classes with poor performance, thereby enabling researchers to utilize the proposed framework for only certain disease categories. Further the localization of the area of interest has an intersection over union of 0.5913 which proves the performance of the localizer and its ability to detect regions of interest. Computation is further reduced by only utilizing the localization algorithm on CXR images classified to contain pathology. Such a research would contribute towards the growing interest in automating medical image classifications and segmentation.

The performance of the proposed algorithm can be further increased in future research wherein the training data utilizes the complete NIH dataset of over 112000 images of different categories. Further improvement is possible by displaying a semantic map which highlights the relative areas of interest and the probability of the region being an anomaly. Such improvements would prove highly convenient and more accurate for medical practitioners.

References

[1] Sepp Hochreiter*;* Jürgen Schmidhuber *(1997).* "Long short-term memory". Neural Computation*. 9 (8): 1735–1780.* doi*:*10.1162

[2] P. Rajpurkar et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning Dec. 2017 [online] Available: https://arxiv.org/abs/1711.05225

[3] Xue, Zhiyun & You, Daekeun & Candemir, Sema & Jaeger, Stefan & Antani, Sameer & Long, L & R. Thoma, George. (2015). Chest X-ray Image View Classification. Proceedings - IEEE Symposium on Computer-Based Medical Systems. 2015. 66-71. 10.1109/CBMS.2015.49.

[4] Cortes, Corinna*; Vapnik, Vladimir N. (1995). "Support-vector networks".* Machine Learning*. 20 (3): 273–297*

[5] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest x-rays using deep convolutional neural networks," arXiv preprint arXiv:1705.09850, 2017.

[6] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision. Springer, 2014, pp. 818–833

[7] Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L., Yang, Y.: Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification.

[8] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," arXiv preprint arXiv:1710.10501, 2017

[9] Li, Z., Wang, C., Han, M., Xue, Y., Wei, W., Li, L., Li, F.: Thoracic disease identification and localization with limited supervision. In: CoRR (2017)

[10] Ren, K. He, R. Girshick, and J. Sun, 2015, Faster R-CNN: Towards real-time object detection with region proposal networks: Advances in Neural Information Processing Systems

[11] https://openi.nlm.nih.gov/

[12] https://www.kaggle.com/nih-chest-xrays/sample

[13] https://www.math.uci.edu/icamp/courses/math77c/demos/hist_eq.pdf

[14] S. Albelwi A. Mahmood "A framework for designing the architectures of deep convolutional neural networks" Entropy vol. 19 no. 6, 2017