Course project in Deep Learning: Chest X-rays

Röngten radiation images of the chest, or more simply chest x-rays, is one of the most common radiographical imaging examinations used in suspicions of and diagnostic of diseases related to the chest. The lungs are the number one organ in focus during these examinations, but also the throat and heart. Even though it exists multiple algorithmic methods and models that are being used during the diagnosis of (for example) lung cancer or dark shadows (or spots) on the lungs, there are still challenges when it comes to diagnosing other lung- or chest-related diagnoses. These challenges often arise due to their observations in the chest x-ray are similar, and thereafter may lead to misdiagnosis or doubt with the x-ray and/or the radiograph that performed the examination. By constructing a deep learning multi-label classification model, we can train this model to identify if a patient has one or more of the 14 most common observations that are seen in a chest x-ray.

Motivation

The motivation to construct such a model is to contribute to radiologists and diagnoses to interpretative chest x-ray better. The model will also act as an alternative 'second opinion' by localizing and categorizing the observations and potential diagnostic. This will make the diagnosis of patients more effective, minimize the potential waiting time for a diagnosis and minimize wrong diagnosis. This potential misdiagnosis can potentially occur due to pressure (in the form of little time or pressure from patients and/or doctors and specialists that have referred the patient to x-ray) or uncertainty due to that many of the observations look similar when it comes to viewing them in an x-ray image. Such a model that can contribute to diagnosing, will potentially minimize this pressure and minimize the uncertainty that occur when two observations are similar in x-rays.

Discloser

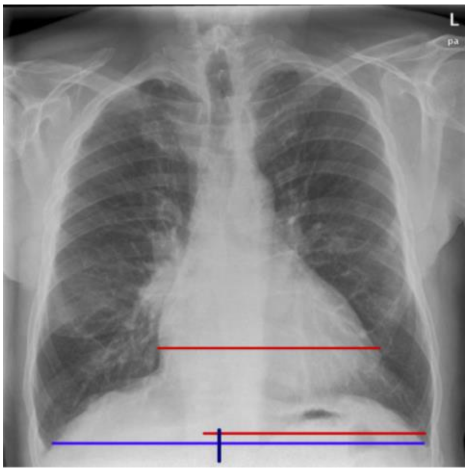
Before I start explaining my progress and how I worked, I would like to point out that I did not achieve as great a success with my deep learning model as I hoped. I have tried many times during this semester to make a good classifier to make the training more efficient and have a higher accuracy on the validation set for making predictions. This I was not successful with, so I tried making multiple classifiers during this duration of the project before it was due. So, what I decided to do was to deliver my progress so far: my best classifiers for if a patient has Pneumothorax (punctured lung) or not and my progress so far on my multilabel classifier.

The datasets

During this project, I have used many different datasets for experimentation, in hopes that it will give me a better understanding, due to my lack of medical and radiographical knowledge, and better models and predictions. For my first classifier, I used the SIIM-ACR Pneumothorax Segmentation dataset, where my main goal was to see if I can build a classifier that gave me more than a percentage of my model's accuracy. My original plan was to use this classifier (marked with to make *Collapsed lung or not* – classifier in the repository) on the larger CheXpert dataset for all 14 observations. The SIIM-ACR dataset consisted of a subset of 15 302 chest x-rays that were weakly labeled, the meaning will be returned to, all the states of pneumothorax, and images that marked as no findings. The second dataset that I had planned on making the predictions on using the first classifier that I made, where called the CheXpert[1] dataset. This dataset consisted of 224 316 X-rays from 65 240 patients. The dataset is also a competition for automatic chest X-ray interpretation, where the pictures contain uncertain labels and references to labels that are given by a radiologist that has followed the standard evaluation set. This is the case in the first dataset as well. The X-rays are labeled with the following categories (please note that the images in this part of the report is not part of the actual dataset):

[1] https://stanfordmlgroup.github.io/competitions/chexpert/

Enlarged Cardio mediastinum[2]:



meaning that the heart of the patient is larger than normal, as one can see in the figure. An enlarged heart is observed when the heart is more than half of the size of the whole chest cavity. Here in the figure, we can see that the heart is more than half the size of the chest cavity. The red line represents the size of the heart, and the blue mark represents the size of the chest cavity, and the dark blue mark marks the middle (50%) of the chest cavity. This one of a type enlargement of the heart.

Cardiomegaly;

is an observation that has the same observations as in the Enlarged Heart, but it also includes other parts of the heart like arteries. This observation includes all kinds of enlargements of the heart and the observations for an enlarged heart.

[2] https://www.google.com/url?sa=i&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FMediastinum&psig=AOvVaw2QE1EQRyfbILAKSNYzySX-&ust=1676722394917000&source=images&cd=vfe&ved=0CBAQjhxqFwoTCPCZ5-fDnP0CFQAAAAAdAAAAABAE

Lung Opacity*[3] *;

areas in the lungs that are supposed to be the lung, meaning that there are no white specks or anything because the air in an x-ray is black or dark, and have white densities there instead of being a dark lung shape.

Lung lesion ;

nodules or also known as incipient cancerous tumors. This observation is observed when we see small white specs on the x-ray or larger ones that indicate a possible tumor.

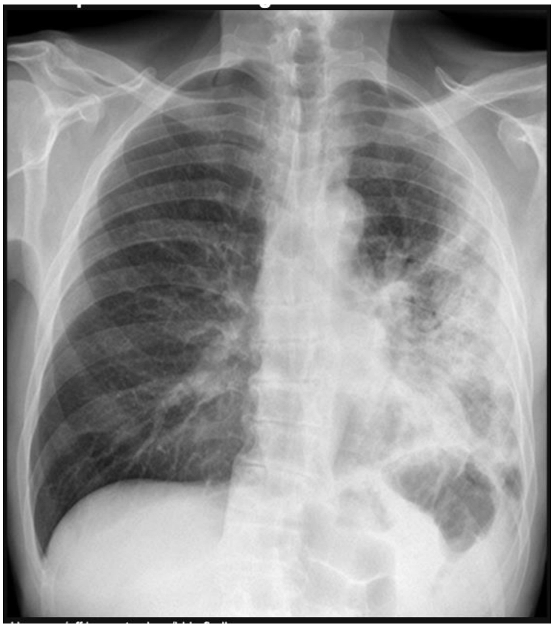
Edema ;

Liquid in the lungs

Consolidation ;

the lungs are full of pus, liquid, blood and/or other neoplastic cells.

[3] https://www.stepwards.com/?page_id=25474



Pneumonia ;

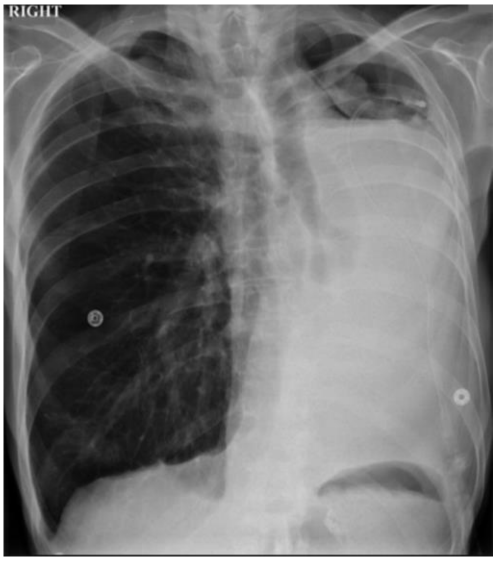
the lungs are filled with pus, an liquid infection or other types of liquids.

Atelectasis ;

collapse of a lung or some kind of closure or blockage in parts of the lung. This results in no air in the lungs, therefore giving the image white parts or a completely white side as shown in the figure to the right.

Pneumothorax ;

lungcollapse. Either caused by an internal puncture or an opening in the chest wall.



Pleural effusion;

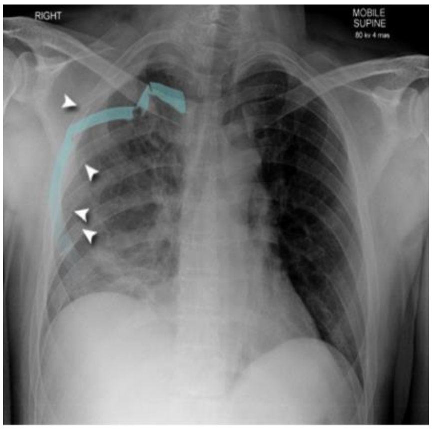
water on the lungs

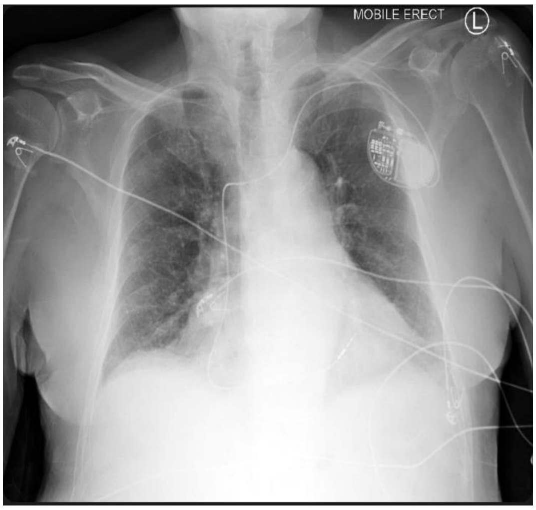
Pleural other;

disturbance or foreign bodies in the lung sac that surrounds the lung.

Fracture;

broken ribs or other kinds of fractures that can happen in the chest, like collarbone for example.



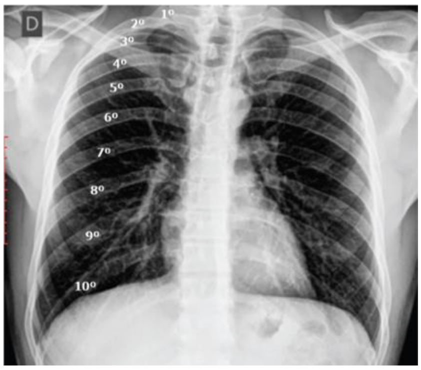


Support devices;

wires or small medical apparatus and devices. In the picture you can see a pacemaker in the upper corner of the left lung and some wires that are connected to some other equipment on outside of the skin.

No finding;

indicating that there where none of the other observations mention before. This means that we are dealing with the image of a pair of normal lungs.



The data and their uncertainty

Since the documentation states that many of the X-rays are marked with one or more observations, indicates that the classification of the X-rays is a semi-supervised problem. The reason behind this comes from how the dataset is made. The dataset is made by taking all the observations that are to be found in the patient journal of the person that the X-ray was taken. In the cases where there is more than one label is due to the uncertainty used in the wording of the journal to the patient. The wording containing words like “possible”, “it is not possible to conclude” or similar, has made this classification task a multilabel classification problem. The uncertainty connected to the data may also come from the x-rays that were taken, contained irregularities, and may not also have been taken in the optimal way of taking an x-ray. The optimal way will be discussed later.

This uncertainty in the data may also lead to uncertain predictions or uncertainty in predictions. But to look at it in a more positive way, the model will be better prepared to handle observations in chest X-rays that are unclear or blurry, taken in a non-optimal way, and/or has irregularities. But if we do not have good enough data to train the model, it will also give us challenges.

My assessments and my troubles

There are many ways to solve this problem regarding making a model that can read an image and name the observations that can lead to a diagnosis. For my first classifier, I had plans to just follow the instructions given in the tutorial given by the webpage to the fastai-library. One of the things I discovered working with that approach was that it seemed to be a bit more complicated than what I intended my classifier to be. So I made a note that I wanted to make a couple of other classifiers just for experimentation´s sake. But I used the tutorial as a guideline for how to build my code and make it more explanatory.

I also had the idea of making my own version of the dataset that I wanted to use as the dataset to make predictions on, but this seemed to be more work than reward. When I realized this I found out that I should have worked more with the dataset. Then I would have realized my big mistake when it came to labeling the data in my try at making predictions on the whole dataset, this can be seen in my code for the Allround classifier in my repository. Here it is obvious to everyone but me, that I am labeling my pictures with the name of the study taken from the patient instead of identifying the observations in the image. And if you cannot see it in my code directly, you see it very clearly in my printout as one works one way through my code.

One of my ideas for solving my problems with my datasets, was to take the one that were given in the task and modify it. My plan was to take the dataset and use those who were labeled more than ones and reorginize all of them into each of the different observations. By this reorganisation I hoped to achive a more broader way to classify them due to that the training were done more than once, like I would have done without the possible reorganization. Fortunatly I soon realised after reading the terms and conditions of usage of the CheXpert dataset, that the data cannot be copied without permission from the university of Standford. This strategy is very timeconsuming and it does not have a great guarantee of huge success, therefore this idea just remained an idea.

I also had a strategy of cleaning the data a lot more, but I wanted to feel the feeling of mastery of working with uncleaned data and get a decent model as a result. My plan, if I came long enough in the progress, was to clean the data so I only was getting frontal images and not just lateral. As one can see from the code in both of my classifiers, there are both frontal and lateral images in both training sets to the classifiers, so this cleaning of the data was not done.

Based on most of the arguments given above, I decided to make a classifier that classifies if a patient has a collapsed lung or not and use this to classify the whole CheXpert dataset afterward. Unfortunately, as you can see, I did not accomplish that.

Ideal X-ray images

The plan was to have all the images in the same format due to a risk of poor alignment[4] in the images, this is very common with dealing with X-rays and the risk of getting them not perfect (perfection is ideal in the situation of a possible diagnosis). This idea I got from the first article that used the CheXpert dataset was recommended to solve these problems early on in the coding process, but I ended up doing it at a later stage than recommended. One of the risks one can run into when diagnosing and classification of the observations is the bad quality of the pictures and that the picture has big disorientation[5], the article recommended rescaling the pictures again and feeding them to a template matching algorithm to localize the lung itself in the X-ray. This I did not do, and the reason is that I do not possess the right amount of knowledge to make such an algorithm, I tried finding a way to incorporate it into my model at the start of the semester in one of the mini-project that we had with not the greatest amount of success. I also tried to make a classification model from scratch but realized soon after my first try that it was not very efficient when it came to training time, learning time, and prediction time. And because of this, the idea was quickly discarded. It was also recommended to do something to handle or reduce the source-dependent variation, but I do not have the knowledge and/or time to accomplish it.

My plans also included the usage of ensembling techniques, but this is something that I again do not have the correct knowledge and the knowledge that I have regarding this technique I am not confident enough to use. Of course, if I used it will give me, according to all the research I did before starting this project, give me a higher prediction score.

Cross-entropy loss[6]

This is a cost function that is used to optimize classification models, therefore it is important to maintain and optimize it if possible. And since it is a loss function, the optimal situation is to have it as close as possible to 0. These measurements are used when we adjust the weights during the training of the model. And again the goal is to minimize the loss, and by doing so we will get a better model. I tried to make a misclassification error[7] at the end of the first classifier that I made. I made this in order to have some kind of metric that gives the percentage of the observations that were incorrectly classified and predicted

[4] https://musculoskeletalkey.com/x-ray-equipment-alignment-and-patient-safety/

[5] https://musculoskeletalkey.com/x-ray-equipment-alignment-and-patient-safety/

[6] https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e

[7] https://www.statology.org/misclassification-rate/

in my model. And again we want these values to be as low as possible, which as you can see in the models I had variable success.

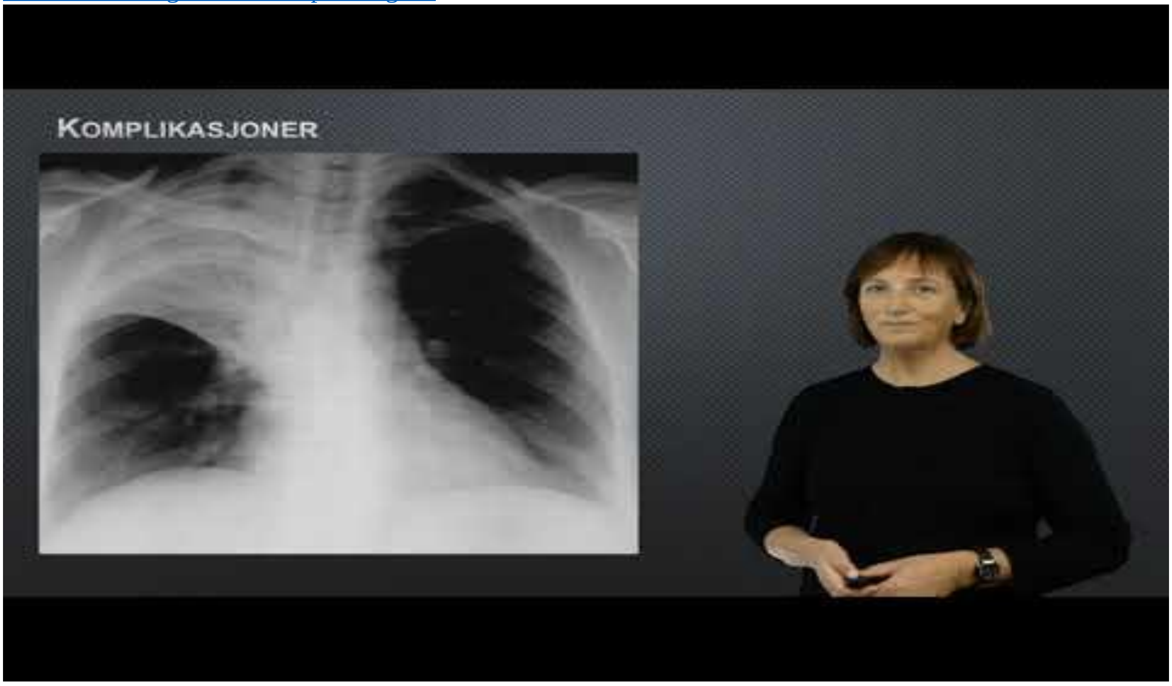
Conclusion

Even thou I had big plans and high hopes for my end results, I know I could have achieved a lot more than what I have had if I discovered my mistake earlier. There are many ways I could have made a much better model and have cleaned the data a lot more and trained it for a longer period of time, but since I thought I had made a good classifier and I just didn't find the way to transfer my classifier in order to predict on other datasets then the one it was trained and validated on. This was very frustrating and I decided that I should maybe start all over again and build a new classifier and then use it directly on the big dataset. This was not a huge success either as I did not pay as much attention to how my model did during the preparation of the dataset for training or when I started to explore the dataset. Again I should have made the discovery of the mistake earlier, but now I know that I need to be more observant when I work with datasets in a field I am not familiar with and in this case, I am not a doctor, nor do I have any medical training or background. But on the bright side, I have learned a lot from this project regarding the process of making a model and having it end up making predictions (even if there is just in this case of one class, disease or no disease) and fun fact: in order to make sure that I get the right predictions, I learned how to read X-rays in preparation for working with the data.

References used

https://arxiv.org/pdf/1911.06475.pdf

https://arxiv.org/pdf/2106.05915.pdf
Hvordan se lungebetennelse på røntgen?



Chest X Rays (CXR) Made Easy! - Learn in 10 Minutes!