Course project in Deep Learning DAT255: Chest X-rays (second try)

Imaging of the chest using Röngten radiation waves, or more commonly known as chest x-rays, is the most common radiographical imaging technique and examination used in suspicion of and diagnostic of diseases related to the chest. The lungs are the organ most in focus during these examinations, but also includes other organs like throat and heart. Despite the existing numerus algorithmic methods and models that are being used during diagnosis of (for instance) lung cancer or dark shadows, or sports, there still exists challenges when it comes to diagnosis of other lung- or chest-related diagnoses. These challenges are most likely to arise due to their observations in the x-ray are similar, and thereafter could lead to wrong diagnosis or create doubt with the x-ray image and/or the radiograph who performed the examination. By creating and constructing a multi-label classification model, one can train the model to identify if a patient has one or more of the most common observations that can be seen in a chest x-ray.

Motivation

The motivation behind the idea to construct a model, as described above in the introduction, is to contribute to radiologists and diagnoses to interpretative chest x-ray images better. The model can, or will if done properly, act as an alternative "second opinion" by localizing and categorizing the observations, and thereafter make the potential diagnostic. This will make diagnosis of patients more effective, minimize the potential waiting time for a diagnosis and minimize the potential for a wrong diagnosis. This potential misdiagnosis can potentially due to time pressure (little time or pressure from patients and/or doctors and specialists that has requested the x-ray), or the uncertainty due to the similarities the observations have with each other when viewing them in an x-ray image. Such a model that have the potential to contribute to diagnosing, can potentially minimize this pressure previously described, and the potential uncertainty that can occur when two similar observations occur in the same x-ray image.

Discloser

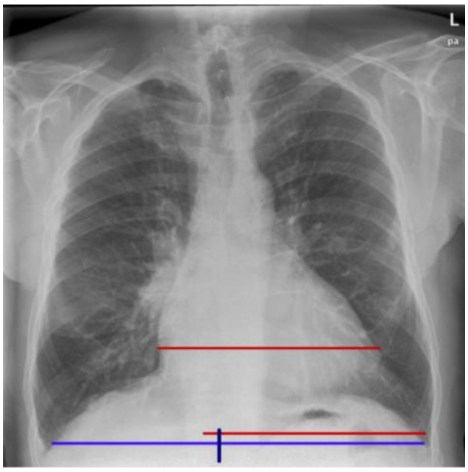
Prior to explaining my progress and my work, I would like to point out that I have already given this assignment a try before, in the last spring semester. But due to the fact that I did not achieve as great success with my model as I had ambitions and hoped to achieve. I tried many times during last semester, to create a good classifier and a good base for my model. A good classifier would make the training of the model more efficient and have a higher accuracy on the validation set for making predictions. That time I was not successful with, even thou I tried to make multiple classifiers to try achieving my goal during the duration of last spring semester. I decided for my try to deliver my progress and what I had achieved then. My last try has now bothered me for almost a year, and I am over the moon that I have gotten a second try at this assignment. My plan is to try and see if I can improve the classifier I delivered last spring, a classifier for if a patient has Pneumothorax, punctured lung, or not. If this does not work or I do not achieve an acceptable accuracy on my new model, I will try to make a classifier for if a patient has pneumonia or not, like other used for their base classifier when classifying the CheXpert-dataset which is the same dataset I will be using.

The dataset

Last spring I used many different datasets for experimentation, in the hopes that it would give me a better understanding of the images, due to the lack of my medical and radiographical knowledge and improve the models and predictions. Last year I used the SIIM-ACR Pneumothorax Segmentation dataset, where I had a goal to see if I could build a classifier which could give me a more accurate accuracy for my model. My original plan at the time was to use this classifier (I marked it with Collapsed lung or not – classifier in the repository in my answer from the last year spring semester) on the larger CheXpert-dataset for all 14 observations, given in the assignment and later in this report. The SIIM-ACR dataset consisted of the subset containing 15302 x-rays that were poorly or weakly labeled, will be defined later in the report, as the stages of pneumothorax and images marked with no finding. The second dataset that I used in last spring, I also plan on using in this try as well, that I plan on making predictions using the classifier that I made. This dataset is called the CheXpert[1] dataset. This is a dataset consisting of 224316 x-rays from 65240 patients. The dataset is also used for a competition for automatic chest x-ray interpretation. Here the images contain uncertain labels and references to labels labeled by a radiologist who followed the standard evaluation set. The same is done for the other dataset as well. The x-rays are labeled with these labels or categories, note that the images in this report is not part of the actual datasets:

[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.

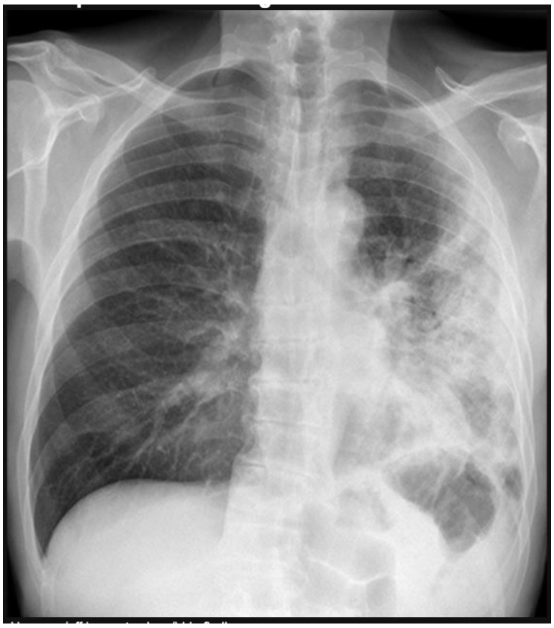
Edema ;

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 ;

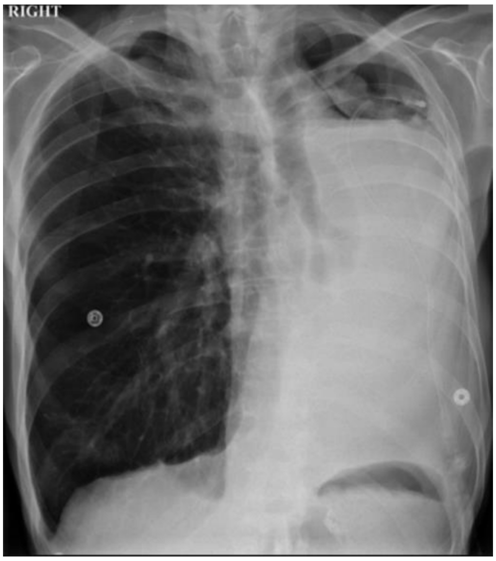
the lungs are filled with pus, an liquid infection or other types of liquids.

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;

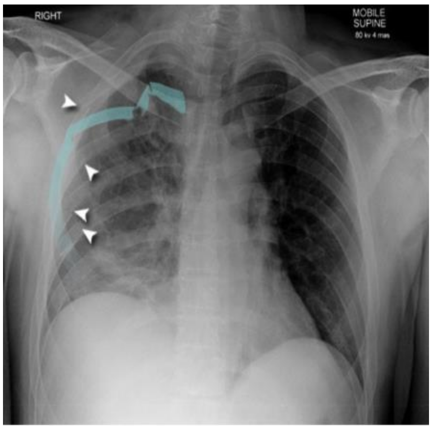
water on the lungs

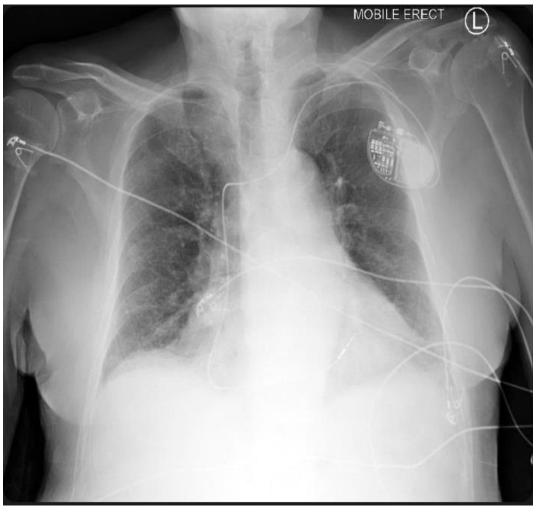
Pleural other;

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.



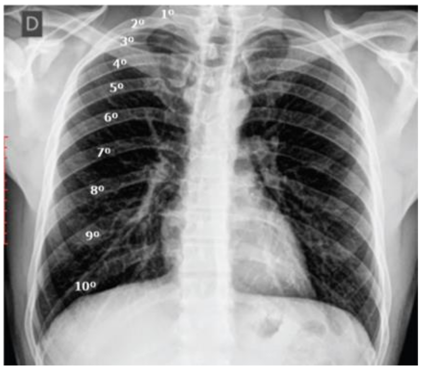


Support devices;

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.



For my second go at this assignment, I decided to do a little bit different approach, where instead of making a single label classifier (is there a specific observation present or not), I will be doing a multilabel classifier instead. This is because I want to try something new, and I want to see if I am more successful.

The data and their uncertainty:

The documentation of the dataset states that many of the of the x-rays are multilabel, meaning that the problem is a semi-supervised classification problem. This comes from how the dataset is made. The dataset is created by taking all the observations that were in the x-rays and that were present in the patient´s journal. In the cases where there was more than one label, comes the uncertainty caused by the wording in the patient journal. Wording containing statements such as “possible”, “it is not possible to conclude” or similar words and statements of uncertainty, makes the problem a multilabel problem. Another uncertainty aspect that comes from the dataset, comes from the imaging techniques of the x-rays; they may contain irregularities, or they may not have been taken in what is referred to as the optimal way. What this is will be returned to later in the report.

The uncertainty coming from the data can lead to uncertainty in the predictions or produce uncertain predictions. There is however a bright side to this, the model will be more capable in handling observations when the x-ray image is unclear, blurry, contain irregularities and/or are taken in a non-optimal way. But despite this, if the data in the dataset is not "good enough", it will lead to challenges.

My assessment and troubles during my first try

When creating a model that can interpret an x-ray, label the observations and use it as an aid to make a diagnosis, there are multiple ways and approaches. In the classifier I created last spring semester, I used a tutorial from a webpage to the fastAI-library. When following the approach on the webpage, I came to realize that it seems to be more complicated than I intended and planned my first classifier to be. What I learned from the realization was that I wanted to experiment with and make other classifiers. My approach was to follow the tutorial as a guideline for building the code and make it more explanatory.

Another idea I had planned to do this last spring was to make a personal version of the dataset and use it to make predictions. But last time it seemed to be more work than reward, however I decided to give this approach a go this semester. However, if I am still unsuccessful, the back-up plan is to use another version of the dataset from the Che-Xpert dataset.

Another thing I learned from the mistakes I made last year; I need to do more work on the dataset and not just the classifier. However as previously mentioned, my main goal with giving the assignment another try is to correct my mistakes caused by labeling the data. My mistake last year was that I was labeling the images with the name of the study from the patient, for example patient 1 – study 4, instead of identifying the observations in the image. This was a very obvious mistake, can be seen in the picture or in the code I did last year. However, I later discovered when I was looking over my code from last year for inspiration and insight regarding my strategy for this year's try, I realized I had done everything as it should have been done. My only mistake was how I navigated the directory, and I found out that the labels are correct regarding which study they are connected to. The problem was that the data are not classified with the observations at all!

Previously I mentioned that one of my plans last year was to modify the dataset given from Standford University and make my own labels. However, this may come at the cost of the reliability of my dataset and model, this is because I do not have any medical training, medical background or any experience with taking x-rays or interpreting them and use the interpretation to give a diagnosis. Another strategy I wanted to try was to reorganize the whole dataset and put them in their own directory and use the reorganized dataset as my dataset. The motivation behind this reorganization was that my model would have seen the images more than once, and therefore had I would have more images in the dataset. There is also another obstacle that prevents me from doing this reorganization, in the terms and conditions of usage for the CheXpert dataset. The terms and conditions of use state that the data cannot be copied without permission from Standford University. This is an approach that is very time-consuming and does not guarantee great success. Because of this, this approach was left as an idea this last semester, due to the lack of time and motivation to give it a try. I was, however, more motivated this semester to give it another try.

A thing I understood more of the importance of last year, was the importance of cleaning the data. And I did not spend a lot of time cleaning the data last year, therefore I intended on doing more cleaning before making the predictions. I have the ambition to remove all the lateral images, x-rays taken from the side of the human body, and only use the frontal images, meaning x-rays taken from the front. This cleaning I did not do on my last try last spring, so I intend to do it on this try.

Based on all the argumentations given above, I will have a similar approach to what I did on my first try. My plan was to make a classifier that is trained on classifying if an image contains a pair of lungs that has pneumonia or not, before using the classifier to classify the whole set.

One thing I also realized was that there were a lot of empty cells, so I had to find a way to deal with them. And due to my lack of medical knowledge, I figured the simplest and "most secure" way would be to simply replace the empty cells with 0 because I assumed that they were supposed to be 0. And to deal with the cells that contained a -1, I interpreted it to be that it was a spelling error in the original patient report, so I assumed it was supposed to be 1. Again, full disclosure, I have no medical background nor training, so all these changes are based on a guess with the hopes they are correct assumptions and will give good results.

The ideal x-ray

My strategy was to have all the x-ray images in the same format due to the of poor alignment[4] in the images. This is a risk that is very common with dealing with x-rays, including that the photographing is not being perfect. Perfection is important and central when giving a possible diagnosis. Together with the risk of big disorientation[5], this is a risk that comes when classification of observations combined with pictures with bad quality. From the article I read last year, that recommended rescaling the images before inserting them to one template matching algorithm to find the lung in the x-ray. Last year I did not know how to make such an algorithm, so my hopes were that I would manage to do the same this year. The same article recommended changes in order to handle or reduce source-dependent variation.

My plan also included using ensembling techniques, but this is something I did 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 the technique, it would give me, according to all the research I did prior to starting this project, give a higher prediction score.

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

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

Cross-entropy loss

Cross-entropy loss[6] is a cost function that can be used to optimize classification models; therefore, it is important to maintain it and optimize it if possible. Due to the function being a loss function, the optimal situation is to have the score as close as possible to 0. These measurements are used when one adjusts the weights during training. The goal is to minimize the loss, and by doing so one will get a better model. I tried to make a misclassification error[7] at to the first classifier I created. I did this because I wanted some kind of metric that could provide the percentage of the observations that were incorrectly classified and predicted by my model. And again, one wants it to as low as possible, which I achieved variable success.

Conclusion

Last year I had big plans and high hopes for my model's end results, and I could have achived a lot more than what I had, but if only I had discovered my mistake that had bothered me for over a year, earlier. There was a plethora of ways I could have then made a better model, done a better job with cleaning the data and trained the data for a longer time. However, since I thought last year, I had made a good classifier, and my troubles came from that I didn´t find the right way to transfer my classifier in order to predict on other datasets that was not the dataset that it was trained and validated on. This made me at the time, and some time after, quite frustrated, and I did want to do a different approach. I should start all over again and build a new classifier before then use it directly on the big dataset CheXpert. This was not a big success either, due to the fact that I did not pay as much attention to how the model performed during the preparation of the dataset for training nor when I started the exploration of the dataset. If I paid more attention to this, I now understand that I would have discovered the mistake earlier. So, I intended to be more observant when I work with datasets in a field, I am not familiar with, like previously mentioned I am no doctor, nor do I have any medical training or background. The bright aspect of this mistake is that I learned a lot from my last try at this project with the focus on regards to the process of making a model and using the model to make predictions (even though it was in that case, disease or no disease). But I did achieve something, in order to make sure that I got the right predictions, I got the skill (just at a very low and basic level) and learned how to read x-rays in preparation for the work with the data.

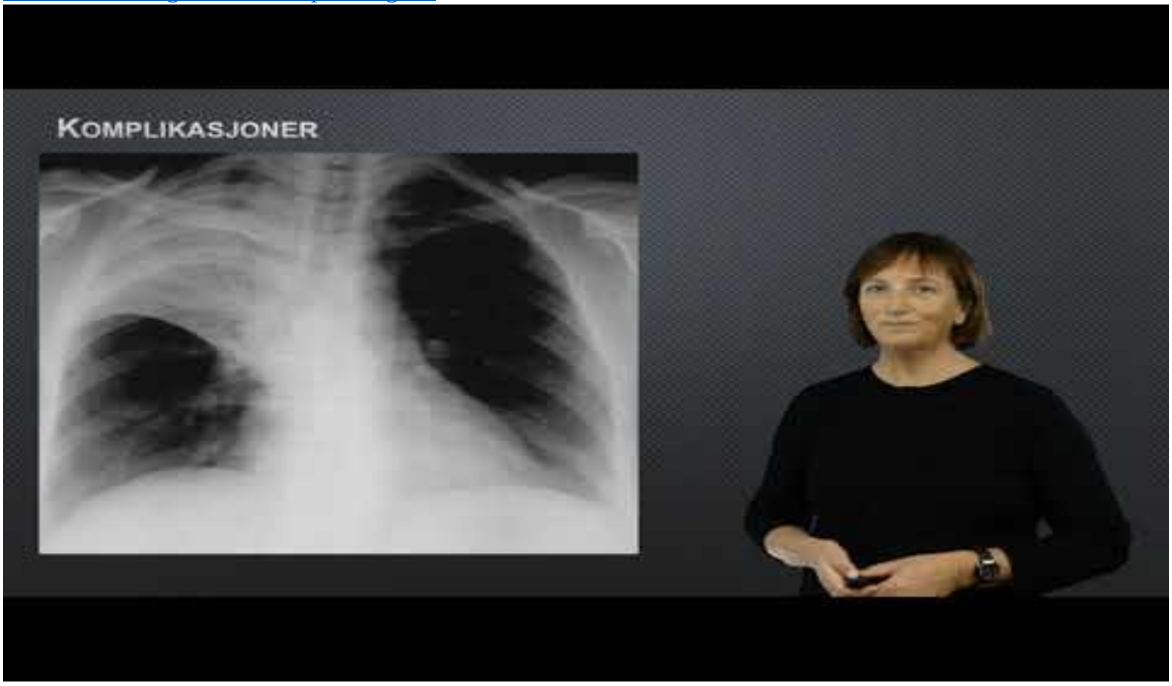
Despite my second try I got, I did not get the greatest success this time either. I got into trouble when it came to creating the dataloader. Here I tried many different strategies, even asking the infamous Chat-GPT for help but this did not give me much other than error messages and a decrease in motivation. I also asked other AI-code-generators, but to no luck there either. So I can safely say that my second try was not very successful either. But I have however come a long way from where I

[6] https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e
[7] https://www.statology.org/misclassification-rate/

started by correcting my mistakes from my previous attempts, and again I am reminded the importance of cleaning and exploring the dataset that one is working with.

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!