I suggest we use a model similar to Baltruschat, the ResNet38 with larger image entrance resolution and patient metadata. On behalf of utilizing larger images, they added a convolutional layer to the original model, and after the feature extraction part, they concatenated the patient metadata to the image feature vector, then all this information is inputted to the ResNet Dense layers. This architecture got the bests results training on the ChestX-ray14 dataset , achieving the state-of-the-art in five of fourteen categories, but now we can train the network on Cohan covid19 dataset, because it’s a aggregation between others datasets and it has metadata like age, sex, day of infection that the x-ray had been taken.

We can perform the agreement analysis between the covid19 dataset aggregations, and between other datasets that we can eventually use. This will tell us how bad or good these different data are biasing our network results, in terms of miss labeling. There is disagreement between Radiologists, Clinicians and automatic labelers.

Other thing that concerns me, besides the labeling disagreement, is the bias generated by the different dataset features, like contrast, shades of gray, angle, size, etc. I’ve seen a paper on this topic, the author trained AlexNet to recognize from which source the x-ray image is from, hiding the lungs, and they got accuracy score very close to one.

If we succeed on this, afterward, we can investigate other models like VGG and Xception, aiming the same modifications as in the ResNet38 by Baltruschat, and we can apply Gradient-weighted Class Activation Mapping (Grad-CAM) to see which characteristics of the image the network took more into account when making the classification.

We can also do Fine-Tuning on ChestX-ray data then do Transfer-Learning on covid19 dataset.

-------------------------------------------------------------------------------------------------------

Performance Assessing

* Accuracy (cross-validation) ROC-AUC
* Find the best threshold to get the minimum False Negative
* Use at least 3 networks and get the average output as the final output
* Agreement between different datasets
* Modify the Network to recognize from which dataset the image is
* Modify the Network to classify the sex and infection day, and regress the age
* Train the network on one dataset and test on the others, for all datasets, just like Cohen did. Thus, we can assess the dataset quality.

Image preprocessing: how can we preprocess the images to get the minimum bias using different datasets?