In this lab I will use the data set which contains pictures of chest x ray from people with pneumonia and people without. In order to algorithm’s performance I will use the confusion matric of my algorithm output and the real values. Since we are handling with medical diagnosis, the case of false negative (where we diagnosis that a person is health where he is actually is not) is worse than false positive because it is better to be on the safe side rather then missing a person with pneumonia. Therefore, I will use the following weight metric for evaluation

When I loaded the data, it was already split into testing set and training one. Here I will go trough the processing of all image from both training and testing set and will make sure that I have acceptable ratio between training and testing. Since our set in large (5,863 images) I will not use cross validation but the split that the data was originally came with. If needed, I will do small modifications.

Now that we have pretty big data set (8340 images) that is quit balanced, I am confidence saying that after shuffling the data and randomly split it into 80% training set and 20% testing set, the performance of the model on the testing set will be very similar to the actual performance of the model on average on any data set.

Like I declared at the beginning, the lower the evaluation value, the better the model is. We can clearly see that the performance of model 2 – the ResNet model, is much better than the performance of the regular convolutional network in model one. Also, from the representations of the confusion metrices, we can see that most of the incorrect predictions of the false negative that occurred in model 1, which are the more expensive mistakes, reduced to 2%. If we also compare the accuracy of the two, we see that model 2 got about 2% higher accuracy than model 1.

We can see that the evaluation value of the MLP model (185) is lower than the one of model 1 but higher than the second model, means that the performance of model 2 is better than the MLP model according to the conditions we set at the beginning of this lab.

In order to take the comparison a little farther, I will use the the McNemar’s statistic test. My null hypothesis will be that model 2 and the MLP model are the same. In order to reject it with a confidence of 90%, I will need a chi square value higher than 2.706.