Report

I had troubles with loading the models I built, consequently, I could not analyze the results of the network.

Here are some observations and improvements that I could have done if I had the time.

1. **Future works from the loss and accuracy curves**

a. Features selection

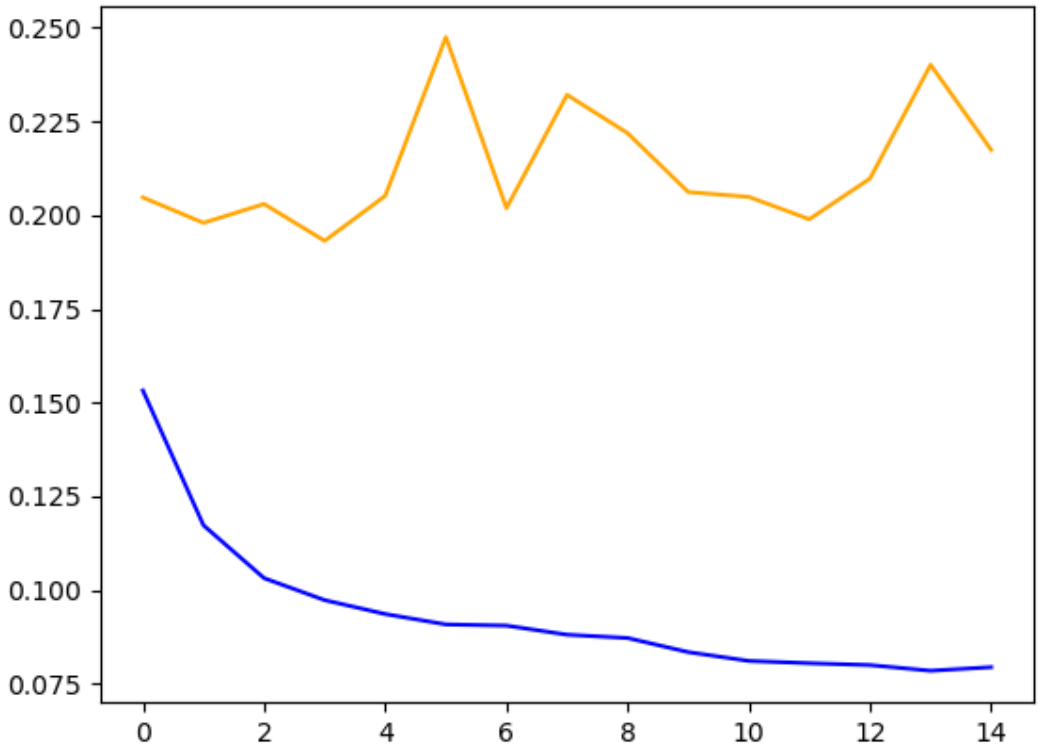

Figure 1: Loss functions across the different epochs. In blue is the training loss and in orange is the validation loss, obtained with a subset of the training set.

We can see that the training loss does not decrease at all. The model could have been too complex for generalization. A first step in its simplification could have been to decrease the number of features of the input vector.

For example, we can see that ribcage, abdos, BodyPos are very correlated, we could have selected only one of these features, or aggregated them. We could also have run a feature selector from sklearn to reduce the number of features.

b. Data augmentation

Also, the validation set can be too small (one fifth of the training set) to be representative. Cross-validation or data augmentation could have fixed this issue.

NB: in this work, the validation set is a subset of the training set, some of its samples can come from recordings that are also in the training set, explaining the high validation accuracy value (as shown in Figure 2). I should have created the validation set separately, with the DataLoader class, as done with the testing set.

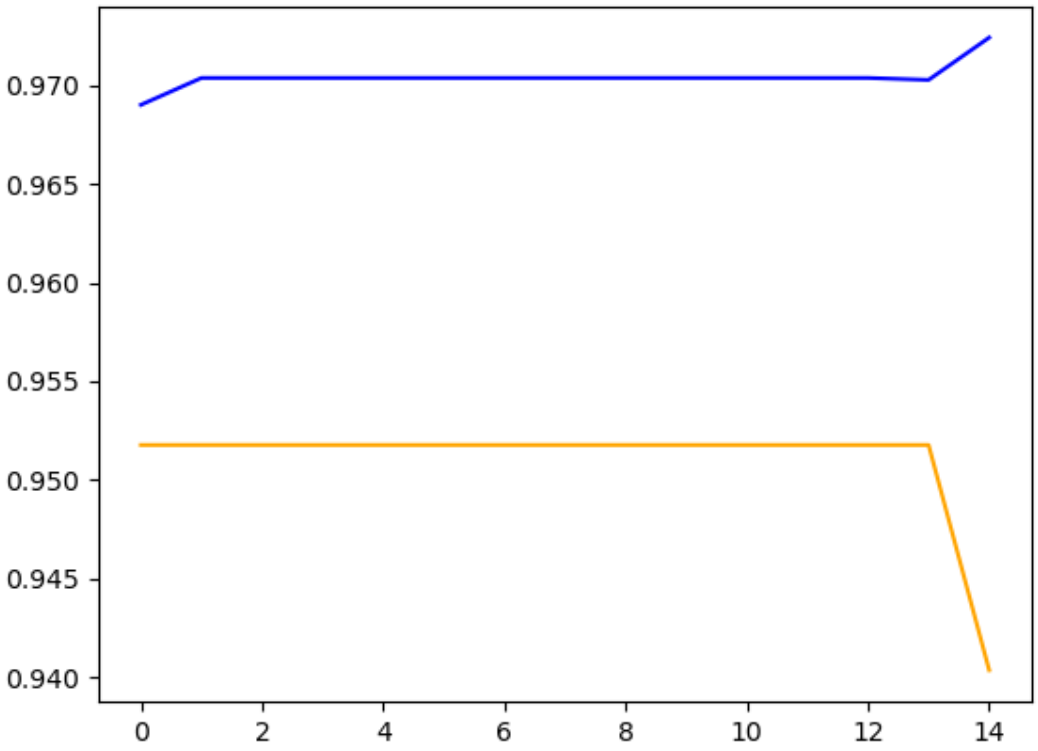

Figure 2: Binary accuracy curve across the different epochs. In blue is the training binary accuracy and in orange is the validation binary accuracy, obtained with a subset of the training set.

2. Future works from data visualization

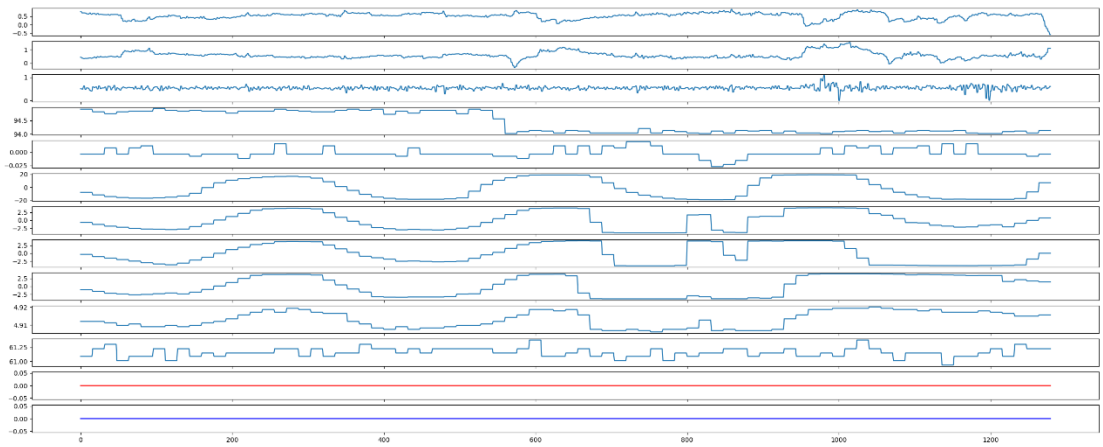

Figure 3: A plot of a testing sample (here, row 22 of the testing set), with time at the x axis. Like many of the samples, no respiratory event happened during the whole session (epoch duration of 10s in this figure).

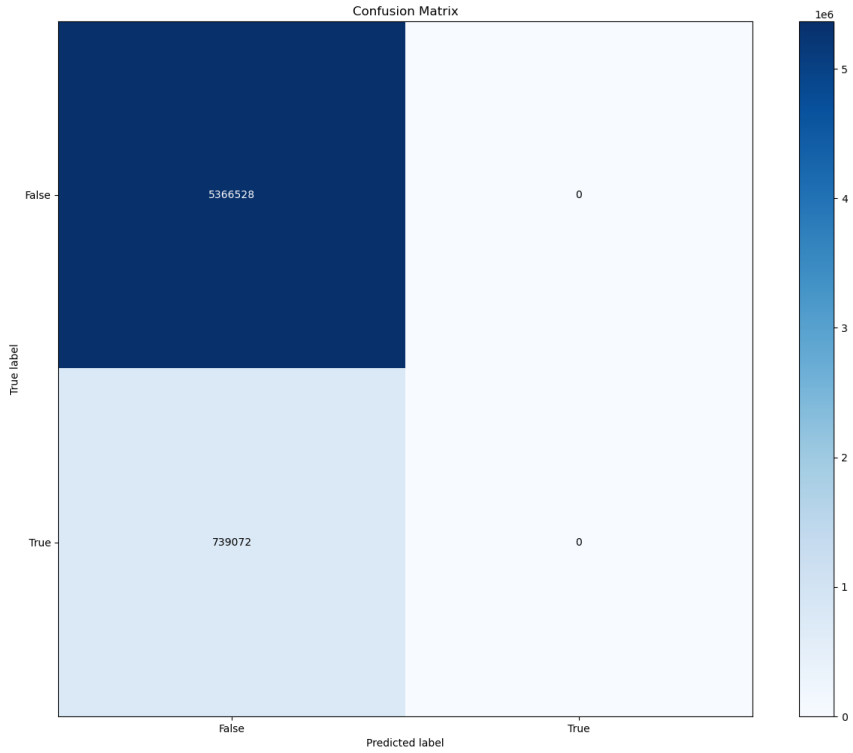

Figure 4: Confusion matrix obtained with a dummy model. It still shows the high imbalanced in the testing set that we could have limited with appropriate data augmentation or loss function.

When ploting rows of the features, predictions and groundtruth with the plot_data function of ResultsVisualizer (Figure 3), we can see that we have a lot of samples for which no respiratory event is detected. This could also have impeded the training.

The analysis of the confusion matrix could have enabled a better troubleshooting. We can already notice that the testing set (and very probably the training set) is highly imbalanced.

Data augmentation with a focus on the respiratory events could have helped, as well as a loss that focuses on the wrongly predicted samples (e.g. BinaryFocalEntropy).

It would also have been possible to increase the epoch duration to 5min (we couldn't due to RAM issues), to increase the chances that one sample had at least one label for a respiratory event.