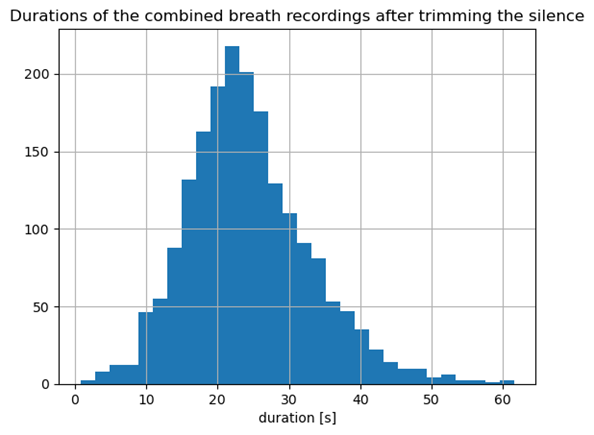
# Detailed Performance Analysis of Breathing Feature Sets

Varying time and frequency resolutions are being tested. Mostly the parameters hop size and fft-window-length were changed. A change in hop size results (when keeping the number of input parameters the same) in a varied length of the input. So if we have 224 time steps of the mel spectrogram and a hop size of 6ms compared to 92ms we get 1.3seconds and 20.6 seconds respectively.

The shorter frame can capture more short-term changes in the signal. The long signal can capture features like breath frequency.

The number of time steps will also be varied and finally it will be attempted to increase performance of shorter time frames with a MIL approach. For the longer time frames, a MIL approach is not expected to yield any improvements, but it can still be tested.

The average length of the recordings is shown below (about 20 seconds).

Notably there are a few hyperparameters not being used/that will be included in later runs after choosing the best time/frequency resolution. Others may just need to be explained (so I remember later). Other settings are not tracked (yet) by these hyperparameters:

## Bag size and n\_MIL\_neurons

Are only being used when using the MIL algorithms. Otherwise these values can be anything and are just not being used.

## Focal Loss:

Similarly the Focal Loss sets the alpha of the loss function. If the loss function used is the BCE with logits (like it is here for the first tests) this value is ignored, whatever it is.

## Lr\_in:

If this is set to a value and not None it determines the learning rate at the input. The regular lr parameter sets the lr at the output. The layers in between get a linearly interpolated lr. If it is None, the entire network is set to “lr”.

## Online Augmentations RandomTransferFunction and RandomGain:

Currently there are two new online augmentations with fixed values. The RandomTransferFunction adds for each frequency bin +/- k dB at maximum. K is now set to 3.

RandomGain adds x dB to every value in the mel spectrogram where x is a random value between -9 and +3dB right now.

## Residual Norm gamma:

This needs to be experimented later. The input into the resNorm layer is added to the output of the layer via a skip connection with a factor gamma. This gamma can either be a hyperparameter or can be learnable(?).

## Time Domain Augmentations:

Augmentations that are done beforehand and saved because they take too much time to do online (pitch shift, timestrech,..). They will also be added afterwards when the better feature sets have been determined.

# 92ms hop size; 184ms FFT window

This is the longest time frame that is being tested. 100 time frames would contain almost 10 seconds of audio. This also means that this is the smallest dataset with about 500MB and it trains the fastest.

A hyperparameter search was conducted with lr, time\_steps, dropout and min\_quality being varied parameters. The tracker name is:

2023-05-14\_resnet18\_logmel\_combined\_breaths\_NEW\_92msHop\_184msFFT\_fmax11000\_224logmel\_baseline\_hyperparams

A screenshot of a table

Description automatically generated with low confidence

## Time Steps

Time steps tested were: 100, 200, 400 (10s, 20s, 40s). 400 time steps/40s seems to cause a bit of a degradation of the performance. Between 100 and 200 there seems to be no great difference.

## Learning rate

['0.008', '0.0005', '0.0008', '0.0001’]

Were the tested rates. The large learning rate performed the worst. The two lower ones outperformed the larger one for the most part in most metrics.

## Dropout

Dropout is the greatest factor. In auc\_roc there is a slight increased performance when NOT using dropout which is curious. Usually auc\_roc was a metric improved by dropout. Dropout slightly improves TPR but this is very much undone by the fact that NOT using dropout improves TNR, F1 and accuracy GREATLY:

A screen shot of a graph

Description automatically generated with low confidence

## Discussion:

Interestingly, the accuracy looks great for many runs but the auc\_roc is a lot less exciting. In the cough analysis the difference was often marginal or the auc roc was even higher than the accuracy. Also the TPR is not very promising but the TNR is amazing. So there is a clear focus on the negative samples. This can be addressed by adding the time domain augmented samples or by increasing the weight of the positive class in the loss function. Usually dropout also increases the TPR at the cost of TNR but interestingly this is not so clear here in this case.

The good performance indicates, that there are some features that are appear on this long term time scale like the breathing frequency which is not captured in for example 3 second time frames.

Also to be noted: The accuracy for the best performing run reaches 98% very early despite the fact that mix-up and online augmentations are applied. A picture containing text, screenshot, plot, line

Description automatically generated

