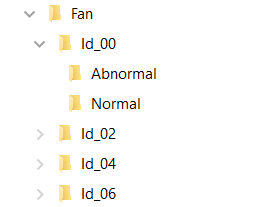
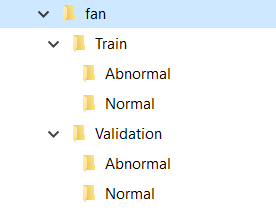
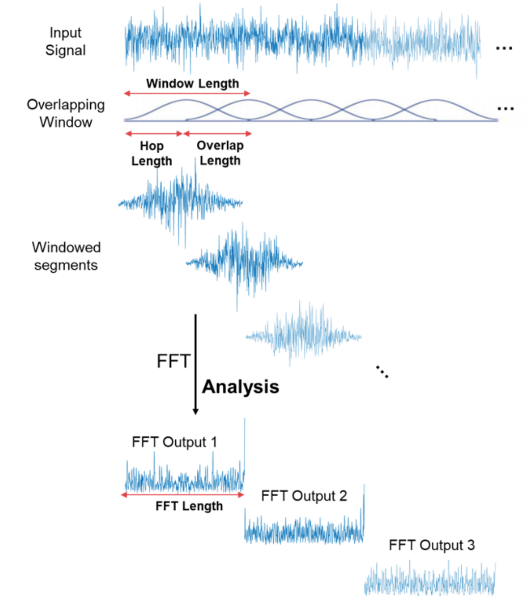
The work consists in training a sound classification model to detect the state of a fan in an industrial environment. The dataset used can be found through this link (https://zenodo.org/record/3384388#.YTi1Tk4YBPZ) . it is arranged as follows:



The layout was then modified as follows:

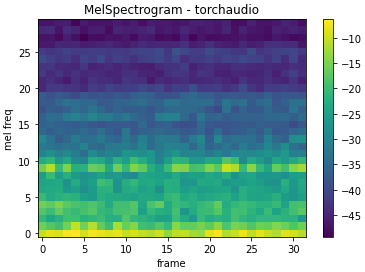


The model behaves like an image classifier: the input nodes are fed by a matrix named Log-Mel spectrogram. In short, the log-mel spectrogram is like a set of FFTs put end to end while applying a scale modification to get closer to the behavior of the human ear.



The parameters of the spectrogram and the structure of the model have been adapted to the limitation imposed by the microcontroller (especially the size of the RAM limited to 96 KB).

The preprocessing consists in preparing a csv file containing the metadata. Moreover, it is a question of normalizing all the files present (same duration, same sampling frequency as well as a mono format) so that the spectrograms obtained are of the same dimension (in our case 30\*32) as shown in the following figure:



Once the model was trained, it was exported in ".pth" file. As this file is not directly compatible with the current version of the STM32 Cube AI suite, I had to convert it to a more generic format ONNX (Open Neural Network Exchange).

The program implemented on microcontroller follows the following flowchart:

