

1 Data description

The dataset contains a training subset (13538 recordings from 35 sensors), and validation subset (4308 recordings from 9 sensors), and a test subset (669 recordings from 48 sensors). Each recording has been annotated using a set of 23 “tags” like “engine presence, machinery presence, non-machinery-impact presence, dog-barking-whining presence”. The sensors in the test set will not disjoint from the training and validation subsets, but the test recordings are displaced in time, occurring after any of the recordings in the training and validation subset. We plan to use the test data to find out the aggregate of mismatch by using a Multi Label classification Machine Learning Model.