The data is collected in EDF file. The project outline is as follows:

1) Read the data and select aportion that has no artifacts:

* Read the file
* Identify the time indices of periods corresponding to hyperventilation, photic stimulation and flat signal
* If one minute of clean data is present, record it as an EDF file; if not, return a specific value. [If more minutes are needed, one can simply modify the value 300, which corresponds to 3 minutes to larger numbers. We only considered the files which have at least 3 continuous minute of clean data to mitigate the fact that segments following the artifacts may be affected of those]

2) Apply minimal pre-processing

* Transform data to numpy array, resample at the required frequency, detrend and normalize

After designing the corresponding functions (which can be found in **reading\_script.py**), one has only to loop over all the files and store the obtained arrays (for example in a .npy file). One major challenge is that multiple files must be created due to the large quantity of data. Another is that many files have problems and usually functions need to provide indications of those to further be assessed manually.

All the functions for the next stages can be found in **neural\_network.py**:

3) Prepare data for training

* To improve model capabilities, we divide the array for each patient into smaller chunks and give them the same label in the training process (i.e. the age of the particular patient)

4) Create functions for training and testing

* Split the data into training-test subsets (validation could also be added), create a function that returns the predictions of the network and one that evaluate the error of predictions

5) Create Neural Network models:

* We defined 10 possible neural network models labeled f1, f2, …, f10 with different layer structure, activation functions and pooling strategies
* Train the model, display error for training and test