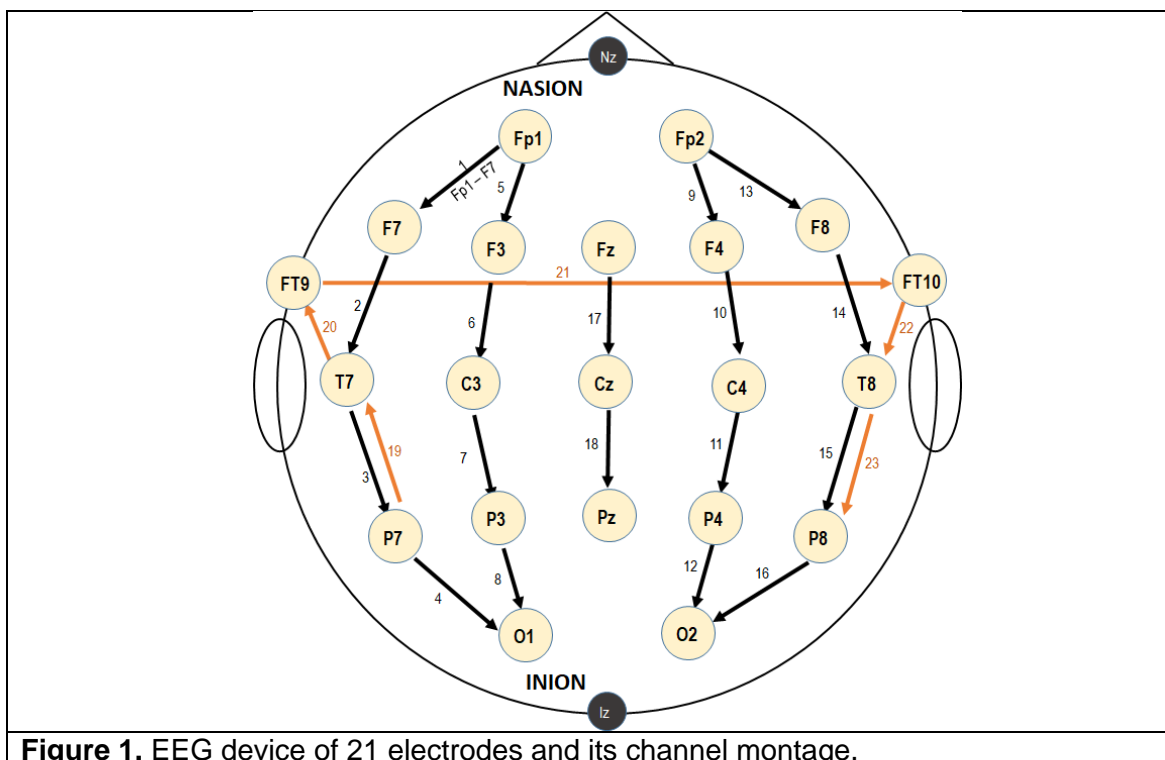


# CHB-MIT EEG Dataset Description

When performing different tasks along the day, people experiment different levels of cognitive state depending on the level of attention required, the difficulty of such task and how many sub-tasks are needed to take care off. Physiological measures provide a more reliable data of cognitive state by measuring physiological dynamic changes which cannot be controlled consciously. Electroencephalogram (EEG) is the preferred one among researchers. The goal is to characterize the cognitive state by means of neural networks from EEG signals. In particular, to identify epileptic seizures.

The data was acquired from the CHB-MIT Scalp EEG Database (<https://physionet.org/content/chbmit/1.0.0/>). This dataset contains EEG recordings from 23 pediatric and one adult patients with intractable seizures. Recordings were made for several hours per patient and records are delivered in files of one hour-long. Records that contain seizures are labelled as seizure records, the rest of records are labeled as non-seizure records. As ground truth, for each seizure records it is provided the start and the end time (in seconds) of the epileptic seizure based on expert judgements. A seizure record can contain one or more epileptic seizures.

Majority of the records in the dataset contains 23 channel of signals (which are the differences between electrodes) recorded by an EEG device of 21 electrodes. Signals were collected at the sampling frequency of 256 Hz. Figure 1 shows the localization of EEG electrodes over the head scalp and the arrows illustrates the montage (longitudinal-bipolar) used to generate the channels. Table 1 shows the selected channels with the calculated differences between electrodes.



**Figure 1.** EEG device of 21 electrodes and its channel montage.

Table1: Channel names

Channel Id	Channel Name
Channel 1	FP1-F7
Channel 2	F7-T7
Channel 3	T7-P7
Channel 4	P7-O1
Channel 5	FP1-F3
Channel 6	F3-C3
Channel 7	C3-P3
Channel 8	P3-O1
Channel 9	FP2-F4
Channel 10	F4-C4
Channel 11	C4-P4
Channel 12	P4-O2
Channel 13	FP2-F8
Channel 14	F8-T8
Channel 15	T8-P8
Channel 16	P8-O2
Channel 17	FZ-CZ
Channel 18	CZ-PZ
Channel 19	T7-FT9
Channel 20	FT9-FT10
Channel 21	FT10-T8

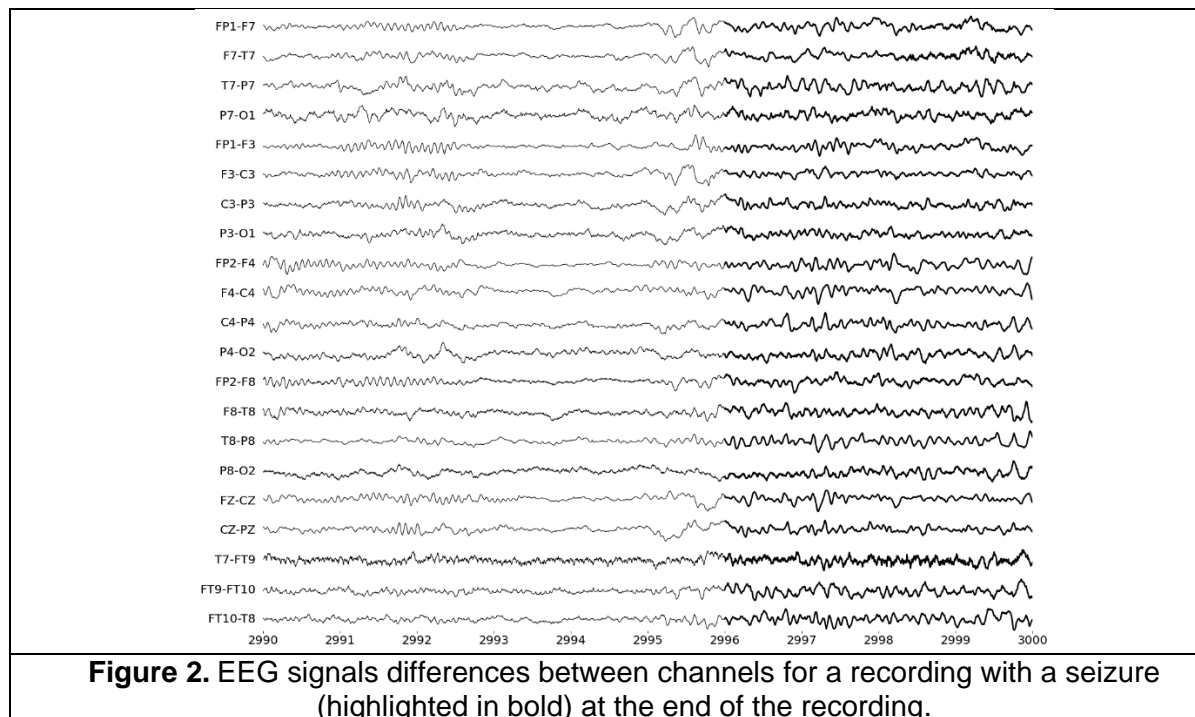
For each patient, the dataset provides EEG recordings of many consecutive hours in EDF files (European Data Format), usually of one hour of duration. Files which have at least on seizure are annotate as seizure files, otherwise they are named as non-seizure files. Besides, the dataset furnishes annotation of the time in which an epilepsy starts and ends (this time is reference to the starting time of the seizure file). Table 2 presents the information of patients and their seizures during recordings.

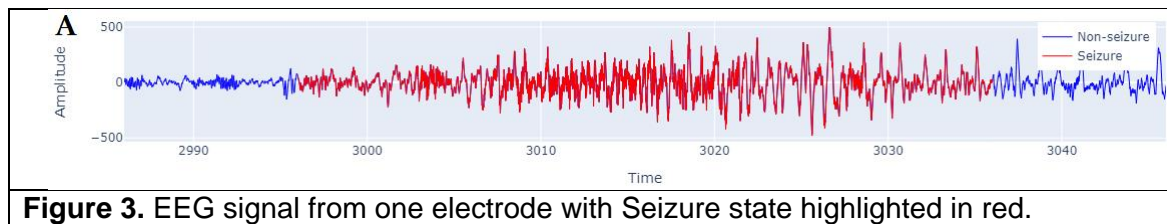
Table 2: The dataset and patient information.

Patient	Sex – Age	Number of seizures	Total duration of seizures (sec)	Total recording time (h)	Total time of seizure files (h)
chb01	F-11	7	442	40.55	6.65
chb02	M-11	3	172	35.27	2.27
chb03	F-14	7	402	38.00	7.00
chb04	M-22	4	378	156.07	10.66
chb05	F-7	5	558	39.00	5.00
chb06	F-1.5	10	153	66.74	25.89
chb07	F-14.5	3	325	67.05	9.04
chb08	M-3.5	5	919	20.01	5.00
chb09	F-10	4	276	67.87	9.58

chb10	M-3	7	447	50.02	14.02
chb11	F-12	3	806	34.79	2.79
chb12	F-2	27	989	20.69	9.68
chb13	F-3	10	440	11.00	7.00
chb14	F-9	8	169	26.00	7.00
chb15	M-16	20	1992	39.01	14.01
chb16	F-7	8	69	17.00	5.00
chb17	F-12	3	293	20.01	3.01
chb18	F-18	6	317	34.63	5.63
chb19	F-19	3	236	28.93	2.93
chb20	F-6	8	294	27.60	5.57
chb21	F-13	4	199	32.83	3.83
chb22	F-9	3	204	31.00	3.00
chb23	F-6	7	424	26.56	8.96
chb24		16	511	21.30	12.00
<b>Total</b>		<b>181</b>	<b>11015</b>	<b>951.93</b>	<b>185.51</b>

Figure 2 illustrates a segment of EEG data from Patient 1 with a seizure starting at the 2996 second highlighted in bold face. Although the shape of seizure can vary across patients due the type of epilepsy, individuals tend to keep similar patterns. For instance, it is noticeable the spatial and temporal consistency in the seizure of Patient 1 shown in Fig.2. Figure 3 shows EEG signal from a single electrode with a seizure interval highlighted in red.





**Figure 3.** EEG signal from one electrode with Seizure state highlighted in red.

### Input Data:

The original MIT data files have been pre-processed to provide EEG signals for each case in parquet format and annotations of each of the recordings

For each patient, the original edf recordings have been packaged in a parquet, EEG recordings with seizure and without seizure. It has been generated using annotation, therefore, they contain both 23 and 18 channel records. Records with other sensor and channel distribution are located in the EDF files. Data has been down-sampled to 128 Hz (128 samples per second) for storage saving and taking into account that detection of seizure does not require higher sampling rate. Files are identified by the patient ID.

Each parquet contains the following information:

1. PatID: patient identifier
2. filename: name of the original .edf recording. This field identifies the case in the df\_annotation\_full.xlsx in order to produce the GT.
3. type: indicates whether the .edf recording contains a seizure or not: normal (if there is no attack) or seizure (if the file has any attack)
4. One column for each EEG channel ('FP1-F7', 'F7-T7', 'T7-P7', 'P7-O1', 'FP1-F3', 'F3-C3', 'C3-P3', 'P3-O1', 'FP2-F4', 'F4-C4', 'C4-P4', 'P4-O2', 'FP2-F8', 'F8-T8', 'T8-P8', 'P8-O2', 'FZ-CZ', 'CZ-PZ', 'T7-FT9', 'FT9-FT10', 'FT10-T8')

The file df\_annotation\_full.xlsx contains the following fields for each .edf file:

- 1.type: indicates whether the .edf recording contains a seizure or not: normal (if there is no attack) or seizure (if the file has any attack)
2. PatID: patient identifier
3. filename: .edf filename that identifies the recording
- 4.seizure\_id: number that identifies the attack in the edf file, 0 if there is no attack, and the attack order in case there is more than one (e.g. if there are several consecutive attacks in the file it would be a 1 for the first a 2 for the second and so on).
5. seizure\_start: time ( seconds) the attack begins in the .edf recording
6. seizure\_end: : time ( seconds) the attack ends in the .edf recording
7. rec\_duration: duration of the .edf recording (seconds)
8. seizure\_duration: duration (seconds) of each attack