**EEG Dataset Description Report**

**Abstract**: This dataset includes the EEG of 6 epileptic patients recorded at the Epilepsy monitoring unit of the American university of Beirut Medical Center between January 2014 and July 2015. The data represents measurements from 21 scalp electrodes, following the 10-20 electrode system, sampled at 500 Hz. All channels have been bandpass filtered between 1/1.6 Hz and 70Hz while filtering out the 50Hz (electrical utility frequency). Some channels have been omitted from specific recordings due to artifact constraints.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset  Characteristics: | Multi-channel Epileptic EEG Time Series | Number of  Patients: | 6 | Total Nb of seizures: | 35 |
| Available Seizure Types | - Complex Partial Seizure  - Electrographic Seizures  - Video-detected Seizures with no visual change over EEG | Seizure Duration | Varying |  |  |

**Source:**

Original Owner:  
  
Dr. Wassim Nasereddine

Associate Professor of Clinical Specialty

Neurology Department  
American University of Beirut Medical Center   
Beirut, Lebanon

**Data Set Information:**

The data was obtained from patients diagnosed with focal epilepsy and undergoing presurgical evaluation with long-term video-EEG monitoring for possible epilepsy surgery. During this evaluation, the anti-seizure medications are stopped in an attempt to record the patient’s habitual seizures. There are 6 patients with a total of 20 recordings. Each recording includes an interictal portion during which no seizures are recorded and a single or multiple ictal sections that consists of a seizure. There are different seizure types, ictal onset zones, and durations that are labeled for each recording.

**Raw Data Information:**

The raw data is stored in European Data Format (EDF) which divides the EEG records into header information and data information.

The header information is as follows:

**ver**: identification code

**patientID:** local subject identification

**recordID:** local recording identification **'**

**startdate** (starting date of the recording)

**starttime** (starting time of the recording)

**bytes**: 5120

**records**: 10799

**duration**: 1

**ns**: number of channels

**label**: channel labels which shows the electrode type and name, example listed below

{'EEGFp2Ref'} {'EEGFp1Ref'} {'EEGF8Ref'} {'EEGF4Ref'} {'EEGFzRef'} {'EEGF3Ref'} {'EEGF7Ref'} {'EEGA2Ref'} {'EEGT4Ref'} {'EEGC4Ref'} {'EEGCzRef'} {'EEGC3Ref'} {'EEGT3Ref'} {'EEGA1Ref'} {'EEGT6Ref'} {'EEGP4Ref'} {'EEGPzRef'} {'EEGP3Ref'} {'EEGT5Ref'} {'EEGO2Ref'} {'EEGO1Ref'} {'ECGEKG'} {'Manual'} {'EDFAnnotations'}

**units**: voltage units for all available channels

**physicalMin**: Physical minimum in units of physical dimension

**physicalMax**: Physical maximum in units of physical dimension

**digitalMin**: Digital minimum for all channels

**digitalMax:** Digital maximum for all channels

**prefilter:** prefilter range for all channels (low pass and high pass cutoff frequencies)

**samples**: sampling rate for all channels

**Labeled Data Information:**

Labeling the EEG records as normal or lesional data is crucial for machine learning applications. Therefore, the available EEG datasets have been processed, separated and labeled as normal or lesional data.

To have uniform information across all records, the data of channels Cz and Pz have been omitted in the labeling process.

The classified data are categorized as matrices of size 19x500. 19 representing the number of channels and 500 respesenting one second duration as the sampling rate is 500Hz. The labeling will be as follows

* **1**: for Complex Partial Seizures

(3034 matrices of size 19x500 corresponding to 3034 seconds of complex partial seizures)

* **2**: for Electrographic Seizures

(705 matrices of size 19x500 correponding to 750 seconds of electrographic seizures)

* **3**: for Video-detected Seizures with no visual change over EEG

(111 matrices of size 19x500 corresponding to 111 seconds of Video-detected Seizures with no visual change over EEG)

* **0:** for Normal data

(3895 matrices of size 19x500 corresponding to 3895 seconds of normal data, 3895 is the total duration of all available seizures to create the balance between normal and lesional data).

The total labeled data will therefore have a total size of 7790x19x500 (7790 is the summation of the above seconds, normal and seizures). The data will be split into train and test data, 7011 (90%) and 779 (10%) respectively. The data is avaible in both numpy (.npy) and Matlab (.mat) format, saved as shown below:

* **x\_train**, size 7011 x 19 x 500 with its respective labels **y\_train**
* **x\_test**, size 779 x 19 x 500 with its respective labels **y\_test**