

Supplementary materials in conjunction with the paper **A Deep Learning Framework for Epileptic Seizure Detection based on Neonatal EEG Signals**

Artur Gramacki & Jarosław Gramacki

10-04-2022

Introduction

This report may be seen as a supplement to the *Replication of the results* Section in the paper (steps 1-5). It was generated using **Knitr** (<https://en.wikipedia.org/wiki/Knitr>), an engine for dynamic report generation with R. You can regenerate this HTML file at any time by running `tutorial_on_how_to_run_R_codes.Rmd` script in RStudio (<https://www.rstudio.com>).

R version used

All the R codes have been tested in R version **4.1.2**. Most likely, however, they can also be launched in other **4.x.x** versions of R. The R version during generating this report is:

```
R.version.string
```

```
## [1] "R version 4.1.2 (2021-11-01)"
```

Preliminary remarks, required packages

We strongly recommend using **RStudio** program to work with R (<https://www.rstudio.com>). Before running our codes, three required packages **must be installed manually**, i.e. `edf`, `png` and `rhdf5`. In order to install the first two use menu **Tools→Install Packages** in RStudio. Type the name of the package and press Install button. The third package is available at <https://bioconductor.org> and to install this package start R (version not older than **4.x.x**) and enter:

```

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

pkgs <- c("edf", "png")
to_install = !pkgs %in% installed.packages()
if(any(to_install)) {
  install.packages(pkgs[to_install])
}

pkg <- c("rhdf5")
to_install = !pkg %in% installed.packages()
if(to_install) {
  BiocManager::install("rhdf5")
}

```

After installing the two packages **load** them using the following commands:

```

library(edf)
library(png)
library(rhdf5)

```

Next steps to be performed

1. Set directory structure

Be sure that you have the following directory structure on your disk and that **R directory is the current working one**.

```

# |---annotations
# |---edf
# |---Python
# |---R          <-- this MUST be set as your working directory
#   |---test_files
# |---working
#   |---acc_loss
#   |---best_models
#   |---hists
#   |---inputs
#   |---logs
#   |---results
#   |---ROC
#   |---waveforms

```

To check your working directory, execute the following function:

```
getwd()
```

```
## [1] "D:/Users/GoogleDrive/Colab Notebooks/EEG_neonatal/__codes__and__data__/R"
```

2. Download the dataset of neonatal EEG recordings

These are available at <https://zenodo.org/record/4940267> (<https://zenodo.org/record/4940267>). There are 79 EDF files and 3 CSV annotations files. The EDF files are approximately 4GB in size.

3. Upload the dataset

Upload 79 EDF files and 3 CSV files to the `edf` and `annotations` directories respectively. Check their content. Must be as below:

```
list.files('../edf')
```

```
## [1] "__read_me__.txt" "eeg1.edf"      "eeg10.edf"     "eeg11.edf"
## [5] "eeg12.edf"      "eeg13.edf"     "eeg14.edf"     "eeg15.edf"
## [9] "eeg16.edf"      "eeg17.edf"     "eeg18.edf"     "eeg19.edf"
## [13] "eeg2.edf"       "eeg20.edf"     "eeg21.edf"     "eeg22.edf"
## [17] "eeg23.edf"      "eeg24.edf"     "eeg25.edf"     "eeg26.edf"
## [21] "eeg27.edf"      "eeg28.edf"     "eeg29.edf"     "eeg3.edf"
## [25] "eeg30.edf"      "eeg31.edf"     "eeg32.edf"     "eeg33.edf"
## [29] "eeg34.edf"      "eeg35.edf"     "eeg36.edf"     "eeg37.edf"
## [33] "eeg38.edf"      "eeg39.edf"     "eeg4.edf"      "eeg40.edf"
## [37] "eeg41.edf"      "eeg42.edf"     "eeg43.edf"     "eeg44.edf"
## [41] "eeg45.edf"      "eeg46.edf"     "eeg47.edf"     "eeg48.edf"
## [45] "eeg49.edf"      "eeg5.edf"      "eeg50.edf"     "eeg51.edf"
## [49] "eeg52.edf"      "eeg53.edf"     "eeg54.edf"     "eeg55.edf"
## [53] "eeg56.edf"      "eeg57.edf"     "eeg58.edf"     "eeg59.edf"
## [57] "eeg6.edf"       "eeg60.edf"     "eeg61.edf"     "eeg62.edf"
## [61] "eeg63.edf"      "eeg64.edf"     "eeg65.edf"     "eeg66.edf"
## [65] "eeg67.edf"      "eeg68.edf"     "eeg69.edf"     "eeg7.edf"
## [69] "eeg70.edf"      "eeg71.edf"     "eeg72.edf"     "eeg73.edf"
## [73] "eeg74.edf"      "eeg75.edf"     "eeg76.edf"     "eeg77.edf"
## [77] "eeg78.edf"      "eeg79.edf"     "eeg8.edf"      "eeg9.edf"
```

```
list.files('../annotations')
```

```
## [1] "annotations_2017_A_fixed.csv" "annotations_2017_B.csv"
## [3] "annotations_2017_C.csv"
```

4. Set required variables

```

# Symbols of human experts.
we <- c( "A", "B", "C")

# Annotations file names, as downloaded from https://zenodo.org/record/4940267
ann.f <- c("annotations_2017_A_fixed.csv",
           "annotations_2017_B.csv",
           "annotations_2017_C.csv")

# Infant IDs which have seizures annotated by all 3 experts
s.IDs <- c(1,4,5,7,9,11,13,14,15,16,17,19,20,21,22,25,31,34,36,38,39,40,41,44,47,50,51,52,62,63)
# Infant IDs which are seizure free.
ns.IDs <- c(3,10,18,27,28,29,30,32,35,37,42,45,48,49,53,55,57,58,59,60,70,72)

# If you are working with a directory structure as shown above, do not change this variable.
dir = "../"

```

5. Load required functions

Finally you can load a script with all required R functions. This script contains all the functions we have written that are required to perform the analyzes discussed in the article. These functions are:

`select_seizure_chunks()` , `generate_montage()` , `generate_samples()` , `generate_eeg_waveforms()` , `read_channel_names()` , `generate_selected_waveforms()` , `read_signal_lengths()` . Below, we execute the `read_signal_lengths()` function whose task is to read the lengths of all raw EDF signals.

```
source("EEG_neonatal_FUNS.R")
```

```

time_elapsed <- system.time({

read_signal_lengths(1:79, dir)

})

```

patient 1: 116m 33s
patient 2: 62m 41s
patient 3: 73m 32s
patient 4: 57m 5s
patient 5: 64m 1s
patient 6: 78m 23s
patient 7: 60m 54s
patient 8: 76m 53s
patient 9: 59m 10s
patient 10: 90m 27s
patient 11: 124m 48s
patient 12: 74m 28s
patient 13: 256m 56s
patient 14: 62m 6s
patient 15: 114m 58s
patient 16: 99m 1s
patient 17: 91m 33s
patient 18: 60m 51s
patient 19: 150m 6s
patient 20: 66m 27s
patient 21: 95m 7s
patient 22: 64m 4s
patient 23: 73m 53s
patient 24: 61m 46s
patient 25: 111m 49s
patient 26: 69m 42s
patient 27: 58m 16s
patient 28: 95m 57s
patient 29: 133m 57s
patient 30: 65m 37s
patient 31: 58m 49s
patient 32: 76m 53s
patient 33: 62m 27s
patient 34: 108m 11s
patient 35: 62m 47s
patient 36: 84m 42s
patient 37: 76m 18s
patient 38: 101m 35s
patient 39: 77m 9s
patient 40: 97m 17s
patient 41: 161m 24s
patient 42: 82m 54s
patient 43: 121m 23s
patient 44: 56m 0s
patient 45: 63m 42s
patient 46: 85m 20s
patient 47: 60m 6s
patient 48: 64m 0s
patient 49: 144m 3s
patient 50: 164m 10s
patient 51: 78m 21s

```
## patient 52: 65m 20s
## patient 53: 57m 50s
## patient 54: 72m 24s
## patient 55: 86m 1s
## patient 56: 63m 51s
## patient 57: 52m 13s
## patient 58: 78m 5s
## patient 59: 65m 23s
## patient 60: 62m 59s
## patient 61: 97m 9s
## patient 62: 97m 32s
## patient 63: 65m 0s
## patient 64: 105m 37s
## patient 65: 80m 0s
## patient 66: 189m 10s
## patient 67: 81m 40s
## patient 68: 60m 15s
## patient 69: 66m 13s
## patient 70: 81m 46s
## patient 71: 71m 16s
## patient 72: 83m 18s
## patient 73: 62m 10s
## patient 74: 72m 44s
## patient 75: 65m 55s
## patient 76: 63m 41s
## patient 77: 69m 53s
## patient 78: 82m 51s
## patient 79: 54m 57s
```

```
## Execution time: 2min 29sec
```

Generate required HDF5 files

1. Introductory notes

To do this run `EEG_neonatal.R` script (this can take a few hour). Make sure that the current working directory is `R`. Set also the `dir` variable to the one indicating the appropriate directory structure in your local computer. The parameters of the `generate_samples()` function can be changed depending on your actual needs. Those that are saved in the `EEG_neonatal.R` script will generate exactly the same HDF5 files that were included in the Electronic Supplements.

After generating, the HDF5 files will be saved in the `working\inputs` directory (along with a couple of diagnostic files that are not directly used in the subsequent calculations). The directory should contain 90 HDF5 data files ready to be fed to the neural network and additionally 184 auxiliary/diagnostic files. The data files have the same logical structure as in Figure 8 and use a uniform naming convention. For example, the file `expert_C_5sec_2chunk_64Hz.hdf5` means that data was generated according to the annotations made by expert C, the window size was set to 5 seconds and the number of contiguous chunks was set to 2 (see Figures 4 and 5).

The similar naming convention is used for all other files in the `working` subdirectories.

Note: we do not put HDF5 files in the regular Electronic Supplements, as their total size is about 16.6GB. However, for your convenience, we included them in separate zip archives, see *Data and code availability* Section in the paper.

2. Test run

For testing purposes first let's try to read and display one annotation file (the first 20 line).

```
options(width = 999)

ann <-
  read.csv(
    paste(dir, "annotations/", "annotations_2017_A_fixed.csv", sep = ""),
    sep = ",",
    header = TRUE,
    stringsAsFactors = F,
    check.names = FALSE,
    encoding = 'UTF-8')

head(ann, 20)
```

```
##      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
## 1  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 2  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 3  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 4  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 5  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 6  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 7  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 8  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 9  0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 10 0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 11 0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 12 0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 13 0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 14 0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 15 0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 16 0 0 0 0 1 0 1 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 17 0 0 0 0 1 0 0 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 18 0 0 0 0 1 0 0 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 19 0 0 0 0 1 0 0 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 20 0 0 0 0 1 0 0 0 0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

The total number of rows in `annotations_2017_A_fixed.csv` file is 15416.

For another test try to generate only one HDF5 file (`expert_A_1sec_1chunk_64Hz.hdf5`). Additionally, also 3 TXT files will be generated (`SEIZURE_A_1sec_1chunk_64Hz.txt` , `NON.SEIZURE_A_1sec_1chunk_64Hz.txt` and `expert_A_1sec_1chunk_64Hz.txt`). We generate TXT files for illustrative purposes only. As for the content, they are fully compatible with the HDF5 binary file. If TXT files are not needed, simply set `write.txt.files = FALSE` .

```
options(width = 999)

time_elapsed <- system.time({

out <- generate_samples(
  which.expert = "A",
  annotations_file = "annotations_2017_A_fixed.csv",
  seizure.IDs = s.IDs,
  non.seizure.IDs = ns.IDs,
  window = 1,
  chunks = 1,
  down.sampling.factor = 4,
  preprocessing = FALSE,
  dir = dir,
  random = FALSE,
  write.txt.files = FALSE,
  write.hdf5.files = TRUE
)

})
```



```

## -----
## annotations file name:      annotations_2017_A_fixed.csv
## seizure patients:          1 4 5 7 9 11 13 14 15 16 17 19 20 21 22 25 31 34 36 38 39 40 41 4
## non-seizure patients:      3 10 18 27 28 29 30 32 35 37 42 45 48 49 53 55 57 58 59 60 70 72
## hdf5 file being generated: expert_A_1sec_1chunk_64Hz.hdf5
## -----
##
## ../edf/eeg1.edf (total number of seizures annotated: 25)
## ../edf/eeg4.edf (total number of seizures annotated: 2)
## ../edf/eeg5.edf (total number of seizures annotated: 5)
## ../edf/eeg7.edf (total number of seizures annotated: 6)
## ../edf/eeg9.edf (total number of seizures annotated: 3)
## ../edf/eeg11.edf (total number of seizures annotated: 3)
## ../edf/eeg13.edf (total number of seizures annotated: 5)
## ../edf/eeg14.edf (total number of seizures annotated: 45)
## ../edf/eeg15.edf (total number of seizures annotated: 19)
## ../edf/eeg16.edf (total number of seizures annotated: 30)
## ../edf/eeg17.edf (total number of seizures annotated: 4)
## ../edf/eeg19.edf (total number of seizures annotated: 9)
## ../edf/eeg20.edf (total number of seizures annotated: 17)
## ../edf/eeg21.edf (total number of seizures annotated: 1)
## ../edf/eeg22.edf (total number of seizures annotated: 8)
## ../edf/eeg25.edf (total number of seizures annotated: 12)
## ../edf/eeg31.edf (total number of seizures annotated: 2)
## ../edf/eeg34.edf (total number of seizures annotated: 1)
## ../edf/eeg36.edf (total number of seizures annotated: 2)
## ../edf/eeg38.edf (total number of seizures annotated: 19)
## ../edf/eeg39.edf (total number of seizures annotated: 6)
## ../edf/eeg40.edf (total number of seizures annotated: 12)
## ../edf/eeg41.edf (total number of seizures annotated: 45)
## ../edf/eeg44.edf (total number of seizures annotated: 7)
## ../edf/eeg47.edf (total number of seizures annotated: 3)
## ../edf/eeg50.edf (total number of seizures annotated: 10)
## ../edf/eeg51.edf (total number of seizures annotated: 4)
## ../edf/eeg52.edf (total number of seizures annotated: 1)
## ../edf/eeg62.edf (total number of seizures annotated: 1)
## ../edf/eeg63.edf (total number of seizures annotated: 5)
## ../edf/eeg66.edf (total number of seizures annotated: 2)
## ../edf/eeg67.edf (total number of seizures annotated: 16)
## ../edf/eeg69.edf (total number of seizures annotated: 14)
## ../edf/eeg71.edf (total number of seizures annotated: 4)
## ../edf/eeg73.edf (total number of seizures annotated: 6)
## ../edf/eeg75.edf (total number of seizures annotated: 1)
## ../edf/eeg76.edf (total number of seizures annotated: 2)
## ../edf/eeg77.edf (total number of seizures annotated: 1)
## ../edf/eeg78.edf (total number of seizures annotated: 22)
## ../edf/eeg79.edf (total number of seizures annotated: 5)
## ../edf/eeg3.edf
## ../edf/eeg10.edf
## ../edf/eeg18.edf
## ../edf/eeg27.edf

```

```
## ../edf/eeg28.edf
## ../edf/eeg29.edf
## ../edf/eeg30.edf
## ../edf/eeg32.edf
## ../edf/eeg35.edf
## ../edf/eeg37.edf
## ../edf/eeg42.edf
## ../edf/eeg45.edf
## ../edf/eeg48.edf
## ../edf/eeg49.edf
## ../edf/eeg53.edf
## ../edf/eeg55.edf
## ../edf/eeg57.edf
## ../edf/eeg58.edf
## ../edf/eeg59.edf
## ../edf/eeg60.edf
## ../edf/eeg70.edf
## ../edf/eeg72.edf
```

```
## Execution time: 2min 29sec
```

3. Generating a complete set of HDF5 files used in the paper

Please note that this is a very time-consuming operation (several hours). Uncomment the codes below when you are ready to do these calculations. When the files generation is complete, there should be 90 HDF5 files and 184 auxiliary text files in the `working \ inputs` directory.

```

options(width = 999)

time_elapsed <- system.time({

# for (i in 1:3) {
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#
#   # We set chunks = 10000 and this way we are sure that the maximum possible set of chunks wi
#   # Our dataset simply doesn't have seizures as long as 10,000 seconds.
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
#   out <- generate_samples(which.expert = we[i], annotations_file = ann.f[i], seizure.IDs = s
# }

})

```

```
## Execution time: 0min 0sec
```

4. Notes

Infant numbers stored in the variables `s.IDs` and `ns.IDs` can of course be freely changed, depending on the current needs. For the purposes of this article, we have adopted the rule that we collect positive training samples (i.e. EEG chunks with seizures) only from infants where all 3 experts have annotated at least one seizure. On the other hand, we collect negative training samples (e.g. EEG chunks which are seizure free) from infants where no expert has annotated any seizure.

Looking at Table 7 we can see that 17 neonates had a seizure annotated by only 1 or 2 expert. In fact, EEG recordings of these infants have been excluded from our research as less certain/unambiguous data. However, nothing prevents from trying to use them in your further analysis.

Generating EEG waveforms based on raw EDF recordings and annotations files

According to Tables 7 and 8 (see the paper) there are 1379 seizure recordings in total annotated by 3 independent experts. To facilitate their more detailed analysis, a dedicated function has been prepared that allows one to visualize all the EEG waveforms. In the below example, the parameter `window = 10` was set, which causes that waveforms with a length of 10 seconds each are generated. One example waveform is shown below. It presents an EEG waveform seizure annotated by expert A for infant number 1 (see the first column in `annotations_2017_A_fixed.csv` file and the first row in Tables 7 and 8.). The presented seizure is 18 seconds long (starts at second 104 and ends at second 121). All 1379 EEG waveforms are saved in the `working\waveforms` directory as `png` files (you can also generate files in `pdf` format, to do this one needs to set the `format = "pdf"` parameter). The file name is

`seizures_expert_A_pat1_no1_win10_len18_from104_to121_64Hz.png`, its individual fragments mean:

- `seizures_expert_A` - a seizure annotated by expert A
- `pat1` - patient (infant) number 1
- `no1` - the sequential number of the epileptic seizure. For infant number 1 and expert A there are 25 annotated seizures in total (see Tables 7 and 8)
- `win10` - the length of EEG waveform presented in the plot
- `len18` - the length of epileptic seizure as annotated by expert (see Table 8)
- `from104` - the annotated seizure begins at 104th second
- `to121` - the annotated seizure ends at 121st second
- `64Hz` - the resulting frequency after down sampling

To regenerate all the 1379 plots uncomment the below codes.

```

# infant IDs where 1 or 2 experts annotated seizures, expert A
s12A.IDs <- c(2,8,23,33,54,68)

# infant IDs where 1 or 2 experts annotated seizures, expert B
s12B.IDs <- c(8,24,64,68,74)

# infant IDs where 1 or 2 experts annotated seizures, expert C
s12C.IDs <- c(6,12,23,26,33,43,46,54,56,61,64,65,74)

time_elapsed <- system.time({

# p.ID <- c(s.IDs, s12A.IDs)
# generate_eeg_waveforms(
#   mode = "seizure",
#   which.expert = we[1],
#   annotations.file = ann.f[1],
#   patient.IDs = p.ID,
#   window = 10,
#   down.sampling.factor = 4,
#   non.seizures.chunks = 10,
#   preprocessing = FALSE,
#   random = FALSE,
#   format = "png",
#   save.to.file = TRUE,
#   dir = dir
# )
#
# p.ID <- c(s.IDs, s12B.IDs)
# generate_eeg_waveforms(
#   mode = "seizure",
#   which.expert = we[2],
#   annotations.file = ann.f[2],
#   patient.IDs = p.ID,
#   window = 10,
#   down.sampling.factor = 4,
#   non.seizures.chunks = 10,
#   preprocessing = FALSE,
#   random = FALSE,
#   format = "png",
#   save.to.file = TRUE,
#   dir = dir
# )
#
# p.ID <- c(s.IDs, s12C.IDs)
# generate_eeg_waveforms(
#   mode = "seizure",
#   which.expert = we[3],
#   annotations.file = ann.f[3],
#   patient.IDs = p.ID,
#   window = 10,
#   down.sampling.factor = 4,

```

```
# non.seizures.chunks = 10,  
# preprocessing = FALSE,  
# random = FALSE,  
# format = "png",  
# save.to.file = TRUE,  
# dir = dir  
# )  
  
})
```

```
## Execution time: 0min 0sec
```

In addition to waveform files, additional files with detailed seizure data for each patient are generated. For example, for patient number 1 and expert A, we have a file `expert_A_pat1_64Hz.csv`.

```
out = read.csv(  
  '../working/waveforms/expert_A_pat1_64Hz.csv',  
  sep = "\t",  
  blank.lines.skip = FALSE  
)  
out[is.na(out)] <- ""  
print(out, row.names = FALSE, na.print = "" , quote = FALSE)
```

| ## | patient | seizure_duration | from_sec | to_sec | from_sample | to_sample |
|----|-----------------|------------------|----------|--------|-------------|-----------|
| ## | 1 | 18 | 104 | 121 | 6593 | 7744 |
| ## | 1 | 135 | 317 | 451 | 20225 | 28864 |
| ## | 1 | 59 | 742 | 800 | 47425 | 51200 |
| ## | 1 | 29 | 954 | 982 | 60993 | 62848 |
| ## | 1 | 31 | 1179 | 1209 | 75393 | 77376 |
| ## | 1 | 49 | 1594 | 1642 | 101953 | 105088 |
| ## | 1 | 57 | 1651 | 1707 | 105601 | 109248 |
| ## | 1 | 23 | 2849 | 2871 | 182273 | 183744 |
| ## | 1 | 87 | 3008 | 3094 | 192449 | 198016 |
| ## | 1 | 23 | 3257 | 3279 | 208385 | 209856 |
| ## | 1 | 31 | 3852 | 3882 | 246465 | 248448 |
| ## | 1 | 93 | 3993 | 4085 | 255489 | 261440 |
| ## | 1 | 104 | 4405 | 4508 | 281857 | 288512 |
| ## | 1 | 34 | 4588 | 4621 | 293569 | 295744 |
| ## | 1 | 25 | 4779 | 4803 | 305793 | 307392 |
| ## | 1 | 24 | 4829 | 4852 | 308993 | 310528 |
| ## | 1 | 78 | 4962 | 5039 | 317505 | 322496 |
| ## | 1 | 122 | 5141 | 5262 | 328961 | 336768 |
| ## | 1 | 74 | 5543 | 5616 | 354689 | 359424 |
| ## | 1 | 333 | 6065 | 6397 | 388097 | 409408 |
| ## | 1 | 19 | 6489 | 6507 | 415233 | 416448 |
| ## | 1 | 23 | 6550 | 6572 | 419137 | 420608 |
| ## | 1 | 99 | 6605 | 6703 | 422657 | 428992 |
| ## | 1 | 15 | 6784 | 6798 | 434113 | 435072 |
| ## | 1 | 17 | 6847 | 6863 | 438145 | 439232 |
| ## | | | | | | |
| ## | Total seizures: | | | | | |
| ## | 25 | | | | | |

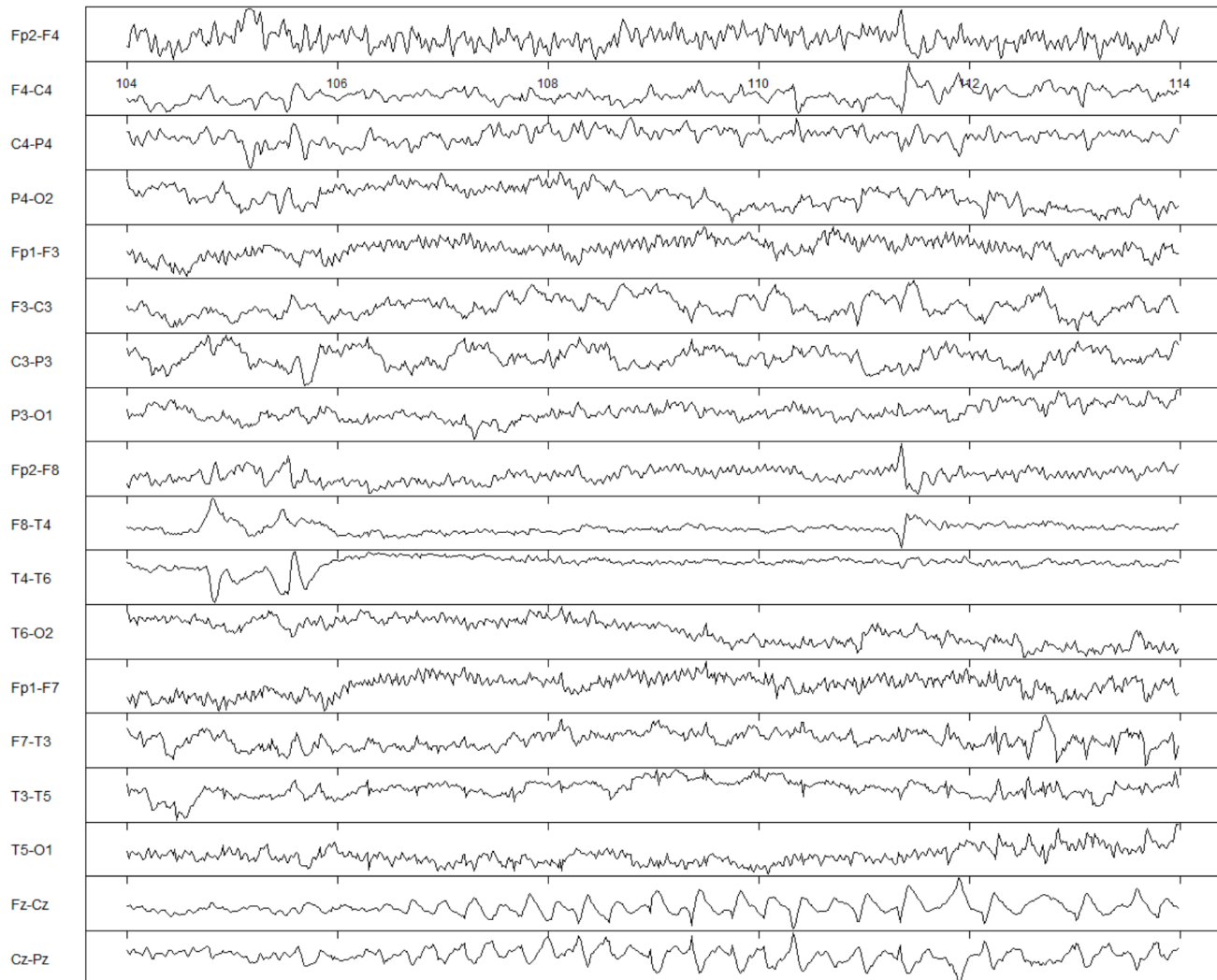
From this file we know that expert A annotated 25 seizures. The corresponding EEG waveforms are written in the following files:

```
## seizures_expert_A_pat1_no1_win10_len18_from104_to121_64Hz.png
## seizures_expert_A_pat1_no10_win10_len23_from3257_to3279_64Hz.png
## seizures_expert_A_pat1_no11_win10_len31_from3852_to3882_64Hz.png
## seizures_expert_A_pat1_no12_win10_len93_from3993_to4085_64Hz.png
## seizures_expert_A_pat1_no13_win10_len104_from4405_to4508_64Hz.png
## seizures_expert_A_pat1_no14_win10_len34_from4588_to4621_64Hz.png
## seizures_expert_A_pat1_no15_win10_len25_from4779_to4803_64Hz.png
## seizures_expert_A_pat1_no16_win10_len24_from4829_to4852_64Hz.png
## seizures_expert_A_pat1_no17_win10_len78_from4962_to5039_64Hz.png
## seizures_expert_A_pat1_no18_win10_len122_from5141_to5262_64Hz.png
## seizures_expert_A_pat1_no19_win10_len74_from5543_to5616_64Hz.png
## seizures_expert_A_pat1_no2_win10_len135_from317_to451_64Hz.png
## seizures_expert_A_pat1_no20_win10_len333_from6065_to6397_64Hz.png
## seizures_expert_A_pat1_no21_win10_len19_from6489_to6507_64Hz.png
## seizures_expert_A_pat1_no22_win10_len23_from6550_to6572_64Hz.png
## seizures_expert_A_pat1_no23_win10_len99_from6605_to6703_64Hz.png
## seizures_expert_A_pat1_no24_win10_len15_from6784_to6798_64Hz.png
## seizures_expert_A_pat1_no25_win10_len17_from6847_to6863_64Hz.png
## seizures_expert_A_pat1_no3_win10_len59_from742_to800_64Hz.png
## seizures_expert_A_pat1_no4_win10_len29_from954_to982_64Hz.png
## seizures_expert_A_pat1_no5_win10_len31_from1179_to1209_64Hz.png
## seizures_expert_A_pat1_no6_win10_len49_from1594_to1642_64Hz.png
## seizures_expert_A_pat1_no7_win10_len57_from1651_to1707_64Hz.png
## seizures_expert_A_pat1_no8_win10_len23_from2849_to2871_64Hz.png
## seizures_expert_A_pat1_no9_win10_len87_from3008_to3094_64Hz.png
```

One selected png file is presented below.

```
img <- readPNG("../working/waveforms/seizures_expert_A_pat1_no1_win10_len18_from104_to121_64Hz.
grid::grid.raster(img)
```


seizures_expert_A_pat1_no1_win10_len18_from104_to121_64Hz.png



Using the function `read_selected_waveforms()` , it is possible to precisely select only the fragments of interest from raw EDF files.

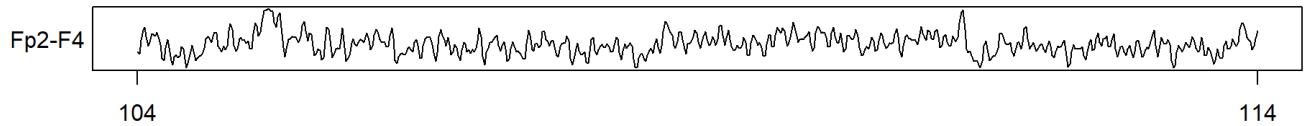
```
range <- c(104, 114)
patient <- 1
channel <- 1

out <- generate_selected_waveforms(
  patient = patient,
  down.sampling.factor = 4,
  channel = channel,
  sec.range = range,
  preprocessing = FALSE,
  dir
)
```

```
## ../edf/eeg1.edf
```

```
plot(out$points, type = "l", xlab = "", ylab = "", yaxt = "n", xaxt = "n", main = paste ("Patient",
ticks <- seq(1, length(out$points), length.out = 2)
axis(1, at = ticks, labels = seq(range[1], range[2], length.out = 2), main = "a")
mtext(text = out$channel.names, las = 1, adj = 0, cex = 1, side = 2, outer = FALSE, line = 3)
```

Patient # 1



Setting `channel = 1` you specify the channel number you want to display. Channel names can be displayed with the `get_channel_names()` function. The names saved in the raw EDF files are shown as well as the names after the so called *montage* procedure.

```
out <- read_channel_names(1)
```

```
## ../edf/eeg1.edf
```

```
# in EDF file
out$sig.names.in.edf.file
```

```
## [1] "EEG_Fp1_REF" "EEG_Fp2_REF" "EEG_F3_REF" "EEG_F4_REF" "EEG_C3_REF"
```

```
# after montage
out$sig.names.after.montage
```

```
## [1] "Fp2-F4" "F4-C4" "C4-P4" "P4-O2" "Fp1-F3" "F3-C3" "C3-P3" "P3-O1" "Fp2-F8" "F8-T4"
```

Building a new CNN model (or models) based on completely new datasets

In this section, we show some tests to help you determine if your EDF files and a seizure description file are correct for our R codes. In the `R\test_files` directory we placed 2 anonymous raw EDF files and a sample annotations file where 4 seizures have been marked.

Read the sample CSV annotations file and display its beginning

```
ann <-  
  read.csv(  
    paste("test_files/sample_annotations.csv", sep = ""),  
    sep = ",",  
    header = TRUE,  
    stringsAsFactors = F,  
    check.names = FALSE,  
    encoding = 'UTF-8'  
  )  
head(ann, 25)
```

```
##      1 2  
## 1  0 0  
## 2  0 0  
## 3  0 0  
## 4  0 0  
## 5  0 0  
## 6  0 0  
## 7  0 0  
## 8  0 0  
## 9  0 0  
## 10 1 0  
## 11 1 0  
## 12 1 0  
## 13 1 0  
## 14 1 0  
## 15 1 0  
## 16 1 0  
## 17 1 0  
## 18 1 0  
## 19 1 0  
## 20 0 0  
## 21 0 0  
## 22 0 0  
## 23 0 0  
## 24 0 0  
## 25 0 0
```

Try to read two sample EDF files. A short summary report is generated

```

IDs <- c(1, 2)

for (i in IDs) {
  filename <- paste("test_files/eeg", i, ".edf", sep = "")
  edf <- read.edf(filename = filename, read.annotations = FALSE, header.only = FALSE)

  n.sigs <- edf[["header.global"]][["n.signals"]]
  f.edf <- edf[["header.signal"]][[1]][["n.samples"]]
  len <- length(edf$signal[[1]]$data)
  # in seconds
  len.secs <- len/f.edf

  sig.names <- NaN
  for (s in 1:n.sigs) {
    sig.names[s] <- edf$header.signal[[s]]$label
  }
  sig.names

  cat("\nEDF file: ", filename, "\n", sep = "")
  cat (
    "  number of signals: ", n.sigs, "\n",
    "  base frequency: ", f.edf, "\n",
    "  number of samples: ", len, "\n",
    "  length in secs: ", len.secs, "\n",
    sep = "")
  cat ("  signal names: ", "\n", sep = "")
  print(sig.names)
}

```

```

##
## EDF file: test_files/eeg1.edf
##  number of signals: 17
##  base frequency: 500
##  number of samples: 458000
##  length in secs: 916
##  signal names:
## [1] "EEG_FP1_F3"      "EEG_FP2_F4"      "EEG_F3_C3"      "EEG_F4_C4"      "EEG_C3_P3"
##
## EDF file: test_files/eeg2.edf
##  number of signals: 17
##  base frequency: 500
##  number of samples: 480500
##  length in secs: 961
##  signal names:
## [1] "EEG_FP1_F3"      "EEG_FP2_F4"      "EEG_F3_C3"      "EEG_F4_C4"      "EEG_C3_P3"

```

Read annotations data and display a simple summary

Patient # 1 has 2 annotations. The first seizure is 10s long, starts at 10th second and ends at 19th second. The second seizure is 20s long, starts at 100th second and ends at 119th second.

Patient # 2 has also 2 annotations entered. The first seizure is 5s long, starts at 900th second and ends at 904th second. The second seizure is 10s long, starts at 952nd second and ends at 961st second.

```
for (k in IDs) {  
  filename <- paste(dir, "edf/eeg", k, ".edf", sep = "")  
  cat("EDF file: ", filename, "\n", sep = "")  
  out <- select_seizure_chunks(data = ann, f.edf, k)  
  print(out)  
}
```

```
## EDF file: ../edf/eeg1.edf  
## $seizures  
##   patient seizure_duration from_sec to_sec from_sample to_sample  
## 1      1           10      10      19      4501      9500  
## 2      1           20     100     119     49501     59500  
##  
## $total.seizures  
## [1] 2  
##  
## EDF file: ../edf/eeg2.edf  
## $seizures  
##   patient seizure_duration from_sec to_sec from_sample to_sample  
## 1      2           5      900     904     449501     452000  
## 2      2          10     952     961     475501     480500  
##  
## $total.seizures  
## [1] 2
```

A problem detected

A closer inspection of the EDF files showed that the last channel (EDF_Annotations) is a bit weird. It contains much less samples and basically only a few dozen unique values have been entered there. Moreover, the frequency is set to 27. Surely, this channel should be excluded from the analysis. The plot confirms the oddity of the data.

```
filename
```

```
## [1] "../edf/eeg2.edf"
```

```
edf[["header.signal"]][["EDF_Annotations"]][["n.samples"]]
```

```
## [1] 27
```

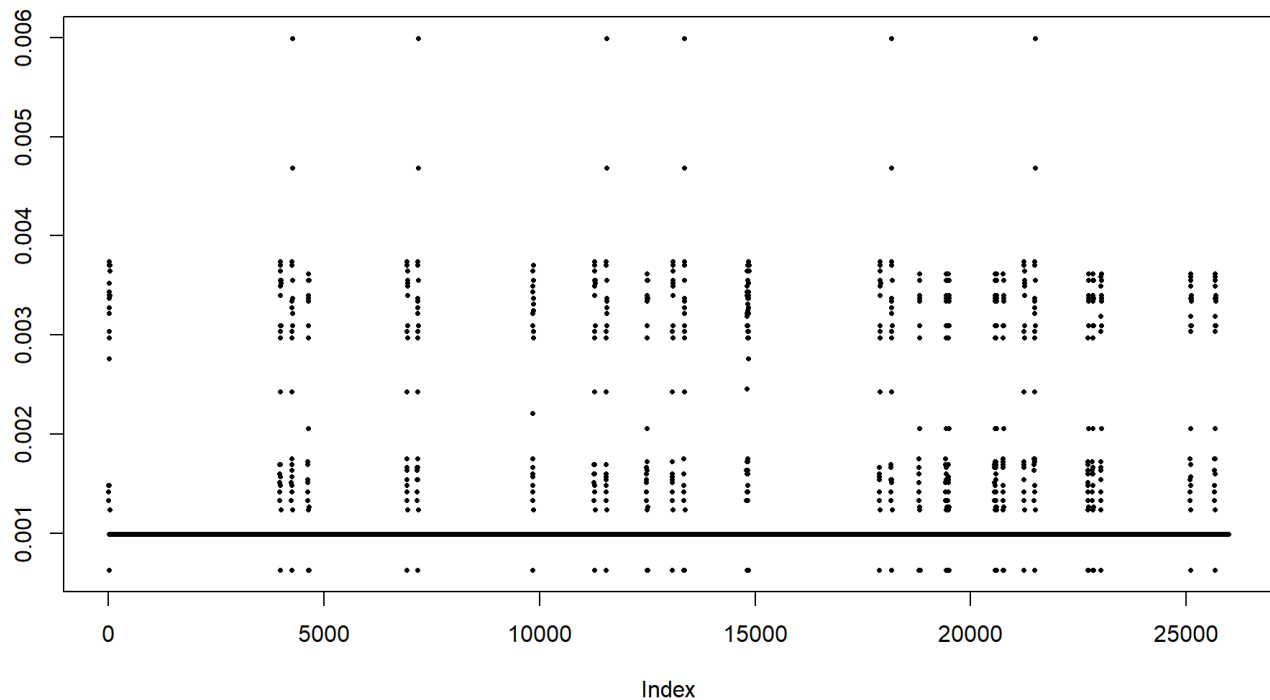
```
length(edf$signal[["EDF_Annotations"]]$data)
```

```
## [1] 25974
```

```
unique(edf[["signal"]][["EDF_Annotations"]][["data"]])
```

```
## [1] 0.0013275349 0.0014801251 0.0014190890 0.0006256199 0.0027618830 0.0029755093 0.0034332
```

```
plot(edf[["signal"]][["EDF_Annotations"]][["data"]], pch = 16, cex = 0.5, ylab = "")
```



Classification of new data using CNN already trained by us

An important note: you must be aware that our CNN network has been trained on a certain dataset (quite specific) and is ready to recognize a certain type of seizures (i.e. neonatals ones). Therefore, it should not be expected that when we provide completely different data to the pre-trained CNN network (e.g. based on elderly patients), the network will correctly classify the data. Also some technical details on EEG recordings must be considered carefully. In our case, signals from 18 EEG channels connected according to the 'double banana' montage were fed to the CNN network. When the new data is not analogous, the classification results can be very questionable. Nevertheless, when the new data is compatible (in the sense as stated above), there are no major contraindications to feed them to our pre-trained CNN network.

As a new dataset, i.e. one that was not used to build and train the CNN network, we will use EEG signals from neonates where there was no consensus between experts. 17 neonates had a seizure annotated by only 1 or 2 experts (infants No. 2, 6, 8, 12, 23, 24, 26, 33, 43, 46, 54, 56, 61, 64, 65, 68, 74). Due to the ambiguity in the expert opinion this subset was excluded from the process of building CNNs.

For classification task we choose the infants where expert C annotated at least 1 seizure (infants No. 6, 12,

23 , 26 , 33 , 43 , 46 , 54 , 56 , 61 , 64 , 65 , 74), see Table 7 in the paper.

To avoid a conflict with the filenames (see `working\inputs` directory), the expert was marked `cc` instead of `C`.

```
time_elapsed <- system.time({  
  
  out <-  
    generate_samples(  
      which.expert = "CC",  
      annotations_file = ann.f[3],  
      seizure.IDs = s12C.IDs,  
      non.seizure.IDs = ns.IDs,  
      window = 10,  
      chunks = 20,  
      down.sampling.factor = 4,  
      preprocessing = FALSE,  
      dir = dir,  
      random = FALSE,  
      write.txt.files = FALSE,  
      write.hdf5.files = TRUE  
    )  
  
  })
```

```

## -----
## annotations file name:      annotations_2017_C.csv
## seizure patients:         6 12 23 26 33 43 46 54 56 61 64 65 74
## non-seizure patients:     3 10 18 27 28 29 30 32 35 37 42 45 48 49 53 55 57 58 59 60 70 72
## hdf5 file being generated: expert_CC_10sec_20chunk_64Hz.hdf5
## -----
##
## ../edf/eeg6.edf (total number of seizures annotated: 4)
## ../edf/eeg12.edf (total number of seizures annotated: 1)
## ../edf/eeg23.edf (total number of seizures annotated: 9)
## ../edf/eeg26.edf (total number of seizures annotated: 7)
## ../edf/eeg33.edf (total number of seizures annotated: 6)
## ../edf/eeg43.edf (total number of seizures annotated: 4)
## ../edf/eeg46.edf (total number of seizures annotated: 4)
## ../edf/eeg54.edf (total number of seizures annotated: 17)
## ../edf/eeg56.edf (total number of seizures annotated: 1)
## ../edf/eeg61.edf (total number of seizures annotated: 3)
## ../edf/eeg64.edf (total number of seizures annotated: 12)
## ../edf/eeg65.edf (total number of seizures annotated: 4)
## ../edf/eeg74.edf (total number of seizures annotated: 6)
## ../edf/eeg3.edf
## ../edf/eeg10.edf
## ../edf/eeg18.edf
## ../edf/eeg27.edf
## ../edf/eeg28.edf
## ../edf/eeg29.edf
## ../edf/eeg30.edf
## ../edf/eeg32.edf
## ../edf/eeg35.edf
## ../edf/eeg37.edf
## ../edf/eeg42.edf
## ../edf/eeg45.edf
## ../edf/eeg48.edf
## ../edf/eeg49.edf
## ../edf/eeg53.edf
## ../edf/eeg55.edf
## ../edf/eeg57.edf
## ../edf/eeg58.edf
## ../edf/eeg59.edf
## ../edf/eeg60.edf
## ../edf/eeg70.edf
## ../edf/eeg72.edf

```

```

## Execution time: 1min 22sec

```

The following files have been created:


```
## expert_CC_10sec_20chunk_64Hz.hdf5  
## non_seizures_CC_10sec_20chunk_64Hz.txt  
## seizures_CC_10sec_20chunk_64Hz.txt  
## seizures_CC_64Hz.txt
```

The classification process is carried out in Python, see the Python Jupyter notebook included in the Electronic Supplements.