

Recognition of epileptic seizures from EEG data

Christian Espinosa Reboredo

Abstract—An electroencephalogram (EEG) is a test that detects electrical activity of the brain. Since 1924 this procedure has been done countless times to obtain brain activity. This paper tries to go a step further to understand electroencephalography better using deep learning algorithms. The data used in this paper is a public dataset CHB-MIT[5] of recordings of paediatric subjects with intractable seizures. Different methods of data management are done and documented to make the most of the algorithms used. The objective is to train an algorithm to acknowledge when the subject is having a seizure.

Keywords— electroencephalogram, deep learning, brain activity, classification, EEG analysis

1 INTRODUCTION

AN epileptic seizure is a period of symptoms due to abnormally excessive or synchronous neuronal activity in the brain. This can cause different effects like uncontrolled shaking movements involving much of the body, parts of the body or subtle momentary loss of awareness. In order to understand this issue, it is important to understand how neurons work and interact with each other to conserve what we call consciousness represented as brain activity and brainwaves.

Neural oscillations are rhythmic or repetitive patterns of neural activity in the central nervous system which can be driven by mechanisms within individual neurons or by interactions. Since 1824 neural oscillations have been observed, fifty years later intrinsic oscillatory behaviour was encountered in vertebrate neurons, but the purpose of these is yet to be fully understood.

As the main objective is to classify seizures from brain activity first of all, it will be needed an inside view on how the brain works to have a hint on how and what is done to extract or intercept information from the neurons to process externally in a computer. This information is available annexed in this paper just to have an overview to further understand the subject, but this matter is not the main purpose of this paper as it specifies in data processing, architecture, model strategies and classification results.

In this TFG we detail the complete process of epileptic sei-

zure detection, from data processing to seizure recognition. Because this TFG works as a pipeline of different stages, an insight view of each is done, starting from data processing from a well-known database (CHB-MIT) of encephalograms collected from 23 subject with interactable seizures that has been used in previous research. Followed by the strategy of the architecture used to classify the signals to finally understand the classified results into seizure and not seizure.

2 RELATED WORK

A lot of research has been made of the brain to further understand it's capabilities using deep learning algorithms. EEG signals can be used in many ways, below there are different projects with similar objectives of mental workload and seizure classification attempting to interpret and process EEG data, using deep learning techniques to develop frameworks to combine and process these signals.

2.1 Mental Workload Detection based on EEG Analysis

A study of mental workload[18] is done in order to work more efficiently, healthier and to avoid accidents since workload compromises both performance and awareness. The use of EEG signals has a high correlation with specific cognitive and mental states such as workload, proposing a binary neural network to classify EEG features across different mental workloads.

Mental workload is defined as “the cognitive and psychological effort to conclude a task”, observing that depending if workload is too heavy or too light this can affect human performance.

- E-mail de contacte: 1459024 @uab.cat
- Menció realitzada: Computació
- Treball tutoritzat per: Aura Hernández Sabaté (Ciències de la Computació)
- Curs 2021/22

There are two main categories to measure workload:

- Subjective measures: Being the most used to assess mental workload, the NASA Task Load Index (TLX) a prominent way to gain insight on perceived workload from the subject based on a weighted average of six sub-variables: mental demand, physical demand, temporal demand, performance, effort and frustration.
- Physiological measures: Providing a more reliable data by measuring physiological dynamic changes which cannot be controlled consciously, so that is why it's more reliable. Readings such as electrocardiogram (ECG), electromyograph, electroencephalogram (EEG)... The combination of inputs reports better accuracy than the analysis if each one independently.

The approach is to investigate the ability of 1D-CNN models to recognise two types of mental load from EEG signals and to generalise the model to a population not seen in the training set. They use N-back test (memory demanding games requiring the resolution of simple arithmetic operations adjusting workload) to induce low and medium workload and to classify a simple neural network (NN) it's trained using only the power spectrum of theta waves. Proposing a personalized model for each individual and a generalist one, perform the models in a dataset of 16 subjects showing outstanding results in a leave-one-out subject test and so generalizing to new unseen subjects.

The method to obtain data all 16 subjects had to first watch a 10-minute relaxing video and afterwards do the N-back-test low, medium and high difficulty. Finally subjects ask a TLX questionnaire for subjective perception of the test difficulty and workload.

To get the data itself EEG recordings were done using EMOTIV EPOC+ headset which has 14 electrodes placed according to the 10/20 system which provide raw data and power spectrum for the main brain rhythms (theta, alpha, beta low, beta high, and), at 128 Hz and 8 Hz, respectively. In conclusion its reported 95% confidence interval for each class computed for all subjects and recall above 90%. It is suggested to use longer windows to capture EEG non stationary nature.

2.2 EEG Signal Dimensionality Reduction And Classification Using Tensor Decomposition And Deep Convolutional Neural Networks

This paper has a similar objective of seizure classification and uses the same dataset as this TFG.

Using convolutional neural networks (CNNs), they still suffer from high dimensionality of the training data. The proposed tensor decomposition-based dimensionality reduction algorithm transforms a large set of slices of the input tensor to a concise set of slices which are called super-slices which handles the artifacts and redundancies of the EEG data and also reduces the dimension of the CNNs training inputs. This proposed framework is tested

on HCB-MIT data and as results show, the approach outperforms other previous studies.

2.3 Epileptic Seizures Detection Using Deep Learning Techniques: A Review

The advent deep learning algorithms in many areas of medicine, such as in the diagnosis of epileptic seizures, has made significant advances. In this study, a comprehensive overview of works focused on automated epileptic seizure detection using DL techniques and neuroimaging modalities is presented. Various methods proposed to diagnose epileptic seizures automatically using EEG and MRI modalities are described. In addition, rehabilitation systems developed for epileptic seizures using deep learning have been analysed, and a summary is provided. The rehabilitation tools include cloud computing techniques and hardware required for implementation of deep learning algorithms. Challenges, advantages and limitations in employing deep learning-based techniques for epileptic seizures diagnosis are presented as well as the most promising models and future works on automated epileptic seizure detection.

2.4 Recognition of mental workload of pilots in the cockpit using EEG signals

The commercial flightdeck is a naturally multi-tasking work environment and so automatic characterization of pilot's workload becomes essential. Electroencephalogram (EEG) signals have shown a high correlation to specific cognitive and mental states like workload. However, there is not enough evidence in the literature to validate how well models generalize in case of new subjects performing tasks of a workload similar to the ones included during model's training. In this paper it's proposed a convolutional neural network to classify EEG features across different mental workloads in a continuous performance task test that measures a portion of working memory and working memory capacity. Our model is valid at a general population level and it is able to transfer task learning to a pilot mental workload recognition in a simulated operational environment.

3 OBJECTIVES

There are two main objectives in this TFG. The first objective is to find the best way to treat and analyse before feeding it to the deep learning algorithms. For example, as a continuous stream of sequence, dividing data per subject... Or trying different test/train strategies such as leave one out, 50/50... The second main objective is to try to find which algorithms architectures have better results. There are different models offered by the research group IAM from the CVC.

To fulfil the objectives is crucial to define smaller objectives to make work easier. Within the first objective there are

several important parts.

- Raw data must be readable, as the data base CHB-MIT[5] is in European Data Format (EDF), a standard file format designed for exchange and storage of medical time series, so all files in the dataset are “.edf”. A script has been programmed to save .edf files into .parquet format.
- Setting different functions to filter data making sure data is okay and fits certain constraints.
- Get the labelling data, to have a ground truth from the recorded data. This part is essential to understand if the model works as expected.
- Define different functions to define how data enters the model to be trained. There are many ways information can be extracted from data.

The next objective after obtaining data is to use different models provided by the research group to IAM:

- Each model needs to be configured to accept the dimensionality of the data fed to it.
- Work different models to choose what models give better answers from input data.
- After all the models results, an overview is done to understand the results and conclude the best way to treat this database, for further investigation.

4 METHODS

4.1 Dataset

The dataset is data collected from the Children’s Hospital Boston, consisting in EEG recordings of subjects with intractable seizures. The folders classify in 23 cases from 22 subjects (case chb21 and chb1 are the same but 1.5 years apart). The subject’s personal information gender and age is in a separate file called SUBJECT-INFO added in this paper as subject_info.csv.

Each case contains between 9 and 42 .edf files. There are .edf files of EEG signals without seizures and others with recordings of seizures, these defined in RECORDS-WITH-SEIZURES. The files with seizures have the extension .edf.seizures which disables the possibility of accessing the file with a normal edf reader library.

Most cases have 1 hour of EEG recordings, but some have 1 to 4 hours depending on the case, split between 9 to 42 edf recordings, recorded at 256Hz in 16 bit resolution.

The data used to train the model in this TFG is data from subject 1 to subject 10, because the database is huge and time and resources are limited. Also only few .edf in each subject’s folders have any seizure. If every .edf was used there would be a lot more data labelled as not seizure than seizure data, so from each subject only the files with

seizure are used, and from these files data is cut to have a balance of labelled data of 50

The model of the CVC uses numpy array as input to work, and other data filtering is usually done in .parquet format, so this TFG uses mainly these two formats to work with data. Starting by .edf files on the database it’s processed to .parquet format, and once is saved again with this format it’s saved again in .npy.

The first script to be executed is ReadEDF.py which gets the raw data from every folder of every subject and filters every .edf from each one. It filters the raw data with the bandwidth desired to work with in the model and sets the labelling of every file. Because of the strategy used in this TFG every subject has an average of 5 .edf files. Each one has 22 different channels, which are the electrodes of the subject. In order to label the data, a new column is created (the 23rd) as seizure with the information of every row being a seizure or not and also a 24th column to set the observation windows for the model.

All this information is then saved as stated previously in parquet format to access if needed, furthermore in order for the model to accept data the parquets are split into two numpy arrays, one of data_x with the number of windows, electrodes and data. Another numpy array data_y with seizure window, being a dimension of data, which defines if a window has or not a seizure in it, and another dimension defining each window.

With this, all the parquets and arrays are specifically saved and ordered to ensure easy access and fast comprehension of the hierarchy of folders. The database folder has every subject folder and a folder to save the model if needed ordered in it by the date of execution. In every folder of the subject there are 3 more sorted by edf, parquet and numpy folders.

4.2 Network

An already done deep learning algorithm is used from the research group IAM from the CVC, which is working on a framework to determine the optimal architecture for cognitive state recognition from EEG signals, with the objective to answer different questions:

- How to combine the signals to create the input features for feature extraction? In this case, having 14 sensors x 5 waves, so 70 raw signals. These signals can be concatenated, or projected. As well, in case of projection, the weights can be equal or learned.
- Which neural network is the best performer?
- Is it better to ensemble the different classifiers before combining the signals?

This framework was originally intended to study brain workload, so, within this framework, the idea is to modify it to fulfil the objective of clinic seizure detection. In this TFG, different strategies are applied on the input data of the algorithm to further study it’s capabilities.

Because in this TFG the CHB-MIT Scalp EEG Database is

being used and all files are in format “.edf” a first script has been needed to change the format from “.edf” to “.parquet”. This script is called “03_ReadEDF.py”, being the first script to execute this TFG. It’s designed to extract the files from a specific folder hierarchy where all the encephalograms are classified by subjects. For simplicity this script obtains, filters, plots, and saves in parquets all input data. Datasets provided by the cvc have also been used, but the scripts used to read, filter and split data is “MainSource_vr1.py”. In the “03_ReadEDF.py” script there are different options on how raw data is imported, and also there are options on the way to execute them. Because during the development of this TFG many tests have been done, there are two different ways to execute the script:

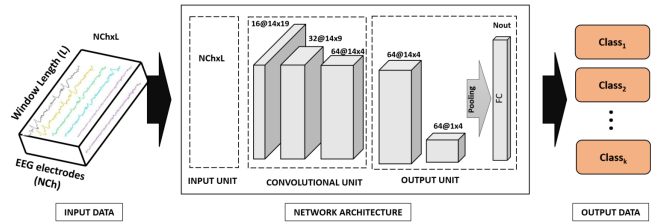
- Single execution, where the subject number and the edf file of the subject needs to be provided to execute the script for this single file.
- Multiple executions, where the number of subjects is provided. The script will go through all the first n subjects defined.

The script will automatically label all raw data using the summary file in each subject’s folder, so it’s important for it to be present or a label execution error will pop up. The files “.edf.seizures” in every subject’s folder were unreadable, even reading the binary was a failure. The script will make sure the file has all the data from the desired electrodes, this is important because hardware problems while recording the edfs some files have gaps or lack some data, if any edf file have this problem it will automatically be excluded and the user will be notified. Filtering data is done by first setting a maximum range from 0.5hz to 50hz, and afterwards only by changing the name of a parameter it can be changed to delta, theta, alpha, beta or gamma’s range frequencies in a dictionary added by default. Everything is modular so it can be changed any time with any range of frequencies. All data is saved in “.parquet” format in a different folder named parquets in every subject’s folder, if plotting is enabled it will plot each subject’s data.

After all desired raw data is filtered and saved, the second script to execute is “04_MExecution.py”, this one is in charge of the execution of the model to obtain the results classifying the data. All hyperparameters are defined at the beginning of the script, and afterwards the module is tested with random data to ensure it’s working as expected. For each parquet it splits it’s content depending of the percentage settled in training and testing, continuing by training the model and testing.

During the implementation of this procedure, a lot of problems of dependency on critical libraries has happened. Starting by cuda, for faster results it is used in all models, but it might not work if the architecture of the graphics card is too old. It is also not compatible with python 3.10 which was the version being worked on at the moment, it had to be switched to an environment with python 3.8 to avoid further issues.

Fig. 1: Scheme of data processing



During the implementation of this procedure, a lot of problems of dependency on critical libraries has happened. Starting by cuda, for faster results it is used in all models, but it might not work if the architecture of the graphics card is too old. It is also not compatible with python 3.10 which was the version being worked on at the moment, it had to be switched to an environment with python 3.8 to avoid further issues.

5 ANNEX

5.1 Neuron and neural activity

To further understand how brain activity works we first need to study a single neuron and its purpose. A neuron is an electrically excitable cell that has the function to communicate with other cells. It does it by nearly touching other cells called synapsis, transmits the message through its axon and delivers the message by synapsis[16] to another cell. Neurons are typically classified into types based on their function:

- Sensory neurons[14]: Which respond to stimuli of the sensory organs and send the signals to the spinal cord or brain.
- Motor neurons[9]: Its axons originate in the brain and spinal cord and innervate the muscles to produce muscle movements.
- Projection fiber[12]: Are neurons found in the central nervous system and only establish synapses with other neurons, these consist of efferent and afferent fibers uniting the cortex with the lower parts of the brain and with the spinal cord.
- Interneuron[6]: Is a neuron of the central nervous system, usually small and with a short axon, that interconnects with other neurons, but never with sensory receptors or muscle fibres, allowing it to perform more complex functions.

Neurons transmit electrical waves originating from a transient change of permeability in the plasma membrane. Their propagation is due to the existence of a potential difference that arises due to different concentrations of ions on either side of the membrane, as described by the Nernst potential, between the inner and outer part of the cell (typically -70 mV). For the transmission of nervous impulses to other neurons, these do it by synapse, being a

structure to pass electrical or chemical signals to another neuron or effector cell, there are two types of synapses[16]:

- Chemical synapse: Electrical activity in the presynaptic neuron is converted into the release of a neurotransmitter that binds to the receptors located in the plasma membrane of the postsynaptic cell.
- Electrical synapse: Is one in which transmission between the first neuron and the second is not by the secretion of a neurotransmitter, but by the passage of ions from one cell to another through gap junctions, small channels formed by the coupling of protein complexes, based on connexins, in closely adherent cells.

These electrochemical processes when large numbers of neurons show synchronized activity, electric fields that they generate can be large enough to be detected outside the skull, and so using electroencephalography (EEG) or magnetoencephalography (MEG) brain activity can be recorded.

5.2 Structure

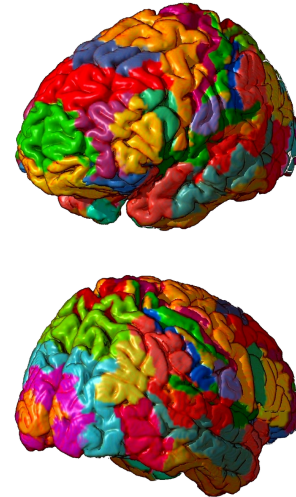
Now that we know where brain activity originates from, we can further study how the brain structures. There are many parts in the brain, but for now we are going to focus on the cerebrum because it initiates and coordinates movement, regulates temperatures, speech, judgement, reasoning, problem-solving, emotions, learning...

The cerebrum[2], it's the largest part of the brain, it's divided by the medial longitudinal fissure in two hemispheres, each of these hemispheres has an outer layer of grey matter, the cerebral cortex and an inner layer of white matter. The fact that these are separated gives the opportunity for lateralisation of brain functions, which is the tendency of neurological functions to specialise in one hemisphere or the other, but even though the cerebrum[2] is separated, these are connected by the corpus callosum.

The cortex is mapped into fifty different functional areas known as Brodmann's areas[1], defined by its cytoarchitecture (cellular composition), or histological structure and organization of cells. One scheme widely used (from Korbinian Brodmann)[1] splits the cortex into 52 different numbered areas of different cellular structure and different functions.

Having clarified this brain structure, obtaining data with electrodes from brain activity, positioning of these is something to keep in touch with depending on what it's being studied. The same goes for defining the dataset train and test data, to feed as input to the deep learning algorithm, opening a new field on how to treat and subdivide data to make the most of it.

Fig. 2: Brodmann areas by colors



6 BIBLIOGRAPHY

REFERÈNCIES

- [1] *Brodmann area*. URL: https://en.wikipedia.org/wiki/Brodmann_area (cons. 20-08-2021).
- [2] *Cerebrum*. URL: <https://en.wikipedia.org/wiki/Cerebrum#Functions> (cons. 20-08-2021).
- [3] *Corpus callosum*. URL: https://en.wikipedia.org/wiki/Corpus_callosum (cons. 20-08-2021).
- [4] *Cytoarchitecture*. URL: <https://en.wikipedia.org/wiki/Cytoarchitecture> (cons. 20-08-2021).
- [5] Ary L Goldberger et al. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals". A: *circulation* 101.23 (2000), e215 - e220.
- [6] *Interneuron*. URL: <https://en.wikipedia.org/wiki/Interneuron> (cons. 20-08-2021).
- [7] *Lateralization of brain function*. URL: https://en.wikipedia.org/wiki/Lateralization_of_brain_function (cons. 20-08-2021).
- [8] Mohsen Joneidi Mojtaba Taherisadr i Nazanin Rahnavard. "EEG SIGNAL DIMENSIONALITY REDUCTION AND CLASSIFICATION USING TENSOR DECOMPOSITION AND DEEP CONVOLUTIONAL NEURAL NETWORKS". A: (). URL: <https://arxiv.org/pdf/1908.10432v1.pdf>.
- [9] *Motor neuron*. URL: https://en.wikipedia.org/wiki/Motor_neuron (cons. 20-08-2021).
- [10] H Nahrstaedt. *Lee-Messer C. holgern/pyedflib* 2017.
- [11] *Neuron*. URL: <https://en.wikipedia.org/wiki/Neuron> (cons. 20-08-2021).

- [12] *Projection fiber*. URL: https://en.wikipedia.org/wiki/Projection_fiber (cons. 20-08-2021).
- [13] *Seizure*. URL: <https://en.wikipedia.org/wiki/Seizure> (cons. 20-08-2021).
- [14] *Sensory neuron*. URL: https://en.wikipedia.org/wiki/Sensory_neuron (cons. 20-08-2021).
- [15] Afshin Shoeibi et al. "Epileptic Seizures Detection Using Deep Learning Techniques: A Review". A: *International Journal of Environmental Research and Public Health* 18.11 (2021), pàg. 5780.
- [16] *Synapse*. URL: <https://en.wikipedia.org/wiki/Synapse> (cons. 20-08-2021).
- [17] Ashish Vaswani et al. "Attention is all you need". A: (2017), pàg. 5998 - 6008.
- [18] M Villaret et al. "Mental Workload Detection Based on EEG Analysis". A: (2021).
- [19] M Villaret et al. "Mental Workload Detection Based on EEG Analysis". A: *Artificial Intelligence Research and Development: Proceedings of the 23rd International Conference of the Catalan Association for Artificial Intelligence*. Vol. 339. IOS Press. 2021, pàg. 268.