A software pipeline for pre-processing and mining EEG signals: application in neurology

(Discussion Paper)

Chiara Zucco[1], Barbara Calabrese[1], Miriam Sturniolo[2], Antonio Gambardella[2] and Mario Cannataro[1]

[1]*Department of Medical and Surgical Sciences, University "Magna Græcia", Catanzaro*
[2]*Mater Domini Polyclinic, University "Magna Græcia", Catanzaro*

Abstract

Psychogenic non-epileptic seizures (PNES) resemble epileptic seizures, but they do not show the characteristic electrical discharges associated with epileptic seizures. Long term video monitoring combined with EEG recording is the gold standard in clinical practice, but this methodology is quite expensive and time-consuming. This paper presents a software pipeline to discriminate short-term interictal EEG from PNES and epileptic patients. The pipeline supports EEG signals pre-processing, features selection and classification. A first case study concerning the classification of 75 EEG (healthy, PNES and epileptic subjects) is under evaluation.

Keywords

EEG, Psychogenic Non-Epileptic Seizures, Machine Learning, pre-processing

1. Introduction

Psychogenic non-epileptic seizures (PNES) are sudden behavioural changes mimicking epileptic seizures but without EEG ictal patterns, caused by psychic alterations [1], [2]. PNES have been associated to a dysfunction in the processing of psychological or social distress, abuse during childhood or severe traumatic events [3], [4]. In [5], the authors reported an aggravation of the seizure frequency in PNES patients, a vulnerable group of people during this pandemic COVID-19.

Misdiagnosis with epilepsy may lead to treatments through antiepileptic drugs (AEDs) with the risk of iatrogenic morbidity and elevated cost for the patient and Health Care Systems [6]. It has been discovered that the correct diagnosis of PNES is usually delayed for an average of seven years [7], with a profound impact on patients and caregivers quality of life. The gold standard for PNES diagnosis is represented by the visual examination of seizures captured during video-electroencephalography (video-EEG), either spontaneously or provoked by intermittent photic stimulation (IPS) suggestion techniques. These methods are time-consuming and ethically

SEBD 2021: The 29th Italian Symposium on Advanced Database Systems, September 5-9, 2021, Pizzo Calabro (VV), Italy
zucco@unicz.it (C. Zucco); calabreseb@unicz.it (B. Calabrese); sturniolom@yahoo.it (M. Sturniolo); a.gambardella@unicz.it (A. Gambardella); cannataro@unicz.it (M. Cannataro)
0000-0003-0048-0457 (C. Zucco); 0000-0003-2123-8221 (B. Calabrese); 0000-0001-7384-3074 (A. Gambardella); 0000-0003-1502-2387 (M. Cannataro)

disputable; thus, since the experts evaluate no definite criterion, visual analysis of EEG signals is insufficient. The International League Against Epilepsy (ILAE), using a consensus review of literature, evaluated fundamental diagnostic approaches, including detailed history and seizures description, EEG, video-EEG, neurophysiology, neuroimaging, hypnosis, and neuro-humoral monitoring [6].

To the best of our knowledge, only a few studies have investigated semi-automatic or automatic machine learning-based approaches for discrimination between epileptic and PNES subjects by only considering EEG recordings. The current study is one of the few in which an EEG dataset, without any correlated video-EEG PNES marker, has been analyzed to discriminate PNES via EEG.

Specifically, this paper proposes a novel and semi-automatic pipeline to discriminate PNES subjects from epileptic subjects based on the extraction of spectral features and classification through a Machine Learning-based approach. To design a software pipeline that allows discrimination, we have implemented different classifiers, i.e. Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Bayesian Network (BN). In this work, we have analyzed EEG recordings of twenty-five PNES subjects, twenty-five healthy volunteer subjects and twenty-five epileptic patients.

The paper is organized as follows: Section 2 presents the background relative to EEG analysis and classification, focusing on PNES discrimination; Section 3 describes experimental EEG acquisitions. Section 4 presents and discusses the proposed software pipeline, highlighting the features of software modules. Finally, Section 5 concludes the paper.

2. Background on EEG analysis and classification

The EEG signals are random and non-stationary and may contain valuable information about the brain state with millisecond temporal resolution. However, it isn't effortless to disentangle information from these signals directly in the time domain by observing them. EEG spectrum contains some characteristic waveforms that fall primarily within five frequency bands: delta (1−-4 Hz), theta (4−-8 Hz), alpha (8-−13 Hz), beta (13-−30 Hz) and gamma (30−60 Hz). Many efforts have been made for automatic and semi-automatic EEG processing by exploiting several algorithms that operate in time, frequency, or the time-frequency domain. Among the time domain (TD) features, Mean Absolute Value (MAV), Zero Crossings (ZC), Slope Sign Changes (SSC) and Waveform Length (WL) are some examples. Fourier Transform (FT) and Power Spectral Analysis (PSD) are the most common features extracted in the frequency domain. The most used methods are Wavelet Transform (WT) and Hilbert - Huang Transform in the time-frequency domain.

There are many studies on EEG data analysis. The work [8] describes a new method to identify seizures in EEG signals using feature extraction in time-frequency distributions (CFDs). In [9], the authors analyze the differences in power spectral density between healthy subjects and schizophrenic patients in various frequency bands. Moreover, frequency and time-frequency approaches have been applied for analysis of patients with brain damage [10], automatic detection of epileptic seizures [11] and, also, early detection of mild cognitive impairment and Alzheimer's disease [12].

Various data mining techniques such as Association Rule, classification, regression, and clustering have been used in the literature. Support vector machine (SVM) is a machine learning model used for classification, and regression analysis [13].

Differential diagnosis cannot rely only on clinical features of PNES because most of the signs fit with epileptic seizures. Longer duration, gradual onset, asynchronous movement, closed eyes, lateralized head movement contribute to a complex clinical pattern of PNES. These events are involuntary and out of the patient's conscious control. PNES patients mimic the different type of epileptic seizures but no epileptic seizures electrophysiological patterns. Thus discrimination would be useful to aid neurologist diagnosis and consequently assign the right pharmacological treatment. The prevalence of PNES is high in selected populations: 5−-20% in outpatient epilepsy populations [14], [15]. 10−-40 % of patients referred to tertiary epilepsy centers for medically refractory seizures [16]. Making PNES diagnosis is soupy with medico-legal hazards, and it is hard to prove [17].

EEG recordings alone are not sufficient to diagnose PNES: an ictal scalp EEG may show no epileptic features during simple partial seizures or mesial frontal lobe seizures. The latter may be easily mistaken for PNES. The differentiation between non-epileptic seizures and healthy subject can be difficult. Therefore, a good knowledge of the semiology of PNES is essential for the early screening of patients for video-EEG monitoring (VEM) and correct interpretation of the examination. PNES should also be differentiated from physiological, non-epileptic events such as syncope, cataplexy, migraine, paroxysmal movement disorders, breath-holding spells, stable ictal heart rate [18].

Additionally, VEM monitoring is highly time-consuming and labour intensive and therefore is relatively expensive and limited in availability; thus, alternative diagnostic procedures are necessary.

Despite many efforts made, no bio-marker of PNES has yet been identified. In [19], the authors sustain patients with PNES have a stable frequency of rhythmic movements, about (5Hz), which produces a stable rhythmic artefact on the EEG with little variation or evolution. Bolen et al. [20] have noted significantly more multifocal abnormalities in frontal, temporal, parietal, occipital, cerebellar and brainstem brain regions PNES patients.

Continuous wavelet transform is used in [21] to process CNT and PNES EEG signal.

In [22], a machine learning (ML) pipeline for classifying EEG epochs of PNES and healthy controls is introduced. This software pipeline consisted of a semi-automatic signal processing technique and a supervised ML classifier to aid the discriminative clinical diagnosis of PNES. Statistical features like the mean, standard deviation, kurtosis, and skewness were extracted from a power spectral density (PSD) map split up into the five conventional EEG rhythms. Finally, we evaluated three different supervised ML algorithms, namely, the Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and Bayesian network (BN), to perform EEG classification tasks for control vs PNES subjects. The algorithm's performance was evaluated on a dataset of 20 EEG signals (10 PNES and 10 control), achieving an average accuracy above 90 %.

3. EEG signals acquisition

The data under evaluation were collected from the Operative Unit of Neurology, Mater Domini Polyclinic, University of Catanzaro, Italy. In this study, we analyzed EEG recordings from 25 patients with PNES, 25 healthy patients referred to as CNT and 25 epileptic patients. PNES diagnosis was made based on a typical episode recorded during video-EEG, with EEG showing neither concomitant ictal activity nor post-ictal. Healthy controls did not suffer from any neurological disorder and had a normal neurological examination. None of the subjects was on chronic medication or had received any drug up to 24 hours before EEG acquisition. The study was conducted following the Declaration of Helsinki and formally approved by the local Medical Research Ethics Committee. Participants were comfortably seated in a semi-darkened room and with open eyes. The technician kept the subject alert in order to prevent drowsiness.

EEG recordings were conducted in poorly light-room using 19 Ag/AgCl surface electrodes placed according to the International 10/20 System. Recordings were performed with a Xltek® Brain Monitor EEG Amplifier with a sampling rate of 256 Hz; high-pass filter at 0.5 Hz and a low-pass filter at 70 Hz, plus a 50 Hz notch filter. All of the EEG signals were recorded using a montage with the following channels layout: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz and Pz, and reference in Pz.

All the electrode-skin impedance has been kept below 5 KΩ. The EEG data were recorded in a resting condition. Each participant was seated in a comfortable chair in an electrically and acoustically shielded room. EEG artefacts were rejected through a semi-automatic procedure: (i) first artefacts components were individuated through clinical (visual) inspection; (ii) all signal segments affected by evident artifactual components are cut off from raw EEG data.

4. EEG software pipeline

The architecture of the proposed EEG software pipeline (written in Python language) presents the following modules:

- Pre-processing: this module performs digital filtering and semi-automatic artefact rejection;
- Features extraction: Power Spectral Density (PSD) function is evaluated for each channel by using Welch's method. From PSD functions, cumulative power coefficients in clinical bands are calculated to build features vector to fed in input to the classifiers;
- Features selection: searches for the most significant features for classification;
- EEG classification: the framework implements different machine learning approaches, such as Support Vector Machine, Bayesian Networks and Linear Discriminant Analysis to discriminate PNES EEG from CNT EEG.

The rest of the section details the features of each module.

4.1. EEG Pre-processing

In general, EEG signals are contaminated by environmental noise or even distorted by artefacts. Removing noise is an important step in EEG signals processing. A correct data cleaning may improve the signal to noise ratio and helps to disentangle the most informative features from the signal. Even if different semi-automatic approaches have been proposed for EEG pre-processing, in clinical practice, the rejection of artefacts is performed manually by trained neurologists by discarding contaminated EEG epochs. Therefore, the pre-processing stage is operator-dependent, tedious and time-consuming.

In this study, each EEG recording was checked by a trained neurologist to marker noise and artefact corrupted epoch. Subsequently, all EEG data were pre-processed using digital filtering techniques. Specifically, we have used a Butterworth band-pass filter (0.1-70 Hz) and a notch filter (cut-off frequency 50 Hz) to reduce high-frequency artefacts and power-line interference.

4.2. EEG Features Extraction

After the pre-processing step, the next step in the EEG software pipeline is the features extraction stage. As mentioned before, features extraction aims to extract relevant information contained in the signals. One of the most used features extractor in EEG signals processing is the Power Spectral Density (PSD) analysis. Different methods for PSD estimation have been reported in the literature. We have chosen Welch's method.

Welch Spectrum Welch's method is a well-known non-parametric method for PSD estimation. Let be $x[n]$, $n = 0, \cdots, N-1$ are the samples from an EEG segment. To estimate the Welch's spectrum of this segment, three basic steps are applied:

- First, divide the original EEG time series into N sections (possibly overlapped O) of equal lengths M;

$$x\lfloor n \rfloor = x\lfloor n + iO \rfloor \qquad i = 0, \ldots K-1, \text{ and } n = 0, \ldots N-1 \tag{1}$$

- Apply a window to each section and then calculate the periodogram on the windowed sections (modified periodograms). The periodogram is defined as:

$$\tilde{S}_{xx}(k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-\frac{2\pi jkn}{N}}\right|^2. \tag{2}$$

- Average the modified periodograms from the K sections in order to obtain an estimator of the spectral density.

$$P_{yx}(f) = \frac{1}{K}\sum_{i=0}^{K-1} P_i(f) \tag{3}$$

Where P_{yx} estimates the cross power spectral density of two discrete-time signals, x and y, the Welch method eliminates the tradeoff between spectral resolution and variance by allowing the segments to overlap. If a high-frequency resolution is desired, one can only split the record into a small number N of segments of length L. In system identification, the number of segments K is typically 2,3, 6. Unfortunately, a small number of segments K implies a higher variance of the estimated power spectrum. Therefore, it is worth looking into methods using overlapping sub-records for reducing the variance as the main processing algorithm remains unchanged. Our analysis uses a segment with 50% overlap for the first step and the Hamming window in the second step.

In this paper, power spectral density (PSD) of classical frequency bands from around 1 Hz to 70 Hz were used as features. Specifically, delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), gamma (30-70 Hz) bands have been considered.

The main processing steps of our feature extraction approach are:

- Power Spectral Density (PSD) was estimated trough Welch method;
- from PSD matrix output, we selected five frequency sub-bands;
- for each band we computed cumulative power.

All features extracted are arranged into a vector, known as a features vector.

4.3. Features Selection and EEG Classification

Features selection is useful in the presence of high-dimensional data as it reduces the space of the features. This module is still under implementation. To discriminate CNT EEG from PNES EEG, we have implemented three supervised Machine Learning algorithms. The classifiers used in this paper are the SVM, LDA, and BNs.

The SVM algorithm is based on the statistical learning theory. An SVM [23] constructs a separating hyperplane or a set of hyperplanes in a high dimensional space. Intuitively, a good separation is achieved by the hyperplane with the largest distance to the nearest training data points of any class (so-called functional margin).

Bayesian networks (BNs) [20], also known as belief networks, belong to probabilistic graphical models (GMs). These structures are used to extract knowledge about an uncertain domain. Each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. The graph dependencies are estimated by using known statistical and computational methods. Hence, Bayesian Networks provide a simple definition of independence between any two distinct nodes.

LDA classifier is one of the classification methods that finds an optimal linear transformation that maximizes the class separability. LDA generates a linear combination of data sets that allows the largest mean differences between the desired classes. It works well when the feature vector is multivariate normally distributed in each class group, and different groups have a common covariance.

5. Conclusions

The main contribution of this paper is a comprehensive software pipeline written in Python for the acquisition, pre-processing and classification of EEG signals. A first experimentation of the pipeline in the classification of PNES vs epileptic subject is under evaluation by the clinical team of the Neurology Unit of the University Magna Græcia of Catanzaro. The preliminary results show that a features selection phase is necessary to improve the performance of the classification.

References

[1] N. M. G. Bodde, J. L. Brooks, G. A. Baker, P. A. J. M. Boon, J. G. M. Hendriksen, O. G. Mulder, A. P. Aldenkamp, Psychogenic non-epileptic seizures–definition, etiology, treatment and prognostic issues: a critical review, Seizure 18 (2009) 543–553.

[2] S. Gasparini, M. Campolo, C. Ieracitano, N. Mammone, E. Ferlazzo, C. Sueri, G. G. Tripodi, U. Aguglia, F. C. Morabito, Information Theoretic-Based interpretation of a deep neural network approach in diagnosing psychogenic Non-Epileptic seizures, Entropy 20 (2018) 43.

[3] G. Baslet, A. Roiko, E. Prensky, Heterogeneity in psychogenic nonepileptic seizures: understanding the role of psychiatric and neurological factors, Epilepsy Behav. 17 (2010) 236–241.

[4] A. A. Uliaszek, E. Prensky, G. Baslet, Emotion regulation profiles in psychogenic non-epileptic seizures, Epilepsy Behav. 23 (2012) 364–369.

[5] A. R. B. G. M. R. d. S. B. P. A. Valente, K.D., The covid-19 outbreak and pnes: The impact of a ubiquitously felt stressor, Epilepsy Behavior 117 (2021).

[6] W. C. LaFrance, Jr, S. R. Benbadis, Avoiding the costs of unrecognized psychological nonepileptic seizures, Neurology 66 (2006) 1620–1621.

[7] M. Reuber, T. Kral, M. Kurthen, C. E. Elger, New-onset psychogenic seizures after intracranial neurosurgery, Acta Neurochir. 144 (2002) 901–7; discussion 907.

[8] C. Guerrero-Mosquera, A. M. Trigueros, J. I. Franco, A. Navia-Vázquez, New feature extraction approach for epileptic EEG signal detection using time-frequency distributions, Med. Biol. Eng. Comput. 48 (2010) 321–330.

[9] Y. Akbar, S. N. Khotimah, F. Haryanto, Spectral and brain mapping analysis of EEG based on pwelch in schizophrenic patients, in: Journal of Physics: Conference Series, volume 694, 2016, p. 012070.

[10] M. P. Davey, J. D. Victor, N. D. Schiff, Power spectra and coherence in the EEG of a vegetative patient with severe asymmetric brain damage, Clin. Neurophysiol. 111 (2000) 1949–1954.

[11] A. T. Tzallas, M. G. Tsipouras, D. I. Fotiadis, Epileptic seizure detection in EEGs using Time–Frequency analysis, IEEE Trans. Inf. Technol. Biomed. 13 (2009) 703–710.

[12] J. C. McBride, X. Zhao, N. B. Munro, C. D. Smith, G. A. Jicha, L. Hively, L. S. Broster, F. A. Schmitt, R. J. Kryscio, Y. Jiang, Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early alzheimer's disease, Comput. Methods Programs Biomed. 114 (2014) 153–163.

[13] I. Guler, E. D. Ubeyli, Multiclass support vector machines for EEG-Signals classification, 2007.

[14] A. A. Leis, M. A. Ross, A. K. Summers, Psychogenic seizures: ictal characteristics and diagnostic pitfalls, Neurology 42 (1992) 95–99.

[15] GATES, J. R, The classification and treatment of nonepileptic events, Epilepsy and behavior (1991).

[16] R. Martin, J. G. Burneo, A. Prasad, T. Powell, E. Faught, R. Knowlton, M. Mendez, R. Kuzniecky, Frequency of epilepsy in patients with psychogenic seizures monitored by video-EEG, Neurology 61 (2003) 1791–1792.

[17] S. R. Benbadis, Nonepileptic behavioral disorders, 2013.
[18] C. Opherk, L. J. Hirsch, Ictal heart rate differentiates epileptic from non-epileptic seizures, Neurology 58 (2002) 636–638.
[19] A. Vinton, J. Carino, S. Vogrin, L. MacGregor, C. Kilpatrick, Z. Matkovic, T. O'Brien, Convulsive nonepileptic seizures have a characteristic pattern of rhythmic artifact distinguishing them from convulsive epileptic seizures, Epilepsia 45 (2004) 1344–1350.
[20] R. D. Bolen, E. H. Koontz, P. B. Pritchard, 3rd, Prevalence and distribution of MRI abnormalities in patients with psychogenic nonepileptic events, Epilepsy Behav. 59 (2016) 73–76.
[21] S. Gasparini, E. Beghi, E. Ferlazzo, M. Beghi, V. Belcastro, K. P. Biermann, G. Bottini, G. Capovilla, R. A. Cervellione, V. Cianci, G. Coppola, C. M. Cornaggia, P. De Fazio, S. De Masi, G. De Sarro, M. Elia, G. Erba, L. Fusco, A. Gambardella, V. Gentile, A. T. Giallonardo, R. Guerrini, F. Ingravallo, A. Iudice, A. Labate, E. Lucenteforte, A. Magaudda, L. Mumoli, C. Papagno, G. B. Pesce, E. Pucci, P. Ricci, A. Romeo, R. Quintas, C. Sueri, G. Vitaliti, R. Zoia, U. Aguglia, Management of psychogenic non-epileptic seizures: a multidisciplinary approach, Eur. J. Neurol. 26 (2019) 205–e15.
[22] G. Varone, S. Gasparini, E. Ferlazzo, M. Ascoli, G. G. Tripodi, C. Zucco, B. Calabrese, M. Cannataro, U. Aguglia, A comprehensive machine-learning-based software pipeline to classify eeg signals: A case study on pnes vs. control subjects, Sensors 20 (2020).
[23] K.-R. Muller, K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, B. Scholkopf, An introduction to kernel-based learning algorithms, 2001.