

ACCEPTED MANUSCRIPT

On the use of patterns obtained from LSTM and feature-based methods for time series analysis: application in automatic classification of the CAP A phase subtypes

To cite this article before publication: Fábio Mendonça *et al* 2020 *J. Neural Eng.* in press <https://doi.org/10.1088/1741-2552/abd047>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2020 IOP Publishing Ltd.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by-nc-nd/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

On the use of Patterns Obtained from LSTM and Feature-Based Methods for Time Series Analysis: Application in Automatic Classification of the CAP A Phase Subtypes

Fábio Mendonça^{1,2}, Sheikh Shanawaz Mostafa^{1,2}, Fernando Morgado-Dias^{2,3} and Antonio G. Ravelo-García^{2,4}

¹ Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal

² Interactive Technologies Institute (ITI/LARSyS) and M-ITI, 9020-105 Funchal, Portugal

³ Faculty of Exact Sciences and Engineering, University of Madeira, 9000-082 Funchal

⁴ Institute for Technological Development and Innovation in Communications, Universidad de Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain

E-mail: fabio.mendonca@tecnico.ulisboa.pt

Received xxxxxx

Accepted for publication xxxxxx

Published xxxxxx

Abstract

The cyclic alternating pattern is a marker of sleep instability identified in the electroencephalogram signals whose sequence of transient variations compose the A phases. These phases are divided into three subtypes (A1, A2, and A3) according to the presented patterns. The traditional approach of manually scoring the cyclic alternating pattern events for the full night is unpractical, with a high probability of miss classification, due to the large quantity of information that is produced during a full night recording. To address this concern, automatic methodologies were proposed using a long short-term memory to perform the classification of one electroencephalogram monopolar derivation signal. The proposed model is composed of three classifiers, one for each subtype, performing binary classification in a one versus all procedure. Two methodologies were tested: feed the pre-processed electroencephalogram signal to the classifiers; create features from the pre-processed electroencephalogram signal which were fed to the classifiers (feature-based methods). It was verified that the A1 subtype classification performance was similar for both methods and the A2 subtype classification was higher for the feature-based methods. However, the A3 subtype classification was found to be the most challenging to be performed, and for this classification, the feature-based methods were superior. A characterization analysis was also performed using a recurrence quantification analysis to further examine the subtypes characteristics. The average accuracy and area under the receiver operating characteristic curve for the A1, A2, and A3 subtypes of the feature-based methods were respectively: 82% and 0.92; 80% and 0.88; 85% and 0.86.

Keywords: A phase subtypes, CAP, LSTM, Recurrence quantification analysis, Sleep quality

1. Introduction

Sleep is a complex physiological process associated with the circadian rhythm [1], which is usually examined by polysomnography (PSG) [2]. This technique comprises the employment of multiple source sensors to attain accurate diagnosis [3]. Among the recorded sensors, the electroencephalogram (EEG) provides unique and significant information about the sleeping brain, working as the reference to define the sleep structure by considering five characteristic

frequency bands: Delta (0.5 to 4 Hz); Theta (4 to 8 Hz); Alpha (8 to 12 Hz); Sigma (12 to 15 Hz); Beta (15 to 30 Hz) [4].

Conventionally, sleep is divided into macrostructure and microstructure [2] [5]. The macrostructure is a stepwise profile that categorizes sleep in stages, according to the prevalent EEG activity, in consecutive thirty second epochs (standardized scoring epoch). This structure oscillates between periods of Rapid Eye Movement (REM) and Non-REM (NREM), which is divided into three stages (N1, N2,

and N3) [2]. On the other hand, the microstructure considers the time structure of phasic EEG events that are observed during the NREM stages [5]. These events have a shorter duration than the standardized scoring epoch; thus the one second duration epochs paradigm was developed for their evaluation [6]. The Cyclic Alternating Pattern (CAP) concept was developed as a tool to evaluate the microstructure. It is characterized by sequences of transient electrocortical events that are dissimilar from the background EEG activity [7].

The CAP is composed of a sequence of transient EEG variations, named A phases, that break the background rhythm of the ongoing sleep stage. Each A phase is directly followed by an intermittent recovery of background activity, identified as the B phase. Both A and B phases duration can range from two to sixty seconds [7].

The amplitude and spectral contents of the EEG signal characterize the A phase and its three subtypes [6] [8]:

- **A1:** Linked to mild or minor polygraphic variations and is characterized by synchronized EEG patterns (i.e. high-voltage slow waves), specifically, sequences of delta bursts or K-complexes in the N2 and N3 stages and with the intermittent alpha rhythm in N1. Although high-voltage delta waves are the most prevalent in this subtype, the desynchronized pattern (i.e. low-amplitude fast rhythms) can still occur, but it must account for less than 20% of the total activation time.
- **A2:** Associated with an increase of cardiorespiratory rate phases and/or muscle tone, with desynchronized EEG patterns, mixed with (or are preceded by) slow high-voltage waves such as arousals with slow wave synchronization, K-alpha, and K-complexes with alpha and beta activities. These rapid activities account for 20% to 50% of the total activation time.
- **A3:** Related with the occurrence of desynchronized EEG patterns which duration exceeds two thirds of the activation time. These are coupled with a significant increase in cardiorespiratory rate phases and/or muscle tone. Therefore, more than the 50% of the total phase has rapid activities (particularly in the beta band).

The incidence of the A phase subtypes is correlated with age, as defined by Terzano et al. [6]. For healthy subjects, A1 is the most prevalent, followed by A2 and A3 is the less predominant. However, as the age progresses, the incidence of A1 decreases (71% for adolescence, 61% for young adults, 62% for mature adults, and 47% for senescence) while A3 increases (9% for adolescence, 11% for young adults, 11% for mature adults, and 18% for senescence). The prevalence of the subtypes is almost the same during young and mature adulthood. The CAP phase's transitions (A phase onset and offset) also exhibit characteristic patterns that can remarkably separate the activation phase from the EEG background activity [9] [10]. However, the detection of onsets was found to be easier to perform (showing values of energy that were on average three times higher than B phases values at all the sleep stages [9]) while the offsets present a smooth transition on the

way to the basal sleep stage oscillations, thus, are harder to identify [11]. It was also reported that subtypes A1 and A3 are easier to separate, while the A2 subtype is similar to the A1 subtype [11].

CAP is strongly related to the dynamic organization of sleep, as it designates a condition of the instability of the vigilance level, translating the brain effort to regulate and maintain the sleep macrostructure [8] [6]. It was suggested that the EEG synchrony, during sleep, is build-up and preserved through a fluctuating process of slow activities, associated with the A1 subtype, combined with a powerful inhibition of rapid EEG shifts linked to the A2 and A3 subtypes [6]. Therefore, the A1 subtype is mostly involved with the construction and consolidation of slow-wave sleep while A2 and A3 subtypes modulate the REM sleep onset [8] [6]. It was also verified a prevalence of the A1 subtype (92%) through the sleep cycles in the descending branch while the A2 and A3 subtypes (respectively, 45% and 19%) are more common in the ascending branch [12] [13].

Taking into consideration the role of CAP in the sleep process, it was proposed that the regulatory mechanisms underlying the CAP recurrent fluctuations may become the pathophysiological source for disordered sleep [6] [14]. Therefore, alterations of the A phase subtypes can be a marker for the occurrence of sleep disorders such as sleep apnea [15], periodic limb movements [16], insomnia [17], narcolepsy [18], and nocturnal frontal lobe epilepsy [19]. Hence, CAP can express the stability of sleep and be considered as a sleep quality marker [20]. It was also observed that the temporal occurrence of the A phases shows consistency of self-affinity properties under different pathologies, suggesting that sleep is itself a resilient process [21].

All approaches found in the state of the art for the automatic A phase subtypes classification are based on the application of tuned thresholds or employing a machine learning classifier using multiclass methods. However, threshold-based methodologies can possibly be problematic to generalize for a broader population. On the other hand, CAP analysis based on machine learning typically suffers from strong imbalance data, where the number of epochs related to the background activity is significantly larger than the epochs associated with the A phase subtypes [22]. Therefore, the subtypes classification usually has a poor performance for at least one class where the sensitivity is low. Taking into consideration that each subtype has relevant information that requires to be individually analyzed to assess metrics (such as the incidence percentage of the subtype or the frequency of occurrence), the output of the multiclass classification is frequently converted to a binary time series that contains the variation of a subtype in the one versus all representation. Therefore, a novel approach was followed in this work, using three individual binary classifications (based on machine learning), one for each subtype, with the goal of investigate if a better performance can be attained when comparing to the other methods (since the model only focus on optimizing the

detection of one class) and provide the most relevant information that is usually evaluated by the physicians when a multiclass output is provided.

It was also verified that most works employ features created by the researchers to improve the classification. However, classifiers such as LSTM can find relevant patterns in the signal that were not previously emphasized by any feature found in the state of the art. Therefore, two approaches were followed in this work: using features selected by a ranking procedure to feed the classifier (feature-based methods); to feed the pre-processed signal to the classifier (methods without an explicit feature creation procedure).

Hence, the objective of this work is to evaluate the approach of using multiple binary classifications (one for each subtype), testing the feature-based methods and the methods without an explicit feature creation procedure. It is also intended to perform a characterization analysis for each subtype to attain a broader view of the characteristic patterns associated with them.

This work is composed of five sections, organized as follows: an overview of the state of the art is presented in section 2; materials and methods are presented in section 3; the performance and results analysis of the developed models is presented in section 4; discussion of the results is presented in section 5; the work is concluded in section 6.

2. State of the art

Several classifiers were previously proposed to perform the CAP phase analysis [22] [23] [24] [25], and it was verified that artificial neural networks attained the best performance. Among these, the Long Short-Term Memory (LSTM), a type of the recurrent neural network, was identified as possibly the most suited to perform the classification [25] [26] since the CAP phases have a strong temporal dependency [7]. Such temporal correlations frequently occurs in the physiological signals [27] and can be identified by the LSTM, capable of finding both short and long-term correlations in the time series [28].

The usual approach for the A phase subtypes evaluation employs features that extract relevant information from the EEG signal. A summary of the features reported in the state of the art is presented in Table I. Most works employ statistical analysis to characterize the subtypes by evaluating the features [29], [30], [31], [32]. However, few works were found, in a literature review that was carried out for this work, performing the epoch based subtypes classification by feeding the features (or the EEG signal) to a classification procedure. Two main approaches for this procedure were found in the state of the art. The first performs the classification by applying tuned thresholds to the features. This methodology was used by Navona et al. [33] to examine a Macro-Micro Structure Descriptor (MMSD). Machado et al. [34], evaluated the Teager Energy Operator (TEO) and a MMSD for each of the characteristic EEG frequency bands. It was verified that, in

general, TEO was the best feature.

Table I. Summary of the features presented in the state of the art, presenting for each feature the usual name (employed to reference the feature), a brief description, and the publications where it was used.

Name	Description	Publications
Discrete time short time Fourier transform	Application of the discrete Fourier transform to the signal divided in windows to further assess other features such as the frequency of maximum energy or the frequency of mean energy	[36], [35]
Empirical mode decomposition	Decompose the signal into intrinsic mode functions, and each function denotes an embedded characteristic oscillation on a disjointed time scale	[36], [35]
Fractal dimension	Counts the number of occurrences of a sequence	[36], [35], [11], [32], [9]
Hjorth activity	Evaluates the variance of the signal's amplitude	[23], [26]
Lempel–Ziv complexity	Metric that evaluates the randomness of a finite sequence	[36], [35], [11], [32], [9], [31]
Macro–micro structure descriptor (band descriptor)	Measure of how the average EEG signal amplitude, in a defined frequency band, differs from its background	[23], [26], [36], [35], [30], [33], [34]
Power spectral density	Describes the power distribution of the signal into the frequency components	[29], [11]
Sample entropy	Measures the regularity of the signal	[11], [32], [9]
Shannon entropy	Evaluates the complexity of a signal by estimating the probability for a given value to occur (high probabilities suggest that the signal has less information, leading to a smaller entropy)	[26], [36], [35]
Statistical features	Statistical based characteristics such as mode, kurtosis, skewness, and standard deviation	[11], [9]
Teager energy operator	Non-linear estimation of the instantaneous energy	[26], [36], [35], [34]
Tsallis entropy	Metric that is a one-parameter generalization of the Shannon entropy	[11], [32], [9]
Variance	Measurement of the signal's spread	[23], [26], [36], [35]
Zero-crossing ratio	Number of baseline crossings and can provide information about the dominant frequency	[36], [35]

The second approach employs a machine learning algorithm to learn the patterns from the features (or the EEG signal) and perform the classification. This methodology was used by Machado et al. [35] [36], evaluating the MMSD, TEO, zero-crossing rate, Lempel–Ziv Complexity (LZC), discrete time short time Fourier transform, empirical mode decomposition, Shannon Entropy (SE), Fractal Dimension (FD), and variance. The features' relevance was assessed by the Minimal-Redundancy-Maximal-Relevance (mRMR) algorithm and fed to a classifier. Quadratic discriminant analysis, k-Nearest Neighbors (k-NN), and Support Vector Machine (SVM) were tested, and it was verified that SVM attained the best results using the top forty features ranked by

mRMR. Hartmann and Baumert [26] used the Hjorth activity, SE, TEO, MMSD, and differential variance as features, fed to an LSTM. It was concluded that the usage of time sequences improved the classification performance.

A group of statistics (standard deviation, mode, kurtosis, and skewness), frequency (power in the characteristic EEG frequency bands) and complexity (LZC, FD, sample entropy, and Tsallis entropy) features were evaluated by Mendez et al. [11]. These features fed a k-NN to perform the classification. It was observed that the A2 subtype is the most commonly misclassified due to their similarity with the other two subtypes. Therefore, it was suggested the need for a deep study to understand how to measure other features that the experts detect, which could possibly not be explicitly defined in the scoring rules. Arce-Santana et al. [37] fed the EEG signal to a Convolutional Neural Network (CNN), training the classifier with 25% of the total activation phases present in the studied population. It was proposed that an ad-hoc classifier, trained for each subject, could possibly provide an alternative view to the fully automatic methods.

Through the state of the art analysis, it was observed that LSTM is a prime candidate for an analysis characterized by a strong temporal dependency. It was also verified that multiple features are relevant for the subtype classification where information is present in both time and frequency domains.

3. Materials and methods

Two methodologies were studied to examine the signal from one EEG monopolar derivation. The first created features from the pre-processed signal, chosen according to a ranking and fed them to the classifier. The second fed the pre-processed signal to the classifier, without an explicit feature creation procedure.

The epoch by epoch evaluation of the A phase subtypes was performed by three individual binary classifications, each executing a one versus all analysis. The algorithms were developed in Python 3, and the block diagram of the experimental procedure is presented in Fig. 1.

3.1 Studied Population

Recordings from fifteen subjects (nine females and six males), were selected from the Physionet CAP Sleep Database [7] [38]. The subjects were free of neurological disorders and were recorded at the Sleep Disorders Center of the Ospedale Maggiore of Parma. A single-channel analysis was performed, evaluating a monopolar derivation (C4–A1 or C3–A2) that is considered to be essential for CAP scoring [7]. The recording's length ranged from 430 to 573.5 minutes, with an average of 499.62 minutes. The subject's age ranged from 23 to 42 years old, with an average of 32.19 years old.

Annotations regarding the A phases were provided by expert neurologists according to the scoring rules defined by Terzano et al. [7]. The distribution of the database labels over all the epochs (considering each one second epoch's

annotation) was 90%, 3%, 2%, and 5%, for “not-A”, “A1”, “A2”, and “A3”, respectively. It was observed that the dataset is strongly unbalanced, having considerably more “not-A” one second epochs than one second epoch referring to an A phase subtype. The distribution of the classified activation events was 61%, 21%, and 18%, for “A1”, “A2”, and “A3”, respectively. It is notorious that the A3 subtype distribution is about 7% higher than what is estimated for the normal population, while the A2 subtype distribution is 7% lower [6]. These distributions can possibly be related to specific characteristics of the population that was used to create the dataset. Fig. 2 presents the variation of the A phase subtypes duration where it is possible to conclude that the A3 subtype is, on average, the longest with the highest variation while the A1 is the shortest with the lowest variation.

The spectral analysis of the A phase subtypes is presented in Fig. 3. The calculation was performed by first predicting the power spectrum of all A phase subtypes for each subject. Afterwards the results of all subjects were combined by considering the average of the power spectrum for each subtype. By evaluating the figure it is possible to verify that the A1 subtype is characterized by a prominent peak in the delta band. This peak is also present in the A2 subtype however, it has lower power and, in the A2 subtype, another significant peak occurs in the alpha band. The A3 subtype has a dominant peak in the same band that is followed by some power expressed in the beta band. These results are in agreement with the definition for each subtype [6] [8].

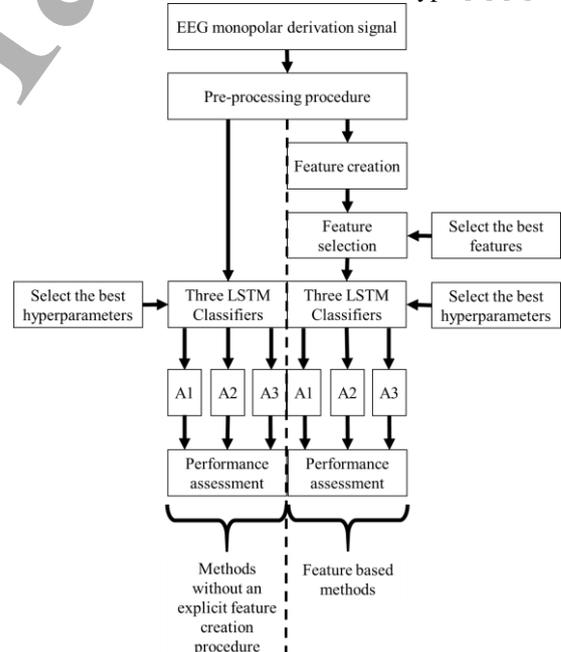


Fig. 1. Block diagram of the experimental procedure.

3.2 Pre-processing procedure

The database signals were recorded with a sampling frequency that varied between 100 Hz and 512 Hz. Therefore, a uniform database was produced by resampling all signals at

the lowest sampling frequency by decimation [39]. A constant factor of reduction for the sampling rate, s , was used and a standard lowpass filter (order eight Chebyshev type I filter, with a normalized cutoff frequency of $0.8/s$ and a passband ripple of 0.05 dB [40]) was employed to downsample the signal and avoid aliasing. Subsequently, the resampling process chooses each sth point from the filtered signal to create the resampled signal. Therefore, each one second epoch contained 100 sampling points.

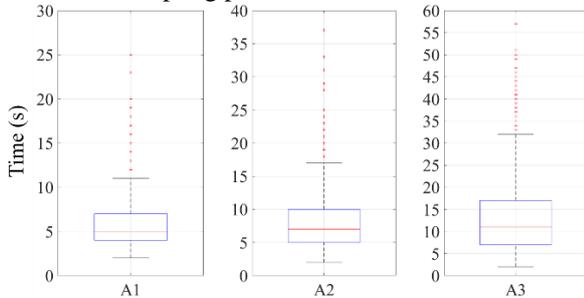


Fig. 2. Box plot of the variation of the A phase subtypes duration. Indication of the subtype in the abscissa and time (in second) in the ordinate.

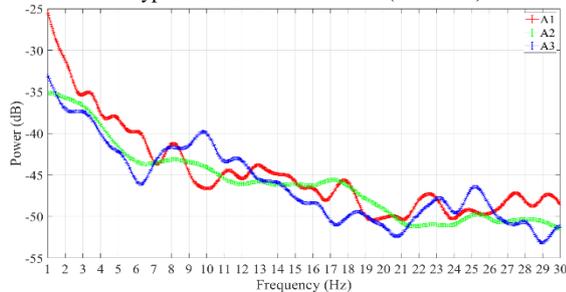


Fig. 3. Average curves and error bars (considering one standard error) of the A phase subtypes' power spectrum analysis. Frequency (in Hz) in the abscissa and power (in dB) in the ordinate. At 1 Hz the power of A1, A2 and A3 is -24 dB, -33 dB, and -35 dB, respectively.

The process of resampling at the lowest resolution (re-quantizing the signals if needed) allows the development of a device-independent estimation of the EEG signal, reducing the subject related variations of the dataset [41]. The resampled signals were subsequently standardized (subtract the mean and divide by the standard deviation) to nullify the effect of systematic variations of the data [42].

Several studies performed the removal of artifacts related to movements during sleep when studying the EEG signals [43]. However, these are frequently associated with the occurrence of a microstructure event, which in turn can be associated with the A phase subtypes that are intended to be classified [44]. Therefore, these events were kept. The eye movement and cardiac field artifacts were also recommended to be removed when studying the CAP [26]. Though these processes require both the electrocardiogram and electrooculogram signals that may not be available, for example, in home monitor devices. Therefore, this approach was not followed in this work to provide a method that is simple to apply in future hardware implementations, using as few sensors as possible (to be suitable for home monitoring devices which can be easily self-assembled). For these reasons, the developed method employs the signal from one EEG monopolar derivation exclusively.

3.3 Feature creation

Contrarily to the methods without an explicit feature creation procedure, the feature-based methods require the creation of a new step between the pre-processing procedure and the classification. This step is responsible for the creation of the feature, which is fed to the classifier.

The estimated features can be divided into three categories: evaluation of the amplitude variation; power in frequency bands; amplitude to frequency power ratio. Hence, the evaluated features allow to assess the A phase characteristic behaviors which present variations in both amplitude and frequency.

A segmentation analysis (based on symbolic dynamics) and an amplitude variation metric were evaluated for the first category. The segmentation method was proposed for this work, and performed a transformation of the signal into a sequence of symbols by considering multiple thresholds for the amplitude of the signal. A segmentation example is presented in Fig. 4. Each threshold was selected to be a multiple of the standard deviation, σ , thus, significantly simplifying the signal to a sequence of symbols (nine was found to be the most suitable number of thresholds using an experimental method described in the next section). The key element of such analysis is to identify the suitable multiplier, M , to apply. From the vector that contained the sequence of symbols, V_s , it is possible to evaluate several statistical features [45] [24]. However, to avoid redundant information (taking into consideration that frequency-based features and features focused on the relation between amplitude and frequency were employed in the other two categories), the vector analysis was based on the number of occurrences of each symbol. Such information can be especially relevant for the LSTM that evaluates the time-based variations over multiple time steps.

The last feature of this category was the amplitude variation metric, A_v given by

$$A_v = \max(E) - [\max(E-1) - \max(E-2)] \quad (1)$$

where \max is the maximal value and E is the epoch. This feature is proposed in this work. It takes into consideration the fact that the amplitude of the phasic activities that initiate an A phase must be 33.3% higher than the background voltage of the previous two seconds [13]. Therefore, this feature evaluates the maximum amplitude variation of the current epoch with respect to the variation of the previous two epochs (given by $\max(E-2) - \max(E-1)$), working as a marker for the activation phase transitions.

The Power Spectral Density (PSD) of each of the five characteristic EEG frequency bands (Delta, Theta, Alpha, Sigma, Beta) was considered as features that composed the second category. These features were previously identified as possibly the most relevant for A phase analysis [22]. The PSD of each frequency band was computed using the Welch's method [46], employing the Hanning window with an overlap of 50%.

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
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

succession of 010 was corrected to 000 and 101 to 111.

3.6 Performance assessment

The performance of the developed models was assessed by the standard performance metrics [52], specifically, the Accuracy (Acc), Sensitivity (Sen) and Specificity (Spe). The diagnostic capability was measured by the Area Under the receiver operating characteristic Curve (AUC) which refers to the probability of a classifier to rank a randomly selected positive instance higher than an arbitrarily selected negative instance [53]. The average value and the standard deviation of the results were presented for each metric. Each simulation was repeated fifty times to attain statistically significant results.

There is no detailed orientation for selecting the optimal hyperparameters for the classifier, such as the number of units in each cell or number of time steps [54]. Therefore, a grid search approach was followed in this work, testing multiple combinations. The model which attained the highest AUC (metric considered relevant to achieve a general measure of predictiveness [53]) was chosen as the best solution. The search procedure evaluated the type of layers, the number of hidden units of each layer, and the number of time steps of the LSTM.

The first layer could be either a LSTM or a Bidirectional LSTM (BLSTM), while the second layer was either a dense layer or another LSTM (cascade of LSTM layers) followed by a dense layer. The number of hidden units of the LSTM layer was increased in steps of 100, starting at 100 while the time steps were increased in steps of 10 (starting with 5 time steps). The iterative search was performed until the AUC decreased or saturated the increment. The number of neurons of the fully connected layer was chosen to be either half, the same or twice the number of hidden units of the previous layer while the number of hidden units of a second LSTM layer, of a cascade model, was chosen to be the same as the number of hidden units of the first LSTM layer.

The two-fold cross-validation scheme [28] was employed to perform the tests since it is a fast and reliable model, reducing the computational time required to perform the large number of simulations needed to complete the grid search. Eight subjects were randomly chosen to compose the training set and the remaining composed the testing set. Subject independent results were attained by only using the data from a subject either in the training or the testing set. The subjects that composed the sets were randomly changed in each simulation.

The Adam algorithm [55] was employed for the network's error optimization. Cost sensitive learning was used to mitigate the effect of class unbalance in the classification since it was found out that it consistently outperforms the sampling methods when the datasets have more than 10000 examples [56] and does not require to change the distribution of the data which could be problematic for a model that considers the

information from previous time steps.

Taking into consideration that each classifier is performing a one-versus-all evaluation (either A1 or not-A1, A2 or not-A2 and A3 or not-A3), an error matrix was computed to further evaluate if the misclassifications of the subtypes were occurring due to the manifestation of another subtype or if the model was confusing the background activity with an activation phase. For each subtype classification a total of five counting variables were created (one for each subtype that is not the one related to the current classification, one for the NREM periods, one for Wake periods, and one for REM periods). Each time a false positive was detected the model verified if an activation phase, related to another subtype, occurred. If that was the case, then the counting variable related to that subtype was increased. Otherwise, the background activity was confused with an A phase and the counting variable related to the current macrostructure label was increased. The same process occurred each time a false negative was detected. The counting variables were then divided by the total number of misclassifications.

3.7 Recurrence and spectral entropy evaluation

The similarity of the A phase subtypes among each other and with the background activity was evaluated by the Recurrence Quantification Analysis (RQA) metrics. In this work, these metrics were used only for characterization purposes and are composed of Recurrence Rate (RR), percent determinism (DET), maximal line length in the diagonal direction (D_{max}), Shannon entropy of the frequency distribution of the diagonal line lengths (ENT), and trend (TND) [57] [58]. Through the evaluation of these metrics, it is possible to characterize the behavior of the EEG signal (that has nonlinear temporal properties) during the activations. The goal of this investigation is to examine changes of the dynamics within EEG signal (event based analysis) and not quantify the dynamics of a whole EEG signal.

Several works have proposed values for the RQA parameters [59]. However, these usually considered time windows with a fixed length. Nevertheless, a unique methodology was followed in this work where each A phase subtype's length can change. Therefore, these parameters required to be properly chosen for each examined window. The recurrence analysis's delay parameter was estimated using the mutual average information function [60] and the first local minima were chosen from an array of multiple possible solutions [61]. The embedding dimension was chosen by finding the first local minimum of the Kennel et al.'s false-nearest neighbor algorithm [62], checking the number of false neighbors, in the phase-space, as a function of the number of embedding dimensions [61].

The recurrence threshold parameter is the most difficult to tune. It is required to find a value that does not produce too many or too few recurrences since information about the dynamics will be lost in both scenarios [63]. The second

possibility (too few) is conceivably a worse case since the recurrences will mainly appear due to variations caused by noise. However, the scenario with too many recurrences includes points into the neighborhood which are simple consecutive points on the trajectory thus, hiding the recurrence structure [63]. Several works have proposed methodologies to estimate the optimal threshold [57]. For physiological time-series, it was indicated that the threshold should be selected with the goal of keeping the percentage of recurrent points lower than 5%. However, this may not be feasible for noisy signals since noise could distort any existing structure [57]. In this work it was followed the recommendation of using five times the signal's standard deviation as reference for the threshold when the noise can be a significant factor [63] [64] [65].

Significance of the results attained by the RQA analysis was evacuated by surrogate testing through the iterative amplitude adjusted Fourier transform surrogates [66]. This algorithm preserves the amplitude distribution and the power spectrum of the original time series in the surrogate data. Therefore, the considered null hypothesis was: there is no difference between the average of an RQA metric estimated from the EEG signal or from the surrogate. The alternative hypothesis (selected when the p-value was less than 0.05) was: the RQA metric's average is higher than the surrogate RQA metric's average (we do not expect the original EEG signal to be less predictable than the surrogate data).

The Spectral Entropy (SE) was computed to measure the spectral power distribution. This metric quantifies the regularity of the power spectrum during a specific period of time where a low entropy indicates the presence of few spikes where energy is concentrated and a high entropy is likely to be found on a broader spectrum with multiple relevant frequencies [67].

4. Results

A total of three tests were performed. The first implemented a characterization analysis to evaluate the distinctive patterns in time and frequency domains of the A phase subtypes. The second and third tests assessed the performance of the methods without an explicit feature creation procedure and the feature-based methods, respectively.

4.1 Characterization analysis

The normalized (by the maximum value) average of the RQA metrics for each subtype is presented in Table III. By examining table, is it possible to verify that the RR increased from the A1 subtype to the A3 subtype, signifying an increase in the probability that a specific state will occur. However, RR was the metric with the lowest support for the alternative hypothesis thus, these results may not be conclusive. The opposite happens for the DET which attained the highest support for the alternative hypothesis, indicating that A1 subtype is more prone to periodic behaviors while the A3

subtype is more likely to be associated with chaotic processes. The D_{\max} is related to the divergence of the trajectory segments thus, it may suggest that A1 subtype presented the most divergent trajectories since it attained the lowest value for this metric (the support for the alternative hypothesis was excellent, advocating the significance of the results).

Nonetheless, it is important to bear in mind that the employed methodology considered the full duration of each individual activation for the RQA metrics calculation, providing a comparative analysis between the A phase subtypes. Since the average duration of each subtype is considerably different, hence, the estimated metrics will be effected by this factor. This effect is particularly relevant for D_{\max} as it indicates how long was the longest diagonal line segment (excluding the main diagonal line of identity). The higher ENT for A1 subtype suggest that it has a higher complexity with the wider distribution of diagonal line lengths. The support for the alternative hypothesis, regarding the ENT, was excellent for all subtypes. For the TND the A3 subtype presented a higher value, suggesting greater nonstationary dynamics. The support for the alternative hypothesis regarding the TND was excellent for all subtypes. Table III. Normalized (by the maximum value) average of the RQA metrics for each subtype. The employed notation presents the average value of the RQA metric followed by the number of subjects which support the alternative hypothesis (in the left) and the average p-value (in the right, separated by dash) between brackets.

Subtype	RQA metrics				
	RR	DET	D_{\max}	ENT	TND
A1	0.53 (12 – 0.05)	1.00 (15 – 0.00)	0.60 (15 – 0.01)	1.00 (15 – 0.00)	0.95 (11 – 0.03)
A2	0.59 (6 – 0.13)	0.80 (15 – 0.00)	0.69 (13 – 0.03)	0.86 (14 – 0.01)	0.87 (10 – 0.07)
A3	1.00 (1 – 0.27)	0.61 (15 – 0.00)	1.00 (14 – 0.01)	0.85 (14 – 0.01)	1.00 (15 – 0.01)

It was observed that the A1 subtype attained on average the highest support for the alternative hypothesis while the A2 subtype attained the lowest. These results can possibly be linked to the definitions of the subtypes where the A1 is associated with mild or minor polygraphic variations with high-voltage slow waves (synchronized EEG patterns) while the A3 is linked to a predominance of low-amplitude fast rhythms (desynchronized EEG patterns). The A2 subtype is in between the other two subtypes [6] [8]. However, it is important to keep in mind that the RQA metrics can change according to the chosen parameters. Nevertheless, it was verified that the RQA metrics are relatively stable over a broad range of values for the recurrence threshold [61].

The normalized (by the maximum value) average SE for the A1, A2, and A3 subtypes was 0.73, 0.85, and 1.00, respectively. It was verified that the A1 subtype has the lowest average SE, indicating a lower spread in the spectrum, while the A3 subtype usually has a low entropy at the beginning that quickly grows and becomes more stabilized, presenting the highest average SE thus, suggesting the presence of a broader

spectrum. Similarly to the recurrence analysis, the A2 subtype behavior is between the other two subtypes, having more variation in the SE than the A1 subtype but with a lower average than the A3 subtype.

The B phases (background activity that lasts between 2 and 60 seconds and is bounded by A phases) were also examined as a comparison point with the activations. The normalized (by the maximum value) average RQA metrics were 0.62, 0.88, 0.66, 0.92, and 0.92 for RR, DET, D_{max} , ENT, and TND, respectively. The support for the alternative hypothesis (with the average p-value between brackets) was 9 (0.07), 15 (0.00), 15 (0.00), 15 (0.00), and 15 (0.01), respectively. Only the RR attained poor support while the support for the remaining metrics was excellent. The normalized (by the maximum value) average SE was 0.80. It was verified that the RQA metrics and SE are usually in between the A1 and A2 subtypes. These results can possibly suggest that B phase events can have characteristics of the A2 and A3 subtypes, suggesting that the A1 subtype will be easier to identify than the A2 and A3 subtypes.

4.2 Classification without an explicit feature creation procedure

Several architectures were examined to determine the most suited for the A phase subtype classification without an explicit feature creation procedure (the pre-processed EEG signal fed the classifier). It was verified that using the LSTM followed by a dense layer, who's number of hidden units was half of the number of hidden units of the LSTM, attained the best relationship between performance and complexity (number of parameters) of the model. The effect of increasing the number of hidden units of the dense layer or using a BLSTM (instead of the LSTM) was not relevant (less than 1% increases in the AUC), while the cascade LSTM models reduced the average performance.

The variation of the AUC and Acc of the models according to the number of time steps is presented in Fig. 5. By evaluating the figure, it is possible to verify that A1 subtype classification was the least sensitive to the number of time steps, reaching the peak of the AUC using 25 time steps and an LSTM with 300 hidden units. On the other hand, the A2 subtype presented a significant variation according to the number of time steps and reached the peak in the AUC using 35 time steps and 100 hidden units. For the A3 subtype it is possible to verify that the LSTM can no longer effectively exploit the information provided by the time steps, attaining the best AUC using 100 hidden units. Therefore, the A3 subtype classification was found to be the most difficult to be performed. The learning curved of the best models for each subtype are presented in Fig. 6. It is possible to observe, in the linear tendency line, that the performance of the models could possibly be improved by using more data, although this improvement is not likely to be significant as the learning curves started to reach a saturation point.

The performance of the best models is summarized in Table IV. It was verified that the classifiers attained a similar accuracy for all subtypes' classification. However, the A1 examination attained a significantly better AUC, demonstrating a balanced performance (similar sensitivity and specificity) while the A3 evaluation presented the lowest AUC, expressing unbalanced results. The performance of the A2 subtype classification is in between the other two subtypes, reaching a higher AUC than the A3 examination but lower than the A1 classification. On the other hand, it was predicted that the specialist agreement in CAP classification, analyzing the same EEG signals, is in the 69% to 78% range [68]. The A1 and A2 subtypes classifications are within this range. However, the A3 subtype evaluation was lower, suggesting that further information should be provided to the classifier (this approach was followed in the next subsection by examining features).

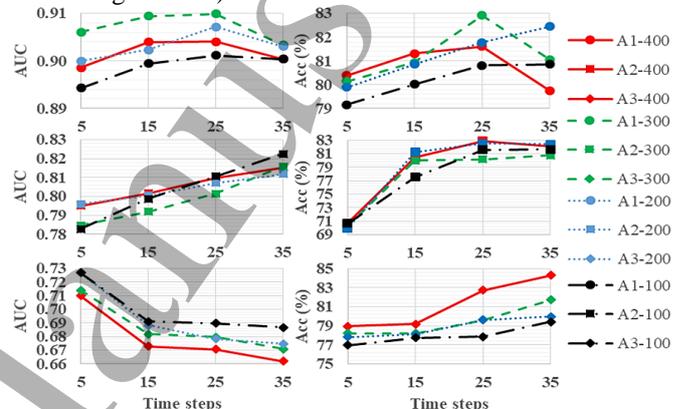


Fig. 5. Variation of the AUC and Acc for different time steps of the models without an explicit feature creation procedure. The activation subtype is presented in the legend followed by the number of hidden units of the LSTM. Number of time steps in the abscissa and variation of the performance metric in the ordinate.

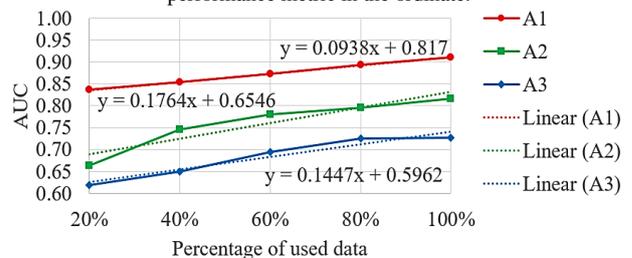


Fig. 6. Learning curves for the best models without an explicit feature creation procedure of each subtype. The linear tendency line and its equation are also presented. Percentage of used data in the abscissa and variation of the AUC in the ordinate.

The error matrix of the results is presented in Table V. By examining the table, it was verified that the A1 subtype was mostly misclassified with a background activity (typically during NREM). Almost all misclassification occurred at the end of the activation phase where the classifier predicted a longer duration (frequently by one or two epochs) of the activation. Thus, the main difficulty for the A1 subtype classification was the accurate detection of end of the activation boundary. This difficulty is common for all subtypes classifications.

Table IV. Performance of the methods without an explicit feature creation procedure for each subtype.

Subtype	Performance (mean \pm standard deviation)			
	Acc (%)	Sen (%)	Spe (%)	AUC
A1	82.92 \pm 3.65	83.11 \pm 7.22	82.90 \pm 4.15	0.910 \pm 0.015
A2	81.66 \pm 5.90	63.97 \pm 11.23	82.03 \pm 6.19	0.822 \pm 0.029
A3	76.97 \pm 7.09	56.03 \pm 8.63	77.69 \pm 7.58	0.727 \pm 0.025

Table V. Error matrix of the methods without an explicit feature creation procedure for each subtype.

	Miss-classified during an activation phase (%)			Miss-classified during background activity (%)		
	A1	A2	A3	NREM	REM	Wake
A1	-	2.11	1.91	88.85	1.69	5.44
A2	7.61	-	2.87	76.84	5.22	7.46
A3	3.84	2.30	-	49.01	14.72	30.13

For the A2 subtype, it was verified that some misclassifications occurred during an activation phase of another subtype, suggesting that the classifier was sometimes confusing the A2 subtype with other activations. This is likely related to the fact that the A2 subtype has characteristic traces of both A1 and A3 subtypes. It was also observed the increase of misclassifications related to REM or wake (after the sleep onset) periods. This tendency was more noticeable in the A3 subtype, suggesting that some of the patterns associated with the A2 and A3 subtypes are similar to the background activity of the brain during these periods. Hence, the identified difficulties for the A2 subtype classification are the boundary detection (of the activation's conclusion) and accurately discriminate between the other subtypes.

The increase of misclassifications related to the background activity periods was identified as the major issue associated with the A3 subtype classification. Almost all wake (after the sleep onset) periods and more than a third of the REM periods were misclassified as an A3 subtype, strongly advocating the need to use specific features to help discriminate the occurrence of the activation phase. It was also observed that some misclassifications occurred when the A1 or A2 subtypes presented desynchronized EEG patterns. Therefore, the A3 subtype detection was found to be the most challenging for the employed classifier and the identified difficulties are the boundary detection (of the activation's conclusion), differentiate the desynchronized EEG patterns of the A3 subtype, when comparing to the patterns of other activation phases, and correctly distinguish the presence of an A3 activation and the occurrence of background activity.

4.3 Feature-based assessment

The mRMR algorithm was used to rank the evaluated features and the results are presented in Fig. 7. Table II presented the specification of the feature associated with the identification number. The higher the rank, the less relevant the feature is for the specific subtype classification.

The number of thresholds considered for the symbolic dynamics was increased until they were considered the less relevant by the mRMR selection. It was verified that the limit was nine thresholds ($M=9$) as V_1 and V_9 (features with the

identification number 1 and 9, respectively) are already in the bottom as the less relevant for the A1 subtype. The optimal number of features was found by testing the 20 possible feature sets (the first set was composed of only the most relevant feature, the second by the two most relevant features, and so one until the last set with all features). The features composing the feature set which attained the highest AUC were selected for the feature-based model assessment.

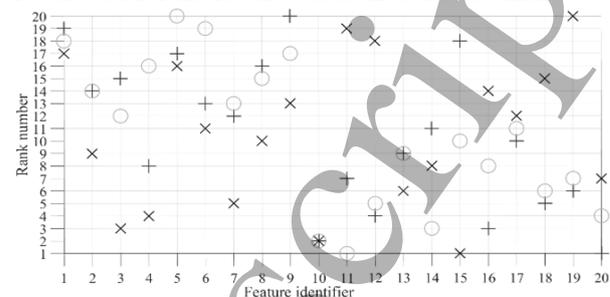


Fig. 7. Features ranked by the mRMR algorithm. A1 is indicated by '+', A2 is indicated by 'o', A3 is indicated by 'x'. Feature identification number (allusion to Table II) in the abscissa and ranking of the feature in the ordinate.

Through a grid search approach it was verified that the LSTM with 300 hidden units followed by a dense layer, whose number of hidden units was half of the number of hidden units of the LSTM, attained the best compromise between performance and complexity (number of parameters) thus, it was used for all classifiers based on features. The variation of the AUC and Acc of the models according to the number of features is presented in Fig. 8.

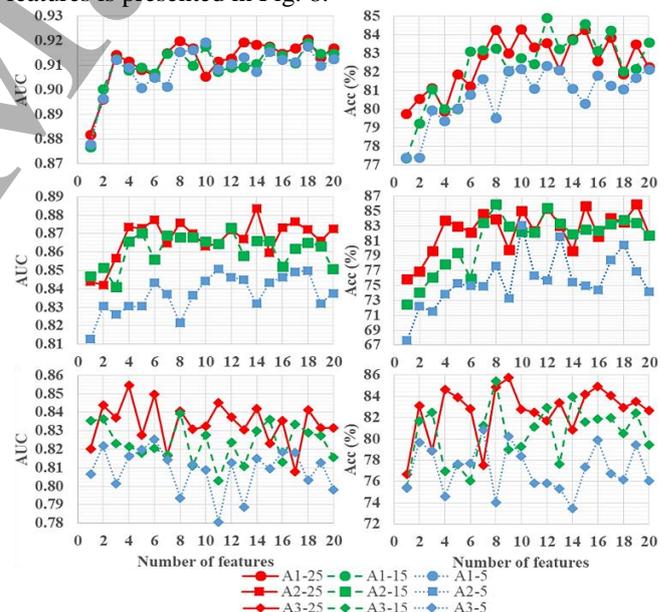


Fig. 8. Variation of the AUC and Acc, for different time steps, according to the number of features selected by mRMR. The activation subtype is presented in the legend followed by the number of time steps of the LSTM. Number of features in the abscissa and variation of the performance metric in the ordinate.

It was observed that 25 time steps provided the best AUC for all models, indicating that the classifiers were able to identify the temporal information provided by the time steps. The sequence of selected features, for A1 was: R_B ; A_V ; R_D ;

$PSD_T; R_A; R_S; PSD_D; V_4; PSD_A; R_T; PSD_S; V_7; V_6; V_2; V_3; V_8; V_5; PSD_B$. For A2 was: $PSD_D; A_1; PSD_S; R_B; PSD_T; R_A; R_S; R_D; PSD_A; PSD_B; R_T; V_3; V_7$. For A3 was: $PSD_B; A_v; V_3; V_4$.

The best models' performance is summarized in Table VI. From the AUC examination, the A1 subtype classification attained the best results, while the A3 subtype assessment is again the most challenging. The A2 subtype classification reported the lowest accuracy but the Sen and AUC are higher than the A3 subtype. By comparing the results of the feature-based models with the results of the methods without an explicit feature creation procedure (Table IV), it is possible to verify the significant increase in the AUC (around 18%) of the A3 subtype classification, attained by the employment of features. This result was due to the major increase in the sensitivity (about 25%) of the model, with an also substantial increase of the specificity (around 10%).

These results strongly advocate the feasibility of using features for the A3 subtype classification, suggesting that the model cannot extract all information when directly fed by the EEG signal. The AUC of the A2 subtype classification was also improved (around 7%) by using the feature-based models while the A1 subtype classification performance was similar for the feature-based models and the methods without an explicit feature creation procedure. It was also observed that the standard deviation of the performance was considerably lower for the feature-based models.

Table VII presents an error matrix created to further evaluate the performance of the feature-based models. When comparing with the methods without an explicit feature creation procedure (Table V) it is possible to observe that the A1 misclassification caused by wake (after the sleep onset) periods was reduced. However, the system still retains the issue related to the boundary detection at the end of the activation. For the A2 subtype, the identified issues for the methods without an explicit feature creation procedure were mitigated. The feature-based models reduced the misclassifications related to the occurrence of wake (after the sleep onset) periods.

The error matrix of the A3 subtype indicated a reduction of the misclassifications related to an A1 subtype's occurrence. The boundary issues were significantly reduced. However, the tradeoff was an increase in the misclassifications related to the occurrence of REM periods. These misclassifications can possibly be linked to the modulation of the REM sleep onset associated with the A2 and A3 subtypes [8] [6]. However, this issue can be significantly diminished by employing a method to remove the REM periods from the classification. This approach was followed by several works that perform the CAP analysis [30]. However, such methodology will require the use of either manual intervention, decreasing the usefulness of the model, or involve the use of a new classification procedure to classify the NREM and REM epochs. The automatic methodologies for sleep stage classification can be challenging since the specialist agreement, for sleep scoring, is probably lower than 90% [69].

Table VI. Performance of the features based methods for each subtype.

Subtype	Performance (mean \pm standard deviation)			
	Acc (%)	Sen (%)	Spe (%)	AUC
A1	81.91 \pm 2.43	87.83 \pm 3.94	81.60 \pm 2.75	0.921 \pm 0.011
A2	79.65 \pm 5.69	81.07 \pm 7.26	79.59 \pm 5.96	0.884 \pm 0.025
A3	84.61 \pm 5.31	70.37 \pm 8.41	85.07 \pm 5.72	0.855 \pm 0.034

Table VII. Error matrix of the features based methods for each subtype.

	Miss-classified during an activation phase (%)			Miss-classified during background activity (%)		
	A1	A2	A3	NREM	REM	Wake
A1	-	4.69	1.66	91.12	1.59	0.94
A2	7.75	-	3.03	79.92	6.26	3.04
A3	2.56	2.46	-	48.49	22.52	23.97

5. Discussion

It was observed that the A1 subtype has a low recurrence with energy concentrated in few frequencies. This is likely related to the predominance of high-voltage slow waves with the presence of few low-amplitude fast rhythms [6] [8]. The A3 subtype presented the highest SE average and recurrence, possibly due to the predominance of rapid activities with low-amplitude [6] [8]. The A2 subtype behavior, which can be seen as a combination of A1 and A3 subtypes behaviors, is likely related to the presence (without having the predominance) of low-amplitude fast rhythms that are mixed with slow high-voltage waves [6] [8].

A direct comparison between the attained results with the works presented in the state of the art is difficult to be performed since all other works used multiclass classification and then presented the results for the subtypes assessment (usually one versus all). A different approach was proposed in this work that consisted of training three classifiers (one for each subtype) that can be tuned with the goal of reaching the best performance for the subtype classification. Nevertheless, a first estimate of the results was attained by evaluating the models' overall capability to discriminate against the subtypes. This comparison is presented in Table VIII.

By evaluating Table VIII, it is possible to conclude that the proposed method attained, on average, the best results, while considering the most challenging methodology of not removing the EEG signal artifacts and the epochs not related to the NREM sleep. By not performing the NREM sleep isolation (removing the epochs not related to the NREM sleep hence, only performing the analysis during the NREM sleep) and doing one versus all classification with the other subtypes and the background activity, it increased the practical usefulness of the models and reduced the manual intervention.

The threshold based approaches proposed by Machado et al. [34] and by Navona et al. [33] reached significant results. Although these methods are computationally efficient and simple to be implemented, the potential of the tuned thresholds for generalization (application of the methods to new populations) can be problematic if there are significant variations between the populations used for training and the new population where the trained classification procedure will be applied. Mendez et al. [11] reported a higher sensitivity

than the proposed methods for the A1 and A3 subtypes detection but a significantly lower sensitivity for the A2 subtype classification. It is also relevant that the results reported were in the format subtype versus other two subtypes which will filter the non-CAP related events. Machado et al. [35], Hartmann and Baumert [26] and Arce-Santana et al. [37] reported a lower performance than the results reached in this work. Although Hartmann and Baumert [26] had reached a better sensitivity for the A3 subtype detection, the proposed method required the use of both electrocardiogram and electrooculogram signals in the preprocessing to remove the cardiac field and eye movement artifacts. Therefore, the process is more complex than the proposed method.

Table VIII. Performance of the methods proposed in the state of the art for the activation subtype classification, presenting the reference to the publication, a brief description of the employed method, the type of classification and the average of the performance metrics.

Publications*	Method	Classification	Average performance*		
			Acc (%)	Sen (%)	Spe (%)
[34]	Tuned thresholds applied to the TEO	A1 vs all	-	80	83
		A2 vs all	-	77	73
		A3 vs all	-	67	74
[33]	Tuned thresholds applied to the MMSD	A1 vs A2 and A3 combined	-	81	81
[11]	Features fed to k-NN	A1 vs A2 and A3 combined	-	90	-
		A2 vs A1 and A3 combined	-	43	-
		A3 vs A1 and A2 combined	-	80	-
		-	-	-	-
[35]	Features fed to SVM	A1 vs all	-	58	-
		A2 vs all	-	44	-
		A3 vs all	-	24	-
[26]	Features fed to LSTM	A1 vs all	-	63	-
		A2 vs all	-	42	-
		A3 vs all	-	71	-
[37]	Features fed to CNN	A1 vs A2 vs A3	77	-	-
		-	-	-	-
This work	Fed the pre-processed EEG to LSTM	A1 vs all	83	83	83
		A2 vs all	82	64	82
		A3 vs all	77	56	78
	Features fed to LSTM	A1 vs all	82	88	82
		A2 vs all	80	81	80
		A3 vs all	84	70	85

* From the method that reported the best results
 "all" specifies the signal with the B phases and non-CAP related events
 "Global" designates the overall performance

With the exception of Machado et al. [35], all machine learning based classifiers employed the NREM sleep isolation approach. However, as observed in the previous section, the practical implementation of these methodologies will require either manual intervention or a new classification procedure (to classify the NREM and REM epochs). It is also important to maintain the REM sleep epochs in the examination as some pathological conditions can generate CAP sequences in REM [35]. As a result, a method that can predict these occurrence

has the potential to be clinically significant to identify such conditions.

6. Conclusion

A new approach for the automatic A phase subtypes classification was proposed, performing a one versus all analysis for each subtype. The proposed method executes the examination by evaluating the signal from only one EEG monopolar derivation without manual manipulation of the signal, without removing artifacts, and without eliminating the epochs that are not related to the NREM sleep (the REM or wake periods).

The characterization analysis allows the assessment of how different the subtypes are from each other. It was verified that the A1 and A3 subtypes are the most dissimilar and that the A2 and A3 subtypes have some similarities with the B phases. The first observation can possibly be the reason why the A2 subtype presented the lowest accuracy (a tendency that was also observed in other works presented in the state of the art). The error matrix of both feature-based methods and the methods without an explicit feature creation procedure pointed out that the A2 and, with a high prevalence, the A3 subtypes were confused with the background activity during wake or REM periods, conceivably supporting the second observation.

It was observed that the A1 subtype classification performance was similar for both feature-based methods and the methods without an explicit feature creation procedure. However, for the A3 subtype classification the feature-based methods were significantly superior, indicating that the methods without an explicit feature creation procedure did not properly identify relevant patterns. Nevertheless, the A3 subtype classification was found to be the most challenging to be properly performed. This observation is in the same line as the findings reported by Mendez et al. [11] where it was predicted that the classification of the CAP phases could be affected by up to 25% of ambiguity and subjectivity, emphasizing the difficulties to perform an analysis of the CAP. As stressed in the scoring rules [6], the activations can present ambiguous limits due to inconsistent voltage changes and, if the EEG events do not clearly meet the characteristics of an activation phase then, they cannot be scored as part of an activation. It was verified that the highest variation in the activation duration was associated to the A3 subtype while the A1 subtype had the lowest variation. The longer variations of the A3 subtype can possibly be linked to an ill-defined phase offset detection. The misclassifications related to the offsets in the A3 subtype can possibly be associated to the longer average duration of this subtype as offsets present a smooth transition on the way to the basal sleep stage oscillations. Thus, the offset recognition can possibly be more ambiguous for a longer subtype.

When comparing the performance attained by the proposed methods with other works presented in the state of the art, it

was verified that the proposed method attained on average the best results, advocating the relevance of the proposal. This is likely due to the possibility of optimizing the classifier for the subtype detection. It is also relevant to notice that the specialist agreement, analyzing the same EEG signals, for CAP analysis, ranges from 69% to 78% [68] and the performance of the proposed models based on features is either equal or slightly superior to the upper bound, advocating the relevance of the developed mythology for clinical applications. The features proposed for this work can also be applied to other fields, providing an added value for the developed work. They have also provided an inside for the underlying physiological processes associated with the CAP phases. The future steps in this research are the evaluation of the developed models in a bigger dataset, employ sensor fusion to combine the information of multiple EEG channels [70] to attest if the developed model's performance can improve, and further examine the characteristic patterns of the subtypes with the goal of providing a deeper understanding of these events that can possibly lead to a reduction of the subjectivity in the CAP analysis.

Acknowledgements

Acknowledgment to the Portuguese Foundation for Science and Technology for their support through Projeto Estratégico LA 9 - UID/EEA/50009/2019. Acknowledgment to ARDITI - Agência Regional para o Desenvolvimento da Investigação, Tecnologia e Inovação under the scope of the Project M1420-09-5369-FSE-000001 - PhD Studentship. Acknowledgement to the Project MITIExcell co-financed by Regional Development European Funds, for the Operational Programme "Madeira 14-20" - EIXO PRIORITÁRIO 1, of Região Autónoma da Madeira, with no. M1420-01-0145-FEDER-000002.

References

- [1] Schulz H and Salzarulo P 2012 Forerunners of REM sleep *Sleep Medicine Reviews* **16** 95–108
- [2] Berry R, Brooks R, Gamaldo C, Harding S, Lloyd R, Marcus C and Vaughn B 2017 *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications* (Darien, Illinois, USA: American Academy of Sleep Medicine)
- [3] Mendonça F, Mostafa S, Ravelo-García A, Morgado-Dias F and Penzel T 2018 Devices for home detection of obstructive sleep apnea: A review *Sleep medicine reviews* **41** 149–160
- [4] Teplan M 2002 Fundamental of EEG Measurement *Measurement Science Review* **2** 1–11
- [5] Kubicki S and Herrmann W 1996 The Future of Computer-Assisted Investigation of the Polysomnogram: Sleep Microstructure *Journal of Clinical Neurophysiology* **13** 285–94
- [6] Terzano M and Parrino L 2000 Origin and Significance of the Cyclic Alternating Pattern (CAP) *Sleep Medicine Reviews* **4** 101–123
- [7] Terzano M, Parrino L, Smerieri A, Chervin R, Chokroverty S, Guilleminault C, Hirshkowitz M, Mahowald M, Moldofsky H, Rosa A, Thomas R and Walters A 2001 Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep *Sleep Medicine* **2** 537–553
- [8] Mariani S, Manfredini E, Rosso V, Mendez M, Bianchi A, Matteucci M, Terzano M, Cerutti S and Parrino L 2011 Characterization of A phases during the Cyclic Alternating Pattern of sleep *Clinical Neurophysiology* **122** 2016–2024
- [9] Mendez M, Chouvarda I, Alba A, Bianchi A, Grassi A, Arce-Santana E, Milioli G, Terzano M and Parrino L 2016 Analysis of A-phase transitions during the cyclic alternating pattern under normal sleep *Medical & Biological Engineering & Computing* **54** 133–148
- [10] Medina D, Mendez M, Chouvarda I and Murguía J 2019 Wavelet singularity analysis for CAP sleep delineation *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE) 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)* (Athens, Greece)
- [11] Mendez M, Alba A, Chouvarda I, Milioli G, Grassi A, Terzano M and Parrino L 2014 On separability of A-phases during the cyclic alternating pattern 36th Annual International Conference of the IEEE engineering in Medicine and Biology Society (Chicago, USA)
- [12] Terzano M, Parrino L, Boselli M, Smerieri A and Spaggiari M 2000 CAP components and EEG synchronization in the first 3 sleep cycles *Clinical Neurophysiology* **111** 283–90
- [13] Parrino L, Milioli G, Melpignano A and Trippi I 2016 The Cyclic Alternating Pattern and the Brain-Body-Coupling During Sleep *Epileptologie* **33** 150–160
- [14] Terzano M and Parrino L 1993 Clinical applications of cyclic alternating pattern *Physiology & Behavior* **54** 807–813
- [15] Korkmaz S, Bilecenoglu N, Aksu M and Yoldas T 2018 Cyclic Alternating Pattern in Obstructive Sleep Apnea Patients with versus without Excessive Sleepiness *Sleep Disorders* **2018** 1–7
- [16] Parrino L, Boselli M, Buccino P, Spaggiari M, Giovanni G and Terzano M 1996 The cyclic alternating pattern plays a gate-control on periodic limb movements during non-rapid eye movement sleep *Journal of Clinical Neurophysiology* **13** 314–23
- [17] Parrino L, Milioli G, Paolis F, Grassi A and Terzano M 2009 Paradoxical insomnia: the role of CAP and arousals in sleep misperception *Sleep Medicine* **10** 1139–45
- [18] Terzano M, Smerieri A, Felice A, Giglia F, Palomba V and Parrino L 2006 Cyclic alternating pattern (CAP) alterations in narcolepsy *Sleep Medicine* **7** 619–26
- [19] Parrino L, Paolis F, Milioli G, Gioi G, Grassi A, Riccardi S, Colizzi E and Terzano M 2012 Distinctive polysomnographic traits in nocturnal frontal lobe epilepsy *Epilepsia* **53** 1178–84
- [20] Parrino L, Ferri R, Bruni O and Terzano M 2012 Cyclic alternating pattern (CAP): the marker of sleep instability *Sleep Medicine Reviews* **16** 27–45
- [21] Mendez M, Murguía J, Alba A, Arce-Guevara V, González-Aguilar H and Guerrero-Mora G 2020 Power-law scaling behavior of A-phase events during sleep: Normal and pathologic conditions *Biomedical Signal Processing and Control* **57** 101757
- [22] Mendonça F, Fred A, Mostafa S, Morgado-Dias F and Ravelo-García A 2018 Automatic detection of cyclic alternating pattern *Neural Computing and Applications* 1–11
- [23] Mariani S, Manfredini E, Rosso V, Grassi A, Mendez M, Alba A, Matteucci M, Parrino L, Terzano M, Cerutti S and Bianchi A 2012 Efficient automatic classifiers for the detection of A phases of the cyclic alternating pattern in sleep *Medical & Biological Engineering & Computing* **50** 359–372
- [24] Niknazar H, Seifpour S, Mikaili M, Nasrabadi A and Banaraki A 2015 A Novel Method to Detect the A Phases of Cyclic Alternating Pattern (CAP) Using Similarity Index 23rd Iranian Conference on Electrical Engineering (Tehran, Iran)
- [25] Mendonça F, Mostafa S, Morgado-Dias F and Ravelo-García A 2019 A Portable Wireless Device for Cyclic Alternating Pattern Estimation from an EEG Monopolar Derivation *Entropy* **21** 1203
- [26] Hartmann S and Baumert M 2019 Automatic A-Phase Detection of Cyclic Alternating Patterns in Sleep Using Dynamic Temporal Information *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27** 1695–703
- [27] Steenkiste T, Groenendaal W, Deschrijver D and Dhaene T 2019 Automated Sleep Apnea Detection in Raw Respiratory Signals using Long Short-Term Memory Neural Networks *IEEE Journal of Biomedical and Health Informatics*
- [28] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Massachusetts: The MIT Press)

- [29] Yeh C and Shi W 2018 Identifying Phase-Amplitude Coupling in Cyclic Alternating Pattern using Masking Signals *Scientific Reports* **8** 2649
- [30] Mariani S, Grassi A, Mendez M, Milioli G, Parrino L, Terzano M and Bianchi A 2013 EEG segmentation for improving automatic CAP detection *Clinical Neurophysiology* **124** 1815–1823
- [31] Yeh C and Shi W 2018 Generalized multiscale Lempel–Ziv complexity of cyclic alternating pattern during sleep *Nonlinear Dynamics* **93** 1899–1910
- [32] Chouvarda I, Mendez M, Alba A, Bianchi A, Grassi A, Arce-Santana E, Rosso V, Terzano M and Parrino L 2012 Nonlinear Analysis of the Change Points between A and B phases during the Cyclic Alternating Pattern under Normal Sleep *34th Annual International Conference of the IEEE EMBS* 34th Annual International Conference of the IEEE EMBS (California, USA)
- [33] Navona C, Barcaro U, Bonanni E, Martino F, Maestri M and Murri L 2002 An automatic method for the recognition and classification of the A-phases of the cyclic alternating pattern *Clinical Neurophysiology* **113** 1826–1831
- [34] Machado F, Sales F, Bento C, Dourado A and Teixeira C 2015 Automatic identification of Cyclic Alternating Pattern (CAP) sequences based on the Teager Energy Operator 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Milan Italy)
- [35] Machado F, Sales F, Santos C, Dourado A and Teixeira C 2018 A knowledge discovery methodology from EEG data for cyclic alternating pattern detection *BioMedical Engineering OnLine* **17** 185
- [36] Machado F, Teixeira C, Santos C, Bento C, Sales F and Dourado A 2016 A-phases subtype detection using different classification methods IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (Florida, USA)
- [37] Arce-Santana E, Alba A, Mendez M and Arce-Guevara V 2020 A-phase classification using convolutional neural networks *Medical & Biological Engineering & Computing* **58** 1003–1014
- [38] Goldberger A, Amaral L, Glass L, Hausdorff M, Ivanov P, Mark R, Mietus J, Moody G, Peng C and Stanley H 2000 PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research *Circulation* **101** 215–220
- [39] Digital Signal Processing Committee I 1979 *Programs for Digital Signal Processing* (New York, USA: IEEE press)
- [40] Phillips C, Parr J and Riskin E 2013 *Signals, Systems, and Transforms* (London, UK: Pearson)
- [41] Chapotot F and Becq G 2010 Automated sleep–wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules *International Journal of Adaptive Control and Signal Processing* **24** 409–23
- [42] Muralidharan K 2010 A Note on Transformation, Standardization and Normalization *The IUP Journal of Operations Management* **9** 116–22
- [43] Urigüen J and Zapirain B 2015 EEG artifact removal – State-of-the-art and guidelines *Journal of Neural Engineering* **12** 031001
- [44] Ferri R, Manconi M, Aricò D, Sagrada C, Zucconi M, Bruni O, Oldani A and Ferini-Strambi L 2010 Acute dopamine-agonist treatment in restless legs syndrome: effects on sleep architecture and NREM sleep instability *Sleep* **33** 793–800
- [45] Niknazar H, Maghooli K and Nasrabadi A 2015 Epileptic Seizure Prediction using Statistical Behavior of Local Extrema and Fuzzy Logic System *International Journal of Computer Applications* **113** 24–30
- [46] Welch P 1967 The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms *IEEE Transactions on Audio and Electroacoustics* **15** 70–3
- [47] Peng H, Long F and Ding C 2005 Feature selection based on mutual information criteria of max-dependency, max-relevance, and minredundancy *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27** 1226–38
- [48] Mostafa S, Morgado-Dias F and Ravelo-García A 2018 Comparison of SFS and mRMR for oximetry feature selection in obstructive sleep apnea detection *Neural Computing and Applications* 1–21
- [49] Hochreiter S and Schmidhuber J 1997 Long Short-term Memory *Neural Computation* **9** 1735–80
- [50] Gers F, Schmidhuber J and Cummins F 2000 Learning to forget: Continual prediction with LSTM *Neural Computation* **12** 2451–71
- [51] Mendonça F, Mostafa S, Morgado-Dias F and Ravelo-García A 2018 Sleep quality estimation by cardiopulmonary coupling analysis *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26** 2233–9
- [52] Sackett D, Haynes R, Guyatt G and Tugwell P 1991 *Clinical Epidemiology: A Basic Science for Clinical Medicine* (Pennsylvania, USA: Lippincott Williams and Wilkins)
- [53] Fawcett T 2006 An introduction to ROC analysis *Pattern Recognition Letters* **27** 861–74
- [54] Le X, Ho H, Lee G and Jung S 2019 Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting *Water* **11** 1387
- [55] Kingma D and Ba J 2015 Adam: A Method for Stochastic Optimization 3rd International Conference on Learning Representations (California, USA)
- [56] Weiss G, McCarthy K and Zabar B 2007 Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs? 2007 *International Conference on Data Mining, DMIN 2007* 2007 International Conference on Data Mining, DMIN 2007 (Las Vegas, Nevada, USA)
- [57] Marwan N and Webber C 2015 Mathematical and Computational Foundations of Recurrence Quantifications *Recurrence Quantification Analysis - Theory and Best Practices* Understanding Complex Systems (Cham, Switzerland: Springer Nature) pp 3–43
- [58] Martín-González S, Navarro-Mesa J, Juliá-Serdá G, Ramírez-Ávila G and Ravelo-García A 2018 Improving the understanding of sleep apnea characterization using Recurrence Quantification Analysis by defining overall acceptable values for the dimensionality of the system, the delay, and the distance threshold *PLoS One* **13** e0194462
- [59] Becker K, Schneider G, Eder M, Ranft A, Kochs E, Zieglgansberger W and Dodt H 2010 Anaesthesia Monitoring by Recurrence Quantification Analysis of EEG Data *PLoS One* **5** e8876
- [60] Thiel M, Romano M, Read P and Kurths J 2004 Estimation of dynamical invariants without embedding by recurrence plots *Chaos* **14** 234–243
- [61] Wallot S 2017 Recurrence Quantification Analysis of Processes and Products of Discourse: A Tutorial in R *Discourse Processes* **54** 382–405
- [62] Kennel M, Brown R and Abarbanel H 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction *Physical Review A* **45** 3403–3411
- [63] Marwan N, Romano M, Thiel M and Kurths J 2007 Recurrence plots for the analysis of complex systems *Physics Reports* **438** 237–329
- [64] Thiel M, Romano M, Kurths J, Meucci R, Allaria E and Arecchi F 2002 Influence of observational noise on the recurrence quantification analysis *Physica D: Nonlinear Phenomena* **171** 138–52
- [65] Marwan N 2011 How to avoid potential pitfalls in recurrence plot based data analysis *International Journal of Bifurcation and Chaos* **21** 1003–17
- [66] Lancaster G, Iatsenko D, Pidde A, Ticcinelli V and Stefanovska A 2018 Surrogate data for hypothesis testing of physical systems *Physics Reports* **748** 1–60
- [67] Helakari H, Kananen J, Huotari N, Raitamaa L, Tuovinen T, Borchardt V, Rasila A, Raatikainen V, Starck T, Hautaniemi T, Myllylä T, Tervonen O, Rytty S, Keinänen T, Korhonen V, Kiviniemi V and Ansakorpi H 2019 Spectral entropy indicates electrophysiological and hemodynamic changes in drug-resistant epilepsy – A multimodal MREG study *NeuroImage: Clinical* **22** 101763
- [68] Rosa A, Alves G, Brito M, Lopes M and Tufik S 2006 Visual and automatic cyclic alternating pattern (CAP) scoring: inter-rater reliability study *Arquivos de Neuro-Psiquiatria* **64** 578–581
- [69] Penzel T and Conradt R 2000 Computer based sleep recording and analysis *Sleep Medicine Reviews* **4** 131–148
- [70] Mendonça F, Mostafa S, Morgado-Dias F and Ravelo-García A 2020 Cyclic alternating pattern estimation based on a probabilistic model over an EEG signal *Biomedical Signal Processing and Control* **62** 102063