A-phases subtype detection using different classification methods

Fátima Machado[1], César Teixeira[1], Clara Santos[2], Conceição Bento[3], Francisco Sales[3] and António Dourado[1], *Member IEEE*

***Abstract*— Cyclic alternating patterns (CAPs) occur during normal sleep, but higher CAP rates are associated with abnormal conditions, such as epilepsy. Efficient automatic classification of CAP A-phase sub-types would be of remarkable importance for the consideration of CAP as a disease bio-marker.**

This paper reports a multi-step methodology for the classification of A-phases subtypes. The methodology encompasses: feature extraction, feature ranking, and classification (Support Vector Machine (SVM), k-Nearest Neighbor (k-NN) and Discriminant Analysis (DA)).

The study was carried out on 30 subjects with nocturnal frontal lobe epilepsy. The best classifier is based on a SVM that achieved an accuracy of 71%. For each A-phase subtype, i.e. A1, A2, and A3, the sensitivities were 55%, 37% and 25%, respectively. The classifiers developed are an innovation compared to what is found on literature, because they are designed to detect all subtypes and achieved better performance values. However, the performance values still need to be improved to achieve a reliable classifier that would not need a human technician supervision.

I. INTRODUCTION

A CAP sequence is composed by a succession of CAP cycles each one composed by two elements: an A and a B-phase. The A-phase of CAP is a transient phenomena translated by an increase amplitude and/or frequency which is clearly distinguishable from background activity, i.e. the B-phase. This phase is related to a brain activation including cortical arousal, and for this reason, it is a potential trigger of somatomotor activities. Each phase could have a duration between 2 and 60 seconds. The mean duration, in young adults, of CAP sequences is two and half minutes, which contains, in average, six CAP cycles [1]. An A-phase can be composed by high-voltage slow waves which are manifestations of EEG synchrony and/or by low-amplitude fast rhythms that are evidence the EEG desynchrony. Depending on the percentage of each waves types present in the transition epoch, the A-phase is classified into three different subtypes: subtype A1 is composed by high-voltage slow waves; subtype A3 is characterized by rapid low voltage rhythms; subtype A2 has elements from subtypes A1 and A3 thus, it is composed by a mixture of fast and slow rhythms.

Different algorithms have been proposed in the last few years for automatic scoring of A-phases. Barcaro et al proposed a detection method for A-phases [2], [3], [4]. In [3] and [4] it is proposed a methodology to distinguish between the A1 subtype versus A2 and A3 subtypes together, because of the resemblance between A2 and A3. The principle of the methodology is the same in [3] and [4]: an amplitude feature is computed for each one of the conventional frequency bands (Delta, Theta, Alpha, Beta, and Gamma) and these features are compared against a threshold. When one of the five features cross the threshold an A-phase is detected. The type assigned to the A-phases detected (A1 or A2/A3) depends on which of our features crossed the threshold. The best reported results are the ones proposed with the methodology presented in [4], which have a correctness of 83.5% for A-phase detection and 73.7% if distinction between subtypes is required. Other approaches were reported to just identify A-phases independently of the subtype, for example: the usage of wavelets and genetic algorithms [5]; the consideration of different classifiers (Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Adaboost [6] and Artificial Neural Networks (ANN)) [7], among others.

The methodologies published so far are mainly focused only in A-phase detection and not in the discrimination between the subtypes. Each subtype has a different pathological importance, thus it is necessary that the algorithm differentiate them. In this paper we describe a multi-class classification methodology to discriminate among different A-phase subtypes.

II. DATABASE AND METHODS

A. Database

The study was carried out on 30 subjects with nocturnal frontal lobe epilepsy, 14 female and 16 male, aged between 14 and 67 years old (mean = 31.03 ±

This work was supported by the Liga Portuguesa Contra a Epilepsia.

[1]Centre for Informatics and Systems, Polo II, University of Coimbra, 3030-290 Coimbra, Portugal cteixei@dei.uc.pt, mariad@dei.uc.pt, dourado@dei.uc.pt

[2]Centro de Medicina do Sono do Hospital Geral de Coimbra claraarsantos@gmail.com

[3]UMES, Centro Hospitalar e Universitário de Coimbra. franciscosales@huc.min-saude.pt

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

11.64). The dataset is available on an online database, the *CAP Sleep Database* [8], includes several one-night polysomnographic recordings , and has been used in several studies [3]–[5], [7]. The recordings were acquired at the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy. The polysomnographic data includes at least three EEG channels (F3 or F4, C3 or C4 and O1 or O2, referred to A1 or A2), two EOG channels, three electromyographic EMG signals, respiration signals and the ECG. The macrostructure scoring was annotated according to the R&K rules, while CAP was detected in agreement with Terzano reference atlas [9].

B. Methodology

The implemented methodology was based on the processing of the whole duration of the EEG signal related to channel C4-A1, and encompasses the steps: filtering, feature extraction, feature pre-processing, feature selection, classification, post-processing and performance evaluation. In the next sections a closer look into the each one of the steps, will be presented.

C. Filtering

In this step the signal is filtered to obtain six signals, each one with a different frequency range using a third-order *Butterworth* filter. At first the signals are filtered to remove undesired frequencies which are normally associated with muscular and ocular artefacts. The clinical practice is to filter the signal between $1-35$ Hz, thus the same approach is applied in this paper. One of the six signals is the raw signal filtered between $1-35$ Hz, the others five are the signal filtered in the different conventional frequency bands (Delta, Theta, Alpha, Beta, and Gamma) mentioned before.

D. Feature Extraction

Each polysomnographic has a duration of a whole night of sleep (± 8 hours). For the majority of the features, with exception of the features obtained from the Discrete Time Short Time Fourier Transform (DTSTFT), a non-overlapping moving window of 1-s duration is applied to the signals obtained from filtering. To mention that empirical mode decomposition is computed for the entire signal and then the resulting time series is segmented in 1-s epochs, aiming to avoid the windowing effect. A window length of 3 s, with an overlap between windows of 2 s was considered for the features obtained from the DTSTFT. Each window contributes with a feature vector for the next processing steps. The different features computed for each signal are listed in Table I. Details on the features and its application to sleep and CAP scoring can be obtained by consulting the related references.

TABLE I: Computed EEG features

Feature		Acronym	Reference
Macro-Micro Structure Descriptor		MMSD	[2]–[4]
Teager Energy Operator		TEO	[10], [11]
Zero-Crossing		ZC	[12]
Lempel-Ziv Complexity		LZC	[13], [14]
DTSTFT	Freq. of max. energy	$freq_{max}$	[15]
	Freq. of mean energy	$freq_{mean}$	
	Area under spect. curve	$freq_{area}$	
Empirical Mode Decomposition		EMD	[16], [17]
Shannon entropy		ShEnt	[18]
Fractal Dimension		FD	[19], [18]
Variance		s_{σ}^2	[7]

E. Features Pre-processing

In first place outliers are removed. Samples four standard deviations away from the mean are considered as outliers, and their value is replaced by the median of the feature. Afterwards, features with exception of MMSD TEO and EMD, were smoothed using a moving average filter. These features are excluded from filtering because they detect changes in amplitude and frequency, and if the smoothing technique is applied important information might be lost. Finally the values of each feature are normalised to be between 0-1.

F. Feature selection

In this work we used the Minimum Redundancy Maximum Relevance (mRMR) algorithm. mRMR ranks the features based on two objectives: obtain the highest correlation between the selected features and class labels and reduce the redundancy between features. This technique is described in detail in [20].

G. Classification

The different classification methods are used in this work using the two feature selection/reduction methods aiming to see for which conditions a better performance is achieved.

1) Discriminant Analysis: Discriminant Analysis (DA) [21] considers a feature vector x of size $(1 \times d)$, and c classes ω_k ($k = 1, 2, ..., c$), to find the best discrimination function, g, by minimizing a quadratic error criterion. The discriminant functions can be linear, $g(x) = wx + w_0$ and in this case the classifier is named by linear discriminant analysis (LDA), or quadratic, $g(x) = w_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d} \sum_{j=1}^{d} x_i x_j w_{ij}$, and in this case is designated by quadratic discriminant analysis (QDA) where w is the weight vector and w_0 is the bias.

2) k-NN: k-NN is a non-parametric method, which means that there is no assumption about the underlying pattern distributions. Unlikely the DA the kNN does not find the best function to divide the space into regions. Instead, the training data is stored in a matrix containing the features and the class assigned to each point. To label a new point, x, it is compared with the k-closest training points. The class assigned is the prevalent class of the nearest k-points [22].

3) Support Vector Machines: In its native formulation Support Vector Machine (SVM) [23] finds a decision hyperplane that maximizes the separation between patterns belonging to two different classes. In this paper we consider cost-sensitive SVM classifiers with Gaussian kernels that by default have as free parameters the cost (C) and the spread of the Gaussian kernel (γ). Discrimination of A-phases is a multi-class problems, thus a one-against-all multi-class approach was considered [24].

H. Post-processing

The unique post-processing implemented was focused on the validation the A-Phase duration. The length of A-phases must be within the interval $[2, 60]$ seconds, then, the A-phases whose duration is less than two seconds are considered as background.

I. Performance evaluation

For the evaluation of the techniques developed the Leave-one-out cross-validation was used. The algorithm output for each patient is compared to the real CAP scoring by filling a confusion matrix.

Considering a given class i versus the others, four measurement can be taken: TP_i, the number of instances correctly classified as class i; FP_i is the number of instances classified as i when in fact they belong to other class; TN_i is the number of instances correctly not classified as i; and FN_i is the number of instances belonging to class i that are assigned to other classes. To evaluate the algorithm's performance sensitivity (SE_i) and accuracy (acc) were computed, which are given by:

$$\mathrm{SE}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i}; \quad acc = \frac{\mathrm{TN}_i + \mathrm{TP}_i}{\mathrm{TN}_i + \mathrm{FP}_i + \mathrm{FN}_i + \mathrm{TN}_i}$$

III. Results

In total, 55 features are computed, i.e. the features listed in Table I for different frequency bands. Using the mRMR method the features are ranked in descending order. An evaluation of the performance using different number of features to build a classifier is performed and the results will be shown in this section. A straight-forward selection is implemented, at start only the two best features are considered, then the three best features (adding the following best feature) until 55 features are used.

In general, the performance, concerning accuracy, monotonically increases with dimensionality and stabilizes between $30 - 40$ features. The higher value of accuracy (61%) is achieved for 30 features for a LDA classifier. For k-NN it was observed that the accuracies values do not differ much with the k value used. Considering a k-NN model with 30 features accuracy is 70%. Analyzing the results obtained with LDA and k-NN, it can be stated that the accuracy values stabilizes after 40 features for both classifiers. Thus, at least 40 features are needed to obtain the maximum possible accuracy.

Since the computational costs increase with the number of dimensions, the SVM models were build using only the first best 40 features. A grid search was performed for $C = 2^{-5}, ..., 2^{15}$ and $\gamma = 2^{-15}, ..., 2^{5}$. With this classifier a best accuracy of 71% is obtained for $C = 2^{-1}$ and $\gamma = 2^{-1}$.

A confusion matrix analysis shows that for SVM with $C = 2^{-1}$ and $\gamma = 2^{-1}$, B-phases are confused around 10% of the times with the A-phase subtypes. Approximately 21% of the A1 phases are confused with A2 but only 3.6% with A3. A2 is more confused with A1, although 20% of these phases are considered by the algorithm as B-phases. A3 is the one with lower sensitivity and the misclassification is quite high, around 30% are confused with A2 or B-phase and 15% with A1. The A3 and A1 are the most distinct types, thus it was expectable that they were less confused.

IV. Discussion

To the best of our knowledge, only two methods that discriminate between A-phases are found in literature, proposed by Barcaro et al ([3] and [4]). In these methods the subtypes A2 and A3 are joined and classified as one, A2&A3. Table II compare the best results obtained with our methodology and the results presented by Barcaro et al. Observing this table it can be seen that the sensitivities of A2&A3 are quite low, in the two versions proposed by Barcaro et al (below 10%). Although, the accuracy of this method is high. This is because of the high sensitivity in detecting the B-phases, which occupies 70% of all the data. Then, due to the high sensitivity of B-phases, the accuracy will be high, even if the sensitivity in detection A-phases is low. The methods proposed in this work for A-phases discrimination are built to detect the three A-phases subtypes, which is an innovative aspect of this paper. All the A-phases sensitivities are higher than other published works. The accuracies relative to the A-phases, in descending order, are: A1, A2, and A3 in all methods, which is interesting since the frequencies in A1 to A3 increase. Most of the A3 are arousals which are events with characteristics

TABLE II: Comparison of the best classifier models proposed in this work with others from the literature.

Method	paper	sensitivity (%)				acc (%)
		B	A1	A2	A3	
LDA	proposed	73	66	37	18	68
k-NN	proposed	81	59	31	16	70
SVM	proposed	76	58	44	24	71
MMSD	Barcaro et al [3]	89	51	6.5		79.4
	Barcaro et al [4]	79.3	57	10		71.9

of the awake sleep stage. Therefore, the model was trained with B-phases similar to A3 phases, which lead to a considerable number of A3 phases being wrongly classified as B-phase. The characteristic waves of the A1-subtype are K-complexes, which are only present on this phase. This fact explains the high performance results for A-phase detection.

V. Conclusion

In this paper we approached different classification methods. It was concluded that using the features ranked by mRMR leads to acceptable results on discriminating all the A-phases sub-types. It was shown that the SVM is the best classification method. Although, the models obtained are not yet enough to classify the A-phases without a supervision.

Future steps will encompass the consideration of the macro-structure classification. This information could be important to discriminate A3 phases, given that it can be confused with the awake state.

References

[1] L. Parrino, R. Ferri, O. Bruni, and M. Terzano, "Cyclic alternating pattern (cap): The marker of sleep instability," *Sleep Medicine Reviews*, vol. 16, pp. 27–45, 2012.

[2] U. Barcaro, C. Navona, S. Belloli, E. Bonanni, C. Gneri, and L. Murri, "A simple method for the quantitative description of sleep microstructure," *Electroencephalography and Clinical Neurophysiology*, vol. 106, p. 429432, 1998.

[3] C. Navona, U. Barcaro, E. Bonanni, F. Martino, M. Maestri, and L. Murri, "An automatic method for the recognition and classification of the a-phases of the cyclic alternating pattern," *Clinical Neurophysiology*, vol. 113, no. 11, pp. 1826 – 1831, 2002.

[4] B. U., E. Bonanni, M. Maestri, L. Murri, L. Parrino, and M. Terzano, "A general automatic method for the analysis of nrem sleep microstructure," *Sleep Medicine*, vol. 5, no. 6, pp. 567 – 576, 2004.

[5] R. Largo, C. Munteanu, and A. Rosa, "Cap event detection by wavelets and ga tuning," in *Intelligent Signal Processing, 2005 IEEE International Workshop on*, Sept 2005, pp. 44–48.

[6] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*. Springer, 1995, pp. 23–37.

[7] S. Mariani, E. Manfredini, V. Rosso, A. Grassi, M. Mendez, A. Alba, M. Matteucci, L. Parrino, M. Terzano, S. Cerutti, and A. Bianchi, "Efficient automatic classifiers for the detection of a phases of the cyclic alternating pattern in sleep," *Medical & Biological Engineering & Computing*, vol. 50, no. 4, pp. 359–372, 2012.

[8] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), [Online; accessed on 6-September-2015].

[9] M. Terzano, L. Hollberg, L. Parrino, A. Sherieri, and R. Chervin, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (cap) in human sleep," *Sleep Medicine*, vol. 2, p. 537553, 2001.

[10] E. Kvedalen, "Signal processing using the teager energy operator and other nonlinear operators," University of Oslo - Department of Informatics, Tech. Rep., May 2003.

[11] F. Duman, A. Erdamar, O. Erogul, Z. Telatar, and S. Yetkin, "Efficient sleep spindle detection algorithm with decision tree," *Expert Systems with Applications*, vol. 36, no. 6, pp. 9980 – 9985, 2009.

[12] M. Carrozzi, A. Accardo, and F. Bouquet, "Analysis of sleep-stage characteristics in full-term newborns by means of spectral and fractal parameters." *Sleep*, vol. 27, no. 7, pp. 1384–1393, 2004.

[13] D. Abásolo, S. Simons, R. Morgado da Silva, G. Tononi, and V. Vyazovskiy, "Lempel-ziv complexity of cortical activity during sleep and waking in rats," *Journal of Neurophysiology*, vol. 113, no. 7, pp. 2742–2752, 2015.

[14] A. Casali, O. Gosseries, M. Rosanova, M. Boly, S. Sarasso, K. Casali, S. Casarotto, M. Bruno, S. Laureys, G. Tononi, and M. Massimini, "A theoretically based index of consciousness independent of sensory processing and behavior," *Science Translational Medicine*, vol. 5, no. 198, pp. 198ra105–198ra105, 2013.

[15] R. Bracewell, *The Fourier Transform and Its Applications*, ser. Electrical engineering series. McGraw Hill, 2000.

[16] N. E. Huang, S. Long, and Z. Shen, "The mechanism for frequency downshift in nonlinear wave evolution," *Advances in applied mechanics*, vol. 32, pp. 59–117C, 1996.

[17] P. Diez, V. Mut, E. Laciar, A. Torres, and E. Avila, "Application of the empirical mode decomposition to the extraction of features from eeg signals for mental task classification," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, Sept 2009, pp. 2579–2582.

[18] I. Chouvarda, M. Mendez, V. Rosso, A. Bianchi, L. Parrino, A. Grassi, M. Terzano, N. Maglaveras, and S. Cerutti, "Predicting eeg complexity from sleep macro and microstructure," *Physiological Measurement*, vol. 32, no. 8, p. 1083, 2011.

[19] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.

[20] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.

[21] A. Izenman, "Linear discriminant analysis," in *Modern Multivariate Statistical Techniques*. Springer, 2008, pp. 237–280.

[22] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[23] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[24] K. Duan and S. Keerthi, "Which is the best multiclass svm method? an empirical study," in *Multiple Classifier Systems*. Springer, 2005, pp. 278–285.