Awake/Sleep Scoring Through Wavelet Analysis Associated to Decision Tree Algorithms

Thiago L. T. da Silveira
Masters' Student in Computer Science (PPGI) and
Undergraduate Student in Information Systems
Universidade Federal de Santa Maria (UFSM)
Santa Maria, Brazil
Email: thiago@inf.ufsm.br

Alice J. Kozakevicius[1] and Cesar R. Rodrigues[2]
[1]Department of Mathematics and LANA and
[2]Department of Electronics and Computing
Universidade Federal de Santa Maria (UFSM)
Santa Maria, Brazil

***Abstract*—Sleep deprivation is a public health problem which must be carefully examined and treated. Several studies have proposed automatic methods aiming to identify when a person falls asleep. The identification of the sleep state can help sleep experts to diagnose and prevent certain disorders such as apnea and insomnia. In the current work, the discrete wavelet transform is employed in order to analyze signals from a single electroencephalogram channel. Statistical features are then extracted from the wavelet coefficients, representing the characteristics of specific frequency ranges of the signal. Afterwards, these features are carried out to the classification procedure. Three classical decision tree algorithms are considered aiming to assess the extracted features' robustness. Results yielding more than 95% of accuracy are achieved in two of the three analyzed classifiers.**

Keywords*—Awake/Sleep stages scoring; discrete wavelet transform (DWT); electroencephalogram (EEG); decision tree algorithms.

I. Introduction

The comprehension of the human sleep can help doctors to diagnose and avert sleep-related disorders such as apnea, narcolepsy and insomnia [1]. Several researches, as [2], [3], have proposed automatic methods to distinguish when a subject is awake or sleeping based on simple wearable devices, reducing therefore health care costs [3]. Cole *et al.* [2] and Tilmanne *et al.* [4] have proposed an awake/sleep scoring method based on actigraph signals, respiration effort and accelerometer signals. In this context, an introductory study is presented here, investigating analysis and classification algorithms, which, in the future, will be embedded in integrated systems for processing electroencephalogram (EEG) signals. However, at the initial development phase, the execution of these algorithms is made in conventional desktop computers over single channel EEG signals from a public polysomnograph (PSG) database.

Since 1983, the distinction of sleep and awake stages through physiological signals is the goal of several researches. In van Luijtelaar and Coenen [5] study, which method's input are the non-invasive EEG and electromyogram (EMG) signals, 93.6% of accuracy is achieved in experiments with rats. Tilmanne *et al.* [4] have employed their methodology in an infant polysomnograph database, reaching around 86.9% of accuracy when considering a healthy group of infants. Cole *et al.* [2] and Karlen & Floreano [3] methods achieved, respectively, 91.9% and 90.4 $\pm$ 3.6% of accuracy also considering healthy subjects. In the current work, signals from healthy patients, aged between 25 and 34, of a public polysomnography database – Physionet Sleep EDF [Expanded] [6] – are considered.

Independently of the used polysomnograph or actigraph signals, in order to be analyzed, they must be divided into small time intervals called epochs. According to both standards for sleep classification - the Rechtschaffen's and Kales' (R&K) recommendations [7] and the American Academy of Sleep Medicine [8] - the length of these epochs is often 20 or 30 seconds. Here, 30-second epochs are adopted. Different techniques to analyze and to extract features of these epochs as well as different classification algorithms can be found in literature. The current study identifies the sleep and awake stages through wavelet analysis of signals from a single EEG channel; other EEG channels as well as electrooculogram (EOG), electrocardiogram (ECG) and EMG signals are discarded. Afterwards the wavelet analysis, three statistical features are applied to the coefficients, generating distinct magnitudes for awake and sleep stages. The robustness of the chosen features is assessed though three decision tree classifiers: random tree, reduced error pruning (REP) tree and classification and regression tree (CART) [9]. The main contribution of this study is to provide a methodology based on a single EEG channel and reduced feature set allied to the usage of low computational cost decision tree algorithms.

The remaining of this work is organized as follows: Section II presents the proposed methodology. The data description and preprocessing procedure are presented in Section II-A. Section II-B briefly describes the chosen discrete wavelet transform and also explains the feature extraction. The results are pointed out and discussed in Section III. Finally, the conclusions are drawn in Section IV.

II. Proposed methodology

The proposed methodology is described in detail in the following subsections.

A. Data description and preprocessing

The experimental data considered in this work was obtained from the Physionet Sleep EDF [Expanded] public database [6]. This database offers recordings from two EEG channel – the Fpz-Cz and Pz-Oz – sampled at 100Hz besides EOG and EMG signals of Caucasian volunteers. Here, only the Pz-Oz EEG channel is used. Signals from all healthy subjects without any sleep-related medication were selected in this study. The total quantity of recordings is 106,376 30-second epochs (about 886.5 hours). In order to balance the data, a subsample of the complete set of epochs was obtained. This subset contains 34,023 epochs in the awake stage and the same quantity of epochs in the sleep stage (which comprises the different patterns of sleep stages S1, S2, S3 and S4 besides rapid eye movement (REM) stage according to the R&K recommendations). Expert's scores, i.e. hypnogram annotations, for each 30-second time segment are jointly provided in the database. These annotations are considered as a correct reference and are used to train and test the decision tree algorithms.

B. *Discrete wavelet transform and feature extraction*

Over each epoch, a four-level Daubechies 2 (Db2) discrete wavelet transform (DWT) [10] is applied in order to separate the signal components approximately into specific cerebral rhythms. The wavelet transform has good representativity in both time and scale domains [10] unlike Fourier transform which temporal mark is easily lost. Furthermore, the wavelets can efficiently analyze non-stationary signals – as the EEG ones – and allow to extract different statistical properties of them [11].

Since the signals of interest were sampled at 100Hz, it is possible to decompose them, using a Db2 wavelet, as follows: the first wavelet coefficient set contains the signal's information relative to 25-50Hz which is close to the low-gamma rhythm (30-50Hz). The second wavelet coefficient set has the 12.5-25Hz frequency range which contains relevant information about the beta rhythm (13-30Hz). The third and fourth wavelet coefficient sets have, respectively, informations about data in 6.25-12.5Hz and 3.125-6.25Hz. The main associated rhythms to these sets are, respectively, the alpha (4-13Hz) and theta (0.5-4Hz). Fig. 1 illustrates the 4-level wavelet decomposition scheme and the main rhythm associated with each wavelet coefficient set.

The knowledge of the signal's behavior in each one of these specific bands, thanks to the wavelet analysis, allows to extract particular characteristics of it, assisting the chosen classifier to accurately recognize the awake and sleep stages. The low-gamma rhythm is associated to activities in the awake stage such as attention [12]. Beta rhythm occurs more intensely in the awake stage [13]. In the transitions from the wakefulness to resting conditions, alpha gradually decreases and theta increases [11].

In order to emphasize these differences, besides reducing the data dimensionality, the standard deviation, skewness and kurtosis are extracted from each one of the wavelet coefficient sets of interest. Statistical measurements have already been successfully used to analyze physiological signals as performed in [14], [15] and commented in [16]. Fig. 2 shows the differences in magnitude of the selected features in each decomposition level. It is possible to notice the sensitivity of specific features per level around the moments in which the subject wakes up. For better visualization, skewness and kurtosis were multiplied by a factor 2.

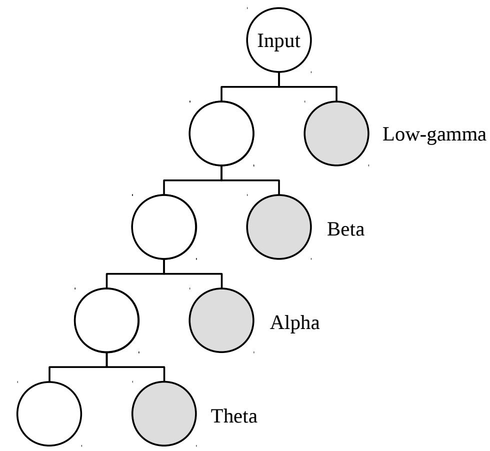

Fig. 1. Four-level wavelet decomposition scheme. The tree's root present the input epoch sampled at 100Hz. The remainder nodes are the scale (in white) and wavelet (in gray) coefficient sets. Main cerebral rhythm related to each wavelet coefficient set is shown on its right side.

C. *Decision tree algorithms*

A decision tree is a computational tree whose internal nodes are tests made on input patterns (the features) and the leaf nodes are classes (in this case, awake or sleep) [17]. There are several algorithms based on decision trees as the random trees, REP trees, CART, J48 tree, ID3 tree and C.45 tree. They can differ in relation to the attribute type (numerical and/or categorical), number of child nodes (binary tree or not), rules for growing themselves, etc. In this study, we focus in the three first tree decision algorithms. The Weka [18] data mining tool is used for the classification task based on the $F = 12$ (standard deviation, kurtosis and skewness for four wavelet coefficient sets) extracted features.

The random trees select randomly $K = \log_2(F) + 1$ features to grow themselves and do not apply pruning techniques. By other hand, the REP trees grow by computing the information gain estimator [1] for each attribute. Furthermore, REP trees apply a reduced error pruning technique. CART grow by estimating the Gini impurity [19] besides applying a pruning technique that prizes at minimal complexity.

All these tree algorithms are grown by considering a relevant to the problem training set. The complete balanced feature dataset is then split randomly into training and testing sets. Fig. 3 illustrates the features' dataset organization. The training set is used to grown the tree algorithms, whilst the testing one is used to evaluate them. Approximately a half of the features in the training and testing sets refer to the awake stage whilst the other half is relative to the sleep stage according to the experts' annotation.

III. Results and discussion

Several metrics can be used to evaluate a classification method, as the accuracy, precision, recall (sensitivity), and Cohen's kappa coefficient [20]. These metrics can be computed through analysis of the confusion matrix. The confusion matrix relates the accepted as truth data (sleep experts' scoring) versus the predicted data (the classifier output) and gives an overview of all hits and misses by class.

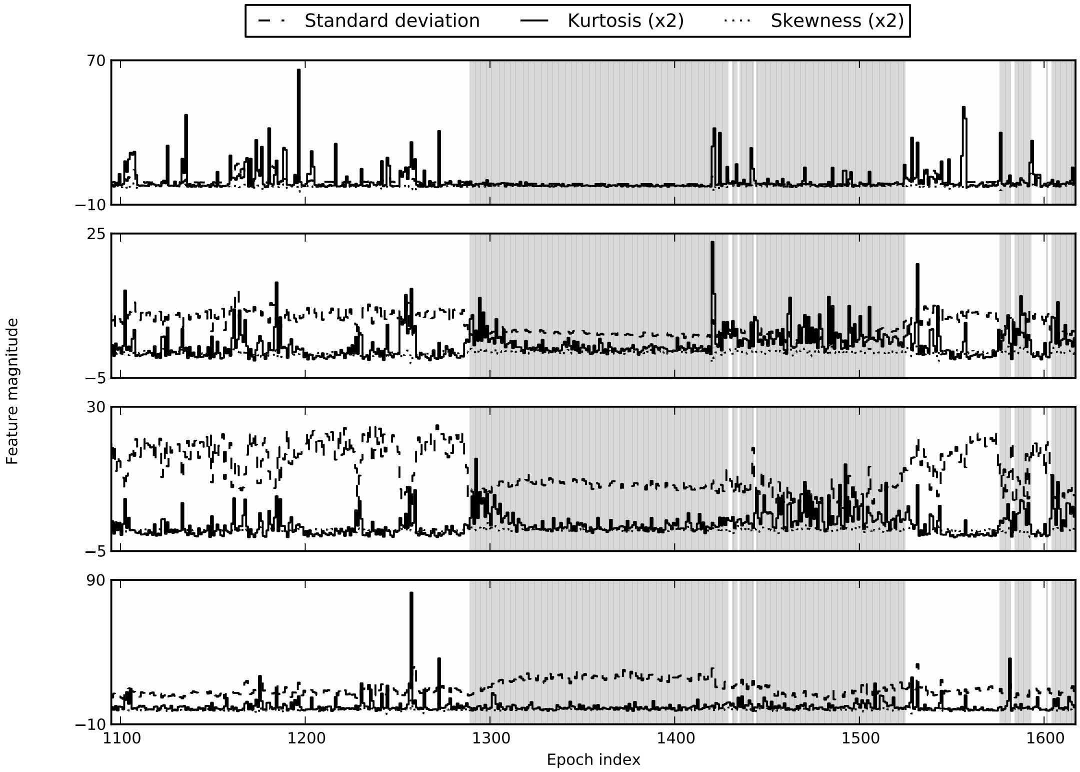

Fig. 2. Standard deviation, kurtosis (multiplied by 2) and skewness (multiplied by 2) for several 30-second epochs of the first night recording of the subject labeled as 05 according to the Physionet. The areas where the background is white are relative to the awake stage whilst those where the background is gray are relative to the sleep stage. These markings are according to the sleep experts.

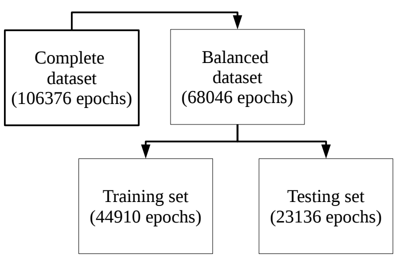

Fig. 3. Data preprocessing and separation in training and testing sets.

Let TP, TN, FP and FN be the quantity of true positives, true negatives, false positives and false negatives. Thus, it is possible to calculate the method's accuracy (ACC) through

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

Similarly, the precision (PRE) and recall (RCL) for each class are given, respectively, by

$$PRE = \frac{TP}{TP + FP} \text{ and} \quad (2)$$

$$RCL = \frac{TP}{TP + FN}. \quad (3)$$

The Cohen's kappa coefficient (κ) can be calculate through

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e}, \quad (4)$$

where π_0 is an observational probability of agreement and π_e is a expected probability by chance [20]. The values of π_0 and π_e are computed respectively through [1]

$$\pi_0 = \frac{\sum_{i=1}^{N} M_{ii}}{\sum_{i=1}^{N}\sum_{j=1}^{N} M_{ij}} \text{ and} \quad (5)$$

$$\pi_e = \frac{\sum_{i=1}^{N}\left(\sum_{j=1}^{N} M_{ij} \sum_{j=1}^{N} M_{ji}\right)}{\left(\sum_{i=1}^{N}\sum_{j=1}^{N} M_{ij}\right)^2}, \quad (6)$$

where $N = 2$ is the quantity of classes of the problem and M_{ij} is the $\{i, j\}$-th value of the confusion matrix.

Tables I, II and III present, respectively, the confusion matrix for the random tree, REP tree and CART algorithms. Its is possible to notice the high performance achieved by all the methods through analysis of the precision and recall values (optimal classifier has 100% of precision and recall).

TABLE I. CONFUSION MATRIX FOR RANDOM TREE

		Physionet's annotation	
		Awake	Sleep
Classifier's score	Awake	**10871**	684
	Sleep	688	**10893**
Precision (%)		94.1	94.1
Recall (%)		94.0	94.1

Accuracies reached 94.07%, 95.79% and 95.72% and Cohen's kappa coefficient, 0.88, 0.92 and 0.91, respectively, for

TABLE II. CONFUSION MATRIX FOR REP TREE

		Physionet's annotation Awake	Physionet's annotation Sleep
Classifier's score	Awake	**11147**	563
	Sleep	412	**11014**
Precision (%)		95.2	96.4
Recall (%)		96.4	95.1

TABLE III. CONFUSION MATRIX FOR CART

		Physionet's annotation Awake	Physionet's annotation Sleep
Classifier's score	Awake	**11126**	558
	Sleep	433	**11019**
Precision (%)		95.2	96.2
Recall (%)		96.3	95.2

the random tree, REP tree and CART algorithms. The average accuracy and Cohen's kappa are also computed after a 10-fold cross-validation [19]. In this case, the complete balanced feature dataset (containing 68,046 30-second epochs) is considered. Results achieved, respectively for random tree, REP tree and CART algorithms, 94.39%, 95.63% and 95.64% of accuracy and 0.89, 0.91 and 0.91 of Cohen's kappa coefficient. According to Landis and Koch [20], methods which kappa coefficient is greater than 0.80 are considered excellent.

In comparison with a recent study, Sano and Picard [21], the current study has comparable or even higher accuracy in relation to the considered EEG (83%), wrist (73%) and combined EEG and wrist feature sets (95%). These authors have assessed their feature sets in k-nearest neighbor (with k=1–4) and support vector machines under a 10-fold cross-validation. Additionally, the results presented in the current study also outperform those reported in Section I.

IV. CONCLUSION

This study performs the awake and sleep stages separation through analysis of a single EEG channel, which in real applications can be acquired by a portable grade consumer EEG headset. Furthermore, the four-level discrete wavelet transform of Daubechies family with two vanish moments (Db2) is considered as signal analyzer and feature extractor. Three statistical measures – the standard deviation, kurtosis and skewness – are computed for all the wavelet coefficient sets. In order to test the selected feature set, three decision tree algorithms are evaluated with the same training and testing sets.

The methodology is assessed through precision, recall, accuracy and Cohen's kappa coefficient measurements. The achieved results suggest that (i) the feature set is robust and (ii) simple decision tree techniques with or without pruning can provide good results in terms of sleep and awake identification.

Future works include the investigation of other features in the time and scale domains besides the test of the methodology presented in this study in a real scenario with a consumer grade EEG headset.

ACKNOWLEDGMENT

The first author would like to thank CAPES for his Master's Scholarship. The second and third authors would like to thank FAPERGS, grant PG n.1873-25.51/13-0.

REFERENCES

[1] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time-frequency analysis of a single eeg channel and random forest classifier," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 10–19, 2012.

[2] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity," *Sleep*, vol. 15, no. 5, 1992.

[3] W. Karlen and D. Floreano, "Adaptive sleep-wake discrimination for wearable devices," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 4, 2011.

[4] J. Tilmanne, J. Urbain, M. V. Kothare, A. V. Wouwer, and S. V. Kothare, "Algorithms for sleep-wake identification using actigraphy: a comparative study and new results," *Journal of Sleep Research*, vol. 18, no. 1, pp. 85–98, 2009.

[5] E. L. J. M. van Luijtelaar and A. M. L. Coenen, "An eeg averaging technique for automated sleep-wake stage identification in the rat," *Physiology and Behavior*, vol. 33, pp. 837–841, 1983.

[6] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 47, pp. 1185–1194, 2000.

[7] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Electroencephalography and Clinical Neurophysiology*, vol. 26, no. 6, p. 644, 1969.

[8] C. Iber, S. Ancoli-Israel, A. L. C. Jr, and S. F. Quan, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, American Academy of Sleep Medicine, 2007.

[9] I. H. Written, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Morgan Kaufmann/Elsevier, 2011.

[10] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992.

[11] A. Subasi, "Automatic recognition of alertness level from eeg by using neural network and wavelet coefficients," *Experts Systems with Applications*, vol. 28, no. 4, pp. 701–711, 2005.

[12] J. X. and K. A., "Gamma rhythms in the brain," 2011.

[13] M. Corsi-Cabrera, Z. Muoz-Torres, Y. del Ro-Portilla, and M. Guevara, "Power and coherent oscillations distinguish REM sleep, stage 1 and wakefulness," *International Journal of Psychophysiology*, vol. 60, no. 1, pp. 59 – 66, 2006.

[14] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel eeg signal," *Computers in Biology and Medicine*, vol. 42, no. 12, pp. 1186–1195, 2012.

[15] M. Asadzadeh, E. Hashemi, and A. Kozakevicius, "Efficiency of combined daubechies and statistical parameters applied in mammography," *Applied and Computational Mathematics: An International Journal*, vol. 12, no. 3, pp. 289–305, 2013.

[16] M. Ronzhina, O. Janoušek, J. Kolářová, P. Honzik, and I. Provaznik, "Sleep scoring using artificial neural networks," *Sleep Medicine Reviews*, vol. 16, pp. 251–263, 2012.

[17] N. J. Nilsson, *Introduction to Machine Learning*, draft ed., 1998.

[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[19] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, 2010.

[20] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–174, 1977.

[21] A. Sano and R. Picard, "Comparison of sleep-wake classification using electroencephalogram and wrist-worn multi-modal sensor data," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 930–933.