

# Sleep Stages Classification Using Spectral Based Statistical Moments as Features

Classificação de Estágios do Sono Utilizando Como Atributos os Momentos Estatísticos Calculados Com Base Em Componentes Espectrais

Eduardo T. Braun<sup>1\*</sup>, Thiago L. T. da Silveira<sup>2</sup>, Alice J. Kozakevicius<sup>3</sup>, Cesar R. Rodrigues<sup>1</sup> and Giovani Baratto<sup>1</sup>

**Abstract:** In the pursuit of portable, efficient and effective sleep staging systems, researchers have been testing a massive number of combinations of EEG features and classifiers. State of the art sleep classification ensembles achieve accuracy in the order of 90%. However, there is presently no consensus regarding the best set of features for identifying sleep stages with a single EEG channel, leading researchers to modify the feature selection according to the number of classification stages. This paper introduces a reduced set of frequency-domain features capable of yielding high classification accuracy (90.9%, 91.8%, 92.4%, 94.3% and 97.1%) for all 6- to 2-state sleep stages. The proposed system uses fast Fourier transform (FFT) to convert data from Pz-Oz EEG channel into the frequency domain. Afterwards, eight statistical features are extracted from specific frequency ranges associated to brain rhythms, feeding a random forest classifier.

**Keywords:** sleep stage classification — random forest — FFT — statistical moments

**Resumo:** Na busca por sistemas portáteis, eficientes e efetivos para classificação de sono, pesquisadores têm testado um grande número de combinações de atributos de sinais EEG e classificadores. O estado da arte dos aplicativos de classificação para estágios de sono atinge taxas de acerto na ordem de 90%. Contudo, não existe um consenso em relação ao melhor grupo de atributos para identificar estágios de sono em sinais EEG de um único canal, levando pesquisadores a modificar a seleção dos atributos de acordo com o número de estágios a serem classificados. Este trabalho apresenta um conjunto reduzido de atributos, capaz de produzir uma classificação com taxas elevadas de acerto (90.9%, 91.8%, 92.4%, 94.3% and 97.1%) para todos os agrupamentos de estágios de sono (de 6 a 2 estágios). O sistema proposto utiliza a transformada rápida de Fourier (FFT) para converter os dados do canal de EEG Pz-Oz para o domínio da frequência. Posteriormente, oito atributos estatísticos são extraídos de faixas de frequência específicas associadas a ritmos cerebrais, alimentando um classificador do tipo floresta aleatória.

**Palavras-Chave:** classificação de estágios de sono — floresta aleatória — FFT — momentos estatísticos

<sup>1</sup> Department of Electronics and Computing, Federal University of Santa Maria, Brazil

<sup>2</sup> Post-Graduate Program in Computing, Federal University of Rio Grande do Sul, Brazil

<sup>3</sup> Department of Mathematics, Federal University of Santa Maria, Brazil

\*Corresponding author: [eduardotbraun@gmail.com](mailto:eduardotbraun@gmail.com)

DOI: <http://dx.doi.org/10.22456/2175-2745.74030> • Received: 06/07/2017 • Accepted: 18/12/2017

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

## 1. Introduction

Monitoring the behavior of human sleep through polysomnography (PSG) allows experts to diagnose important sleep related disorders such as apnea, insomnia and narcolepsy [1]. This monitoring task is currently performed by visually inspecting polysomnographic recordings such as electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG) and electroencephalogram (EEG), in which characteristic waveform patterns to each sleep stage are associated [2].

Among non-invasive techniques, EEG is considered one of the most reliable source for sleep staging [3], supporting an increasing number of works [4, 5, 6, 7]. Their main attempts are both to enhance classification performance and reducing PSG acquisition channels, seeking highly accurate and efficient portable sleep classification systems.

According to the Rechtschaffen and Kales' (R&K) recommendations [8], short time EEG intervals, called epochs, can be classified as awake (W) or one of the following five possible sleep stages, labeled as S1, S2, S3, S4 and rapid

eye movement (REM). Another recognized guideline for the human sleep classification was proposed by the American Academy of Sleep Medicine (AASM) [2], defining instead 5 stages of sleep, in which S3 and S4 are merged into the slow wave sleep (SWS) state [9]. When joining S1 and S2 into a single S1/S2 stage, a group of 4 states is obtained: W, S1/S2, SWS and REM. The 3-state sleep stage classification considers S1/S2 and SWS as a unique state, namely non-rapid eye movement (NREM). Finally, when REM and NREM are grouped together, classification is performed into 2 stages: W and sleep (SLP) [7].

Both standards, R&K and AASM, address directives based on certain waves and events (as well as their durations) within the recorded signal. According to Moser *et al.* [9], the sleep stages are equivalent in both standards, which allows a comparison between the resulting R&K and AASM classifications.

Independently of the chosen classification standard, the sleep scoring procedure is a time consuming process [10], involving the analysis of a huge amount of data related to whole night recordings. According to Ronzhina *et al.* [11], 8 hours of recorded signal demand from 2 to 4 hours of visual analysis. Since the introduction of the R&K standard, a variety of computational methodologies has been proposed [4, 12, 13, 14, 11, 7] with the objective of supporting automatic identification of sleep stages. Such methods are usually based on: EEG preprocessing, feature extraction and classification.

In the current work, as in [4, 15, 16], the EEG preprocessing is performed through the application of the fast Fourier transform (FFT) [17] on only one channel of the many possibilities of a PSG, namely the Pz-Oz channel. Thus, as one relevant contribution, the proposed sleep staging method relies on the selection of a reduced and fixed set of spectral-based features, which here are the first, third, and fourth statistical moments (mean, skewness and kurtosis, respectively) from Fourier coefficients corresponding to sleep-related frequencies of the brain (delta, theta, alpha, sigma, beta, and gamma). Unlike here, in other works [4, 11, 7], different feature sets are considered depending on the number of stages to be classified. The classification procedure itself is performed by a random forest, which is considered in [12, 13] also in the context of brain pattern classification.

The remainder of this paper is organized as follows. A concise review of the literature, presenting some basic concepts of this field and addressing state of the art works, is given in Section 2. The proposed methodology is detailed in Section 3. In Section 4, the achieved results are presented, discussed and compared to those of well established studies. Finally, some conclusions are drawn in Section 5.

## 2. Related Work

R&K and AASM standards give directives to classify the sleep mainly based on the form and duration of the waves in EEG signals, i.e. the brain rhythms. Table 1 presents those rhythms that are known to be related to sleep events [8]. The

delta rhythm (0.5–4 Hz) appears more intensely during the S3 and S4 stages, presenting 20–50% and more than 50%, respectively, of the spectrum energy [11]. When people fall asleep, the theta rhythm (4–8 Hz) increases while the alpha rhythm (8–13 Hz) decreases [18]. Sigma oscillations occur mainly during S2 stage [19]. Corsi-Cabrera *et al.* [20] claim that the beta rhythm (13–30 Hz) is primarily associated to the awake stage. Finally, the gamma rhythm (30–80 Hz) is related to the working memory and attention [21] and, thus, arises mainly during the awake stage.

**Table 1.** Brain rhythms and associated frequency ranges.

Brain rhythm	Frequency range (Hz)
Delta	0.5 – 4
Theta	4 – 8
Alpha	8 – 13
Sigma	11 – 15
Beta	13 – 30
Gamma	30 – 80

Based on the sleep characterizations defined by R&K and AASM, several computational methodologies were proposed, aiming to assist the sleep experts' task of scoring, respecting the chosen standards. Those methods invested efforts on the information sources (e.g. PSG channels selection), on the preprocessing step (e.g. data normalization), on the feature extraction method (e.g. domain transformation), on the extracted features themselves, and, of course, on the classification algorithms. For a comparative review between state of the art methods regarding preprocessing, number of channels, domain, feature selection, and classifiers, the reader can refer to [22].

Krakovská and Mezeiová [23] computed up to 74 features from EEG, EMG, EOG and ECG signals, selected a subset of them and applied for classification an artificial neural network (ANN). Charbonnier *et al.* [24] went in the same direction of [23], increasing the final accuracy with a similar ANN-based approach. Hsu *et al.* [15] also considered ANNs as classifiers although they used band-delimited energy of a single-channel EEG signal. By the other hand, Zhu *et al.* [7] considered horizontal (HVG) and difference visibility graphs (DVG) mapping from single-channel EEG signals, and classified the resulting features using a support vector machine (SVM). Liang *et al.* [14] employed linear discriminant analysis (LDA) along with the computation of the entropy of EEG signals in multiple scales. Random forest (RF) is fed with Renyi's entropy in Fraiwan *et al.* [13] and statistical moments in wavelet domain in [12], both extracted from a single EEG channel. In [12, 13], among many classification algorithms, RF was the best-performing choice.

Statistical information suppress effects of individuals' variations on the analyzed signal whilst capture its main behavior. By using feature selection algorithms, Şen *et al.* [25] surprisingly ranked statistical moments extracted from EEG in the

time domain among the worst-performing in a set of 41 features. In opposition to them, the work of Ronzhina *et al.* [11] rated statistical moments as important parameters in signal analysis at time domain. More recently, Hassan and Bhuiyan [26] and Silveira *et al.* [12] extracted statistical information after decomposing the signals in time-frequency domain, and used them as classification attributes for sleep staging. Nevertheless in [26] the preprocessing stage applied an expensive data-driven Ensemble Empirical Mode Decomposition (EEMD) method for obtaining the frequency bands. The high classification performances obtained in both works highlight the potential in associating time-frequency information with statistical tools for defining features. Still related to spectral-based pattern recognition, statistical moments were successfully employed to detect microcalcifications in mammograms [27], suggesting their effectiveness for image and signal analysis in the frequency domain. The current study aims to contribute to the discussion addressed by the aforementioned authors with respect to the choice of the feature set and its impact on the classifier performance, as further discussed in the upcoming sections.

### 3. Materials and Methods

In the current paper, the preprocessing step comprises two tasks: (i) Each of the Pz-Oz signals is normalized with respect to their variance and then split into smaller sets (epochs of 30s); (ii) FFT is applied on each epoch. The FFT complexity is  $\mathcal{O}(n \log n)$ , for data of size  $n$ . In the feature selection stage, again for each epoch, spectral-based statistical moments are computed from the FFT amplitudes. And finally, for the classification procedure, these statistical moments form a feature vector carried to a random forest classifier.

Each one of these steps is described and justified in the following sections. The data set used here is described in Section 3.1. Section 3.2 addresses the preprocessing step: signal normalization, splitting, and transformation. Section 3.3 motivates and describes the feature extraction step. Finally, the chosen classification algorithm as well as training details are described in Section 3.4.

#### 3.1 Data Description

As done in [6, 12, 7], the EEG data used in the current study were obtained from the PhysioNet Sleep-EDF (Expanded) database [28, 29]. We used all available signals from 20 subjects recorded during two nights, with the exception those signals recorded in the second night from the subject labeled as 13, which were not available in the database. All the thirty nine (39) analyzed signals were obtained from healthy volunteers, aged between 25 and 34, who stated not having used any sleep-related medication [30].

Since sleep can be seen as a succession of phenomena that begins in the brain stem, extending to all regions of the brain, the several EEG channels can contain redundant information [31]. Therefore it is well accepted by the scientific community that one EEG channel shall sufficiently provide

information to classify the sleep stages as stated in [11], and also explored in [5] and [32]. The recordings analyzed here are formatted in the EDF standard and contain two EEG channels, Pz-Oz and Fpz-Cz, sampled at 100Hz. In this study, as in [7, 11, 4], only the Pz-Oz channel is used since better results with respect to Fpz-Cz were obtained when applying different methodologies.

Together with the two EEG channels, each signal from the Sleep-EDF database has an associated hypnogram, containing a sleep stage classification. These annotations (scorings) are provided for each 30-second epoch of the recorded signals, and were performed manually by sleep experts, following the R&K recommendations [8]. Aiming to agree with the provided annotations and to enable the correct method evaluation, the chosen length of the analyzed epochs is also 30 seconds. Epochs whose hypnograms were labeled as "movement time" or "not scored" were removed from the analysis since both situations are not relevant for the sleep staging problem.

#### 3.2 Preprocessing of EEG Signals

As mentioned previously, the input set contains the data from the Pz-Oz channel of the 39 EEG signals, denoted by  $S_j$ , each one containing  $k = 1, \dots, N$  samples. Since each signal  $S_j$  ( $j = 1, \dots, 39$ ) is recorded in different conditions and is prone to individual signal variations between subjects [33], we start by normalizing each one of the input signals, striving for a zero mean and unit variance signal  $\hat{S}_j$ .

To accomplish that, the signal mean, denoted by  $\bar{s}_j$  in (eq. (1)) and its standard deviation  $\sigma_j$  (eq. (2)) are calculated. Thereafter, the normalized signal  $\hat{S}_j$  (eq. (3)) is obtained by subtracting  $\bar{s}_j$  from each sample  $S_{j,k}$  and dividing the result by the corresponding  $\sigma_j$ . This sort of normalization is also seen on other time-series related applications [34], because of the advantage of diminishing the biases of the data.

$$\bar{s}_j = \frac{\sum_{k=1}^N S_{j,k}}{N}, \quad (1)$$

$$\sigma_j = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (S_{j,k} - \bar{s}_j)^2}, \quad (2)$$

$$\hat{S}_{j,k} = \frac{S_{j,k} - \bar{s}_j}{\sigma_j}, \quad k = 1, \dots, N. \quad (3)$$

Afterwards, the normalized signals  $\hat{S}_j$  ( $j = 1, \dots, 39$ ) are split into  $Q$  epochs of 30 seconds length aiming to match the annotation interval provided by the database. Here the epochs are denoted by  $E^i$ ,  $i = 1, \dots, Q$ , omitting the index  $j$  to simplify the notation. Given the signal sample rate of  $r = 100$  Hz, the epoch time interval of  $t = 30$  s and the signal length  $N$ , the number of epochs for each signal  $\hat{S}_j$  is given by  $Q = \lfloor \frac{N}{p} \rfloor$ , being  $p = t \cdot r = 3000$  and  $\lfloor \cdot \rfloor$  the floor function. The FFT is then applied to each epoch, generating as output a vector of complex numbers (eq. 4).

$$f^i = FFT(E^i), \quad i = 1, \dots, Q. \quad (4)$$

Because of the symmetry of the FFT coefficients [17], we use only half of them,  $p/2$ , that are associated to the positive

frequencies. For being implemented in portable systems, this simple property of the FFT represents a significant advantage in comparison to data-driven techniques, that rely on adaptive procedures to obtain convergence to the searched results. To complete the preprocessing stage, the modulus (or the norm of the complex values) of the FFT coefficients is computed, given by eq. (5) and eq. (6), assuming  $\Re(z)$  and  $\Im(z)$  as the real and imaginary parts of a complex number  $z$ . The variable  $F_m^i$  in eq. (6) is proportional to amplitude of single-side EEG epoch spectrum.

$$F^i = \{F_0^i, F_1^i, \dots, F_{p/2}^i\}, \quad i = 1, \dots, Q, \quad (5)$$

$$F_m^i = \sqrt{\Re(f_m^i)^2 + \Im(f_m^i)^2}, \quad m = 0, 1, \dots, p/2. \quad (6)$$

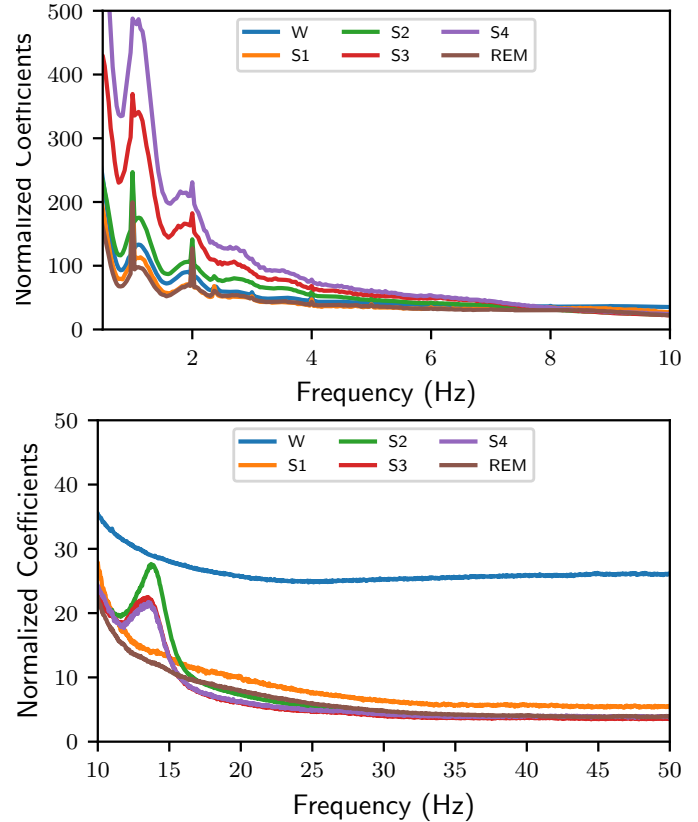
### 3.3 Feature Extraction

As the sleep staging with EEG signals is considered a typical problem of classification with extracted features [32, 4, 25], selection of the classifier inputs is the most important component of pattern classification since even the best classifier will perform poorly if the inputs are not selected well [35]. Moreover, it can be observed that different classifiers provide distinct performances for the same feature set, indicating that matching both may enhance results. In other words, the performance can be enhanced by taken the operation principle of the classifier into account and choosing a feature set that favors the classification process. In classification processes relying on decision trees, the classifier uses one or more features to recursively discriminate set of individuals into different classes, obtaining as many pure sets as possible thus leading to entropy reduction. The feature set proposed in this paper, takes the decision trees' classification strategy into account by systematically providing means to discriminate one or more classes at time.

To illustrate the way in which features obtained from different patterns can be employed to separate each of six sleep stages defined by the R&K standard [8], the corresponding mean-value curves computed from all the  $F^i$  values,  $i = 1, \dots, 39 \cdot Q$  are presented in Figure 1<sup>1</sup>. Figure 1 (left) shows the normalized values for frequencies from 0.5 to 10 Hz and Figure 1 (right), for 10 to 50 Hz. The  $F^i$  signals are grouped according the provided annotations (hypnogram provided by the dataset), and the mean curves are computed component-wise, given as functions of the frequencies. As observed in Figure 1 (right), by using the mean of Fourier coefficients between 30-50Hz is enough to separate a pure class W. Moreover, taking the average of coefficients corresponding to the sigma rhythm allows for separating S2, S3, and S4 from S1 and S3. This kind of reasoning stemmed from empirical observations on several spectral features guided the construction of the method proposed in this work.

The set of attributes proposed in the current work for the classification of sleep stages is computed from the vectors  $F^i$ ,

<sup>1</sup>The whole range of frequencies of interest of this study (0.5–50Hz) is presented in two plots because of the dissimilar scales in the vertical axis.



**Figure 1.** Mean of the Fourier coefficients. Left panel for frequencies from 0 to 10Hz. Right panel for frequencies from 10 to 50 Hz.

which contain the modulus from the Fourier coefficients of each epoch  $i$ . Again the notation is simplified, omitting the index  $j$  to specify each one of the 39 input signals, totalizing  $39 \cdot Q$  vectors  $F^i$  to be analyzed. For the construction of this set of attributes (denoted by  $V$ ), three statistical moments for certain frequency bands are computed: mean, skewness and kurtosis. The chosen frequency bands are associated to the main brain rhythms, what contributes for emphasizing the difference among them. In this sense, the computation of skewness (which is a measure of the degree of asymmetry of the data distribution) and kurtosis (which describes the peakedness of a distribution) are considered to strength this distinction.

The chosen eight attributes as well as the related frequency ranges are shown in Table 2. The reader can note that, according to the Nyquist theorem [17], we can only represent frequencies up to 50Hz.

#### 3.3.1 Mean of the FFT Coefficients from 30 Hz to 50 Hz

The first feature of the vector of attributes,  $V_1^i$ , is given in eq. (7). It is associated to index  $q = 1$  of Table 2, and it is computed for each epoch  $i$ .

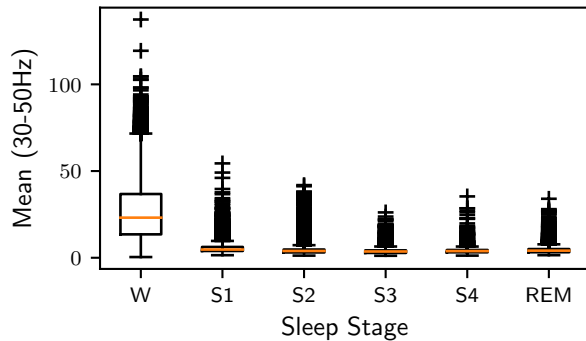
$$V_1^i = \frac{\sum_{k=n_q}^{N_q} F_k^i}{N_q - n_q}, \quad i = 1, \dots, Q. \quad (7)$$

**Table 2.** Eight proposed features computed from the modulus of the Fourier coefficients according to the indicated frequency range.

Index $q$	Measure	Frequency range (Hz)
1	Mean	30-50
2	Kurtosis	11-50
3	Skewness	11-50
4	Mean	0.5-4 (Delta rhythm)
5	Mean	4-8 (Theta rhythm)
6	Mean	8-13 (Alpha rhythm)
7	Mean	13-30 (Beta rhythm)
8	Mean	11-15 (Sigma rhythm)

The values  $n_q$  and  $N_q$ , with  $q \in \{1, \dots, 8\}$ , are defined respectively as the indexes of the first and the last Fourier coefficient for each selected range of frequency, presented in Table 2.

As can be seen in Figure 1, the average value of the coefficients at frequencies between 30Hz and 50Hz is roughly constant for all the stages. However, for the awake stage curve (W), the average value is significantly larger when compared to the other stages.



**Figure 2.** Box plot of the mean of the modulus of the FFT coefficients from 30 Hz to 50 Hz, according to the six sleep stages, computed based on each epoch and each signal.

In Figure 2, the information presented in Figure 1 is now depicted in terms of the box plot specifically for the selected range of frequency (30-50 Hz), highlighting the main differences among the awake stage and all the remainders. As suggested in [26], the non-overlapping and non-collinear notches of the box plots suggest good inter-class variation of the selected features. The mean of the FFT coefficients from 30 to 50Hz allows the classifier to distinguish the awake stage from the others. Thus, the next attributes should be chosen such that they could contribute for distinguishing all other sleep stages (S1 to REM) from each other. In this scenario, the further quantities to be analyzed are skewness and kurtosis (third and fourth statistical moments) also computed based on the values of  $F^i$  restricted to specific frequency ranges.

### 3.3.2 Kurtosis and Skewness of the FFT Coefficients from 11 to 50Hz

The next two chosen attributes are the third and fourth moments about the mean, the skewness, given in eq. (8), and kurtosis, given in eq. (9). Skewness indicates the shape deformations of a data distribution when compared with the normal distribution, and kurtosis measures how outlier-prone a distribution is [36]. These two figures of merit are now computed from all values  $F^i$ , again grouped for each R&K sleep range and restricted to a selected range of frequencies, delimited by  $n_q$  and  $N_q$ ,  $q = 2$  and  $q = 3$ , according:

$$V_2^i = \frac{\frac{1}{(N_q - n_q)} \sum_{k=n_q}^{N_q} (F_k^i - \bar{F}^i)^3}{\sigma_q^3}, \quad i = 1, \dots, Q; \quad (8)$$

$$V_3^i = \frac{\frac{1}{(N_q - n_q)} \sum_{k=n_q}^{N_q} (F_k^i - \bar{F}^i)^4}{\sigma_q^4}, \quad i = 1, \dots, Q; \quad (9)$$

The quantity  $\sigma_q$  is the standard deviation computed only for the selected frequency ranges, i.e. for indexes between  $n_q$  and  $N_q$ ,  $q = 2, 3$ . These two attributes allow to distinguish the sleep stages W, S1 and REM from S2, S3 and S4. Figures 3a and 3b present the average kurtosis and skewness for each sleep stage assuming all signals from the data base, as previously explained. As can be seen in Figure 1, the shape of the curves for stages S2, S3 and S4 are different compared to the other stages, strongly suggesting that kurtosis and skewness are effectively good options to discriminate sleep stages.

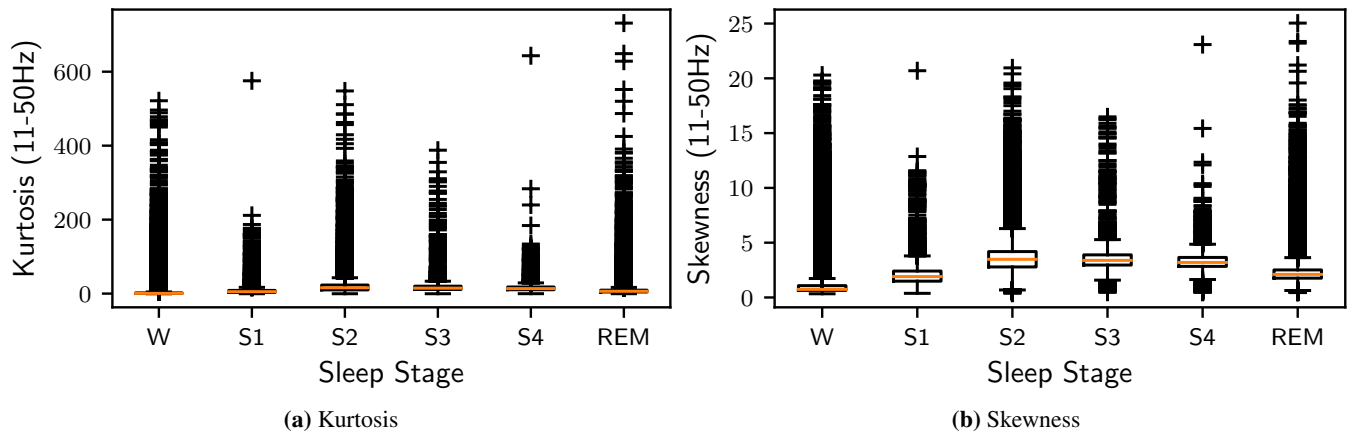
### 3.3.3 Mean of the FFT Coefficients of the Brain Rhythms

The feature vector  $V$  considered here, aggregates also the mean values computed from the modulus of the FFT coefficients for the sleep-related rhythms. Again, these quantities are restricted to the range of selected frequencies, as indicated in Table 2, delimited by  $n_q$  and  $N_q$ . Figure 4 present the mean values of  $F^i$  grouped in the six R&K sleep stages for the sleep-related rhythms. Once more, it is possible to observe the sensibility of the proposed features.

## 3.4 Classification Algorithm

The feature vector  $V$ , considered as the input data for the classification procedure, containing the eight features presented in the previous sections, is computed for each 30-second epoch indexed by  $i$ ,  $i = 1, \dots, 39 \cdot Q$ . Since the associated hypnograms are also informed in the database, the correct classification label is known for each epoch, enabling the evaluation of the robustness of the selected feature set for a specific classification algorithm.

Here, the considered classification procedure is the RF algorithm, introduced by [37] and already used in [12, 13] to classify sleep stages, but assuming different features. The RF consists in a set of  $T = 10$  decision tree predictors that classify individually an input instance and gives their votes for



**Figure 3.** Box plot of the kurtosis and skewness of the modulus of the FFT coefficients from 11 to 50hz of each epoch from the EEG signals.

total computation. The most voted class is the random forest output. In the current study, WEKA’s [38] implementation is used, and each one of the  $T$  tree predictors is a random tree. In [13], the random trees were also used as individual predictors although, as stated by them, other decision tree algorithms, such as C4.5, CART, CHAID and REP tree, could be used.

The process of building the random forest begins with a training set selection phase. Each random tree is grown with a smaller number  $F$  of features randomly selected from the original amount, given by  $F = \lfloor \log_2(M) + 1 \rfloor$  [37]. Here  $M$  is the total of the instances’ features. Furthermore, a replicated subset of the training set – obtained through bootstrap sampling – is used as a training set for each tree predictor growing during the classification process. Since for each epoch, the complete number of features  $M = 8$ , each tree is grown with exactly  $F = 4$  random features.

The tree growing, i.e. how the features are disposed in the tree body, is guided by the information gain measure [13]. Each tree grows as much as possible and no pruning technique is applied. Aiming to assess the robustness of the classification algorithm used in this study, the robust  $k$ -fold cross validation [39], with  $k = 10$ , is applied. In this technique, the complete dataset is divided into  $k = 10$  parts. In a given moment,  $k - 1$  subsets are used to train the classifier and the remaining one is considered to test the built model. This procedure is repeated  $k$  times and each part is used for testing just once. The average accuracy of all  $k$  executions is taken as the final evaluation.

In fact, preliminary tests with four decision tree algorithms also available in WEKA – namely random tree, C4.5, REP tree and CART – were also performed. The average results for the 6-state sleep staging problem, following the 10-fold cross validation are, respectively, 86.5%, 88.6%, 89.1% and 89.4%. However, as indicated in [12], RF has a better trade-off between accuracy and runtime.

### 4. Results and Discussion

The response of each extracted feature to W, S1, S2 and S3 R&K sleep stages is depicted for the second night recording of subject 00 in Figure 5. It illustrates how their complementary behavior contributes to the six sleep stages identification.

To provide a basic framework for assessment of the proposed methodology, the confusion matrices for the 2- to 6-state sleep stages are computed. A confusion matrix presents an overview of all hits and misses for each class performed by some classifier. These hits and misses presuppose the knowledge of correct instances, that are in this case the given hypnograms of each epoch. From a given confusion matrix, it is possible to obtain classes’ metrics as precision and sensitivity besides the overall method’s accuracy. Let  $tp$ ,  $fp$ ,  $tn$  and  $fn$  be the amount of true positives, false positives, true negatives and false negatives, respectively. The accuracy computation is given by

$$\text{accuracy} = \frac{tp + tn}{tp + tn + fp + fn}.$$

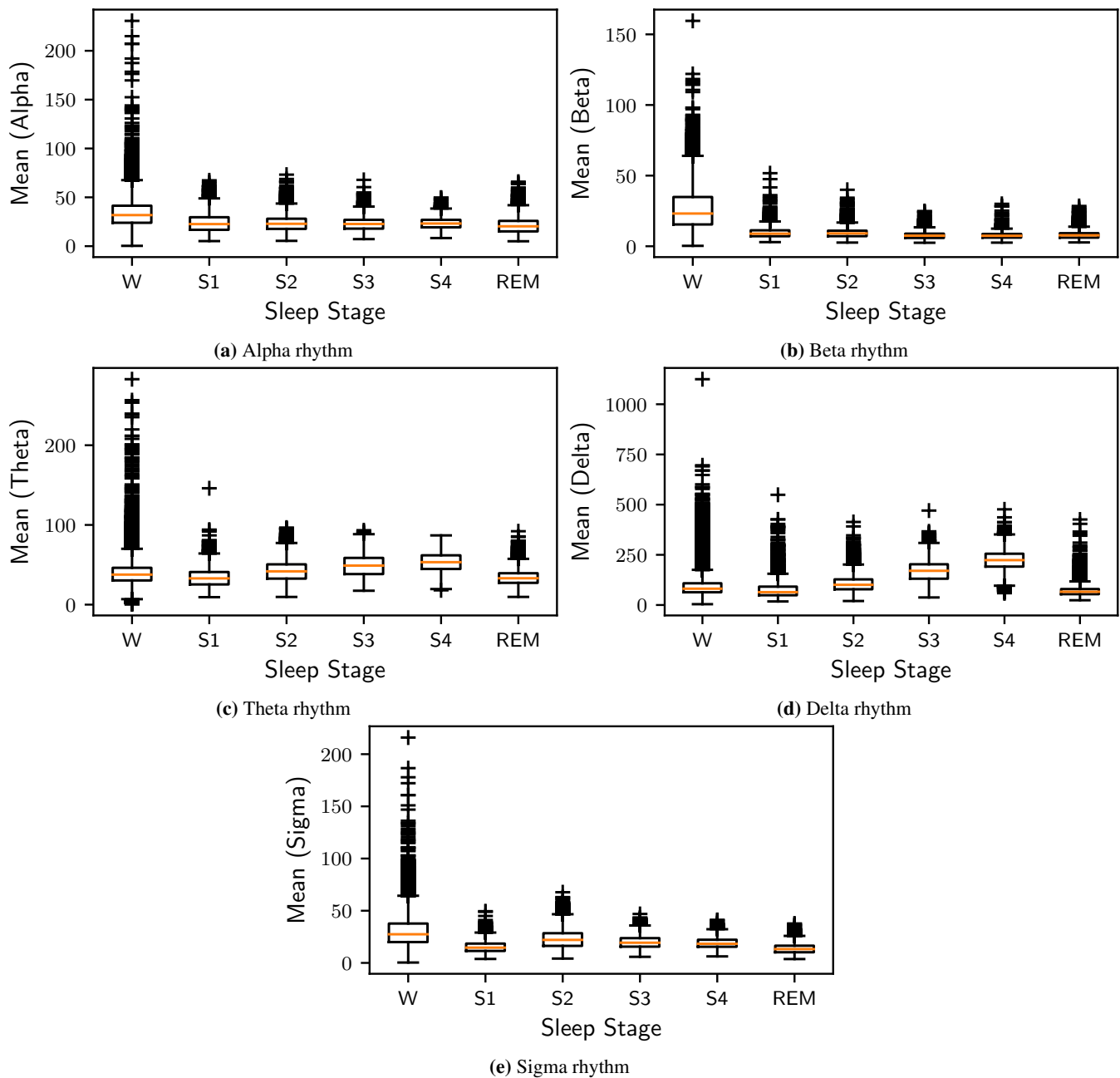
Similarly, precision and sensitivity measures for each class are obtained through

$$\text{precision} = \frac{tp}{tp + fp}$$

and

$$\text{sensitivity} = \frac{tp}{tp + fn}.$$

According to [40], precision is the rate of relevant classified instances, i.e. the correct scores; whilst the sensitivity is the rate of relevant instances which were correctly classified. The resulting confusion matrices and class metrics for the 6- to 2-state sleep stages are disposed in Tables 3, 4, 5, 6 and 7. Not all related works exhibit this complete information source, and therefore, we compare the obtained results with those we had access to.



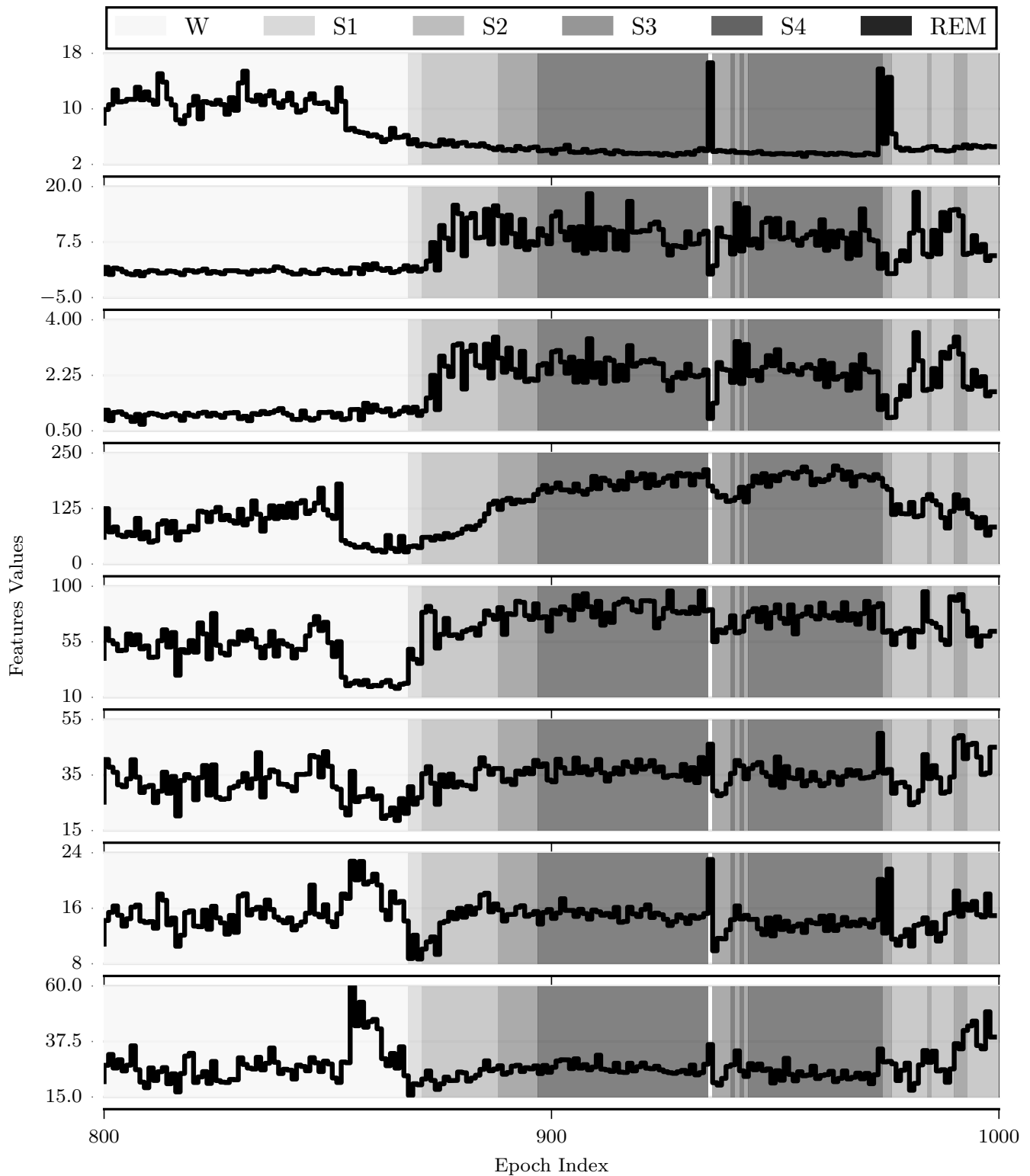
**Figure 4.** Box plots of the mean values computed from the modulus of the FFT coefficients for the alpha, beta, theta, delta and sigma rhythms, assuming data grouped in the six R&K sleep stages.

The proposed method provides a sensitivity enhancement of respectively 9.2% and 10.4% on S1 and REM staging, with respect to our previous work [12], in the 6-states case (cf. Table 3). Those two classes are prone to be misclassified while staging, with S1 normally exhibiting the poorest classification performances [32].

For the group with 5 classes (cf. Table 4), the following comparisons can be made. The proposed method performed 7% and 41.6% better in terms of precision for W and S1 stages when compared to [4] with their own data set being considered. In the case where Physionet signals are employed,

the proposed methodology yields a 13.8% higher sensitivity for W. Respective precisions for W and SWS are 7.1% and 2.4% higher, and sensitivity of S2 is 1.8% better in comparison to [13]. When compared to [15], our methodology achieves 25.9% and 14% higher precision and sensitivity values for the W stage. Still, sensitivity of SWS is 14.9% higher for the introduced methodology. In comparison with [14], the proposed methodology achieves 8.2%, 37.5% higher precision values for W and S1.

For the 4-state sleep stage classification (cf. Table 5), the proposed method achieved 11.7% and 2.7% higher precision



**Figure 5.** Behavior of the eight extracted features for several sleep stage transitions from the second night recording of subject 00 according to the Physionet Bank notation. From top to bottom, the mean in 30–50Hz, the kurtosis and skewness in 11–50Hz and the mean of delta, theta, alpha, sigma and beta rhythms.



**Table 3.** Confusion matrix of the 6-state sleep stages classification

		Proposed method's scoring					
		W	S1	S2	S3	S4	REM
Expert's scoring	W	70143	134	152	8	31	404
	S1	971	439	477	6	2	909
	S2	601	82	15529	518	98	971
	S3	126	1	1164	1594	475	10
	S4	82	0	199	487	1562	3
	REM	594	133	885	7	4	6094
Precision		96.7%	56.1%	84.5%	61.2%	72.2%	72.5%
Sensitivity		99.0%	15.0%	87.2%	47.9%	66.1%	79.2%

**Table 4.** Confusion matrix of the 5-state sleep stages classification

		Proposed method's scoring				
		W	S1	S2	SWS	REM
Expert's scoring	W	70145	127	147	53	400
	S1	984	429	473	15	903
	S2	602	72	15353	775	997
	SWS	198	1	1176	4316	12
	REM	598	118	895	20	6086
	Precision		96.7%	56.4%	85.1%	83.5%
Sensitivity		99.0%	15.3%	86.5%	75.3%	78.9%

for the S1/S2 and SWS stages. In the context of 3 classes (cf. Table 6), our method was 1.7% more sensitive for the awake stage. Both results are regarded to [7]. Although we do show Table 7, it was not possible to compare the sensitivity and precision for the proposed method, in the 2-sleep stage context, once considered studies do not present this information.

**Table 5.** Confusion matrix of the 4-state sleep stages classification

		Proposed method's scoring			
		W	S12	SWS	REM
Expert's scoring	W	70044	417	52	359
	S12	1484	16870	753	1496
	SWS	187	1202	4297	17
	REM	567	1352	16	5782
	Precision		96.9%	85.1%	84.0%
Sensitivity		98.8%	81.9%	75.3%	75.1%

Summarizing, the awake stage is the most identifiable stage by the proposed methodology for all the cases. The same behavior is observed in [7]. Furthermore, as occurred in [7] and [20], the S1 stage is easily classified as REM and its adjacent stages, i.e. awake and S2. The authors in [11] have also reported similar issue in their methodology, which still remains a challenge for the researchers in this field [32].

Table 8 shows the comparison of the proposed method

**Table 6.** Confusion matrix of the 3-state sleep stages classification

		Proposed method's scoring		
		W	NREM	REM
Expert's scoring	W	69995	523	354
	NREM	1595	23195	1516
	REM	563	1388	5766
Precision		97.0%	92.4%	75.5%
Sensitivity		98.8%	88.1%	74.9%

**Table 7.** Confusion matrix of the 2-state sleep stages classification

		Proposed method's scoring	
		W	SLP
Expert's scoring	W	69761	1111
	SLP	1850	32173
Precision		97.4%	96.7%
Sensitivity		98.4%	94.6%

with other well established studies in terms of overall accuracy for classical AASM 5-state sleep stage scoring.

One can note that the proposed methodology is reliable even considering more than 100,000 epochs. Moreover it achieves the best performance among the considered state-of-the-art studies. The introduced methodology has 6.4% higher accuracy when compared with [13], that used the same classifier and approximately five times less epochs for analysis.

Finally, Table 9 compares the performance of our method with other methods that focus in major of the 2- to 6-state sleep stage classifications, and use the same dataset. Once [11] presented several ANN architectures, those reported by the authors as performing better are considered here for comparison. Note that the proposed method achieved the best accuracies for the 6-, 5- and 4-state sleep stages. For the problems of 3 and 2 classes, the proposed method achieves the second and third best results. It is important to highlight again that [4, 11, 7] do not consider a single feature set for the five classification problems, as we do.

## 5. Conclusion

The current paper presents a novel methodology for classification of sleep stages based on the analysis of a single EEG channel in frequency domain. A positive aspect is the assumption of a single set of just 8 features for scoring all the 2- to 6-state sleep stages in accordance with the Rechtschaffen and Kales' recommendations. Another advantage of the proposed methodology, is the inexpensive treatment given for

**Table 8.** Comparison with other methods for 5-state sleep stages

Methodology	Features	Classifier	Quantity of epochs	Accuracy
Krakovska and Mezeiova [23]	EEG-based features	ANN	18,058	74.0%
Liang <i>et al.</i> [14]	Multiscale entropy	LDA	3,708	76.9%
Fraiwan <i>et al.</i> [13]	Renyi's entropy	RF	20,269	83.0%
Charbonnier <i>et al.</i> [24]	EEG, EMG and EOG-based features	ANN	62,399	85.5%
Hsu <i>et al.</i> [15]	Energy of EEG	ANN	2,880	87.2%
Zhu <i>et al.</i> [7]	DVG and HVG	SVM	14,963	89.0%
Silveira <i>et al.</i> [12]	Statistical moments	RF	106,376	91.5%
Proposed method	Statistical moments	RF	104,895	91.8%

**Table 9.** Accuracy comparison with other methods for sleep stage classification

Sleep states	Zhu <i>et al.</i> [7]	Ronzhina <i>et al.</i> [11]	Berthomier <i>et al.</i> [4]	Silveira <i>et al.</i> [12]	Proposed method
2	97.9%	96.9%	95.4%	97.3%	97.1%
3	92.6%	90.3%	88.3%	93.9%	94.3%
4	89.3%	81.4%	74.5%	92.3%	92.4%
5	88.9%	-	71.2%	91.5%	91.8%
6	87.5%	76.7%	-	90.5%	90.9%

the preprocessing step, specially when compared with EMD-based techniques to obtain ranges of frequencies. These are signal-driven techniques which might need a great amount of iterations until convergence has been obtained and features can be actually computed. In the recently published work [26], 5000 EEMD iterations were considered for generating the localized time-frequency signal estimations, which might be considered as a drawback in the design of portable systems.

More than 100,000 epochs of 30 seconds from an open access database are analyzed highlighting the robustness of the preformed analysis, when considering a large amount of data. In the current study, WEKA's random forest implementation was used as the classification procedure, whose performance is also compared with other classifiers also at disposal in the WEKA software. The introduced procedure achieves the best performance for classifying the sleep stages in 5 states when compared with several state-of-the-art studies, that in general considered a much smaller set of analyzed epochs. Furthermore, the proposed method achieves best results for 4- to 6-state sleep stages classification scenarios.

Through the presented scoring experiments, the strength of the proposed methodology is enhanced and an significant contribution in distinguishing S1 and REM is obtained. The statistical moments computed from Fourier coefficients restricted to frequency ranges not obviously related to sleep rhythms show themselves as an effective feature choice capable of increasing the accuracy of sleep stages classification procedures.

As a future work, we intend to investigate measures to mitigate the effects of class imbalance. and also evaluate the performance of the random forest algorithm based not on random trees but on other algorithms like REP tree, CART and C4.5. Motivated by the results obtained by the individual predictors (cf. Section 3.4), we do believe that better results can be achieved. Other ensemble configurations, besides pruning can be explored aiming to enforce a good compromise between performance and runtime.

## Acknowledgements

The authors would like to thank FAPERGS PG n.1873-25.51/13-0, UFSC/PIPE, CNPq and CAPES for partial support.

## Author contributions

The paper shows the results of a project cooperatively developed by the authors in the last years in the field of classification of biological signs such as ECG and EEG. Eduardo Braun made the implementations. Thiago da Silveira suggested the Random Forest classifier. Giovanni Baratto indicated the area of application. Prof. Cesar Ramos Rodrigues gave the idea to classify sleep stages. Prof. Alice Kozakevicius suggested the use of statistical moments.

## References

- [1] COLE, R. J. et al. Automatic sleep/wake identification from wrist activity. *Sleep*, v. 15, n. 5, p. 461 – 469, 1992.
- [2] IBER, C. et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. 1. ed. Darien, USA, 2007.
- [3] JAP, B. T. et al. Using EEG spectral components to assess algorithms for detecting fatigue. *Experts Systems with Applications*, v. 36, n. 2, p. 2352–2359, 2009.
- [4] BERTHOMIER, C. et al. Automatic analysis of single-channel sleep EEG: Validation in healthy individuals. *Sleep*, v. 30, n. 11, p. 1587–1595, 2007.

- [5] SILVEIRA, T. da; KOZAKEVICIUS, A. de J.; RODRIGUES, C. R. Drowsiness detection for single channel EEG by DWT best m-term approximation. *Research on Biomedical Engineering*, v. 31, n. 2, p. 107 – 115, 2015.
- [6] SILVEIRA, T. L. T. da; KOZAKEVICIUS, A. J.; RODRIGUES, C. R. Automated drowsiness detection through wavelet packet analysis of a single EEG channel. *Expert Systems With Applications*, v. 55, n. 3, p. 559–565, 2016.
- [7] ZHU, G.; LI, Y.; WEN, P. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. *IEEE Journal of Biomedical and Health Informatics*, v. 18, n. 6, p. 1813–1821, 2014.
- [8] RECHTSCHAFFEN, A.; KALES, A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Electroencephalography and Clinical Neurophysiology*, v. 26, n. 6, p. 644, 1969.
- [9] MOSER, D. et al. Sleep classification according to AASM and rechtschaffen & kales: effects on sleep scoring parameters. *SLEEP*, v. 32, n. 2, p. 139–149, 2009.
- [10] EBRAHIMI, F. et al. Automated sleep stage classification based on EEG signals by using neural networks and wavelet packet. In: *IEEE 30th Annual International IEEE EMBS Conference*. Vancouver, Canada: IEEE, 2008. v. 1, n. 1.
- [11] RONZHINA, M. et al. Sleep scoring using artificial neural networks. *Sleep Medicine Reviews*, v. 16, n. 3, p. 251–263, 2012.
- [12] SILVEIRA, T. L. T. da; KOZAKEVICIUS, A. J.; RODRIGUES, C. R. Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Medical & Biological Engineering & Computing*, v. 55, n. 2, p. 343–352, 2017.
- [13] FRAIWAN, L. et al. 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*, v. 108, n. 1, p. 10–19, 2012.
- [14] LIANG, S. F. et al. Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models. *IEEE Transactions on Instrumentation and Measurement*, v. 61, n. 6, p. 1649–1657, 2012.
- [15] HSU, Y.-L. et al. Automatic sleep stage recurrent neural classifier using energy features of EEG signals. *Neurocomputing*, v. 104, n. 1, p. 105–114, 2013.
- [16] PAPADELIS, C. et al. Indicators of sleepiness in an ambulatory EEG study of night driving. In: NEUMAN, M. (Ed.). *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. New York, USA: IEEE, 2006. v. 25, n. 3.
- [17] BRIGGS, W. L.; HENSON, V. E. *The DFT: An Owners' Manual for the Discrete Fourier Transform*. 1. ed. Philadelphia, USA: Society for Industrial and Applied Mathematics, 1995. v. 1. (Miscellaneous Bks, v. 1).
- [18] SUBASI, A. Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Experts Systems with Applications*, v. 28, n. 4, p. 701–711, 2005.
- [19] KRYGER, M. H. *Atlas of Clinical Sleep Medicine: Expert consult*. 2. ed. Philadelphia, USA: Saunders, 2009.
- [20] CORSI-CABRERA, M. et al. Power and coherent oscillations distinguish REM sleep, stage 1 and wakefulness. *International Journal of Psychophysiology*, v. 60, n. 1, p. 59–66, 2006.
- [21] JIA, X.; KOHN, A. Gamma rhythms in the brain. *PLoS Biology*, v. 9, n. 4, p. 1–4, 2011.
- [22] BOOSTANI, R.; KARIMZADEH, F.; NAMI, M. A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer Methods and Programs in Biomedicine*, v. 140, n. Supplement C, p. 77 – 91, 2017.
- [23] KRAKOVSKÁ, A.; MEZEIOVÁ, K. Automatic sleep scoring: A search for an optimal combination of measures. *Artificial Intelligence in Medicine*, v. 53, n. 1, p. 25–33, 2011.
- [24] CHARBONNIER, S. et al. Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging. *Computers in Biology and Medicine*, v. 41, n. 6, p. 380 – 389, 2011.
- [25] ŞEN, B. et al. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *J Med Syst*, v. 38, n. 3, p. 1–21, 2014.
- [26] HASSAN, A. R.; BHUIYAN, M. I. H. Automated identification of sleep states from eeg signals by means of ensemble empirical mode decomposition and random under sampling boosting. *Computer Methods and Programs in Biomedicine*, v. 140, n. Supplement C, p. 201 – 210, 2017.
- [27] ASADZADEH, M.; HASHEMI, E.; KOZAKEVICIUS, A. Efficiency of combined daubechies and statistical parameters applied in mammography. *Applied and Computational Mathematics: An International Journal*, v. 12, n. 3, p. 289–305, 2013.
- [28] GOLDBERGER, A. L. et al. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, v. 101, n. 23, p. 215–220, 2000.
- [29] PHYSIONET. *The Sleep-EDF-X Database*. 2013. <<http://www.physionet.org/physiobank/database/sleep-edfx>>. Online; Accessed on May 12th 2015.
- [30] KEMP, B. et al. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEGkemp2000a. *IEEE Transactions on Biomedical Engineering*, v. 47, n. 9, p. 1185–1194, 2000.

- [31] AMIN, H. U. et al. A novel approach based on data redundancy for feature extraction of EEG signals. *Brain Topography*, v. 29, n. 2, p. 207–217, 2016.
- [32] ABOALAYON, K. A. I. et al. Sleep stage classification using EEG signal analysis: A comprehensive survey and new investigation. *Entropy*, v. 18, n. 9, p. 272–302, 2016.
- [33] FREEMAN, W. J. Origin, structure, and role of background EEG activity. part 1. analytic amplitude. *Clinical Neurophysiology*, v. 115, n. 9, p. 2077–2088, 2004.
- [34] VIKKI, O.; LAURILA, K. Cepstral domain segmental feature vector normalization for noise robust speech recognition. *Speech Communication*, v. 25, n. 1–3, p. 133 – 147, 1998.
- [35] ÜBEYLI, E. D. Statistics over features: Eeg signals analysis. *Computers in Biology and Medicine*, v. 39, n. 8, p. 733 – 741, 2009.
- [36] DECARLO, L. T. On the meaning and use of kurtosis. *Psychological Methods*, v. 2, p. 292–307, 1997.
- [37] BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.
- [38] HALL, M. et al. The WEKA data mining software: An update. *SIGKDD Explorations*, v. 11, n. 1, p. 1–9, 2009.
- [39] ALPAYDIN, E. *Introduction to Machine Learning*. 2nd. ed. Cambridge, USA: The MIT Press, 2010. v. 1.
- [40] HAN, J.; KAMBER, M. *Data Mining: Concepts and Techniques*. 3. ed. Burlington, USA: Morgan Kaufmann/Elsevier, 2012. v. 1.