INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. (2009)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/acs.1147

Automated sleep–wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules

Florian Chapotot[1,3,*,†] and Guillaume Becq[2,3]

[1]*Department of Medicine, University of Chicago, Chicago, IL, U.S.A.*
[2]*GIPSA-Lab, Department Image Signal, Grenoble, France*
[3]*PhiTools SARL, Strasbourg, France*

SUMMARY

The classification of sleep–wake stages suffers from poor standardization in scoring criteria and heterogeneous conditioning of polysomnographic signals. To improve applicability of fully automated sleep staging, we have designed a formal classification framework to rigorously (1) select robust candidate features, (2) emulate artificial neural network classifiers, and (3) assign sleep–wake stages using flexible decision rules. An extensive database of 48 PSG records scored in 20 s epochs by two independent clinicians was used. A small subset of 2 s elementary epochs representative of each stages with unequivocal expert scores was selected to form a limited set of learning exemplars. From 16 statistical, spectral and non-linear candidate features extracted in 2 s epochs from EEG and EMG signals, a sequential forward search selected an optimal set of five features with a 22% error rate. Multiple layer perceptrons were trained from this optimal feature set while classification accuracy was assessed using the unequivocal instance subset. A simple majority vote among 10 consecutive classifier outputs ensured a final scoring resolution comparable to that of the experts. Poor classification performance was obtained for movement time, wakefulness, and intermediate sleep stages with a $36 \pm 15\%$ error rate (Cohen's kappa 0.48 ± 0.18). In contrast, deep and paradoxical sleep was classified with an 82% accuracy not far from inter-expert expert agreement ($83 \pm 3\%$). Significant improvements should be expected using a larger learning set compensating for a high inter-individual variability, and decision rules incorporating more domain-knowledge. Copyright © 2009 John Wiley & Sons, Ltd.

Received 1 April 2008; Revised 7 May 2009; Accepted 27 August 2009

KEY WORDS: data mining; decision rule; machine learning; polysomnography; signal processing; sleep

1. INTRODUCTION

Polysomnography (PSG) is a psycho-physiological method for the assessment of sleep and wake states. It is based on the concurrent recording of brain electroencephalographic (EEG), chin electromyographic (EMG) and electro-oculographic (EOG) signals collected in human individuals using non-invasive surface electrodes. PSG allows for the description of different sleep–wake states, which may exhibit abnormal qualitative and quantitative changes with clinical conditions and environmental situations. It is the golden standard in the diagnosis of sleep disorders and in the evaluation of psychotropic and other drugs potentially impacting sleep and vigilance.

In human, the scoring of sleep–wake stages is usually performed by a highly trained human expert on

*Correspondence to: Florian Chapotot, Department of Medicine, University of Chicago, Chicago, IL, U.S.A.
†E-mail: fchapotot@uchicago.edu

Contract/grant sponsor: PhiTools SARL

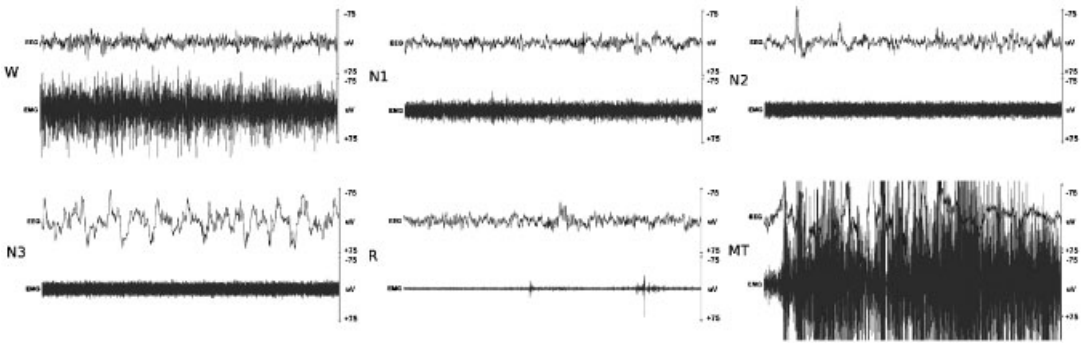

Figure 1. Typical scoring epochs obtained in different sleep–wake stages from human PSG recording. A 20 s epoch including a central EEG (top) and a chin EMG (bottom) derivation is illustrated for each stage (W for wakefulness, N_1 for transitional sleep, N_2 for shallow sleep, N_3 for deep sleep, R for paradoxical sleep, and MT for movement time periods).

the basis of an epoch-by-epoch visual interpretation of the PSG signals according to a set of conventional rules originally defined in the Rechtschaffen and Kale's manual [1], and recently updated by the American Academy of Sleep Medicine [2]. In animal research, different scoring rules exist and vary according to laboratories and species under study. Visual interpretation of PSG records uses a fixed epoch duration, usually 10, 20, or 30 s, and allows for the recognition of different sleep and wake stages (Figure 1). Such a manual task can be fastidious and time-consuming, lasting between 20 and 60 min for a usual nighttime human study. As it involves recognition, judgment and decision from a human operator, it is also an inherently subjective process.

With the emergence of data mining and new pattern recognition technics, automation of the sleep–wake staging process has been attempted repeatedly [3–6]. However, automatic analysis remains poorly applicable and has never been used efficiently on a very large scale. Limitations in the generalizability of automatic methods can mostly be accounted for by a sizable heterogeneity in the classifier inputs, and by a lack of standardization in its output preventing user full satisfaction. Indeed, substantial differences exist in the signal conditioning and quality of a constellation of acquisition devices from numerous PSG hardware manufacturers. Moreover, human and animal sleep is not always described using the same scoring rules and epoch durations [1, 7]. For instance, considering human sleep alone, interpretation is based on a different number of stages according to the guidelines adopted by the laboratory [1, 2].

A review of computer-based sleep analysis [8] reported the existence of the afore-mentioned challenges compromising applicability of the existing solutions. Indeed, as previously reported [4], the results of automatic sleep–wake stagers, when assessed using a different set of records, can demonstrate substantial discrepancies. Therefore, a strategic goal of the present study was to provide an applicable and productive framework to automate the scoring of human and animal sleep–wake stages by operating regardless of the devices collecting PSG signals, and independently of the desired time resolution of the analysis.

In this study, we focussed on extracting the most relevant information, selecting adequate learning exemplars, and combined the advantages of an additional post-classification decision layer. Artificial neural networks are widely used, suitable and well adapted in handling biological data. Based on the reported advantages and demonstrated superiority of this kind of classifiers in sleep and wake staging [9–11], we opted for a specific classifier, the multiple layer perceptron (MLP), as first introduced in sleep analysis by Schaltenbrand *et al.* [3, 12]. In parallel, an independent study showed that, in the purpose of automatic classification, relevant information can be captured from PSG signals using relatively simple sets of features [13]. We further improved these

DOI: 10.1002/acs

preliminary results by addressing potential issues in signal conditioning and bio-calibration related to the use of different PSG recording devices. Finally, a highly desirable but still missing feature, i.e. temporal flexibility, has been introduced in the present study, enabling automatic analysis to operate at variable time resolution.

We thus describe the methods developed and the results obtained (1) while training machine classifiers using a limited selection of instances and attributes, and (2) while assessing realistic performance of the system combining machine classification and post-processing decision rules using a larger data set. A description of the algorithms corresponding to these different processes is also provided.

2. DATA

2.1. Collection

PSG signals were recorded continuously in 13 healthy subjects (19–47 years old, 11 males and 2 females) for a total of 48 nights, either during baseline sleep or after sleep deprivation. A complete description of the experimental protocol can be found in Chapotot *et al.* [14]. A multi-channel ambulatory recording device was used to collect four EEG derivations (C_3-A_2, P_3-A_2, C_4-A_1, and P_4-A_1), a transversal EOG channel, and a chin EMG channel, filtered between 0.3 and 35 Hz. Bio-electrical signals were digitized at a sampling frequency of 128 Hz using 8 bit quantization between -500 and 500 μV, and stored into computer files using the standard EDF data format [15]. EEG cup-electrodes were attached onto the scalp of the subjects according to the international 10–20 system for electrodes placement [16].

2.2. Visual interpretation

PSG records were visually and independently interpreted by two sleep physicians blind as to the subjects and experimental conditions using the PRANA biosignal processing software (PhiTools, Strasbourg, France). Sleep–wake stages were scored according to conventional criteria [1] using a fixed epoch duration of 20 s resulting in 84 040 scoring epochs. Each epoch was classified exclusively amongst six possible stages: wakefulness (W), transitional sleep (N_1), shallow sleep (N_2), deep sleep (N_3), paradoxical sleep (R) and movement time (MT).

3. METHODS

3.1. Candidate feature extraction

A preselection of candidate features used as attributes in machine learning and classification was achieved using a mixture of conventional spectral, statistical, and more advanced non-linear parameters characterizing PSG signals in time and frequency domains. Given the large diversity of existing PSG recorders, and, actually, of the important heterogeneity in corresponding hardware specifications and acquisition settings, candidate features were preselected according to potential independence regarding changes or differences in signal conditioning and calibration. The practical purpose here was to achieve a robust capture of information without limitation from the biosignal recording device brand or model. The rationale in selecting candidate features was based on the mathematical properties involved in the computation of each feature.

Since sleep can usually be assessed visually disregarding eye movements, the EOG channel, generally not recorded in animal, was discarded from the present study. Consequently, a set of 16 different features (list given in Table I) has been extracted from the right central EEG (C_4-A_1) or the chin EMG channels in the 48 PSG records. A fixed temporal window of 2 s, during which signals could reasonably be expected as either stationary or contaminated by recording artifacts, ensured a successive non-overlapping time-varying extraction. The PRANA software (PhiTools, Strasbourg, France) was also used with the corresponding plug-in to extract all candidate features. A total amount of 840 400 elementary epochs were then processed.

Shannon Entropy: Shannon entropy (ShE) represents a measure of the information contained in a signal [17]. Since its computation does not rely on the signal amplitude range, ShE was supposed to be an adequate candidate.

DOI: 10.1002/acs

Table I. Candidate features extracted for their potential independence regarding differences in PSG acquisition settings and signal conditioning.

Feature	Source	Unit	Label
Shannon entropy	EEG	bit	$\mathrm{ShE}_{\mathrm{EEG}}$
Sample entropy	EEG	—	$\mathrm{SaE}_{\mathrm{EEG}}$
[2]Hjorth activity	EEG	μV	$\mathrm{Act}_{\mathrm{EEG}}$
[1]Hjorth mobility	EEG	Hz	$\mathrm{Mob}_{\mathrm{EEG}}$
Hjorth complexity	EEG	—	$\mathrm{Cpx}_{\mathrm{EEG}}$
Hurst exponent	EEG	—	H_{EEG}
Spectral edge frequency 95%	EEG	Hz	$\mathrm{SEF}_{95,\mathrm{EEG}}$
Delta relative power	EEG	%	$P_{\delta,\mathrm{rel,EEG}}$
Theta relative power	EEG	%	$P_{\theta,\mathrm{rel,EEG}}$
Alpha relative power	EEG	%	$P_{\alpha,\mathrm{rel,EEG}}$
[5]Sigma relative power	EEG	%	$P_{\sigma,\mathrm{rel,EEG}}$
[4]Beta relative power	EEG	%	$P_{\beta,\mathrm{rel,EEG}}$
Gamma relative power	EEG	%	$P_{\gamma,\mathrm{rel,EEG}}$
Shannon entropy	EMG	bit	ShEEMG
[3]Spectral edge frequency 95%	EMG	Hz	$\mathrm{SEF}_{95,\mathrm{EMG}}$
Gamma relative power	EMG	—	$P_{\gamma,\mathrm{rel,EMG}}$

Superscripted values indicate the order of feature selection resulting from a sequential forward selection algorithm.

The computation of ShE is numerically defined by

$$\mathrm{ShE}(x)=\sum_{k=1}^{N} p_k \log p_k \tag{1}$$

where p_k is the probability of a sample to be at quantification level k in a range of N levels. The probability p_k was estimated by summing up the number of samples in the range $[x_k, x_{k+1}]$, with $x_0=-250\,\mu\mathrm{V}$ and $x_k=x_0+500/(2^{-16}-1)k$ adapted to the signal amplitude range of the study.

Sample Entropy: Sample entropy (SaE) represents a more complex measure of similarity in the information carried by a signal [18]. SaE varies between 0 and 1 regardless of the signal amplitude range and so was chosen as a candidate feature.

The computation of SaE is based on an estimate of the number of near vectors in an embedding space created from the signal. Let m be the dimension of the embedding space (here m was 2), and N the signal dimension depending on the number of samples (using 2 s epochs sampled at 128 Hz, N was 256). Let $B_i(r)$ be the number of vectors $x(j)$, near to $x(i)$ in the embedding space at level r:

$$B_i(r)=\|\{j\,|\,d(x(i),x(j))<r\}\| \tag{2}$$

with $\|.\|$ the number of indices j for neighbors of $x(i)$, d the Chebyshev distance $d(x,y)=|y-x|$, and r set to $r=0.2\cdot SD$, with SD being the standard deviation of the signal. Let C be a statistical measure of this number of elements verifying the condition in the embedding space:

$$C_i^m(r)=\frac{B_i(r)}{N-m+1} \tag{3}$$

and C an average over all these elements such as:

$$C^m=\frac{1}{N-m+1}\sum_{i=1}^{N-m} C_i^m \tag{4}$$

An approximate of the Kolmogorov entropy [18] is then given by:

$$\mathrm{SaE}=-\log\frac{C^{m+1}}{C^m} \tag{5}$$

Hurst exponent: The Hurst exponent (H) characterizes dependance or self-similarity of a signal [19, 20]. H varies between 0 and 1 and when it exceeds 0.5, the corresponding signal is said persistent with similar consecutive trends. In contrast, when H displays values below 0.5, anti-persistence is present and indicates a

DOI: 10.1002/acs

changing trend in the analyzed signal. When H approximates 0.5, the signal randomly switches from one trend to another like in a brownian motion.

The computation of H, as described in [21], is based on the analysis of the rescaled range R/S, with R being the range of the cumulative sum of the centered signal X_T on a scale containing T samples, and S the standard deviation of the signal. R/S is the mean value of R over S for all bins of a signal containing N samples. H is then defined as:

$$R/S = (a \cdot T)^{\text{Hurst}} \tag{6}$$

In pratice, H was evaluated by determining the slope of $\log(R/S)$ over $\log(T)$ with T varying between 10 and $[N/2]$, to discard noise inference and consider at least two signal segments.

Hjorth's descriptors: The three EEG descriptors elaborated by Hjorth, activity (Act), mobility (Mob) and complexity (Cpx), and defined in [22] were selected as candidate features since they have been widely used in the sleep community.

Euler approximations were used to obtain derivatives. Act is simply the variance of the signal and was obtained by:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x(i) \tag{7}$$

$$\text{Act} = {\sigma_a}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x(i)-\bar{x})^2 \tag{8}$$

Mob is the standard deviation of the slope of the signal normalized by its standard deviation. It was obtained by:

$$x_d(i) = \frac{x(i+1)-x(i)}{T_e} \quad \forall i \in [2, N] \tag{9}$$

$$\text{Mob} = \frac{\sigma_d}{\sigma_a} = \frac{\sqrt{\frac{1}{N-2}\sum_{i=2}^{N}(x_d(i)-\bar{x}_d)^2}}{\sigma_a} \tag{10}$$

Cpx can be seen as the normalized standard deviation of the second derivative of the signal. It is a measure of the complexity of spectral behaviors. It displays values below 1 for signals more complex than a sine wave and was obtained by:

$$x_{dd}(i) = \frac{x(i)-2x(i-1)+x(i-2)}{T_e^2} \quad \forall i \in [3, N] \tag{11}$$

$$\text{Cplx} = \frac{\sigma_{dd}/\sigma_d}{\sigma_d/\sigma_a} = \frac{\sqrt{\frac{1}{N-3}\sum_{i=3}^{N}(x_{dd}(i)-\bar{x}_{dd})^2}}{\sigma_a} \tag{12}$$

Spectral relative powers and edge frequency: Spectral relative powers and edge frequency depend on the spectral composition of the signal. They are expressed as relative and frequency units, respectively, and thus do not rely on the signal amplitude range and sampling frequency. Relative spectral powers in the $N_b = 6$ traditional frequencies bands, δ [0.5, 4.5 Hz], θ [4.5, 8.5 Hz], α [8.5, 12.5 Hz], σ [12.5, 15.5 Hz], β [15.5, 22.5 Hz], and γ [22.5, 35 Hz], are very common in the field of quantitative EEG.

Let $S_{xx}(f_i)$ be the power spectral density computed at frequency f_i for the signal x. The relative power in the frequency band $[f_1, f_2]$ is given by:

$$P_{xx,\text{rel}}(f_1, f_2) = \frac{\sum_{f_i=f_1}^{f_2} S_{xx}(f_i)\cdot\Delta f}{P_{xx,\text{tot}}} \tag{13}$$

$\Delta f = F_s/N$, with F_s being the sampling rate and N the number of samples in x. $P_{xx,\text{tot}}$ is the total power in N_b different frequency bands:

$$P_{xx,\text{tot}} = \sum_{k=1}^{N_b} P_{xx,\text{rel}}(f_{1,k}, f_{2,k}) \tag{14}$$

Spectral edge frequency (SEF) can be seen as a robust summary of spectral activity, indicating the frequency value at which α% of the spectral power of a signal is obtained [23].

The SEF at level α% can be defined as:

$$\text{SEF}_\alpha = \max_f \{ f \mid P_{xx,\text{rel}}(0, f) \leqslant \alpha/100 \} \tag{15}$$

where $P_{xx,\text{tot}}$ is evaluated in the frequency band $[0, F_s/2]$ and F_s the sampling frequency of the signal.

3.2. Machine learning

A classification framework for automatic sleep–wake staging first requires designing a relevant machine

DOI: 10.1002/acs

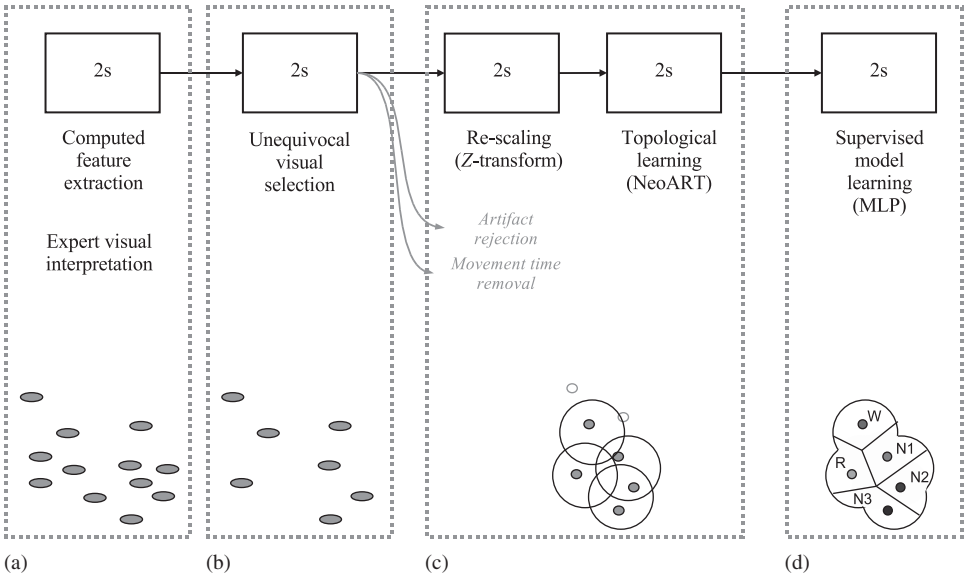

Figure 2. General architecture of the automatic sleep–wake stage learning scheme. Classifiers were obtained using: (a) an instance database of PSG-derived features extracted at short time intervals and their corresponding expert visual scores; (b) the selection of few representative exemplars discarding equivocal scoring epochs; (c) the attribute re-scaling and topological learning to reject outliers; and (d) the canonical model learning of sleep–wake stages using a supervised training.

learning scheme. The general architecture of the learning scheme implemented in the present study is summarized in the diagram of Figure 2. In parallel to an independent and visual interpretation of the sleep–wake stages by two medical experts, candidate features have been extracted automatically at short time intervals from a large database of PSG records. Subsequently, a limited subset of exemplars has been selected to train machine classifiers and achieve the instance-based learning of a canonical model associating PSG-derived features with unequivocal expert scores.

3.2.1. Instance selection. In order to minimize the impact on machine learning of inaccuracies related to the judgment of human experts, equivocal scoring epochs, i.e. 20 s epochs with a different score between the two experts, were discarded. The remaining data subset resulted in 69 585 non-equivocal scoring epochs and represented 82.8% of the total number of epochs scored by both experts. For each 2 s elementary epoch included in this unequivocal subset, the extracted candidate features and the corresponding expert score were gathered together into a new database containing 695 850 instances.

To train classifiers using typical exemplars, machine learning was further achieved after reducing the number of instances and pruning noisy ones. To do so, a limited set of elementary epochs representative of each sleep–wake stages was visually identified by an expert in 10 individual PSG records (s101t1, s101t2, s105t1, s121t2, s122t2, s123t2, s124t1, s125t2, s141t2, s142t2). The resulting 5 040 instances with each stage equally represented formed the exemplar learning set used in the present study.

3.2.2. Attribute scaling. To enable uniform handling of numerical data by the classification algorithm, all inputs were re-scaled to a smaller and similar range using the Z-transform. Transformed data were obtained using $Z=(X-\mu_j)/\sigma_j$, where μ_j is the mean and σ_j the standard deviation of the estimated feature j.

DOI: 10.1002/acs

3.2.3. Outlier rejection. In a final step prior to actual machine classification, an additional procedure was required to automatically clean the classifier inputs and reject outliers related to the presence of recording artifacts or movement time periods [12]. A NeoART artificial neural network based on the adaptive resonance theory [24] was used in this purpose.

NeoART topological learning consists of generating a dictionary of weighted attributes W according to a distance criterion. For each instance X in the exemplar learning set, distances were evaluated from each of the M generated weighted attributes W. The minimum distance d_i was compared with the threshold d_q. When attributes were below that threshold, they were considered close to W_i and the weighted instances were then adjusted using $W_i = W_i + a \cdot (X - W_i)$. Otherwise, new weighted attributes were generated again using $W_{M+1} = X$, and so on. In order to scale attributes according to the dimension of the input, the empirical threshold $d_q = M + 1.96 \cdot \mathrm{SD}$, with M being the median distance within attributes and SD the standard deviation, was defined with a a criterion of 10%.

3.2.4. Classifier training. Following topological learning, the classification of instances into sleep–wake scores was achieved using a supervised artificial neural network consisting of a feedforward back-propagation MLP.

A set of classifiers was generated according to an MLP architecture using three layers and a varying number of neurons in the hidden layer. The MLP included J neurons in the input layer, V neurons in the hidden layer, and K neurons in the output layer. The dimension J was imposed by the number of relevant features extracted from the analyzed signals while K corresponded to the number of outputs the classifier was intended to produce. The dimension V in the hidden layer was chosen after a series of tests including a varying number of neurons. Owing to random initialization, 10 MLP in each set were evaluated. Tangent sigmoid $(y = (2/(1+e^{2x})) - 1)$, linear $(y = x)$ and log-sigmoid $(y = 1/(1+e^{x}))$ transfer functions were defined for the input, the hidden, and the output layers, respectively. MLP weights were randomly initialized using the Nguyen–Widrow initialization function and optimization obtained using a Levenberg–Marquardt method [25].

An optimal architecture could be selected according to the average performance of the set of 10 MLP with different number of neurons in the hidden layer. MLP with 20 neurons in the hidden layer showed optimal performance [10] and were elected for use in the present study.

3.2.5. Feature selection. A selection of candidate features previously extracted from the analyzed 2 s signal intervals was achieved using a sequential forward selection (SFS) algorithm [26]. According to the minimum length description principle, the aim of this task was to obtain an optimal model with a minimal number of features to avoid overtraining and achieve better generalization performance when applied to unknown data.

The SFS algorithm starts searching the attribute space with an empty set of features, then tests instances according to the set of candidate features, and selects features one after each other by minimizing a given criterion. In the first step, if d disposable features are enabled (d was 16 in this study), SFS starts by learning d models with one feature (d MLP with one feature in the input layer) and selects the one with feature (i_{r_1}) that maximizes the performance criterion. At step two, SFS tests $d-1$ models constructed using candidate feature (i_{r_1}) and one of the $d-1$ remaining features. At the end of the process, d subsets are presented with their associated performances $(\{i_{r_1}\}, \{i_{r_1}, i_{r_2}\}, \ldots, \{i_{r_1}, i_{r_2}, \ldots, i_{r_d}\})$.

Analysis of the selection progression allowed defining a subset of features satisfying Occam's razor principle, which considers a model of minimal dimension d^* whose performance poorly improves when the model complexity further increases. SFS was achieved according to the algorithm provided in Appendix A.1.

3.3. Machine classification

In the first part of this study, an optimal set of PSG-derived features has been identified, and a machine classifier has been trained using a limited set of

DOI: 10.1002/acs

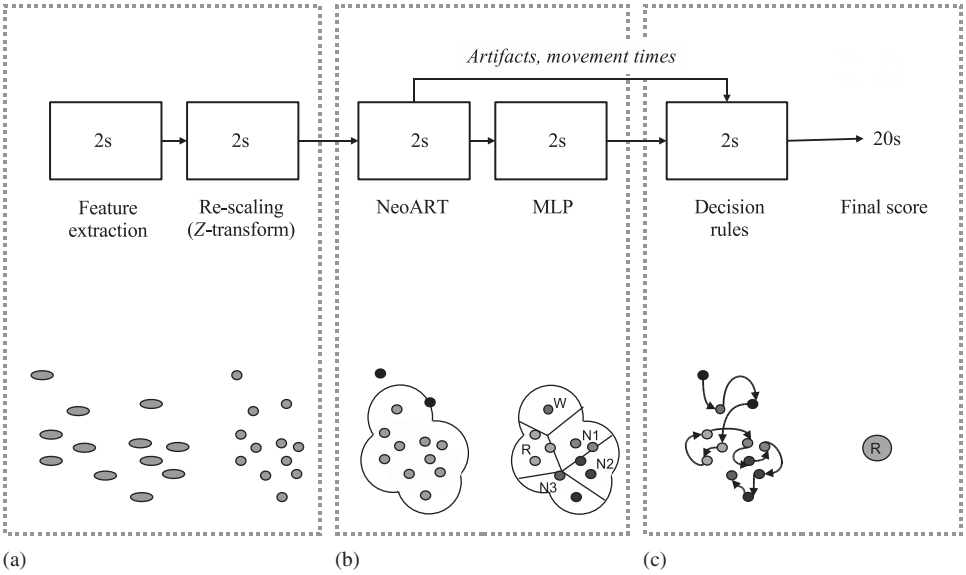

Figure 3. General architecture of the automatic sleep–wake stage classification scheme. Automated scoring was achieved by successively: (a) re-scaling an optimal subset of PSG-derived features extracted at short time intervals; (b) identifying outliers and classifying new instances using a pretrained machine classifier; and (c) integrating final scores using simple decision rules.

exemplars. A classification scheme was then implemented using the selected features and the pretrained MLP. The general architecture of the proposed classification scheme (Figure 3) includes three distinct processing modules. When applied successively, the three modules perform an automated data processing from the PSG signals to the final sleep–wake scores.

3.3.1. Preprocessing and classification. The preprocessing and classification modules (a and b in Figure 3) were implemented using the optimal set of candidate features previously identified from one EEG and one EMG signals by the SFS algorithm, and a pretrained classifier resulting from the initial machine learning scheme.

Together two preprocessing sub-modules enable the extraction and scaling of a set of attributes with an elementary temporal resolution of 2 s. An intermediate classification module subsequently identifies outliers such as recording artifacts and movement time periods using NeoART, and performs initial pattern recognition of sleep–wake stages using a trained and optimally performing MLP.

3.3.2. Post-processing. A post-processing module (c in Figure 3) was finally introduced into the classification scheme and implemented as a set of inference rules allowing integration of the classifier-generated outputs into final sleep–wake scores. In addition to temporal flexibility, this decisional layer was supposed to compensate for the effect of abrupt changes in successive classifier outputs.

The present study used a rudimentary decision rule consisting of a majority vote inspired by the 50% rule of the standard staging criteria [1]. Such a rule states that when more than one sleep or wake stages are present in a given epoch, the score is that of the stage occupying most of the time in the actual epoch. Here, among k successive 2 s elementary scores, a final score was obtained every $k \cdot 2$ s. With an actual value of $k = 10$, a decision was generated every 20 s allowing further comparisons with the experts. In case of equality, a decision was drawn uniformly from conflictual outputs.

DOI: 10.1002/acs

3.4. Performance assessment

The entire PSG records of all individual were submitted to full automatic analysis according to the previously described classification scheme. Final sleep–wake scores obtained automatically were then compared to those interpreted unequivocally by the two experts.

To estimate the error rate, a concordance matrix P, with the intra- and inter-class agreements P_{ij}, was computed on the automatic and expert score series. The individual error rate P_e for each record was used in evaluating individual performance. Global concordances, obtained by summing up the individual matrix values, provided an evaluation of the general performance of the classification scheme for each sleep–wake stage.

The Cohen's Kappa coefficient, which takes into account agreement obtained by pure chance [27], finally ensured statistical assessment of the machine classification. By doing so, random classifications were indicated by coefficient values close to zero and non-random ones by values not far from one. The statistics was computed using:

$$\kappa = \frac{f_0 - f_E}{N - f_E} \tag{16}$$

with f_0 being the frequency of diagonal observations due to classification, f_E the expected frequency of diagonal due to chance, and N the number of observations:

$$\begin{aligned} f_0 &= \sum_{i=j} P_{ij}, \quad P_i = \sum_j P_{ij} \\ P_j &= \sum_i P_{ij}, \quad f_E = \frac{1}{N} \sum_{i=j} P_i \cdot P_j \end{aligned} \tag{17}$$

4. RESULTS

4.1. Feature selection

The results of candidate feature selection obtained in applying the SFS algorithm to a limited set of unequivocal exemplars are illustrated in the performance curve of Figure 4.

An optimal set was obtained with five features. As indicated in Table I, $\mathrm{Mob}_{\mathrm{EEG}}$ was the first candidate selected with an empirical error rate of 43%. Candidate features $\mathrm{Act}_{\mathrm{EEG}}$, $\mathrm{SEF}_{95,\mathrm{EMG}}$, $P_{\beta,\mathrm{rel},\mathrm{EEG}}$, and $P_{\sigma,\mathrm{rel},\mathrm{EEG}}$ were then selected consecutively, achieving an optimal error rate of 22%, after which adding more features did not substantially improve performance.

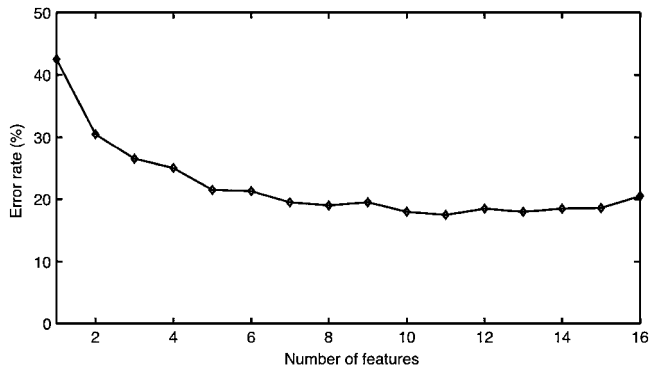

Figure 4. Performance curve resulting from the sequential forward selection algorithm during candidate feature search. Progression of the classification global error is given as a function of the feature subset dimension.

4.2. Classification performance

4.2.1. Inter-expert agreement. The global agreement in the sleep–wake stage scoring between the two experts involved in the present study was 83±3%. The inter-expert agreement and the sources of error for each sleep–wake stages are provided in Table II. The highest expert classification agreements were obtained for sleep stages R, N_3, and N_2, respectively. The intra-class agreements for stages MT, W, and N_1 were the lowest, each of these stages being mostly misclassified between each other.

4.2.2. Global performance. The global error rate calculated from the concordance matrix given in Table III was 42%. The relative error rate, i.e. the mean error rate by class, was 49%.

Wakefulness (W) showed a correct machine classification of 34% only and was misclassified mostly as stages N_1 (41%) and R (13%). Transitional sleep stage N_1, with 43% of correct machine classifications, was confounded with stage R (48%). Shallow sleep stage N_2, displayed a classification agreement of 51% and was confounded mostly with stages N_1 (12%) and

DOI: 10.1002/acs

Table II. Concordance matrix between the scoring of sleep–wake stages by the two human experts involved in the study.

		Expert AB					
		W (%)	N_1 (%)	N_2 (%)	N_3 (%)	R (%)	MT (%)
Expert ES	W	80	10	1	0	2	7
	N_1	13	45	20	0	16	6
	N_2	0	1	91	6	1	1
	N_3	0	0	7	93	0	0
	R	0	2	3	0	94	1
	MT	7	1	6	0	1	85

Table III. Concordance matrix between machine and unequivocal expert scoring.

		Machine					
		W (5%)	N_1 (13%)	N_2 (28%)	N_3 (16%)	R (36%)	MT (2%)
Experts	W (4%)	34%	41%	2%	1%	13%	9%
	N_1 (3%)	4%	43%	3%	2%	48%	0%
	N_2 (46%)	2%	12%	51%	5%	31%	0%
	N_3 (16%)	2%	0%	13%	82%	3%	0%
	R (21%)	1%	15%	1%	1%	82%	0%
	MT (10%)	22%	16%	17%	7%	26%	13%

R (31%). Deep sleep stage N_3 was classified with a high rate of correct classification (82%). It was mostly misclassified with stage N_2 (13%). Paradoxical sleep stage (R), also with a high rate of correct classification (82%), was confounded with stage N_1 and displayed a low error rate of only 15%. Epochs scored as movement time (MT) were poorly classified (13% of agreement) and misclassified as stages R (26%), W (22%), N_2 (17%), and N_1 (16%).

4.2.3. Individual performance. All individual results are provided in Table IV. The individual error rates, without consideration of the stage distribution, averaged $36 \pm 15\%$ with a κ of 0.48 ± 0.18. The best classification was achieved with an error rate of only 14%. This PSG record ($s105t6$) had a small number of stage N_2 epochs (42%), which were scarcely misclassified as stage R (12%). The lowest individual agreement was obtained for record $s141t7$ with an error rate of 66%. This record displayed an important number of stage N_2 epochs (62%), among which 58% were misclassified as stage R.

5. DISCUSSION

Few studies on automatic sleep–wake staging have been undertaken using a database as large as the one used in the present study. To the best of our knowledge, none of them have ever consider the applicability of automatic analysis to various signals from heterogeneous PSG recording devices. According to the methods developed and the results presented in this study, evaluating the system performance was not trivial and was thus considered in two stages. On the one hand, we reported the performance observed in the learning stage while optimizing the classifier inputs and architecture. On the other hand, the accuracy achieved while automatically scoring all PSG records from the whole database with an epoch duration similar

DOI: 10.1002/acs

Table IV. Individual classification performance obtained with all records of the PSG database: ID, P_e and κ represents the record identifier, the error rate and the Cohen's κ coefficient, respectively.

ID	Pe (%)	κ	ID	Pe (%)	κ
s101t1*	49	0.3	s101t2*	36	0.4
s101t6	36	0.4	s101t7	45	0.3
s105t1*	16	0.7	s105t2	14	0.7
s105t6	14	0.7	s105t7	15	0.7
s121t1	45	0.4	s121t2*	60	0.2
s121t6	22	0.6	s121t7	28	0.5
s122t1	27	0.5	s122t2*	33	0.5
s122t6	24	0.6	s122t7	22	0.6
s123t1	47	0.3	s123t2*	33	0.5
s123t6	47	0.3	s123t7	30	0.5
s124t1*	32	0.4	s124t2	51	0.3
s124t6	45	0.4	s124t7	45	0.3
s125t1	27	0.6	s125t2*	31	0.5
s125t6	21	0.6	s141t1	58	0.2
s141t2*	54	0.2	s141t6	61	0.1
s141t7	66	0.1	s142t1	51	0.3
s142t2*	42	0.4	s142t6	29	0.5
s142t7	38	0.4	s143t6	14	0.7
s143t7	22	0.6	s144t1	63	0.1
s144t2	58	0.2	s144t6	43	0.3
s145t1	54	0.2	s145t2	42	0.3
s145t6	27	0.6	s145t7	36	0.4
s146t1	22	0.6	s146t2	24	0.6
s146t6	14	0.7	s146t7	21	0.6

PSG records used in the learning scheme are indicated by *.

to that of the experts was presented in an attempt to understand the different sources of error. Finally, potential alternatives to improve the performance and applicability of the proposed classification framework were suggested in the view of the results obtained.

Based on empirical evaluation of the classifier accuracy in the machine learning stage, the observed performance showed a global error rate of 22%. Accordingly, one could conclude that automatic classification showed an agreement close to that obtained between experts ($83 \pm 3\%$). However, the data set used in this preliminary stage was the same than the one used for training. This unrealistic assessment was however useful in achieving a selection of relevant candidate features using the SFS algorithm. Indeed, limiting the number of features used as classifier inputs not only avoids overtraining but also accelerates their extraction, which represents the most time-consuming task in automated analysis (on a standard personal computer, approximately 15 min per 8 h of PSG record with all the 16 candidate features).

In practice, it is important to note that PSG makes use of numerous recording device brands and models with various hardware characteristics and acquisition settings. Also, an adequate bio-calibration of the collected signals is not always achieved correctly, and changes in the electrode impedance may occur throughout the recording, which may lead to uncontrolled changes in signal amplitudes. In the present study, we have tried as far as possible to develop a system that could operate independently of such heterogeneity in signal conditioning by considering appropriate features. Although this property remains to be demonstrated using PSG records collected with multiple devices, encouraging results were obtained.

The three EEG descriptors elaborated by Hjorth (Act, Mob and Cpx) have been included in our set of candidate features. However, activity (Act) can greatly

DOI: 10.1002/acs

differ when computed from the raw signals of recording devices operating at various sampling frequencies. The inclusion of this feature might well adversely impact robustness of the method and should thus be avoided. The possibility to compute Hjorth's descriptors by estimating the frequency moments [22] would represent another alternative. The features sample entropy (SaE) and the Hurst exponent (H) were of particular interest since they are dimensionless. Here, they have been partially optimized for computation on EEG and EMG signals, but were not selected in the SFS search. In our implementation, the estimation of these features could still be biased by signal quantization and sampling frequency, and, consequently, should be re-evaluated in further studies. As a general principle, device-independent estimation of features describing PSG signals should require operating at constant quantization scale and sampling rate. This could easily be achieved by re-sampling and re-quantisizing all signals at lower resolutions, for example, at 100 Hz on 8 bits, before classification.

With respect to the order of candidate selection in the selective forward search, the optimal set of features identified in the present study consisted of $\mathrm{Mob}_{\mathrm{EEG}}$, $\mathrm{Act}_{\mathrm{EEG}}$, $\mathrm{SEF}_{95,\mathrm{EMG}}$, $P_{\beta,\mathrm{rel,EEG}}$, and $P_{\sigma,\mathrm{rel,EEG}}$. In a concurrent study [28], a subset of different features ($P_{\beta,\mathrm{rel,EEG}}$, $\mathrm{Mob}_{\mathrm{EMG}}$, $P_{\alpha,\mathrm{rel,EEG}}$, $P_{\sigma,\mathrm{rel,EEG}}$, $\mathrm{ShE}_{\mathrm{EEG}}$, $\mathrm{ShE}_{\mathrm{EMG}}$, and $P_{\theta,\mathrm{rel,EEG}}$) has been selected from a combination of EEG and EMG signals. Performance of both studies was surprisingly similar with a reported accuracy rate of 78 versus 80%, respectively. Together, these results obtained from the same PSG records reveal the redundancy and the competing nature of some features. In our optimal feature set, Hjorth's activity ($\mathrm{Act}_{\mathrm{EEG}}$), which varies according to changes in the analyzed signal amplitude, probably accounted for the recognition of deep sleep (stage N_3) characterized by the appearance of high-amplitude EEG slow waves. In the study of Zoubek *et al.* it might well have been Shannon's entropy ($\mathrm{ShE}_{\mathrm{EEG}}$), which displays highest levels during deep sleep. The two features commonly selected in both studies were relative EEG powers in the sigma and beta frequency bands. From a neurophysiological point of view, the former is a well-known marker of the occurrence of sleep spindles, which represent a salient feature of shallow sleep (stage N_2), while the latter reflects bioelectrical activity in the high-frequency range characterizing activated brain states such as wakefulness and paradoxical sleep (stage W and R, respectively).

The innovative aspect of our study was to take into account rapid changes in PSG signals and, by doing so, the micro-structural aspect of sleep. Indeed, from our combined approach using short-time feature extraction and additional decision rules, new properties emerged from the classification system, and notably that of producing results at a flexible time resolution matching various expert requirements (human and animal researchers as well as clinicians). Because of both the rapid alternance of vigilance states and the presence of intermingling recording artifacts, the selection of short learning exemplars representative of each sleep–wake stages from the instance database was easier than that of larger epochs. Actually, the careful selection of a small amount of 2 s learning epochs was a key factor in achieving good performance. Given its cost-sensitive nature, our learning scheme ensured the acquisition of relevant structural descriptions from the feature space and prevented discovering spurious, contingent or accidental irregularities. On one PSG record, the individual error rate was as low as 14% indicating that the method could work quite fairly. However, as reflected in the testing stage by a large global error rate (42%) obtained using the unequivocal instance database, our classifier training apparently suffered from a limited set of exemplars, which could have been insufficient considering a large inter-individual variability in EEG signals. Consequently, the accuracy of our classification method may probably benefit from a larger number of individuals in the learning set.

Another crucial aspect of this study was the implementation of a self-organized module (NeoART, see Section 3.2.3) to process attributes and reject outliers before they served as inputs to the actual sleep–wake stage classifier. Using an adaptive threshold, this module detects extreme attribute values and, by doing so, enables the identification of elementary epochs contaminated by recording artifacts and the subsequent detection of movement times (MT). Actually, while the experts scored an average of 10% of MT epochs in the testing set, only 2% were obtained from automatic analysis, with misclassifications distributed in

DOI: 10.1002/acs

almost every other stages, and mostly in wakefulness, transitional, shallow and paradoxical sleep (22, 17, 16, and 26% for W, N_1, N_2, and R, respectively). Many outliers were thus automatically filtered out, the experts being more selective than the machine in scoring MT each time a movement-related artifact of short duration occurred in the actual 20 s epochs. While the standard criteria for staging MT are ill-defined in the existing guidelines, this problem could be addressed by introducing a specific post-processing rule for MT.

The analysis of the concordance matrices revealed that expert-scored wakefulness epochs were classified not only as MT, but also as transitional and paradoxical sleep, with error rates of 9, 41, and 13%, respectively. It has to be noted that wakefulness and movement time were scored by the experts with an accuracy in the 80% range only. In fact, the stage W described in the sleep scoring guidelines brings together multiple behavioral states such as active wakefulnes and passive wakefulness with the eyes either opened or closed. Distinct EEG and EMG patterns characterize these different states, which most probably explains the errors.

To a similar extent, fully automated analysis demonstrated similar issues in discriminating transitional from paradoxical sleep, with reciprocal error rates of 48 and 15% between stages N_1 and R. As surprisingly as it could be, the inter-expert scoring agreement for stage N_1 was as low as 45%. From a waveform point of view, these two sleep stages exhibit similar activated EEG patterns not easily discriminated even by an expert. Whether our approach of excluding the EOG channel could have decreased classification accuracy remains questionable. Conflicting results have been published in this respect with reports of either an excellent performance in classifying sleep stages without EOG [6], or of a 25% increase in the discrimination accuracy of transitional sleep without impacting that of paradoxical sleep [13]. However, the low accuracy of our system in classifying transitional sleep indicates the lack of an adequate feature, derived from either an EEG or EOG channel.

While deep sleep (stage $N3$) and paradoxical sleep (stage R) were both classified with levels of agreement comparable to that of the two human experts (82 versus 93% and 82 versus 94%, respectively), our system was relatively inaccurate in classifying shallow sleep (stage N_2), which was misclassified as transitional sleep (12%) and paradoxical sleep (31%). According to the individual error rates, it was clear that the individual sleep structure accounted for most errors. Thus, our classifier with poor performance for stage N_2 led to poor performance when applied to individuals sleeping dominantly in this stage. Shallow sleep is a very heterogeneous stage that is frequently scored when all the criteria for other sleep–wake stages are not fully met. Also, when assessed in 2 s epochs, shallow sleep shows many transitions to or from other sleep stages, but the simple inference rule used in the present study was not designed to account for such rapid changes.

In this study, we have chosen to classify and post-process 2 s elementary segments of PSG recorded EEG and EMG signals to fully automate the scoring of sleep and wake stages. While temporal segmentation [29–31] could have been used in this regard, a flexible decisional module ensured suitability for both human and animal studies, which use various but fixed scoring epoch durations (4, 5, 10, 20, 30, or 60 s). Even though a simple inference rule alone was expected not to perform optimally, a majority vote has been retained in this proof-of-concept study. Increasing the performance of the proposed classification scheme can reasonably be expected by including additional decision rules based, for example, on the use of contextual information.

Various biomedical applications involving human expertise in pattern recognition of time series generated by linear or non-linear biological processes could benefit from the method developed in the present study. Similar machine learning and classification schemes have indeed been used in brain–computer interfaces [32]. In the light of our results, the machine classification of signals of different nature, such as accelerometric and magnetic data in the analysis of motor manifestations related to degenerative diseases or epileptic seizures [33], and vital signs for patient monitoring and hemorrhagic shock detection [34], would, for example, also merit further developments.

6. CONCLUSIONS

We have shown that combining short-time feature extraction, supervised artificial neural networks, and

DOI: 10.1002/acs

simple decision rules provides a flexible classification framework capable of achieving appreciable performance in the automation of sleep–wake stage scoring. Further research and optimization in this direction, with the incorporation of additional domain-knowledge using more sophisticated expert-based rules, may well lead to increasing performance up to a satisfactory level of applicability on heterogeneous PSG records.

APPENDIX

A.1. Sequential forward search algorithm

```
Start
  Let I_r := {∅}, indices set of retained features
  Let I_d := {1, ···, d}, indices set of disposable features
While (I_d ≠ {∅}) do
  For i ∈ I_d do
    I_c(i) := I_r ∪ i, indices set of running features to
      evaluate
    Create S_{I_c(i)}, set of examples with features in I_c(i)
    Compute Ξ(i), performance criterion for I_c(i)
  End For
  Select i_r that minimizes Ξ
  Update :
  I_r := I_r ∪ i_r
  I_d := I_d \ i_r
End While
End
```


ACKNOWLEDGEMENTS

The authors thank PhiTools SARL (www.phitools.com) for funding Dr Guillaume Becq postdoctoral studies and Dr Raymond Cespuglio for providing office space. This work was based on the clinical sleep scoring expertise of Drs Alain Buguet and Emilia Sforza.

REFERENCES

1. Rechtschaffen A, Kales A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. US Government Printing Office: Washington, 1968.
2. American Academy of Sleep Medicine. *AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine: Westchester, IL, 2007.
3. Schaltenbrand N, Lengelle R, Macher JP. Neural network model: application to automatic analysis of human sleep. *Computers and Biomedical Research* 1993; **26**:157–171.
4. Chapotot F, Becq G, Sforza E, Buguet A. Automatic sleep-wake stage scoring sing artificial neural networks: optimisation and evaluation. *Journal of Sleep Research* 2004; **13**(1):132.
5. Anderer P, Gruber G, Parapatics S, Woertz M, Miazhynskaia T, Klosch G, Saletu B, Zeitlhofer J, Barbanoj MJ, Danker-Hopfe H, Himanen SL, Kemp B, Penzel T, Grozinger M, Kunz D, Rappelsberger P, Schlogl A, Dorffner G. An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the somnolyzer 24×7 utilizing the Siesta database. *Neuropsychobiology* 2005; **51**(3):115–133.
6. Berthomier C, Drouot X, Herman-Stoica M, Berthomier P, Prado J, Bokar-hire D, Benoit O, Mattout J, d'Ortho M-P. Automatic analysis of single-channel sleep EEG: validation in healthy individuals. *Sleep* 2007; **30**(11):1587–1595.
7. Timo-Iaria C, Negrao N, Schmidek WR, Hoshino K, Lobato de Menezes CE, Leme da Rocha T. Phases and states of sleep in the rat. *Physiology and Behavior* 1970; **5**(9):1057–1062.
8. Penzel T, Conradt R. Computer based sleep recording and analysis. *Sleep Medicine Reviews* 2000; **4**(2):131–148.
9. Robert C, Guilpin C, Limoge A. Review of neural network applications in sleep research. *Journal of Neuroscience Methods* 1998; **79**:178–193.
10. Becq G, Charbonnier S, Chapotot F, Buguet A, Bourdon L, Baconnier P. Comparison between five classifiers for automatic scoring of human sleep recordings. *Softcomputing in Knowledge Discovery Methods and Applications*: *Part B—Classification and Clustering for Knowledge Discovery*, Halgamuge S, Wang L (eds). Springer: Berlin, 2005; 113–127.
11. Becq G, Chapotot F, Sforza E, Buguet A, Cespuglio R. Automatic sleep–wake scoring combining artificial neural networks, feature extraction and post-processing inference rules. *EU 6th Framework Integrated Project SENSATION, International Conference on Monitoring Sleep and Sleepiness—From Physiology to New Sensors*, Basel, Switzerland, 2006.
12. Schaltenbrand N, Lengelle R, Toussaint M, Luthringer R, Carelli G, Jacqmin A, Lainey E, Muzet A, Macher JP. Sleep stage scoring using neural network model: comparison between visual and automatic analysis in normal subjects and patients. *Sleep* 1996; **19**(1):26–35.
13. Zoubek L, Charbonnier S, Lesecq S, Buguet A, Chapotot F. Feature selection for sleep–wake stages classification using data driven methods. *Biological Signal Processing and Control* 2007; **2**:171–179.
14. Chapotot F, Pigeau R, Canini F, Bourdon L, Buguet A. Distinctive effects of modafinil and d-amphetamine on the homeostatic and circadian modulation of the human waking. *EEG and Psychopharmacology* 2003; **166**: 127–138.
15. Kemp B, Varri A, Rosa AC, Nielsen KD, Gade J. A simple format for exchange of digitized polygraphic recordings.

DOI: 10.1002/acs

Electroencephalography and Clinical Neurophysiology 1992; **82**(5):391–393.
16. Jasper HH. Appendix to report to committee on clinical examination in EEG: the ten–twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology* 1958; **10**:371–375.
17. Shannon CE. A mathematical theory of communication (Part 1). *Bell System Technical Journal* 1948; **27**:379–423.
18. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology—Heart and Circulatory Physiology* 2000; **278**:H2039–H2049.
19. Acharya R, Faust O, Kannathal N, Chua T, Laxminarayan S. Non-linear analysis of EEG signals at various sleep stages. *Computer Methods and Programs in Biomedicine* 2005; **80**:37–45.
20. Natarajan K, Acharya R, Alias F, Tiboleng T, Puthusserypady SK. Nonlinear analysis of EEG signals at different mental states. *BioMedical Engineering OnLine* 2004; **3**(7):1–11.
21. Rolo-Naranjo A, Montension-Otero ME. A method for the correlation dimension estimation for on-line condition monitoring of large rotating machinery. *Mechanical Systems and Signal Processing* 2005; **19**:939–954.
22. Hjorth B. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology* 1970; **29**:306–310.
23. Schwender D, Daunderer M, Mulzer S, Klasing S, Finsterer U, Peter K. Spectral edge frequency of the electroencephalogram to monitor depth of anaesthesia with isoflurane or propofol. *British Journal of Anaesthesia* 1996; **77**:179–184.
24. Carpenter GA, Grossberg S. The ART of adaptive pattern recognition by a self-organizing neural network. *Computer* 1988; **21**:77–88.
25. Demuth H, Beale M, Hagan M. *Neural Network Toolbox User's Guide*. The MathWorks Inc.: Natick, MA, U.S.A., 2006.
26. Fukunaga K. *Introduction to Statistical Pattern Recognition*. Academic Press: New York, 1972.
27. Howell DC. Méthodes statistiques en sciences humaines. De Boeck, translated from *Statistical Methods for Psychology*. Language Learning, 1997.
28. Zoubek L, Charbonnier S, Lesecq S, Buguet A, Chapotot F. A two-step sleep/wake stages classifier taking into account artefacts in the polysomnographic signals. *Proceedings of the 17th World Congress, The International Federation of Automatic Control*, Seoul, Korea, 6–11 July 2008; 5227.
29. Agarwal R, Gotman J. Computer-assisted sleep staging. *IEEE Transactions on Biomedical Engineering* 2001; **48**(12): 1412–1423.
30. Feng L, Ju K, Chin KH. A method for segmentation of switching dynamic modes in times series. *IEEE Transactions on Systems, Man and Cybernetics Part B* 2005; **35**(5): 1058–1064.
31. Charbonnier S, Becq G, Biot L. On-line segmentation algorithm for continuously monitored data in intensive care units. *IEEE Transactions on Biomedical Engineering* 2004; **51**(3):484–492.
32. Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG based brain-computer interfaces. *Journal of Neural Engineering* 2007; **4**:R1–R13.
33. Becq G, Bonnet S, Minotti L, Antonakios M, Guillemaud R, Kahane P. Collection sand exploratory analysis of attitude sensor data in an epilepsy unit. *Proceedings of the 29th IEEE Engineering in Medicine and Biology Society* (*EMBS*) *Annual International Conference*, vol. 1, Lyon, France, 2007; 2775–2778.
34. Becq G, Charbonnier S, Bourdon L, Baconnier P. Evaluation of a device scoring classes of hemorrhagic shock. *Twenty-Sixth IEEE EMBS Annual International Conference*, vol. 1, San Francisco, SA, 2004; 470–473.

DOI: 10.1002/acs