Computers in Biology and Medicine ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging

S. Charbonnier [a,*], L. Zoubek [b], S. Lesecq [c], F. Chapotot [d]

[a] Gipsa-lab, Control System Department, BP 46, F-38 402 Saint Martin d'Hères Cedex, France
[b] Department of Information and Communication Technologies, Pedagogical Faculty, University of Ostrava, Českobratrská 16, 701 03 Ostrava, Czech Republic
[c] CEA-LETI MINATEC, 17 rue des Martyrs, F-38054 Grenoble Cedex 9, France
[d] Department of Medicine, University of Chicago, Chicago, IL, USA

ARTICLE INFO

Article history:
Received 23 February 2010
Accepted 1 April 2011

Keywords:
Decision support system
Sleep/wake stages
Neural network
Classification

ABSTRACT

An automatic sleep/wake stages classifier that deals with the presence of artifacts and that provides a confidence index with each decision is proposed. The decision system is composed of two stages: the first stage checks the 20 s epoch of polysomnographic signals (EEG, EOG and EMG) for the presence of artifacts and selects the artifact-free signals. The second stage classifies the epoch using one classifier selected out of four, using feature inputs extracted from the artifact-free signals only. A confidence index is associated with each decision made, depending on the classifier used and on the class assigned, so that the user's confidence in the automatic decision is increased. The two-stage system was tested on a large database of 46 night recordings. It reached 85.5% of overall accuracy with improved ability to discern NREM I stage from REM sleep. It was shown that only 7% of the database was classified with a low confidence index, and thus should be re-evaluated by a physiologist expert, which makes the system an efficient decision-support tool.

1. Introduction

Polysomnography is a diagnostic method used to analyze human sleep. It consists in simultaneously recording several physiological parameters during a whole night sleep. The set of measured signals is denoted as a polysomnographic (PSG) recording. The most relevant signals are the electroencephalogram (EEG), the electrooculogram (EOG) and the electromyogram (EMG). The analysis of the polysomnographic recordings aims at classifying the whole recording into a succession of sleep/wake stages, to obtain a hypnogram. The hypnogram provides an overall representation of the sleep architecture and presents the chronological distribution of the sleep/wake stages. It reveals the internal architecture of sleep and the alternation of NREM and REM sleep phases, which makes the discrimination between normal and abnormal sleep much simpler. PSG is thus a powerful tool in the diagnosis of sleep disorders, which are rather common with about 5% of the general population affected [1].

To build the hypnogram, the physician or medical expert typically splits the polysomnographic recordings into segments of constant duration (named epochs), generally of 20 or 30 s length. Each epoch is classified into sleep/wake stages using rules initially proposed in the conventional Rechtschaffen and Kales (R&K) human sleep/wake stage scoring manual [2] and recently updated by The American Academy of Sleep Medicine. In general, six stages are recognized in human sleep, namely, wakefulness, non-rapid eye movement (NREM) sleep stages I, II, III and IV and REM sleep or paradoxical sleep (PS). NREM stages III and IV represent the slow wave sleep (SWS) which is why they frequently form one united stage.

Since 1970 and the growth of computerized methods, automated systems have emerged in order to automatically score PSG recordings, so as to avoid the expert to spend too much time on this tedious and time-consuming work. The R&K visual interpretation of PSG recordings is a typical pattern recognition task [3]. Physicians look at the signals and classify successive epochs from the shape, frequency, amplitude and monotony of their traces. Special attention is given to the identification of artifacts polluting the signals that may lead to misclassifications. Artifacts are modifications observed on a signal. They are caused by external causes, such as interferences from another biological signal monitored or environmental perturbations. They must not be confused with a real physiological evolution.

The field of artificial intelligence provides a broad range of methods and algorithms that were tested in the last years in order to propose reliable automatic systems of sleep/wake stages classification. Several studies focused on the selection of the appropriate classifier, most often a neural network, with various

* Corresponding author.
E-mail address: Sylvie.Charbonnier@gipsa-lab.grenoble-inp.fr (S. Charbonnier).

0010-4825/$ - see front matter © 2011 Published by Elsevier Ltd.
doi:10.1016/j.compbiomed.2011.04.001

features used as inputs (for instance [4–9]) Classifiers using contextual information, such as Hidden Markov Models, were also proposed [10,11]. Methods to process the polysomnographic signals were presented in [12–14]. Grozinger et al. [13] assessed the improvement brought by some non-conventional EEG features, such as the highest Lyapunov exponent, the correlation dimension and the spectral entropy for classification of REM sleep. Zoubek et al. [15] suggested applying data mining methods to select the most appropriate signal processing methods.

Though several classifiers were proposed in the last decades, which showed interesting classification rates, none of them has been widely accepted by the medical community, probably because of a lack of confidence in the decision made by the systems. In our opinion, several problems should be solved to obtain automatic classifiers able to obtain results similar to human experts and accepted by the medical community. First, systems should be able to deal with the presence of artifacts in the polysomnographic signals, in the same way as the expert does. Indeed, the presence of artifacts can generate inappropriate numerical values of the features extracted which may lead to classification errors. Then, the polysomnographic signals should be processed with adequate techniques so as to obtain inputs to the classifier (i.e. features) which are the most similar to the visual information used by the expert. Finally, to increase the expert confidence in the system decision, the system should be able to evaluate the reliability of each decision it makes and provide a confidence index associated with each decision. In this way, the expert could use the automatic classifier as a decision-support tool. He could accept the automatic decisions made with a high confidence index and visually analyze and classify the epochs classified with a lower confidence index.

In this paper, a sleep/wake stages classification system that deals with the possible presence of artifacts and that provides an auto evaluation of the decision it makes is proposed. An analysis of the quality of the polysomnographic signals (EEG, EOG, EMG) is performed before using specific classifiers. The proposed system consists of two processing stages (Fig. 1). For each epoch to be classified, the polysomnographic signals are at first analyzed to detect artifacts and consequently, the signals that can be used for the classification are selected. Then, the features are extracted and classified into sleep/wake stages, using one classifier selected among a set of four different classifiers. The classifiers differ one from the others by the input signals they use. The idea is to use a different classifier for each epoch to be classified, depending on the polysomnographic signals available, i.e. the artifact-free signals. During the classifiers conception phase, data mining techniques were applied to select the most relevant features for each classifier. Finally, the classifier classifies one epoch of signals into the five sleep/wake stages and provides information on the reliability of the decision, through a confidence index computed from the performances obtained by each classifier on a control database. The value of this index depends on the phase recognized and on the classifier used to make the decision.

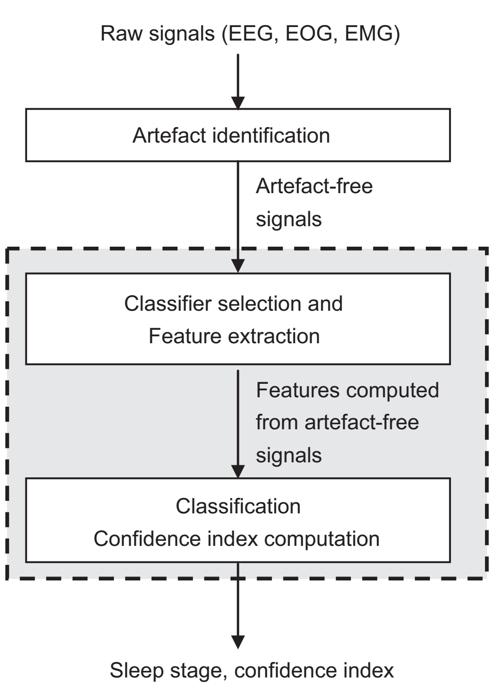

Fig. 1. Scheme of the two-stages classification system.

The outline of this paper is the following. In Section 2, the data set used in this study is presented. In Section 3, the two stages of the classifier proposed are described. Finally, the results are presented and discussed in Section 4.

2. Materials

A large database of polysomnographic recordings has been used for this study. The full database contains 46 night-time polysomnographic recordings obtained from 13 healthy adult subjects (19- to 47-year-old). Recordings were made continuously during the night sleep (typically 8 h between 22:00 h and 06:00 h). Each polysomnographic recording contains four electroencephalogram (EEG) channels (C3-A2, P3-A2, C4-A1 and P4-A1), one transversal electrooculogram (EOG), one chin electromyogram (EMG) and one electrocardiogram (ECG). The analog signals were then digitized with an 8-bit A/D converter at the sampling frequency fs = 128 Hz. Only the EEG C3-A2 channel, the EOG and the EMG signals were analyzed. The EEG leads were attached onto the scalp according to the International 10-20 EEG System of Electrodes Placement [16]. The protocol of the investigation is described in [17].

All the 46 PSG recordings were visually scored by two independent sleep physicians. Visual sleep/wake stage scoring was performed with constant epoch duration of 20 s according to the conventional rules of the R&K manual [1]. Each epoch was thus classified into one of the five sleep/wake stages: wakefulness, NREM sleep stage I, NREM sleep stage II, slow wave sleep and REM sleep.

To avoid the introduction of expert inaccuracies in the database, only epochs identically classified by both experts were considered. Consensual epochs represented 87% of the original PSG recording database. Therefore, only that subset was used to form our study database. The total number of epochs included for the study was 66 164. Its repartition in the five sleep stages is presented in Table 1.

3. Presentation of the two-stage system

3.1. Classification strategy

An automatic classifier is a decision system whose input is a predefined set of features, extracted from monitored signals.

Table 1
Repartition of the epochs in the five sleep stages.

Number of epochs	**Awake state**	**NREM I**	**NREM II**	**SWS**	**REM**
Both experts-test	5298	1981	32 462	11 210	15 213

Please cite this article as: S. Charbonnier, et al., Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging, Comput. Biol. Med. (2011), doi:10.1016/j.compbiomed.2011.04.001

When an artifact is detected by an identification procedure carried out on the signal monitored separately from the classification task, some parts of the signals are suppressed. It results in missing values in the input set of features, which is an issue most classifiers are unable to handle. A two-stage classification system is proposed here as a solution to deal with missing values.

In the first stage, the three signals (EEG, EOG and EMG) are analyzed to determine if any artifact is present in the epoch to be classified. If only a part of the epoch is artifacted, features are calculated using parts of the signal that are not artifacted. If too large a part of a signal is artifacted during the epoch, the signal is removed from the classification process. Thus, the number of signals used for classification varies from one epoch to the other.

In the second stage, the decision is made using the artifact-free signals only. The decision system is composed of four neural networks. The four neural networks use different inputs (the so-called features), extracted from different combinations of signals, respectively, EEG only, EEG and EOG, EEG and EMG, EEG and EOG and EMG. For each epoch to classify, the appropriate classifier is selected. When all the signals are detected to be artifact-free, the EEG–EOG–EMG classifier is used. If the EOG and/or EMG signals are artifacted and cannot be used (at all), the classification is carried out by the classifier that does not use this signal. Let us note that the EEG signal is known to be crucial for sleep/wake staging [4]. Thus, if the EEG signal is fully artifacted, no classification can be made.

With each decision, a confidence index is provided. It depends on the classifier used and on the class recognized. The confidence index value is defined from the performances obtained by each classifier in the elaboration phase, using a control data set.

The structure of the automatic system is shown in Fig. 1. To summarize the two-stage classification system, for each epoch, an artifact identification procedure is first achieved and the combination of artifact-free signals is determined. Then, the appropriate neural network classifier is selected from the set of classifiers, the relevant features are extracted from the artifact-free signals and the epoch is scored with a confidence index.

3.2. Artifact identification stage

In this stage, artifact detection methods inspired by [18] and [19] are implemented. To reduce the amount of data lost, the time resolution of the artifact detection algorithm is reduced from 20 to 2 s. Each original 20-s epoch is split into a succession of ten 2 s sequences and each 2-s segment is analyzed to detect the occurrence of an artifact. A lucid strategy is employed to decide whether the entire 20-s epoch is to be marked as artifacted or not. If more than 20% of the epoch duration contains any kind of artifact then the entire 20-s epoch is marked as "artifacted". The threshold value of 20% of epoch duration corresponds to two segments with a length of 2 s. A 20-s epoch marked as "artifacted" is excluded from the sleep/wake stage scoring. On the contrary, if the number of artifacted segments is less or equal to two, the corresponding 20-s epoch is marked as artifact-free and is used for classification. However, all the 2-s segments contaminated with artifact are cut off from the epoch trace. This artifact detection strategy avoids undesirable loss of data: it provides an adequate rejection of artifacted segments (brief artifacts) or entire epochs (large, long-lasting artifacts). The ratio of 20% has been set in order to keep intervals of artifact-free signals long enough to provide an accurate estimation of the features.

The algorithms that were implemented to detect the artifacts most frequently present in the polysomnographic signals are detailed below. They are tuned in accordance with an expert's point of view. Two strategies are used. The first one is based on a simple comparison of a given parameter, extracted on the 2-s segment of the signal, to a fixed threshold. It detects values that are over a normal range. The second one computes the ratio of the value of a given parameter calculated on a 2-s segment of the signal over the median of N values of this parameter calculated in the vicinity of the 2-s segment analyzed. A high ratio detects a parameter value higher than the values measured locally. In total, seven different artifacts are identified:

1. an overflow artifact is detected if the maximal absolute value of the signal is greater than a fixed threshold depending on the amplifier active range;
2. a flat-line artifact is detected whenever the maximal peak to peak amplitude of the signal is lower than a fixed threshold depending on the amplifier analog-to-digital resolution;
3. a power-line artifact corresponds to interferences generated by the main power line (50 Hz in Europe). It is detected whenever the maximal peak to peak amplitude of the band-passed signal in [45;64] Hz is above a fixed threshold;
4. a high-frequency artifact is detected whenever the 95th spectral edge frequency is higher than a fixed value. Spectral edge frequency is defined further in Section 3.3.2
5. an ECG artifact consists in a sharp peak similar to the original QRS complex of the electrocardiogram. The algorithm computes the ratio of two parameters calculated from the first derivative of the signal, on a 2-s segment of the signal. The first parameter is the peak-to-peak amplitude of the signal first derivative. The second one is the interquartile range computed from the signal first derivative. The ratio is then compared with a fixed threshold. The detector can be also used to detect other sudden and unwanted sharp peaks in the signals.
   The last two detectors of artifacts use ratios. The window used to estimate the local value of the parameter to be compared is a symmetric window of 60 s, therefore corresponding to 3 consecutive scoring epochs;
6. a low-frequency artifact is detected whenever the ratio of the maximal peak to peak amplitude of the filtered signal in the range [0;2] Hz calculated on the 2-s segment analyzed over the median value of the peak to peak amplitude of the filtered signal in [0;2] Hz calculated on elementary intervals of 2 s on a symmetric window of 60 s is above a given threshold;
7. a muscular activity artifact consists in a burst of (high amplitude) spikes in the signal trace. It is detected whenever the ratio of the variance of the signal filtered in [5;64] Hz calculated on the 2-s segment considered on the median value of the variance of the signal filtered in [5;64] Hz calculated on elementary intervals of 2 s on a symmetric window of 60 s is above a given threshold.

The different detectors are tuned from an expert's point of view and the settings can be applied on any other database. The settings for overflow, flat-line and power-line detection depend on the amplification factor used by the data acquisition system. To detect overflow, the threshold should be set to 99% of the maximal amplitude range, to detect flat lines, to 1% of the maximal amplitude range. Power line is detected whenever the maximal peak to peak amplitude of the filtered signal is above 50% of the maximal range. High frequency artifacts are detected whenever the 95th spectral edge frequency is strictly above 30 Hz.

ECG, low frequency and muscular activity thresholds were tuned from a small set of artifacts. The thresholds were fixed so that all the artifacts present in the set were detected. Since they express a ratio of information present in the signal and thus do not depend on the amplification of the signal used by the acquisition system, the same thresholds could be applied on any new signals. They were fixed, respectively, to 13, 7.5 and 3.5.

Please cite this article as: S. Charbonnier, et al., Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging, Comput. Biol. Med. (2011), doi:10.1016/j.compbiomed.2011.04.001

3.3. Classification stage

3.3.1. Set of classifiers

The set of classifiers is composed of four multi-layer perceptrons (MLP) [20]. For each MLP, the number of neurons in the first layer is defined by the number of features extracted from the epoch to be processed. Classifier 1 makes use of features extracted from the EEG signal, classifier 2 uses features from EEG and EOG signals, features used by classifier 3 are obtained from EEG and EMG signals, and classifier 4 exploits features computed with EEG, EOG and EMG signals. The transfer function of the neurons in the first layer is a hyperbolic tangent function. The second layer of the network contains 6 neurons and the transfer function is a logarithmic sigmoid function. The output layer of the network consists of 5 neurons, each one corresponding to one sleep/wake stage. The transfer function of each neuron in this layer is a hyperbolic tangent. The methods used to design and train the neural networks and to select the input features for each neural network are described in the following section.

3.3.2. Input features

3.3.2.1. Feature selection strategy. For each of the four classifiers, a feature selection strategy, the Sequential Forward Selection (SFS), was used to select the most relevant features for the classification problem considered. SFS is an iterative technique that selects at each iteration the subset of features that maximizes a criterion J based on classification accuracy, presented in Eqs. (1) and (2). A detailed presentation of the method can be found in [17]. To perform SFS on the problem at hand, a data subset, named S, was extracted from the data set described in Section 2. S is composed of 3500 artifact-free 20-s epochs, equally distributed in the 5 sleep/wake stages to avoid class misrepresentation. The epochs forming S were collected in any of the 46 recordings and thus could come from any of the 13 patients. S is further split into seven subsets S_k, $S=\{S_1, S_2, \ldots, S_7\}$. Each subset S_k contains 500 epochs where each sleep/wake stage is represented with the same number of epochs, that is 100 epochs. The choice to use a subset of 500 epochs was made in accordance to the study presented in [9], which showed that 100 epochs per class were sufficient for a correct training and validation of the MLP.

The MLP described in Section 3.3.1 is trained on one subset S_k and validated on the 6 other subsets $S_{\bar{k}}$, $S_{\bar{k}} \in \bar{S}_k$, with $\bar{S}_k = S - S_k$.

An accuracy function is calculated on each of the 7 subsets $S_{\bar{k}}$ as

$$Acc(k,\bar{k}) = \frac{card[\{epoch(i) \in S_{\bar{k}} / C(epoch(i)) - E(epoch(i)) = 0\}]}{card[S_{\bar{k}}]} \quad (1)$$

where $epoch(i)$ is an epoch belonging to $S_{\bar{k}}$, $C(epoch(i))$ is the class assigned to $epoch(i)$ by the MLP, trained on the subset k. $E(epoch(i))$ is the class assigned by the experts to $epoch(i)$.

To obtain the global accuracy performance, a 7 fold cross validation is used. A circular permutation is performed on the 7 subsets S_k. The classifier is trained 7 times using the different data sets S_k, which provides 42 values of $Acc(k,\bar{k})$ (7 training applied on 6 different validation sets). Its mean value, ACC, is used as the criterion for SFS

$$ACC = \frac{1}{42} \sum_{k=1}^{7} \sum_{\bar{k} \notin k} Acc(k,\bar{k}) \quad (2)$$

Thus, in the SFS procedure, a new feature is added to the set of relevant features whenever the mean value of $Acc(k,k^{-})$, ACC, is significantly increased when this feature is added to the MLP inputs. The decision to stop the feature selection process is made using a mean comparison test between the ACC value obtained at the previous SFS step and the ACC value obtained at the current step.

3.3.2.2. List of candidate features. The list of features extracted from each 20 s epoch which are candidate for the feature selection process is presented below. Features can be classified in two groups:

- a first group containing the features that represent information in the frequency domain, computed by means of Fourier transform:
  1. the spectral activity of EEG in traditional frequency bands δ (delta, [0.5;4.5] Hz), θ (theta, [4.5;8.5] Hz), α (alpha [8.5;11.5] Hz), σ (sigma, [11.5;15.5] Hz) and β (beta, [15.5;32.5] Hz). The features are calculated using Welsh's periodogram Fourier transformation [21] on a 20-s window of signal, split into 2-s sequences. Relative powers, $Prel_j$, $j \in \{\delta, \theta, \alpha, \sigma, \beta\}$, are computed in the five frequency bands by dividing the absolute power in each frequency by the sum of powers in the [0.5;32.5] Hz frequency band;
  2. the relative power of EMG in the high frequency band [12.5;32] Hz is calculated. The total frequency band is defined as [8;32] Hz;
  3. the spectral edge frequency 95 ($SEF95_i$, $i = EEG, EMG, EOG$) indicates the highest frequency below which 95% of the total signal power is located. The spectral edge frequency function used in the present work is described in [22]. It is calculated on the three signals (EEG, EMG, EOG);
- a second group containing features computed in the time domain, all of them calculated on EEG, EOG and EMG:
  1. the entropy ($entr_i$, $i = EEG, EMG, EOG$) of the signal that measures the signal variability thanks to the distribution of its amplitude values. The algorithm used in this project was published in [23]
  2. a set of three parameters defined by Hjorth [24]: activity (act_i), mobility (mob_i) and complexity ($comp_i$), $i = EEG, EMG, EOG$;
  3. the 75th percentile ($prctile75_i$, $i = EEG, EMG, EOG$) defines the value below which 75% of the random variable values are located;
  4. the standard deviation (std_i), the skewness ($skew_i$) and the kurtosis ($kurt_i$), $i = EEG, EMG, EOG$.

The whole set of features contains 33 features that characterizes each epoch. Before the set of features is used for classification, each feature is transformed and normalized in order to reduce the extreme and outlying values, using the transformation strategy described in [25] and [15].

3.3.3. Design of the classifier

The structure of the neural network (NN) was selected prior to feature selection. The feature selection procedure is iterative: at each step the number of inputs of the neural network is increased. Though the optimal structure for the neural network may change when the number of inputs changes, and more specifically the optimal number of neurons in the hidden layer may vary, to select the most adequate neural network structure at each step of the feature selection process would be an intractable problem. The choice was made to select the most adequate NN structure when the whole set of possible inputs is used and to keep this structure during the search. This choice is based on the assumption that the NN designed to provide the best classification accuracy when the classification problem is the most complex should provide adequate results when the number of inputs decreases.

The selection of the best structure was carried out using an empirical incremental procedure. S presented in Section 3.3.2.1 was randomly split in two subsets, one being the training set and the other one the testing set. The number of neurons in the hidden layer was increased by one at each step until the mean squared error on the testing set obtained by the neural network learnt on the learning set stopped to significantly decrease. Different transfer functions were also tested for neurons in the input and hidden layers. The structure obtained was then kept as the final structure for the neural network.

The training of the neural network at each step of the feature selection process is carried out as follows. As presented in Section 3.3.2.1, the NN is learnt 7 times, on each of the subsets S_i, and validated on the six other subsets to obtain the classification accuracy criterion. For each subset S_i, the learning process is the following. Subset i is presented several times to the neural network and the mean squared error obtained on subset i is calculated after each presentation. The mean squared error on subset $i+1$, which serves as a testing set, is also calculated. To avoid over-training, the learning phase is stopped when the mean squared error on the testing set cease to decrease. The initial weights of the NN are randomly assigned. To avoid getting trapped in a local minimum, the learning phase is repeated 10 times, with 10 random different initializations and the final weights corresponding to the best solution are finally used to calculate the classification accuracy on the remaining subsets. The learning algorithm used is the Levenberg–Marquardt backpropagation algorithm.

3.3.4. *Classifier performances*

The results obtained by each classifier, that is for each combination of signals, are presented in the following way:

– the relevant features selected by the SFS algorithm are listed in selection order and the corresponding averaged classification accuracy (*Acc*) is presented.
– the confusion matrix obtained when the optimal set of features is used is displayed. Its columns represent the stages predicted by the classifier and its rows represent the stages determined by the experts. Each cell (i, j) corresponds to the number of examples classified as i by the experts and j by the machine, expressed as a percentage of the examples classified as i by the experts. The last row shows the confidence index presented in Section 3.3.5

3.3.4.1. Classifier 1: EEG signal. The initial set of candidate features contains 14 features. A subset of four relevant features was selected: $Prel_\beta$, $entr_{EEG}$, $Prel_\sigma$ and $Prel_\alpha$. The classification accuracy is $74.70 \pm 1.19\%$. The confusion matrix is presented in Table 2.

3.3.4.2. Classifier 2: EEG and EOG signals. The initial set of candidate features contains 23 features, 14 of them extracted from the EEG and 9 features from the EOG signal. A subset of 7 relevant features was selected: $Prel_\beta$, mob_{EOG}, $Prel_\alpha$, $entr_{EEG}$, $Prel_\sigma$, $kurt_{EOG}$ and $Prel_\theta$. The classification accuracy is $80.71 \pm 1.25\%$. The confusion matrix is presented in Table 3.

3.3.4.3. Classifier 3: EEG and EMG signals. The initial set of candidate features contains 24 features, 14 features extracted from the EEG and 10 features extracted from the EMG. A subset of 6 relevant features was selected : $Prel_\beta$, mob_{EMG}, $Prel_\alpha$, $Prel_\sigma$, $entr_{EEG}$ and $Prel_\theta$. The classification accuracy is $80.34 \pm 1.02\%$. The confusion matrix is presented in Table 4.

3.3.4.4. Classifier 4: EEG, EOG and EMG signals. The initial set of candidate features contains 33 features. A subset of 7 relevant features was selected: $Prel_\beta$, mob_{EMG}, $Prel_\alpha$, $Prel_\sigma$, $entr_{EOG}$, $entr_{EEG}$ and $kurt_{EOG}$. The classification accuracy is $82.52 \pm 1.21\%$. The confusion matrix is presented in Table 5.

3.3.4.5. Results analysis. An analysis of the results obtained with the different combinations of signals shows that the lowest classification accuracy is obtained when EEG only is used. It improves by nearly 10% when the two other signals (EMG and EOG) are added. Though wake, NREM II and SWS are correctly classified with EEG, the confusion matrix reveals high

Table 2
Confusion matrix. Relevant features extracted from the EEG signal.

%	**Classifier**				
	Wake	**NREM I**	**NREM II**	**SWS**	**REM**
Expert					
Wake	79.77	11.87	3.92	0.61	3.83
NREM I	12.34	49.61	7.35	0.61	30.09
NREM II	1.83	5.75	85.37	6.09	0.96
SWS	0.11	0.04	4.52	95.33	0
REM	3.60	30.56	2.40	0.39	63.05
Confidence index	0.8	0.5	0.8	0.9	0.6

Table 3
Confusion matrix. Relevant features extracted from the EEG and EOG signals.

%	**Classifier**				
	Wake	**NREM I**	**NREM II**	**SWS**	**REM**
Expert					
Wake	84.43	9.95	2.54	0.37	2.71
NREM I	8.86	72.11	5.32	0.37	13.34
NREM II	0.47	6.56	85.27	6.17	1.53
SWS	0.28	0.13	4.35	95.24	0
REM	3.53	27.90	1.79	0.37	66.41
Confidence index	0.9	0.6	0.9	0.9	0.8

Table 4
Confusion matrix. Relevant features extracted from the EEG and EMG signals.

%	**Classifier**				
	Wake	**NREM I**	**NREM II**	**SWS**	**REM**
Expert					
Wake	82.70	10.98	3.39	0.74	2.19
NREM I	10.46	55.64	6.08	0.61	27.21
NREM II	1.19	4.69	85.01	5.90	3.21
SWS	0.09	0	4.89	95.02	0
REM	1.66	13.34	1.64	0.39	82.97
Confidence index	0.9	0.7	0.8	0.9	0.7

Table 5
Confusion matrix. Relevant features extracted from the EEG, EMG and EOG signals.

%	**Classifier**				
	Wake	**NREM I**	**NREM II**	**SWS**	**REM**
Expert					
Wake	84.64	8.73	3.04	0.59	3.00
NREM I	9.01	69.55	5.91	0.50	15.03
NREM II	0.68	5.32	84.61	7.09	2.30
SWS	0.13	0.06	3.38	96.39	0.04
REM	2.35	18.33	1.70	0.32	77.30
Confidence index	0.9	0.7	0.9	0.9	0.8

Please cite this article as: S. Charbonnier, et al., Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging, Comput. Biol. Med. (2011), doi:10.1016/j.compbiomed.2011.04.001

disagreements between NREM I and REM sleep stages. Only about 50% of NREM I epochs are correctly scored. When EOG is added, the classification accuracy of the NREM I stage is increased of about 20%. The other stages are also slightly improved. When EMG is added to EEG, the classification of REM sleep epochs is highly improved: the classification accuracy of REM sleep stage is increased of about 20%.

These results are in concordance with the hypothesis that both EOG and EMG signals are helpful to classify these two stages, which are characterized by a similar EEG activity. During REM sleep, rapid eye movements can be observed, which can be measured by EOG. EMG measures the muscular activity (muscle tone) during the night sleep, which is very low during REM sleep and higher during wake and REM I stages.

The detailed analysis of the sets of relevant features selected shows that the set ($Prel_\beta$, $entr_{EEG}$, $Prel_\sigma$ and $Prel_\alpha$), which estimates the frequency contents of EEG, was selected whatever the combinations of signals (EEG+EOG, EEG+EMG, EEG+EOG+EMG). This fact is in concordance with the manual scoring performed by the physician. During a manual scoring, the physician analyzes at first the EEG signal trace and focuses on the information contained in the EOG and/or EMG signals only when his decision is not obvious. Data mining methods selected a set of four features representing the core of the information stored in the EEG. The features selected from the other signals are additional information used to precise the scoring.

Finally, the concordance between the automatic discovery made by the NNs used in the feature selection process and the physiologists' knowledge on sleep staging, as well as the satisfactory accuracies reached by the neural networks whatever the input signals are, validate the choice of a unique structure for all the neural networks.

3.3.5. Confidence index

The classification accuracy of the 5 stages depends strongly on the signals used. This means that the reliability of the decision made varies with the classifier used to make the decision. In this section, the results obtained during the elaboration phase of the classifiers are used to provide a confidence index associated with each decision made. This index is equal to the positive predictive value obtained for each class by each classifier. The positive predictive value is the percentage of epochs classified in one stage and that actually belong to this stage. It estimates the chance that, when a decision is made, this decision is correct. It is calculated from the confusion matrix as the number of epochs in the cell (i, i) divided by the sum of the epochs in column i.

The value of the confidence index associated with each decision made by the two-stages system is presented in Table 6. A color is associated with each value so as to provide visual information on the reliability of the decision. It shows that SWS is classified with a high confidence index whatever the classifier used. It can be easily discerned from the other stages by its typical EEG frequency contents. NREM II and wake are stages correctly classified with EEG too, but one can be all the more confident in the decision when EOG is used in the classification stage. Finally, NREM I and REM are classified with a poor confidence index when EEG only is used. This confidence index is increased when EMG is used to classify NREM I stage and when EOG is used to classify REM stage.

A hypnogram can be built in a user-friendly way, from Table 6, displaying the different sleep/wake stages recognized in various colors, depending on the reliability of the decision. This makes it easy for the physician to analyze the succession of epochs classified by the system and to discern the epochs for which the confidence index is lower. These epochs can be later manually classified by the expert.

4. Results and discussion

The two-stages decision system described in the previous section was implemented and tested on the database presented in Section 2, composed of 66 164 20-s epochs, collected on 13 different patients. The four MLP classifiers used in the classification stage were trained once on one of the seven subsets described in Section 3.3.2.1, using the training procedure described in Section 3.3.3. The training set is composed of 500 epochs recorded on different patients with each class being represented by the same number of epochs to avoid classification errors due to class misrepresentation. The training set represents 0.8% of the database. The figures presented in the following section correspond to results obtained on the whole database.

4.1. Artifact detection

The results obtained by the artifact detectors presented in Section 3.2 on the database are displayed in Table 7. It shows the number of 20-s epochs artifacted, i.e. polluted by more than 4 s of artifacts, for each polysomnographic signal. The total number of epochs in the database is 66 164. The number of epochs with at least one signal artifacted is 17 541, which represents 26.5% of the database. One can see that the EMG signal presents the highest rate of artifacts, with nearly 12% of the epochs artifacted. EMG artifacts are mostly overflow and muscular activity (see artifacts 1 and 7 in Section 3.2), while EEG and EOG are mainly polluted by high frequency artifacts (artifact number 4 in Section 3.2). As said previously, the artifact detection algorithm is tuned according to expert knowledge. Though no quantitative evaluation of the detectors evaluation can be made, since artifacts were not visually expertized by a physiologist, one can nevertheless assume that the use of sensible thresholds enables a realistic assessment of the number of artifacts. Thus, the high rate of artifacts detected in the database justifies by itself the necessity to develop classification systems able to deal with them.

Table 7
Percentage of artifacted epochs present in the database.

Both experts	EEG	EOG	EMG
Artifacted	5.9%	10.6%	11.9%

Table 6
Confidence index associated with each decision; green: higher than 0.9, blue between 0.7 and 0.8, red strictly smaller than 0.7.

	Confidence Index				
	Wake	NREM I	NREM II	SWS	REM
EEG	0.8	0.5	0.8	0.9	0.6
EEG+EOG	0.9	0.6	0.9	0.9	0.8
EEG+EMG	0.9	0.7	0.8	0.9	0.7
EEG+EOG+EMG	0.9	0.7	0.9	0.9	0.8

4.2. Performances of the two-stage classifier

Performances obtained by the two-stage system are presented in this section. At first, Table 8 shows how many epochs were scored by each of the four classifiers or excluded. The first line of Table 8 shows the absolute numbers of epochs processed by each individual classifier. In the second line of the table, the same information is expressed using percentage values.

From Table 8, it can be seen that 3765 epochs (5.7%) were excluded from the classification because of artifacts detected on the EEG signal. The rest of the database, 62 399 epochs (94.3%) in total, was scored by one of the four classifiers implemented. The majority of the data (48 623 epochs, i.e. 80% of the epochs classified) was scored using classifier 4 which uses features computed from all the three signals (EEG, EOG and EMG). Only 1 525 epochs (about 2.5%) have both EOG and EMG signals artifacted and thus were scored by classifier 1 using the features extracted from the EEG signal only.

The overall classification accuracy obtained by the two-stage system, computed over the 62 399 epochs, is 85.5%, which is quite acceptable.

The detailed analysis of the results is presented in Table 9 which shows the corresponding confusion matrix. The number in case (i, j) represents the percentage of epochs in stage i classified by the automatic system in stage j.

The two-stage system is very successful in classification of NREM sleep stage II and SWS. Classification accuracy of these stages exceeds 85% (about 87% and 95%, respectively). These stages are traditionally well classified during automatic classification. The classification accuracy of wake and REM sleep stages is slightly below 80%. The lowest classification accuracy is obtained for NREM I stage and reaches only about 65%. This stage is still confused with wake and REM sleep.

The approach proposed to deal with missing values seems to be effective. It allows the correct classification of epochs that would be excluded because of the presence of artifacts in the EOG or EMG signals. The classification accuracy reached for these 13 776 artifacted epochs is 80.7%. This value alone is high enough to conclude on the interest of the method proposed in this paper.

A detailed analysis of the epochs contained in the database is presented in Table 10. The table shows the absolute number of epochs processed by each individual classifier depending on individual sleep/wake stages. The color associated with each case reminds the reader of the confidence index associated with the corresponding decision. Epochs excluded from the classification are associated with the red color because these epochs will have to be evaluated afterwards by a physiologist.

A total of 4699 epochs are classified with a low confidence index (below 0.7 or excluded), 18 252 epochs were classified with an average confidence index (between 0.7 and 0.8) and 43 213 epochs with a high confidence index (0.9 or more). When expressed in percentage, this means that only 7% of the epochs constituting the database was classified with a low confidence index and should be re-evaluated by a physician. On the contrary, 65% of the epochs was classified with a high confidence index and could be accepted by the physician without further analysis. This shows how effective a decision-support tool the two-stage classifier is. If the user (the physician) decides to accept only the epochs classified with a very high confidence index (i.e. the decision index is above 0.9), he will have to analyze only one-third of the whole night sleep, which will result in a significant saving in time.

An example of a colored hypnogram obtained with the classification system is presented in Fig. 2, on the upper part. The corresponding hypogram scored by an expert is presented on the lower part. Code 1 corresponds to the wake stage, 2 to NREM I, 3 to NREM II, 4 to SWS and 5 to REM. Phases NREM II and SWS are evaluated as green by the system, which means that it is confident in its decision. When comparing with the expert's decision, these phases are correctly classified. Transitions between phases NREM II and SWS are not as sharp as the transitions from the expert but these transitions correspond to periods when the disagreement between different experts is high (the overall disagreement between 2 experts is about 13% on the database used). REM sleep is classified with an averaged confidence. This stage may be confused with NREM I but the classification is more reliable when EOG and/or EMG signals are available, which is the case in the hypnogram displayed. One can see on the expert hypnogram that most of these epochs are correctly classified. NREM I classification is less reliable, which is visible because it is classified on the hypnogram with either an

Table 8
Number of epochs classified per classifier.

	EEG	EEG and EOG	EEG and EMG	EEG and EOG and EMG	Excluded
	Classifier 1	Classifier 2	Classifier 3	Classifier 4	
Number of epochs	1525	7717	4534	48623	3765
%	2.3	11.7	6.8	73.5	5.7

Table 9
Confusion matrix. Performance of the two-stage classification system.

%	Classifier				
	Wake	NREM I	NREM II	SWS	REM
Expert					
Wake	**78.07**	12.79	3.40	2.10	3.64
NREM I	8.14	**64.77**	7.13	0.63	19.33
NREM II	1.80	4.65	**86.92**	4.80	1.83
SWS	0.15	0	5.09	**94.75**	0.01
REM	2.13	16.30	1.75	0.49	**79.33**

Table 10
Numbers of epochs classified by each individual classifier–distribution into sleep/wake stages.

	EEG	EEG & EOG	EEG & EMG	EEG & EOG & EMG	excluded
wake	445	329	535	1,162	2,827
NREM I	73	268	207	1,345	88
NREM II	358	4,131	1,587	25,931	455
SWS	56	1,295	96	9,678	85
REM	593	1,694	2,109	10,507	310

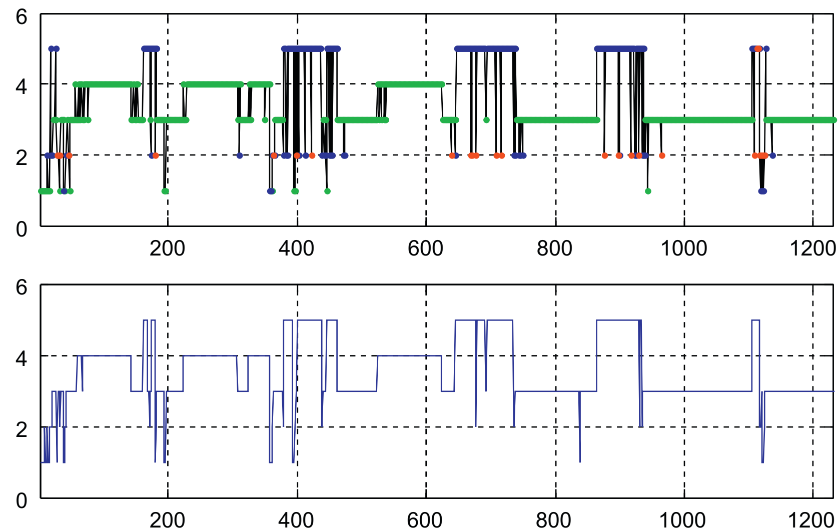

Fig. 2. On the upper part, a colored hypnogram obtained with the two-stage classifier. On the lower part, the corresponding hypnogram built by one of the 2 experts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

averaged or a low confidence, depending on the signals available. These epochs correspond to part of the hypnogram that the user should score manually. Wake stage is globally correctly classified with a high or averaged confidence index. The few errors made by the system correspond to a misclassification in the NREM I phase, its adjacent phase. Let us note that wake and NREM I are stages for which the inter-scorer disagreement is usually high.

4.3. Comparison with single classifiers

For a better analysis of the two-stage classification system performances, it is now compared with two single-stagers composed of only one MLP instead of a set of four.

4.3.1. Single classifier using artifact identification

The first stager is equipped with the same artifact identification stage as the two-stage system. It processes the epochs where all the three signals (EEG, EOG and EMG) are artifact-free and excludes the other ones. The classifier implemented is classifier number 4 used in the two-stage system.

Its confusion matrix is presented in Table 11. The overall classification accuracy of the stager is 86.8%, which is slightly higher than the two-stage system, with an increased ability to classify stage I (69% of concordance against 64.7% for the two-stage system). However, this automatic classifier is able to score 48 623 epochs out of the 66 164 epochs contained in the whole database, that is only 73.5% of the data. The rest of the database, 17 541 epochs (26.5%), is not classified.

4.3.2. Simple classifier without artifact identification

The second single sleep stager is not equipped with any artifact identification stage. It makes a decision whether the signals are artifacted or not, without excluding any epoch.

A complete elaboration phase was achieved to build this classifier, from the selection of relevant features to the training of the neural network. The selection of relevant features was achieved using the selection strategy presented in Section 3.3.2.1 on seven data subsets containing 500 possibly artifacted epochs. SFS selected seven relevant features: $Prel_\beta$, $entr_{EMG}$, $Prel_\sigma$, $entr_{EOG}$, $entr_{EEG}$, $Prel_\alpha$ and $Prel_\theta$.

The overall classification accuracy obtained on the whole database of polysomnographic recordings is 83.24%, which is

Table 11
Confusion matrix obtained with the single classifier using artifact identification.

%	Classifier				
	Wake	NREM I	NREM II	SWS	REM
Expert					
Wake	**79.95**	12.82	2.84	0.43	3.96
NREM I	6.32	**69.07**	7.51	0.45	16.65
NREM II	1.82	4.15	**87.33**	4.93	1.77
SWS	0.16	0	4.98	**94.85**	0.01
REM	1.91	15.04	1.68	0.14	**81.23**

slightly lower than the classification accuracy of the two-stage system. However, this single classifier classifies the whole base of 66 164 epochs. The confusion matrix is presented in Table 12. The comparison between this matrix and the confusion matrix obtained for the two-stage system (Table 9) shows that the classification accuracy of the REM sleep obtained with the single classifier has decreased by 10%. REM epochs are wrongly classified as NREM I stage. This increase in the number of REM sleep epochs misclassified as NREM sleep stage I is slightly counterbalanced by a decrease in the number of NREM I epochs wrongly classified in REM by 1.5%.

When analyzing the absolute numbers of epochs misclassified, it becomes obvious that the two-stage system using a set of classifiers performs a better discrimination between REM sleep and NREM I stages. Indeed, the system without artifact processing misclassifies 4201 epochs of stages NREM I and REM sleep. When the two-stage system is used, the number of misclassified epochs scored by both experts as NREM I and REM sleep decreases to 2762 epochs. The improved ability to discern NREM I and REM sleep stages is mainly due to the system not using artifacted signals. Indeed, a detailed analysis of the artifacted segments showed that a high number of EOG and EMG artifacted epochs were found in NREM I and REM sleep stages.

The confusion matrix of the single stager without artifact identification also shows a slight increase of the classification accuracy in the wake stage. The classification accuracy of wake increased by 3% compared to the two-stage system. It could be explained by the high artifact contamination of the wake stage. About half of the wake epochs are detected as artifacted by overflow or muscular activity EMG artifacts. It seems that the sleep stager, learnt on possibly artifacted data, misinterpreted

Table 12
Confusion matrix. Performance of the classifier without artifact identification.

%	Classifier				
	Wake	NREM I	NREM II	SWS	REM
Expert					
Wake	**80.99**	14.01	1.94	1.78	1.28
NREM I	5.71	**67.64**	8.58	0.25	17.82
NREM II	1.29	4.98	**86.96**	4.86	1.91
SWS	0.74	0.01	5.15	**93.95**	0.15
REM	1.49	25.29	2.41	0.55	**70.26**

Table 13
Classification results obtained for the different classifiers.

Type	Accuracy (%)	Number of epochs
Complex two-stage classifier	85.48	62399
Classifier with artifact identification	86.82	48623
Classifier without artifact identification	83.24	66164

these high-amplitude artifacts as a true high amplitude signal, typical of the wake stage. This artifact manifestation then facilitated the classification of artifacted epochs into wake stage.

4.4. Discussion

Table 13 summarizes the accuracies reached by the three automatic classifiers, as well as the numbers of epochs classified.

The highest classification accuracy is reached when a unique classifier using all the three signals is used in combination with an artifact identification strategy. However, this high classification accuracy is ransomed by a high number of data that cannot be classified by the unique classifier. More than 25% of the whole data are not classified because of the presence of artifacts in at least one signal. This high number of excluded epochs limits the usability of the system.

When the two-stage classification system is used, only a slight decrease of classification accuracy is noticed. However, the number of processed epochs is much higher. About only 6% of the whole data cannot be classified by the system. So, the strategy using a set of classifiers enables the classification of 13 776 epochs excluded when the single classifier is used. The accuracy reached by the two-stage system on these 13 776 epochs is equal to 80.7%.

When the classification is performed without artifact processing, the final classification accuracy computed over the whole database is the lowest, with 83.2% of global accuracy, although the classifier is fed only with features computed from the three signals.

If the global accuracy is considered as a unique assessment criterion, the advantage of the two-stage system is not so obvious with 85.5% accuracy to classify 94% of the whole database compared to 83.2% accuracy to classify 100% of the whole database. However, by using polysomnographic signals not polluted by artifacts, the two-stage classifier has an improved reliability. The confidence one user can have in a system that makes a decision using signals known to be quite often artifacted without analyzing their quality is comprehensively low. This is overcome by the two-stage system that makes decision using selected signals. Moreover, the automatic selection of the most relevant features depending on the available signals that was performed on a control data subset makes it sure that the most adequate signal processing methods are used by the classifier. Finally, the use of a confidence index alerting the user that the decision may be erroneous because the required relevant signals were not used to make the decision increases its reliability and thus the confidence one user can have in the system.

The analysis of the confusion matrices makes it clear that the two-stage system has an improved ability to discern NREM I and REM stages. NREM I and REM stages are stages difficult to discriminate because the spectral content of EEG is very similar in these two stages. Accurate features extracted on EOG and EMG are necessary to improve their classification. Both the two stage classifier and the single stager without artifact detection use EMG and EOG. However, by selecting EMG and EOG signals not polluted by artifacts, the two-stage system obtains better performances, which proves the interest to analyze the quality of the signals prior classification. In the same way, the figures obtained by the single stager without artifact identification to classify the wake stage can be discussed. The accuracy of this stage is artificially increased because of the many EMG overflow and muscular activity artifacts present in the database during this phase. The high energy of the EMG signal during these artifacted epochs was learnt as a main characteristic of the wake stage by the single stager. The good results obtained on validation with similarly artifacted epochs should decrease significantly when using the classifier on less artifacted polysomnographic signals and the user would not be warned of a possible misclassification of the phases.

To conclude, it is difficult to compare the performances of the system presented with classifiers published in the literature. As discussed above, the global accuracy cannot be used as an absolute comparison criterion and this is all the more true when performances are evaluated on different databases. Performances presented in the literature vary between 70% and 95% of overall accuracy depending on the number of classes to discriminate, on the signals used to make the classification and on the way epochs forming the validation set were selected. Most of them were performed on a database much smaller than the one used in this paper, which is one of the largest used. With 85% of global accuracy, the system presented here obtains satisfactory results.

5. Conclusion

A two-stage sleep/wake classification system was proposed and evaluated on a large database of polysomnographic recordings. The first stage of the system checks the 20-s epoch of polysomnographic signals (EEG, EOG and EMG) for the presence of artifacts and selects the artifact-free signals. The second stage classifies the epoch using one classifier selected out of four, using feature inputs extracted from the artifact-free signals only. A confidence index, associated with each automatic decision, has been proposed. It was developed from the predictive positive values obtained by each specific classifier on a control sub-database.

The set of four classifiers is used as a means to classify epochs where EMG or EOG are artifacted, which would be discarded by a single classification system using an artifact identification strategy. This strategy enables the classification of more than 20% epochs of the database, containing artifacts on EMG and/or EOG signal(s), with 80% of accuracy. The overall accuracy reached on the whole database is 85.5%. It is higher than the performances obtained when a single stager without artifact identification is used, with a better discrimination between NREM I and REM stages.

The index provided with each decision enables an interaction between the physiologist and the machine and thus can be a support to the human decision. The user can easily build a hypnogram by accepting all the decisions made with a high confidence and re-evaluating the epochs classified with a lower confidence index. Considering that only 7% of the 66 164 epochs of the database were classified with a low confidence index, the saving in time is tremendous. Moreover, the proposal of a confidence index may increase the acceptance of automatic sleep stagers by physiologists.

Conflict of interest statement

None declared.

Acknowledgement

The authors are grateful to Alain Buguet and Emilia Sforza for providing and scoring the data.

References

[1] M. Reite, D. Buysse, C. Reynolds, W. Mendelson, The use of polysomnography in the evaluation of insomnia, Sleep 18 (1995) 58–70.
[2] A. Rechtschaffen, A. Kales, A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects, US Government Printing Office, Washington, 1968.
[3] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Elsevier Science, 2008.
[4] M. Schwaibold, J. Schochlin, A. Bolz, Automated sleep stage detection with a classical and a neural learning algorithm–methodological aspects, Biomedizinische Technik, Berlin 47 (2002) 318–320.
[5] C. Robert, C. Guilpin, A. Limoge, Review of neural network applications in sleep research, Journal of Neuroscience Methods 79 (1998) 187–193.
[6] N. Schaltenbrand, R. Lengelle, J.P. Macher, Neural network model: application to automatic analysis of human sleep, Computers and Biomedical Research 26 (1993) 157–171.
[7] M. Grozinger, J. Roschke, B. Kloppel, Automatic recognition of rapid eye movement (REM) sleep by artificial neural networks, Journal of Sleep Research 4 (1995) 86–91.
[8] C. Robert, P. Karasinski, R. Natowicz, A. Limoge, Adult rat vigilance state discrimination by artificial neural network using a single EEG channel, Physiology and Behavior 59 (6) (1996) 1051–1060.
[9] G. Becq, S. Charbonnier, F. Chapotot, A. Buguet, L. Bourdon, P. Baconnier, Comparison between five classifiers for automatic scoring of human sleep recordings, in: Saman K. Halgamuge, Lipo Wang (Eds.), Studies in Computational Intelligence (SCI), Volume 4: Classification and clustering for knowledge discovery, vol. 4, Springer Verlag, 2005, pp. 113–127.
[10] A. Flexer, G. Gruber, G. Dorffner, A reliable probabilistic sleep stager based on a single EEG signal, Artificial Intelligence in Medicine 33 (3) (2005) 199–207.
[11] P. Andere, G. Gruber, S. Parapatics, M. Woertz, T. Miazhynskaia, G. Klosch, B. Saletu, J. Zeitlhofer, M.J. Barbanoj, H. Danker-Hopfe, S.L. Himanen, B. Kemp, T. Penzel, M. Grozinger, D. Kunz, P. Rappelsberger, A. Schlogl, G. Dorffner, An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24 $\times$ 7 utilizing the Siesta database, Neuropsychobiology 51 (3) (2005) 115–133.
[12] E. Oropoesa, H.L. Cycon, M. Jobert, Sleep Stage Classification using Wavelet Transform and Neural Network, International Computer Science Institute, 1999 (TR-99-008).
[13] M. Grozinger, J. Fell, J. Roschke, Neural net classification of REM sleep based on spectral measures as compared to nonlinear measures, Biological Cybernetics 85 (2001) 335–341.
[14] J. Fell, J. Röschke, K. Mann, C. Schaffner, Discrimination of sleep stages: a comparison between spectral and nonlinear EEG measures, Electroencephalography and Clinical Neurophysiology 98 (1996) 401–410.
[15] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, F. Chapotot, Feature selection for sleep/wake stages classification using data driven methods, Biomedical Signal Processing and Control 2 (3) (2007) 171–179.
[16] H.H. Jasper, Appendix to report to Committee on Clinical Examination in EEG: the ten-twenty electrode system of the International Federation, Electroencephalography and Clinical Neurophysiology 10 (1958) 371–375.
[17] A. Buguet, A. Montmayeur, R.A. Pigeau, P. Naitoh, Modafinil, d-amphetamine and placebo during 64 hours of sustained mental work. Effects on two nights of recovery sleep, Journal of Sleep Research 4 (1995) 229–241.
[18] P. Anderer, S. Roberts, A. Schlogl, G. Gruber, G. Klosch, W. Herrmann, P. Rappelsberger, O. Filz, M.J. Barbanoj, G. Dorffner, B. Saletu, Artifact processing in computerized analysis of sleep EEG – a review, Neuropsychobiology 40 (3) (1999) 150–157.
[19] D. Brunner, R. Vasko, C. Detka, J. Monahan, C. Reynolds, D. Kupfer, Muscle artifacts in the sleep EEG: Automated detection and effect on all-night EEG power spectra, Journal of Sleep Research 5 (1996) 155–164.
[20] S. Haykin, in: 2nd edition, in: S. Haykin (Ed.), Neural Networks: A Comprehensive Foundation, 842, Prentice Hall Inc., New York, USA, 1998?.
[21] P.D. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Transactions on Audio Electroacoustics AU-15 (1967) 70–73.
[22] I.J. Rampil, F.J. Sasse, N.T. Smith, B.H. Hoff, D.C. Fleming, Spectral edge frequency-a new correlate of anesthetic depth, Anesthesiology 53 (1980) 512–517.
[23] R. Moddemeijer, On estimation of entropy and mutual information of continuous distributions, Signal Processing 16 (1989) 233–246.
[24] B. Hjorth, EEG analysis based on time domain properties, Electroencephalography and Clinical Neurophysiology 29 (1970) 306–310.
[25] T. Gasser, P. Bächer, J. Möchs, Transformations towards the normal distribution of broad band spectral parameters of the EEG, Electroencephalography and Clinical Neurophysiology (1982) 119–124.