Biomedical Signal Processing and Control 49 (2019) 14–23

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Multi-modality of polysomnography signals' fusion for automatic sleep scoring

Rui Yan[a,b], Chi Zhang[a], Karen Spruyt[c], Lai Wei[d], Zhiqiang Wang[d], Lili Tian[e], Xueqiao Li[e], Tapani Ristaniemi[b], Jihui Zhang[f,*], Fengyu Cong[a,b,**]

[a] School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China
[b] Faculty of Information Technology, University of Jyväskylä, 40014, Jyväskylä, Finland
[c] Lyon Neuroscience Research Center, INSERM U1028-CNRS UMR 5292-Waking Team, University Claude Bernard, School of Medicine, Lyon, France
[d] Otolaryngology Department, Affiliated Zhongshan Hospital of Dalian University, China
[e] Department of Psychology, University of Jyväskylä, 40014, Jyväskylä, Finland
[f] Department of Psychiatry, The Chinese University of Hong Kong, Shatin, Hong Kong, China

ARTICLE INFO

Article history:
Received 28 March 2018
Received in revised form
11 September 2018
Accepted 6 October 2018

Keywords:
Polysomnography
Multi-modality analysis
Rules of R&K
Automatic sleep scoring

ABSTRACT

Objective: The study aims to develop an automatic sleep scoring method by fusing different polysomnography (PSG) signals and further to investigate PSG signals' contribution to the scoring result.
Methods: Eight combinations of four modalities of PSG signals, namely electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG) were considered to find the optimal fusion of PSG signals. A total of 232 features, covering statistical characters, frequency characters, time-frequency characters, fractal characters, entropy characters and nonlinear characters, were derived from these PSG signals. To select the optimal features for each signal fusion, four widely used feature selection methods were compared. At the classification stage, five different classifiers were employed to evaluate the validity of the features and to classify sleep stages.
Results: For the database in the present study, the best classifier, random forest, realized the optimal consistency of 86.24% with the sleep macrostructures scored by the technologists trained at the Sleep Center. The optimal accuracy was achieved by fusing four modalities of PSG signals. Specifically, the top twelve features in the optimal feature set were respectively EEG features named zero-crossings, spectral edge, relative power spectral of theta, Petrosian fractal dimension, approximate entropy, permutation entropy and spectral entropy, and EOG features named spectral edge, approximate entropy, permutation entropy and spectral entropy, and the mutual information between EEG and submental EMG. In addition, ECG features (e.g. Petrosian fractal dimension, zero-crossings, mean value of R amplitude and permutation entropy) were useful for the discrimination among W, S1 and R.
Conclusions: Through exploring the different fusions of multi-modality signals, the present study concluded that the multi-modality of PSG signals' fusion contributed to higher accuracy, and the optimal feature set was a fusion of multiple types of features. Besides, compared with manual scoring, the proposed automatic scoring methods were cost-effective, which would alleviate the burden of the physicians, speed up sleep scoring, and expedite sleep research.

1. Introduction

Sleep covers almost one-third of the human lifespan [1]. Adequate and high-quality sleep is vital to our physical and mental well-being [2]. To study sleep dynamic, Rechtschaffen and Kales [3] introduced rules for labelling each sleep segment of 30 s as wakefulness (W), NREM stage (S1, S2, S3, or S4) or REM stage (R). Each sleep stage has a certain proportion and plays a vital role in the recuperation of living organisms. Studies have found that the distortion of sleep structure could lead to catastrophic outcomes. For example, REM disturbance slows down the perceptual skill improvement [4], deprivation of slow wave sleep is associated with Alzheimer's disease [5], insufficient sleep duration has detrimental effects on metabolic health [6], etc. Therefore, assessing sleep behavior and analyzing sleep structure are crucial in clinical applications.

* Corresponding author.
** Corresponding author at: School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China.
E-mail addresses: jihui.zhang@cuhk.edu.hk (J. Zhang), cong@dlut.edu.cn (F. Cong).

https://doi.org/10.1016/j.bspc.2018.10.001

A golden tool for quantitatively assessing sleep is PSG test that records simultaneously tens of physiological signals containing EEG, EOG, EMG, ECG, etc. Generally, according to the rules of Rechtschaffen & Kales (R&K) [3] and the recently updated American Academy of Sleep Medicine rules (AASM) [7], these PSG recordings are scored mutually by at least one registered sleep technologist (RST) to get the sequence of sleep stages. However, the manual scoring process is rather time-consuming [8] and subjective to some extent [9,10]. By contrast, automatic sleep scoring has shown advantages of cost-effective and preferable scoring performance. According to signal types employed in previous researches, these scoring methods can be divided into two categories, single-modality processing and multi-modality processing.

Single-modality scoring methods mainly based on EEG since EEG signals contained valuable and interpretable information resembling brain activities. Şen et al. [11] extracted 41 parameters from channel C3-A2 achieving a classification accuracy of 97.03%, which was fairly high among other related studies. However, as Diykh et al. [12] claimed, it was challenging to achieve such high accuracy. Instead of traditional linear features, Liang et al. [13] employed multiscale entropy and autoregressive models for single-channel EEG and obtained a good performance. Another new attempt appeared in Dimitriadis and his colleagues' study [14], in which sleep stages were evaluated by calculating cross-frequency coupling of predefined frequency pairs. In addition, to improve the performance of automatic sleep scoring, multiple EEG channels were investigated in existing studies. By fusing joint collaborative representation and joint sparse representation algorithms [15], a two-stage multi-view learning algorithm was constructed achieving a mean scoring accuracy of $81.10 \pm 0.15\%$.

According to the R&K rules, multi-modality of signals (such as EEG, EOG and EMG) were required for technologists to score sleep stages. Therefore, numerous studies considered multiple PSG signals. In Estrada et al.'s research [16], it concluded that EOG and EMG served as an important switching index of different sleep stages. In addition, using EEG, EMG and EOG recordings of five healthy subjects, Özşen [17] developed five different artificial neural network architectures to train each sleep stage separately. This separation of training procedure exhibited its superiority as Özşen claimed. Similarly, the multiple classifiers were used in Zhang and his colleagues' study [18] which achieved a high accuracy of 91.31%.

In previous studies, ECG signals, as vital physiological measures, were mainly used in home sleep monitoring systems [19–21]. To our knowledge, very limited articles explored it together with EEG, EOG and EMG in automatic sleep scoring algorithms. In Šušmáková and Krakovská's study [22], ECG was considered to be negligible compared with EEG, EOG and EMG. By extracting 74 measures from the signals of EEG, EMG, EOG and ECG, Krakovská and Mezeiová [23] found the ECG feature named zero-crossing rate performed well in automatic sleep scoring, but was still inferior to those from the other three signals. Whereas, four distance-based ECG features, related to the similarity of a baseline ECG epoch to the rest of epochs, were considered important in Gharbali et al.'s study [24]. Therefore, new evidence is needed to clear the PSG signals' contribution to automatic sleep scoring.

Based on the above issues, the present study aims to develop a multi-modality sleep scoring method and to detect PSG signals' contribution to sleep scoring. Besides, different feature selectors and classifiers are compared to provide a reference for future studies in this area. The main contributions of this work are presented as following:

Table 1
Distribution of sleep stages on dataset.

Sleep stages	W	R	S1	S2	S3	S4	Total
Number of epochs	449	1405	280	2162	570	1181	6047

1 Developing an automatic sleep scoring method by fusing four modalities of PSG signals,
2 Analyzing four types of PSG signals' contribution to distinguishing sleep stages,
3 Comparing the performance of different signals' fusions in sleep scoring,
4 Evaluating the discrimination ability of the optimal features selected from different signals' fusions,
5 Assessing the effect of different feature selectors and classifiers.

The article is organized as follows. Section 2 explains the details of experimental data and methodology of this study, together with a brief description of feature selection methods and classifiers employed in this study. Section 3 demonstrates the performance of proposed method, different feature selectors and classifiers. Section 4 provides discussions of results and limitations of this study. Finally, Section 5 gives conclusions of this paper.

2. Methodology

2.1. Data description

All-night PSG sleep recordings were provided by PhysioBank [25] (The CAP Sleep Database [26]). The CAP Sleep Database was a collection of 108 polysomnographic recordings registered at the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy, which included EEG channels, EOG channels, submental EMG channel and other electrophysiological signals. A detailed description and definition of CAP Sleep Database was given in Terzano et al.' s study [27]. In the present study, four types of PSG signals (EEG, EOG, EMG and ECG) were required. However, these measurements were not contained in each recording of the database. Given that, a total number of 6 healthy subjects that contained requisite four types of PSG signals were selected. The total duration of all recordings combined was 50 h, 33 min and 30 s with 8.5 h' average sleep time for each subject. The age of subjects ranged from 23 to 37 years, with a mean of 32 years and a standard deviation of 5.4 years.

For each overnight sleep recording, a traditional hypnogram followed the R&K rules was available, which represented the manual classification of sleep stages by the experts on 30 s non-overlapping segments. The hypnograms were used as a reference to evaluate automatic classification results. For each subject, the following four modalities of signals were analyzed: one EEG channel (C4, referred to the left mastoids A1 following the 10–20 international electrode placement system), one relative EOG channel (ROC, referred to LOC), one submental EMG channel and one ECG channel.

All signals were sampled or resampled to 512 Hz. In order to remove noise and artifacts, a notch filter at 50 Hz, a high-pass filter with a cut-off frequency of 0.3 Hz and a low-pass filter with a cut-off frequency of 30 Hz were applied to the signals of EEG, EOG and ECG [26]. In terms of EMG, a notch filter at 50 Hz, a high-pass filter with a cut-off frequency of 10 Hz and a low-pass filter with a cut-off frequency of 75 Hz were performed [22]. Afterwards, all the signals were divided into 30-second epochs, each epoch corresponding to a single sleep stage in hypnogram. Table 1 presented the distribution of sleep stages in the present data set.

Table 2
List of features from EEG signal.

P	Feature	P	Feature	P	Feature	P	Feature
1	minV	11	The 25_{th} percentile	37	Power spectral density	61	Approximate entropy
2	maxV	12	The 50_{th} percentile	38	Spectral edge	62	Permutation entropy
3	SD	13	The 75_{th} percentile	39-42	Absolute power spectral	63	Spectral entropy
4	Mean	14	The 90_{th} percentile	43-46	Relative power spectral		
5	Variance	15	HA	47-55	Power ratios		
6	Skewness	16	HM	56	PFD		
7	Kurtosis	17	HC	57	Mean teager energy		
8	Median	18-20	AR coefficients for EEG epoch	58	Energy		
9	ZCs	21-32	AR coefficients for rhythm waves	59	Mean curve length		
10	The 5_{th} percentile	33-36	Energy for rhythm waves	60	Hurst exponent		

Table 3
List of features from EOG signal.

P	Features	P	Features	P	Features
64	minV	74	HM	86	Spectral entropy
65	maxV	75	HC	87	Permutation entropy
66	Mean	76-78	AR coefficients		
67	SD	79	Power spectral density		
68	Variance	80	Spectral edge		
69	Skewness	81	PFD		
70	Kurtosis	82	Hurst exponent		
71	Median	83	Energy		
72	ZCs	84	Mean teager energy		
73	HA	85	Approximate entropy		

Table 4
List of features from EMG signal.

P	Features	P	Features	P	Features
88	minV	98	HM	110	Spectral entropy
89	maxV	99	HC	111	Permutation entropy
90	Mean	100-102	AR coefficients		
91	SD	103	Power spectral density		
92	Variance	104	Spectral edge		
93	Skewness	105	PFD		
94	Kurtosis	106	Hurst exponent		
95	Median	107	Energy		
96	ZCs	108	Mean teager energy		
97	HA	109	Approximate entropy		

2.2. Methodology and algorithm description

A total of 232 features from 7 different categories (time, frequency, time-frequency, fractal, entropy, nonlinearity and mutual-based features) are extracted in the feature extraction phase. The following section introduces the details of feature extraction methods. All the features and their corresponding origins are listed in Tables 2–5.

2.2.1. Time domain features

Statistical parameters, containing minimum value (minV), maximum value (maxV), standard deviation (SD), arithmetic mean (Mean), variance, skewness, kurtosis and median are derived from the segments of EEG, EOG, EMG and ECG. These statistical parameters are good indicators of the amplitude and distribution of time series. Details of these computations can be found in Şen et al.'s research [11].

Hjorth parameters (i.e., activity, mobility and complexity) are often used in the analysis of EEG signals [28]. In this article, they are derived from the segments of EEG, EOG, EMG and ECG.Hjorth activity:

$$HA = \sigma_0^2 \tag{1}$$

Hjorth mobility:

$$HM = \sigma_1/\sigma_0 \tag{2}$$

Hjorth complexity:

$$HC = \frac{\sigma_2/\sigma_1}{\sigma_1/\sigma_0} \tag{3}$$

where σ_0, σ_1 and σ_2 respectively denote the standard value of time series x_n, its first derivative $\dot{x}_n$, and its second derivative $\ddot{x}_n$.

AR coefficients, encoding a signal into several coefficients, are capable of undermining time-domain dynamics of signals which cannot be revealed by other features [29]. The following equation illustrates an AR model.

$$x(n) = -\sum_{i=1}^{p} a_i x(n-i) + e(n) \tag{4}$$

where $x(n)$ is a time series, a_i denoting AR coefficients, and $e(n)$ indicates prediction error. In the present study, only the first three coefficients were considered.

Zero-crossings is a time-based feature widely used in electronics, mathematics, image processing and signal processing [11]. Zero-crossings is an indicator of signal's noise ratio.

$$(x_{n-1} < 0 \,\&\, x_n > 0) \,\|\, (x_{n-1} > 0 \,\&\, x_n < 0) \,\|\, (x_{n-1} \neq 0 \,\&\, x_n == 0) \tag{5}$$

Percentile analysis provides information about the distribution of signal's amplitude, which is helpful to the discernment of sleep stages. The 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of signal's amplitude are calculated in this study.

Table 5
List of features from ECG signal.

P	Feature	P	Feature	P	Feature	P	Feature
112	Energy	124	Petrosian Fractal dimension	134	Mean RRIs	144	MAD of detrended RRIs changes
113	The 4^{th} order power	125	ZCs	135	Median RRIs	145	SD of RRIs difference
114	Curve length	126	HA	136	Mean detrended RRIs	146	Approximate entropy
115	Mean teager energy	127	HM	137	Standard deviation of RRIs	147	Permutation entropy
116-118	AR coefficients	128	HC	138	Difference between Max and Min RRIs	148	Spectral entropy of RRIs
119	Power spectral density	129	Mean of R_{amp}	139	Inter-quartile range		
120	Spectral edge	130	SD of R_{amp}	140	Mean absolute deviation (MAD)		
121	Mean-PSD	131	Mean HR	141	RMS of RRIs changes		
122	Median-PSD	132	Mean detrend HR changes	142	Mean of RRIs changes		
123	Spectral entropy	133	Mean nondetrend HR changes	143	MAD of nondetrended RRIs changes		

2.2.2. Frequency and time-frequency features

Power spectral density: Each epoch is segmented into fifteen non-overlapping parts (1024 points each) by hamming window. Then, the fifteen vectors are zero-padded to the length of 2048 points, respectively. The final spectral density is achieved by averaging spectral densities of the fifteen segments [22].

Spectral edge is defined as the frequency corresponding to 90% of the total spectral power [30]:

$$\sum_{f=f_{min}}^{edge} P(f) = 0.9 \sum_{f=f_{min}}^{30Hz} P(f) \tag{6}$$

where f_{min} is 0.3 Hz in terms of EEG, EOG and ECG, and 10 Hz in EMG.

Absolute and relative spectral power are obtained from four frequency bands of EEG, namely, 0.3–4 Hz (delta), 4–8 Hz (theta), 8–16 Hz (alpha) and 16–30 Hz (beta). Here, maximum overlap discrete wavelet is employed to decompose EEG signals. Afterwards, power spectral density is calculated within each frequency band. Then the relative spectral power is defined as the ratio of spectral power within the specific frequency band to the total spectral power. The total spectral powers of EEG, EOG and ECG signals are computed within the range of 0.3–30 Hz. Whereas, the total spectral power of EMG is calculated within the range of 10–30 Hz.

Power ratios are computed based on relative spectral powers in aforementioned frequency bands. The following power ratios are computed: delta/theta, delta/alpha, delta/beta, theta/alpha, theta/beta, alpha/beta, alpha/(theta + delta), delta/(theta + alpha) and theta/(beta + delta).

2.2.3. Nonlinear-based features

Energy is deemed as a reliable indicator to discern different activities of sleep stages [11]. It is defined as,

$$E = \sum_{n=1}^{N} x(n)^2 \tag{7}$$

where $x(n)$ is time series, and N denotes the length of time series.

Mean teager energy (MTE) is a non-linear operator that can effectively track the energy of signals. It can be derived by the following formula [31],

$$MTE = \frac{1}{N}\sum_{n=3}^{N}\left(x(n-1)^2 - x(n)x(n-2)\right) \tag{8}$$

where $x(n)$ is time series, and N denotes the length of time series.

Mean curve length (MCL) was proposed by Esteller et al. [32] to provide an estimation of Katz fractal dimension. It is widely used in the identification of EEG signals' activities [11]. MCL is defined as

$$MCL = \frac{1}{N}\sum_{n=2}^{k}\left|x(n) - x(n-1)\right| \tag{9}$$

where $x(n)$ is time series, N denoting the length of time series, and k is the last sample in the epoch.

Hurst exponent is used in time series analysis to present non-stationary or antistatic signal states observed in sleep [33]. It is defined as

$$H = \log\left(R/S\right)/log(T) \tag{10}$$

where T is the duration of the time series, R/S the value of rescaled range, R the difference between maximal and minimal "accumulated" values, and S is the standard deviation of observed series $x(n)$ [34].

2.2.4. Fractal feature

Fractal dimension is a chaotic parameter elucidating the complexity of signals. Petrosian fractal dimension (PFD) facilitates the rapid calculation of fractal dimension [11]. The rapid calculation is achieved through transforming the signal into a binary sequence. It can be estimated by the following formula,

$$PFD = log_{10}N/\left(log_{10}N + log_{10}\left(N/(N + 0.4N_\sigma)\right)\right) \tag{11}$$

where N is the length of time series, and N_σ is the number of sign changes in the signal derivative.

2.2.5. Entropy-based features

Approximate entropy (ApEn) is a measure used to quantify the unpredictability or randomness of signals. It has been reported that the mean value of approximate entropy changed significantly with different sleep stages[33].

Permutation entropy (Pen) is a complexity measure of time series based on comparing neighboring values [11]. More details of this measure can be found in Liu and Wang's study [35].

Spectral entropy is computed based on the relative power spectral density P_{ref} [22]. It is defined as

$$SEN = \frac{1}{ln(N)}\sum_{f=f_{min}}^{30Hz} P_{ref}(f)\, ln\left(P_{ref}(f)\right) \tag{12}$$

where N is the length of time series, and f_{min} is set as 0.3 Hz in terms of EEG, EOG and ECG signals, and 10 Hz of EMG.

2.2.6. Mutual-based features

To evaluate the relationship between two signals, the spectral coherence and phase angle are computed between every two epochs from four modalities of signals. To elaborate frequency bands, aforementioned four frequency bands (0.3–4 Hz, 4–8 Hz, 8–16 Hz and 16–30 Hz) are further divided as following: 0.3–2 Hz, 2–4 Hz, 4–6 Hz, 6–8 Hz, 8–10 Hz, 10–12 Hz, 12–14 Hz, 14–16 Hz and 16–30 Hz. For EMG signals, spectral measures are only computed in frequency bands higher than 10 Hz because a high-pass filter is used in the pre-processing stage. More details about coherence and phase angle can be found in Šušmákovác and Krakovská's study [22].

Mutual information that measures the mutual dependence of two variables is derived from the signals of EEG, EOG, EMG and ECG. It is calculated based on marginal entropy and joint entropy of two variables. The definition of mutual information can be found in Šušmákovác and Krakovská's study [22].

2.2.7. Heart-beat related features

To extract heartbeat-related features, RR intervals (RRI, the time interval between consecutive heartbeats or R peaks) are required. In order to highlight R wave, the median filters of 200 ms and 600 ms are used to ECG epochs successively. In the present study, Pan and Tompkins' QRS detector [36] is employed to locate R peaks. Then, the amplitude of R waves, heart rates, RR intervals and the change of RR intervals are calculated. A total of 37 features are extracted from ECG segments, a detailed description of these measures which can be found in Noviyanto et al.'s study [37].

2.3. Feature normalization and selection

2.3.1. Feature normalization

After all features extracted, a feature set, with the dimension of 6047 × 232 for four modalities of PSG signals' fusion, is achieved. In order to balance numerical ranges and to avoid numerical difficul-

ties in classification, each feature is separately normalized to [0, 1] by the following formula,

$$\tilde{p}_{i,j} = \left[p_{i,j} - min\left(p_{:,j}\right)\right] / \left[max\left(p_{:,j}\right) - min\left(p_{:,j}\right)\right] \quad (13)$$

where p_j denotes a vector of each independent feature, and $p_{i,j}$ is an element in the j_{th} feature vector.

2.3.2. Feature selection methods

Feature selection is a process of selecting an effective subset from canonical features to reduce dimension, shorten training time and simplify learning model. Feature selection methods are highly dependent on their defined objective function that heavily influences selection results. Therefore, to find the most discriminative features, four different feature selectors are considered in the present article. The employed feature selectors are ReliefF algorithm, improved distance-based evaluation methods (IDE), genetic algorithms (GA) and forward selection process (FSP). A brief description of these methods is provided below.

ReliefF is a supervised feature-weighting algorithm of the filter model that searches for the nearest neighbors of instances from different classes. *ReliefF* weights features according to how well they differentiate instances of different classes [38]. ReliefF is robust and also able to deal with incomplete and noisy data [39]. Therefore, ReliefF, as a widely used feature selector in the multi-modality analysis, is employed as a prime selector in the present study.

Improved distance-based evaluation methods (IDE) was developed by Lei et al. [40]. The method is especially useful in feature selection for the purpose of classification. It grades features between [0, 1] in where a higher value indicates a higher discriminative capability. The discriminative feature set can be selected by setting a threshold for graded results. IDE method has the advantage of simplicity and reliability as other distance-based feature selectors.

Genetic algorithms (GA) is a population-based technique. Instead of single potential solutions, it uses a population of potential solutions. That strategy is particularly suitable for multi-objective optimization. GA has motivated an increasing number of applications in engineering and related fields due to its capability of finding global optima and solving discontinuity and noise problems [41].

Forward selection process (FSP) is the simplest method among sequential strategies. It is a greedy search algorithm that determines iteratively the effective feature subset by adding one feature per iteration, on the condition that the newly added feature increases the value of objective function [42]. Once the termination condition is satisfied, the selected number of features, which reaches the lowest error at the first time, will be chosen as the dimension of optimal features set [43]. The main drawback of the sequential approach is that it gravitates toward local minima due to the inability to re-evaluate the effectiveness of features. Once a feature is added or discarded from the final set of features, the results would be irreversible.

2.4. Classification

In order to capture characters of sleep stages and to predict new instances, five different classifiers are employed, namely k-nearest neighbor classifier (KNN), binary decision tree (BDT), naive Bayes (NB), random forest (RF) and support vector machine (SVM). Their performances are compared to ascertain the optimal classifier for automatic sleep scoring. Introduction of these classifiers is given in the following.

K-nearest neighbor classifier (KNN) is a nonparametric classification approach. It has been pervasively used in the fields of science and engineering as a benchmark classifier due to its robust performance [18]. An instance is classified based on the majority votes of its closest training neighbors. The number of closest training neighbors k is crucial for classification results. Generally, the increasing of k value would attenuate the noise effect in the classification, whereas it would also obscure the boundaries between classes. In the present study, different k values are examined. It turns out $k = 5$ is the best choice. A peculiarity of the KNN algorithm is its sensitivity to the local structure of data, especially when the class distribution is skewed.

Binary decision tree (BDT) is a decision-support tool that used a tree-like graph or model of decisions. It takes booleans as inputs and produces booleans as output. Due to its simplicity and ease of understanding, BDT is widely used in classification, data mining and machine learning [44]. It adopts multi-stages or consecutive approaches in the classification procedure. Trees are generated at the first stage of classification. At the second stage, the test sample is discriminated from the root node to the child node with higher probability. The process is recursively executed until the sample is assigned to the leaf node corresponding to a specific category. The defect of the BDT algorithm is that it is sensitive to noise.

Naive Bayes (NB) is a probabilistic classifier based on Bayes theorem (from Bayesian statistics) with strong (naive) independence assumptions between features. It assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature [45]. Naive Bayes classifier is highly scalable, which requires a linear relationship between the number of instances and the number features. Maximum-likelihood training, one of the probability model of Naive Bayes classifiers, evaluates a closed-form expression that requires linear time, rather than expensive iterative approximation used in other classifiers. An advantage of NB is that it requires only a small number of training data to weight the importance of parameters in classification.

Random forest (RF) is an ensemble learning method for classification. It constructs a multitude of trees at the training period, and the final classification results achieved by the most votes in the forest [1]. The training algorithm of random forest applies the general technique of bootstrap aggregating which selects random instances with replacement from the training set. The bootstrapping procedure ensures that even though a single tree's decision is highly sensitive to noise, the average decision of multiple trees would not be influenced, as long as these trees are not correlated.

Support vector machine (SVM) is a supervised method with associated learning algorithms [46]. Generally, SVM implicitly constructs a hyperplane or a set of hyperplanes in a high- or infinite-dimensional space for its inputs in terms of kernel function. A satisfactory separation is often achieved in the hyperplane with the largest distance to the nearest training data point (so-called functional margin). The distance of functional margin negatively correlates with classifier's error rate. Compared with other types of tree algorithms, SVM is capable of classifying complicated problems via different kernels. The main advantage of SVM is its predominant generalization capability in statistical learning.

3. Performance assessment

For a given set of features, the following 10-fold cross-validation procedure was performed:

1 All samples were randomly divided into ten equal sized subsets.
2 In ten subsets, a single subset was retained as test data, and the remaining nine subsets were used as training data.
3 The cross-validation process was then repeated ten times, with each of ten subsets used exactly once as the test data.

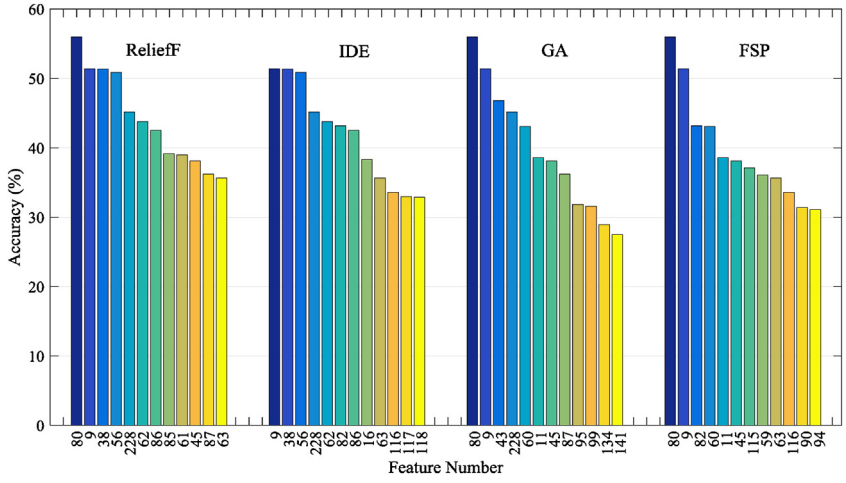

Fig. 1. Selected features of different feature selection methods.
* 1-63: EEG features; 64-87: EOG features; 88-111: EMG features; 112-148: ECG features; 149-187: Coherence; 188-226: Phase angles; 227-232: Mutual information.

All assessment indexes were calculated based on the total results of 10-fold cross-validation. The following indexes used to evaluate the performance of the proposed methods.

Accuracy which indicates the fraction of the total number of correct detections in the sleep scoring. It is defined as,

$$Acc = \frac{TP + TN}{TP + FN + FP + TN}(\%) \tag{14}$$

where, *TP*, *TN*, *FP* and *FN* respectively denote true positives, true negatives, false positives and false negatives [11].

Sensitivity which represents the fraction of positive epochs that are correctly identified by the algorithm [11].

$$Sen = \frac{TP}{TP + FN}(\%) \tag{15}$$

Specificity which denotes the fraction of corresponding negative epochs being correctly rejected [11].

$$Spe = \frac{TN}{FP + TN}(\%) \tag{16}$$

Positive predictive value which is the fraction of correct detections of positive epochs with respect to the total number of positive epochs [11].

$$Ppv = \frac{TP}{TP + FP}(\%) \tag{17}$$

As described above, a total of 232 features were extracted from the signals of EEG, EOG, EMG and ECG. After normalization, these features were fed into four different feature selectors in order to pick out discriminative features. Section 3.1 compared the performance of different feature selectors. In Section 3.2, five different classifiers were compared to ascertain the optimal one for automatic sleep scoring. In Section 3.3, different signals' fusions were compared to explore PSG signals' contribution and to highlight discriminative features of different signals' fusions. Detailed comparative analysis was provided below.

3.1. Performance evaluation of different feature selectors

In order to determine the effective feature selector, a grid search was carried out in terms of the results obtained from four feature selectors (ReliefF, IDE, GA and FSP). The features, selected by feature selectors, were fed into RF classifier one by one to evaluate its capability of distinguishing sleep stages. The top twelve outcomes and its corresponding classification accuracy were displayed in Fig. 1. As Fig. 1 described, the EOG feature named spectral

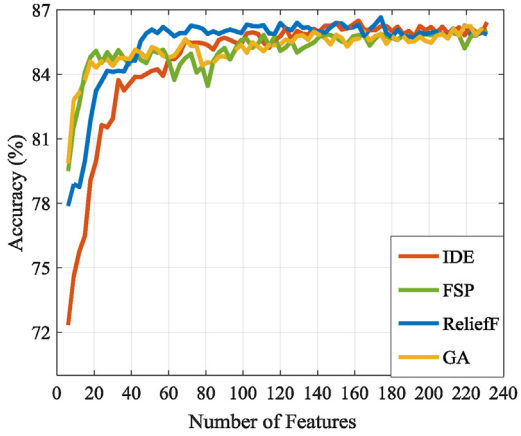

Fig. 2. Mean value of 10-fold cross-validation for specific feature selector.

edge (P: 80; Acc.: 55.88%) achieved the highest accuracy that indicated its outstanding capability of distinguishing sleep stages. The comparison results (showed in Fig. 1) demonstrated that the optimal feature set was obtained by ReliefF, since its elements showed higher discriminative accuracy than those from other selectors. The top twelve features in the optimal feature set were respectively EEG features named zero-crossings, spectral edge, relative power spectral of theta, Petrosian fractal dimension, approximate entropy, permutation entropy and spectral entropy, and EOG features named spectral edge, approximate entropy, permutation entropy and spectral entropy, and the mutual information between EEG and submental EMG. The fact that most of discriminative features were generated from EEG and EOG signals revealed its indispensable role in automatic sleep scoring.

Fig. 2 showed the mean classification accuracy of 10-fold cross-validation corresponding to the numbers of features from four feature selectors. As can be seen from Fig. 2, the automatic sleep analysis required at least four or five features. Meanwhile, an increasing number of features would contribute to classification accuracy. Fig. 2 also revealed that GA, SFP and ReliefF achieved a higher classification accuracy at the initial phase that meant they could select the most discriminative features in the first several steps. However, GA and SFP showed a clumsy increasing of accuracy when the number of features exceeded 20. On the contrary, ReliefF showed sustained growth, and its performance was more stable after the number of features exceeded 60. Based on the results

Table 6
Performance of the proposed method for various feature selection methods and classifier.

Classifiers	Feature selection methods			
	ReliefF	IDE	GA	FSP
KNN	85.26%	82.69%	81.97%	81.41%
Tree	77.08%	76.25%	75.67%	74.93%
Bayes	64.51%	56.95%	70.49%	74.48%
RF	86.24%	85.29%	84.74%	85.17%
SVM	83.00%	82.37%	78.39%	80.73%
Num.Feature	56	70	20	20

of Fig. 1 and Fig. 2, ReliefF was chosen as a predominant feature selector in the following experiments.

3.2. Evaluation of different classifiers

After the optimal feature selector determined, the performance of different classifiers was evaluated in this section. Table 6 depicted the mean classification accuracy of 10-fold cross-validation obtained from five classifiers and their corresponding feature selectors. Here, the number of features was determined based on Fig. 2. In order to decrease the consumption of CPU and memory, we chose as few features as possible in the condition of no significant improvement of accuracy. Table 6 revealed that the optimal performance with the highest accuracy was achieved by random forest classifier (RF) with 56 features selected by ReliefF. The superior performance of RF has been verified in many studies [20,29].

Table 7 presented the confusion matrix of the optimal classification result. NREM-stage S4 was identified with the highest classification sensitivity of 93.3%. Wakefulness, REM and NREM-stage S2 were correctly classified with a sensitivity above 80%, whereas, stage S3 was correctly classified with a sensitivity of 77.6%. Stage S3 was misclassified as its neighbor NREM stage S2 in 27.9% of the instances and misclassified as stage S4 in 11.6% instances. The least recognizable was stage S1 with a sensitivity of 66.3%. Stage S1 was considered as a transient state between wakefulness and "real" sleep. Stage S1 commonly accounted for about 5% of the total sleep duration of a healthy subject. Therefore, the consensus in scoring S1 among experts was quite obscure, compared with other stages [23]. The confusion matrix showed that 8.5% instances of stage S1 were misclassified as REM sleep, 28.6% as S2, and 17.8% as wakefulness. According to our interpretation, the low classification accuracy of stage S1 and stage S3 may also attribute to the imbalanced instances which consequently made the result inclined to the labels with greater numbers.

To demonstrate the performance of proposed methods, we compared the proposed algorithms with those from the other three studies. These algorithms were applied to the present dataset, as described in Section 2.1. The comparison results were shown in Table 8, wherein the classification accuracy of 10-fold cross-validation was shown by the mean value along with the standard deviation in parentheses. As shown in Table 8, the proposed methods, with ReliefF as a feature selector and RF as a classifier, generated the optimal performance in each signal fusion. However, the results generated by referred algorithms were less satisfactory than those reported in the cited studies. The divergence of results might be influenced by different datasets and limited instances. The number of instances in the present article is the least compared to those in the three references.

3.3. Comparison of different signals' fusions

Different fusions of EEG, EOG, EMG and ECG signals were performed to clarify signals' contribution. Table 9 and Fig. 3 summarized the results of different signals' fusions. Table 9 presented the top twelve features from six different signals' fusions. These features were selected by ReliefF. Closer inspection of Table 9 showed that features extracted from EEG and EOG achieved a higher classification accuracy when compared with those from EMG and ECG signals. Meanwhile, the features, namely entropy, spectral edge, Petrosian fractal dimension and zero-crossings showed a crucial role in automatic sleep scoring. In addition, some traditional features named arithmetic mean, standard deviation and Hjorth parameters also showed good discrimination in single modality process.

Fig. 3 showed the accuracy of different signals' fusions. EEG showed the outstanding contribution to sleep scoring. Furthermore, the addition of a few more characteristics from other modalities of signals would still be worth considering. The richness of signal modalities contributed to the increasing accuracy of sleep stages classification. According to the results shown in Fig. 3, the fusion of two modalities PSG signals showed significant improvement in the accuracy. The addition of the third and the fourth modality of signal improved classification accuracy to some extent. No significant difference was found among signal types. The slight improvement, generated by the addition of the fourth signal, might attribute to information saturation. However, the definitive conclusion would require further investigations. Besides, the addition of signals and features was a challenge for feature selector so that a preeminent feature selector was requisite.

To further elucidate multi-modality signals' contributions to automatic sleep scoring, a feature selection process, for four modalities of signals (EEG, EOG, EMG and ECG), was performed to distinguish every pair of sleep stages. The top 15 features selected by ReliefF were shown in Table 10, wherein features sorted in ascending order of discriminative capability. As can be seen from Table 10, the features from EEG accounts for a large proportion. Meanwhile, the features from ECG signal showed its contribution to the distinguishing among W, S1 and R. For mutual-based features, coherence and mutual information between two signals performed well in the discrimination of S1-S2, S1-R and S3-S2. In term of EMG

Table 7
Scoring agreement between manual scoring and the proposed method.

	Technologists' score stage						Sen.
	Stage	W	R	S1	S2	S3	S4	
Proposed	W	394	8	50	35	2	2	80.2%
	R	6	1299	24	99	0	0	90.9%
	S1	26	8	126	30	0	0	66.3%
	S2	21	88	80	1931	159	15	84.2%
	S3	0	0	0	57	343	42	77.6%
	S4	2	2	0	10	66	1122	93.3%
Ppv		87.8%	92.4%	45.0%	89.3%	60.2%	95.0%	
Spe.		99.0%	97.7%	97.4%	93.8%	95.9%	98.8%	
Acc.								86.24%

Table 8
Algorithm performance with different signals' fusions and its comparison with three references.

	EEG	EEG, EOG&EMG	EEG, EOG, EMG&ECG
Reference	59.2715% (±1.4580%) [10]	82.9519% (±1.4870%) [2]	76.1726% (±3.3857%) [23]
Proposed	76.0531% (±0.9056%)	85.3068% (±0.8244%)	86.244% (±1.0725%)

Table 9
Selected features of different signals' fusions.

EEG		EOG		EMG		ECG		EEG & EOG & EMG		EEG & EOG & EMG & ECG	
P	Acc.	P	Acc.	P	Acc.	P	Acc.	P	Acc.	P	Acc.
9	51.43%	80	56.01%	104	40.57%	120	38.88%	80	56.01%	80	56.09%
38	51.38%	81	47.41%	96	37.46%	125	36.74%	228	45.15%	9	51.43%
56	50.91%	72	47.37%	105	37.20%	124	36.56%	62	43.79%	38	51.38%
62	43.79%	82	43.19%	95	31.85%	116	33.59%	82	43.19%	56	50.91%
60	43.09%	86	42.53%	99	31.57%	117	32.95%	86	42.53%	228	45.15%
61	38.96%	85	39.17%	90	31.42%	118	32.89%	85	39.17%	62	43.79%
11	38.60%	74	38.93%	98	30.33%	129	29.24%	61	38.96%	86	42.53%
16	38.35%	87	36.23%	109	30.05%	127	28.89%	74	38.93%	85	39.16%
45	38.10%	66	34.52%	91	30.01%	130	28.39%	11	38.60%	61	38.96%
13	36.76%	64	33.10%	106	29.41%	146	27.42%	45	38.10%	45	38.10%
4	35.65%	67	30.60%	111	28.25%	148	26.84%	87	36.23%	87	36.23%
63	35.64%	75	29.60%	110	28.06%	147	26.77%	63	35.64%	63	35.64%

*1-63: EEG features; 64-87: EOG features; 88-111: EMG features; 112-148: ECG features; 149-187: Coherence; 188-226: Phase angles; 227-232: Mutual information.

Table 10
Selected features for distinguishing specific pair of sleep stages.

Top fifteen features	S4-S3	S4-S2	S4-S1	S3-S2	S3-S1	S2-S1	S4-R	S3-R	S2-R	S1-R	S4-W	S3-W	S2-W	S1-W	W-R
Top.1	80	80	9	80	9	62	9	80	80	80	9	9	80	61	38
Top.2	9	9	38	9	38	82	38	228	228	62	38	38	9	125	61
Top.3	38	56	56	56	56	60	56	62	82	85	56	56	56	44	16
Top.4	56	81	228	228	43	85	228	82	86	125	228	62	81	124	125
Top.5	81	72	62	62	228	61	62	60	85	124	62	60	72	59	44
Top.6	72	228	82	82	62	87	82	86	74	64	82	86	85	116	124
Top.7	228	62	86	60	82	129	60	85	45	95	60	61	61	117	116
Top.8	86	82	61	86	60	123	86	11	52	129	86	16	74	118	117
Top.9	85	86	11	85	86	231	61	45	55	110	61	44	16	40	64
Top.10	61	85	16	61	61	110	11	52	125	187	11	87	44	129	118
Top.11	74	61	45	13	11	147	16	55	124	147	16	46	87	229	95
Top.12	11	74	55	87	16	181	45	13	87	181	45	116	46	123	129
Top.13	13	11	13	129	13	175	13	87	95	175	44	117	116	231	229
Top.14	87	13	63	175	87	169	4	95	129	169	46	118	117	232	231
Top.15	63	63	129	169	129	220	63	129	147	220	63	129	129	147	232

*1-63: EEG features (color: yellow); 64-87: EOG features (color: green); 88-111: EMG features ((color: blue)); 112-148: ECG features (color: red); 149-187: Coherence (color: gray); 188-226: Phase angles (color: gray); 227-232: Mutual information (color: gray).

features, median amplitude (P: 95) and spectral entropy (P: 110) behaved well in distinguishing R stage from W and S1.

4. Discussion

In general, a sleep study records several kinds of signals. Taking full advantage of multi-modality signals is advisable to perform a comprehensive sleep assessment. Some studies have reported that features from multi-modality signals are beneficial to the improvement of scoring accuracy[29]. The effectiveness of multi-modality signals' fusion was illustrated in Fig. 3, Table 8 and Table 10. More specifically, EEG signals contain valuable and interpretable information resembling brain activities. The changes of rhythm waves (such as delta, theta, alpha, beta) reflect the alternation of sleep stages leading to the cyclic pattern of sleep [3,7]. As demonstrated in Table 10, EEG features contributed to the discrimination of most stages. The eyes movement, recorded by EOG, may be very frequent in stage W and REM while being rare during NREM stages[16]. Therefore, EOG features have good performance in differentiating NREM stages from stage W and REM. The stage W presents the highest muscular activity in contrast to REM stage which has the lowest EMG activity[16]. As a result, EMG features are good at distinguishing stage W and stage R. Heart rate decreases with less variability in NREM stages, while increases and becomes more unstable during REM sleep[19]. ECG features are useful for differentiating stage R and W from the others, as shown in Table 10. With the deepening of sleep, the frequency of EEG signals attenuates gradually along with rare eye movements, low EMG activity and slow heart rate. There-

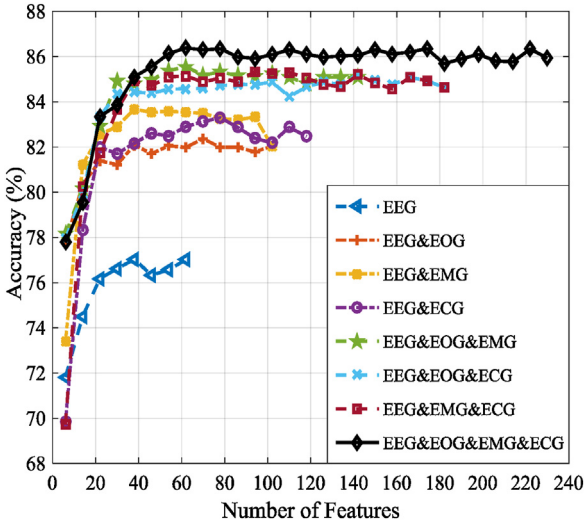

Fig. 3. Mean value of 10-fold cross-validation from different signals' fusions.

fore, four modalities of PSG signals contribute to the discriminant of NREM stage (S1, S2, S3 and S4). Features extracted from the signals of EEG, EOG, EMG and ECG can reveal sleep status from different aspects, which contributes greatly to a higher scoring accuracy in the identification of multi-modality signals.

The present study investigated 232 features (time-based, frequency-based, statistical and nonlinear features) coming from four modalities of PSG signals (EEG, EOG, EMG and ECG) to identify the most discriminative features fusion. The results point to the view that the optimal feature set can be reached by joint fusion of spectral measures (e.g. spectral edge), entropy measures (e.g. approximate entropy, spectral entropy, permutation entropy), fractal measures (e.g. Petrosian fractal dimension) and statistical time-domain features. The good performance of spectral edge has been reported in several studies [23,30,47]. In Fell and his colleagues' study [47], it claimed that delta power and spectral edge were two clinically well established measures used for monitoring sleep cycles. Meanwhile, the time-domain feature named zero-crossings is crucial in automatic sleep scoring [23,47,48]. It has been observed that the zero-crossings, as a rough estimate of average frequency, decreases as sleep goes deeper [48]. Besides, the relative spectral power of theta wave with the frequency band of 4–8 Hz (P: 45) is also important for the automatic sleep scoring since it is a marker of sleep onset [23].

It has been agreed that the classification of S1 is an enormous challenge for virtually every sleep scoring method. From the neurophysiological standpoint, S1 is a transition phase and a mixture of wakefulness and sleep, which is likely to result in the obscurity of neuronal oscillation between S2 and wakefulness (W). Besides, in the REM stage, the cortex generates 40–60 Hz gamma waves, which also occurs in the awake stage [49]. Due to the resemblance of wakefulness and REM, stage S1 is often misclassified as wakefulness or REM by both automatic sleep scoring and human technologist scorers [9], as shown in Table 7. Besides, the unbalanced instances also account for the poor classification accuracy of S1. The duration of each sleep stage varies. As a result, different stages may contain different numbers of epochs. Specifically, as described in Table 1, the number of epochs in S1 and S3 is relatively smaller when compared with other stages. It can be seen that sleep data invariably suffer from the imbalance of instances. As a result, traditional classification models are inclined to classify instances into large groups [50].

5. Conclusion

This study proposed an automatic sleep scoring method, which achieved an encouraging accuracy by employing multiple features from four modalities of polysomnography signals (EEG, EOG, EMG and ECG). In addition, different signals' fusions were investigated and the optimal features were highlighted by searching a large scale of features covering statistical characters, frequency characters, time-frequency characters, fractal characters, entropy characters and nonlinear characters. The present study concluded that the optimal feature set was a joint fusion of multiple characteristics and that the fusion of multi-modality PSG signals contributed to the increasing of classification accuracy. Furthermore, through comparing the performance of different selectors and classifiers, ReliefF and random forest classifier turned out to be reliable candidates for automatic sleep scoring.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 81471742 & 91748105), the Fundamental Research Funds for the Central Universities [DUT16JJ(G)03] in Dalian University of Technology in China, and the scholarships from China Scholarship Council (Nos. 201606060227).

References

[1] S.M. Isa, I. Wasito, A.M. Arymurthy, A. Noviyanto, Kernel dimensionality reduction on sleep stage classification using ECG signal, Int. J. Comput. Sci. Issues 8 (1) (2011) 1178–1181.
[2] S. Khalighi, T. Sousa, G. Pires, U. Nunes, Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels, Expert Syst. Appl. 40 (no. 17) (2013) 7046–7059.
[3] A. Rechtschaffen, A. Kales, A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects, US National Inst. Heal. Publ., Washington DC, 1968.
[4] A. Karni, D. Tanne, B. Rubenstein, J. Askenasy, D. Sagi, Dependence on REM sleep of overnight improvement of a perceptual skill, Science 265 (5172) (1994) 679–682.
[5] J.E. Kang, et al., Amyloid-β dynamics are regulated by orexin and the sleep-wake cycle, Science 326 (no. 5955) (2009) 1005–1007.
[6] A.V. Nedeltcheva, M.C. Program, C. Disorders, Metabolic effects of sleep disruption, links to obesity and diabetes, Curr. Opin. Endocrinol. Diabetes Obes. 21 (no. 4) (2014) 293–298.
[7] R.B. Berry, et al., Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events, J. Clin. Sleep. Med. 8 (no. 5) (2012) 597–619.
[8] T. Penzel, R. Conradt, Computer based sleep recording and analysis, Sleep. Med. Rev. 4 (2) (2000) 131–148.
[9] A.R. Hassan, M.I.H. Bhuiyan, A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features, J. Neurosci. Methods 271 (2016) 107–118.
[10] A.R. Hassan, M.I.H. Bhuiyan, Automatic sleep scoring using statistical features in the EMD domain and ensemble methods, Biocybern. Biomed. Eng. 36 (1) (2016) 248–255.
[11] B. Şen, M. Peker, A. Çavuşoğlu, F.V. Çelebi, A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms, J. Med. Syst. 38 (3) (2014).
[12] M. Diykh, Y. Li, P. Wen, EEG sleep stages classification based on time domain features and structural graph similarity, IEEE Trans. Neural Syst. Rehabil. Eng. 24 (11) (2016) 1159–1168.
[13] S.F. Liang, C.E. Kuo, Y.H. Hu, Y.H. Pan, Y.H. Wang, Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models, IEEE Trans. Instrum. Meas. 61 (6) (2012) 1649–1657.
[14] S.I. Dimitriadis, C. Salis, D. Linden, A novel, fast and efficient single-sensor automatic sleep-stage classification based on complementary cross-frequency coupling estimates, bioRxiv 7 (July4) (2017) 1–41.
[15] J. Shi, X. Liu, Y. Li, Q. Zhang, Y. Li, S. Ying, Multi-channel EEG-based sleep stage classification with joint collaborative representation and multiple kernel learning, J. Neurosci. Methods 254 (2015) 94–101.
[16] E. Estrada, H. Nazeran, J. Barragan, J.R. Burk, E.A. Lucas, K. Behbehani, EOG and EMG: Two important switches in automatic sleep stage classification, Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc. (2006) 2458–2461.
[17] S. Özşen, Classification of sleep stages using class-dependent sequential feature selection and artificial neural network, Neural Comput. Appl. 23 (5) (2013) 1239–1250.

[18] J. Zhang, Y. Wu, J. Bai, F. Chen, Automatic sleep stage classification based on sparse deep belief net and combination of multiple classifiers, Trans. Inst. Meas. Control. 38 (4) (2016) 435–451.
[19] F. Ebrahimi, S.K. Setarehdan, H. Nazeran, Automatic sleep staging by simultaneous analysis of ECG and respiratory signals in long epochs, Biomed. Signal. Process. Control. 18 (2015) 69–79.
[20] I. Hermawan, M.S. Alvissalim, M.I. Tawakal, W. Jatmiko, An integrated sleep stage classification device based on electrocardiograph signal, Adv. Comput. Sci. Inf. Syst. (ICACSIS), 2012 Int. Conf. (2012) 37–41.
[21] B. Yılmaz, et al., Sleep stage and obstructive apneaic epoch classification using single-lead ECG, Biomed. Eng. Online 9 (39) (2010) 39.
[22] K. Šušmáková, A. Krakovská, Discrimination ability of individual measures used in sleep stages classification, Artif. Intell. Med. 44 (3) (2008) 261–277.
[23] A. Krakovská, K. Mezeiová, Automatic sleep scoring: A search for an optimal combination of measures, Artif. Intell. Med. 53 (1) (2011) 25–33.
[24] A.A. Gharbali, S. Najdi, J.M. Fonseca, Investigating the contribution of distance-based features to automatic sleep stage classification, Comput. Biol. Med. 96 (March) (2018) 8–23.
[25] R.S. Vasan, D. Levy, PhysioBank, PhysioToolkit, and PhysioNet components of a new research resource for complex physiologic signals, Heart Fail. (2000) 2118–2121.
[26] M. Giovanni Terzano, et al., Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep, Sleep. Med. 3 (2) (2002) 185.
[27] M.G. Terzano, D. Mancia, M.R. Salati, G. Costani, A. Decembrino, L. Parrino, The cyclic alternating pattern as a physiologic component of normal NREM sleep, Sleep 8 (2) (1985) 137–145.
[28] C. Vidaurre, N. Krämer, B. Blankertz, A. Schlögl, Time domain parameters as a feature for EEG-based brain-computer interfaces, Neural Netw. 22 (no. 9) (2009) 1313–1319.
[29] R. Boostani, F. Karimzadeh, M. Nami, A comparative review on sleep stage classification methods in patients and healthy individuals, Comput. Methods Programs Biomed. 140 (2017) 77–91.
[30] S.A. Imtiaz, E. Rodriguez-Villegas, A low computational cost algorithm for REM sleep detection using single channel EEG, Ann. Biomed. Eng. 42 (11) (2014) 2344–2359.
[31] M.K. Uçar, M.R. Bozkurt, C. Bilgin, K. Polat, Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques, Neural Comput. Appl. (2016) 1–16.
[32] R. Esteller, J. Echauz, T. Tcheng, B. Litt, B. Pless, Line length: an efficient feature for seizure onset detection, Annu. Int. Conf. IEEE Eng. Med. Biol. 2 (January 2015) (2001) 1707–1710.
[33] R. Acharya, U.O. Faust, N. Kannathal, T. Chua, S. Laxminarayan, Non-linear analysis of EEG signals at various sleep stages, Comput. Methods Programs Biomed. 80 (1) (2005) 37–45.
[34] S. Vorobyov, A. Cichocki, Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis, Biol. Cybern. 86 (4) (2002) 293–303.
[35] X.F. Liu, Y. Wang, Fine-grained permutation entropy as a measure of natural complexity for time series, Chin. Phys. B 18 (7) (2009) 2690–2695.
[36] J. Pan, W.J. Tompkins, A Real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. BME-32 (3) (1985) 230–236.
[37] A. Noviyanto, S.M. Isa, I. Wasito, A.M. Arymurthy, Selecting features of single lead ecg signal for automatic sleep stages classification using correlation-based feature subset selection, Int. J. Comput. Sci. Issues 8 (1) (2011) 1178–1181.
[38] L. Yu, H. Liu, Efficient feature selection via analysis of relevance and redundancy, J. Mach. Learn. Res. 5 (2004) 1205–1224.
[39] M. Robnik-Siknja, I. Kononeko, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learn. 53 (2003) 23–69.
[40] Y. Lei, Z. He, Y. Zi, A new approach to intelligent fault diagnosis of rotating machinery, Expert Syst. Appl. 35 (4) (2008) 1593–1600.
[41] C.M. Fonseca, P.J. Fleming, Genetic algorithms for multiobjective optimization: formulation, discussion and generalization, Icga 93 (July) (1993) 416–423.
[42] A.W. Whitney, A direct method of nonparametric measurement selection, IEEE Trans. Comput. C-20 (9) (1971) 1100–1103.
[43] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[44] B. Şen, E. Uçar, D. Delen, Predicting and analyzing secondary education placement test scores: A data mining approach, Expert Syst. Appl. 39 (10) (2012) 9468–9476.
[45] S. Isa, M. Fanany, W. Jatmiko, A. Murini, Feature and model selection on automatic sleep apnea detection using ECG, Int. Conf. Comput. Inf. Syst. (February) (2010) 357–362.
[46] C. Cortes, V. Vapnik, Support vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[47] J. Fell, J. Röschke, K. Mann, C. Schäffner, Discrimination of sleep stages: A comparison between spectral and nonlinear EEG measures, ElectroEncephalogr. Clin. Neurophysiol. 98 (no. 5) (1996) 401–410.
[48] F. Machado, Automatic Sleep Staging Based on Classification Methods, 2015.
[49] J. Horne, Why REM sleep? Clues beyond the laboratory in a more challenging world, Biol. Psychol. 92 (2) (2013) 152–168.
[50] C. Seiffert, T.M. Khoshgoftaar, J. Van Hulse, A. Napolitano, RUSBoost: A hybrid approach to alleviating class imbalance, IEEE Trans. Syst. 40 (1) (2010) 13.