Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2019) 27: 605 – 614
© TÜBİTAK
doi:10.3906/elk-1710-28

Optimal set of EEG features in infant sleep stage classification

Maja CIC[1], Mario MILICEVIC[2,*], Igor MAZIC[2]
[1]Ericsson Nikola Tesla d.d., Zagreb, Croatia
[2]Department of Electrical Engineering and Computing, University of Dubrovnik, Dubrovnik, Croatia

Received: 03.10.2017 • **Accepted/Published Online:** 22.10.2018 • **Final Version:** 22.01.2019

Abstract: This paper evaluates six classification algorithms to assess the importance of individual EEG rhythms in the context of automatic classification of infant sleep. EEG features were obtained by Fourier transform and by a novel technique based on the empirical mode decomposition and generalized zero crossing method. Of six evaluated classification algorithms, the best classification results were obtained with the support vector machine for the combination of all presented features from four EEG channels. Three methods of attribute ranking were assessed: relief, principal component analysis, and wrapper-based optimized attribute weights. The outcomes revealed that the optimal selection of features requires one feature from every significant frequency band, either a spectral feature or a frequency dynamic feature. This means that reducing the number of features will have a minimal impact on the classification accuracy.

Key words: Empirical mode decomposition, generalized zero crossing, sleep classification, feature selection, support vector machine

1. Introduction

Sleep has a strong influence in numerous areas of life, for infants as well as for children and adults. Sleep structures reflect predispositions or a state of the whole organism. To support visual sleep stage classification, experts have been working on developing the methods of automatic sleep stage classification for years. Polysomnographic (PSG) recordings including electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG) recordings [1] are visually scored by experts using the Rechtschaffen and Kales guidelines [2]. As the literature indicates, the EEG signal is one of the most important and frequently used signals for analyzing sleep staging, regardless of whether the scoring is manual or automatic. The recording is first divided into 20- or 30-s epochs, which are subsequently classified as wakefulness (W), rapid eye movement (REM) sleep, and non-REM (NREM). NREM sleep can be further divided into stages S1, S2, S3, and S4. Because PSG recordings include multiple signal channels and are visually examined, sleep stage scoring is expensive, prone to human error, and often tedious and time-consuming [1]. The extensive literature could be viewed through three criteria: papers oriented towards the evaluation of a certain feature extraction method, papers oriented towards finding the optimal set of features from a large variety of features, and papers oriented towards the evaluation of classifiers. In recommendations for the visual scoring of sleep, sleep stages are characterized with predominant frequency waves, but it is possible to directly connect one sleep stage with one EEG feature [3]. There are papers that address EEG characteristics over specific frequency bands together with some other measures, but it is not possible to reach a definite conclusion about the significance of each

*Correspondence: mario.milicevic@unidu.hr

EEG frequency band. The spectral characteristics of the EEG signal together with the fractional dimension and nonlinear measures were derived in [4], where the authors studied all possible combinations of the characteristics on a single EEG channel. They found that the most suitable characteristics for sleep stage separation are the spectral entropy, the fractional dimension, and the delta power (power in 0.5–4 Hz frequency band). A single channel EEG-based automated sleep scoring method was also proposed in [5], where the authors first decomposed EEG signal segments into subbands using tunable-Q wavelet transform (TQWT). In [6] the authors chose five research papers and practically implemented appropriate methods on a publicly available sleep EEG dataset. According to the achieved results, entropy of wavelet coefficients along with the random forest classifier were chosen as the best feature and classifier, respectively. The mentioned feature and classifier provide 87.06% accuracy for healthy subjects. Review paper [1] summarizes the variety of methods that have been used for the detection of sleep stages using EEG signals, including an overview of the processing techniques. Comparing different studies is very difficult mainly because different EEG databases were used in them. In addition, many studies used fewer than 10 recordings, which makes it difficult to properly implement machine learning methods. It should also be mentioned that deep learning algorithms have been used recently, in studies where the classification is based on the visualization of the EEG signal (e.g., by spectrogram). The obtained results, such as in [7], show that this approach still does not yield better results compared to classical methods.

The capacities of 74 individual features to differentiate among various sleep stages were evaluated in [8]. The most successful features were the fractal exponent and delta/beta power ratio. On the basis of the same extracted features, the authors aimed to identify an optimal mix of several of these 74 features in [9]. In a comparative study on the classification of sleep stages based on EEG signals [10], the 25 listed features included entropy-based features, nonlinear features, statistical features of time characteristics, and Hjorth parameters. Another 16 features were obtained over four frequency bands by discrete wavelet transform. In three conducted experiments, only the 4–8 Hz band was selected to be among the first 10 features. We also looked for proof of the importance of EEG frequency bands from investigations of EEG signals in other areas. In the largest systematic study of seizure detection features [11], 65 different features were derived. The best performing features were line length and relative power in the 12.5–25 Hz frequency band. Another example is a comprehensive comparison of features for the cue-based brain–computer interface (BCI) [12]. The band power was the best feature for many BCI subjects, because the standard alpha (8–12 Hz) and beta (16–24 Hz) bands are not necessarily optimal for all subjects. It is obvious that, in these papers, different EEG frequency bands were selected to be placed among the most suitable features. The aim of this paper is to assess the importance of individual rhythms of EEG signals in infant sleep through the evaluation of six machine learning algorithms and three feature relevance ranking methods.

2. Materials and methods

In our previous paper [13] we established a novel approach for EEG signal quantification based on the empirical mode decomposition (EMD) process. Following the established approach, in this work features are calculated for every frequency band and taken independently, so that the significance of the individual frequency band can be assessed through a method for ranking the relevance of attributes. The results then could point toward the importance of individual rhythms of EEG signals that would mark out an optimal set of EEG features for automatic sleep stage classification. A method for a sleep stage classification was created through the following steps: 1) EEG signal recording, filtering, and artifact removing; 2) feature extraction; 3) sleep stage classification; 4) feature attribute selection; 5) final evaluation.

2.1. EEG signal

In this study, the features were evaluated from EEG sleep signals of 32 healthy babies aged 3 months. Informed parental consent was obtained and the procedure was approved by the Ethics Committee of the Clinical Hospital Split (Croatia). An EEG was recorded during one sleep cycle, in the absence of the wake phase. The dataset was created separating the EEG signal in segments of 30-s nonoverlapping periods (epochs). Artifact segments and outliers were removed during the preprocessing step. Sleep was visually scored in REM and NREM sleep stages, as a baby's brain is not yet fully matured at this age. The EEG channels were sampled at 256 Hz. A high-pass filter of 0.3 Hz, a low-pass filter of 70 Hz, a notch filter of 50 Hz, and sensitivity of 7 µV/mm were used to record the EEG. EEG caps with 12 gold electrodes, placed according to the International '10–20' system, were used. In [14], the authors reported that quiet sleep was perceptible mainly on the C3 and C4 electrodes and other states on the T3 and T4 electrodes. Thus, the measured signals were analyzed only from channels Fp1–C3, Fp2–C4, Fp1–T3, and Fp2–T4. This dataset contained 1988 segments per EEG channel, where 572 segments belonged to the NREM sleep stage, so stratified sampling was used.

Here is a brief description of three types of EEG features: two features based on EMD, and the power spectral density (PSD) feature based on Fourier transform.

2.2. Empirical mode decomposition (EMD)

The EMD method [15] is an adaptive decomposition method with which any signal can be disassembled to the intrinsic mode functions (IMFs) and the residue. An IMF function has the following two properties: 1) the number of extrema and zero crossings are either equal or differ at most by one; 2) the mean value of two envelopes associated with the local maxima and minima is zero.

To decompose the signal into IMF functions, a procedure known as sifting is performed in four iterative steps [13]:

1) Identifying all local maximums and minimums of the input signal,

2) Creating the upper envelope by interpolation between the maximums and lower envelope by interpolation of the minimums,

3) Calculating the mean value of the upper and lower envelopes,

4) Subtracting the signal obtained after step three from the input signal.

The resulting signal is an IMF candidate. Next, the IMF candidate is tested to determine if it satisfies the two IMF properties mentioned above. If it does not satisfy the properties of an IMF, the process is repeated with the resulting signal as the input. After the IMF properties have been fulfilled, the obtained function is declared an IMF and it is subtracted from the input signal. The residue of the subtraction constitutes a new input signal, and the sifting process begins again. Because each subsequent IMF contains lower frequencies than the previous IMF functions, the EMD procedure can be continued until all IMFs are found, or stopped when a satisfying number of IMFs is obtained. In this study, the EMD procedure was stopped when an EEG signal segment had been decomposed into seven IMFs, discarding the IMFs of frequencies less than 0.3 Hz.

2.3. Generalized zero crossing (GZC)

The GZC method [16] was employed to calculate the instantaneous frequencies. The GZC method defines all zero crossings and local extrema points as critical points. Their numbers and positions are used for signal analysis and decision-making as follows. For each series of signal points between two successive critical points, seven different period values are calculated [13]:

1) The time between two consecutive zero crossings of the same type (positive to negative or negative to positive) or between two consecutive maximum or minimum values can be treated as a single period. For any point on the time axis, four different values of this type of period can be determined and marked as P_1.

2) The time between two consecutive zero crossings or two consecutive extrema points (maximum to minimum or vice versa) is considered a half-period. For any point on the time axis, two different values of the half-period can be calculated and marked as P_2.

3) The time between the extrema point and zero crossing (or vice versa) can be considered a quarter-period, which can be calculated and marked as P_4.

For a given series of signal points between two critical points, the instantaneous frequency can be computed as follows:

$$\bar{\omega} = \frac{1}{7}\left\{\frac{1}{4P_4} + \left(\frac{1}{2P_2^1} + \frac{1}{2P_2^2}\right) + \left(\frac{1}{P_1^1} + \frac{1}{P_1^2} + \frac{1}{P_1^3} + \frac{1}{P_1^4}\right)\right\} \tag{1}$$

2.4. EMD-based feature extraction using GZC

As mentioned above, the EMD decomposition method was applied to every 30-s epoch, so the result is seven IMFs per epoch. Each IMF lasts 30 s. Two types of features were extracted from these seven IMFs: median frequency (IMF_{med}) and number of instantaneous frequency changes (IMF_{ch}). The GZC method was used for calculating instantaneous frequencies for each IMF. The average for instantaneous frequencies for each IMF represents 7 IMF_{med} features and the number of instantaneous frequency changes for each IMF represents 7 IMF_{ch} features. Justification for the second feature are Rechtschaffen and Kales rules that illustrate the frequency characteristics of each sleep stage. The different frequency dynamics of an EEG signal in the NREM and REM sleep stages can be correlated with instantaneous frequency changes in IMFs.

2.5. Relative power spectral density (RSD)

In standard frequency bands, known as delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), sigma (12–16 Hz), and beta (16–24 Hz), the power of the EEG signal was calculated in the frequency domain using the following equation:

$$P = \frac{1}{M \cdot N} \cdot \sum_{k=1}^{M} \left|X_k\left(e^{j\omega}\right)\right|^2 \tag{2}$$

Here, N is 7680 recorded samples in one epoch, M is 8192 points for fast Fourier transform, and $X_k(e^{j\omega})$ are samples of EEG signals in the frequency domain. For each segment, five RSD features were calculated by dividing the obtained power with the sum of powers in all five standard frequency bands. These five RSD features represent the energy distribution among the standard frequency bands.

2.6. Combined features

Three combined features were created by joining the spectral features and the time-frequency features. The first combined feature was the combination of all three types of used features: RSD, IMF_{med}, and IMF_{ch} (altogether 19 features, "all" for short). The second combination was RSD and IMF_{med} and the third combination was RSD and IMF_{ch}.

2.7. Classification algorithms

In this paper we studied the performances of six commonly used classification algorithms, which are well known and were described in detail in the referenced literature. The original support vector machine (SVM) algorithm was proposed by Vapnik and the standard implementation was published by Cortes and Vapnik [17] in 1995. The basic idea is to find the best hyperplane (the one with the largest margin) that separates n-dimensional data into two classes introducing transformed kernel induced feature space, which translates the data into a higher dimensional space where the data are separable. Random forests (RFs) are a combination of tree predictors, such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [18]. An artificial neural network (NN) is defined as a computing system made up of a number of simple, highly interconnected processing elements (neurons) with linear or nonlinear transfer functions. These elements process information by their dynamic state response to external inputs [19]. K nearest neighbors (k-NN) is a type of instance-based learning [20], or lazy learning, where the function is only approximated locally and all computation is deferred until classification. A case is classified by a majority vote of its neighbors, where the case is assigned to the class most common among its K nearest neighbors measured by a distance/similarity function. A naive Bayes (NB) classifier [21] is based on Bayes' theorem with (naive) independence assumptions between predictors. The repeated incremental pruning to produce error reduction (rule induction) algorithm incrementally generates rule sets directly from training datasets [22].

2.8. Performance metrics

The sleep stage classification of babies is considered a two-class problem (REM and NREM stages). The sensitivity and specificity are given for the REM stage. The performance metrics used in this paper are calculated as follows:

Accuracy ACC = (TN + TP)/(TN + TP + FP + FN),

Sensitivity SE = TP/(TP + FN), and Specificity SP = TN/(TN + FP)

TP is the number of true positively classified segments of REM stage, TN is the number of true negative classified segments (the number of accurately classified NREM segments), FP is the number of false positives, and FN is the number of false negatives.

2.9. Feature selection

The process of selecting a subset of relevant features is performed to simplify the model, to make it more comprehensible and easier to interpret by researchers. In addition, enhanced generalization can be obtained by reducing overfitting along with shorter training times. Feature selection techniques imply that the dataset contains features that are either not equally relevant or redundant. Our combined feature created from all presented features has 19 (RSD(5), IMF_{med}(7), IMF_{ch}(7)) features for one EEG channel. In order to determine the importance of frequency bands from all utilized features, three methods of feature ranking were assessed: the relief algorithm, PCA, and wrapper-based approaches.

The relief algorithm [23] estimates a feature according to its ability to discriminate instances based on the relevance weight assigned to each feature. The algorithm randomly samples instances from the training set and calculates the relevance values based on the distance between the selected instance and the two nearest instances of the same and opposite class. The principal component analysis (PCA)-based approach [24] uses the eigenvector corresponding to the eigenvalue with the highest magnitude to rank original features and to choose the most sensitive features from the original set. These filter approaches are usually very efficient, but, in the context of the target machine learning method, the selected feature subset may not perform well. Therefore, the wrapper-based optimized attribute weights approach is introduced [25]. It determines the fitness of a feature subset by actually training a classifier (in this case with an SVM algorithm), usually with better performances at the cost of higher computational complexities.

3. Results and discussion

The first tests in this study were statistical analysis of features. One-way ANOVA analysis of variance revealed a statistically significant difference (with significance level $\alpha = 0.001$) in all extracted features between NREM and REM sleep stages. The results in terms of the obtained $F_{1,1986}$ values for each feature means of extracted features for the whole EEG dataset are all $F_{1,1986} > 18.46$ for NREM and REM sleep stages at four EEG channels. Thus, the corresponding P-value for all features is zero. These results indicate the ability of the extracted features to distinguish between sleep stages. In this study each of the described feature types was fed into the classifiers separately and jointly, seriatim: five RSD features, seven IMF_{med} features, seven IMF_{ch} features, the combination of RSD and IMF_{med} features, the combination of RSD and IMF_{ch} features, and the combination of all 19 features. The Figure shows 10-fold cross-validation accuracy results of combined features for the Fp2–T4 channel. The best results in all three evaluation parameters were obtained with the SVM classifier (from 81.94% to 89.09% accuracy).

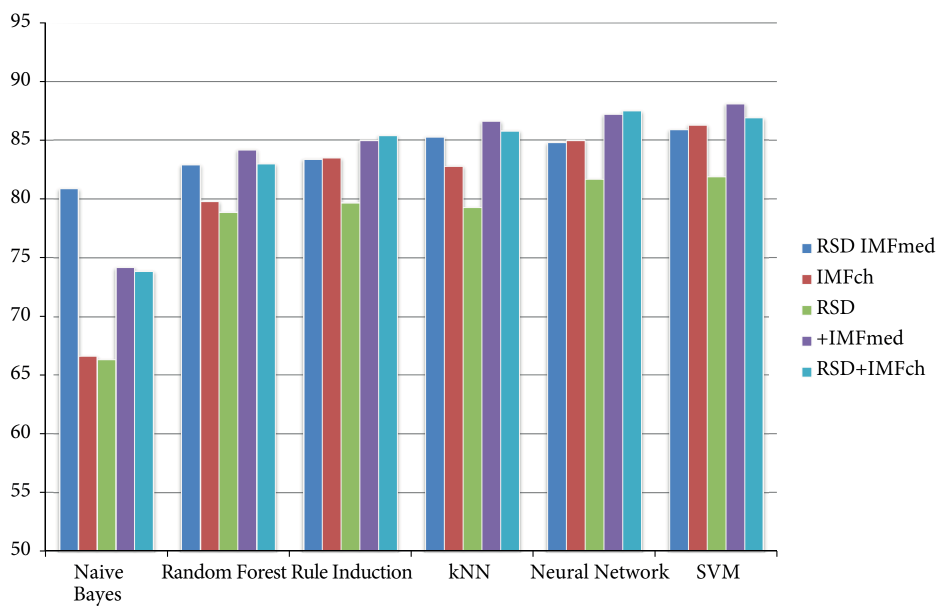

Figure. Accuracy of the six classification methods for all features at Fp2–T4 channel.

Combining the IMF_{ch} with the RSD feature, the NN was able to classify sleep stages with 87.48% accuracy, which is an even better result than that obtained with the SVM algorithm (86.92%). The best result of 89.09% accuracy was achieved by the SVM with the combination of all types of features (RSD, IMF_{med}, and IMF_{ch}).

After the systematic testing of different classifiers, three methods for an optimal feature combination assessment were evaluated. As can be seen in the Figure, the SVM classification method provided the highest values; therefore, the optimized attribute weights (forward selection) wrapper method was implemented for the SVM algorithm. Ranking of the features acquired through this method, as well as through relief and PCA, is presented in Table 1. From the gained results we discerned that, for a computationally not demanding set and yet a set that would bring more accuracy than visual scoring, the optimal set could include the seven features. For such a set of features that would suitably differentiate sleep stages, two of the reduction methods selected five features from IMF frequency characteristics and only two features from the power spectral density feature type. This confirms the quality of our method for describing the EEG frequency dynamics based on the EMD process.

Table 1. The seven best ranked features from tree methods.

	Relief	PCA	Optimized attribute weight
1	theta	sigma	IMF1 med
2	delta	beta	IMF3 ch
3	IMF1 med	alfa	beta
4	IMF3 ch	delta	IMF5 med
5	IMF1 ch	IMF1 ch	IMF6 ch
6	IMF2 ch	theta	IMF7 ch
7	IMF3 med	IMF6 ch	sigma

Of the three evaluated methods, the optimized attribute weight method selected the set of seven features that achieved the best SVM classification accuracy of 87.33% (Table 2), which is the result with an only 2% lower value than the best SVM result achieved with the combination of all 19 features. Looking at the frequency details of the selected set, two chosen RSD features represent sigma and beta frequency bands, i.e. frequencies greater than 12 Hz, and which carry a smaller amount of spectral power compared to any of the three remaining bands. However, the IMF frequency characteristics were taken from the 1st, 3rd, 5th, 6th, and 7th IMFs, whose ranges spread over the whole EEG frequency spectrum.

Table 2. The classification results (accuracy, sensitivity, and specificity) for the SVM classification with the use of the seven best ranked features, at Fp2–T4 channel.

	Relief	PCA	Optimized attribute weight
acc	86.77%	86.52%	87.33%
sp	72.55%	80.42%	78.32%
se	92.51%	88.98%	90.96%

The set of the seven features gained through the relief method enclose the frequency dynamic characteristics from the 1st to the 4th IMF, which constitute the higher part of the EEG frequency range. However, two selected RSD features represent theta and delta frequency bands (frequencies lower than 4 Hz), thus supplementing the part of the EEG frequency range not covered with IMFs. The set of the seven features gained through the PCA method enclose all five RSD features, taking into account the whole EEG frequency range. Two features of IMF frequency dynamics are of the 1st and 6th IMF.

All these results indicate the importance of the whole EEG signal frequency spectrum. The seven features gained through the optimized attribute weight method listed in Table 1 were evaluated through six classification algorithms. The results in Table 3 show the same ranking of the algorithms in terms of the classificatory success in accuracy. The SVM algorithm again achieved the best results, with 2.22% enhancement in accuracy with regard to NN, k-NN, or the rule induction algorithm.

Table 3. The performance details of the six classification methods for the best ranked seven features gained through optimized attribute weight method, at Fp2–T4 channel.

Classifier	acc	sp	se	Statistical difference	P-value
Naive Bayes	72.99	41.43	85.73	extremely stat. signif.	less than 0.0001
Random forest	84.36	61.01	93.79	extremely stat. signif.	less than 0.0001
Rule induction	85.21	72.55	90.32	**not** statist. significant	0.2678
k-NN	85.11	74.65	89.43	**not** statist. significant	0.7713
NN	85.11	80.07	87.15	extremely stat. signif.	less than 0.0001
SVM	87.33	78.32	90.96	**not** statist. significant	0.8501

The EEG signal analysis from four channels obtained the best SVM classification result from the Fp1–C3 channel and showed similarity of results over channels in the values of accuracy, specificity, and sensitivity with standard deviation of 0.6%, 1.87%, and 0.86%, respectively, as presented in Table 4. Additionally, the classification methods were tested for the selection of the entire feature space. The feature vector was generated from all three feature types containing 19 features and for all four EEG channels, thus becoming a vector of 76 (19×4) features. Again, the best results were obtained with the SVM method (91.45% accuracy, 95.13% sensitivity, and 82.34% specificity).

Table 4. Accuracy, sensitivity, and specificity of SVM classification with the combined features of all 19 attributes, at four EEG channels.

EEG channels	acc	sp	se	Statistical difference	P-value
Fp2-T4	89.09	80.24	92.66	**not** statist. significant	0.5871
Fp1-T3	88.78	83.74	90.82	**not** statist. significant	0.1080
Fp2-C4	88.48	81.29	91.38	**not** statist. significant	0.3549
Fp1-C3	89.89	84.09	92.23	**not** statist. significant	0.2042

Finally, the optimized attribute weight forward method was applied to these 76 features. The sleep stage classification results of the SVM method for the first seven features are 90.09% accuracy, 93.15% sensitivity, and 82.52% specificity.

After executing the automatic classification of sleep stages, for testing the statistical significance of the results, McNemar's nonparametric test of the difference of the results for two dependent samples (visual scoring and automatic classification) was performed. For the best ranked features, McNemar's test P-values are provided in relevant tables (Tables 3 and 4) with other performance metrics.

McNemar's nonparametric test for statistical significance suggested that the NB, RF, and rule induction algorithms have extremely statistically significant differences in visual and automatic scoring results, which gives them low priority as options for a sleep stage classification algorithm. The NN algorithm showed a not statistically significant difference in results for IMF$_{med}$, IMF$_{ch}$, and the combination of all 19 features. The NN algorithm showed a very to extremely statistically significant difference in results for RSD and the combinations of frequency dynamic features with RSD. However, the SVM algorithm showed opposite results: a statistically very significant difference in the results for individual feature types, and a not statistically significant difference in the results for all tested combinations of features. Besides the different statistical significance of the results for different algorithms, the results gained in this study bring new insights about the accuracy of the algorithms used for sleep stage classification and about the significance of individual EEG frequency bands for sleep staging. To highlight this, we cite two comparative studies on the classification of sleep stages based on EEG signals.

In [9], Krakowská and Mezeiová evaluated 74 features in the search for the optimal set, and in the end they concluded that for different subject types a new search for the optimal selection of classifying measures may be necessary. In [10], Şen et al. sought to determine a method for the identification of features that can best represent the EEG data and they stated that a set of the selected features made it possible to identify the classification algorithm that provides the highest accuracy. In contrast, our results show that for all the presented features (relative PSD, IMF frequencies, and combined features), the ranking of the tested algorithms in terms of the classificatory success in accuracy remained the same. This way, the study exposes feature extraction as the most important phase of the complete signal classification process. The results of the feature selection procedures in the literature showed a big difference in the set of the selected attributes depending on the feature selection process, but in this paper the outcomes from three different selection methods bring an analogous conclusions: the optimal set of seven features should spread over the whole EEG frequency range.

4. Conclusions

EEG features in this study were obtained in a classical way by Fourier transform and an EMD-GZC-based method. The best result was obtained with all features from three presented types of features from four EEG channels; the classification accuracy, sensitivity, and specificity were 91.45%, 95.13%, and 82.34%, respectively. With seven features optimally selected from the set of 76 features from all analyzed channels, the classification accuracy, sensitivity, and specificity were 90.09%, 93.15%, and 82.52%, respectively. High sensitivity reveals that the proposed method is effective to classify REM and NREM stages using mixed types of features. The study outcome shows that the reduction from 76 features to the optimal combination of seven features from four channels does not significantly affect classification performance, and an assessment about the importance of individual rhythms of EEG signals for sleep staging could be brought. The outcomes of the three different feature selection methods emphasized that the optimal set of features requires one feature from every EEG frequency band, either a spectral feature or a frequency dynamic feature to represent each frequency band.

References

[1] Aboalayon KAI, Faezipour M, Almuhammadi WS, Moslehpour S. Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation. Entropy 2016; 18: 272.

[2] Rechtschaffen A, Kales A. A Manual of Standardized Terminology, Techniques and Scoring System of Sleep Stages in Human Subjects. Bethesda, MD, USA: National Institutes of Health.

[3] Lajnef T, Chaibi S, Ruby P, Aguera PE, Eichenlaub JB, Samet M, Kachouri A, Jerbi K. Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. J Neurosci Methods 2015; 250: 94-105.

[4] Piryatinska A, Woyczynski WA, Scher MS, Loparo KA. Optimal channel selection for analysis of EEG-sleep patterns of neonates. Comput Meth Prog Bio 2012; 106: 14-26.

[5] Hassan AR, Subasi A. A decision support system for automated identification of sleep stages from single-channel EEG signals. Knowl-Based Syst 2017; 128: 115-124.

[6] Boostani R, Karimzadeh F, Torabi-Nami M. A comparative review on sleep stage classification methods in patients and healthy individuals. Comput Meth Prog Bio 2017; 140: 77-91.

[7] Vilamala A, Madsen KH, Hansen LK. Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring. In: MLSP 2017; USA. New York, NY, USA: IEEE. pp. 1-6.

[8] Šušmáková K, Krakovská A. Discrimination ability of individual measures used in sleep stages classification. Artif Intell Med 2008; 44: 261-277.

[9] Krakovská A, Mezeiová K. Automatic sleep scoring: a search for an optimal combination of measures. Artif Intell Med 2011; 53: 25-33.

[10] Şen B, Peker M, Çavuşoğlu A, Çelebi FVA. Comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. J Med Syst 2014; 38: 18.

[11] Logesparan L, Casson AJ, Rodriguez-Villegas E. Optimal features for online seizure detection. Med Biol Eng Comput 2012; 50: 659-669.

[12] Boostani R, Graimann B, Moradi MH, Pfurtscheller G. A comparison approach toward finding the best feature and classifier in cue-based BCI. Med Biol Eng Comput 2007; 45: 403-412.

[13] Čić M, Šoda J, Bonković M. Automatic classification of infant sleep based on instantaneous frequencies in a single-channel EEG signal. Comput Biol Med 2013; 43: 2110-2117.

[14] Gerla V, Paul K, Lhotska L, Krajca V. Multivariate analysis of full-term neonatal polysomnographic data. IEEE T Inf Technol Biomed 2009; 13: 104-110.

[15] Huang NE, Shen Z, Long S, Wu M., Shih H, Zheng Q, Yen NC, Tung C, Liu H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proc R Soc Lond A 1998; 454: 903-995.

[16] Huang NE. Computing frequency by using generalized zero-crossing applied to intrinsic mode functions. United States Patent No. US 6990436 B1, 2006.

[17] Cortes C, Vapnik V. Support vector networks. Mach Learn 1995; 20: 273-297.

[18] Breiman L. Random forests. Mach Learn 2001; 45: 5-32.

[19] Caudill M. Neural network primer: Part I. AI Expert 1989; 2: 46-52.

[20] Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. Mach Learn 1991; 6: 37-66.

[21] John GH, Langley P. estimating continuous distributions in Bayesian classifiers. In: Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence; 18–20 August 1995; Montreal, Canada. pp. 338-345.

[22] Cohen WW. Fast effective rule induction. In: Proceedings of the Twelfth International Conference on Machine Learning 1995. pp. 115-123.

[23] Kira K, Rendell L. The feature selection problem: traditional methods and a new algorithm. In: Proceedings of the Tenth National Conference on AI; 1992. pp. 129-134.

[24] Malhi A, Gao R. PCA-based feature selection scheme for machine defect classification. IEEE T Instrum Meas 2004; 53: 1517-1525.

[25] Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell 1997; 97: 273-324.