33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# Efficient Feature Selection for Sleep Staging Based on Maximal Overlap Discrete Wavelet Transform and SVM

Sirvan Khalighi, Teresa Sousa, Dulce Oliveira, Gabriel Pires, Urbano Nunes

*Abstract* — **In this paper, a novel algorithm is proposed with application in sleep/awake detection and in multiclass sleep stage classification (awake, non rapid eye movement (NREM) sleep and REM sleep). In turn, NREM is further divided into three stages denoted here by S1, S2, and S3. Six electroencephalographic (EEG) and two electro-oculographic (EOG) channels were used in this study. The maximum overlap discrete wavelet transform (MODWT) with the multi-resolution Analysis is applied to extract relevant features from EEG and EOG signals. The extracted feature set is transformed and normalized to reduce the effect of extreme values of features. A set of significant features are selected by mRMR which is a powerful feature selection method. Finally the selected feature set is classified using support vector machines (SVMs). The system achieved 95.0% of average accuracy for sleep/awake detection. As concerns the multiclass case, the average accuracy of sleep stages classification was 93.0%.**

## I. Introduction

The study of sleep pattern through whole night sleep recordings has consistently been an important research topic. Scoring of sleep stages was done on the basis of Rechtschaffen and Kales standard (R&K) until recent dates [1]. The American Academy of Sleep Medicine (AASM) determined new criteria in the scoring of sleep based on the R&K rules. Sleep-wake cycle is categorized in awake, non rapid eye movement (NREM) and rapid eye movement (REM) sleep stages. NREM sleep is further divided into three stages: S1, S2 and S3 [2]. Sleep scoring by experts is a very time consuming task and normally may require hours to classify a whole night recording. It is also a somewhat subjective procedure in which the concordance between human experts can vary greatly [3]. Accordingly, the development of automatic systems is highly desirable to save time and improve agreement levels of sleep stage scoring. Several studies have reported the development of automatic sleep stage classification (ASSC) methods based on electroencephalographics (EEG) records, sometimes in combination with electro-oculogram (EOG) and electromyogram (EMG) records. Hilbert-Huang Transform and Wavelet Transform were applied to extract harmonic parameters from EEG signals in [4]. Ebrahimi *et al.* [5] used neural networks and wavelet packet coefficients to discriminate between different sleep stages. Doroshenkov *et al.* [6] have developed a classification algorithm based on Hidden Markov Models using only EEG signals, achieving the best accuracy result for REM stage. Zoubek *et al.* [7] suggested feature selection algorithms to find the most relevant features from polysomnography (PSG) signals. In another work, Gunes *et al.* [8] proposed a K-means clustering method with a reported agreement of 55.88% to 82.15% in discriminating six sleep stages. Jo *et al.* [9] proposed a genetic fuzzy classifier applied to discriminate four stages: wakefulness, shallow sleep, deep sleep, and REM stages. Values of classification accuracy vary widely among ASSC methods reported in scientific publications. Rigorous comparisons between the reported systems cannot be done since they differ in recording conditions and validation procedures.

An ASSC algorithm has been developed aiming to improve sleep stage classification accuracy in two applications: sleep/awake detection and multiclass sleep stage classification. In both cases the classification is based on six EEG and two EOG channels by using temporal, parametric and time-frequency features. The maximal overlap discrete wavelet transform (MODWT) uses a multi-resolution analysis (MRA) to decompose EEG and EOG signals at different resolutions. A support vector machine (SVM) classifies transformed and normalized features that are selected by a minimum-redundancy maximum-relevance (mRMR) algorithm [10]. Furthermore, a median filter is used to enhance the classification accuracies.

This work has been supported by the QREN funded project SLEEPTIGHT, with FEDER reference CENTRO-01-0202-FEDER-011530, and by the Portuguese Technology Science Foundation (FCT) under project grant RIPD/ADA/109661/2009.

The authors are with the Institute for Systems and Robotics (ISR-UC), University of Coimbra, 3030-290 Coimbra, Portugal.
e-mails: {skhalighi,tsousa,doliveira,gpires,urbano}@isr.uc.pt

## II. Materials and Methodology

The proposed system can be organized in various interoperating parts as illustrated in Fig. 1.

### A. Data Acquisition

Data from all-night PSG records, each with a duration around 8 hours (acquired by a SomnoStar Pro; Viasys SensorMedics), were provided by the Laboratory of Sleep from Hospital Centre of Coimbra. All EEG and EOG recordings were performed with a sampling rate of 200 Hz. The dataset comprises data from fourteen subjects, ten males and four females with ages between 22 and 79 years old (mean = 56 years; std = 17.11 years). The international 10-20 standard electrode placement system was used for EEG recording. Six EEG and two EOG channels were used in our evaluation: F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, O2-A1, right EOG (R-EOG)-A1 and left EOG (L-EOG)-A2 for all the subjects.

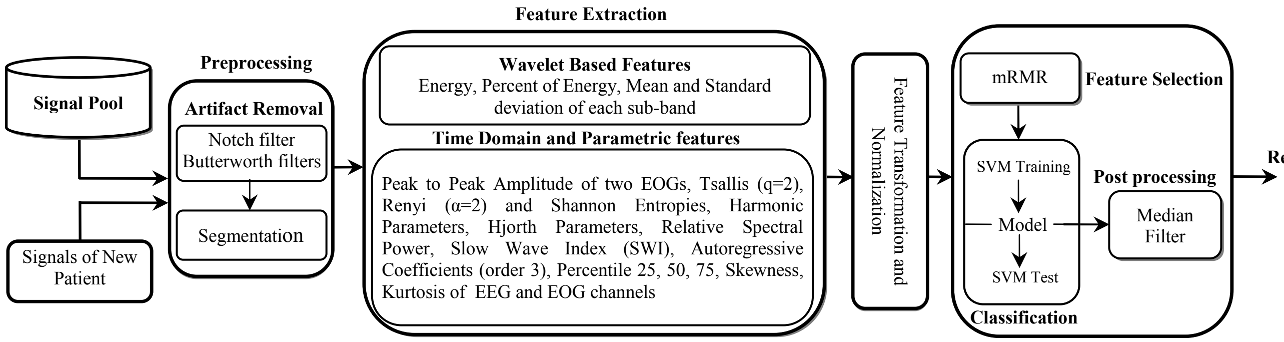


Fig.1 System Architecture

### B. Preprocessing and Feature Extraction

Before computing the feature vectors from the sampled EEG signals, the recorded signals are filtered to eliminate noise and undesired background EEG, by using a notch filter at 50 Hz and a bandpass Butterworth filter with lower cutoff of 0.5 Hz and higher cutoff of 45 Hz. The signals were segmented in 30 seconds epochs. EEG is traditionally analyzed in the frequency domain, since each sleep stage is characterized by a specific pattern of frequency contents. However further useful information can be extracted from temporal analysis of EOG and EEG signals. Moreover, EEG signals are non-stationary; therefore time-frequency transformations like wavelets are very useful. Thus, after preprocessing, some features are extracted using several methods in the time-frequency, temporal and frequency domain.

#### 1) Wavelet Based Features

The discrete wavelet transform (DWT) generates coefficients that are local in both time and frequency. The maximum overlap discrete wavelet transform (MODWT) [11] is a DWT in which the operation of subsampling from an output filter is omitted. By giving up of the orthogonality property, the MODWT gained new features; although losing efficiency in computation, this transform does not have any restriction on the sample size and it is shift invariant. As a result, in the MODWT, the wavelet and scaling coefficients must be rescaled to retain the variance preserving property of the DWT. Although the components of MODWT are not mutually orthogonal, their sum is equal to the original time series. Additionally, the detail and smooth coefficients of a MODWT are associated with zero phase filters. This means that temporal events and patterns in the original signal are meaningfully aligned with the features in the multi resolution analysis. Furthermore, the MODWT is invariant to circularly shifting the original time series. Hence, shifting the time series by an integer unit will shift the wavelet and scaling coefficients by the same amount. This property does not hold for the DWT because of the subsampling involved in the filtering process. In addition, the MODWT does not induce the phase shifts within the component series. The MODWT wavelet variance estimator is also preferred because it has been shown to be asymptotically more efficient than an estimator based on the DWT [12].

In our study a MODWT of depth 6 with Daubechies order four (db4) is applied to every 30 second epochs with a sampling rate of 200 Hz. The frequency ranges are broken down within δ range (<4 Hz), θ range (4–8 Hz), α range (8–13 Hz) and β range (13–30 Hz). A set of statistical wavelet based features (see Fig.1) are extracted to represent the time–frequency distribution of the EEG and EOG signals.

#### 2) Frequency and Temporal Features

Regarding the importance of spectral and temporal analysis, some features are extracted as suggested in [4], [7], [13], [14], [15]. These features are discussed in the experimental results section.

### C. Feature Transformation and Normalization

The extracted features are transformed and normalized in order to reduce the influence of extreme values. The transformation methods applied to each feature are described in [16]. It was verified that some of those transformations improved the classification results. After a thorough experimental evaluation of each transform operator over extracted features, it was empirically verified that the best classification results were attained with the transform $\boldsymbol{X} = \log(\boldsymbol{Y})$, where $\boldsymbol{Y}$ denotes the feature matrix, and $\boldsymbol{X} = \{x_{ij}; i = 1,2,\dots,N \text{and } j = 1,2,\dots,M\}$ (where $N$ and $M$ denote the number of subjects and the number of features respectively) is the transformed feature matrix. Thereby this transform was adopted in the overall sleep staging system. To avoid features in greater numeric ranges dominating those in smaller numeric ranges, and numerical difficulties during the classification; in the selection process, each feature of the transformed matrix $\boldsymbol{X}$ is independently normalized to the [0, 1] range by applying

$$\bar{x}_{ij} = x_{ij}/(max(\boldsymbol{x}_i) - min(\boldsymbol{x}_i)) \quad (1)$$

where $\boldsymbol{x}_i$ is a vector of each independent feature [17]. Effects of transformation and normalization on the classification process are discussed in the experimental results section.

### D. Feature Selection, Classification and Post processing

Larger numbers of high-dimensional feature vectors make the classification process, more complex and less reliable due to feature's redundancy. So there is the need of reducing the number of features which can be effectively done by mRMR feature selection method [18]. Our experimental results lead us to the same conclusion. In our study, SVM

TABLE I

SELECTED FEATURES USING mRMR. ALL FEATURES ARE EXTRACTED FROM 6 EEG AND 2 EOG CHANNELS EXCEPT THE PEAK TO PEAK AMPLITUDE THAT WAS EXTRACTED ONLY FROM THE EOG CHANNELS.

| Feature Extraction Methods | Selected / Total | Feature Extraction Methods | Selected / Total |
|---|---|---|---|
| Wavelet based features [7] | 47/160 | Skewness [7] | 2/8 |
| Harmonic parameters [4] | 39/120 | Percentile25,50,75 [7] | 1/24 |
| Relative power [7] | 32/40 | Kurtosis [7] | 0/8 |
| Spectral analysis [13] | 26/104 | Rényi entropy [15] | 0/8 |
| Hjorth parameters [14] | 14/24 | Tsallis entropy [15] | 0/8 |
| AR coefficients [15] | 10/48 | Peak to peak amplitude [21] | 0/2 |
| Shannon entropy [15] | 5/8 | | |

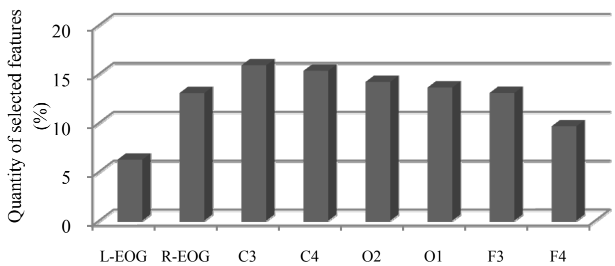


Fig.2 Proportion of selected features of each channel (EEG and EOG) in a total of 176 selected features.

[19] were adopted to handle the classification process. In general, when input data is too noisy, SVM shows reliable classification results. Moreover, an SVM can provide a good generalization performance for classification problems despite that it does not incorporate the problem-domain knowledge. In order to eliminate non stationary short-term transients, a postprocessing stage by means of a median filter is applied as described in [6].

## III. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the proposed algorithm was assessed using the fourteen subjects dataset mentioned in section II.*A*. In our experiments, a fourth order Daubechies with MODWT decomposition was adopted. Libsvm toolbox [20] with sigmoid kernel was used in the classification phase. The sigmoid degree and C parameters were set to 0.13 and 1.25 respectively, as they produced the best empirical results. The classification accuracy was determined by using Leave-One subject-Out Cross-validation (LOOCV).

The extracted feature sets and corresponding selected features, using mRMR method, are presented in Table I. The methods that were applied to extract features from the spatial and frequency domains are also shown in Table I. A total of 570 (71 per EEG channel plus 72 per EOG channel) features were extracted for each subject. The transformed and normalized feature matrix is fed into the feature selector. The total number of selected features by mRMR method was 176, which has provided the best average accuracy when applying a grid search. As illustrated in Table I most relevant features are extracted from MODWT decomposition (47 selected features) and Harmonic parameters (39), and the least effective ones are Kurtosis, Rényi and Tsallis entropies and Peak to Peak amplitude.

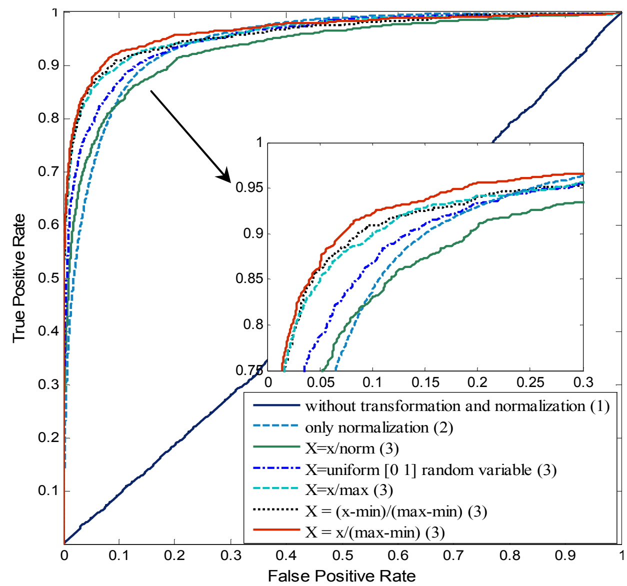


Fig.3 ROC curves corresponding to (1) without any transformation and any normalization; (2) without transformation but applying normalization $\bar{x}_{ij} = x_{ij}/(max(x_i) - min(x_i))$; (3) with different normalization methods over the transformed features by log(x).

TABLE II

AGREEMENT EPOCH MATRIX OBTAINED WITH THE AUTOMATIC SLEEP STAGE CLASSIFICATION USING THE SELECTED SET OF FEATURES

| | | *Visual scoring* | | | | |
|---|---|---|---|---|---|---|
| | **Stages** | *Awake* | *S1* | *S2* | *S3* | *REM* |
| *Auto scoring* | *Awake* | 2974 | 152 | 1 | 21 | 104 |
| | *S1* | 182 | 695 | 218 | 3 | 139 |
| | *S2* | 9 | 347 | 3081 | 332 | 82 |
| | *S3* | 0 | 0 | 254 | 2115 | 15 |
| | *REM* | 48 | 117 | 27 | 0 | 844 |

According to the AASM visual scoring rules, the frontal electrodes are the best for detecting K-complexes, central electrodes for detecting sleep spindles and occipital electrodes for alpha waves [2]. Considering the total features for each EEG channel (71 features) and for each EOG channel (72 features), we found that the most discriminative channels were C3, C4 and O2, as shown in Fig.2.

Receiver operating characteristic (ROC) curves in Fig.3 show the effectiveness of applying different normalization methods on extracted features. As it is shown, the performance of the system improved when normalization (1) was applied over all features.

Table II gives a detailed comparison between automatic and visual scoring. The number of false positives (FP) was the highest in classification of S2 followed by S1 and stage awake. The maximum incidence of false negatives (FN) was in S1, followed by S2 and S3. On the other hand, the number of true positives (TP) was the highest in S2, and the number of true negatives (TN) was the highest in REM stage due to the smaller number of epochs in this stage.

In Table III, the results of statistical analysis of our ASSC are presented, namely sensitivity, specificity, accuracy and the confidence interval (CI). The results were obtained stage by stage based on the values of Table II. The best sensitivity

TABLE III
STATISTIC ANALYSIS RESULTS OF MULTICLASS CLASSIFICATION

| | *Stages* | | | | | | |
|---|---|---|---|---|---|---|---|
| | *Awake* | *S1* | *S2* | *S3* | *REM* | *Total* | *CI (95%)* |
| Sensitivity | 92.561 | 53.013 | 86.037 | 85.593 | 71.284 | 77.698 | 12.416 |
| Specificity | 96.747 | 94.813 | 90.586 | 97.104 | 98.185 | 95.487 | 2.351 |
| Accuracy | 95.604 | 90.153 | 89.201 | 94.685 | 95.476 | 93.024 | 2.425 |

TABLE IV
AGREEMENT EPOCH MATRIX OBTAINED WITH THE AUTOMATIC AWAKE VERSUS SLEEP CLASSIFICATION USING THE OPTIMAL SET OF FEATURES

| | | *Visual scoring* | |
|---|---|---|---|
| | **Stages** | *Awake* | *Sleep* |
| *Auto scoring* | *Awake* | 2857 | 233 |
| | *Sleep* | 356 | 8314 |

value corresponds to stage awake, which means that our algorithm has a good ability to detect the awake stage. S1 in most of the cases is misclassified as S2, awake and REM, which leads S1 to attain the lowest sensitivity (53%). Specificity has the highest and lowest values for REM and S2 stages respectively. The mean accuracy of sleep stages classification is 93.024% with Cohen's Kappa (*k)* of 0.815.

Our algorithm shows better performance in subjects with a large number of awake epochs, because the classification results are biased by this stage. The classifier fails to discriminate the stages that show similar neurophysiological features. For instance, stage awake and S1 have both alpha rhythms; stages S1 and S2 have a common frontal prominence of beta and slow rolling eye movements; stages S2 and S3 frequently show the same delta activity.

Considering only sleep/awake classification, the achieved accuracy was 94.99% ± 7.119 (*k* = 0.872), as computed from Table IV. For five of the subjects the accuracy was above 97%, which is a high value even when considering manual scoring agreement between experts.

## IV. CONCLUSION AND FUTURE WORK

Several feature extraction methods have been applied and combined with an SVM classifier. Features with higher positive impact in classification accuracy were the MODWT decomposition and harmonic parameters, associated with central and occipital areas. Transformation and normalization in the feature domain played an important role in the remarkably improvement of classification accuracy.

The multiclass classification based on LOOCV reached an average accuracy of 93.02%. The classification fails mainly on stages S1 and S2 because of their similarities. This issue can be improved by applying some heuristic rules in the classification algorithms. In sleep/awake detection it was achieved an average accuracy of 95% (agreement with expert visual scoring). To a more robust performance assessment, the classification algorithm has to be validated in a larger database. As a future work it will be pursued the integration of heuristic rules in the ASSC.

## REFERENCES


[1] A. Rechtschaffen, A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects", *Bethesda, MD, U.S. National Institute of Neurological Diseases and Blindness, Neurol. Inform. Netw.*, 1968.

[2] C. Iber, Ancoli-Israel S, Chesson A., Quan SF, "The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications", *1th: Westchester, Illinois: American Academy of Sleep Medicine*, 2007.

[3] N. Nicolaou, J. Georgiou, "The use of permutation entropy to characterize sleep electroencephalograms", *Clinical EEG and Neuroscience*, 42(1), pp. 24–28, Jan. 2011.

[4] WC. Tang, SW. Lu, CM. Tsai CY. Kao, HH. Lee, "Harmonic parameters with HHT and wavelet transform for automatic sleep stages scoring", *International Journal of Biological and Medical Sciences*, pp.176–179, 2007.

[5] F. Ebrahimi, M. Mikaeili, E. Estrada, H. Nazeran, "Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients", *in IEEE EMBS'08,* 2008.

[6] L. Doroshenkov, V. Konyshev, S. Selishchev, "Classification of human sleep stages based on EEG processing using hidden markov models", *in Biomedical Engineering,* Vol.41(1), pp. 25–28, 2007.

[7] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, F. Chapolot, "Feature selection for sleep/wake stages classification using data driven methods", *Biomedical Signal Processing and Control*, Vol. 2, issue 3, 2007, Page 171-179.

[8] S. Gunes, K. Polat, S. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting", *Expert Systems with Applications*, Vol.37,Issue 12, 2010.

[9] H. Jo, J. Park, C. Lee, S. Ann, S. Yoo, "Genetic fuzzy classifier for sleep stage identification", *Computers in Biology and Medicine,* Vol. 40(7), pp. 629–634, July 2010.

[10] H. Peng, F. Long, C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-Redundancy*", IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27(8), pp. 1226–1238, 2005.

[11] Percival, D. B., and Walden, A. T., *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2000.

[12] Saif Ahmad, "Temporal Pattern Identification and Summarization Method for. Complex Time Serial Data," Ph.D. dissertation, Information Extraction and Multimedia Group Dep. of Comp. School of Elec. and Phy. Sci. Uni.of Surrey Guildford, Surrey GU2 7XH, UK.

[13] A. Yilmaz, A. Alkan, M. Asyali, "Applications of parametric spectral estimation methods on detection of power system harmonics", *Electric Power Systems Research*, Vol.78 (4), pp. 683-693, 2008.

[14] K. Ansari-Asl, G. Chanel, T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood", *in, EUSIPCO'07*, pp. 1241–1245, Sept. 2007.

[15] E. Arbabi, MB. Shamsollahi, "Comparation between different types of features used for classification in BCI", *12th International Conference on Biomedical Engineering*, *ICBME'05,* Dec. 2005.

[16] G. Becq, S. Charbonnier, F. Chapotot, A. Buguet, L. Bourdon, P. Baconnier, "*Comparison between five classifiers for automatic scoring of human sleep recordings*", In: S.K. Halgamuge, L. Wang (Eds.), Studies in Computational Intelligence (SCI), Vol.4: Classification and Clustering for Knowledge Discovery, Springer–Verlag, pp. 113–127, 2005.

[17] S. Aksoy, R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval", *Pattern Recognition Letters*, 22, pp. 563–582, 2001.

[18] C. Yun, D. Shin, H. Jo, J. Yang, S. Kim, "An experimental study on feature subset selection methods,", *in 7th IEEE International Conference on Computer and Information Technology*, 2007.

[19] J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, Vol. 2, 1998.

[20] C. Chang and C.J. Lin, LIBSVM: a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[21] J. Virkkala, J. Hasan, A. Värri, SL. Himanen, K. Müller, "Automatic detection of slow wave using two channel electro-oculography", *Journal of Neuroscience Methods*, 160, pp. 171–177, 2007.