Improving Sleep Stage Classification from Electroencephalographic Signals by Fusion of Contextual Information

Iosif Mporas, Anastasia Efstathiou, and Vasileios Megalooikonomou, *Member*, *IEEE*

***Abstract*— In this article we present a fusion architecture for the automatic classification of sleep stages. The architecture relies on time and frequency domain features which are processed by dissimilar classifiers. The initial predictions of each classifier are refined by using fusion of the prediction estimations together with temporal contextual information of the electroencephalographic signal. The experimental results showed that the proposed architecture achieved approximately 95% sleep stage classification accuracy, which corresponds to an improvement of 5% comparing to the best performing single classifier.**

I. Introduction

In modern society the often stressful lifestyle, causes mental distress at the office and in other daily activities. Sleep problem becomes one of the popular healthy problems and except influencing the physical health also affects the mental health of people. Effective sleep results to comfort and relief from stress, while the quality of sleep is more important than the quantity, and the falling asleep period is important in obtaining good quality sleep [1]. The inability to sleep, or to stay awake, is a distress that can disrupt a normal way of living.

The understanding of the sleep mechanism is essential for the diagnosis as well as the therapy of sleep disorders. Sleep is a losing state as temporary, partial, and periodic in the form of that can be returned with various forced stimulus of the communication of the organism with the environment [2]. It can also be defined by the decreasing of motor activity, the decreasing of response with stimulus, and to be easy recycling as behavioral [3, 4].

From the last century, neuroscientists describe the human sleep as a succession of five recurring stages: stage REM (rapid eye movement) and NREM (non-rapid eye movement) stage N1, stage N2, stage N3 and stage N4, without including the awake stage. These stages are characterized by rapid changes in the amplitude and the frequencies (rhythms) of the electroencephalographic (EEG) signal [5]. Specifically, N1 presents a frequency transition of the brain from alpha waves (approx. 8–13 Hz, common in the awake state) to theta waves (approx. 4-7 Hz). N2 is characterized by sleep spindles (approx. 11-16 Hz) and K-complexes [4, 6]. In N3 stage a minimum of 20% delta waves (approx. 0.5-2 Hz) appear, having a peak-to-peak amplitude of more than 75 μV, while in N4 stage delta waves reach 50%. REM stage accounts for 20–25% of total sleep time duration in most human adults and presents a rapid low-voltage EEG (higher frequency saw-tooth waves). Sleep scientists consider that the study and analysis of the sleep stages for the treatment of sleep related disorders is of great importance.

The analysis of the sleep EEG waveforms is performed by expert neurologists. Since the typical duration of sleep recordings is 6-8 hours manual/visual investigation of the EEG signals is a tedious and time-consuming procedure, which will heavily rely on the degree of experience of the sleep expert thus resulting to subjective analysis results (i.e. biased to the specific neurologist). Due to the disadvantages of manual analysis of sleep recordings and after the progress in signal processing, pattern recognition and data mining, computer-based solutions have been proposed for automatic sleep stage scoring.

In general, the proposed architectures are based on the extraction of short-time frames (epochs) obtained from EEG signals, described by spectral analysis methods [2, 12, 14, 15], wavelet transformation [7, 8,], autoregressive analysis [5, 11], band-specific energy [9, 10], entropy [10, 13]. Analysis of sleep stages has also been performed using electrocardiogram (ECG) [1, 9, 16] signals as well as polysomnographic (PSG) data [10, 15], where the study of sleep and wakefulness is based on concurrent recordings of multiple sensors including EEG, ECG, electro-oculogram (EOG) and electromyogram (EMG). The labeling of the corresponding feature vectors is typically based on machine learning algorithms for classification, such as support vector machines (SVMs) [9, 13], artificial neural networks (NNs) [7, 8, 10, 12, 16], decision trees (DTs) [2] and others such as the Gaussian mixture models (GMMs) [14], the hidden Markov models (HMMs) [17], the Kullback-Leibler divergence (KL) [5] the Bayesian classifier (BN) and the k-nearest neighbors (knn) [10].

In this article we present a two-leveled architecture for automatic classification scoring of sleep stages from electroencephalographic data. In particular, we initially perform short-time analysis of EEG using widely used time-domain and frequency-domain features and process them with different base classification algorithms. In the second level we fuse the produced score results of the base classifiers as well as we append to the fusion scheme contextual information from previous and optionally from next epochs.

The remaining of this article is organized as follows: In Section 2 we explain the proposed architecture for automatic sleep stage classification. In Section 3 we offer description of the data used in the evaluation and describe the experimental

I. Mporas is with the Dept. of Computer Eng. and Informatics, Univ. of Patras, 26500 Rion, Greece and with the Computer and Informatics Eng. Dept., TEI of Western Greece (corresponding author to provide phone: +30-2610-996993; fax: +30-2610-969018; e-mail: imporas@ upatras.gr).

A. Efstathiou is with the Dept. of Computer Eng. and Informatics, Univ. of Patras, Greece (e-mail: nefstathiou@upatras.gr).

V. Megalooikonomou is with the Dept. of Computer Eng. and Informatics, Univ. of Patras, Greece (e-mail: vasilis@ceid.upatras.gr).

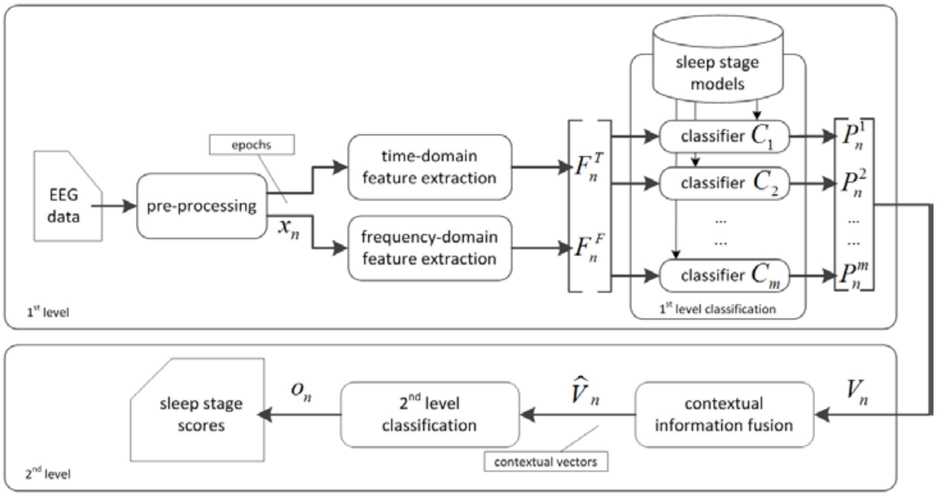

Figure 1. Block diagram of the proposed two-leveled architecture for sleep stage classification from EEG data using contextual information.

setup. In Section 4 we describe the experimental results and conclude in Section 5.

II. Sleep Stage Classification using Contextual Information

The proposed architecture for sleep stage classification is based on the use of time and frequency based EEG features in order to exploit the differences in the frequencies and the amplitudes of the EEG signal among the sleep stages, as described in Section 1. Additionally, the architecture exploits the different discriminative ability of dissimilar classification algorithms, which may offer complementary information. Moreover, it uses the temporal context for each epoch in order to reduce outliers in results and improve performance at the time intervals around the sleep stages changes. The concept of the architecture is illustrated in Figure 1.

As can be seen in Figure 1, the EEG signal is initially preprocessed with frame blocking of constant length w samples and time-shift $t \le w$ samples, thus decomposing the EEG signal to a sequence of N epochs $X = \{x_n\}$, $x_n \in \mathbb{R}^w$, with $1 \le n \le N$. Here we consider single channel EEG analysis. Each epoch is processed by time domain and frequency domain signal parameterization algorithms producing corresponding time, $F_n^T \in \mathbb{R}^{|F_n^T|}$, $1 \le n \le N$, and frequency, $F_n^F \in \mathbb{R}^{|F_n^F|}$, $1 \le n \le N$, based feature vectors.

At the 1-st level of the architecture we employ a number of dissimilar classifiers, C_m, $1 \le m \le M$, which use as input the concatenated time-frequency features produced above, i.e. $F_n \in \mathbb{R}^{|F_n^T|+|F_n^F|}$, $1 \le n \le N$. Each classifier produces for every epoch the corresponding probability of belonging to one of the K classes (i.e. the sleep stages), S_k, $1 \le k \le K$, i.e. the output of the m-th classifier is $P_n^m \in \mathbb{R}^K$.

At the 2-nd level, in order to exploit the complementary information of the dissimilar classifiers the 1-st level's sleep stage estimations are concatenated into one fusion vector for each epoch, i.e. $V_n \in \mathbb{R}^{K \cdot M}$. The fusion vector is appended with temporal contextual information from the previous r and the succeeding r epochs, resulting to $\hat{V}_n \in \mathbb{R}^{K \cdot M \cdot (2 \cdot r+1)}$. The final estimation of the sleep stage is produced by the 2-nd level's classification function f, i.e.

$$o_n = f\left(\hat{V}_n, r\right) \tag{1}$$

where $o_n \in \{S_1, ..., S_k, ..., S_K\}$, which thus labels each epoch with a sleep stage category.

The above architecture for sleep stage scoring can be extended to more channels without any loss of generality. The selection of the contextual width, r, can be performed either using only previous estimations for online (real-time) operation or using both previous and succeeding ones for offline analysis of existing sleep recordings.

III. Experimental Setup

The sleep stage classification architecture presented in Section 2 was evaluated using "The Sleep EDF Database" [18] provided by Physionet [19]. It is a collection of 61 polysomnograms (PSGs) with accompanying hypnograms. The recordings were obtained from Caucasian males and females (21 - 35 years old) without any medication and they contain Fpz-Cz and Pz-Oz EEG, sampled at 100 Hz. The Fpz-Cz channel was used here. Hypnograms were manually scored according to Rechtschaffen & Kales based on Fpz-Cz/Pz-Oz EEG.

During pre-processing the EEG recordings were frame blocked to epochs of 15, 30, 45 and 60 seconds length, without time-overlap between successive epochs. For each epoch well-known and widely used time-domain and frequency domain features were extracted [20]. In particular, the EEG waveform was parameterized using the following features: (i) time-domain features: minimum value, maximum value, mean, variance, standard deviation, percentiles (25%, 50%-median and 75%), interquartile range, mean absolute deviation, range, skewness, kyrtosis, energy, Shannon's entropy, logarithmic energy entropy, number of positive and negative peaks, zero-crossing rate, and (ii) frequency-domain features: 6-th order autoregressive-filter (AR) coefficients, power spectral density, frequency with maximum and minimum amplitude, spectral entropy, delta-theta-alpha-beta-gamma band energy, discrete wavelet transform coefficients with mother wavelet function Daubechies 16 and decomposition level equal to 8, thus resulting to a feature vector of dimensionality equal to 55.

Both for the 1-st and 2-nd level classification we relied on dissimilar algorithms in order to maximize complementarily. Specifically, we used the k-nearest neighbor classifier (knn), the C4.5 decision tree (C4.5), the multilayer perceptron neural network (MLP) with 3-layers, and the support vector machine (SVM) using radial basis function kernel. The selection of the parameters C and γ for the kernel were empirically selected to C equal to 15 and γ equal to 0.1 after grid search. For the implementation of the algorithms we relied on the WEKA toolkit implementations [21].

IV. EXPERIMENTAL RESULTS

The sleep stage classification architecture using contextual information presented in Section 2 was evaluated following the experimental setup described in Section 3. In order to evaluate the performance of the the proposed sleep stage classification architecture we examined the percentage of correctly classified epochs. In order to avoid overlap between training and test datasets we followed a ten-fold cross validation setup.

As a first step, we examined the performance of the four evaluated classification algorithms with different size, w, of frame length (epoch). The results (in percentages) are tabulated in Table I and are considered here as baseline. The best performing setup is shown in bold. In all evaluated epoch sizes there is no overlap between successive epochs.

TABLE I. SLEEP STAGE CLASSIFICATION ACCURACY FOR DIFFERENT CLASSIFICATION ALGORITHMS USING DIFFERENT EPOCH LENGHTS

Classifier	w **=15s**	w **=30s**	w **=45s**	w **=60s**
k-nn	85.94	87.50	87.10	86.32
C4.5	84.13	86.31	85.92	84.73
SVM	87.66	**89.44**	88.84	88.31
MLP	87.12	88.02	87.63	87.09

As can be seen in Table I, the best performing algorithm is the SVM with best achieving accuracy equal to 89.44%. The best performing SVM is followed by the second discriminative algorithm, MLP, with 88.02%. The other two algorithms achieved significantly worse performance. With respect to the size to the epoch, the 30 seconds length achieved the best performance across all evaluated classification algorithms. The experimental results show that the use of longer epochs slightly decreases the performance, while the use of shorter epochs significantly increases the error rate. The superiority of SVMs is probably owed to the fact that they circumvent the "curse of dimensionality" and the fact that SVM training always finds a global solution [22], in contrast to other classifiers (e.g. neural networks) where many local minima usually exist and thus are sensitive to the distributional specifics of the training data.

In a second step we examined the degree of complementarity of information between the four evaluated algorithms. Specifically, we examined the percentage of the instances (epochs) that were misclassified for each algorithm, and was correctly classified by one of the other classification algorithms. The results are tabulated in Table II. The value of the cell at row i and column j shows the percentage of the epochs that were misclassified from the i-th classifier but were correctly classified by the j-th classifier.

TABLE II. COMPLEMENTARITY OF INFORMATION BETWEEN THE FOUR EVALUATED CLASSIFIERS

Classifier →	k-nn	C4.5	SVM	MLP
k-nn	-	47.51	48.75	44.61
C4.5	53.44	-	55.65	53.25
SVM	36.96	38.51	-	17.02
MLP	42.26	45.06	26.67	-

As seen in Table II, significant percentage of the epochs that are misclassified by one sleep stage scorer can be correctly scored by another classifier. This is an indication that a proper fusion of these dissimilar classifiers can improve the overall performance and reduce the number misclassified to wrong sleep stage labels epochs.

Next, we applied the contextual information fusion architecture presented in Section 2. The performance of it was evaluated using the best performing 30 seconds epoch size. Two scenarios were examined, namely the offline and the online. In the offline scenario we applied the contextual fusion scheme using proceeding and succeeding context information, while in the online scenario we assume that no future contextual information is available, thus only previous epoch's information is used.

The experimental results for the offline contextual information fusion setup are tabulated in Table III. The contextual interval is denoted as $\left[-r, r\right]$. The best sleep stage classification performance is indicated in bold.

TABLE III. OFFLINE CONTEXTUAL INFORMATION FUSION SETUP

Classifier	[0]	[-1,+1]	[-2,+2]	[-3,+3]
k-nn	89.38	90.76	91.43	91.10
C4.5	88.21	89.30	90.87	90.12
SVM	92.87	93.56	**94.57**	94.13
MLP	89.45	89.96	90.24	90.10

The experimental results of Table III show a significant improvement of the sleep classification accuracy comparing to the results of the 1-st level (see Table I). The best performing fusion algorithm is the SVM with 94.57%, which corresponds to an accuracy increase of 5.13% in terms of absolute performance. The best performing contextual setup was found for r =2. The use of contextual information allows the sleep stage scorer to correct misclassified epochs which are positioned among correctly classified ones. The use of higher values of r does not improve the classification accuracy, probably due to the fact that changes in more than 60 seconds from the current epoch serve as noise to the classifier rather than improve robustness.

The experimental results for the online contextual information fusion setup are tabulated in Table IV. The best sleep stage classification performance is also indicated in bold.

TABLE IV. ONLINE CONTEXTUAL INFORMATION FUSION SETUP

Classifier	[0]	[-1,0]	[-2,0]	[-3,0]
k-nn	88.61	89.54	90.70	90.41
C4.5	87.72	88.20	89.13	88.92
SVM	90.48	91.06	**92.15**	91.93
MLP	88.97	89.03	89.76	89.70

As shown in Table IV, for the online scenario the lack of succeeding information from the fusion scheme results to slight reduction of the accuracy comparing to the offline case. However, still the performance is significantly higher than the baseline one (see Table I).

V. CONCLUSION

We presented a fusion architecture for the robust classification of sleep stages from EEG signals. The architecture exploits time-domain and frequency domain features and at the fusion step utilizes temporal contextual information. The experimental results showed the validity of the proposed concept, since the accuracy was improved by more than 5% comparing to the baseline sleep stage classification scheme. The use of temporal contextual information reduces the errors caused to isolated misclassification of epochs, probably due to microevents within the duration of the epoch which instantly change the amplitude and spectral characteristics of the EEG signal.

ACKNOWLEDGMENT

The research reported in the present paper was partially supported by the ARMOR Project (FP7-ICT-2011-5.1-287720) “Advanced multi-paRametric Monitoring and analysis for diagnosis and Optimal management of epilepsy and Related brain disorders”, co-funded by the EC under the 7th Framework Programme and the BioMed-Mine Project "Mining Biomedical Data and Images: Development of Algorithms and Applications" funded through the Operational Program "Education and Lifelong Learning" of the NSRF-Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

REFERENCES

[1] H. Hagiwara, "Estimation of sleep stage in the falling asleep period using a Lorenz plot of ECG RR intervals", In Proc. of the 31st Annual International Conference of the IEEE EMBS, 2009, pp. 2510-2513.

[2] 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, 2010, pp. 7922–7928.

[3] K. Polat, S. Yosunkaya, S. Gunes, "Comparison of different classifier algorithms on the automated detection of obstructive sleep apnea syndrome", J. of Medical Systems, vol. 32(3), 2008, pp. 243–250.

[4] A. Rechtschaffen, A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subject", Washington, DC: US Government Printing Office, National Institute of Health Publication, 1968.

[5] I. Zhovna, I.D. Shallom, "Automatic detection and classification of sleep stages by multichannel EEG signal modeling", In Proc. of the 30th IEEE EMBS Conference, 2008.

[6] I. Mporas, P. Korvesis, E.I. Zacharaki, V. Megalooikonomou, "Sleep spindle detection in EEG signals combining HMMs and SVMs", Artificial Intelligence: Methods and Applications Lecture Notes in Computer Science, vol. 8445, 2014, pp. 442-447.

[7] C.-C. Chiu, B. Huy Hai, S.-J. Yeh, "Sleep stages recognition based on combined artificial neural network and fuzzy system using wavelet transform features", In proc. of the 4th Int. Conf. on Biomedical Engineering, IFMBE Proceedings, vol. 40, 2013, pp. 72–76.

[8] 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 Proc. of the 30th IEEE EMBS Conference, 2008.

[9] S. Yu, Xi Chen, B. Wang, X. Wang, "Automatic sleep stage classification based on ECG and EEG features for day time short nap evaluation", In Proc. of the 10th World Congress on Intelligent Control and Automation, 2012.

[10] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, F. Chapotot, "Feature selection for sleep/wake stages classification using data driven methods", Biomedical Signal Processing and Control, vol. 2, 2007, pp. 171–179.

[11] I. Zhovna, I. Shallom, "Multichannel Analysis of EEG Signal Applied to Sleep Stage Classification", Recent Advances in Biomedical Engineering, 2009, ISBN: 978-953-307-004-9.

[12] N. Kerkeni, F. Alexandre, M.H. Bedoui, L. Bougrain, M. Dogui, "Automatic classifcation of sleep stages on a EEG signal by artificial neural networks", In proc. of the 5th WSEAS International Conference on Signal, Speech and Image Processing, WSEAS SSIP'05, 2005.

[13] K.A.I. Aboalayon, H.T. Ocbagabir, M. Faezipour, "Efficient sleep stage classification based on EEG signals", In Proc. of the IEEE Systems, Applications and Technology Conference, LISAT, 2014.

[14] U.R. Acharya, E.C. Chua, K.C. Chua, L.C. Min, T. Tamura, "Analysis and automatic identification of sleep stages using higher order spectra", Int. J. Neural Syst., vol. 20(6), 2010, pp. 509-21.

[15] P.A. Estevez, C.M. Held, C.A. Holzmann, C.A. Perez, J.P. Perez, J. Heiss, M. Garrido, P. Peirano, "Polysomnographic pattern recognition for automated classification of sleep-waking states in infants", Med Biol Eng Comput., vol. 40(1), 2002, pp. 105-13.

[16] A. Noviyanto, A.M. Arymurthy, "Sleep stages classification based on temporal pattern recogniton in neural network approach", In Proc. of IEEE World Congress on Computational Intelligence (WCCI), 2012.

[17] M. Langkvist, L. Karlsson, A. Loutfi, "Sleep stage classification using unsupervised feature learning", Advances in Artificial Neural Systems, Vol. 2012.

[18] B. Kemp, A.H. Zwinderman, B. Tuk, H.A.C. Kamphuisen, J.J.L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG", IEEE Transactions on Biomedical Engineering, vol. 47(9), 2000, pp. 1185-1194.

[19] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals", Circulation, vol. 101(23), 2000 (June 13).

[20] I. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis, M. Richardson, V. Megalooikonomou, "Seizure detection using EEG and ECG signals for computer-based monitoring, analysis and management of epileptic patients", Expert Systems with Applications, vol. 42(6), 2015, pp. 3227–3233.

[21] I.H. Witten, E. Frank, "Data mining: practical machine learning tools and techniques", 2nd ed, Morgan-Kaufman Series of Data Management Systems, San Francisco: Elsevier, 2005.

[22] C. Burges, "A tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, vol. 2(2), 1998, pp. 121-167.