The 7th IEEE International Conference on E-Health and Bioengineering - EHB 2019
Grigore T. Popa University of Medicine and Pharmacy, Iasi, Romania, November 21-23, 2019

Automatic Sleep Stage Detection using a Single Channel Frontal EEG

Alexandra-Maria Tăuţan[1,2], Alessandro C. Rossi[2], Ruben de Francisco[2], Bogdan Ionescu[1]
[1]*Research Center CAMPUS, University Politehnica of Bucharest*
Bucharest, Romania, alexandra.tautan@upb.ro
[2]*Onera Health*
5617 AB Eindhoven, The Netherlands

***Abstract*— Sleep stage detection algorithms can significantly reduce the workload of manual sleep staging and in improving sleep disorder diagnostics. In this paper, we focus on the automatic detection of sleep stages from a frontal channel EEG using expert defined features in both time and frequency domain, fed to a random forest classifier. The proposed approach shows that using a single frontal channel EEG signal as input to automated sleep scoring algorithms is as effective as using EEGs recorded from the central and occipital regions. Mean overall accuracy, precision and recall were respectively of 72.98%, 79.75% and 71.83%, when validating our method on the MGH (Massachusetts General Hospital), *You snooze, you win* dataset.**

Keywords — Sleep Scoring; Single Channel EEG; Random Forests.

I. Introduction

Sleep disorders are diagnosed and characterized clinically through the analysis of polysomnographic (PSG) data. In a PSG study, the physiological parameters of the patient are monitored during sleep. Several biomedical signals are included in typical PSG studies: one or more electroencephalographic (EEG) channels, electrooculograms (EOG), electromyograms (EMG) placed on the submental and tibialis muscles, electrocardiogram leads (ECG), oxygen saturation (SaO2), several signals used for tracking respiration either by measuring the respiratory effort through chest and abdominal belts or by tracking airflow through cannulas or masks.

Normal sleep in adult humans can be characterized by several sleep stages. Anomalies observed during sleep stages can provide clues for the diagnosis and treatment of sleep disorders. The golden standard for determining sleep stages is the manual scoring of PSG recordings by clinicians. This is performed via a tedious visual analysis of the PSG data divided into 30 second epochs according to recognized sleep scoring guidelines. The most recent AASM (American Academy of Sleep Medicine) guidelines propose 4 stages of sleep: NR1 (light sleep), NR2 (non-REM sleep), NR3 (deep sleep), and REM (rapid eye movement) [1].

Automated sleep staging can ease the burden on clinicians and help them provide a faster diagnosis. Several algorithms for automatic sleep staging have already been proposed. An overview of the current state of the art is presented in [2].

Although using multiple EEG channels as input to automatic scoring algorithms has proven to give good results [3], using single channel EEGs has also proven effective. Most algorithms for automatic sleep scoring with single channel EEG as input focus on central or parietal EEG channels as these positions are recommended in the AASM guidelines. Fraiwan et. al [4] use the C3 channel with a Random Forest classifier in a hold-out validation resulting in an accuracy over all classes of 83%. Hassan et. al [6] focus on the Pz EEG channel, also with a random forest classifier with a result of 90.38% accuracy. Most studies use different features extracted from the signals and feed them into a classification algorithm. Other approaches like the one of Tsinalis et. al [5], use raw EEG data with a convolutional neural network resulting in an accuracy of 76%.

The present work focuses on the automatic detection of sleep stages from single channel frontal EEG. Obtaining enough accuracy from automatic sleep staging from a single EEG channel opens the possibility of using reduced electrode montages for PSG studies. This in turn could facilitate the use of wearable EEG headsets and remote/home monitoring of sleep disorders. Frontal EEGs are particularly interesting as data can be collected from areas proven to result in better quality signals with less preparation time and more stable connections between the sensor and the skin. The results are compared with the performance from other two EEG channel locations from the central and occipital areas.

The paper is organized as follows. Section II provides an overview of the algorithm, along with a detailed description of the features extracted. Section III discusses the results obtained on the selected database and the evaluation framework. Section IV and last section analyzes the obtained results in the context of what has previously been proposed in literature and proposes future improvements for the current work.

978-1-7281-2603-6/19/$31.00 ©2019 IEEE

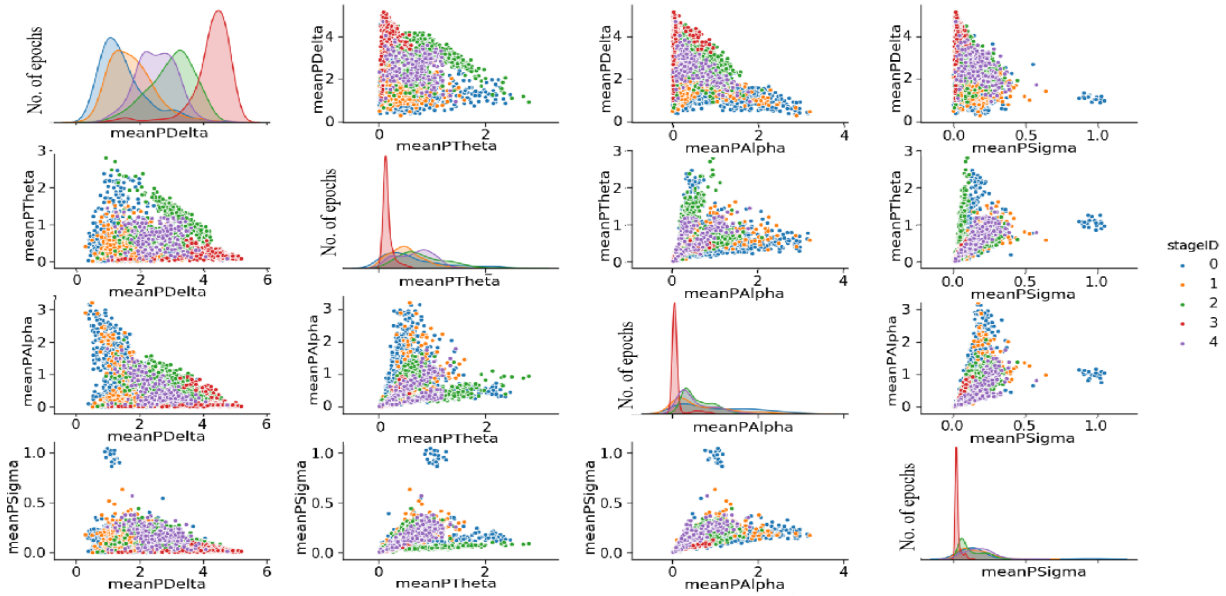

Fig. 1. Mean power spectral density for each EEG band extracted for the annotated epochs of subject tr03-0134; stageID – 0 – Wake, 1 – NR1, 2 – NR2, 3 – NR3, 4 – REM. Each color represents the features extracted from a specific sleep stage. All features are measured in V^2/Hz. The plots on the diagonal represent histograms of each specific feature, while the other plots are pair-wise feature plots.

II. Proposed Approach

A. Overview

The automatic sleep scoring algorithm is based on a supervised learning technique. The input data is represented by 30s epochs from a frontal EEG channel each labeled into one of five classes. The five classes represent the 4 sleep stages (classes 1, 2, 3, 4) defined by the AASM guideline while class 0 represents the wakefulness state. Features are extracted from both time and frequency domains and are fed into a random forest classifier.

B. Feature Extraction

A total of 28 features were extracted from the single channel EEG data. Each feature was extracted from 30 s of input data corresponding to sleep-stage annotated windows. The extracted features are listed in Table I.

Time domain features describe signal amplitude levels and higher order statistics. The frequency domain representation of the signal was obtained using Welch's method on the entire epoch length. Five EEG bands were considered as detailed in Table I. Similarly, to the expert-defined features used in [3], we extracted the mean power spectral density ratios between delta and theta, theta and alpha, as well as delta and alpha.

Figure 1 shows a representation of the mean power spectrum features extracted from each EEG band on one of the subjects from the selected database. The diagonal plots represent histograms for each feature, while the other plots are pair-wise plots between the features. Ideal features would provide a very clear separation between the classes, both in the density distribution plots but also in pair-wise representations. In this case, some of the stages can clearly be differentiated, while for others there is a significant amount of overlap.

TABLE I. List of Extracted Features

Name	**Description**	
Time Domain		
meanA	Mean amplitude in the time domain	
maxA	Max amplitude in the time domain	
skewS	Skew of the signal in time domain	
kurtosisS	Kurtosis of signal in time domain	
stdS	Standard deviation of signal in time domain	
Frequency Domain – EEG bands		
meanP	*Delta – 0.5-4 Hz* *Theta – 4-8 Hz* *Alpha – 8-12 Hz* *Sigma – 12-20 Hz*	Mean spectrum value in EEG band
maxP		Maximum spectrum value in EEG band
minP		Minimum spectrum value in EEG band
stdF		Standard deviation of EEG band spectrum
kurtosisF		Kurtosis of EEG band spectrum
Frequency Domain – EEG Bands Ratio		
delta/theta	Ratio between mean spectrum power in delta and theta band	
theta/alpha	Ratio between mean spectrum power in theta and alpha band	
delta/alpha	Ratio between mean spectrum power in delta and alpha band	

C. Classification

Random Forests (RF) are an ensemble learning method that combine the output of several decision trees [11]. Each node represents a decision on a specific attribute of the data which leads to other nodes. The final nodes are called leaves and they represent the classification labels. In RF, the outputs of several decision trees are merged into a final classification result.

For our experiments, we chose to use a number of decision trees equal to 10. We experimented with the minimum number of samples that are required for a decision to be given by a leaf.

III. Results

A. Dataset

The dataset used for training and testing the algorithm was the MGH dataset used in the "*You Snooze, You Win:* PhysioNet/Computing in Cardiology Challenge from 2018" [9],[10]. The training set contains AASM sleep stage annotations that we used as ground truth for our experiments.

The AASM guidelines have been recently introduced and is aimed to replace the R&K (Rechtschaffen and Kales [7]. In the AASM guideline, two of the deep sleep stages described in R&K are merged. Many studies focused on automatic sleep scoring are developed either using datasets that are manually annotated in the R&K guideline or that convert the annotations to the AASM guideline by merging the two deep sleep stages (stages 3 and 4). The latter approach might lead to the loss of information which might further lead to erroneous classifications. Therefore, we decided to contextualize our results with studies that exclusively use AASM labeling (Table II).

The selected dataset contains PSG data from 994 subjects. Six EEG channels are included: F3M2, F4M1, C3M2, C4M1, O1M2, O2M1. For our study, we have considered data from the F3M2, C3M2 and O1M2 EEG channels. All subjects were included in the analysis.

B. Evaluation methods

The algorithm is validated on the selected dataset by using k-fold cross-validation. In k-fold cross validation, the data is divided into k-folds, out of which k-1 are used for training and the k^{th} fold is used as a test set. The process is repeated until each one of the folds has been used for testing. In our case, we used 10-fold cross-validation.

For multi-class classification problems, the most encompassing metric for algorithm performance is the confusion matrix. The confusion matrix provides an overview of all the true positives (Tp), true negatives (Tn), false positives (Fp) and false negatives (Fn) obtained for all the classes, which in our case are the sleep stages. We use a class normalized confusion matrix as an indication for algorithm performance.

Once the confusion matrix is computed, we extract accuracy, precision and recall values for each class, as defined below:

$$Accuracy = (Tp + Tn) / (Tp + Tn + Fp + Fn) \quad (1),$$

$$Recall = Tp / (Tp + Fn) \quad (2),$$

$$Precision = Tp / (Tp + Fp) \quad (3).$$

C. Results and Discussion

The performance of the proposed framework for automatic sleep scoring while varying the minimum number of samples per leaf (MSL) of the random forest classifier is presented in Figure 2. The best performance was obtained when the random forest classifier has a MSL of 10 samples per leaf.

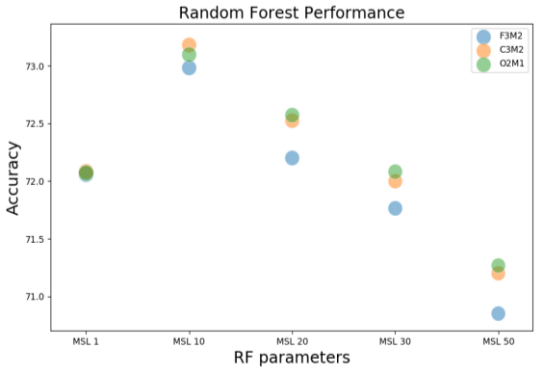

Fig. 2. Accuracy of the automated sleep scoring algorithm when varying the minimum number of samples per leaf (MSL).

When using only one MSL, a slightly lower performance is obtained due to overfitting. When increasing the MSL, the performance has a decreasing trend as the requirement for a minimum number of observations on the terminal node does not allow the decision trees to branch out sufficiently decreasing the prediction performance.

The mean accuracy, precision and recall over all classes in this case was of 72.98%, 79.75% and 71.83% respectively for the frontal channel F3M2. For the EEG channels from the central and occipital regions, a similar performance was obtained as shown in Table II. From these results, it can be concluded that the automatic sleep staging performed with the proposed approach on F3M2 is equivalent in performance when using single EEG channel data from different locations on the scalp (C3M2 and O2M1). The standard deviation of accuracy across the 10 folds validation framework is also reported for each one of the channels (Table II). The variation is equivalent across all three channels and is very low with respect to the mean accuracies.

Table II also provides a comparison between our results and other algorithms available in literature. All studies presented have used datasets annotated according to the AASM guidelines or have converted the R&K annotations to the newer standard. Our results are in line with other experiments on sleep scoring using a single channel EEG as input. Although other approaches have reported slightly better results, the different datasets and validation techniques used should be considered. For instance, Biswal et. al used several algorithms on the entire dataset of 10.000 patients recorded at MGH [3]. Using all 6 EEG channels as input, the performance was approximately 2% higher using expert defined features and a random forest classifier.

Figure 3 shows the normalized confusion matrix for the best performing model using the frontal channel. The wakefulness state was classified with the highest accuracy, while the REM stage with the lowest. A total of approximately $2*10^6$ annotated epochs were considered. The division in the five classes is presented in Figure 4. The dataset considered is not balanced as most epochs belong to the NR2 sleep stage. However, this does not seem to be reflected in the performance of the algorithm: the accuracy for NR2 stage is of 69% while for the wakefulness state, which is represented by a

TABLE II. COMPARISON WITH OTHER APPROACHES IN LITERATURE

Study	Input Channel	Dataset	Classification Algorithm	Validation method	Accuracy [%]
Fraiwan et. al 2012 [4]	C3A1	Private dataset	Random Forest	Hold out	83
Tsinalis et. al 2016 [5]	FpzCz	Sleep-EDF	CNN	20-fold cross-validation	71-76
Mohammadi et. al 2016 [8]	C3A2	Private dataset	SVM	Hold out	83.6
Biswal et. al 2017 [3]	6 EEG channels	MGH extended	Random Forest	Hold out	75.67
Current Study	F3M2	MGH	Random Forest	10-fold cross validation	72.98 ± 3.5
	C3M2				73.18 ± 3.8
	O1M2				73.09 ± 3.9

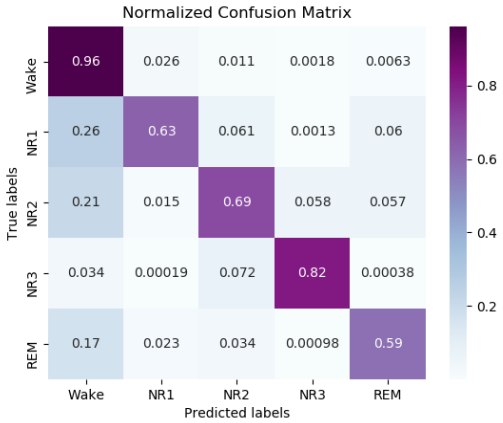

Fig. 3. Normalized confusion matrix for the Random Forest algorithm with 10 minimum samples per leaf resulting in the highest overall accuracy.

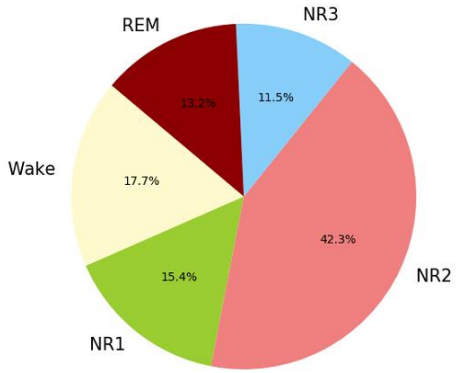

Fig. 4. Percentage of available 30s epochs from each sleep stage.

smaller number of samples, it is of 96%. Therefore, the features extracted might not properly reflect the NR1, NR2 and REM sleep stage which presented lower performance.

IV. CONCLUSIONS

In this paper we investigated automatic sleep stage detection through a Random Forest model that uses features extracted from a single frontal EEG channel. The accuracy over all classes was of 72.98%. The performance is similar to the results obtained from central and occipital electrodes. This shows that using a frontal EEG channel for sleep scoring can be as effective as using single EEG channels from other locations on the scalp. The performance obtained using a frontal EEG channel is comparable to that obtained with input from 6 EEG channels.

Further work for incremental improvements might involve a) the use of multi modal signals from PSG recordings and b) data balancing techniques across the 5 classes. Feature relevance should also be studied to potentially reduce computational power.

REFERENCES

[1] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, R. M. Lloyd, S. F. Quan, M. M. Troester and B. V. Vaughnh, *The AASM Manual for the Scoring of Sleep and Associated Events. Version 2.4*, American Academy of Sleep Medicine, 2017.

[2] L. Fiorillo, A. Puiatti, M. Papandrea, P. L. Ratti, P. Favaro, C. Roth, P. Bargiotas, C. Bassetti and F. D. Faraci, "Automated sleep scoring: A review of the latest approaches [pre-proof]," *Sleep Medicine Reviews*, 2019.

[3] S. Biswal, J. Kulas, H. Sun, B. Goparaju, M. B. Westover, M. T. Bianchi and J. Sun, "SLEEPNET: Automated Sleep Staging System via Deep Learning," *arXiv*, 2017.

[4] L. Faiwan, K. Lweesy, N. Khasawneh, H. Wenz and H. Dickhaus, "Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 10-19, 2012.

[5] O. Tsinalis, P. M. Matthews, Y. Guo and S. Zafeiriou, "Automatic Sleep Stage Scoring with Single-Channel EEG Using Convolutional Neural Networks," *arXiv*, 2016.

[6] A. R. Hassan and M. I. H. Bhuiyan, "A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features," *Journal of Neuroscience Methods*, vol. 271, pp. 107-118, 2016.

[7] A. Rechtschaffen and A. Kales, "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects," US Government Printing Office, Washington DC.

[8] M. Mohammadi, S. Kouchaki, M. Chavami and S. Sanei, "Improving time-frequency domain sleep EEG classification via singular spectrum analysis," *Journal of Neuroscience Methods*, pp. 96-106, 2016.

[9] M. M. Ghassemi, B. E. Moody, L.-w. H. Lehman, C. Song, Q. Li, H. Sun, R. G. Mark, M. B. Westover and G. D. Clifford, "You Snooze, You Win: the PhysioNet/Computing in Cardiology Challenge 2018," *Computing in Cardiology*, vol. 45, pp. 1-4, 2018.

[10] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Invanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng and H. Stanley, "PhysioBank, PhysioToolkit and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. 215-220, 2003.

[11] F. Schroff, A. Criminisi and A. Zisserman, "Object Class Segmentation using Random Forest," in *Proceedings of the British Machine Vision Conference* , Leeds, 2008.