Personalized Sleep Spindle Detection in Whole Night Polysomnography

Personalized Sleep Spindle Detection in Whole Night Polysomnography

Stefano Scafa[1,*], Luigi Fiorillo[2,3], Marta Lucchini[2], Corinne Roth[5], Valentina Agostini[1], Alberto Vancheri[2] and Francesca D. Faraci[2]

***Abstract*— The present study proposes a new personalized sleep spindle detection algorithm, suggesting the importance of an individualized approach. We identify an optimal set of features that characterize the spindle and exploit a support vector machine to distinguish between spindle and non-spindle patterns. The algorithm is assessed on the open source DREAMS database, that contains only selected part of the polysomnography, and on whole night polysomnography recordings from the SPASH database. We show that on the former database the personalization can boost sensitivity, from 84.2% to 89.8%, with a slight increase in specificity, from 97.6% to 98.1%. On a whole night polysomnography instead, the algorithm reaches a sensitivity of 98.6% and a specificity of 98.1%, thanks to the personalization approach. Future work will address the integration of the spindle detection algorithm within a sleep scoring automated procedure.**

I. INTRODUCTION

Sleep spindles are electroencephalography (EEG) oscillations in non-rapid-eye-movement (NREM) sleep stages [1], and are implicated in sleep-related cerebral plasticity [2]. According to the official definition by the American academy of sleep medicine (AASM) [3], a sleep spindle is a train of distinct waves with frequency in the range of 11 to 16 Hz and with duration $\geq$ of 0.5 s, usually maximal in amplitude using central EEG derivations. Spindles are a distinctive pattern of sleep stage NREM2, and can also be present in NREM3; their identification assists the sleep scoring procedure. Many spindle detection algorithms have been developed in the last decades [4], and present very good performance. However, when applied to a different database the performance often deteriorates, as highlighted in [5], where six different automated detectors are compared.
Presently, an automated sleep spindle detector flexible enough to be routinely used in the everyday practice and adaptable to different database is not yet available.

In general, the automated spindle detection procedure consists of several phases: pre-processing, feature extraction, feature selection and spindle recognition with a classifier. We refer the reader to [6] for an extensive analysis of the various methodologies. Very recently, also deep learning [7] has been successfully tested in sleep spindle identification. Spindle density and characteristics such as mean oscillation frequency, amplitude and duration appear stable over time in the same night for the same subject but may vary considerably between subjects [8]. Therefore, an individualized approach has been already proposed by some authors. In [9] the specific mean peak spindle amplitude is calculated so as to improve the detection. In [10] the EEG is bandpass-filtered according to individually adjusted frequency criteria, then spindles are detected if the rectified signals exceed the amplitude criteria for more than 0.5 s. As highlighted in [11], it is quite difficult to compare different results, because of different metrics, different databases and different criteria used for sleep spindles identification across studies. In general, recent efforts claim to reach a sensitivity and a specificity both up to 98% [9].

In the present work we investigated how a spindle detection algorithm can be improved with a more flexible technique and individualized information. In our approach all the features are tailored to the specific set of data, and we let a classifier deciding how to better differentiate spindles from non-spindles. The advantage is that the adaptability of the detection algorithm is enhanced, without a priori constraints or threshold. We also examined how the algorithm performance changes in a whole night polysomnography (PSG). In the vast majority of the publications the detection algorithms are applied on a recording belonging only to NREM2 stage, and not to the entire PSG. A better characterization of the false positives and false negatives could improve many existing detection algorithms. Moreover, a good spindle detector could be integrated and could support a sleep scoring automated algorithm, enhancing its performance.

II. DATA

In order to be comparable with existing work, we tested our detection algorithm on the commonly used DREAMS database [11]. The database contains recordings of eight patients, with spindles scored by two experts. It consists of 30 minutes of EEG (channel C3-A1 or Cz-A1) extracted from whole-night PSG recordings. The data were acquired in a sleep laboratory of a Belgian hospital using a digital 32-channel polygraph. They consist of eight polysomnographic recordings belonging to patients with different pathologies. The second expert only scored the first six recordings, whilst the first expert only scored the first part of the 30-minute recordings. Therefore only the first six subjects were used, and the union of the two scoring was considered, as done in [11]. We re-sampled all the recordings so as to consider a sampling rate of 200 Hz.

*This work was partially co-funded by the Swiss State Secretariat for Education, Research and Innovation and the European Union Organization

*Corresponding author: scafa.stefano@gmail.com

[1]Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy

[2]Institute for Information Systems and Networking, University of Applied Sciences and Arts of Southern Switzerland, Manno, Switzerland

[3]Institute of Informatics, University of Bern, Bern, Switzerland

[4]Neurology Department, Inselspital, Bern University Hospital, Bern, Switzerland

TABLE I
DREAM DATABASE AND SPASH DATABASE.

	Recording	Sampling Frequency	Scored by expert 1	Scored by expert 2	Spindles Union
DREAMS	subject(1)	100 Hz	52	115	134
	subject(2)	100 Hz	50	62	76
	subject(3)	50 Hz	5	44	44
	subject(4)	200 Hz	44	25	63
	subject(5)	200 Hz	56	86	103
	subject(6)	200 Hz	72	97	117
	subject(7)	200 Hz	18	–	–
	subject(8)	200 Hz	48	–	–
SPASH	subject(1)	200 Hz	487	–	–

We also investigated the goodness of our spindle detector on a whole night PSG recording (channel C4-M1), extracted from the SPASH database. SPASH database belongs to the Sleep-Wake-Epilepsy-Centre (SWEZ) of the InselSpital of Bern (Switzerland) and contains 60 whole night PSG recordings of healthy subjects, sampled at 200 Hz. Presently only one PSG has been fully scored by the sleep experts, including spindles, K-complexes and vertex waves.

Further details on the two databases are reported in Table I.

III. METHODS

Our first challenge was to identify the optimal set of features to characterize the spindle. From the literature it appears clear that many signal characteristics have to be taken into account to gather complementary information. Then, the features were employed to train a support vector machine (SVM) classifier to distinguish a spindle from any other events in the EEG recording. We propose a personalized sleep spindle detection (PSSD) procedure where the features are tightly adapted to the subject sleep spindle characteristics.

A. Feature Extraction and Feature Selection

We identified 14 promising features, with complementary information about the spindle physiology. The 14 features were the following: sample entropy (SpEn), maximum, minimum, variance, standard deviation, phase amplitude coupling (PAC), instantaneous frequency, energy ratio ($\text{Energy}_{11\text{-}16\text{Hz}}$), kurtosis, skewness, power peak (PWR_{peak}), power ratio ($\text{PWR}_{\text{ratio}}$), interquartile range (IQR) and zero crossing.

The EEG signal is filtered with a FIR filter between 0.3 Hz and 35 Hz. All the features were computed starting from the filtered signal, and most of them within a time window of 0.5 seconds (i.e, the minimum duration of a spindle [3]). The window was centered on the pattern detected by the sleep physician. Sample entropy, maximum, minimum, variance, standard deviation, phase amplitude coupling, instantaneous frequency and energy ratio were directly computed on the filtered signal. In particular, instantaneous frequency and energy ratio were obtained using Hilbert-Huang transform. Kurtosis, skewness, power peak and power ratio values were computed on the power spectral density (PSD) of the signal. Interquartile range and zero crossing were calculated on the first mode of the empirical mode decomposition (EMD) of the filtered signal.

The feature selection procedure has been done using the minimum redundancy maximum relevance (MRMR) algorithm [12]. The following features have been finally selected.

SpEn. The sample entropy was computed as in [13]. The parameter was calculated on a time series of 50 samples, with a tolerance of 5% of the standard deviation of the signal. The sample entropy detects the complexity of a signal. In fact, unlike the other features, is not limited to extract information from the 0.5 second window used for spindle detection but it is used to calculate a value relative to the adjacent signal. This allows to characterize the environment in which a spindle occurs. Since the spindle usually occurs during the NREM2 phase, it is expected to find low frequency fluctuations in its surrounding, where delta (0.3-3 Hz) and theta (4-8 Hz) activities are prevalent. A very different background from the ones that can be found in a awake or REM sleep stages.

PAC. Phase amplitude coupling evaluates the degree of spindle amplitude modulation by low frequency activity. Low frequency waves and sleep spindles are two predominant features of NREM sleep. Sleep spindles are short trains of waves that often have a waxing-and-waning shape, a frequency activity of 11-16 Hz modulated by a low frequency (0.1-1.5 Hz)[14]. The feature was computed by expanding the observation time window from 0.5 seconds to 4 seconds to gather the dynamics of the low-frequency components. Hence, the PAC allows to detect a synchronization between the phase of low frequency components and spindle amplitude.

$\text{Energy}_{11\text{-}16\text{Hz}}$. The feature is the ratio between the energy of the signal in band 11-16 Hz and the energy of the signal in the whole band. In presence of a spindle, a high value of this parameter is expected.

PWR_{peak}. The power peak value was computed on the PSD of the signal; it was the maximum power value identified in the sleep spindle frequency band (11-14 Hz). This feature can better differentiate sleep spindles from other EEG oscillations in the same frequency band.

$\text{PWR}_{\text{ratio}}$: The parameter was computed on the PSD of the signal; it is defined as the ratio between the total power of the spindle spectral band and the total power of the low frequency band (0.3-8 Hz).

IQR. The interquartile amplitude range was calculated on the first mode of the EMD. Through the decomposition of the signal into monocomponents, it is possible to isolate the higher frequency components in the first mode and to have an estimate of the amplitude. The hypothesis is that the activity of the spindle, characterizing an NREM2 stage, occurs at a higher frequency than the signal in the background.

B. Classification Procedure

The SVM classifier was trained on a dataset composed by 30 samples from the same PSG: 15 spindles, randomly

selected among the one identified by the scorer, and 15 non-spindles. The selected features were extracted from these samples, in particular they were extracted from time windows of 0.5 seconds. Then the trained classifier was tested on the whole recording. The test set consists of sliding windows of 0.5 seconds, shifted on the whole signal with steps of 0.1 second. The classifier predicts the presence or non-presence of a spindle in each window.
After the classification procedure, a post-processing phase is needed. With the windowing procedure a single spindle is detected by consecutive overlapping windows. Therefore it is necessary to aggregate them in just one window enclosing the whole spindle pattern. On the other hand, when a spindle is detected only by one window, it is almost certainly a non-spindle (false positive).

The SVM classifier was tested also with a non-personalized procedure. The 30 samples of the training set, spindles and non-spindles, are randomly selected among all the different subjects. SPASH data were used to train the SVM used for DREAMS patients and vice versa.

C. Performance Metrics

In order to compare the performance with existing detection algorithm, the sensitivity, the specificity and the false positive rate (FPR) were computed.

$$Sensitivity = \frac{TruePositive}{(TruePositive + FalseNegative)} \quad (1)$$

$$Specificity = \frac{TrueNegative}{(TrueNegative + FalsePositive)} \quad (2)$$

$$FPR = \frac{FalsePositive}{(FalsePositive + TrueNegative)} \quad (3)$$

IV. RESULTS AND DISCUSSION

We firstly tested our algorithm on the open source DREAMS database in a personalized and non-personalized approach. The results are in Table II. Our approach reaches a sensitivity of 89.8%, definitely higher than the one reported in the state of the art [11], despite a 1.9% FPR. In order to evaluate the enhancement that the personalization can have, we tested our algorithm on the DREAMS database also in a non-personalized approach. We found as expected a significantly low sensitivity, down to an average of 84.2% ($p<0.05$), with higher variability. Then we tested our algorithm on a whole night PSG recording from the SPASH database, of around seven hour length. We repeated the test several times, extracting every time randomly a different training set of 15 spindles. We obtained very good averaged results, with low variability, over the whole duration of the EEG signal: 93.0% sensitivity, 98.1% specificity and 1.9% false positive rate. Also for SPASH database the non-personalized approach was tested, using

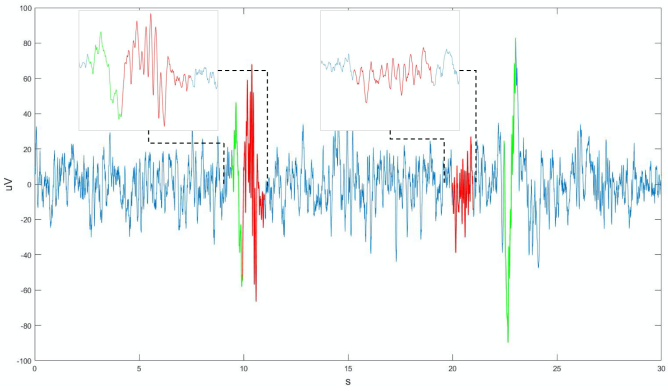

Fig. 1. An EEG epoch of 30 seconds with highlighted in green the K-Complexes and in red the spindles.

data from the DREAMS database. Sensitivity and specificity values decreased to 92.5% and 97.7% respectively ($p<0.05$). The results present a certain variability, implicit in the way that the algorithm is built. The classifier performance depends strongly on the training set: the authenticity of the spindles and the precision with which they are selected within the window are both very important. If the training set is selected with the due accuracy the on a whole night polysomnography the algorithm can reach a sensitivity of 98.6% and a specificity of 98.1%. Spindle characteristics may vary slightly during the night within the same subject. Often the spindles may be overlapped with other events and this can introduce biases in the feature extraction process. Fig. 1 shows example of two spindles in a 30-second epoch, the first spindle is overlapped with a K-complex. Although there is a certain intrinsic variability in the performance of the PSSD algorithm, the results are very satisfactory.

From the distribution of false positives in Table III emerged that about 30% of these were in sleep stages where sleep spindles should not be present. Therefore, we constructed a mask to remove W, NREM1 and REM epochs, without removing epochs close to a transition. In fact, the scorer has annotated spindles in epochs close to transitions that are not labeled neither as NREM2 or as NREM3. After applying this mask, the increase in specificity was minimal. Approximately 95% of false positives generated in the awake and REM stages are placed in the transitional periods and therefore wre not removed with this mask. Another observation can be made on the 70% of false positives belonging to the two stages NREM2 and NREM3: it is possible that some activities, maybe very short but recorded at least in two adjacent windows (0.6 seconds of activity), were actually sleep spindles activities not marked by the scoring sleep expert. The results are reported as PSSD* in Table II.

V. CONCLUSIONS

In the present work we introduced the PSSD algorithm, that with six well defined features can extract complementary information about the spindles of a single subject. Our algorithm presents a better performance than the state-of-the-art, and highlights the improvement that is possible to

TABLE II

PERFORMANCE OF THE PSSD CLASSIFIER WITH A PERSONALIZED AND NON-PERSONALIZED APPROACH ON THE DREAMS AND SPASH DATABASE. PERFORMANCE OBTAINED FROM THE RANDOM SELECTION OF 500 TRAINING SETS (AVERAGE VALUE) $\pm$ (STANDARD DEVIATION).

DATABASE	ALGORITHM	PERSONALIZED	SENS.	SPEC.	FPR
DREAMS	[11]	NO	0.702	0.986	0.014
	PSSD	NO	0.842 ± 0.104	0.976 ± 0.013	0.024 ± 0.013
		YES	$\mathbf{0.898 \pm 0.067}$	0.981 ± 0.008	0.019 ± 0.008
SPASH	PSSD	NO	0.925 ± 0.055	0.977 ± 0.012	0.023 ± 0.012
		YES	$\mathbf{0.930 \pm 0.056}$	0.981 ± 0.008	0.019 ± 0.008
	PSSD*	YES	0.930 ± 0.056	$\mathbf{0.986 \pm 0.005}$	$\mathbf{0.014 \pm 0.005}$

TABLE III

DISTRIBUTION OF FALSE POSITIVES DETECTED ON SPASH DATABASE.

Database	W	NREM1	NREM2	NREM3	REM
SPASH	12%	11%	46%	23%	8%

reach through personalization. We have noticed a very strong dependency on the quality of the data. Our approach can be successfully utilized only if the few spindles used for the training set are accurately identified.

In future the PSSD algorithm could be integrated in a scoring automated procedure: the human scorer would have to well determine a set of spindles manually for each subject, let the algorithm find all the others and use this information to improve the scoring.

REFERENCES

[1] De Gennaro, Luigi, and Michele Ferrara. "Sleep spindles: an overview." (2003): 423-440.

[2] Puentes-Mestril, Carlos, et al. "How rhythms of the sleeping brain tune memory and synaptic plasticity." Sleep 42.7 (2019): zsz095.

[3] Iber, Conrad, and Conrad Iber. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Vol. 1. Westchester, IL: American Academy of Sleep Medicine, 2007.

[4] Bódizs, R., Körmendi, J., Rigó, P., and Lázár, A. S. (2009). The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. Journal of neuroscience methods, 178(1), 205-213.

[5] Warby, Simon C., et al. "Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods." Nature methods 11.4 (2014): 385.

[6] Motamedi-Fakhr, Shayan, et al. "Signal processing techniques applied to human sleep EEG signals—A review." Biomedical Signal Processing and Control 10 (2014): 21-33.

[7] Kulkarni, Prathamesh M., et al. "A deep learning approach for real-time detection of sleep spindles." Journal of neural engineering 16.3 (2019): 036004.

[8] Esther, et al. "Spindle frequency activity in the sleep EEG: individual differences and topographical distribution." Electroencephalography and clinical neurophysiology 103.5 (1997): 535-542.

[9] Ray, Laura B., et al. "Validating an automated sleep spindle detection algorithm using an individualized approach." Journal of sleep research 19.2 (2010): 374-378.

[10] Bódizs, Róbert et al. "The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm." Journal of neuroscience methods 178.1 (2009): 205-213.

[11] Devuyst, Stéphanie, et al. "Automatic sleep spindles detection—overview and development of a standard proposal assessment method." 2011 Annual international conference of the IEEE engineering in medicine and biology society. IEEE, 2011.

[12] Ding, C., and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. Journal of bioinformatics and computational biology, 3(02), 185-205.

[13] Song, Y., Crowcroft, J., and Zhang, J. (2012). Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. Journal of neuroscience methods, 210(2), 132-146.

[14] Li, D., Ni, M., and Dun, S. (2015, October). Phase-amplitude coupling in human scalp EEG during NREM sleep. In 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI) (pp. 219-223). IEEE.