Automatic arousal detection using heart rate from a single-lead electrocardiogram

Franz Ehrlich, Johannes Bender, Hagen Malberg, Miriam Goldammer

Institute of Biomedical Engineering, TU Dresden, Dresden, Germany

Abstract

Arousals during sleep give deep insights into the pathophysiology of sleep disorders and sleep quality. Detecting arousals is a time-consuming process manually performed by a trained expert. The required measurement is performed on an inpatient basis and is uncomfortable for the patient. As arousals relate to the autonomic nervous system, they also reflect in the electrocardiogram, which is therefore a promising alternative biosignal. In this study, we developed a deep learning model for automatic detection of sleep arousals from heart rate.

We developed our algorithm using 5323 recordings from the Sleep Heart Health Study. 1003 of them were held-out as test data. We derived RR intervals from the ECG and interpolated them into a 4 Hz signal. Next, we developed a convolutional neural network (CNN) for end-to-end event detection. Model output is a continuous arousal probability with a frequency of 1 Hz.

The optimization resulted in a twelve-layer CNN that achieved a Cohen's kappa of 0.47, an area under the precision-recall curve of 0.54 on hold-out test data.

This study demonstrates the ability of machine learning to detect arousals during sleep from heart rate. As our approach uses only the heart rate, it is potentially transferable to other signals, e.g. the photoplethysmogram.

1. Introduction

Good quality and quantity of sleep have a great impact on health and overall quality of life [1]. Arousals during sleep provide deep insights into the pathophysiology of sleep disorders and sleep quality [2]. According to the American Academy of Sleep Medicine (AASM) guidelines, an arousal is an abrupt shift in the frequency of the electroencephalogram (EEG) that lasts at least three seconds and requires ten seconds of preceding sleep [3]. During rapid eye movement (REM) sleep, scoring also requires a simultaneous increase in the submental electromyogram (EMG) [3]. Arousals are spontaneous or may occur in response to sleep-disturbing events such as apneas, hypopneas, respiratory effort-related arousals (RERAs), and periodic leg movements. These transient waking reactions lead to a more fractured sleep [4].

Scoring of arousals is part of the standard procedure for diagnosing sleep-related diseases. The medical gold standard of arousal detection is a time-consuming process performed manually by a trained expert. Arousals are visually detected in the EEG and EMG, which are part of polysomnography (PSG). A scored arousal can be seen in figure 1. Measuring the PSG requires an inpatient setup and is uncomfortable for the patient. Since arousals are coupled with the autonomic nervous system, they are also reflected in the electrocardiogram (ECG) and heart rate (HR) in general, which therefore represents a promising alternative biosignal for the detection of arousals.

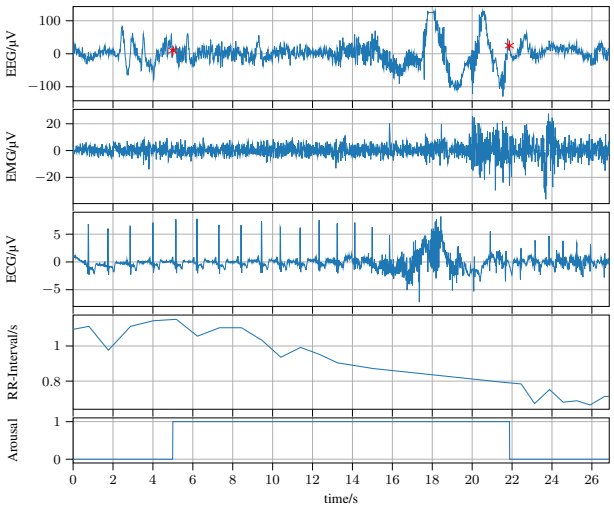

Figure 1: A manually scored cortical arousal with some relevant biosignals from the PSG. The start and end of arousal are indicated by the red stars in the EEG. ECG and RR intervals demonstrate the effects of arousal on these signals.

Recently, convolutional neural networks (CNNs) have shown promising results in processing raw biosignals [5]. They use kernels to store local features in the signal with-

out requiring manual feature extraction. In our previous work, RR intervals (RRI) were successfully used to classify sleep stages [6]. In this study, we propose a deep CNN to detect arousals with a resolution of one second using RRI.

2. Methods

The underlying data were collected as part of the Sleep Heart Health Study (SHHS) [11, 12]. SHHS is a cohort study investigating sleep-disordered breathing as a risk factor for the development of cardiovascular disease. The study involved 6600 adults aged forty years and older. The first part of the study yielded 5804 polysomnograms (PSG), which are publicly available and used for the following research. A detailed breakdown of patients can be seen in Table 2. Certified raters manually scored the sleep stages and sleep events. The arousal annotations were scored using the ASDA [4] criteria. The large number of records and the quality assurance of the study provide a solid foundation for training a machine learning algorithm.

The included ECG was digitized with a sampling frequency of 125 Hz. We use the heart rate variability based on beat-to-beat (RR) interval times, which we have adopted from previous work [6]. The R-peaks are extracted from the ECG using a QRS detector [13]. Records are discarded if the arousal annotation is missing or more than 25 % of the initially found RR intervals were removed by the adaptive filter [6]. This results in 5323 nightly records, which we split into a training set and a test set, with the test set containing 1003 records.

We interpolated the RR intervals into a signal with a sampling rate of 4 Hz. The different types of arousals are merged for binary classification. To capture autonomic features associated with arousals, the annotation is extended. We added two seconds before the start of annotation and ten seconds after the end based on [8]. The same procedure was used in the PhysioNet/Computing in Cardiology Challenge 2018 [14], so the results are comparable to the algorithms that participated in the challenge.

An extensive grid search led to a twelve-layer CNN model architecture. A convolution layer is followed by a BatchNormalization layer and a ReLU activation. The first two convolutional layers reduce the input length by using a stride of two and expand the dimension to a total of sixteen. We use a kernel size of 128 in the first layer and 64 in the other layers. The last ten convolutional layers have an identical structure with 32 filters and a kernel size of 64. The last layer is a fully connected layer to generate the arousal prediction. A sigmoid activation is used to produce outputs ranging from zero to one, representing the arousal probability. The overall model architecture with optimized parameters is shown in Table 3.

The training data were divided into five folds of equal size and trained separately. After ten epochs of training resulted in no improvement in the area under the precision-recall curve (AUPRC), training was discontinued. Each fold produces a trained model that calculates the arousal probability. The final model uses all five models by averaging the output of each model. To calculate the confusion matrix, a threshold must be defined. Based on the training data and maximization of Cohen's kappa coefficient (κ) a threshold of 0.30 was determined as the operating point. If the model probability was above the threshold, the segment was marked as arousal.

3. Results

The model achieved an AUPRC of 0.54 and an area under the receiver operating characteristic (AUROC) of 0.86 on hold-out test data. The prediction of arousals on the test data yields the confusion matrix presented in Table 4. This results in a Cohen's kappa coefficient of 0.47, a precision of 0.59, a recall of 0.46, and an F1 score of 0.53.

There are few machine learning approaches for detecting arousal based on cardiac features [15]. In 2007, Basner

Table 1: Comparison of approaches using ECG features or the full PSG with our approach. The approaches are listed with different types of validation metrics and the size of the test set (n).

	Models	Year	Type	Signals	Test Set	n	$AUPRC$	$AUROC$	$F1$
ECG	Basner et al. [7]	2007	handcrafted	HRV-Features	own dataset	10	-	0.91	-
	Olsen et al. [8]	2018	FFNN	Features	WSC	77	-	-	**0.67**
	Li et al. [9]	2020	CNN-LSTM	ECG	MESA	311	**0.62**	**0.93**	**0.67**
					SHHS-2	785	0.54	0.91	0.69
	proposed model	2022	CNN	RRI	SHHS-1	1003	0.54	0.87	0.53
PSG	DeepSleep [10]	2021	U-Net	PSG	SHHS1	250	0.63		
		2021		PSG	SHHS2	250	**0.70**		

Table 2: Statistical information on the first part of the SHHS, divided into all records and records in our test set.

		Count	Mean	STD	Min	Max
All Data	Records	5804				
	Male	2765				
	Female	3039				
	Age		63.13	11.22	39	90
	Hypertension	2478				
	Pacemaker	57				
	BMI		28.16	5.09	18	50
	ESS		7.77	4.40	0	24
	AHI		12.13	11.74	0	111.38
	ArI		19.16	10.66	0	110.39
Test set	Records	1003				
	Male	473				
	Female	530				
	Age		62.35	10.79	39	90
	Hypertension	401				
	Pacemaker	10				
	BMI		28.26	5.28	18	50
	ESS		7.81	4.20	0	24
	AHI		12.07	11.67	0	107.88
	ArI		19.27	10.00	0	85.24

STD: Standard deviation, BMI: Body mass index in kg/m, ESS: Epworth Sleepiness Scale AHI: Apnea-Hypopnea-Index in count/hour, ArI: Arousal-Index in count/hour

Table 3: The final twelve layer CNN architecture as a result of the grid search.

	Layer-Type	Output-Shape	Filter	Stride	Kernel
	InputLayer	28800×8			
	Conv1d	14400×8	8	2	128
	BatchNormalization	14400×8			
	ReLU-Activation	14400×8			
	Conv1d	7200×16	16	2	64
	BatchNormalization	7200×16			
	ReLU-Activation	7200×16			
$\times 10$	Conv1d	7200×32	32	1	64
	BatchNormalization	7200×32			
	ReLU-Activation	7200×32			
	Dense	7200×1			
	Sigmoid-Activation	7200×1			

et al. [7] conducted a study of 56 healthy patients from a study examining the effects of aircraft noise on sleep. They used statistical methods to create an algorithm. In 2019, Olsen et al. [8] created an algorithm using 258 patients from the Wisconsin Sleep Cohort. They trained a feed-forward network with hand-crafted features that also included manually scored sleep stages. In 2020, Li et al. [9] developed *DeepCAD*, a machine learning algorithm using CNN and *long short term memory* (LSTM). They used the raw ECG signal to train their algorithm on 1547 patients from the Multi-Ethnic Study of Atherosclerosis. In addition, the network was also trained on the 2nd part of the *SHHS* data.

Due to the recent Physionet Challenge 2018 [14], there are well-comparable approaches to arousal detection using the full PSG. The *DeepSleepNet* [10] achieved highest results on the challenge data. Since they also trained and tested their model on the *SHHS* data, their approach is sufficient to put our approach into perspective. They used 250 subjects to validate their algorithm.

The results of the different approaches are shown in table 1. By only using HRV, we obtained similar results on a comparable dataset from the leading approach *DeepCad* [9]. The approaches using only the ECG or features from the ECG have an *AUPRC* that is approximately 0.14 lower than the state-of-the-art arousal detection using all signals from the PSG.

Table 4: Confusion matrix for the test set with a total of 25,524,000 one-second segments

		Actual	
		Non Arousal	Arousal
Predicted	Non Arousal	85.15%	4.58%
	Arousal	4.96%	5.32%

4. Conclusions

We developed a machine learning algorithm that predicts arousals on a second-by-second basis. By using only heart rate variability, we were able to minimize the information. The SHHS dataset provides enough data to develop a state-of-the-art algorithm. Stacking CNN layers seems to be an effective method to find a suitable architecture.

Research is still needed to develop robust and accurate arousal detectors. The sum of false positives and false negatives still exceeds the number of true positives. This may be due to the reliability of the arousal labels, which is only "modest" even in the SHHS data [16]. Automatic algorithms have the advantage of performing scoring quickly, cost-effectively and consistently.

Our approach requires only RRI, which can be derived from the ECG or possibly from other signals, such as

photoplethysmogram (PPG) by similar P-wave intervals. Since RRI does not change rapidly, the signal does not need to be high-resolution, which reduces the complexity of machine learning and the resulting algorithm. Due to these advantages, an application in the field of mobile and wearable devices is conceivable. On the other hand, using RRI requires an additional step of preprocessing and a good QRS detector compared to using the raw ECG.

Although the SHHS dataset includes over 5800 participants, they are all over 40 years old. Therefore, no conclusions can be drawn about how the algorithm performs for younger people. The approach should also be tested on other datasets, including the second part of the SHHS, to allow better comparison.

Acknowledgements

The Sleep Heart Health Study (SHHS) was supported by National Heart, Lung, and Blood Institute. The National Sleep Research Resource was supported by the National Heart, Lung, and Blood Institute.

This research was partly funded by the European Regional Development Fund with the project 100346021 Tele-Schlaf-Medizin.

References

[1] Mukherjee S, Patel SR, Kales SN, Ayas NT, Strohl KP, Gozal D, Malhotra A. An official american thoracic society statement: the importance of healthy sleep. recommendations and future priorities. American journal of respiratory and critical care medicine 2015;191(12):1450–1458.

[2] Halász P, Terzano M, Parrino L, Bódizs R. The nature of arousal in sleep. Journal of sleep research 2004;13(1):1–23.

[3] Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus C, Vaughn BV, et al. The aasm manual for the scoring of sleep and associated events. Rules Terminology and Technical Specifications Darien Illinois American Academy of Sleep Medicine 2012;176:2012.

[4] Scoring E. Eeg arousals: scoring rules and examples: a preliminary report from the sleep disorders atlas task force of the american sleep disorders association. Sleep 1992; 15(2):174–184.

[5] Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 1d convolutional neural networks and applications: A survey. Mechanical systems and signal processing 2021; 151:107398.

[6] Goldammer M, Weber L, Malberg H, Zaunseder S. Individualized sleep stage classification from cardiorespiratory features. In 2019 13th International Symposium on Medical Information and Communication Technology (ISMICT). IEEE, 2019; 1–6.

[7] Basner M, Griefahn B, Müller U, Plath G, Samel A. An ecg-based algorithm for the automatic identification of autonomic activations associated with cortical arousal. Sleep 2007;30(10):1349–1361.

[8] Olsen M, Schneider LD, Cheung J, Peppard PE, Jennum PJ, Mignot E, Sorensen HBD. Automatic, electrocardiographic-based detection of autonomic arousals and their association with cortical arousals, leg movements, and respiratory events in sleep. Sleep 2018;41(3):zsy006.

[9] Li A, Chen S, Quan SF, Powers LS, Roveda JM. A deep learning-based algorithm for detection of cortical arousal during sleep. Sleep 2020;43(12):zsaa120.

[10] Li H, Guan Y. Deepsleep convolutional neural network allows accurate and fast detection of sleep arousal. Communications biology 2021;4(1):1–11.

[11] Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O'Connor GT, Rapoport DM, Redline S, Robbins J, Samet JM, et al. The sleep heart health study: design, rationale, and methods. Sleep 1997;20(12):1077–1085.

[12] Zhang GQ, Cui L, Mueller R, Tao S, Kim M, Rueschman M, Mariani S, Mobley D, Redline S. The national sleep research resource: towards a sleep data commons. Journal of the American Medical Informatics Association 2018; 25(10):1351–1358.

[13] Vidaurre C, Sander TH, Schlögl A. Biosig: the free and open source software library for biomedical signal processing. Computational intelligence and neuroscience 2011; 2011.

[14] Ghassemi MM, Moody BE, Lehman LWH, Song C, Li Q, Sun H, Mark RG, Westover MB, Clifford GD. You snooze, you win: the physionet/computing in cardiology challenge 2018. In 2018 Computing in Cardiology Conference (CinC), volume 45. IEEE, 2018; 1–4.

[15] Qian X, Qiu Y, He Q, Lu Y, Lin H, Xu F, Zhu F, Liu Z, Li X, Cao Y, et al. A review of methods for sleep arousal detection using polysomnographic signals. Brain Sciences 2021;11(10):1274.

[16] Whitney CW, Gottlieb DJ, Redline S, Norman RG, Dodge RR, Shahar E, Surovec S, Nieto FJ. Reliability of scoring respiratory disturbance indices and sleep staging. Sleep 1998;21(7):749–757.