Comparison of Signal Combinations for Cardiorespiratory Sleep Staging

Miriam Goldammer[1], Sebastian Zaunseder[2], Franz Ehrlich[1], Hagen Malberg[1]

[1] Institute of Biomedical Engineering, TU Dresden, Dresden, Germany
[2] Department of Information Technology, FH Dortmund, Dortmund, Germany

Abstract

This work investigates the benefit of using multiple signals and preprocessing strategies for sleep staging from cardiorespiratory signals.

We modified our previous Neural Network model to take different signal combinations as input. To that end, we added oxygen saturation and different respiratory signals to the electrocardiogram. We further invoked different preprocessing strategies that have been described previously for such signals, namely using downsampled signals vs. using time series of breath-to-breath intervals. We trained and tested our model variations with 4784 polysomnograms from the Sleep Heart Health Study.

We found the best combination of signals to be heart rate together with a downsampled respiratory signal. The classification resulted in a κ of 0.68 on hold-out test data, which outperforms our previous results and state of the art for cardiorespiratory sleep staging.

We observe that combinations of cardiorespiratory signals can improve classification performance for automatic cardiorespiratory sleep staging. As there are generally more cardiorespiratory signals available and many more options for preprocessing them, we expect that further research in this area will show even more improvements.

1. Introduction

Sleep staging from cardiorespiratory signals has improved significantly during the last two decades. Increasingly, feature-based classifiers are replaced by approaches that use signals or time series as inputs to Neural Networks. So far, usage of one-dimensional inputs from single signals (primarily electrocardiogram (ECG)) has been the focus of investigations, even though we know from feature-based approaches that information from respiration and cardiorespiratory coupling can result in significant classification improvements. Also, implementations using the photoplethysmogram (PPG) at 64 Hz [1] and respiratory inductance plethysmography (RIP) at 10 Hz [2] for sleep staging, have shown that Neural Networks can learn relevant sleep features from signals that contain heart beats and breaths as raw waves.

Currently, many approaches based on the ECG apply QRS detection first, to gain a binary respresentation (e. g. [2]) or the instantaneous heart rate (e. g. [3]). When using the PPG or RIP, downsampling is often used as preprocessing (e. g. [1, 2, 4]). Very few works, one exception is e. g. [2], use more than one input signal.

Concerning the effort of applying QRS detection, our hypothesis is that a lowpass filtered version of the ECG together with respiration contains enough information for sleep scoring. For more simplification of the cardiac input signal, Casal et al. [5] showed that even the mere heart rate (HR) from the pulse oximeter (PO) contains enough information for sleep-wake distinction. This would simplify preprocessing for cardiorespiratory sleep stage classification from QRS detection plus feature generation to mere lowpass filtering or resampling.

This contribution investigates signal combinations with a focus on simple preprocessing variants for cardiorespiratory sleep staging. We will compare (i) generating R-R-interval time series and breath-to-breath-interval time series from ECG and RIP (our previous approach [6]), (ii) lowpass filtering ECG and RIP with cutoff frequency 2 Hz, and (iii) resampling HR and oxygen saturation (SpO_2) from the PO.

2. Methods

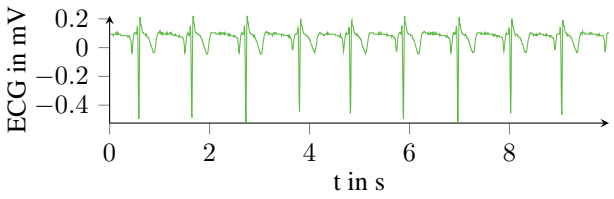
2.1. Data

For our investigations, we used data from the first part of the Sleep Heart Health Study [7–9]. From these 5804 polysomnograms (PSGs) of different study participants, we excluded 1012 PSGs due to low signal quality of the ECG. The remaining PSGs were split patient-wise into 916 PSGs as hold-out test data and 3867 PSGs for training and validation.

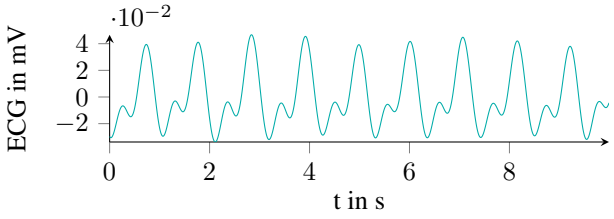
The PSGs contain manually assigned sleep stage labels according to Rechtschaffen & Kales. We summarized S3 and S4 into one stage to yield labels closer to the scoring rules of the American Academy of Sleep Medicine [10]. Other lables, like Movement, were replaced by the label of

Table 1: Model architecture.

Layer	Output Shape	Kernel × Step
Input	240 × 1200 × 1/2/3	
Conv1D	240 × 1185 × 64	16 × 1
Conv1D	240 × 585 × 64	16 × 2
Conv1D	240 × 570 × 64	16 × 1
Conv1D	240 × 278 × 64	16 × 2
Conv1D	240 × 263 × 64	16 × 1
Conv1D	240 × 124 × 64	16 × 2
Conv1D	240 × 109 × 64	16 × 1
Conv1D	240 × 47 × 64	16 × 2
Flatten	240 × 3008	
Dropout	240 × 3008	(rate = 0.3)
Dense	240 × 400	
Dropout	240 × 400	(rate = 0.3)
Bidirectional LSTM	240 × 80	
Dense (Output)	240 × 5	(Softmax)



(a) ECG



(b) ECG after lowpass filtering.

Figure 1: Filtering the ECG using a lowpass filter with cutoff frequency 2 Hz.

the succeeding sleep epoch.

2.2. Previous Model and Preprocessing

Our previous model for sleep stage classification [6] was a Convolutional Recurrent Neural Network that used two parallel inputs: (i) the R-R-interval time series (RRI) from QRS detection on the ECG (specifically using the function *swt_detector* [11] from the Python package *py_ecg_detectors* [12]) and (i) the breath-to-breath-interval time series (BBI) (breath detection according to *respdetect* [13]). Both time series were interpolated with 4 Hz, normalized with z-normalisation, and cut into overlapping segments of 300 s, centered around each 30 s sleep epoch. Due to the recurrent architecture of the Neural Network, 240 consecutive segments are classified as a sequence at the same time (thus classifying a full night in three to four turns). A detailed description of RRI and BBI preprocessing is found in our previous publications [6, 14].

Neither model architecture nor hyperparameters were modified compared to [6], except for the number of parallel input signals. For more details, see Table 1 and [6].

2.3. Training and Evaluation

Important hyperparameters of our training process were Adam Optimizer with a learning rate of 0.001, Categorical Cross-Entropy as loss function, and Early Stopping with patience of ten training epochs.

For each input (resp. input combination), we trained an ensemble of ten models with a data split similar to ten-fold cross-validation. Therefore, each of these models is independent of the hold-out test data. We used the ensembles to classify the test data by majority vote. All presented results are mean values from the 916 PSGs in the test data.

As a key metric for evaluating classification quality, we chose Cohen's Kappa κ [15]. κ is a measure of inter-rater agreement that considers both the observed agreement and the expected agreement by chance. The value of κ ranges from -1 to 1, with values less than 0 being worse than chance. For better interpretability, κ is categorized to show slight (0.01-0.2), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) and almost perfect (0.81-1.0) agreement. [15, 16]

2.4. New Preprocessing

The preprocessing for the new inputs is very simple. To gain dECG and dRIP, ECG and RIP were filtered using a lowpass filter with cutoff frequency 2 Hz. Note that ECGs and RIPs in the SHHS are prefiltered with a highpass filter with cutoff frequencies 0.15 Hz (ECG) and 0.05 Hz (RIP). An exemplary segment of the ECG after lowpass filtering is displayed in Figure 1. Afterwards, the signals were downsampled to 4 Hz. HR and SpO_2 are supplied by the PSGs in the SHHS at a sampling rate of 1 Hz. Therefore, we upsampled them to 4 Hz by sample-and-hold technique. Similar to our previous preprocessing, these new inputs were normalized, cut into overlapping segments of 300 s, and centered around each 30 s sleep epoch.

Due to the secondary aim of using few sensors, we only combined inputs from two of our three sensors: ECG, RIP, and PO. This resulted in the following input combinations, that we considered most interesting: (i) RRI, (ii) dECG, (iii) HR, (iv) HR & SpO_2, (v) RRI & BBI, (vi) dECG & dRIP, (vii) HR & dRIP, (viii) HR & SpO_2 & dRIP.

Table 2: Confusion matrix of hold-out test data classified from the input signals HR & dRIP into five sleep stages according to Rechtschaffen & Kales: W, REM, S1, S2, S3, and S3+S4. A total number of 866 351 epochs was classified in the hold-out test data.

		Label					Precision
		W	REM	S1	S2	S3+S4	
Prediction	W	199 068	2 183	8 637	12 923	922	88.98 %
	REM	3 907	103 902	2 959	13 720	412	83.19 %
	S1	1 023	483	1 594	819	5	40.62 %
	S2	25 648	15 852	18 302	307 812	54 153	72.98 %
	S3+S4	722	204	55	31 083	59 963	65.16 %
Sensitivity		86.41 %	84.73 %	5.05 %	84.02 %	51.94 %	
Accuracy							77.61 %

Table 3: Mean Cohen's Kappa κ on hold-out test data for different inputs and input combinations when classifying into five sleep stages.

Input	Signal/ Sensor	κ
RRI	ECG	0.58
dECG	ECG	0.61
HR	PO	0.60
HR & SpO_2	PO	0.63
RRI & BBI	ECG, RIP	0.64
dECG & dRIP	ECG, RIP	0.65
HR & dRIP	PO, RIP	0.68
HR & SpO_2 & dRIP	PO, RIP	0.68

Table 4: Comparison of our results to results from literature.

Source	Preprocessing	Sensor	κ
[1] (2020)	Downsampling	PPG	0.51
[2] (2020)	Binary Representation (ECG), None (RIP)	ECG, RIP	0.53
[3] (2020)	Instantaneous Heart Rate	ECG	0.61*
[4] (2021)	Downsampling	PPG	0.60
[5] (2021)	None	PO	0.74**
This work	Up-/ Downsampling	PO, RIP	0.68

* classification into four sleep stages (Wakefulness, N1+N2, N3, REM)

** classification into two sleep stages (Wakefulness, Sleep)

3. Results

As Table 3 shows, using just one cardiac signal, dECG and HR yield equal or better results than RRI around 0.6 for κ. Adding any respiration signal (BBI, dRIP) or SpO_2 to these cardiac signals improves these κ classification results by 0.03 to 0.08. The best combination of input signals is HR and dRIP yielding κ of 0.68, see detailed classification results in Table 2. The addition of SpO_2 as a third input does not change this classification quality. This new signal combination outperforms our previous approach by 0.04.

When further summarizing the sleep stages into three classes - Wakefulness, NonREM sleep, and REM sleep - this best signal combination of HR & dRIP yields κ of 0.81. This is an almost perfect agreement with the manual annotation [16] and also slighty outperforms our previous results of 0.80 with RRI & BBI [6].

Comparing these results to results from literature (see Table 4), we see that most of our input combinations, but especially HR & dRIP, outperform those results.[1] Remarkably, there is only one other approach (by Sun et al. [2]) that combines two input signals. Note that κ usually increases distinctly when summarizing sleep stages into fewer classes, therefore only the same number of classes should be compared directly (see e. g. detailed results in [6] for illustration, with κ of 0.68 for four classes and 0.80 for 3 classes).

4. Conclusion

In our investigation, input signals with very simple preprocessing yield just as good (and better) classification results as the more elaborate RRI and BBI. Mere filtering and downsampling proved to be a simple yet powerful preprocessing strategy for the ECG regarding sleep stage classification with Convolutional Recurrent Neural Networks. This preprocessing is significantly easier to implement and faster than QRS detection. To the best of our knowledge,

[1] Note, that there is a preprint by Kotzen et al. with an outstanding κ of 0.75 for classification into four sleep stages from the PPG. Apparently, the preprint is submitted and currently under peer-review. https://arxiv.org/abs/2202.05735v4

dECG or a comparable approach has not been used as input for sleep stage classification or any classification task.

We assume that dECG and HR (measured by PO with 1 Hz) are more robust than RRI and therefore yield better results. Additionally, even though this HR contains seemingly less information than RRI, the underlying patterns might be easier to learn by a Convolutional Neural Network. dRIP seems to contain the same information as SpO_2 (and more), as the comparisons of HR & SpO_2, HR & dRIP, and HR & dRIP & SpO_2 show.

New signals and signal combinations are very promising to improve classification quality. Unfortunately, signal combinations are barely investigated yet (see Table 4), even though we know from feature-based classification approaches that cardiorespiratory coupling provides relevant features for sleep stage classification.

Overall, the variety of inputs in our investigation suggests that there might be many more sensors, signals and preprocessing strategies, that are suitable for sleep stage classification. As our approach demonstrates, these new signals do not necessarily require new model architectures.

Acknowledgments

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

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

References

[1] Korkalainen H, Aakko J, Duce B, Kainulainen S, Leino A, Nikkonen S, Afara IO, Myllymaa S, Töyräs J, Leppänen T. Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea. Sleep nov 2020;43(11).

[2] Sun H, Ganglberger W, Panneerselvam E, Leone MJ, Quadri SA, Goparaju B, Tesh RA, Akeju O, Thomas RJ, Westover MB. Sleep staging from electrocardiography and respiration with deep learning. Sleep jul 2020;43(7).

[3] Sridhar N, Shoeb A, Stephens P, Kharbouch A, Shimol DB, Burkart J, Ghoreyshi A, Myers L. Deep learning for automated sleep staging using instantaneous heart rate. npj Digital Medicine aug 2020;3(1).

[4] Huttunen R, Leppänen T, Duce B, Oksenberg A, Myllymaa S, Töyräs J, Korkalainen H. Assessment of obstructive sleep apnea-related sleep fragmentation utilizing deep learning-based sleep staging from photoplethysmography. Sleep oct 2021;44(10).

[5] Casal R, Di Persia LE, Schlotthauer G. Classifying sleep–wake stages through recurrent neural networks using pulse oximetry signals. Biomedical Signal Processing and Control jan 2021;63.

[6] Goldammer M, Zaunseder S, Brandt MD, Malberg H, Gräßer F. Investigation of automated sleep staging from cardiorespiratory signals regarding clinical applicability and robustness. Biomedical Signal Processing and Control jan 2022;71.

[7] Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O'Connor GT, Rapoport DM, Redline S, Robbins J, Samet JM, Wahl PW. The Sleep Heart Health Study: Design, Rationale, and Methods. Sleep dec 1997;20(12):1077–1085.

[8] 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 oct 2018; 25(10):1351–1358.

[9] National Sleep Research Resource. Sleep Heart Health Study. URL `https://doi.org/10.25822/ghy8-ks59`.

[10] Berry RB, Quan SF, Abreu AR, Others; for the American Academy of Sleep Medicine. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Version 2.6 edition. Darien, IL: American Academy of Sleep Medicine, 2020.

[11] Kalidas V, Tamil L. Real-time QRS detector using Stationary Wavelet Transform for Automated ECG Analysis. In 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE). oct 2017; 457–461.

[12] Howell L, Porr B. py-ecg-detectors 1.0.2, 2019.

[13] Vidaurre C, Sander TH, Schlögl A. BioSig: The Free and Open Source Software Library for Biomedical Signal Processing. Computational Intelligence and Neuroscience mar 2011;2011.

[14] Goldammer M, Zaunseder S, Malberg H, Gräßer F. Specializing CNN Models for Sleep Staging Based on Heart Rate. In 2020 Computing in Cardiology Conference (CinC), volume 47. sep 2020; .

[15] Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement apr 1960; 20(1):37–46.

[16] Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics mar 1977; 33(1):159–174.

Address for correspondence:

Technische Universität Dresden
Institut für Biomedizinische Technik
Miriam Goldammer
01062 Dresden
Germany
miriam.goldammer@tu-dresden.de