Sleep-Wake Classification for Home Monitoring of Sleep Apnea Patients

Dorien Huysmans[1], Eva Heffinck, Ivan D Castro[2], Margot Deviaene[1], Pascal Borzée[3], Bertien Buyse[3], Dries Testelmans[3], Sabine Van Huffel[1], Carolina Varon[1,4]

[1] KU Leuven, Department of Electrical Engineering (ESAT), STADIUS, Leuven, Belgium
[2] imec, Leuven, Belgium
[3] UZ Leuven, Department of Pneumology, Leuven, Belgium
[4] TU Delft, Circuits and Systems (CAS) group, Delft, the Netherlands

Abstract

Sleep apnea is a common sleep disorder, whose diagnosis can strongly benefit from home-based screening. As the total sleep time is essential to assess the sleep apnea severity, a sleep-wake classifier was developed based on heart rate and respiration. These two signals were selected as they can be measured using unobtrusive sensors. A 1D convolutional neural network (CNN) was designed to classify 30s epochs of tachograms and respiratory inductance plethysmography (RIP) signals. The input based on beat-to-beat variability allows the use of different sensor types. A dataset of 56 patients with an apnea-hypopnea index (AHI) below 10 was used to train and validate the network. This CNN was applied to an independent test set of ECG and RIP signals of 25 subjects. Of these, 8 subjects were simultaneously monitored using an unobtrusive capacitive-coupled ECG (ccECG) sensor integrated in a mattress. Artefact removal and data correction was performed on this acquired data. The performance on the independent dataset of ECG and RIP is comparable to state-of-the-art, with $\kappa = 0.48$*. However, application on the ccECG data resulted in a drop in performance, with* $\kappa = 0.30$*. This was caused by a low amount of remaining wake epochs after data cleaning. Importantly, the network classified 30s segments of sleep apnea patients, without relying on past or future information for feature extraction.*

1. Introduction

Obstructive sleep apnea (OSA) has the highest prevalence among sleep disorders. As it is closely associated with obesity and advancing age, the prevalence is expected to further increase [1]. The disorder is characterized by repeated breathing interruptions, resulting in hypoxaemia and arousals from sleep. Overall, these events fragment the patient's sleep and reduce phases of rapid eye movement and slow wave sleep, leading to excessive daytime sleepiness as well as cardiovascular problems. Patients suspected of OSA are referred to the hospital for diagnosis. There, the severity of sleep apnea is assessed by the Apnea-Hypopnea Index (AHI), which is the number of respiratory events (apneas and hypopneas) per hour of sleep. Thus, the amount of sleep needs to be quantified to not underestimate the AHI, as would be the case when counting the full time in bed. In clinical practice, EEG signals are analysed for sleep scoring. As such, the overall procedure poses a high level of obtrusiveness for the patient and is furthermore expensive and laborious. Therefore, it is desired to perform sleep apnea screening at home with unobtrusive sensors, allowing a normal sleeping environment and follow up over multiple days. Many emerging unobtrusive sensor technologies for sleep monitoring are based on cardiac and respiratory signals. As a consequence, the development of novel algorithms for automated sleep staging based on these unobtrusive signals is an active topic of research. However, most classification algorithms so far, were developed for healthy people. But in OSA patients, altered heart rate variability and sleep fragmentation complicates algorithm design and validation.

Often, feature-based approaches were implemented to differentiate between sleep stages when expert knowledge was available [2, 3]. Thus, one disadvantage was this required level of prior knowledge. Another disadvantage was the extensive processing for feature extraction. Furthermore, these algorithms classified 30s or 60s epochs, although the features were based on longer surrounding signal segments. As such, they included contextual information, but required long segments of good quality. In contrast, a neural network automatically derives features by representation learning from a raw signal of any length. Malik et al. performed sleep-wake classification solely using the instantaneous heart rate (IHR) and a one-dimensional convolutional neural network (1D CNN) [4]. In the current study, a sleep-wake classifier based on this 1D CNN was developed for sleep apnea patients. It was not only fed with tachograms, but also respiratory inductance plethysmography (RIP) signals from polysomnography (PSG). The use of tachograms allows a straightforward application of other sensors capturing the beat-to-beat variability. Additionally, the network was designed to process independent 30s epochs, to alleviate the re-

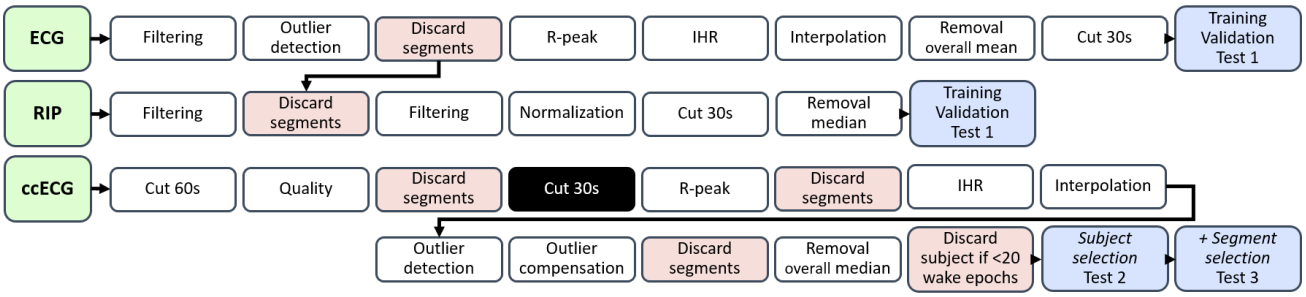


Figure 1: Data processing flow and data selection.

quirement of successive good quality segments. Furthermore, data from capacitive-coupled ECG acquired from a mattress-integrated sensor was applied to test the developed algorithm [5].

2. Methods

2.1. Dataset

The dataset comprised 81 patients with AHI<10, who were referred to the sleep lab of UZ Leuven for a diagnostic PSG. Out of these, 56 (70%) patients were used for training and validation, and 25 (30%) patients were kept for testing. Within the test patients, 8 were recorded with a ccECG. Additionally, the algorithm was applied to a dataset of 158 subjects with a full AHI range of 29.5 $\pm$ 23.6.

2.2. Data processing and selection

In this study the network was developed based on PSG signals, and later applied onto an unobtrusive signal. This resulted in development of different preprocessing steps, and these are illustrated in Figure 1. On the one hand, full-night recordings of ECG and RIP were extracted from the PSG. Therefore, preprocessing was performed over the full signal and afterwards cut into non-overlapping windows of 30s. On the other hand, the ccECG sensor was a multi-electrode array in which the best available channels were selected window-based. Additionally, the ccECG data was preprocessed using a 60s window to apply a state-of-the-art quality indicator [6]. If the segment quality did not meet the standard, it was discarded and the remaining segments were further cut into 30s. Every 30s ccECG segment was further processed individually and additional data cleaning was required as the sensor was prone to movement artefacts. The detailed processing steps of every signal were the following:

ECG: First, the signal was low-pass filtered with a cut-off frequency of 125 Hz and downsampled to 250 Hz. Then, artefacts were detected due to movement or bad skin contact, resulting in saturation of the signal or small amplitude noise. A segment of 30s was removed if the peak-to-peak (PP) amplitude of the segment was equal or smaller than 3% of the maximum PP amplitude. Following, R-peak detection was performed using the Pan-Tompkins algorithm [7]. From the RR-interval, the resulting IHR was derived and expressed in beats per minute (BPM). The unevenly sampled IHR data points were interpolated at 4 Hz by a piecewise cubic hermite interpolating polynomial. The overall mean was removed from the IHR per patient. This way, inter-subject variability was removed but the inter-sleep stage variability retained. Finally, the signal was cut into non-overlapping windows of 30s. These signals were used for training, validation and *Test 1*.

RIP: The RIP signal was filtered similarly as the ECG signal and corresponding artefact segments were removed. It was further low-pass filtered at 2 Hz and downsampled to 4 Hz. Then, the complete signal was normalized by removing the overall mean and dividing by the standard deviation. Finally, the signal was cut into non-overlapping windows of 30s, followed by the subtraction of the median per segment. These signals were used for training, validation and *Test 1*.

ccECG: First, R-peak detection was performed on 30s segments, with the method proposed in [8]. Segments with less than 15 peaks were discarded. From the remaining segments, the IHR was derived and expressed in BPM, and further interpolated at 4 Hz. Outliers were identified if the IHR value was outside the range of 40 to 180 BPM, or the segment's median value $\pm$ 20 BPM, or the segment's median value $\pm$ (3 $\times$ the standard deviation). These outliers were indicated with NaN and compensated if the interval of subsequent NaNs was no longer than 2.5s. This gap was filled by mirroring the values preceding the gap. As the neural network cannot process NaN values, every segment containing NaNs (i.e. outliers which could not be filled in) was discarded. Therefore, outlier correction was important to not discard epochs with minor artefacts and retain as many segments as possible. Finally, the median of all remaining segments was subtracted from every segment. After this data processing, several subjects had an insufficient amount of remaining wake epochs to obtain a representative classification result. Therefore, only subjects with at least 20 wake epochs were included and these subjects defined *Test 2*. The subjects of *Test 2* and their remaining segments defined *Test 3*.

In the case of the ccECG data, *Test 2* and *Test 3* were

identical. However, the ECG and RIP signals were also sent through this "ccECG pipeline" to make fair comparisons. First, the PSG signals of *Test 2* were processed from phase "cut 30s", and were applied for testing. Secondly, from these processed PSG segments, only the segments defined by *Test 3* were applied for testing.

2.3. Network architecture

The network consisted of a convolutional part, to extract features, and a dense part, to classify the segments (see Figure 2). Two separate unimodal networks were first optimized using the cardiac or respiratory signal, based on [4]. Once these networks were trained, they were combined in a multimodal network. All networks consisted of four types of layers. A first type were the convolutional layers, $(f, k, s) - Conv$, with a depth f, a kernel size k, and a stride s. After the convolutional block, dense layers, $(n) - Dense$, with n neurons were included. A third type were dropout layers, $(p) - Dropout$, where p% of the weights equal to zero in every training step to avoid overfitting. The output layer is a softmax layer, $Softmax(1, c)$, delivering posterior class probabilities for every one of the c classes. Weights of the convolutional layers of the optimized unimodal networks were frozen, in order to only train the dense layers, using both ECG and RIP. In order to assign an input segment to the sleep or wake class, the threshold of posterior class probability was set at 0.5.

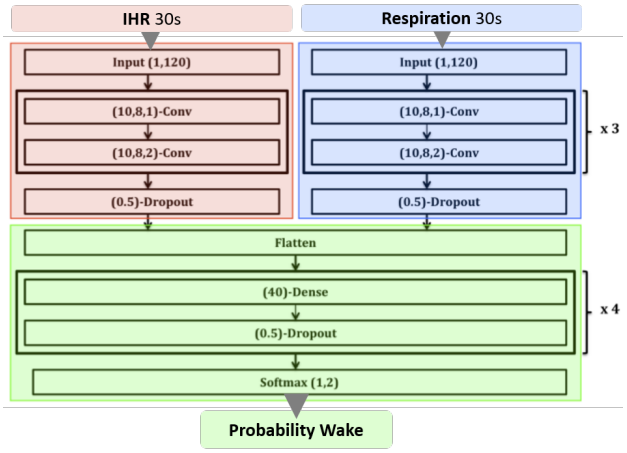


Figure 2: CNN-based classifier.

3. Results and discussion

With the rapid improvement of wearable and unobtrusive sensors, existing sleep staging algorithms were designed to function with wearable input data by focusing on cardiac and respiratory information. Similarly, the network in this study was developed using tachograms, which offers the possibility of application with different types of sensors capturing the beat-to-beat variability. Moreover, the current design also took into account the fact that wearable data is often distorted and only short segments of good quality are delivered. Therefore, the network input was purely based on single 30s segments, to learn features which did not require past or future information. Also, the data processing techniques were developed for these short term segments. The investigation of different processing and data selection mechanisms was therefore important in the light of applying existing algorithms to wearable data. As such, the current algorithm was compared with literature, and further on, the data processing and selection steps were evaluated.

Table 1 displays the results of sleep-wake classification using different data selections and preprocessing techniques, and application on either the unimodal or multimodal network, including comparison with literature.

The networks were optimised based on the training and validation sets. Here, it was observed that the unimodal network reached a higher performance using respiratory input compared to cardiac information. Although, when combining the input signals, the latter seemed to have additional information, which could increase the general performance. However, application on different test sets did not confirm this. Respiratory signals seemed to retain all useful information, and the addition of a cardiac signal resulted in a reduction of performance.

Training the network on low AHI data diminished the effect of apnea segments on sleep-wake classification. When the network was applied to data from the full AHI range, the performance decreased. This was expected, as heart rate and respiration are not only affected by sleep stages, but also by apneic events and arousals. As these apneic events most drastically disrupted the RIP signal, the classification performance using only RIP decreased most, from κ=0.51 to κ=0.44.

Furthermore, representation learning proved to be very valuable compared to feature based classification, as the one proposed in [2]. Although the algorithm proposed in [3] presented state-of-the-art results based on ECG only, features were complex and required several epochs. Additionally, a cross validation scheme was used. Nevertheless, the sleep-wake classification in this work did outperform the one of Malik et al. [4], in which the CNN reached a κ of 0.26 using a 30s tachogram input. Additionally, the studied population was not suffering from OSA.

Comparison of the PSG performance with ccECG-type preprocessing on *Test 2* showed negligible differences, despite variations in processing pipeline. The main difference was the outlier-gap correction, although, as few artefacts were present in the PSG signals, few corrections were carried out.

Further on, a large performance drop was observed from ccECG-type Test 2 to Test 3, although both were using PSG ECG signals. This could be accounted to the fact that only half of the epochs were kept and the percentage of wake epochs drastically dropped. With a low number of wake epochs left, the sleep-wake classification problem

Table 1: Classification results with different datasets. Values are indicated as mean ± standard deviation over patients, except the last row. Additionally, results are compared with literature. Application of ccECG is indicated as (cc).

Process.	Subset	#	AHI	# Epochs	% Wake	Cohen's Kappa score		
						ECG	RIP	ECG + RIP
Full rec.	Train.	44	5.7 ± 5	1067 ± 126	20 ± 12	0.34 ± 0.01	0.54 ± 0.07	0.59 ± 0.03
Full rec.	Valid.	12	5.4 ± 2	1085 ± 108	22 ± 11	0.40 ± 0.01	0.43 ± 0.04	0.47 ± 0.01
Full rec.	Test 1	25	4.7 ± 3	1058 ± 115	15 ± 7	**0.29 ± 0.18**	**0.51 ± 0.12**	**0.48 ± 0.13**
Full rec.		158	29.5 ± 24	1064 ± 125	22 ± 12	0.27 ± 0.16	0.44 ± 0.17	0.42 ± 0.16
ccECG	Test 2	4	5.8 ± 2	981 ± 192	24 ± 9	0.32 ± 0.14	0.51 ± 0.09	0.47 ± 0.15
ccECG	Test 3	4	5.8 ± 2	482 ± 186	16 ± 8	0.14 ± 0.13	0.36 ± 0.08	0.29 ± 0.19
ccECG	Test 3	4	5.8 ± 2	482 ± 186	16 ± 8	**0.11 ± 0.14 (cc)**	**0.36 ± 0.08**	**0.30 ± 0.17**
[2] Feature ($\gg$60s)		25	24.2 ± 20	831 ± 63	23 ± 11	-	-	0.37
[3] Feature ($\gg$30s)		51	>5	960 ± 60	19 ± 13	0.60 ± 0.15[1]	-	-
[4] CNN (30s)			<5	20 102	17.1	0.26	-	-

[1] Performance of 4-class problem: Wake, REM, N1/N2, N3

became more difficult. Indeed, mostly wake and N2 signals were discarded by quality assessment, as the signal quality of ccECG dropped when a person moved. As future work, a different quality assessment will be explored, which only considers the beat extraction from a ccECG epoch and not the correspondence of wave morphology with ECG. This could increase the number of remaining wake epochs.

Lastly, using the exact same processing and data selection with a different input, being either ECG or ccECG signals, similar performances were obtained. The main difference between both datasets was the more frequently used outlier-gap compensation in the case of ccECG signals. This procedure, however, did not influence the classification results, and could therefore be considered as an adequate procedure for this type of data and network.

4. Conclusion

A sleep-wake classifier was designed for application with wearable and/or unobtrusive sensors, to enable home monitoring of sleep apnea patients. Using PSG signals, the performance was comparable to existing literature with a similar dataset. More important, the network classified short-term 30s segments, without relying on past or future information for feature extraction. This increased potential application on wearable signals, as large parts of data are often discarded due to inferior quality. However, application on a cleaned dataset of ccECG and RIP resulted in a drop in performance. This was due to a low number of non-discarded wake epochs per subject recording. An extended ccECG data collection, including unobtrusive respiration monitoring is expected for further development and validation.

Acknowledgements

Agentschap Innoveren en Ondernemen (VLAIO): 150466: OSA+ ; Agentschap voor Innovatie door Wetenschap en Technologie (IWT): O&O HBC 2016 0184 eWatch ; This research received funding from the Flemish Government (AI Research Program). Sabine Van Huffel, Margot Deviaene, Carolina Varon and Dorien Huysmans are affiliated to Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium.

References

[1] Senaratna CV, et al. Prevalence of obstructive sleep apnea in the general population: A systematic review. Sleep Medicine Reviews 2017;34:70 – 81.

[2] Willemen T, et al. Probabilistic cardiac and respiratory based classification of sleep and apneic events in subjects with sleep apnea. Physiol Meas 2015;36(10):2103.

[3] Radha M, et al. Sleep stage classification from heart-rate variability using long short-term memory neural networks. Scientific Reports 2019;9(1):1–11.

[4] Malik J, et al. Sleep-wake classification via quantifying heart rate variability by convolutional neural network. Physiol Meas 2018;39(8):085004.

[5] Castro ID, et al. Capacitive multi-electrode array with real-time electrode selection for unobtrusive ECG & BIOZ monitoring. In EMBC. IEEE, 2019; 5621–5624.

[6] Castro ID, et al. Unobtrusive, through-clothing ECG and bioimpedance monitoring in sleep apnea patients. In CinC. IEEE, 2020; Page–1.

[7] Pan J, Tompkins WJ. A real-time qrs detection algorithm. IEEE Trans Bio Med Eng 1985;(3):230–236.

[8] Varon C, et al. A novel algorithm for the automatic detection of sleep apnea from single-lead ECG. IEEE Trans Bio Med Eng 2015;62(9):2269–2278.

Address for correspondence:

Dorien Huysmans
ESAT/STADIUS/KU Leuven
Kasteelpark Arenberg 10, bus 2446, 3001 Leuven, Belgium
dorien.huysmans@esat.kuleuven.be