Cascading detection model for apnea-hypopnea events based on nasal flow and arterial blood oxygen saturation

Hui Yu [1], Chenyang Deng [2], Jinglai Sun [3], Yanjin Chen [4] and Yuzhen Cao [5*]
[1] Department of Biomedical Engineering, Tianjin University, Tianjin, China
[2] Department of Biomedical Engineering, Tianjin University, Tianjin, China
[3] Department of Biomedical Engineering, Tianjin University, Tianjin, China
[4] Tianjin Hospital of ITCWM Nankai Hospital, Tianjin, China
[5] Department of Biomedical Engineering, Tianjin University, Tianjin, China, Tel: 086-022-27406546, E-mail: yuzhencao18@126.com

Abstract

Purpose

Sleep apnea and hypopnea syndrome (SAHS) seriously affects sleep quality. In recent years, much research has focused on the detection of SAHS using various physiological signals and algorithms. However, there are still some limitations such as low detection resolution.

Methods

A 60 s detector and a 10 s detector were cascaded to improve the detection resolution for apnea-hypopnea (AH) events. Random forests were adopted for classification of data segments based on morphological features extracted from nasal flow and arterial blood oxygen saturation (SpO_2). The start and end time of every AH event and the AH index (AHI) could be both predicted.

Results

For the more than 280,000 10 s segments in the database, the cascading detection model reached an accuracy of 89.0%. For more than 1800 AH events, it achieved a sensitivity of 82.8%, while the Pearson's correlation coefficient between estimated AHI and reference AHI was 0.98. In the diagnosis of SAHS severity, the proposed method exhibited a performance with a Cohen's kappa coefficient of 0.83.

Conclusions

The cascading detection model is able to predict AH events and provide an estimate of AHI. The results indicate that it has the potential to be a useful tool for SAHS diagnosis.

Keywords

Sleep apnea and hypopnea syndrome, Apnea-hypopnea index, Polysomnography, Cascading detection model, Apnea-hypopnea events

Introduction

Sleep apnea and hypopnea syndrome (SAHS) is a prevalent sleep breathing disorder in middle-aged people. The gold standard for diagnosis of SAHS is to perform polysomnography (PSG) in a laboratory. However, PSG requires patients to sleep with many sensors for at least one night, the scoring of apnea-hypopnea (AH) events can take a long time. Therefore, many researchers hope to simplify or replace PSG by using a limited number of physiological signals. Electrocardiogram (ECG) was first studied for this purpose. McNames et al. [1] found that heart rate, S-pulse amplitude, and pulse energy were correlated with SAHS. Bsoul et al. [2] cut the ECG into 60 s segments and used support vector machine (SVM) for real-time detection of SAHS. However, many other diseases except SAHS also affect ECG. Hence, nasal flow (NF) [3-6], arterial blood oxygen saturation (SpO_2) [7], snoring [8], or a combination of these signals [9,10] have been adopted more recently. Gutierrez et al. [4] used the overall features of single-channel NF for the diagnosis of SAHS severity. Xie et al. [10] utilized a combination of classifiers to achieve real-time detection of SAHS based on ECG and SpO_2. All the above studies can be roughly divided into two categories: those that predict the AH index (AHI) based on the detection of AH events [3,7,10,2,5,9,11], and those that predict AHI based on the overall signal features [4,12,6,13,8,1]. The latter approach cannot provide time information for each AH event, whereas most studies in the former [2,11,7,10] only involve 60 s segment identification, which may lead to errors in the estimation of AHI. On the other hand, the methods mentioned above include threshold [9,7,5], SVM [2,11,10], and neural network [11,3], which require a large number of hyperparameters to be set by experience. Therefore, we utilized random forests composed of classification and regression trees (CART) based on morphological features extracted from NF and SpO_2 for the real-time detection of SAHS. A 60 s detector and a 10 s detector were cascaded to improve the resolution of detection for AH events.

Materials and Methods

Subjects

The database used in this work was the St. Vincent University Hospital/Dublin University College Sleep Apnea Syndrome Database (UCDDB) [14] public on Physionet [15]. The database contains 25 subjects' PSG data, including EEG, electrooculogram, submental electromyography, NF, ribcage and abdomen movements, SpO_2, snoring, and body position. All signals were obtained using a Jaeger–Toennies system. The annotation files consisted of onset time and duration of respiratory events provided by an experienced specialist. The cutoff value for AHI is commonly set to 5, 15, or 30 events/h [16,7,4,3,17]. There were data for two non-SAHS subjects, twelve mild-SAHS subjects, five moderate-SAHS subjects, and six severe-SAHS subjects in the database. To balance the number of subjects in each class, we randomly selected two, four, four, and five subjects from each category for training and testing. The sleep-related parameters of the subjects are summarized in Table 1.

Table 1. Summary of sleep-related parameters (mean ± standard deviation)

	Non-SAHS	Mild SAHS	Moderate SAHS	Severe SAHS
Age (years)	52.0±15.6	49.0±1.6	57.5±7.2	46.6±5.5
AHI (events/h)	4.1±5.7	12.4±1.8	25.6±3.6	43.8±16.3
Epworth Sleepiness Score	7.0±8.5	14.5±3.9	9.3±6.2	12.4±7.9

According to the American Academy of Sleep Medicine (AASM) manual [16], apnea is scored when there is a more than 90% drop in the peak signal of the pre-event baseline for NF with a duration longer than 10 s. Hypopnea is scored by the following rules: 1) there is a more than 30% drop in the peak signal of the pre-event baseline for nasal pressure with a duration longer than 10 s, accompanied by 2) more than 3% arterial oxygen desaturation or an arousal. As a result, we selected NF and SpO_2 for SAHS detection.

Study design

The cascading detection model based on AH event detection is shown in Fig. 1. It comprises the following main steps. 1) Removal of invalid data, signal filtering, segmentation with a sliding window, and SpO_2 alignment. 2) Extraction of a specific feature set from each segment. 3) The cascading detection model predicts each segment and outputs a sequence of results. 4) The event detector corrects the invalid results in the sequence and calculates the AHI.

Fig. 1 Design of cascading detection model based on AH event detection

Signal preprocessing

Signal preprocessing comprises the following four steps. 1) Removal of invalid data. Any SpO_2 values lower than 80% were considered to be artifacts and removed from the analysis (5.8% of the data). 2) Signal filtering: a four-point sliding average filter and a third-order Butterworth high-pass filter with a cut-off frequency of 0.05 Hz were used to prevent baseline drift and high-frequency noise caused by artifacts. 3) Segmentation: the original signals were segmented using a 60 s window and a 10 s window, respectively. In both cases, the step was set to 1 s. Segments were categorized into two classes: AH and N (normal). The segments containing more than 5 s SAHS were labeled as AH. Other cases were labeled as N. 4) SpO_2 alignment. As SpO_2 responds slowly to AH events [18], we moved SpO_2 forward τ s ($0<\tau<30$). The results showed that the model performed best with τ set to 23 s. After preprocessing, we obtained a total of 35,309 AH segments and 249,977 N segments.

Feature extraction

NF feature set

According to the AASM definition of AH events, the amplitude of the NF provides important information. Therefore, we first extracted the maximum (F_p) and minimum (F_v) points from the NF data. Then the tidal volume F_T per breath was calculated as the difference between two adjacent extreme points. The mean, standard deviation, and range of the tidal volume were extracted for each segment. We calculated the baseline of the tidal volume every 30 s using equation (1):

$$F_b(i) = \max\{\mathrm{abs}[F_p(i-30:i), F_v(i-30:i)]\} \quad (1)$$

where i represents the ith segment. The number of breaths with tidal volume reductions of more than 30% (f_{ap}) and 70% (f_{ha}), and the number of normal breaths (f_{nor}) were calculated using equations (2)–(4).

$$f_{ap}(i) = FREQUENCY\{F_T(i) < F_b(i) \times 0.3\} \quad (2)$$

$$f_{ha}(i) = FREQUENCY\{F_T(i) < F_b(i) \times 0.7\} \quad (3)$$

$$f_{nor}(i) = FREQUENCY\{F_T(i) > F_b(i) \times 0.85\} \quad (4)$$

Besides, owing to the cessation of breathing, there will be fluctuations in the breathing rate during AH events. One normal breath lasts for 3–5 s, energy will be concentrated with a peak in the corresponding frequency. As a result, we took the fourth statistical moment (f_{kur}) in 0.2–0.4 Hz of NF's frequency spectrum as another feature.

SpO_2 feature set

There are always fluctuations in SpO_2 during AH events. Hence, we first calculated the standard deviation and range coefficients of SpO_2 in each segment. The slope of SpO_2 in each segment (sp_{slope}) was also calculated. Two other commonly used features were adopted: the time that SpO_2 stays below 92% and 91% [19,20] in each segment, respectively. We took the maximum (Sp_{b1}) and average (Sp_{b2}) SpO_2 value every 30 s as the baseline. The duration and level of oxygen desaturation were calculated for each segment using equations (5)–(8).

$$sp_{un1}(i) = FREQUENCY\{Sp(i) < Sp_{b1}(i) \times 0.98\} \quad (5)$$

$$sp_{un2}(i) = FREQUENCY\{Sp(i) < Sp_{b2}(i) \times 0.98\} \quad (6)$$

$$sp_{de1}(i) = Sp_{b1}(i) - mean\{Sp(i)\} \quad (7)$$

$$sp_{de2}(i) = Sp_{b2}(i) - mean\{Sp(i)\} \quad (8)$$

All the above features were extracted from each segment; the total feature set is shown in Table 2.

Table 2. Features and their definitions

Index	Name	Definition
1	f_{mean}, f_{std}, f_{ran}	Average, standard deviation, and range of tidal volume
2	f_{ha}, f_{hap}	Number of breaths with a reduction more than 30% in tidal volume and its ratio to total number of breaths
3	f_{ap}, f_{app}	Number of breaths with a reduction more than 70% in tidal volume and its ratio to total number of breaths
4	f_{nor}, f_{norp}	Number of normal breaths and its ratio to total number of breaths
5	f_{kur}	Fourth statistical moment in 0.2 – 0.4 Hz of NF's frequency spectrum
6	sp_{std}, sp_{ran}	Standard deviation and range of SpO_2
7	sp_{slope}	Slope of SpO_2
8	sp_{un1}, sp_{un2}	Duration of SpO_2 desaturation
9	sp_{de1}, sp_{de2}	Level of SpO_2 desaturation
10	sp_{92}, sp_{91}	Duration of SpO_2 staying below 92% and 91%

Design of cascading detector

The cascading detector contained two parts. The first was a random forest consisting of 10 CARTs for the prediction of 60 s segments. This could screen out most of the N segments while retaining the AH segments. The second part was a random forest consisting of 20 CARTs for the prediction of 10 s segments. Based on the results of the 60 s detector, the 10 s detector was able to locate the start and end time of AH events.

Note that the 60 s detector was trained using a feature set composed of features 2, 6, and 8, in order to improve the training speed; the results indicated that there was almost no effect on the performance. Owing to the imbalance in the number of AH and N segments, the weights for the two classes in CARTs were set to inverse ratio of their numbers.

A twofold cross-validation was used in the test. Each time, half of the segments were used for training with the remaining half used for testing. The cascading detector output the sequence composed of the prediction results of the 10 s segments. The detector was trained on a computer with an i5-7600k CPU and 8 G RAM.

Design of event detector

The sequence predicted by the cascading detector was then fed into the event detector to correct invalid results following two rules. 1) Only an event lasting longer than 10 consecutive AH segments was considered to be a valid AH event. As the original data were segmented by a 10 s window, and one AH event lasts at least 10 s, one AH event corresponded to at least 10 consecutive AH segments. Any AH segment which did not meet the rule was modified to N. 2) The number of N segments between two adjacent AH segments was supposed to be more than five. This was also determined by the nature of the data segmentation. Any segment that did not meet the rule was reset to AH.

Results

The cascading detection model was able to estimate AHI and provide the time information for each AH event. We analyzed its performance with respect to two aspects: segments and AHI.

Segment analysis

Table 3. Results for segments and events

			Reference		ACC (%)	SEN (%)	SPE (%)
			AH	N			
Segments	Estimated	AH	25942	21898	89.0	73.5	91.2
		N	9367	228079			
	Wrong		Right		Total	PPV (%)	SEN (%)
AH events	593		1513		1828	71.8	82.8

ACC accuracy, *SEN* sensitivity, *SPE* specificity, *PPV* positive predictive value

The test set contained data for 15 subjects, a total of 285,286 10 s segments. The prediction results for the above data are shown in Table 3. The cascading detection model achieved an accuracy of 89.0%, a sensitivity of 73.5%, and a specificity of 91.2%. Table 3 also summarizes the prediction results for AH events. The cascading detection model detected 1513 of 1828 AH events, achieving a sensitivity of 82.8% with a positive predictive value (PPV) of 71.8%.

AHI analysis

Fig. 2(a) shows a scatter plot of the AHI (AHI_{est}) estimated by the model and the AHI (AHI_{ref}) determined from PSG. The solid line fitted shows a high correlation (Pearson's correlation coefficient 0.98) between AHI_{est} and AHI_{ref}. Fig. 2(b) shows the Bland–Altman plot of AHI_{est} and AHI_{ref}. The average error of AHI_{est} and AHI_{ref} was -1.7 events/h, and the error range was -5.7 to 2.3 events/h (95% confidence interval).

Fig. 2 a Scatter plot of AHI_{est} and $AHI_{ref.}$ **b** Bland–Altman plot of AHI_{est} and AHI_{ref}

Table 4 summarizes the classification results for SAHS severity. The mean values for sensitivity, specificity, PPV, and accuracy were 100.0%, 87.8%, 87.1%, and 93.3%, respectively, for AHI thresholds of 5, 15, and 30 events/h.

Table 4. SAHS severity classification and diagnostic performance

		Determined from PSG					AHI cutoff (events/h)			
		Non	Mild	Moderate	Severe		≥5	≥15	≥30	AVE
Estimated	Non	2	0	0	0	SEN(%)	100.0	100.0	100.0	100.0
	Mild	0	3	0	0	SPE(%)	100.0	83.3	80.0	87.8
	Moderate	0	1	2	0	PPV(%)	100.0	90.0	71.4	87.1
	Severe	0	0	2	5	ACC(%)	100.0	93.3	86.6	93.3

ACC accuracy, *SEN* sensitivity, *SPE* specificity, *PPV* positive predictive value

Discussion

We proposed a cascading detection model that could predict AHI based on AH event detection. Compared with PSG, only NF and SpO_2 were used. Previously, the original signals were commonly cut into 60 s segments for AH event detection [2,11,7,10]; however, this approach had one limitation: the detection resolution was not high enough. It can only determine whether there was AH in the segment, however, in some cases it cannot distinguish between two AH events with a short interval, which may lead to an error in AHI estimation. Therefore, some researchers [9,7] cut the signals into shorter segments for detection. However, it is difficult to extract effective features from a segment shorter than 10 s, because there will be no more than five complete breaths in one segment in most cases. As a result, we proposed a cascading detection model composed of a 60 s detector and a 10 s detector to predict AH events precisely. Table 3 shows the classification results for the segments. Notably, the model tended to make false positive errors. Of these errors, around 10.3% actually met the rules recommended by AASM in 2012 (Fig. 3(a)). In the case of false negative errors, approximately 85.2% did not meet the rules in the AASM manual (Fig. 3(b)). These segments might have been annotated based on signals from other channels, such as ribcage or abdomen movements. The cascading detection model achieved a sensitivity of 94.9% and a specificity of 92.1% when these segments were excluded.

Fig. 3 a False positives in the prediction results. **b** False negatives in the prediction results

As illustrated in Fig. 2, AHI_{est} showed high correlation with AHI_{ref} (Pearson correlation coefficient 0.98). The performance of the model also showed good consistency among different subjects. On the other hand, AHI_{est} was slightly higher than AHI_{ref}. Consequently, SAHS severity was overestimated for three subjects; for the remaining 12 subjects, the model gave the correct prediction (Table 4). The average kappa coefficient of the cascading detection model for diagnosis of SAHS severity was 0.83, indicating that this method represents a powerful screening tool for SAHS.

We also tested the speed of the cascading detection model. Training required 13.9 s, while only 9.1 s was needed to provide results for all segments and to predict AHI for all 15 subjects. It took 37 ms to predict one segment and 0.6 s to diagnose one subject on average. This implies that the model could be used for real-time SAHS detection.

Compared with the results of other studies, our method exhibited a good sensitivity but not very good specificity. We aim to improve the specificity in future work. More importantly, the model could not only predict the severity of SAHS but could also provide time information for each AH event. Furthermore, compared with other methods such as convolutional neural networks, a smaller number of hyperparameters and less computation were required by our random forest based approach, and the CARTs provided better interpretability for clinical detection.

Table 5. Comparison with other studies

Related work	Method	Signal	AHI cutoff	ACC (%)	SEN (%)	SPE (%)
Choi et al. [3]	Convolutional neural networks	Nasal pressure	5	96.2	100.0	84.6
			15	92.3	98.1	86.5
			30	96.2	96.2	96.2
Gonzalo et al. [4]	AdaBoost-Linear discriminant analysis	Nasal flow	5	86.5	87.1	80.0
			15	81.0	85.9	72.9
			30	82.5	74.2	90.6
Our study	Cascade of random forests	Nasal flow and SpO_2	5	100.0	100.0	100.0
			15	93.3	100.0	83.3
			30	86.7	100.0	80.0

ACC accuracy, *SEN* sensitivity, *SPE* specificity

However, there were some limitations to this study. First, we did not further classify AH events into apnea events and hypopnea events. Second, the model was not tested in an online environment. We hope to confirm the usability of our method online in the future.

Conclusion

The purpose of this study was to propose a model for real-time detection of SAHS. Based on the morphological features of NF and SpO_2, the cascade of a 60 s detector and 10 s detector could not only predict AH, but could also provide time information for each AH event. Compared with previous research, the cascading detection model based on random forests provides better interpretation with reduced computational complexity. Therefore, it is expected to be an effective tool for SAHS diagnosis.

Compliance with ethical standards

Formal consent was not required for this study, as the data had already been de-identified.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. McNames JN, Fraser AM, Ieee I (2000) Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram. In: Computers in Cardiology 2000, Vol 27, vol 27. Computers in Cardiology. pp 749-752. doi:10.1109/cic.2000.898633
2. Bsoul M, Minn H, Tamil L (2011) Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG. Ieee Transactions on Information Technology in Biomedicine 15 (3):416-427. doi:10.1109/titb.2010.2087386
3. Choi SH, Yoon H, Kim HS, Kim HB, Kwon HB, Oh SM, Lee YJ, Park KS (2018) Real-time apnea-hypopnea event detection during sleep by convolutional neural networks. Computers in Biology and Medicine 100:123-131. doi:https://doi.org/10.1016/j.compbiomed.2018.06.028
4. Gutierrez-Tobal GC, Alvarez D, del Campo F, Hornero R (2016) Utility of AdaBoost to Detect Sleep Apnea-Hypopnea Syndrome From Single-Channel Airflow. IEEE Trans Biomed Eng 63 (3):636-646. doi:10.1109/tbme.2015.2467188
5. Lee H, Park J, Kim H, Lee K-J (2016) New Rule-Based Algorithm for Real-Time Detecting Sleep Apnea and Hypopnea Events Using a Nasal Pressure Signal. Journal of Medical Systems 40 (12). doi:10.1007/s10916-016-0637-8
6. Nakano H, Tanigawao T, Furukawa T, Nishima S (2007) Automatic detection of sleep-disordered breathing from a single-channel airflow record. European Respiratory Journal 29 (4):728-736. doi:10.1183/09031936.00091206
7. Jung DW, Hwang SH, Cho JG, Choi BH, Baek HJ, Lee YJ, Jeong DU, Park KS (2018) Real-Time Automatic Apneic Event Detection Using Nocturnal Pulse Oximetry. IEEE Trans Biomed Eng 65 (3):706-712. doi:10.1109/tbme.2017.2715405
8. Sola-Soler J, Antonio Fiz J, Morera J, Jane R (2012) Multiclass classification of subjects with sleep apnoea-hypopnoea syndrome through snoring analysis. Medical Engineering & Physics 34 (9):1213-1220. doi:10.1016/j.medengphy.2011.12.008
9. Huang W, Guo B, Shen Y, Tang X (2017) A novel method to precisely detect apnea and hypopnea events by airflow and oximetry signals. Computers in Biology and Medicine 88:32-40. doi:10.1016/j.compbiomed.2017.06.015
10. Xie B, Minn H (2012) Real-Time Sleep Apnea Detection by Classifier Combination. IEEE Transactions on Information Technology in Biomedicine 16 (3):469-477. doi:10.1109/TITB.2012.2188299
11. Hoa Dinh N, Wilkins BA, Cheng Q, Benjamin BA (2014) An Online Sleep Apnea Detection Method Based on Recurrence Quantification Analysis. Ieee Journal of Biomedical and Health Informatics 18 (4):1285-1293. doi:10.1109/jbhi.2013.2292928
12. Jung DW, Hwang SH, Lee YJ, Jeong D-U, Park KS (2017) Apnea-Hypopnea Index Prediction Using Electrocardiogram Acquired During the Sleep-Onset Period. IEEE Trans Biomed Eng 64 (2):295-301. doi:10.1109/tbme.2016.2554138
13. Timus O, Dogru Bolat E (2017) k-NN-based classification of sleep apnea types using ECG. Turkish Journal of Electrical Engineering and Computer Sciences 25 (4):3008-3023. doi:10.3906/elk-1511-99
14. St. Vincent's University Hospital University College Dublin Sleep Apnea Database (2008).
15. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet - Components of a

new research resource for complex physiologic signals. Circulation 101 (23):E215-E220. doi:10.1161/01.CIR.101.23.e215
16. Berry R, Budhiraja R, Gottlieb D, Gozal D, Iber C, Kapur V, Marcus C, Mehra R, Parthasarathy S, Quan S, Redline S, Strohl K, Davidson Ward S, Tangredi M (2012) Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. J Clin Sleep Med 8 (5):597-619
17. Qureshi A, Ballard RD, Nelson HS (2003) Obstructive sleep apnea. Journal of Allergy and Clinical Immunology 112 (4):643-651. doi:https://doi.org/10.1016/j.jaci.2003.08.031
18. Selvaraj N, Narasimhan R, Ieee (2013) Detection of Sleep Apnea on a Per-Second Basis Using Respiratory Signals. In: 2013 35th Annual International Conference of the Ieee Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Conference Proceedings. pp 2124-2127
19. Olson LG, Ambrogetti A, Gyulay SG (1999) Prediction of sleep-disordered breathing by unattended overnight oximetry. J Sleep Res 8 (1):51-55. doi:10.1046/j.1365-2869.1999.00134.x
20. Magalang UJ, Dmochowski J, Veeramachaneni S, Draw A, Mador MJ, El-Solh A, Grant BJB (2003) Prediction of the apnea-hypopnea index from overnight pulse oximetry. Chest 124 (5):1694-1701. doi:10.1378/chest.124.5.1694