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

**Abstract**

*Background*

Sleep apnea and hypopnea syndrome (SAHS) seriously affects people’s sleep quality. In recent years, many researches have been made for the detection of SAHS using various physiological signals and algorithms. However, there are still some limitations such as low detection resolution.

*Method*

We proposed a cascading detection model for precise detection of apnea-hypopnea (AH) events based on morphological features extracted from nasal flow (NF) and arterial blood oxygen saturation (SpO2).

*Results*

For the more than 280,000 ten-second segments in the database, the cascading detection model reached an accuracy of 89.0%. As for more than 1800 AH events, it achieved a sensitivity of 82.8%, while the Pearson’s correlation coefficient between estimated apnea-hypopnea index (AHI) and reference AHI was 0.98. Besides, for 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 estimation of AHI and the result indicates 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 laboratory. However, PSG requires patients to sleep with many sensors for at least one night meanwhile a long time are required for the score of apnea-hypopnea (AH) events. Therefore, many researchers hope to simplify or replace PSG by using a limited number of physiological signals. Electrocardiogram (ECG) was firstly taken into study, McNames et al.[1] found that heart rate, S-pulse amplitude and pulse energy are correlated with SAHS. Bsoul et al. [2] cut the ECG into 60s segments and utilized Support Vector Machine (SVM) to realize real-time detection of SAHS. However ECG is also correlated with many other kinds of diseases, hence nasal flow (NF)[3-6], arterial blood oxygen saturation (SpO2) [7] , snoring[8] or a combination of these signals[9, 10] have been adopted lately. Gutierrez et al.[4] used the overall features of single-channel NF for the diagnosis of SAHS severity. B.Xie et al. [10] utilized a combination of classifiers to achieve real-time detection of SAHS based on ECG and SpO2. All the above studies can be roughly divided into two categories. One is to predict AHI based on the detection of AH events[2, 3, 5, 7, 9-11], one is to predict AHI based on the overall signal features[1, 4, 6, 8, 12, 13]. The latter cannot provide time information of each AH event, while most studies in the former[2, 7, 10, 11] are only for 60s segment identification which may lead to an error in the estimation of AHI. On the other side, the methods used above include threshold[5, 7, 9], SVM[2, 10, 11] and neural networks[3, 11] which acquire a large number of hyperparameters to be set by experience. Therefore, we utilized random forest composed of CART decision trees based on morphological features extracted from NF and SpO2 for the real-time detection of SAHS. A 60s detector and a 10s detector are cascaded to improve the resolution of detection for AH events.

**Materials and Methods**

**Subjects**

The database utilized in this paper is St. Vincent University Hospital/Dublin University College Sleep Apnea Syndrome Database (UCDDB)[14] public on Physionet[15]. The database contains 25 subjects’ PSG including electroencephalogram (EEG), electrooculogram (EOG), submental electromyography (EMG), NF, ribcage and abdomen movements, SpO2, snoring and body position. All signals were obtained by Jaeger-Toennies system. The annotation files consisted of onset time and duration of respiratory events provided by an experienced specialist. The cutoff value of AHI is commonly set to 5, 15, 30 events/h[3, 4, 7, 16, 17]. There are 2 non-SAHS subjects, 12 mild-SAHS subjects, 5 moderate-SAHS subjects and 6 severe-SAHS subjects in the database. To balance the number of subjects in each class, we randomly select 2, 4, 4, 5 subjects in each category for the following 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）

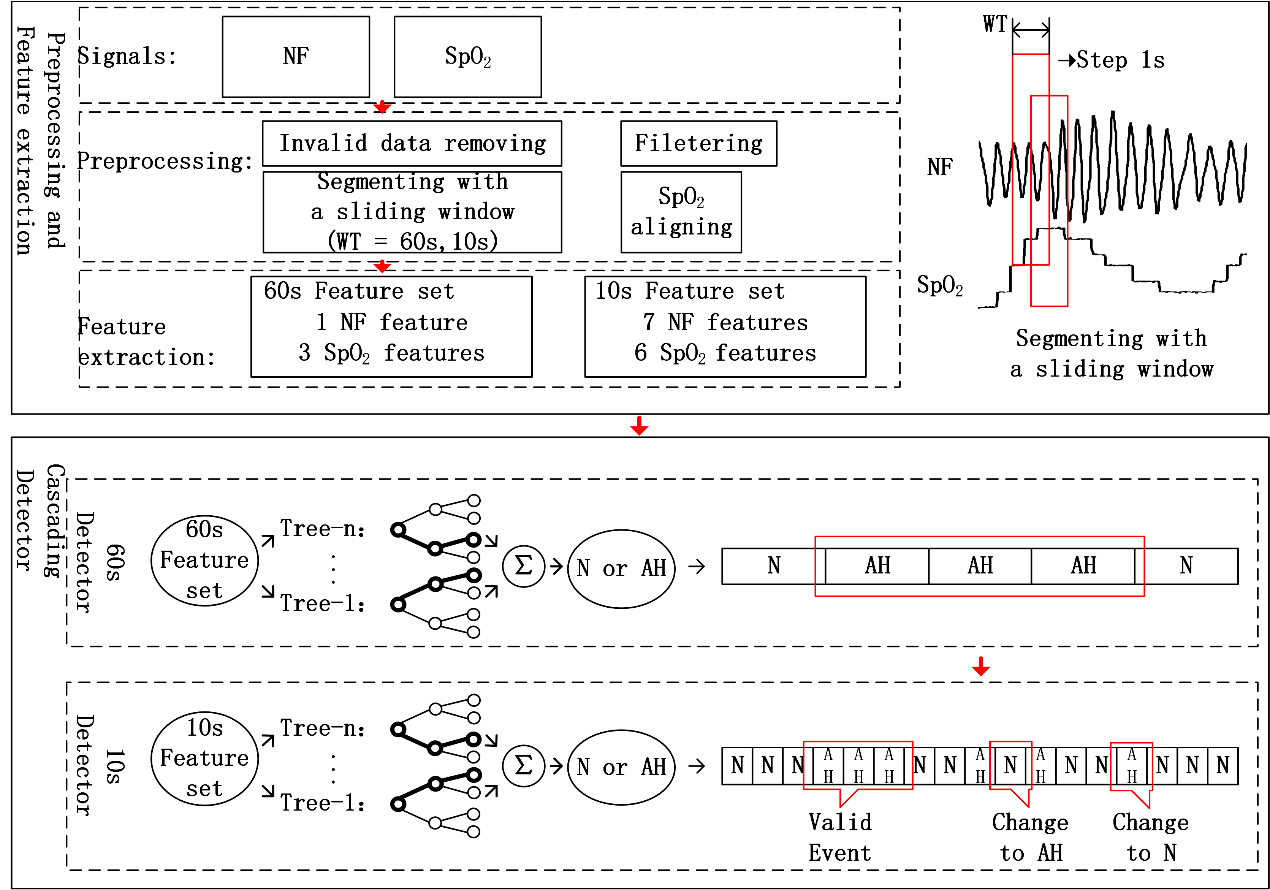
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Non-SAHS | Mild SAHS | Moderate SAHS | Severe SAHS |
| Age(years) | 52.0±15.6 | 49.0±1.6 | 57.5±7.2 | 46.6±5.5 |
| AHI(events/h) | 4.1±5.7 | 12.4±1.8 | 25.6±3.6 | 43.8±16.3 |
| Epworth Sleepiness Score | 7.0±8.5 | 14.5±3.9 | 9.3±6.2 | 12.4±7.9 |

According to American Academy of Sleep Medicine (AASM) manual[16], apnea is scored when there is a more than 90% drop in the peak signal of pre-event baseline with a duration longer than 10s in NF. Hypopnea is scored by the following rules: 1) there is a more than 30% drop in the peak signal of pre-event baseline with a duration longer than 10s in nasal pressure, accompanied by 2) more than 3% arterial oxygen desaturation or an arousal. As a result, we selected NF and SpO2 for SAHS detection.

**Study design**

The cascading detection model based on AH event detection is shown in Fig.1. It mainly includes the following four steps: 1) Invalid data removing, filtering, segmenting with a sliding window and SpO2 aligning. 2) Specific feature set extracted from each segment. 3) Cascading detection model predicts every segment and outputs a sequence of results. 4) Event detector corrects the invalid results in the sequence and calculates the AHI.

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



**Signal preprocessing**

Signal preprocessing includes the following four steps: 1) Invalid data removing. Normal people’s SpO2 tends to stay around 98%, therefore any SpO2 values lower than 80% were considered as artifacts and removed from analysis (5.8% of data). 2) Filtering. A 4-point sliding average filter and a third-order Butterworth high-pass filter with a cut-off frequency of 0.05Hz were adopted to avoid the baseline drift and high-frequency noise caused by artifacts. 3) Segmenting. The original signals were segmented by a 60s window and a 10s window respectively. The step of both windows was set to 1s. Segments were categorized into two classes: AH and N (normal). The segments containing more than five seconds’ SAHS were labeled as AH. Other cases were labeled as N. 4) SpO2 aligning. Considering SpO2 responses slowly to AH events [18], we moved SpO2 forward τ seconds (0<τ<30). It was shown in the results that the model performed best with τ set to 23s. After preprocessing, we totally obtained 35,309 AH segments and 249,977 N segments.

**Feature extraction**

**NF Feature set**

According to AASM’s definition of AH events, there is important information in NF’s amplitude. Therefore, we firstly extracted the maximum () and minimum points in NF. Then the tidal volume of per breath was calculated by the difference between adjacent two extreme points. The mean, standard deviation and range coefficients of tidal volume were extracted from each segment. Meanwhile we calculated the baseline of tidal volume every thirty seconds following equation (1).

（1）

In the equation, represents the segment. Then the number of breaths with a tidal volume reduction more than 30% (and 70% (and the number of normal breaths were calculated following the equations (2) ~ (4).

（2）

（3）

（4）

Besides, due to the cessation of breathing, there will be fluctuations in the breathing rate during AH events. While one normal breath lasts for three to five seconds and its energy will be concentrated with a peak in the corresponding frequency. As a result, we took the kurtosis in 0.2-0.4 Hz of NF’s FFT as another feature.

**SpO2 Feature set**

There are always fluctuations in SpO2 during AH events, hence we firstly calculated the standard deviation and range coefficients of SpO2 in each segment. Meanwhile the slope of SpO2 in each segment (was also calculated. At the same time the commonly used features: the time of SpO2 stays below 92% and 91% [19, 20] in each segment were calculated as another two features. Then we took the maximum ( and average SpO2 value in every thirty seconds as the baseline. Then the duration and level of oxygen desaturation were calculated in each segment following equations (5) – (8).

（5）

（6）

（7）

（8）

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

Table 2. Features and their definitions

|  |  |  |
| --- | --- | --- |
| Index | Name | Definition |
| 1 | , , | Average, standard deviation and range of tidal volume |
| 2 |  | The number of breaths with a reduction more than 30% in tidal volume and the ratio of it to total number of breaths |
| 3 |  | The number of breaths with a reduction more than 70% in tidal volume and the ratio of it to total number of breaths |
| 4 |  | The number of normal breaths and the ratio of it to total number of breaths |
| 5 |  | The kurtosis in 0.2-0.4Hz of NF’s FFT |
| 6 | , | Standard deviation and range of SpO2 |
| 7 |  | Slope of SpO2 |
| 8 |  | Duration of SpO2 desaturation |
| 9 |  | Level of SpO2 desaturation |
| 10 |  | Duration of SpO2 stays below 92% and 91% |

**Design of Cascading Detector**

The cascading detector contains two parts. The first is a random forest consisted of ten CART decision trees for the prediction of 60s segments. It can screen out most of N segments while maintain the AH segments. The second is a random forest consisted of twenty CART decision trees for the prediction of 10s segments. Based on the results of 60s detector, the 10s detector is able to locate the start and end time of AH events.

It should be noted that the 60s detector was trained using the feature set composed of feature 2, 6 and 8 which aims to improve the training speed. While the results illustrated that there was almost no effect on the performance. Besides, considering the imbalance in the number of AH and N segments, the weights for the two classes in CART trees were set to inverse ratio of their numbers.

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

**Design of Event Detector**

The sequence predicted by cascading detector will then be fed into event detector to correct the invalid results following the two rules: 1) Only longer than ten consecutive AH segments were considered to be a valid AH event. Since the original data was segmented by a 10s window while one AH event lasts at least ten second, which determines that one AH event corresponds to at least ten consecutive AH segments. Any AH segment which did not meet the above rule was modified to N. 2) The N segments between two adjacent AH segments should be longer than five. This is also determined by the way of data segmentation. Any segment does not meet the rule was reset to AH. The total number of AH sequence after the event detector was considered to be the number of AH events, while the first and last segment in each AH sequence were considered to be the start and end time of AH event.

**Results**

The cascading detection model is able to estimate AHI and provide a diagnosis of SAHS severity based on AH events detection. We analyzed the performance of it in two aspects: segments and AHI.

**Segments analysis**

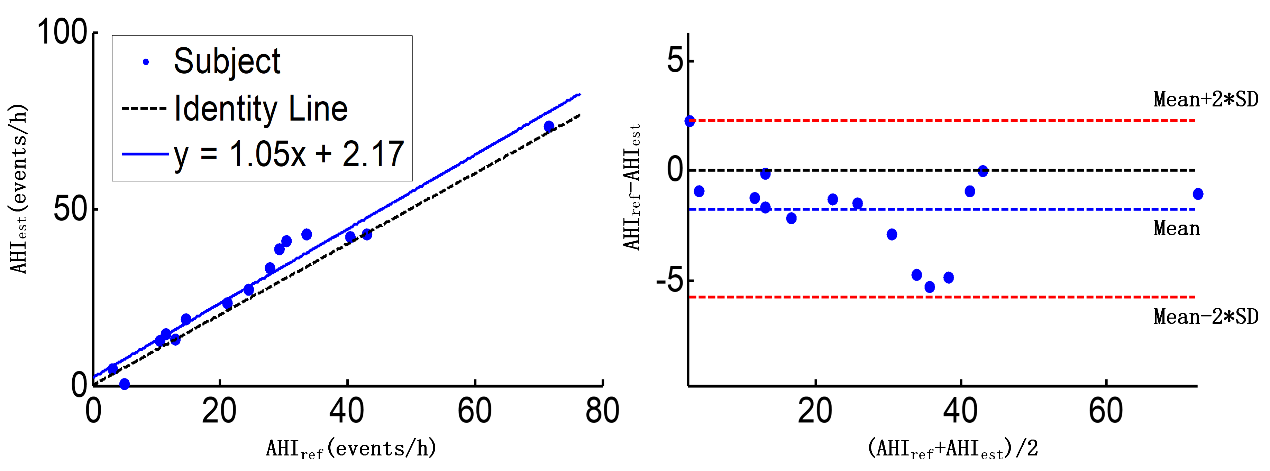
Table 3. Results for segments and events

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

The test set contains fifteen subjects’ over one hundred hours of data, totally 285,286 ten-second segments. The prediction results for the above data are shown in Table 3. The cascading detection model achieved an accuracy (ACC) of 89.0%, a sensitivity (SEN) of 73.5%, and a specificity (SPE) of 91.2%. Table 3 also summarizes the predicition results for AH events. The cascading detection model detected 1513 from 1828 AH events, achieved a sensitivity of 82.8% accompanied by a positive predictive value (PPV) of 71.8%.

**AHI analysis**

Fig. 2(a) shows a scatter plot of the AHI estimated (AHIest) by the model and the AHI marked (AHIref) by PSG. The solid line fitted shows a high correlation (Pearson’s correlation coefficient = 0.98) between AHIest and AHIref. Fig. 2(b) shows the Bland-Altman plot of AHIest and AHIref. The average error of AHIest and AHIref is -1.7 events/h, and the error range is -5.7 to 2.3 events/h with a 95% confidential coefficient.



(a)

(b)

Figure 2. (a) Scatter plot of AHIest and AHIref; (b) Bland-Altman plot of AHIest and AHIref

Table 4 summarizes the classification results for SAHS severity. The mean values for sensitivity, specificity, PPV, accuracy are 100.0%, 87.8%, 87.1%, 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 |

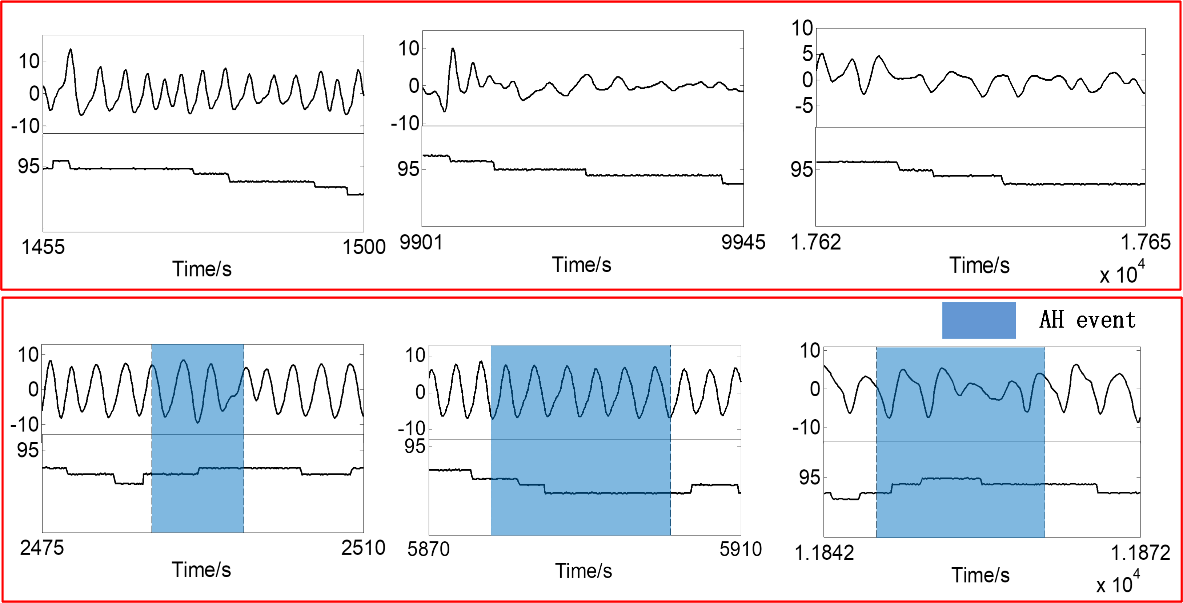
**Discussion**

This paper proposed a cascading detection model that can predict AHI based on AH events detection. Compared with PSG, only NF and SpO2 were adopted. In the past, the original signals were commonly cut into 60s segments for AH event detection[2, 7, 10, 11], however there are two limitations. One is the detection resolution is not enough. The detection by non-overlapping window can only determine whether there is AH in the segment but cannot predict the start and end time of the event. Meanwhile in some cases it cannot distinguish two AH events with a short interval which may lead to an error in AHI estimation. Therefore, some researchers[7, 9] cut the signals into shorter segments for detection. But it is difficult to extract effective features from a shorter than 10s segment because there are no more than five complete breaths in one segment in most cases. As a result, we proposed a cascading detection model composed of 60s detector and 10s detector to predict AH events precisely. Table 3 shows the classification results of segments. It should be noted that the model tended to make false positive errors. In these errors around 10.3% actually met the rules recommended by AASM in 2012 (Fig. 3(a)) but the annotations in the database did not annotate. As for the false negative errors, approximately 85.2% actually did not meet the rules in AASM’s manual (Fig. 3(b)). These segments might be annotated according to signals from other channels like ribcage or abdomen movements. The cascading detection model achieved a sensitivity of 94.9% and a specificity of 92.1% without these controversial segments.

Figure 3. (a)False positives in the prediction results. (b)False negatives in the prediction results.

(a)

(b)



It is illustrated in Fig. 2 that the AHIest shows a high correlation with AHIref (Pearson correlation coefficient = 0.98). Meanwhile the model performed a good consistency among different subjects. On the other side the AHIest is slightly higher than AHIref. Consequently, there were three subjects’ SAHS severity overestimated, as for the rest twelve 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, which means this method can be used as a powerful tool for screening SAHS.

We also tested the speed of the cascading detection model. It cost 13.9s to get trained while only 9.1s to provide the results of all segments and all fifteen subject’s AHI and SAHS severity prediction. 37ms were cost for predicting one segment and 0.6s for diagnosis of one subject on average. Which implies that the model can be applied for real-time SAHS detection.

Compared with other studies, it can be seen that our method illustrates a good sensitivity but not very good specificity. This is where we want to improve in the future. More importantly, the model not only can predict the severity of SAHS but also can provide the time information for each AH event. Meanwhile compared with other methods such as convolutional neural networks, a smaller number of hyperparameters and less computation are required in random forest. And the CART trees in the random forest can provide a better interpretability in clinical detection.

Table 5. Comparison with other studies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Related work | Method | Signal | AHI cutoff | ACC(%) | SEN(%) | SPE(%) |
| San Ho 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 C 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 SpO2 | 5 | 100.0 | 100.0 | 100.0 |
| 15 | 93.3 | 100.0 | 83.3 |
| 30 | 86.7 | 100.0 | 80.0 |

At the same time, there are still some limitations in our research. First, we have no further classified AH events into apnea events and hypopnea events. In addition, the model has not been tested in an online environment. We hope to confirm the usability of our method online in the future.

**Conclusion**

The purpose of this study is to propose a model for real-time detection of SAHS. Based on the morphological features of NF and SpO2, the cascade of 60s detector and 10s detector can not only predict the AHI and SAHS severity, but can also provide the time information of each AH event. Compared with previous research, the cascading detection model based on random forest is able to provide better interpretation and effectively reduces the computational complexity. Therefore, it is expected to be an effective tool for SAHS diagnosis.