1. Introduction

Sleep apnea syndrome (SAS) is a sleep disorder characterized by breathing pauses or breathing volume reduction during sleep. Apnea is defined as a cessation of breathing for at least 10 s, and hypopnea is defined as more than 50% reduction in the amplitude of nasal pressure or respiratory effort associated with a more than 3% reduction in oxyhemoglobin saturation for at least 10 s. SAS is classified into three types: obstructive (OSA), central (CSA), and a combination of these two. Patients with apnea are likely to develop lifestyle-related diseases such as hypertension and type II diabetes (Peppard et al. 2000, Reichmuth et al. 2005). SAS is also recognized as a risk factor for cardiovascular diseases, motor vehicle collisions, and deterioration in daytime work efficiency (Sassani et al. 2004, Monahan and Redline 2011).

Objective: Obstructive sleep apnea (OSA) is a common sleep disorder; however, most patients are undiagnosed and untreated because it is difficult for patients themselves to notice OSA in daily living. Polysomnography (PSG), which is the gold standard test for sleep disorder diagnosis, cannot be performed in many hospitals. This fact motivates us to develop a simple system for screening OSA at home. Approach: The autonomic nervous system changes during apnea, and such changes affect heart rate variability (HRV). This work develops a new apnea screening method based on HRV analysis and machine learning technologies. An apnea/normal respiration (A/N) discriminant model is built for respiration condition estimation for every heart rate measurement, and an apnea/sleep ratio is introduced for final diagnosis. A random forest is adopted for the A/N discriminant model construction, which is trained with the PhysioNet apnea-ECG database. Main results: The screening performance of the proposed method was evaluated by applying it to clinical PSG data. Sensitivity and specificity achieved 76% and 92%, respectively, which are comparable to existing portable sleep monitoring devices used in sleep laboratories. Significance: Since the proposed OSA screening method can be used more easily than existing devices, it will contribute to OSA treatment.
Polysomnography (PSG) is the gold standard test for sleep disorder diagnosis (Kapur et al. 2017), and it records multiple channels, e.g. electroencephalograms (EEG), electrocardiograms (ECG), electrooculograms (EOG), electromyograms (EMG), airflow, and oxygen saturation simultaneously during sleep. Although PSG can definitively diagnose apnea, only a limited number of hospitals can perform PSG due to economic constraints. Portable sleep monitoring devices are suitable for apnea diagnosis from the viewpoint of medical resource conservation; however, they are expensive and require operation skills (Chesson et al. 2003). Usually, type IV devices monitor one or two channels, such as oxygen saturation or airflow (Flemons et al. 2003). Their screening performance is not always high (Kadotani et al. 2011, Matsuo et al. 2016) since they measure only a few channels. In addition, a guideline for OSA diagnosis has reported that sleep apnea-focused clinical prediction rules and questionnaires lack sufficient diagnostic accuracy (Kapur et al. 2017).

It is desirable to develop a high-performance apnea screening system so that potential patients with sleep apnea, who have not been treated, can use the system easily at home and have an opportunity for treatment. In particular, this study focuses on screening OSA because it is the most prevalent apnea.

The effect of sleep apnea on heart rate variability (HRV), which is the fluctuation of the RR interval (RRI) in an ECG, has been reported (Somers et al. 1995, Bauer et al. 1996, Narkiewicz et al. 1998, Dingli et al. 2003, Kufoy et al. 2012, Aeschbacher et al. 2016, Gong et al. 2016). HRV is widely recognized as a non-invasive method for quantifying activities of the autonomic nervous system (ANS). Since it is well known that apnea affects ANS, and such changes in ANS also alter HRV during apnea (Gula et al. 2003, Kufoy et al. 2012, Taranto Montemurro et al. 2014), patients with apnea may be screened by monitoring HRV during sleep.

We propose an HRV-based algorithm for realizing a precise apnea screening system, which monitors the standard HRV features during sleep as measured by a wearable heart rate sensor and discriminates between apnea and normal respiration during sleep. An apnea/normal respiration (A/N) discriminant model is used for respiration condition estimation, which is trained with HRV data both from patients with apnea and from healthy persons during sleep. In order to decide whether users have OSA or not, we define an apnea/sleep (AS) ratio that is calculated from the estimated respiration condition. Although any binary classification method can be used, a random forest (RF) (Ho 1998, Breiman 2001) is adopted for A/N model construction in this work. In the proposed apnea screening method, HRV features which can be extracted from the RRI data are used as input information. Thus, the proposed method is simple and easy to use even at home because the developed algorithm uses only RRI data measured by the wearable heart rate sensor and avoids input of additional information such as subject age, sex, and other profiles.

We used the PhysioNet apnea-ECG database (Goldberger et al. 2000, Penzel et al. 2000) for algorithm construction, and original clinical PSG data for validation. This study aims to show the possibility of future clinical applications of the proposed apnea screening algorithm. Although a preliminary version of this work has already been reported in Nakayama et al. (2015), only small amounts of clinical data were analyzed therein (Nakayama et al. 2015), and other modeling methods were not tried. We have collected a large amount of clinical data and discuss the proposed method based on the clinical data collected.

2. Methods

Figure 1 shows a schematic diagram of the proposed apnea screening method. The RRI data during sleep are recorded using a wearable heart rate sensor, which is stored in a storage medium. After the user awakes, the proposed method extracts HRV features from the stored RRI data and analyzes them to diagnose whether the person is a ‘potential patient’ or a ‘healthy person’.

The algorithm consists of four parts: (1) HRV feature extraction from the RRI data; (2) pre-processing of the extracted HRV features; (3) apnea or normal respiration discrimination; and (4) apnea diagnosis based on the respiration discrimination results. In this work, a total of 11 short-term HRV features are used for apnea screening: six time-domain HRV features (meanNN, SDNN, RMSSD, Total Power, NN50, and pNN50) and five frequency-domain features (LF, HF, LF/HF, LFnu, and HFnu), which can be extracted from 2–3 min RRI data (Malik et al. 1996). Although long-term HRV features such as very low-frequency or entropy features are described in the HRV analysis guideline (Malik et al. 1996, Richman and Moorman 2000), they usually require 5–10 min RRI data or 24 h RRI data for calculation. This work does not adopt these long-term HRV features because they may miss sleep-related events occurring in a short time. The HRV features adopted in this study are explained in the appendix.

In the following sections, the nth sampling of HRV features can be denoted by $x_n = [x_{n,1}, x_{n,2}, \cdots, x_{n,M}]^T$, where $x_{n,m}$ is the nth sample of any mth HRV feature, where $M = 11$ in this work. In addition, $X \in \mathbb{R}^{N \times M}$ is a matrix whose nth row is $x_n^T$. 
2.1. A/N discriminant model

In the proposed apnea screening method, the A/N discriminant model is constructed using a machine learning technique. Although any classification method can be utilized for modeling, RF (Ho 1998, Breiman 2001) is adopted. RF is a well-known ensemble learning technique that constructs multiple decision trees as weak classifiers, which can work fast and avoid overfitting to a learning dataset. In order to build a good classification model in RF, the number of trees $T$ needs to be appropriately tuned as a parameter.

Algorithm 1 is adopted for the A/N discriminant model construction. In this procedure, $z^{(i)} = [z_1^{(i)}, \ldots, z_t^{(i)}, \ldots]_T (i = 1, \ldots, I)$ are the RRI data collected during sleep from the $i$th subject (a patient or a healthy person), $t$ denotes the number of RRI, and $I$ is the total number of subjects used for modeling. In addition, each RRI $z_t^{(i)}$ must be labeled as apnea $A$ or normal respiration $N$ using PSG records. A respiration label vector of the $i$th subject $y^{(i)}$ is defined as $y^{(i)} = [y_1^{(i)}, \ldots, y_t^{(i)}, \ldots]_T$ where $y_t^{(i)} = \{A, N\}$.

Eleven HRV features are extracted from $z^{(i)}$ in step 3. The extracted HRV features of the $i$th subject are arranged as $X^{(i)}$. In step 5, $X^{(i)} (i = 1, \ldots, I)$ are merged lengthwise into one matrix $X$ as follows:

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(I)} \end{bmatrix}. \quad (1)$$

The merged matrix $X$ is normalized with zero mean and a standard deviation of one, which is referred to as $\tilde{X}$, in step 6. Respiration label vectors $y^{(1)}, \ldots, y^{(I)}$ are merged into one vector $y$ in step 7 as

$$y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(I)} \end{bmatrix}. \quad (2)$$

Finally, using $X$ and $y$ as modeling data, the A/N discriminant model $h$ is built by RF. At this time, the number of trees $T$ used should be tuned, for which cross-validation can be adopted.

In step 6, the merged HRV matrix $X$ is normalized because the value ranges of the HRV features are significantly different from each other. In order to make the model training easier, the standard deviations of the input variables should be uniform.

**Algorithm 1. A/N discriminant model construction.**

1: for all $i$ such that $1 \leq i \leq I$ do
2: Collect RRI data $z^{(i)}$ and respiration label data $y^{(i)}$.
3: Derive the $i$th HRV feature data $X^{(i)}$ from $z^{(i)}$.
4: end for
5: Merge matrices $X^{(1)}, \ldots, X^{(I)}$ into one matrix $X$.
6: Standardize $X$ to $\tilde{X}$.
7: Merge vectors $y^{(1)}, \ldots, y^{(I)}$ into one vector $y$.
8: Construct the A/N discriminant model $f$ from $\{X, y\}$ using RF.

2.2. Sleep apnea screening

The A/N discriminant model classifies the respiration condition of every normalized HRV sample into normal or apnea. The A/N discriminant model is written as $\hat{y}_n = h(\tilde{x}_n)$ where $h(\cdot)$ is a function, $\tilde{x}_n$ is the $n$th normalized HRV sample, and $\hat{y}_n$ is its corresponding estimated respiration condition: $\hat{y}_n = \{A, N\}$.
Although the A/N discriminant model $h$ estimates respiration condition $\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_n, \cdots$ for every HRV sample $x_n$, each estimate does not always reflect the real respiration condition directly. The effect of apnea on ANS activities starts several minutes before an apnea onset and remains for several minutes after apnea stops (Vanninen et al 1996), and HRV features are affected by such ANS activation. Thus, the proposed method combines these estimates into one index for apnea screening: the AS ratio $A$

$$A = 100 \times \frac{T_a}{T_s} \%$$

(3)

where $T_s$ is the total sleep time, and $T_a$ denotes the sum of apnea duration calculated from the estimates of the A/N discriminant model $h$. That is, $T_a$ increases when the HRV samples are classified into $A$ by $h$. A user is diagnosed as a ‘potential patient’ when the calculated AS ratio exceeds the predefined threshold $A$, and otherwise as a ‘healthy person’.

The apnea screening procedure is described in algorithm 2. The RRI data $z \in \mathbb{R}^{N+1}$ are collected during sleep and stopped when the subject wakes in steps 2 and 3. The total sleep time $T_s$ is calculated as $T_s = \sum z_0$ in step 4. The HRV samples $x_1, x_2, \cdots, x_N$ are extracted from the collected RRI data $z$ in step 5, and are normalized to $\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_N$ in step 6. $N$ is the number of extracted HRV samples from $z$. $N_{TV}$ denotes the number of RRI data points required for the initial HRV feature extraction, that is, $[z_{TV+1}, z_{TV+2}, \cdots, z_{TV+2}, z_{TV}]$ are used for extraction of the first HRV sample $x_1$.

Steps 7 and 8 classify the respiration condition of the $n$th sample $x_n$ into apnea $A$ or normal respiration $N$ by using the A/N discrimination model $h$. When $x_n$ is classified into $A$, the sum of apnea duration $T_a$ increases to $T_a = T_a + z_n$ where $z_n$ is the $n$th RRI corresponding to $x_n$ in steps 9–12. The AS ratio is calculated for final diagnosis in steps 14–19.

In order to calculate $T_s$ in step 4, sleep onset has to be detected, and HRV-based sleep stage estimation methods can be utilized for sleep onset detection (Xiao et al 2013, Takeda et al 2015).

**Algorithm 2. Apnea screening.**

1: Set $T_a \leftarrow 0$.
2: Start collection of RRI data $z$ when asleep.
3: Stop RRI data collection when awake.
4: Calculate $T_s = \sum z_0$.
5: Derive the HRV data $x_n (n = 1, 2, \cdots, N)$ from $z$.
6: Standardize $x_n$ to $\tilde{x}_n$.
7: for all $n$ such that $1 \leq n \leq N$ do
8: Classify respiration condition into $A$ or $N$ from $\tilde{x}_n$ by $h$.
9: if $\tilde{x}_n \in A$ then
10: Extract the $n$th RRI $z_n$ from $z$.
11: $T_a = T_a + z_n$.
12: end if
13: end for
14: Calculate the AS ratio $A = 100 \times \frac{T_a}{T_s}$.
15: if $A \geq A$ then
16: Output ‘potential patient with apnea’.
17: else
18: Output ‘healthy’.
19: end if

3. Results and discussion

This section provides a report on the results of applying the proposed apnea screening method to the clinical data and a discussion thereof.

3.1. Data description

The present work used the PhysioNet apnea-ECG database (https://physioNet.org/content/apnea-ecg/1.0.0/) (Penzel et al 2000, Goldberger et al 2000) to construct the OSA screening algorithm. The A/N discriminant model was constructed from a learning dataset of the PhysioNet data, and the optimal threshold of AS ratio $A$ was tuned based on a test dataset of the PhysioNet data. In order to evaluate an exact apnea screening performance of the proposed method, it was validated by using a different, original set of clinical PSG data collected at the Shiga University of Medical Science (SUMS) hospital.
3.1.1. PhysioNet data
The PhysioNet apnea-ECG database consists of a learning dataset of 35 records (a01–a20, b01–b05, and c01–c10) and a test set of 35 records (x01–x35), and each record contains an ECG signal during sleep, a profile, and apnea annotations labeled by experts. The ECG lead in the PhysioNet data is unknown. Although some records include additional signals like oronasal airflow, chest and abdominal wall movements for respiratory efforts, and oxygen saturation, we used only ECG signals and apnea annotations. In this analysis, the records with AHI < 15 were regarded as healthy records, while the records with AHI ⩾ 15 were regarded as apnea records.

The records where R waves could not be detected appropriately by a first derivative-based peak detection algorithm due to strong artifacts or arrhythmia were eliminated from the analysis. The numbers of records retained in the learning set and the test set were 25, respectively. Supplemental tables S.1 and S.2 show their profiles (stacks.iop.org/PM/40/125001/mmedia).

The ECG signals were clipped from each record, and the R waves in the ECG signal were detected using a first derivative-based peak detection algorithm, and each RRI was calculated. Finally, 11 HRV features were extracted.

3.1.2. SUMS data
The PSG data of patients with apnea and healthy persons during sleep were collected at the SUMS hospital. The Research Ethics Committee of the SUMS hospital approved data collection and analysis. Written informed consent was obtained from each participant who was involved in the prospective evaluation.

A video, EEG, ECG, EOG, EMG, oronasal airflow, thermistor, chest and abdominal wall movements for respiratory efforts, and oxygen saturation data of the participants were simultaneously recorded during sleep for about 6–7 h using a PSG system (Alice 5, Philips) with a sampling frequency of 200 Hz. The ECG lead was the standard lead II, and measurement items of the SUMS data were compatible with those of the PhysioNet data. These tests were conducted in an EEG recording shield room with a technician in attendance. A sleep specialist certified by the Japanese Society of Sleep Research annotated the PSG data and evaluated AHI. Participants with AHI < 15 were regarded as healthy persons, while participants with AHI ⩾ 15 were regarded as patients. ECG data where R waves could not be detected appropriately by a first derivative-based peak detection algorithm due to strong artifacts or arrhythmia were eliminated. The clinical SUMS data consisted of 25 patients with apnea (P1–P25) and 36 healthy persons (H1–H36) whose profiles are listed in supplemental table S.3.

The RRI data obtained from healthy person H10 (male, 19 y.o., AHI = 0.2) and patient P7 (male, 46 y.o., AHI = 21.9) are shown in figure 2, in which colored bands denote apnea or hypopnea periods. Figures 3 and 4 show parts of the obtained HRV features extracted from H10 and P7. The HRV features of the patient recorded during apnea periods were larger than those during normal respiration periods. However, HRV features of healthy person H10 also fluctuated regardless of their respiration condition. Some of these fluctuations might occur in association with micro-arousal or other sleep-related events. His arousal index was 8.6 events per hour. Researchers have previously reported that sleep stage transition and micro-arousal affect HRV (Sforza et al 2000, Gosselin et al 2002).

These results indicate that it is difficult to screen apnea by analyzing an individual HRV feature and that multiple features should be monitored together.

3.2. A/N discriminant model construction
The A/N discriminant model was constructed from the learning set in the PhysioNet data by following algorithm 1. This study used a classification and regression tree (CART) as a decision tree in RF, and the number of CARTs used in RF was $T = 30$, which was determined by cross-validation.

The constructed A/N discriminant model was applied to all records in the test set of the PhysioNet data by following algorithm 2. In the proposed algorithm, the threshold of the AS ratio, $\bar{A}$, has to be determined. Figure 5 shows the receiver operating characteristic (ROC) curve of the test set in the PhysioNet data, and the area under the ROC curve (AUC) was 0.91. The threshold of the AS ratio $\bar{A}$ was determined based on the Youden index of the ROC curve (Youden 1950) drawn by the test set, and $\bar{A} = 37\%$.

3.3. Application to SUMS data
The apnea screening algorithm was applied to the SUMS data. We used sleep onset and wakening times labeled in the PSG data. The ROC curve was illustrated by the clinical data in figure 5, and its AUC was 0.84. Because the two ROC curves for the test set of the PhysioNet data and the SUMS data were close to each other, it was concluded that the constructed algorithm avoided overfitting, which might be by virtue of RF.

The final apnea screening results of the clinical data are shown in figure 6. The A/N discriminant model identified 19 out of 25 patients as ‘potential patients’, and 33 out of 36 healthy persons as ‘healthy’. Therefore, the sensitivity and specificity of the proposed method were 76% and 92%, respectively.
3.4. Discussion
The developed apnea screening method correctly identified 19 out of 25 patients as ‘potential patients’, and 33 out of 36 healthy persons as ‘healthy’. Figure 7 shows the A/N discrimination results of HRV data in figures 3 and 4. Vertical colored bands denote apnea periods, and horizontal lines are the respiration conditions discriminated by the A/N discriminant model. The discriminated respiration conditions of patient P7 still indicated apnea even after recovery from apnea, which suggests that the HRV of patients with apnea differs from that of healthy persons even when the patients are breathing normally. Vanninen et al. reported that the effect of apnea on ANS activities starts several minutes before an apnea onset and remains for several minutes after apnea recovery (Vanninen et al. 1996). Our analysis results are consistent with their findings.

Our apnea screening method did not correctly diagnose six patients: P2, P5, P12, P14, P15, and P21; and three healthy persons: H18, H21, and H27. According to the PSG records of patients P2 (female, 66 y.o., AHI = 15.3) and P5 (male, 69 y.o., AHI = 19.9), their apnea was not severe. A patient with AHI ≥ 30 is significantly associated with cardiovascular disease and type II diabetes (Qaseem et al. 2014). Thus, their diagnostic errors were not fatal from the viewpoint of prognosis. Patient P14 (male, 28 y.o., AHI = 30.8) was the youngest patient. On the other hand, all of the patients in the learning set were more than 30 years old, according to table S.A1. Since HRV differs according to age group (Laitinen et al. 1998), the characteristics of the HRV data of patient P14 might be different from the learning data. We could not specify the causes of incorrect diagnosis of patients P12 (male, 42 y.o., AHI = 30), P15 (male, 58 y.o., AHI = 31.7), and P21 (male, 42 y.o., AHI = 49.4).
Figure 4. HRV features extracted from patient P7.

Figure 5. ROC curves illustrated by the test set in PhysioNet data and clinical data.

Figure 6. AS ratio of clinical data.
Healthy persons H18 (male, 42 y.o., AHI = 0.7), H21 (male, 38 y.o., AHI = 0.9), and H27 (female, 23 y.o., AHI = 1.8) did not have cardiovascular disease and arrhythmia that may affect HRV; however, the body mass index of H21 was 28.7, and his arousal index was 27 events per hour. Since arousal significantly affects HRV (Sforza et al 2000, Gosselin et al 2002), his HRV might be affected by frequent arousal. On the other hand, the arousal indexes of H18 and H27 were fewer than ten events per hour. Thus, arousal may cause incorrect diagnosis in HRV-based apnea screening.

We had another set of clinical data of a patient (female, 23 y.o., AHI = 75.8) with CSA, which was applied to the proposed screening algorithm. Her AS ratio was 0.25, so she was diagnosed as a healthy person. It is reported that the HRV of patients with CSA during apnea is different from that of patients with OSA (Szollosi et al 2007), and there were no HRV data of CSA patients in the learning dataset. That is, her data were completely different from the learning data, and the proposed apnea screening method did not function.

In order to screen patients with severe OSA (AHI $\geq 30$), we defined an extra threshold of the AS ratio based on the Youden index calculated from the test set of the PhysioNet data, and $\bar{A}' = 58\%$. The extra threshold identified nine out of 13 severe patients as ‘potential severe patients’, and 41 out of 48 moderate patients and healthy persons as ‘healthy or potential moderate patients’ in the SUMS data. The sensitivity and specificity of the proposed method were 70% and 85%, respectively. Thus, our methodologies can also screen severe OSA by tuning the threshold of the AS ratio.

To confirm the reliability of the proposed method, we tried to construct the A/N discriminant model from the SUMS data and to define the threshold of the AS ratio based on the learning dataset in the PhysioNet data in the same manner as described in section 2. The application results of the SUMS-based A/N discriminant model to the test dataset in the PhysioNet data showed that 13 out of 19 patients and seven out of nine healthy persons were classified as ‘potential patients’ and ‘healthy’, respectively. That is, the sensitivity and the specificity were 68% and 78%, respectively. Although the screening performance of the SUMS-based model was slightly worse than that of the PhysioNet-based model, it may be improved through further parameter tuning. This result suggested that our apne screening methodology is reliable.

Although input variables used in the proposed apnea screening algorithm should be limited to the HRV features described in the guideline (Malik et al 1996) from the viewpoint of future medical device approval, we tried additional HRV features that are not described in the guideline in order to improve the OSA screening performance of the proposed method. Poincaré plot-based HRV features (Kamen et al 1996, Hoshi et al 2013), SD1, SD2, SD1/SD2, were used in addition to 11 HRV features in the apnea screening algorithm. However, the screening performance did not change; the sensitivity and the specificity were 71% and 95%, respectively. This result indicates that the Poincaré plot-based HRV features do not contribute to apnea screening.

Since any binary classification algorithm can be used for A/N discriminant model construction, the other well-known machine learning techniques of linear discriminant analysis (LDA) and support vector machine (SVM), were tried for comparison. LDA constructs a linear discriminant axis that determines to which group a sample belongs so that the between-group variance is maximized and the within-group variance is minimized, simultaneously (McLachlan 1992). SVM is a nonlinear classification technique, which was originally developed for classifying data into two classes (Cristianini and Shawe-Taylor 2000). When the modeling data consist of
two classes, SVM constructs an optimal separation hyperplane which has the maximum margin. The margin is defined as the distance between the separation hyperplane and its closest sample.

We constructed A/N discriminant models using LDA and SVM instead of RF and determined the thresholds of the AS ratio by following the same procedure as in algorithm 1. In SVM, the Gaussian kernel with a parameter of $\sigma = 3$ was used.

The sensitivity and the specificity of SVM were 73% and 82%, and those of LDA were 50% and 18%, respectively. RF achieved better performance than SVM; however, further performance improvement of SVM may be achieved when appropriate kernels and tuning parameters are found. The performance of LDA was the lowest. Since LDA is a linear method, it may be difficult to model a complicated phenomenon like the relationship between apnea and HRV. These results show that RF is appropriate for the A/N discriminant model construction.

Alvarez-Estevez and Moret-Bonillo have reported that frequency-domain features of HRV are useful for apnea screening (Alvarez-Estevez and Moret-Bonillo 2016). Lado et al also developed an HRV-based OSA screening algorithm (Lado et al 2011); however, the screening performance of their method was not always high because their algorithm uses only the average value of a specific HRV feature during sleep. The use of the QRS complex area (integration of the area under the peak of an R wave in the interval between a Q wave and an S wave) because peaks of R waves are high and sharp. The use of HRV for apnea screening is much easier than methods that use ECG signals directly.

Various types of apnea screening algorithms utilizing ECG directly have been investigated instead of HRV (Penzel et al 2002). For example, ECG-derived respiration (EDR) components have been used (Heneghan et al 2008, Varon et al 2015, Song et al 2016, Jung et al 2017), and Varon et al reported that the sensitivity and the specificity of their EDR-based apnea screening algorithm were both 84% (Varon et al 2015). Since EDR analysis requires a highly accurate ECG measurement, subjects are required to put ECG electrodes in appropriate positions by themselves when they use EDR-based methods at home; however, it is difficult for untrained persons to place electrodes appropriately. Moreover, good ECG signals are not always obtained due to body motion during sleep or electrode contact failure. It is easy to detect R waves stably even when motion artifacts are contaminated because peaks of R waves are high and sharp. The use of HRV for apnea screening is much easier than methods that use ECG signals directly.

Various HRV-based health monitoring methods have been developed for applications other than sleep apnea screening (Kleiger et al 1987, Malliani 1991). Since excessive neuronal activities before epileptic seizures significantly affect the ANS, epileptic seizures can be predicted by monitoring HRV (Fujiwara et al 2016). Also, an HRV-based drowsy driving detection method has been proposed by utilizing the same framework as epileptic seizure prediction (Abe et al 2016, Fujiwara et al 2019). These various applications of HRV analysis show its usefulness for health monitoring, particularly for physiological changes related to the ANS. The proposed apnea screening method can be easily implemented in mobile computers such as a smartphone because the computational load is much lighter than methods that need to process ECG signals directly. An HRV-based epileptic seizure prediction smartphone app has already been developed and tested in hospitals (Fujiwara et al 2016). These are the advantages of the use of HRV analysis for medical device development.

An RRI measurement device is needed to realize an HRV-based medical device that can be used in daily living. Many types of wearable devices such as smartwatches have a photoplethysmogram (PPG) sensor for pulse detection; however, it is notably difficult for PPG to derive RRI sufficiently precisely to carry out HRV analysis because blood flow changes significantly with body motion (Lu and Yang 2009). Yamakawa et al developed a wearable heart rate sensor that can easily measure precise RRI based on ECG, and that can be manufactured for less than 100 US dollars (Yamakawa et al 2014). Tsukada et al developed a new wearable textile electrode using a conductive fiber (Tsukada et al 2012), and a smart shirt woven with textile electrodes has been developed for ECG measurement. Therefore, it will be easy to measure RRI when the smart shirt becomes available. If an HRV-based apnea screening algorithm can be implemented in such devices, an economic apnea screening system becomes available.

It is concluded that the proposed HRV-based sleep apnea screening method is more promising than conventional portable sleep monitoring devices and other HRV-based screening methods with respect to practical home use.

The limitations of this study include the limited number of patients and healthy persons in the clinical data. In particular, all subjects in the clinical data were Japanese. However, our analysis results show that the proposed method functioned well with both the PhysioNet data and the clinical data which were collected under different conditions. Accordingly, the proposed method can deal with other populations as well as Japanese people.
4. Conclusions

The present work developed a new simple sleep apnea screening method by utilizing HRV analysis and machine learning. In the proposed algorithm, RF constructs an A/N discriminant model from HRV data measured both in patients and in healthy people during sleep, which classifies each RRI collection as apnea or normal respiration. The AS ratio, which is the ratio of the sum of the apnea period estimated by the A/N discriminant model to the total sleep time, is used for diagnosing potential OSA. The results of applying the constructed A/N discriminant model to the clinical data showed a sensitivity of 76% and a specificity of 92%. The results demonstrated that the developed method functioned well. Thus, the developed method has the potential to improve quality of life and prevent the future development of lifestyle-related diseases because users are provided with opportunities for OSA treatment.

We are presently developing an apnea screening system by combining the proposed method implemented in a smartphone app and a wearable heart rate sensor. In this system, the wearable sensor measures RRI data of users and sends them to the smartphone wirelessly. The app classifies a user as a ‘potential patient with apnea’ or a ‘healthy person’ and notifies the user of the result.

In future works, additional clinical data will be collected to improve the apnea screening performance, and the system under development will be tested in hospitals. The long-term HRV features that were not adopted in this work should be evaluated from the viewpoint of OSA screening. In addition, we will try to detect other sleep-related events like leg movement or micro-arousal by combining HRV and machine learning.

Acknowledgments

This work was supported in part by JST PRESTO #JPMJPR1859 and JSPS KAHENHI #17H00872.

Appendix. HRV analysis

This appendix introduces the HRV analysis. The highest peak in an ECG is called the R wave, and the RRI (ms) is defined as the interval between an R wave and the next R wave. HRV is defined as the fluctuation of RRI, which reflects ANS activities. HRV features are classified into time-domain features and frequency-domain features (Malik et al 1996). The HRV features adopted for SAS screening are as follows:

- **Time-Domain Features**
  - meanNN: Mean of RRI.
  - SDNN: Standard deviation of RRI.
  - RMSSD: Root mean square of the difference of adjacent RRI.
  - Total Power: Variance of RRI.
  - NN50: The number of pairs of adjacent RRI whose difference is more than 50 ms within a given length of measurement time.
  - pNN50: The number of pairs of adjacent RRI whose difference is more than 50 ms, divided by the total number of RRI.

- **Frequency-Domain Features**
  - LF: Power of the low-frequency band (0.04 Hz–0.15 Hz) in a power spectrum. LF reflects the activities of both the sympathetic and parasympathetic nervous systems.
  - HF: Power of the high-frequency band (0.15 Hz–0.4 Hz) in a power spectrum. HF reflects the parasympathetic nervous system activity.
  - LF/HF: Ratio of LF to HF. LF/HF expresses the balance between the sympathetic nervous system activity and the parasympathetic nervous system activity.
  - LFnu: Normalized LF defined as LF/(LF+ HF).
  - HFnu: Normalized HF defined as HF/(LF+ HF).

Although time-domain features can be calculated directly from the RRI data, frequency-domain features are defined based on the power spectrum density of the resampled RRI data.

This work uses a rectangular moving window whose window size is 3 min. The time-domain features are calculated directly from the raw RRI data. In frequency-domain feature extraction, the RRI data are resampled so
that their sampling points are arranged at equal intervals, which are interpolated by the third-order spline, and 4 Hz resampling is adopted. An autoregressive model of order 40 was used to calculate frequency-domain features.

**ORCID IDs**

- Koichi Fujiwara [https://orcid.org/0000-0002-2929-0561](https://orcid.org/0000-0002-2929-0561)
- Manabu Kano [https://orcid.org/0000-0002-2325-1043](https://orcid.org/0000-0002-2325-1043)
- Hiroshi Kadotani [https://orcid.org/0000-0001-7474-3315](https://orcid.org/0000-0001-7474-3315)

**References**

Abe E et al 2016 Development of drowsiness detection method by integrating heart rate variability analysis and multivariate statistical process control SICE J. Control Meas. Syst. Integr. 9 10–7

Aeschbacher S et al 2016 Heart rate variability and sleep-related breathing disorders in the general population Ann. J. Cardiol. 118 912–7

Alvaraz-Estevéz D and Moret-Bonillo V 2016 Spectral heart rate variability analysis using the heart timing signal for the screening of the sleep apnea–hypopnea syndrome Comput. Biol. Med. 71 14–23

Bauer T et al 1996 Heart rate variability in patients with sleep-related breathing disorders Cardiology 87 492–6

Bratton D J et al 2015 CPAP versus mandibular advancement devices and blood pressure in patients with obstructive sleep apnea: a systematic review and meta-analysis JAMA 313 2280–93

Breiman L 2001 Random forests Mach. Learn. 45 5–32

Chesson A J et al 2003 Practice parameters for the use of portable monitoring devices in the investigation of suspected obstructive sleep apnea in adults Sleep 26 907–13

Cristianini N and Shawe-Taylor J 2000 An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods (Cambridge: Cambridge University Press)

Dingli K et al 2003 Spectral oscillations of rr intervals in sleep apnoea/hypopnoea syndrome patients Eur. Respir. J. 22 943–50

Flemens W W et al 2003 Home diagnosis of sleep apnea: a systematic review of the literature Chest 124 1543–79

Fujiki K et al 2019 Heart rate variability-based driver drowsiness detection and its validation with EEG IEEE Trans. Biomed. Eng. 66 1769–78

Fujiki K et al 2016 Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features IEEE Trans. Biomed. Eng. 63 1321–32

Giles T L et al 2006 Continuous positive airways pressure for obstructive sleep apnoea in adults Cochrane Database Syst. Rev. 19 CD001106

Goldberger A L et al 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals Circulation 101 e215–20

Gong X et al 2016 Correlation analysis between polysomnography diagnostic indices and heart rate variability parameters among patients with obstructive sleep apnea hypopnea syndrome PLoS One 11 e015628

Gosselin N et al 2002 Age difference in heart rate changes associated with micro-arousals in humans Clin. Neurophysiol. 113 1517–21

Gula J I, Krahn A D, Skanes A, Ferguson K A, George C, Yee R and Klein G J 2003 Heart rate variability in obstructive sleep apnea: a prospective study and frequency domain analysis Ann. Noninvasive Electrocardiol. 8 144–4

Haneghan C et al 2008 Electrocardiogram recording as a screening tool for sleep disordered breathing J. Clin. Sleep Med. 4 223–8

Ho T K 1998 The random subspace method for constructing decision forests IEEE Trans. Pattern Anal. Mach. Intell. 20 832–44

Hoshi R, Pastre C M, Vanderlei L C and Godoy M F 2013 Poincaré plot indexes of heart rate variability: relationships with other nonlinear variables Auton. Neurosci. 177 271–4

Jean R E et al 2017 Improvement in physical activity in persons with obstructive sleep apnea treated with continuous positive airway pressure J. Phys. Act. Health 14 176–82

Jung D W et al 2017 Apnea–hypopnea index prediction using electrocardiogram acquired during the sleep-onset period IEEE Trans. Biomed. Eng. 64 295–301

Kadotani H et al 2011 Durability, safety, ease of use and reliability of a type-3 portable monitor and a sheet-style type-4 portable monitor Sleep Biol. Rhythms 9 86–94

Kamen P W, Krum H and Tonkin A M 1996 Poincaré plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans Clin. Sci. 91 201–8

Kapur V K et al 2002 Underdiagnosis of sleep apnea syndrome in U.S. communities Sleep Breath. 6 49–54

Kapur V K et al 2017 Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline J. Clin. Sleep Med. 13 480–504

Kleiger R E et al 1987 Decreased heart rate variability and its association with increased mortality after acute myocardial infarction Am. J. Cardiol. 59 256–62

Kufoy E et al 2012 Changes in the heart rate variability in patients with obstructive sleep apnea and its response to acute CPAP treatment PLoS One 7 e37069

Lado M J et al 2011 Detecting sleep apnea by heart rate variability analysis: assessing the validity of databases and algorithms J. Med. Syst. 35 473–81

Laitinen T, Hartikainen J, Vanninen E, Niskanen L, Geelen G and Lansimies E 1998 Age and gender dependency of baroreflex sensitivity in healthy subjects J. Appl. Physiol. 84 576–83

Lu G and Yang F 2009 Limitations of oximetry to measure heart rate variability measures Cardiovascular Eng. 9 119–25

Malik M et al 1996 Heart rate variability—standards of measurement, physiological interpretation, and clinical use Eur. Heart J. 17 354–81

Malliani A 1991 Cardiovascular neural regulation explored in the frequency domain Circulation 84 482–92

Marin J M et al 2005 Long-term cardiovascular outcomes in men with obstructive sleep apnea–hypopnoea with or without treatment with continuous positive airway pressure: an observational study Lancet 365 1046–53

Matsuo M et al 2016 Comparisons of portable sleep monitors of different modalities: potential as naturalistic sleep recorders Frontiers Neurol. (https://doi.org/10.3389/fneur.2016.00110)
McLachlan G J 1992 Discriminant Analysis and Statistical Pattern Recognition (New York: Wiley)
Mender M O et al 2009 Sleep apnea screening by autoregressive models from a single ECG lead IEEE Trans. Biomed. Eng. 56 2838–50
Mender M O et al 2010 Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis Physiol. Meas. 31 273–89
Monahan K and Redline S 2011 Role of obstructive sleep apnea in cardiovascular disease Curr. Opin. Cardiol. 26 541–7
Nakayama-C Ashida Y et al 2008 Sleep–disordered breathing in the usual lifestyle setting as detected with home monitoring in a population of working men in Japan Sleep 31 419–25
Nakayama C et al 2015 Development of sleep apnea syndrome screening method by using heart rate variability analysis and support vector machine IEEE EMBC 2015 (Milan, Italy) pp 8165–8
Narkiewicz K et al 1998 Altered cardiovascular variability in obstructive sleep apnea Circulation 98 1071–7
Penzel T, McNames J, Murray A, de Chazal P, Moody G and Raymond B 2002 Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings Med. Biol. Eng. Comput. 40 402–7
Penzel T et al 2000 The Apnea-ECG database Comput. Cardiol. 27 255–8
Peppard P E et al 2000 Prospective study of the association between sleep–disordered breathing and hypertension New Engl. J. Med. 342 1378–84
Qaseem A et al 2014 Diagnosis of obstructive sleep apnea in adults: a clinical practice guideline from the American College of Physicians Ann. Intern. Med. 161 210–20
Reichmuth K J et al 2005 Association of sleep apnea and type II diabetes: a population-based study Am. J. Respir. Crit. Care Med. 172 1590–5
Richman J S and Moorman J R 2000 Physiological time-series analysis using approximate entropy and sample entropy Am. J. Physiol. Heart Circ. Physiol. 278 H2039–49
Sassani A et al 2004 Reducing motor-vehicle collisions, costs, and fatalities by treating obstructive sleep apnea syndrome Sleep 27 453–8
Sforza E et al 2000 Cardiac activation during arousal in humans: further evidence for hierarchy in the arousal response Clin. Neurophysiol. 111 1611–9
Somers V K et al 1995 Sympathetic neural mechanisms in obstructive sleep apnea J. Clin. Invest. 96 1897–909
Song C et al 2016 An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals IEEE Trans. Biomed. Eng. 63 1352–41
Szollosi I et al 2007 Sleep apnea in heart failure increases heart rate variability and sympathetic dominance Sleep 30 1509–14
Takeda T, Mizuno O and Tanaka T 2015 Time-dependent sleep stage transition model based on heart rate variability Proc. 37th Int. Conf. of the IEEE Eng. Med. Biol. Soc. 2015 2343–6
Taranto Montemurro L, Floras J S, Picton P, Kasai T, Alshaer H, Gabriel J M and Bradley T D 2014 Relationship of heart rate variability to sleepiness in patients with obstructive sleep apnea with and without heart failure J. Clin. Sleep Med. 10 271–6
Tsukada S et al 2012 Conductive polymer combined silk fiber bundle for bioelectrical signal recording PLoS One 7 e33689
Vanninen E et al 1996 Cardiac sympathovagal balance during sleep apnea episodes Clin. Physiol. Funct. Imaging 16 209–16
Varon J et al 2015 A novel algorithm for the automatic detection of sleep apnea from single-lead ECG IEEE Trans. Biomed. Eng. 62 2269–78
Xiao M, Yan H, Song J, Yang Y and Yang X 2013 Sleep stages classification based on heart rate variability and random forest Biomed. Signal Process. Control 8 624–33
Yamakawa T et al 2014 Real-time heart rate variability monitoring employing a wearable telemeter and a smartphone APSIPA ASC (Kaohsiung, Taiwan) pp 1–4
Youden W J 1950 Index for rating diagnostic tests Cancer 3 32–5