An Obstructive Sleep Apnea Detection Approach Using a Discriminative Hidden Markov Model From ECG Signals

Changyue Song, Kaibo Liu, *Member, IEEE*, Xi Zhang*, *Member, IEEE*, Lili Chen, and Xiaochen Xian

***Abstract*—Obstructive sleep apnea (OSA) syndrome is a common sleep disorder suffered by an increasing number of people worldwide. As an alternative to polysomnography (PSG) for OSA diagnosis, the automatic OSA detection methods used in the current practice mainly concentrate on feature extraction and classifier selection based on collected physiological signals. However, one common limitation in these methods is that the temporal dependence of signals are usually ignored, which may result in critical information loss for OSA diagnosis. In this study, we propose a novel OSA detection approach based on ECG signals by considering temporal dependence within segmented signals. A discriminative hidden Markov model (HMM) and corresponding parameter estimation algorithms are provided. In addition, subject-specific transition probabilities within the model are employed to characterize the subject-to-subject differences of potential OSA patients. To validate our approach, 70 recordings obtained from the Physionet Apnea-ECG database were used. Accuracies of 97.1% for per-recording classification and 86.2% for per-segment OSA detection with satisfactory sensitivity and specificity were achieved. Compared with other existing methods that simply ignore the temporal dependence of signals, the proposed HMM-based detection approach delivers more satisfactory detection performance and could be extended to other disease diagnosis applications.**

***Index Terms*—Electrocardiogram (ECG), hidden Markov model (HMM), obstructive sleep apnea (OSA), temporal dependence.**

I. Introduction

OBSTRUCTIVE sleep apnea (OSA) syndrome is a chronic disease that clinically features abnormal reductions (hypopnea) or cessations (apnea) in airflow in breathing during sleep. Symptoms of OSA include daytime somnolence, slow reaction, and heavy snoring, among others. Approximately 14% of men and 5% of woman in the US suffer from OSA syndrome, and the disease shows an increasing incidence in various populations worldwide [1]. In clinical practice, the severity of OSA is usually measured by the number of apnea and hypopnea events per hour during sleep; this parameter is known as the apnea-hypopnea index (AHI). A subject with an AHI above 5 combined with other clinical symptoms such as excess daytime sleepiness and impaired cognition, is typically diagnosed as an OSA patient [2]. AHI is generally calculated through overnight polysomnography (PSG) collected from suspected OSA patients.

Currently, PSG is the most widely used diagnostic tool for the OSA; it requires subjects to sleep in a laboratory for one or two nights and records approximately 16 major physiological signals including electrocardiogram (ECG), electroencephalogram (EEG), respiratory effort, airflow signal, and oxygen saturation (SaO2) [3]. However, OSA diagnosis through PSG often results in an uncomfortable experience because several wires and electrodes must be attached to suspected OSA patients during signal recording. This procedure may influence the accuracy of the PSG diagnosis with nonnegligible measurement noises, and the result is sensitive to the complex testing procedure. Moreover, PSG also requires a dedicated laboratory, special equipment and an attending nurse, and thus, considerable capital investment for equipment, bed space, and specialized technical support must be available. As a result, these disadvantages impede the wider implementation of PSG for general use in public health applications.

Adancements in wearable sensors have prompted the development of remote monitoring and diagnosis techniques and encouraged researchers to study a reduced set of physiological signals instead of analyzing all available PSG channels. These signals including ECG [4], [5], SaO_2 [6], [7], snoring signal [8], [9], midsagittal jaw movement [10], respiratory sound [11], thoracic effort [12], and EEG signals [13] or a combination of these [14], have been found in the literature. Among them, ECG signal is a particularly interesting signal type to study because it allows physiological demonstration of OSA occurence and is convenient for recording with wearable devices. Specifically, when an apnea event occurs, the blood oxygen levels decrease and the cardiovascular system is prompted to maintain adequate oxygen supply to the body. Thus, the observation of abnormal heart activities or high heart rate variability may provide evidence of OSA occurrence. In this study, we propose a new method for OSA detection with ECG signals based on hidden Markov models (HMM).

The rest of this paper is organized as follows. Section II reviews the related studies of OSA detection with ECG signals. In Section III, we describe our motivation and summarize the proposed approach. Section IV gives a brief summary of our dataset,

Manuscript received March 4, 2015; revised June 13, 2015, September 20, 2015, and October 30, 2015; accepted October 31, 2015. Date of publication November 5, 2015; date of current version June 16, 2016. This work was supported in part by the US National Science Foundation under Grant CMMI-1362529 and the National Natural Science Foundation of China under Grant 71201002 and Grant 71571003. *Asterisk indicates corresponding author.*

C. Song, K. Liu, and X. Xian are with the Department of Industrial and Systems Engineering, College of Engineering, University of Wisconsin-Madison.

*X. Zhang is with the Department of Industrial Engineering and Management, College of Engineering, Peking University, Beijing 100871, China (e-mail: xi.zhang@coe.pku.edu.cn).

L. Chen is with the Department of Industrial Engineering and Management, College of Engineering, Peking University.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2015.2498199

and Section V provides a detailed description of the proposed approach. Results and performance measures are demonstrated in Section VI. In Section VII, a discussion of the results is presented. The concluding remarks are drawn in Section VIII.

II. Related Works

Most of the OSA detection methods practiced today focus on extracting the time domain, frequency domain, and other nonlinear features from various ECG-based signals and then constructing classifiers with these features to determine OSA occurrence. Studying the complete ECG of a subject during his/her sleep is a popular research topic. For example, Khandoker *et al*., adopted a support vector machine (SVM) for OSA patient identification based on various features extracted from RR invervals (R wave to R wave intervals obtained from ECGs) and ECG-derived resporatory (EDR) signals by wavelet decomposition; in their work, over 90% of the subjects in the testing set were correctly classified [15]. Unfortunately, these methods are unable to provide detailed insights, such as the time of occurrence of apnea events and the disease severity of suspected patients, to supplement and facilitate the accurate diagnosis of OSA by physicians.

To address this issue, several alternative classification methods that extract features from each equal-length segment of ECG signals have been developed for the OSA diagnosis. These methods are capable of determining the occurrence of OSA in each segment as well as the general OSA severity of a patient. For example, Shouldice *et al*., proposed a classification method by integrating minute-by-minute segmented ECG signals into quadratic discriminant analysis (QDA) to detect OSA in pediatric subjects, thereby achieving an accuracies of 72.1% on a per-segment basis and 84% on a per-subject basis [16]. Medez *et al*., studied over 70 features from each segment, and distinguished apnea segments from normal segments with a classification accuracy of 88% by the k -nearest neighbor (KNN) and neural networks [17]. Bsoul *et al*., developed a real-time apnea monitoring system based on smart phones; 63 features from RR intervals and 48 from EDR signals were extracted minute by minute and SVMs were employed for apnea detection [18].

Two common problems in a number of the OSA detection studies lie in the high-dimension feature space and the black-box decision making process. In the literature, various feature selection methods including statistical evaluation [18], wrapper methods [17], and principle component analysis [19], have been applied to reduce the dimension of the feature space. For the decision making process, Sannino *et al.*, proposed a method to extract explicit IF . . . THEN rules and detect OSA by comparing the value of extracted features with predetermined parameters, thus, resulting in clear explanation of the detection result [20].

While the segment-based classifiers described above perform effectively by considering each segment separately, none of them explicitly considered the intrinsic relevance between ECG segments; however, temporal dependence does exist among segments in real cases, as will be demonstrated in the next section.

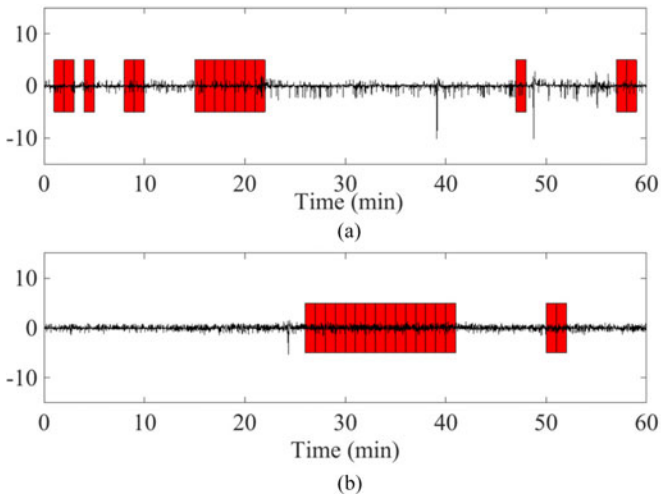


Fig. 1. Example of apnea events patterns determined by PSG from two OSA patients with a similar severity.

III. Proposed Approach

In real ECG recordings, temporal dependence among segments is frequently observed. Fig. 1 shows two pieces of ECG signals of 60 min from real recordings; here, each minute of the ECG signal is labeled as either normal or apnea by physicians and the apnea minutes have been marked by red rectangles. As illustrated in Fig. 1, apnea segments are noticeably more likely to be concatenated over time rather than randomly dispersed, which reflects a high dependency in the time series of segment states (normal/apnea). Furthermore, the patterns revealed by the segments significantly vary between the subjects. For example, the apnea events of Subject 1 in Fig. 1 tend to be more dispersive than those of Subject 2. This finding indicates that not only the state of normal or apnea is highly temporally dependent, but also the state transition patterns differ from one subject to another. Although these results are frequently observed in annotated data sequences, to the best of our knowledge, few studies have yet addressed this issue. As such, developing a systematic approach to capture the temporal dependence of segmented signal sequence as well as the differences among individual subjects is necessary to achieve more accurate OSA detection with ECG signals.

This paper proposes a novel OSA detection method based on Markov chains that considers the temporal dependence of segmented signals. In this approach, we treat the features obtained from each ECG segment as an observation, and the condition (normal or apnea) of the segment that is not directly observed as a Markov state; thus, a HMM could be established. An important assumption in this method is that while the features obtained from ECGs depend only on the state of the segment, which is consistent with most of the segment-based classifiers mentioned above, the transition probabilities between states are subject-specific in order to characterize the subject-to-subject differences (i.e., different subjects may have different transition patterns as shown in Fig. 1). Thus, our proposed method constructs a specific model for each ECG recording and is able to capture the individual differences among potential OSA patients. Conventional methods for parameter estimation and state

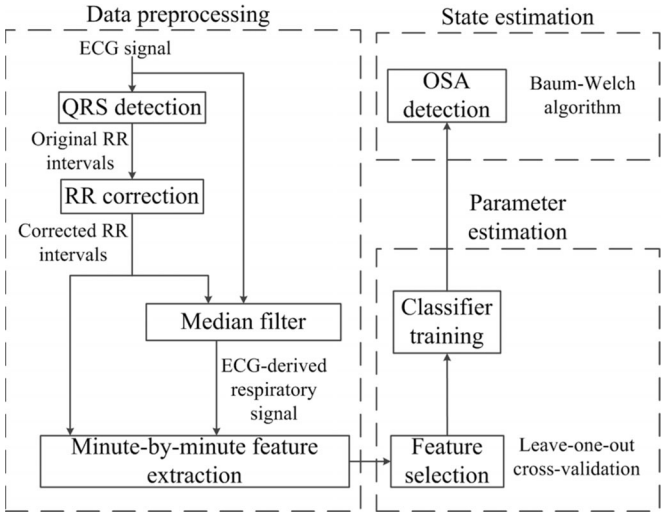


Fig. 2. Scheme of the proposed OSA screening approach.

prediction of HMM s require that both the transition probabilities and the features are subject-specific, and these features must conform to certain distributions in order to obtain analytical solutions. On the contrary, in our study, the features extracted from ECGs are considered subject-independent and may originate from a variety of distributions. To address this problem, a new learning and prediction procedure is further proposed based on a discriminative HMM. The scheme of our proposed method is demonstrated in Fig. 2. First, RR intervals and EDR signals are obtained from ECGs; these data are used to extract features and establish the feature pool for OSA detection. The best feature subset and the observation-to-state classifier are then provided simultaneously by training the discriminative HMM. In the detection procedure, the Baum–Welch algorithm is used to estimate the state of each segment. The proposed method is evaluated by comparing with the results of existing approaches.

IV. Data Description

The data in this study were taken from the Physionet Apnea-ECG database [21], [22]. A total of 70 nighttime recordings of single continuous ECG signals running for approximately 8 h were acquired for analysis. These ECG signals were obtained from complete PSG recordings with a sampling rate of 100 Hz, 16-bit resolution, and modified lead V2 electrode configuration. The recordings were segmented on a minute-by-minute basis, and each segment was labeled as either normal or apnea by physicians.

These recordings were originated from a variety of subjects. Ages of the subjects range from 27 to 63 (mean $\pm$ standard deviation: 43.8 ± 11.1 years); weights from 53 to 135 kg (mean $\pm$ standard deviation: 86.3 ± 22.4 kg); and heights from 158 to 184 cm (mean $\pm$ standard deviation: 175.3 ± 6.1 cm). The duration of these 70 ECG signal recordings varies from 401 to 578 minutes (mean $\pm$ standard deviation: 432 ± 32 min) and the AHI ranges from 0 to 93.5.

The recordings consisted of two sets, each of which contained 35 recordings. The first set (released set) was used for model construction and parameter estimation. The second set (withheld set) was used to assess and validate our method. A total of 17 268 segments were included in the withheld set; of these segments, 10 718 were annotated as "normal" and the remaining 6550 were annotated as "apnea." In this study, recordings with AHI ≥ 5 were defined as OSA positive; otherwise, they were categorized as OSA negative. In this way, the withheld set contained 24 OSA positive recordings and 11 OSA negative recordings.

V. Methodology

A. Preprocessing

Previous studies have revealed that RR intervals and EDR signals may contain critical information about OSA occurrence [26], [28]. Hence, a preprocessing procedure is applied to acquire RR intervals and EDR signals from the original ECGs. The RR interval generally measures the duration of a heartbeat cycle and is obtained by calculating the time span between adjacent QRS (Q wave, R wave, and S wave) peaks. An external package from the BIOSIG-toolbox was employed in this study to locate the peaks of QRS complexes [23], [24]. Because of the existence of physiologically uninterpretable points within the generated RR intervals, a median filter proposed by Chen *et al*., was adopted to eliminate those underlying abnormal values [25].

Besides RR intervals, EDR signals were also derived and studied in the literature [15], [26], [27]. EDR signals generally reflect respiratory activities in terms of relative motions between the electrodes and the heart since the electrical impedance of the thoracic cavity changes during breathing cycles [23]. In this study, we followed the procedure described in [26] to obtain these signals.

B. Feature Extraction

Various features which may provide critical information for OSA detection can be extracted from RR intervals and EDR signals through a number of extraction methods [26], [29], [30]. In our study, all features in the feature pool, including 24 features from RR intervals and 8 features from EDR signals, were extracted on a minute-by-minute basis; these features are listed as follows:

1) Mean, standard deviation, skewness, and kurtosis of RR intervals.
2) The first five serial correlation coefficients of RR intervals.
3) The NN50 measure (variant 1), defined as the number of pairs of adjacent RR intervals where the first RR interval exceeds the second one by more than 50 ms.
4) The NN50 measure (variant 2), defined as the number of pairs of adjacent RR intervals where the second RR interval exceeds the first one by more than 50 ms.
5) Two pNN50 measures, defined as each NN50 measure divided by the total number of RR intervals.
6) The SDSD measure, defined as the standard deviation of the differences between adjacent RR intervals.

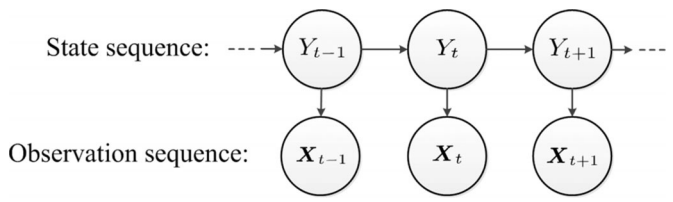


Fig. 3. Illustration of a hidden Markov model.

7) The RMSSD measure, defined as the square root of the average of the squares of differences between adjacent RR intervals.
8) The Allan factor $A(T)$ evaluated at a time scale T of 5, 10, 15, 20, and 25 s

$$A(T) = E\left\{\frac{[N_{i+1}(T) - N_i(T)]^2\}}{2E\{N_{i+1}(T)\}}\right\} \tag{1}$$

where $N_i(T)$ is the number of QRS detection points in a window of length T stretching from iT to $(i+1)T$ and E is the expectation operator.
9) Normalized very low frequency (VLF), low frequency (LF), and high frequency (HF) components of RR intervals where the total power is the sum of the three components;
10) The ratio of LF to HF of RR intervals;
11) Mean, standard deviation, skewness, and kurtosis of EDR signal;
12) Normalized VLF, LF, and HF of EDR signal where the total power is the sum of the three components;
13) The ratio of LF to HF of EDR signal.

It should be noticed that before feature extraction, RR intervals and EDR signals were normalized to mean '0' and standard deviation '1' in each recording to reduce the difference among subjects. After feature extraction, all features were linearly scaled into the same range by the following equation:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{2}$$

where x is a record of a feature to be scaled, $x_{\min}$ is the minimum value of the feature in the dataset, $x_{\max}$ is the maximum value of the feature in the dataset, and x^* is the record of the feature after scaling. There are two reasons for feature scaling. First, some classifiers are sensitive to the scale of features such as KNN which was used in this study. Scaling all features into the same range avoids features in greater numeric ranges dominating those in smaller numeric ranges. Second, gradient descent algorithm converges much faster with feature scaling, and thus, the computational time required for training the classifiers can be saved.

C. Hidden Markov Model

In this section, we briefly introduce the HMM and the corresponding parameter estimation algorithm for OSA detection. An HMM is a statistical Markov model with two sequences: the unobservable Markov states and the observations. Fig. 3 illustrates an HMM in which an observation X_t depends on the corresponding Markov state Y_t and a Markov state Y_t depends on the previous state Y_{t-1}. The notations in the HMM are listed as follows:

$Y_{1:T}$: The Markov state sequence up to time T, $Y_{1:T} = \{Y_1, Y_2, \ldots, Y_T\}$;
$\boldsymbol{X}_{1:T}$: The observation sequence up to time T, $\boldsymbol{X}_{1:T} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_T\}$;
S: The Markov state space, $S = \{0, 1, \ldots, K-1\}$ where K is the total number of different states in the state space S;
Y_t: The Markov state at time t, $Y_t \in S$;
$\boldsymbol{X}_t$: The observation at time t, $\boldsymbol{X}_t$ is a continuous random vector;
$\boldsymbol{\pi}$: The initial probability distribution of the Markov state, $\pi = (\pi_0, \pi_1, \ldots, \pi_{K-1})$ where $\pi_i = P(Y_1 = i)$;
$\boldsymbol{A}$: The transition probability matrix, $\boldsymbol{A} = [a_{i,j}]_{K\times K}$ where $a_{i,j} = P(Y_{t+1} = j | Y_t = i)$;
$\mu_i(\boldsymbol{x})$: The emission distribution of the observations, $\mu_i(\boldsymbol{x}) = P(\boldsymbol{X}_t = \boldsymbol{x} | Y_t = i)$;
$\boldsymbol{p}$: The marginal probability distribution of Markov states, $\boldsymbol{p} = (p_0, p_1, \ldots, p_{K-1})$ where $p_i = P(Y_t = i)$; and
$f_i(\boldsymbol{x})$: The observation-to-state classifier, $f_i(\boldsymbol{x}) = P(Y_t = i | \boldsymbol{X}_t = \boldsymbol{x})$.

According to the probability definition, the following equations must be satisfied:

$$\sum_{i=0}^{K-1} \pi_i = 1, \sum_{i=0}^{K-1} p_i = 1, \sum_{i=0}^{K-1} a_{j,i} = 1, \forall j \in S. \tag{3}$$

As the conditional probability of the observations can be calculated by Bayesian rule

$$P(\boldsymbol{X}_t = \boldsymbol{x} | Y_t = i) = \frac{P(Y_t = i | \boldsymbol{X}_t = \boldsymbol{x}) P(\boldsymbol{X}_t = \boldsymbol{x})}{P(Y_t = i)} \propto \frac{P(Y_t = i | \boldsymbol{X}_t = \boldsymbol{x})}{P(Y_t = i)} \tag{4}$$

the emission distribution μ can be decomposed into the classifier f and the marginal probability distribution $\boldsymbol{p}$. As a result, an HMM can be generatively characterized by the parameters $\boldsymbol{\Theta} = (\boldsymbol{\pi}, \boldsymbol{A}, \mu)$ or discriminatively by the parameters $\tilde{\boldsymbol{\Theta}} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{p}, f)$ [31]. Although these two expressions are theoretically equivalent, discriminative HMMs are preferable to generative HMMs in our study. The advantages of discriminative HMMs are twofold. First, it reduces the complexity of estimating the multivariate emission distribution μ in generative HMMs, especially when the dimension of observation $\boldsymbol{X}_t$ is large and the state space S is relatively small as in OSA detection. Second, discriminative HMMs are compatible with various existing segment-based classifiers from previous studies; thus, they are more applicable for analyzing segmented ECG signals.

In this study, each minute-by-minute segment of the ECG signals has a hidden Markov state Y_t with state space $S = \{0(\text{normal}), 1(\text{apnea})\}$, and the corresponding observation $\boldsymbol{X}_t$ is a vector composed of the features extracted from the segment. Each ECG recording is modeled as a Markov chain. According to our assumptions, i.e., the distribution of features is subject-independent and the transition probabilities between states are

subject-specific, different ECG recordings have the same classifier f but different transition probability matrices $\boldsymbol{A}$.

The likelihood function of a Markov chain with known Markov states $Y_{1:T}$, known observations $\boldsymbol{X}_{1:T}$, and unknown model parameters $\boldsymbol{\pi}$, $\boldsymbol{A}$, f can be written as (see Appendix I for more details)

$$L\left(Y_{1:T}=y_{1:T},X_{1:T}=x_{1:T}|\tilde{\boldsymbol{\Theta}}\right) \propto \pi_{y_1}\cdot\prod_{t=1}^{T}f_{y_t}\left(x_t\right)\prod_{t=2}^{T}a_{y_{t-1},y_t}. \tag{5}$$

Maximum likelihood estimation (MLE) is used to obtain these unknown parameters. Since these three unknown parameters in the likelihood function are assumed to be mutually independent, the three components of the likelihood function can be separately maximized:

1) Without loss of generality, the first component, π_{y_1} can be set to 1. This step agrees with the empirical fact that all Markov chains begin with the same state (normal) since apnea does not usually occur at the beginning of the overnight ECG signals when subjects are not totally in sleep.
2) For the second component $\prod_{t=1}^{T}f_{y_t}(\boldsymbol{x}_t)$, we establish a classifier to predict the state y_t of each segment given the observation x_t. The associated parameters are estimated by training the classifier with the extracted features.
3) The third component $\prod_{t=2}^{T}a_{y_{t-1},y_t}$ can be maximized by:

$$\hat{a}_{i,j}=\frac{n_{i,j}}{n_{i,\cdot}} \tag{6}$$

where $n_{i,j}$ is the empirical transition frequency from state i to state j, and $n_{i,\cdot}$ is the total empirical frequency for leaving from state i in the Markov chain (see Appendix II for more details).

D. Feature Selection

In clinical practice, only parsimonious features with reasonable interpretations are accepted by physicians and supplemented for OSA diagnosis. Thus, a feature selection step is needed to exclude insignificant features and reduce the risk of incorrect detection. Here, a scheme of stepwise feature selection with leave-one-out cross-validation (LOOCV) was adopted. The feature selection procedure is described as follows:

Step 1 (Initialization): Denote the whole set of the feature pool as Ω. Set the best feature subset $B=\varnothing$, and the discarded feature subset $D=\varnothing$.

Step 2 (Forward selection): For each feature $l\in\Omega$ and $l\notin D$, construct a temporary feature subset $E_l=B\cup\{l\}$, which is obtained by adding the feature l to B. Conduct LOOCV within subset E_l and put the feature $l_{\max}$ which leads to the maximum LOOCV accuracy into the best feature subset $B=B\cup\{l_{\max}\}$.

Step 3 (Backward selection): For each feature $l\in B$, construct a temporary feature subset $E_l=B-\{l\}$, which is obtained by removing the feature l from B. Conduct LOOCV within subset E_l and check the feature $l_{\max}$ from E_l which leads to the maximum LOOCV accuracy. If the accuracy is higher than the original LOOCV accuracy based on B, remove the feature out of the best feature subset $B=B-\{l_{\max}\}$ and put it into the discarded feature subset $D=D\cup\{l_{\max}\}$. Repeat this step until subsets B and D do not change.

Step 4: Iterate Steps 2 and 3 until $B\cup D=\Omega$. Then, features in B are selected for constructing classifiers.

E. OSA Detection

During OSA detection, we employed the Baum–Welch algorithm to estimate the unobservable Markov states Y_t and the unknown parameters in our Markov model, including the initial probability distribution π and the transition probability matrix $\boldsymbol{A}$ [32], [33]; here, $\boldsymbol{A}$ is unknown because of the assumption of subject-specific transition probabilities. Define the forward variable $\alpha_t(i)$ as the probability of partial observation sequence up to time t with state i at time t given model parameter set $\tilde{\boldsymbol{\Theta}}$:

$$\alpha_t\left(i\right)=P\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_t,y_t=i|\tilde{\boldsymbol{\Theta}}\right). \tag{7}$$

The backward variable $\beta_t(i)$ is defined as the probability of partial observation sequence from time $t+1$ to T, given state i at time t and model parameter set $\tilde{\boldsymbol{\Theta}}$:

$$\beta_t\left(i\right)=P\left(\boldsymbol{x}_{t+1},\boldsymbol{x}_{t+2},...,\boldsymbol{x}_T|y_t=i,\tilde{\boldsymbol{\Theta}}\right). \tag{8}$$

Then, OSA detection based on the discriminative HMM can be conducted according to the following procedures:

Step 1: Set the initial estimation of $\boldsymbol{\pi}$ and $\boldsymbol{A}$ as $\boldsymbol{\pi}^{(0)}=(\pi_0^{(0)},\pi_1^{(0)},\ldots,\pi_{K-1}^{(0)})$ and $\boldsymbol{A}^{(0)}=[a_{i,j}^{(0)}]_{K\times K}$, where K is the total number of different states in the state space. And set $n=0$.

Step 2: Initialize forward and backward variables by

$$\alpha_1\left(i\right)\;=\;\pi_i^{(n)}\frac{f_i\left(\boldsymbol{x}_1\right)}{p_i} \tag{9}$$

$$\beta_T(i)\;=\;1 \tag{10}$$

for $\forall i\in S$, where $S=\{0,1,\ldots,K-1\}$ is the state space.

Step 3: Calculate all forward variables and backward variables recursively by

$$\alpha_{t+1}\left(i\right)\;=\;\left[\sum_{k=0}^{K-1}\alpha_t\left(k\right)a_{k,i}^{(n)}\right]\frac{f_i\left(\boldsymbol{x}_{t+1}\right)}{p_i} \tag{11}$$

$$\beta_t\left(i\right)\;=\;\sum_{k=0}^{K-1}a_{i,k}^{(n)}\beta_{t+1}\left(k\right)\frac{f_k\left(\boldsymbol{x}_{t+1}\right)}{p_k} \tag{12}$$

for $\forall i\in S$ and $t=1,2,\ldots,T-1$.

Step 4: Calculate the probability of state i at time t and state j at time $t+1$ given all observations by

$$\xi_t(i,j)=\frac{\alpha_t(i)a_{i,j}\beta_{t+1}(j)}{P(\boldsymbol{X}_{1:T}=\boldsymbol{x}_{1:T}|\tilde{\boldsymbol{\Theta}})}\frac{f_j(\boldsymbol{x}_{t+1})}{p_j} \tag{13}$$

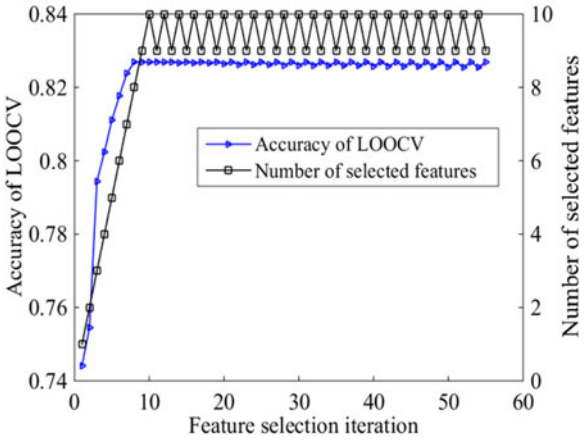


Fig. 4. Accuracy and the number of selected features in each LOOCV iteration.

for $\forall i, j \in S, t = 1, 2, \ldots, T-1$, and

$$P\left(\boldsymbol{X}_{1:T} = \boldsymbol{x}_{1:T} | \tilde{\boldsymbol{\Theta}}\right) = \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \alpha_t(i)\, a_{i,j} \beta_{t+1}(j) \frac{f_j(\boldsymbol{x}_{t+1})}{p_j}. \quad (14)$$

Step 5: Calculate the probability of state i ($i \in S$) at time t ($t = 1, 2, \ldots, T-1$) given all observations by

$$\gamma_t(i) = \sum_{j=0}^{K-1} \xi_t(i,j). \quad (15)$$

Step 6: Update the estimation of $\boldsymbol{\pi}$ and $\boldsymbol{A}$ by

$$\pi_i^{(n+1)} = \gamma_1(i) \quad (16)$$

$$a_{i,j}^{(n+1)} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}. \quad (17)$$

Step 7: If the estimation of $\boldsymbol{\pi}$ and $\boldsymbol{A}$ converges to stable values, stop the algorithm; otherwise, set $n = n + 1$ and go back to step 2.

It should be stated that the latest updated estimation of $\gamma_t(i)$ presents the posterior probability distribution of Markov states for each segment, and thus, it can be used to detect OSA events.

VI. Results

A. Feature Selection

In theory, due to the generality of our proposed approach, a variety of self-designed classifiers can be employed in feature selection. Here, as SVM has been widely used in classification problems, we chose an SVM with Gaussian radial basis function kernel in this study. Fig. 4 provides the selection procedure with accuracy of LOOCV as well as the number of selected features at each iteration. The performance of the classifier improves with increasing number of selected features and reaches a stable level after nine iterations, at which a maximum accuracy of 82.7% is achieved.

TABLE I
Per-Segment OSA Detection Performance Metrics in the Withheld Set

	Accuracy(%)	Sensitivity(%)	Specificity(%)	AUC[a]
SVM	81.2	75.5	84.7	0.889
HMM+SVM	86.2	82.6	88.4	0.940
LR	81.2	74.4	85.4	0.883
HMM+LR	86.2	80.0	89.9	0.939
LDA	80.5	83.1	78.9	0.881
HMM+LDA	85.3	77.5	90.1	0.933
KNN	80.7	75.3	83.9	0.881
HMM+KNN	84.5	74.0	90.8	0.924

[a] AUC: Area under receiver operating characteristic curve.

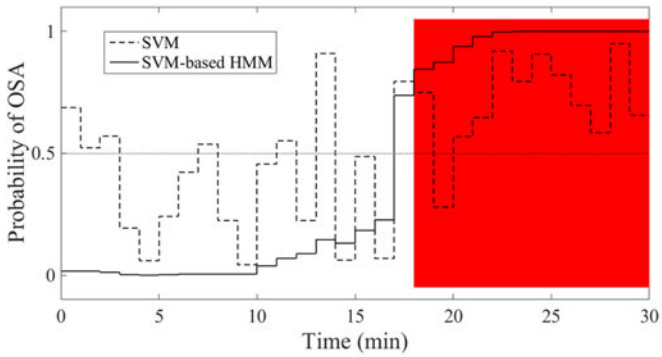


Fig. 5. Posterior probability produced by SVM and SVM-based HMM in 30 min of a recording in the withheld set.

The selected features are listed as follows:

1) Mean value of RR intervals.
2) The second and third correlation coefficients of RR intervals.
3) The pNN50 measure (variant 2).
4) The SDSD measure.
5) Normalized VLF of RR intervals.
6) Normalized VLF, LF, and the ratio of LF to HF of EDR signal.

B. Per-Segment OSA Detection

Upon determination of the selected features, we employed several popular classifiers including SVM, logistic regression (LR), linear discriminant analysis (LDA) and KNN, to evaluate the performance of the proposed OSA detection method. The overall performance of these classifiers and HMMs on the withheld set, including accuracy, sensitivity, specificity, and area under curve (AUC), were calculated and compared in Table I. Results reveal that the detection performance of the model considering the temporal depenence via the proposed HMM-based detection approach, is generally better than that of all other classifiers that do not consider temporal dependence.

In order to compare the performance of segment-based classifiers and HMMs visually, the posterior probabilities of the OSA produced by the SVM and SVM-based HMM within 30 min of a recording in the withheld set are shown in Fig. 5. The actual annotations of the first 18 min segments are

TABLE II
PER-RECORDING CLASSIFICATION METRICS IN THE WITHHELD SET

	Acc.(%)	Sen.(%)	Spec.(%)	AUC	Corr.
SVM	80.0	100	36.4	1.00	0.839
HMM+SVM	97.1	95.8	100	1.00	0.860
LR	74.3	100	18.2	1.00	0.847
HMM+LR	97.1	95.8	100	1.00	0.875
LDA	68.6	100	0.00	1.00	0.833
HMM+LDA	97.1	95.8	100	1.00	0.837
KNN	91.4	100	72.7	1.00	0.810
HMM+KNN	91.4	87.5	100	0.985	0.811

normal, and the remaining 12 min are OSA as shadowed by a red rectangle. As illustrated in Fig. 5, the posterior probability produced by SVM-based HMM accurately identifies the OSA events with an increasing probability as time goes, which further demonstrates the advantage of the proposed SVM-based HMM over the conventional SVM approach that ignores the temporal dependency.

C. Per-Recording Classification

For per-recording classification, we compare the expected number of minutes with apnea per hour based on our proposed approach with the clinical standard of AHI as the threshold. Specifically, if the expected number of minutes with apnea per hour of a certain recording is greater than 5, the recording is considered as an OSA positive recording; otherwise the recording is considered as OSA negative. Since both the classifiers and HMMs provide a probability of apnea at each segment, the expected number of minutes with apnea per hour can be calculated as

$$\widehat{\mathrm{NMA}}_f = \frac{60}{T}\sum_{t=1}^{T} f_1(\boldsymbol{x}_t) \tag{18}$$

$$\widehat{\mathrm{NMA}}_{\mathrm{HMM}} = \frac{60}{T}\sum_{t=1}^{T} \gamma_t(1) \tag{19}$$

where $T/60$ is the number of hours as the ECG signals are segmented minute by minute. The per-recording classification performance metrics of segment-based classifiers and HMMs, including accuracy, sensitivity, specificity, and AUC, are listed in Table II. It is worth stating that almost every classifier obtained an AUC of 1. This is mainly because the per-recording classification is based on combining the results of per-segment OSA detection, and all the segment-based classifiers and HMMs attain a high accuracy in per-segment OSA detection. To better demonstrate the performance of the segment-based classifiers and HMMs in the per-recording classification, we additionally provide the correlation between the actual AHI and the calculated number of minutes with apnea per hour in Table II. It can be observed that the correlation measures of HMMs are always higher than these conventional classification methods without considering temporal dependency.

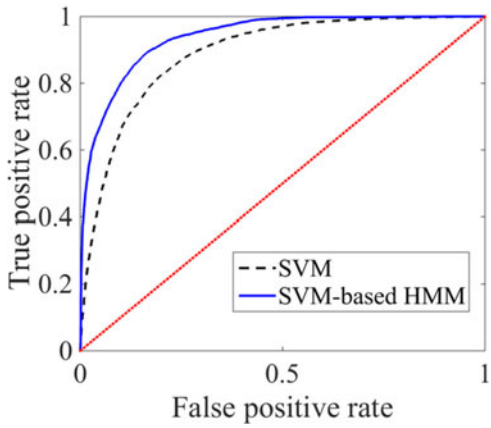


Fig. 6. ROC curves of SVM and HMM on the withheld set.

VII. DISCUSSION

A. Effects of Temporal Dependency

The results obtained from both per-segment OSA detection and per-recording classification reveal that our proposed approach presents significantly improved detection performance. To further evaluate the detection capability of our proposed approach, receiver operating characteristic (ROC) curves were generated. Considering the limited size of our paper, only the ROC curves of the SVM and SVM-based HMM are shown in Fig. 6. The ROC curve of the SVM-based HMM is consistently found above the ROC curve of SVM. This finding demonstrates the superiority of our proposed method for OSA detection when we consider the temporal dependence of segmented signals. It can also be shown theoretically that the segment-based classifiers are special cases of HMMs, and when the segmented signals exhibit strong autocorrelation, HMMs are a better option for OSA detection than conventional classifiers (see Appendix III for more details).

B. Robustness Evaluation

To verify the performance of the proposed approach in different datasets, tenfold cross-validation was conducted. Seventy recordings were divided randomly into ten sets. In every iteration, the SVM and SVM-based HMM were trained on nine selected sets, and then, tested on the remaining set; per-segment testing accuracies were also calculated. Fig. 7 shows the boxplots of the ten accuracy values obtained for SVM and SVM-based HMM. While SVM accuracy ranged from 76.9% to 84.4% (mean $\pm$ standard deviation, $81.9\% \pm 2.35\%$), SVM-based HMM ranged from 80.7% to 90.4% (mean$\pm$standard deviation, $86.4\% \pm 3.61\%$). We can, thus, conclude that our method performs consistently and significantly better in different datasets.

C. Comparison With Previous Work

Recall that our SVM-based HMM can achieve an accuracy of 86.2% for per-segment OSA detection and an accuracy of 97.1% for per-recording classification. These results were

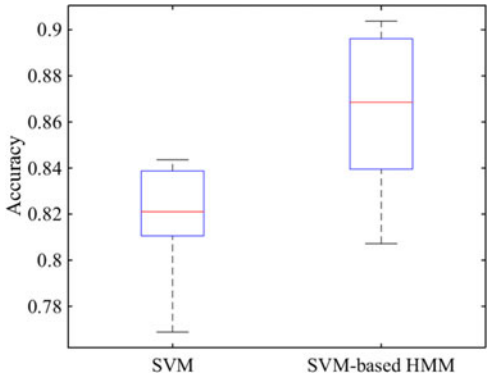


Fig. 7. Boxplots of OSA detection accuracy in tenfold cross-validation.

compared with the existing literatures. For per-segment OSA detection, the results obtained in [17] where an accuracy of 88% were achieved, which is slightly higher than our result, was reported. However, their work requires high-quality dataset and the reported result is based on the partial dataset in which 25 ECG recordings were removed before analysis since the criteria of data quality were not satisfied. Other studies including [18], [20], and [34] also develivered high per-segment OSA detection accuracies which are 89.9%, 88.6%, and 99.2%, respectively. However, among them, [18] and [20] used the released dataset for the validation instead of the whole dataset; [20] detected OSA events in ECG segments based on the historical data from the same ECG recordings; and [34] eliminated one third of the data with a high noise in the preprocessing step. In this regard, our approach is more robust and performs comparatively well on independent datasets without requiring preselection of high-quality data because the quality of the collected signals is not always satisfactory in practice. Furthermore, another advantage of our method is the generality and flexibility since various existing detection algorithms can be directly integrated into our HMM framework.

The results of the present study and other recent studies in terms of per-recording classification are compared in Table III. Among the listed studies, the per-recording classification tasks of [27] and [38] are different from ours. In these studies, recordings were defined as control cases, borderline cases, or apnea cases based on the number of minutes with apnea. The per-recording classification task was to classify nonborderline cases into control cases and apnea cases. In order to compared with [27] and [38], we tested the proposed method in distinguishing control cases with apnea cases on the same dataset. As a result, 29 of the 30 nonborderline cases in the withheld set were successfully classified with an accuracy of 96.7%. Combined with Table III, this result demonstrates that our method achieves a high accuracy and sensitivity.

TABLE III
COMPARISON OF PER-RECORDING CLASSIFICATION RESULT BETWEEN OUR METHOD AND SEVERAL OTHER STUDIES

Reference	Year	Classifier	Accuracy	Sensitivity	Specificity
Macros *et al.* [35]	2009	QDA, LDA, KNN, LR	87.6%	91.1%	82.6%
Khandoker *et al.* [15]	2009	SVM	92.9%	92.4%	93.8%
Alvarez *et al.* [36]	2010	LR	89.7%	92.0%	85.4%
Morillo and Gross [37]	2013	PNN	93.9%	92.4%	95.9%
De Chazal *et al.*[a] [27]	2000	LDA	100%	100%	100%
McNames and Fraser[b] [38]	2000	visual inspection	100%	100%	100%
Our approach		HMM+SVM	97.1%	95.8%	100%

[a,b]The per-recording classification tasks of these two studies are different from ours.
PNN: probabilistic neural networks

VIII. CONCLUSION

We investigated the temporal dependence existed in the segmented ECG signals and its influence on the OSA detection. A discriminative HMM was established to capture this temporal dependence. Under the assumption of subject-independent emission distribution and subject-specific transition probabilities, we constructed a detection procedure to capture individual differences among a suspect population. Features for our HMM-based approach were suggested through an LOOCV feature selection process. The proposed method was verified in terms of accuracy, sensitivity, and specificity by using real ECG recordings, and accuracies of 97.1% for per-recording classification and 86.2% for per-segment OSA detection were observed. These results reveal that our approach achieves more satisfactory performance than previous segment-based classifiers and existing approaches. Our HMM-based approach is compatible with various segment-based classifiers, and thus, provides a basis for future development of diagnostic systems that are not limited to OSA detection. Since only ECG signals are used in our model, the proposed approach promotes the home-based OSA screening and the development of reliable wearable devices for home healthcare services.

Future research study may focus on two aspects. First, the emission distribution in the HMM may be varied for different suspects because each subject presents unique physiological conditions. A well-estimated emission distribution determined by the subject's physiological characteristics may provide more accurate diagnostic information. Second, we observed that emission distributions may depend on the sleeping stages; thus, the Markov chain in our proposed model can be divided into several segments according to the sleeping stage with the help of domain knowledge to elucidate more critical information regarding OSA occurence.

APPENDIX I

Appendix I here derives the likelihood function of the Markov chain

$$L\left(Y_{1:T}=y_{1:T},\boldsymbol{X}_{1:T}=\boldsymbol{x}_{1:T}|\tilde{\boldsymbol{\Theta}}\right)$$
$$=\pi_{y_1}\cdot\prod_{t=1}^{T}\mu_{y_t}\left(\boldsymbol{x}_t\right)\cdot\prod_{t=2}^{T}a_{y_{t-1},y_t}$$

$$= \pi_{y_1} \cdot \prod_{t=1}^{T} \frac{f_{y_t}(\boldsymbol{x}_t) \cdot P(\boldsymbol{X}_t = \boldsymbol{x}_t)}{p_{y_t}} \cdot \prod_{t=2}^{T} a_{y_{t-1},y_t}$$

$$\propto \pi_{y_1} \cdot \prod_{t=1}^{T} f_{y_t}(\boldsymbol{x}_t) \prod_{t=2}^{T} a_{y_{t-1},y_t}. \quad (20)$$

It is worth noting that the marginal probability distribution $\boldsymbol{p}$ represents the prior information about the proportion of normal and apnea segments. Therefore, the value of $\boldsymbol{p}$ should be determined by domain knowledge or other methods. If no prior information is available, a method proposed in Appendix IV can be used to estimate $\boldsymbol{p}$.

APPENDIX II

Appendix II here obtains the upper bound of the third component in the likelihood function of the Markov chain. By the inequality of arithmetic and geometric means,

$$\prod_{t=2}^{T} a_{y_{t-1},y_t} = \prod_{t=2}^{T} n_{y_{t-1},y_t} \prod_{t=2}^{T} \frac{a_{y_{t-1},y_t}}{n_{y_{t-1},y_t}}$$

$$\leq \prod_{t=2}^{T} n_{y_{t-1},y_t} \left(\frac{1}{T-1} \sum_{t=2}^{T} \frac{a_{y_{t-1},y_t}}{n_{y_{t-1},y_t}} \right)^{T-1}$$

$$= \prod_{t=2}^{T} n_{y_{t-1},y_t} \left(\frac{1}{T-1} \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} a_{i,j} \right)^{T-1}$$

$$= \prod_{t=2}^{T} n_{y_{t-1},y_t} \left(\frac{K}{T-1} \right)^{T-1} \quad (21)$$

where $n_{i,j}$ is the empirical frequency for the Markov process to transit from state i to state j, and K is the total number of different states in the state space. The equal sign holds if and only if

$$\frac{a_{y_1,y_2}}{n_{y_1,y_2}} = \frac{a_{y_2,y_3}}{n_{y_2,y_3}} = \cdots = \frac{a_{y_{T-1},y_T}}{n_{y_{T-1},y_T}} = k_c \quad (22)$$

where k_c is a constant. For any i, j that $n_{i,j} \neq 0$, this simplifies to

$$\frac{a_{i,j}}{n_{i,j}} = k_c. \quad (23)$$

Since $\sum_{j=0}^{k-1} a_{i,k} = 1$, the original term is maximized when:

$$\hat{a}_{i,j} = \frac{n_{i,j}}{n_{i,\cdot}}. \quad (24)$$

APPENDIX III

Appendix III here shows that segment-based classifiers are special cases of HMMs. In HMMs, the posterior probability distribution of states given observations can be written as

$$P_{\mathrm{HMM}}(Y_{1:T}|\boldsymbol{X}_{1:T})$$

$$= \frac{P(\boldsymbol{X}_{1:T}|Y_{1:T})\,P(Y_{1:T})}{P(\boldsymbol{X}_{1:T})}$$

$$= \frac{\prod_{t=1}^{T} P(\boldsymbol{X}_t|Y_t) \cdot P(Y_1) \prod_{t=2}^{T} P(Y_t|Y_{t-1})}{P(\boldsymbol{X}_{1:T})}$$

$$= \prod_{t=1}^{T} \frac{P(Y_t|\boldsymbol{X}_t)\,P(\boldsymbol{X}_t)}{P(Y_t)} \cdot \frac{P(Y_1) \prod_{t=2}^{T} P(Y_t|Y_{t-1})}{P(\boldsymbol{X}_{1:T})}$$

$$= \frac{\prod_{t=1}^{T} P(Y_t|\boldsymbol{X}_t) \cdot \prod_{t=1}^{T} P(\boldsymbol{X}_t) \cdot \prod_{t=2}^{T} P(Y_t|Y_{t-1})}{P(\boldsymbol{X}_{1:T}) \cdot \prod_{t=2}^{T} P(Y_t)}. \quad (25)$$

On the other side, this posterior probability in classifiers can be written as

$$P_f(Y_{1:T}|\boldsymbol{X}_{1:T})$$

$$= \frac{P(\boldsymbol{X}_{1:T}|Y_{1:T})\,P(Y_{1:T})}{P(\boldsymbol{X}_{1:T})}$$

$$= \frac{\prod_{t=1}^{T} P(\boldsymbol{X}_t|Y_t) \cdot \prod_{t=1}^{T} P(Y_t)}{P(\boldsymbol{X}_{1:T})}$$

$$= \frac{\prod_{t=1}^{T} P(Y_t|\boldsymbol{X}_t) \prod_{t=1}^{T} P(\boldsymbol{X}_t) \cdot \prod_{t=1}^{T} P(Y_t)}{\prod_{t=1}^{T} P(Y_t)\,P(\boldsymbol{X}_{1:T})}$$

$$= \frac{\prod_{t=1}^{T} P(Y_t|\boldsymbol{X}_t) \prod_{t=1}^{T} P(\boldsymbol{X}_t)}{P(\boldsymbol{X}_{1:T})}. \quad (26)$$

Hence,

$$P_{\mathrm{HMM}}(Y_{1:T}|X_{1:T}) = P_f(Y_{1:T}|\boldsymbol{X}_{1:T}) \frac{\prod_{t=2}^{T} P(Y_t|Y_{t-1})}{\prod_{t=2}^{T} P(Y_t)}. \quad (27)$$

In this way, we proved that the classifiers are special cases of HMMs. When

$$\frac{\prod_{t=2}^{T} P(Y_t|Y_{t-1})}{\prod_{t=2}^{T} P(Y_t)} = 1, \quad (28)$$

HMMs reduce to classifiers (e.g., when $P(Y_t|Y_{t-1}) = P(Y_t)$ for all $t = 2, 3, \ldots, T$). Since HMMs approximate $P(Y_1, Y_2, \ldots, Y_T)$ by $P(Y_1) \prod_{t=2}^{T} P(Y_t|Y_{t-1})$ rather than $\prod_{t=1}^{T} P(Y_t)$ as is in the classifiers, HMMs are better choice than segment-based classifiers when the state sequence demonstrates strong autocorrelation.

APPENDIX IV

Appendix IV here provides a method to estimate the marginal distribution when no prior information is available. Our idea is to select a marginal probability distribution that lead to the maximum OSA detection accuracy if we apply HMMs to the training data

$$\hat{\boldsymbol{p}} = \underset{\boldsymbol{p}}{\operatorname{argmin}} ||y_{1:T} - \hat{y}_{1:T}\left(\tilde{\boldsymbol{\Theta}}\right)||_0 \quad (29)$$

where the parameter $\tilde{\boldsymbol{\Theta}} = (\boldsymbol{\pi}, A, \boldsymbol{p}, f)$. Here, $\boldsymbol{\pi}, A, f$ can be estimated by MLE as mentioned in the paper, and

$$\hat{y}_{1:T}\left(\tilde{\boldsymbol{\Theta}}\right) = \underset{\hat{y}_{1:T}}{\operatorname{argmax}} L\left(Y_{1:T} = \hat{y}_{1:T}|\boldsymbol{X}_{1:T} = \boldsymbol{x}_{1:T}, \tilde{\boldsymbol{\Theta}}\right), \quad (30)$$

in which Viterbi algorithm can be used for OSA detection given parameter $\tilde{\boldsymbol{\Theta}}$ and observations $\boldsymbol{x}_{1:T}$ [31], [32].

ACKNOWLEDGMENT

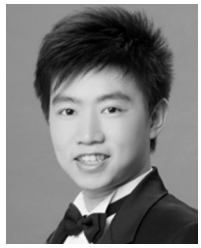
The authors would like to thank Dr. Thomas Penzel of Phillips-University for providing the data.

REFERENCES

[1] P. E. Peppard *et al.*, "Increased prevalence of sleep-disordered breathing in adults," *Amer. J. Epidemiol.*, vol. 177, no. 9, pp. 1006–1014, May 2013.
[2] M. R. Mannarino *et al.*, "Obstructive sleep apnea syndrome," *Eur. J. Intern. Med.*, vol. 23, no. 7, pp. 586–593, Oct. 2012.
[3] K. E. Bloch, "Polysomnography: A systematic review," *Technol. Health. Care*, vol. 5, no. 4, pp. 285–305, Oct. 1997.
[4] T. Penzel *et al.*, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 10, pp. 1143–1151, Oct. 2003.
[5] M. Baumert *et al.*, "Changes in RR and QT intervals after spontaneous and respiratory arousal in patients with obstructive sleep apnea," in *Proc. Comput. Cardiol.*, Durham, NC, USA, 2007, pp. 677–680.
[6] R. Hornero *et al.*, "Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 107–113, Jan. 2007.
[7] J. V. Marcos *et al.*, "Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis," *Med. Biol. Eng. Comput.*, vol. 48, no. 9, pp. 895–902, Sep. 2010.
[8] A. K. Ng *et al.*, "Snore signal enhancement and activity detection via translation-invariant wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 10, pp. 2332–2342, Oct. 2008.
[9] A. Azarbarzin and Z. Moussavi, "Automatic and unsupervised snore sound extraction from respiratory sound signals," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1156–1162, Apr. 2011.
[10] F. Senny *et al.*, "Midsagittal jaw movement analysis for the scoring of sleep apneas and hypopneas," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 87–95, Jan. 2008.
[11] A. Azarbarzin and Z. Moussavi, "Snoring sounds variability as a signature of obstructive sleep apnea," *Med. Eng. Phys.*, vol. 35, no. 4, pp. 479–485, Apr. 2013.
[12] O. Fontenla-Romero *et al.*, "A new method for sleep apnea classification using wavelets and feedforward neural networks," *Artif. Intell. Med.*, vol. 34, no. 1, pp. 65–76, May 2005.
[13] U. R. Abeyratne *et al.*, "Interhemispheric asynchrony correlates with severity of respiratory disturbance index in patients with sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2947–2955, Dec. 2010.
[14] C. Guilleminault *et al.*, "Heart rate variability, sympathetic and vagal balance and EEG arousals in upper airway resistance and mild obstructive sleep apnea syndromes," *Sleep Med.*, vol. 6, no. 5, pp. 451–457, Sep. 2005.
[15] A. H. Khandoker *et al.*, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 37–48, Jan. 2009.
[16] R. Shouldice *et al.*, "Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features." *Sleep*, vol. 27, no. 4, pp. 784–792, Jun. 2004.
[17] M. O. Mendez *et al.*, "Sleep apnea screening by autoregressive models from a single ECG lead," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2838–2850, Dec. 2009.
[18] M. Bsoul *et al.*, "Apnea medassist: Real-time sleep apnea monitor using single-lead ECG," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 416–27, May 2011.
[19] S. M. Isa *et al.*, "Sleep apnea detection from ECG signal: Analysis on optimal features, principal components, and nonlinearity," in *Proc. 5th Int. Conf. Bioinformat. Biomed. Eng.*, Wuhan, China, 2011, pp. 1–4.
[20] G. Sannino *et al.*, "An automatic rules extraction approach to support OSA events detection in an mhealth system," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1518–1524, Mar. 2014.
[21] T. Penzel *et al.*, "The apnea-ECG database," in *Proc. Computers in Cardiol.*, Cambridge, MA, USA, 2000, pp. 255–258.
[22] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
[23] G. B. Moody *et al.*, "Derivation of respiratory signals from multi-lead ECGs," *Comput. Cardiol.*, vol. 12, pp. 113–116, 1985.
[24] V. X. Afonso *et al.*, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 192–202, Feb. 1999.
[25] L. Chen *et al.*, "An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 1, pp. 106–115, Jan. 2015.
[26] P. De Chazal *et al.*, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 686–696, Jun. 2003.
[27] P. De Chazal *et al.*, "Automatic classification of sleep apnea epochs using the electrocardiogram," in *Proc. Comput. Cardiol.*, Cambridge, MA, USA, 2000, pp. 745–748.
[28] G. B. Moody *et al.*, "Clinical validation of the ECG-derived respiration (EDR) technique," *Comput. Cardiol.*, vol. 13, pp. 507–510, 1986.
[29] A. Camm *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north American society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.
[30] L. J. Gula *et al.*, "Heart rate variability in obstructive sleep apnea: A prospective study and frequency domain analysis," *Ann. Noninvasive Electrocardiol.*, vol. 8, no. 2, pp. 144–149, Apr. 2003.
[31] B. B. McShane, "Machine learning methods with time series dependence," Ph.D. dissertation, Wharton School, Univ. Pennsylvaniya, Philadelphia, PA, USA, 2010.
[32] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[33] J. A. Bilmes, "A gentle tutorial of the em algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," Int. Comput. Sci. Inst., Berkeley, CA, USA, Tech. Rep. TR-97-021, Apr. 1998.
[34] C. M. Travieso *et al.*, "Building a Cepstrum-HMM kernel for Apnea identification," *Neurocomputing*, vol. 132, pp. 159–165, May 2014.
[35] J. V. Marcos *et al.*, "Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnea diagnosis from nocturnal oximetry," *Med. Eng. Phys.*, vol. 31, no. 8, pp. 971–978, Oct. 2009.
[36] D. Alvarez *et al.*, "Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2816–2824, Jul. 2010.
[37] D. S. Morillo and N. Gross, "Probabilistic neural network approach for the detection of SAHS from overnight pulse oximetry," *Med. Biol. Eng. Comput.*, vol. 51, no. 3, pp. 305–315, Nov. 2013.
[38] J. N. McNames and A. M. Fraser, "Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram," in *Proc. Comput. Cardiol.*, Cambridge, MA, USA, 2000, pp. 749–752.

Changyue Song received the B.S. degree in industrial engineering from Tsinghua University, Beijing, China, in 2012. He received the M.S. degree in management science and engineering from Peking University, Beijing, China, in 2015. He is currently working toward the Ph.D. degree at the Department of Industrial & Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA.

His research interests include process modeling and data fusion.



Kaibo Liu (GS'13–M'14) received the B.S. degree in industrial engineering and engineering management from the Hong Kong University of Science and Technology in 2009, the M.S. degree in statistics, and the Ph.D. degree in industrial engineering from the Georgia Institute of Technology in 2011 and 2013.

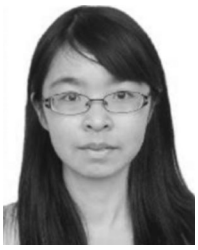
He is currently an Assistant Professor at the Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA. His research interests include data fusion for process modeling, monitoring, diagnosis, and prognostics.

Mr. Liu is a Member of the ASQ, INFORMS, and IIE.

Xi Zhang (M'15) received the B.S. degrees in mechanical engineering and automation from Shanghai Jiaotong University, Shanghai, China, in 2006, and the Ph.D. degree in industrial engineering from the University of South Florida, Tampa, FL, USA, 2010.

He is currently an Assistant Professor at the Department of Industrial Engineering and Management, Peking University, Beijing, China. His research interests include data-driven modeling and analysis for complex dynamic systems including physiological-data-driven analysis for healthcare delivery.

Dr. Zhang is a Member of the ASQ, INFORMS, and IIE.



Lili Chen received the B.S. degree in industrial engineering from Nankai University, Tianjin, China, in 2011. She is currently working toward the Ph.D. degree at the Department of Industrial Engineering and Management, Peking University, Beijing, China.

Her research interests include statistically modeling of physiological data and complex systems.

Xiaochen Xian received the B.S. degrees from the Department of Mathematics, Zhejiang University, Hangzhou, China, in 2014. She is currently working toward the Ph.D. degree at the Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA.

Her research interests include high-dimensional process monitoring and quality improvement.