This is a repository copy of *A Dirichlet process mixture model for autonomous sleep apnea detection using oxygen saturation data*.

Version: Accepted Version

Proceedings Paper:

A Dirichlet Process Mixture Model for Autonomous Sleep Apnea Detection using Oxygen Saturation Data

Zhenglin Li⋆, Mahnaz Arvaneh⋆, Heather E. Elphick†, and Ruth N. Kingshott†, Lyudmila S. Mihaylova⋆

⋆*Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK*
†*Department of Sleep Medicine, Sheffield Children's Hospital NHS Foundation Trust, Sheffield, UK*
Email: {zhenglin.li, l.s.mihaylova, m.arvaneh}@sheffield.ac.uk, {h.elphick, ruth.kingshott}@nhs.net

Abstract—**Sleep apnea is a sleep disorder which is common in many children and adults. It is characterised by abnormal breath pauses or shallow breathing during sleep. Traditional diagnosis of apnea requires special equipment for data collection in clinical conditions and manual analysis by clinicians which is expensive and time-consuming. This paper presents a framework for autonomous detection of sleep apnea, using peripheral blood haemoglobin oxygen saturation (SpO_2) data based on the fusion of multiple features and Dirichlet process mixture model. The SpO_2 signals are segmented into overlapping sub-sequences and several features are extracted from each segment. The distributions of features extracted from disorder and normal segments are modelled by two Gaussian mixture models, respectively, with the Dirichlet process as the prior. The advantage of the framework is that the number of clusters within mixture models can be learned from training data without strong assumptions, which contributes to accurate estimation of the distributions. The proposed framework is subject-independent and it is trained and tested on two publicly available databases with 10-fold cross-validation. It obtains accuracy of 84.89% on the St. Vincent's University Hospital Sleep Apnea Database and accuracy of 97.01% on the Apnea-ECG Database, outperforming state-of-the-art approaches. The results show that the proposed model is capable of representing the distributions of features independently of subjects and can accurately classify segmented signals from patients with symptoms of different severity. The results show the potential of the developed classification framework to support clinicians in their decision making.**

Index Terms—**Sleep apnea-hypopnea syndrome, oxygen saturation (SpO_2) data, Dirichlet process mixture model, classification, sleep disorder diagnostics, decision making**

I. Introduction

Sleep apnea (cessation of airflow) and hypopnea (reduction in airflow) are sleep disordered breathing (SDB) during sleep which has become a major health issue all over the world [1]. It happens if a person experiences pauses in breathing or overly shallow breathing during sleep. The prevalence of SDB in middle-aged adults is estimated to be 9% in women and 24% in men, with the apnea-hypopnea index (AHI) of 5/hr or higher [2]. Around 2% of women and 4% of men can be diagnosed with the sleep apnea and hypopnea syndrome (defined as AHI$\geq$ 5/hr with daytime hypersomnolence), which is a common cause of sleepiness and neurocognitive impairment, and is related to cardiovascular disease as well [3].

Overnight polysomnography (PSG) has been recognised as the gold standard method for a definitive diagnosis of SDB [4]. Figure 1a shows the diagram of such a system for diagnosis. However, it requires the sleeping process of patients to be monitored in laboratories with much professional equipment, of which the high cost and availability limit its application. Additionally, the obtained overnight data needs to be analysed by professional doctors or clinical technicians according to some widely accepted guidance, e.g. the American Academy of Sleep Medicine (AASM) manual [5]. The lengthy scoring process and limited number of trained medical experts results in long waiting times for a diagnosis of sleep disordered breathing. In addition, the intrusive nature of the test and sheer number of sensors that are attached make PSG an uncomfortable process for patients

To solve this problem, the study of automatic SDB detection has attracted significant attention. Research focuses on detection with fewer channels of signals or portable devices. Among them, detection based on peripheral blood haemoglobin oxygen saturation (SpO_2), electrocardiogram (ECG) and sounds have been widely studied [4]. ECG and SpO_2 based approaches have achieved more accurate detection than the sound-based methods according to the reviews [4]. As the SpO_2 signals can be measured easily by nocturnal pulse oximetry, the frameworks based on them are convenient for home sleep health monitoring [6]. The diagram in Figure 1b shows such a setup where data can be acquired both at home and in hospitals, transferred to a cloud server via the Internet and processed to support decision makers.

There are two types of SDB detection approaches: subject-based and epoch-based detection approaches, respectively [4]. The former group of approaches provides a result indicating if a patient has the sleep apnea and hypopnea syndrome based on the overnight data, while the latter ones detect the existence of SDB events in each epoch. In the epoch-based detection, segments with at least one SDB event (whole event or part of it) in them are defined as 'apnea' ones, while those without any respiratory disorder are set as 'normal' segments. This paper presents an epoch-based detection framework for nearly real-time detection of sleep disorders.

The approaches of SDB detection can be divided into three main groups, namely: *i*) clinical rule-based approaches, *ii*) features and classifiers based approaches, and *iii*) end-to-end detection approaches. A majority of the clinical rule-based methods detect the SDB events according to the rules related to decreases of SpO_2 values or decreasing durations [7], [8]. These approaches are clinically explainable, but they are usually not robust enough to individual differences.
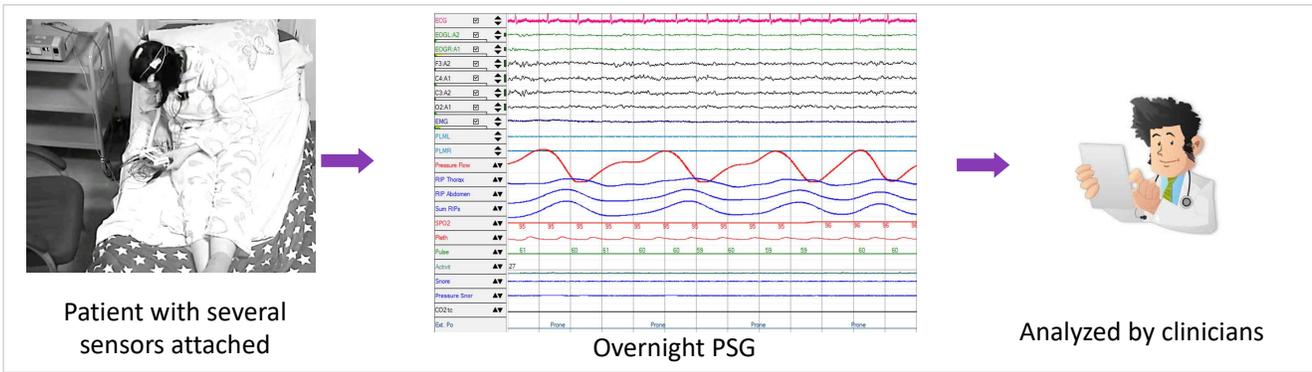
(a) Polysomnography based apnea-hypopnea detection diagram.
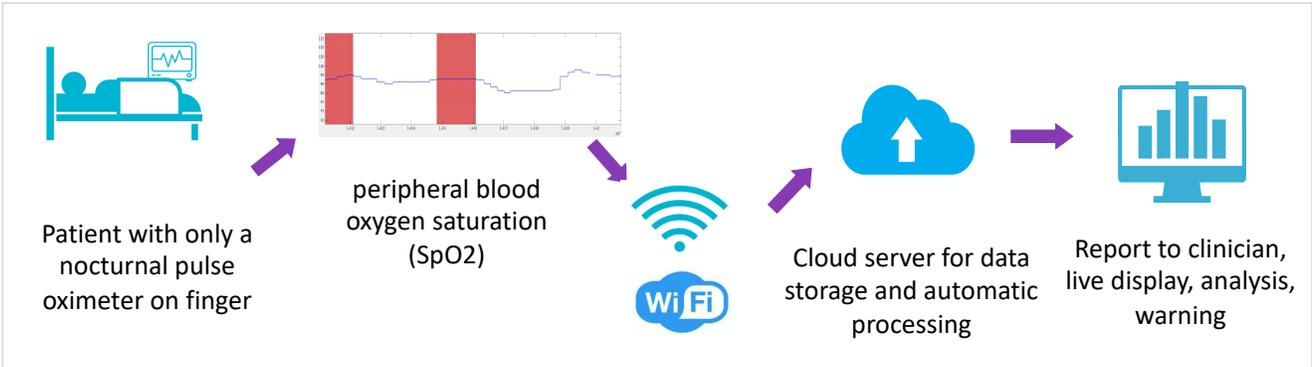
(b) Automatic sleep apnea-hypopnea event detection diagram.

Fig. 1: Diagram of sleep apnea-hypopnea detection.

The approaches using features extracted from SpO_2 signals together with classifiers have been widely applied to automatic detection of SDB events in the past decades. To describe the properties of disordered signals, various features have been extracted, such as classic oximetric indices, features in time and frequency domain, and non-linear features. Conventional features include the minimum and average baseline of SpO_2 values, oxygen desaturation indices (ODIs) and cumulative time (CT) indices. Among them, the ODIs and CT indices measure the times of SpO_2 values dropping below a certain threshold and the time percentage of signals lower than specific thresholds, respectively [7], [9]. Similarly to the rule-based approaches, these features do not work effectively in detecting sleep disorders with signals from different patients.

Time-domain features include the delta index [7] and statistics of SpO_2 values. The mean, variance, skewness and kurtosis are among the most widely used features defined in time domain [9] as they can describe the changes of SpO_2 caused by apnea or hypopnea events. Besides the time domain, features extracted in the frequency domain can also help to distinguish disordered signals from normal ones. Statistics and entropy of the power density function calculated by the discrete Fourier transform (DFT) are often employed as features. The maximum and relative power of the apnea related frequency band (0.014-0.033Hz) can work as discriminative features as well [7].

However, the features in the frequency domain are mostly used in subject-based SDB detection since long sequences are needed to achieve fine resolution in the frequency domain. Additionally, the differences between the changing patterns of disordered signals and normal ones can be described by non-linear features such as the central tendency measure (CTM), Lempel-Ziv complexity (LZC) and sample entropy [4]. The features mentioned above can be extracted from both overnight and segmented signals and classified by various algorithms, e.g. Adaboost, support vector machine (SVM), nearest neighbour (NN) algorithm, linear discriminant analysis (LDA) and logistic regression [4], [10].

In addition to the feature-and-classifier based approaches, the group of 'end-to-end' approaches has become popular recently and has shown good detection results, among which are based on deep neural networks and Gaussian process regression [11]. These methods use segmented signals as inputs directly and output decisions based on the temporally changing patterns. Pathinarupothi et al. employ long short term memory (LSTM) networks for SDB event detection [12], while a deep belief network is used by Mostafa et al. in [13]. However, these approaches may have the problem of over-fitting when the training data is limited. In [14], a state-space model is combined with a Gaussian process to model the changing patterns of oxygen saturation signals and other data for SDB detection. Also, this model is not subject independent and needs to be

trained and tested on data from the same patients.

To achieve more accurate detection of SDB events, this paper proposes a near real-time framework based on the Dirichlet Process mixture model (DPMM) and fusion of multiple features. The distributions of features extracted from both apnea and normal signals can be modelled accurately with limited training data. Specifically, the distributions are represented by Gaussian mixture models with a Dirichlet process (DP) as the prior, of which the numbers of clusters can be learned from training data. This makes the model robust to individual differences. Additionally, novel features based on the Haar wavelet transform are proposed to distinguish the changing patterns of SDB signals from normal ones. The features are extracted from overlapping segments instead of non-overlapping ones which are common in related research, to deal with the delays of SpO_2 values decreasing from the moment when SDB events happen.

The paper is organised as follows. In Section II, we briefly introduce the features extracted from SpO_2 signals for further processing. Subsequently, the DP background knowledge is provided in Section III. Section IV presents the novel framework based on the DPMM and multiple features for automatic SDB detection, while the corresponding results and discussions are provided in Section V. In Section VI we summarise the main results and ideas for future work.

II. FEATURES EXTRACTED FROM SpO_2 SIGNALS

To achieve nearly real-time SDB event detection, SpO_2 signals are segmented into overlapping sub-sequences of the same length in the proposed framework, and multiple features are extracted from each of them for further processing. In the next subsection we describe the features used in the classification process, which are defined in time and wavelet domain.

A. Time-domain Features

1) Variance: The variance of each segment in the time domain can be calculated and employed as a feature, because an SDB segment tends to have a larger variance compared with normal subsequences [9].

2) Range: The range is calculated as the difference between the maximum and minimum signal values within a segment. The SpO_2 values usually decrease after an SDB event while those in a normal segment do not change much. Therefore, the range is chosen as a feature for SDB detection.

3) Minimum Value: As mentioned before, SpO_2 values usually decrease to relatively low levels after SDB events, which seldom happens in normal sleeping periods.

B. Wavelet Transform Based Features

The discrete wavelet transform (DWT) can capture both frequency and temporal location information, contributing to its advantages over the discrete Fourier transform (DFT) which has been widely used for frequency domain features. To describe the changing patterns of SpO_2 signal segments, the DWT can extract features of frequency with a fine resolution in time domain. Figure 2 shows the plots of both 'apnea' and 'normal' SpO_2 segments, together with their corresponding detail coefficients (from the high-pass filter) of level 2 and 3 obtained by Haar wavelet transform [15]. The Haar wavelet is chosen because its step-like wavelet function is suitable for describing the common abrupt changes of SpO_2 during apnea events. The SpO_2 values usually decline continuously to a low level and then recover to normal subsequently when an SDB event happens, while in a 'normal' segment the SpO_2 signal keeps stable values or fluctuates in a narrow range. These differences result to different coefficients of an 'apnea' segment and a 'normal' one. This means that many peaks are observed in the wavelet detail coefficients of 'apnea' segments, while only a limited number of peaks are observed in those of segments corresponding to 'normal' breath patterns, as shown in Figure 2. In the wavelet domain, we use the following features:

1) Number of Points With Large Wavelet Coefficients: In the level 2 and 3 detail coefficients from the Haar wavelet transform, the number of points larger than thresholds can be employed as features to distinguish 'apnea' segments from 'normal' ones of SpO_2.

2) Mean Energy of Wavelet Coefficients: Similarly, the mean energy of the wavelet detail coefficients is different between the 'apnea' and the 'normal' segments. This can be used as another distinguishing feature. The mean energy f_e of a segment is calculated as

$$f_e = \frac{1}{M} \sum_{i=1}^{M} d(i)^2, \tag{1}$$

where $d(i)$ is the i-th element of the wavelet coefficients of a segment and M is the length of the coefficients. In the experiment, only the mean energy of level 2 detail coefficients is used for computational efficiency because employing the mean energies of both level 2 and 3 only leads to small improvement on detection performance.

III. THE DIRICHLET PROCESS MODEL

In this section, the definition of the DP and a constructing scheme of a DP, called the stick-breaking process are briefly described before introducing the proposed framework.

A. Model Setting and Definition of Dirichlet Process

Within the DP framework, it is assumed that each observation $\mathbf{x}_i$ is generated from a distribution with parameter(s) $\boldsymbol{\theta}_i$. The parameter $\boldsymbol{\theta}_i$ is generated from a prior distribution G, which can be set as a DP. Thus, the model is as follows:

$$\boldsymbol{\theta}_i | G \sim G \qquad \text{for each } i, \tag{2}$$
$$\mathbf{x}_i | \boldsymbol{\theta}_i \sim F(\boldsymbol{\theta}_i) \qquad \text{for each } i, \tag{3}$$

where $F(\boldsymbol{\theta}_i)$ is the distribution of $\mathbf{x}_i$ given $\boldsymbol{\theta}_i$ and different $\boldsymbol{\theta}_i$s are not necessarily of distinct values.

A DP is defined as a distribution of a probability measure G over a measurable space [16], satisfying the condition that for any finite measurable partition $(A_1, ..., A_K)$ of the

(a) A 'normal' segment of SpO₂ signal.

(b) An 'apnea' segment of SpO₂ signal.

(c) Level 2 detail coefficients of Haar wavelet transform of a 'normal' segment.

(d) Level 2 detail coefficients of Haar wavelet transform of an 'apnea' segment.

(e) Level 3 detail coefficients of Haar wavelet transform of a 'normal' segment.

(f) Level 3 detail coefficients of Haar wavelet transform of an 'apnea' segment.

Fig. 2: 'Normal' and 'apnea' segments of SpO₂ signal and their corresponding level 2 and level 3 detail coefficients of the wavelet transform.

space, $(G(A_1), ..., G(A_K))$ follows a Dirichlet distribution with parameters of $(\alpha_0 G_0(A_1), ..., \alpha_0 G_0(A_K))$, where G_0 is a probability measure in the measurable space and α_0 is a positive real parameter [17]. It can be written as:

$$(G(A_1), ..., G(A_K) \sim \text{Dir}(\alpha_0 G_0(A_1), ..., \alpha_0 G_0(A_K)). \quad (4)$$

When G follows a Dirichlet process, it is denoted as $G \sim DP(\alpha_0, G_0)$ with a concentration parameter α_0 and a base distribution G_0.

B. The Stick-breaking Process

The stick-breaking process provides a way of constructing a DP as below

$$\beta_k \sim \text{Beta}(1, \alpha_0), \quad (5)$$

$$\boldsymbol{\theta}_k^* \sim G_0, \quad (6)$$

$$\pi_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_l) = \beta_k \left(1 - \sum_{l=1}^{k-1} \pi_l\right), \quad (7)$$

$$G = \sum_{k=1}^{\infty} \pi_k \boldsymbol{\theta}_k^*, \quad (8)$$

where $\boldsymbol{\pi} = \{\pi_k\}_{k=1}^{\infty}$ is a sequence of mixture weights, $\boldsymbol{\theta}^*$s are distinct latent parameters drawn from G_0, and G is the constructed DP [17], [18]. Here k denotes the index of a component. The distribution over $\boldsymbol{\pi}$ can also be expressed as $\boldsymbol{\pi} \sim \text{GEM}(\alpha_0)$, which comes from the initials of Griffiths, Engen and McCloskey [17].

IV. The New Framework for Automatic Sleep Apnea Detection Based on a Dirichlet Process Mixture Model and Multiple Features

Several features are extracted and combined from each segment of SpO$_2$ signals, including both 'apnea' and 'normal' ones. The distributions of features extracted from disordered and normal segments are believed to be different, thus decisions can be made by comparing the probabilities of each testing segment generated from the models of 'apnea' and 'normal'.

The distributions of features from 'apnea' and 'normal' segments can be modelled by two Gaussian mixture models (GMM), as a GMM can approximate any distribution accurately by setting a proper component number and adjusting its parameters. The model settings of the two GMMs are the same and the only difference is in the training data.

Denote the features extracted from the i-th segment as $\mathbf{x}_i$. Its distribution can be expressed as

$$p(\mathbf{x}_i) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_i; \boldsymbol{\theta}_k^*), \qquad (9)$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution and the parameters of the k-th component of the mixture model are denoted as $\boldsymbol{\theta}_k^* \triangleq \{\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*\}$. Here $\boldsymbol{\mu}_k^*$ is the mean vector and $\boldsymbol{\Sigma}_k^*$ is the variance matrix of the k-th Gaussian component.

The theory of mixture models [19], [20] assumes that each $\mathbf{x}_i$ is generated by first choosing a cluster indexed by an assignment variable z_i according to a categorical distribution of $\boldsymbol{\pi} = [\pi_1, ..., \pi_K]$. Then the observation $\mathbf{x}_i$ is generated from the chosen component with the parameter $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{z_i}^*$. However, the number of components K and distribution weight $\boldsymbol{\pi}$ are not available, when we only have the observations. Instead of setting the cluster number K empirically, the proposed framework sets the prior as a DP to solve this problem. Combined with the stick-breaking process defined in Section III-B, the generative model can be represented as follows

$$\boldsymbol{\pi} \sim \text{GEM}(\alpha_0), \qquad (10)$$
$$\boldsymbol{\theta}_k^* \sim G_0, \qquad (11)$$
$$z_i \sim \boldsymbol{\pi}, \qquad (12)$$
$$\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\theta}_{z_i}^*), \qquad (13)$$

where $\{\boldsymbol{\theta}_k^*\}_{k=1}^{\infty}$ are distinct values of the parameters $\boldsymbol{\theta}_k^*$s, sampled independently from the base distribution $G_0(\boldsymbol{\theta}^*|\lambda)$ (λ is the hyper-parameter of G_0) and the distribution of $\boldsymbol{\pi}$ is shown in Eq. (7).

Then the generative model can be interpreted as an 'infinite mixture model', which does not have a fixed number of clusters. Instead, the cluster number may increase with more training data given. In practical applications, the number of clusters cannot be infinite, with a limited number of observations. The model will have a finite number of clusters, which can be learned from training data by Bayesian inference.

Denote $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k^*\}_{k=1}^{\infty}$ and $\boldsymbol{\beta} = \{\beta_k\}_{k=1}^{\infty}$ as the sets of variables $\boldsymbol{\theta}_k^*$s and β_ks, respectively. The random variables β_ks are drawn independently from a Beta distribution as defined in Eq. (5). Let $\mathbf{z} = \{z_i\}_{i=1}^{N}$ be the cluster assignment variables of N training features $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$ and $\mathbf{W} = \{\boldsymbol{\beta}, \boldsymbol{\Theta}, \mathbf{z}\}$ be the collection of all the latent variables.

Given the features $\mathbf{X}$ for training and a new sample $\mathbf{x}'$ for testing, the probability of $\mathbf{x}'$ being generated from the trained model can be derived as

$$p(\mathbf{x}'|\mathbf{X})$$
$$= \int p(\mathbf{x}'|z', \mathbf{W}, \mathbf{X}) p(z'|\mathbf{W}, \mathbf{X}) p(\mathbf{W}|\mathbf{X}) dz' d\mathbf{W} \qquad (14)$$
$$= \int p(\mathbf{x}'|z', \boldsymbol{\beta}, \boldsymbol{\Theta}, \mathbf{z}, \mathbf{X}) p(z'|\boldsymbol{\beta}, \boldsymbol{\Theta}, \mathbf{z}, \mathbf{X}) p(\mathbf{W}|\mathbf{X}) dz' d\mathbf{W}$$
$$\qquad (15)$$
$$= \int p(\mathbf{x}'|z', \boldsymbol{\Theta}) p(z'|\boldsymbol{\beta}) p(\mathbf{W}|\mathbf{X}) dz' d\mathbf{W} \qquad (16)$$
$$= \int p(\mathbf{x}'|\boldsymbol{\theta}_{z'}^*) p(z'|\boldsymbol{\beta}) p(\mathbf{W}|\mathbf{X}) dz' d\mathbf{W}, \qquad (17)$$

where z' is the assignment variable of the testing data $\mathbf{x}'$.

The first term $p(\mathbf{x}'|\boldsymbol{\theta}_{z'}^*)$ in Eq. (17) can be calculated according to (13), while $p(z'|\boldsymbol{\beta})$ can be calculated by (7) and (12). However, the posterior of $\mathbf{W}$ is difficult to be calculated analytically because of the integral in the denominator of

$$p(\mathbf{W}|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{W})p(\mathbf{W})}{\int p(\mathbf{X}|\mathbf{W})p(\mathbf{W})d\mathbf{W}}. \qquad (18)$$

Therefore, a variational distribution in Eq. (19) is employed to approximate the posterior $p(\mathbf{W}|\mathbf{X})$. It is designed as a family of factorized distributions according the idea of mean-field variational inference [21]

$$q(\mathbf{W}; \phi) = \prod_{k=1}^{K} [q(\beta_k; \phi_k^\beta) \ q(\boldsymbol{\theta}_k^*; \phi_k^{\theta^*})] \prod_{i=1}^{N} q(z_i), \qquad (19)$$

where $q(z_i)$s are categorical distributions, and ϕ_k^β and $\phi_k^{\theta^*}$ are parameters of distributions $q(\beta_k)$ and $q(\boldsymbol{\theta}_k^*)$, with $\phi_k = \{\phi_k^\beta, \phi_k^{\theta^*}\}$. Moreover, K is not a fixed number in Eq. (19) but can increase with more training data, corresponding to the infinite mixture model of the distribution. It is assumed that all the parameters ϕ_k are tied with $k > T^*(T^* \ll K)$ and equivalent to the prior.

The variational distributions are assumed to be in an exponential family which is a common choice within the Bayesian nonparametric framework, because of the availability of their analytical solutions. Specifically, it is assumed that

$$p(\beta_k|\alpha) = \text{Beta}(\alpha_1, \alpha_2), \qquad (20)$$
$$q(\beta_k; \phi_k^\beta) = \text{Beta}(\phi_{k,1}^\beta, \phi_{k,2}^\beta), \qquad (21)$$
$$p(\mathbf{x}|\boldsymbol{\theta}^*) = h(\mathbf{x}) \exp\{(\boldsymbol{\theta}^*)^\mathsf{T} \mathbf{x} - a(\boldsymbol{\theta}^*)\}, \qquad (22)$$
$$p(\boldsymbol{\theta}^*|\lambda) = h(\boldsymbol{\theta}^*) \exp\{\lambda_1 \boldsymbol{\theta}^* + \lambda_2[-a(\boldsymbol{\theta}^*)] - a(\lambda)\}, \qquad (23)$$
$$q(\boldsymbol{\theta}_k^*; \phi_k^{\theta^*}) = h(\boldsymbol{\theta}_k^*) \exp\{\phi_{k,1}^{\theta^*} \boldsymbol{\theta}_k^* + \phi_{k,2}^{\theta^*}[-a(\boldsymbol{\theta}_k^*)] - a(\phi_k^{\theta^*})\}, \qquad (24)$$

where $\alpha = \{\alpha_1, \alpha_2\}$, $\lambda = \{\lambda_1, \lambda_2\}$ are hyperparameters of the prior, and $\phi_k^\beta = \{\phi_{k,1}^\beta, \phi_{k,2}^\beta\}$, $\phi_k^{\theta^*} = \{\phi_{k,1}^{\theta^*}, \phi_{k,2}^{\theta^*}\}$ are variational parameters to be optimized. $a(\cdot)$ is the logarithmic

normalizer in the definition of exponential family, ensuring the distribution integrates to one [22].

The probability $q(z_i = k)$ can be calculated as

$$q(z_i = k) = \frac{\exp(E_{i,k})}{\sum_{j=1}^{\infty} \exp(E_{i,j})}, \qquad (25)$$

where $E_{i,k}$ is defined as

$$E_{i,k} = \mathbb{E}_{q_\beta}[\log p(z_i = k|\boldsymbol{\beta}))] + \mathbb{E}_{q_{\theta_k^*}}[\log p(\mathbf{x}_i|\boldsymbol{\theta}_k^*)]. \qquad (26)$$

Other parameters are updated as

$$\phi_{k,1}^\beta = \alpha_1 + \sum_{i=1}^{N} q(z_i = k), \;\; \phi_{k,2}^\beta = \alpha_2 + \sum_{i=1}^{N} \sum_{j=k+1}^{\infty} q(z_i = j), \qquad (27)$$

$$\phi_{k,1}^{\theta^*} = \lambda_1 + \sum_{i=1}^{N} q(z_i = k)\mathbf{x}_i, \;\; \phi_{k,2}^{\theta^*} = \lambda_2 + \sum_{i=1}^{N} q(z_i = k). \qquad (28)$$

The latent variable z_i, and variational parameters ϕ_k^β and $\phi_k^{\theta^*}$, are updated iteratively by Eqs. (25), (27) and (28) until the free energy is minimised. Details of the derivation of the variational inference are available in [23].

Then $p(\mathbf{x}'|\mathbf{X})$ can be rewritten as:

$$p(\mathbf{x}'|\mathbf{X}) = \int p(\mathbf{x}'|\boldsymbol{\theta}_{z'}^*)p(z'|\boldsymbol{\beta})q(\mathbf{W};\phi)dz'd\mathbf{W} \qquad (29)$$

$$= \int p(\mathbf{x}'|\boldsymbol{\theta}_{z'}^*)p(z'|\boldsymbol{\beta}) \prod_{k=1}^{K}[q(\beta_k;\phi_k^\beta) \; q(\boldsymbol{\theta}_k^*;\phi_k^{\theta^*})]$$

$$\prod_{i=1}^{N} q(z_i) \; dz' \; d\beta \; d\boldsymbol{\theta}^* d\mathbf{z}, \qquad (30)$$

which is possible to be calculated analytically.

Given two sets of training features $\mathbf{X}^1 = \{\mathbf{x}_i^1\}_{i=1}^{N_1}$ and $\mathbf{X}^0 = \{\mathbf{x}_i^0\}_{i=1}^{N_0}$ extracted from 'apnea' and 'normal' segments, respectively, the probabilities of $\mathbf{x}'$ generated by the two models can be calculated according to Eq. (30). It is classified as an 'apnea' segment if

$$\log \frac{p(\mathbf{x}'|\mathbf{X}^1)}{p(\mathbf{x}'|\mathbf{X}^0)} \geq c, \qquad (31)$$

where c is a threshold influencing the balance of sensitivity and specificity. Otherwise, the segment is recognised as 'normal'. The parameter c is set as 0 in the experiments of this paper.

V. PERFORMANCE VALIDATION AND EVALUATION

A. Sleep Apnea Datasets

The data for the experiments are from two publicly available databases, the St. Vincent's University Hospital Sleep Apnea Database [24] and the Apnea-ECG Database [25], respectively. The former database contains 25 full overnight polysomnograms, including the SpO_2 signals employed in this paper, while the Apnea-ECG Database has 8 recordings containing SpO_2 signals. The data of both databases are from adults.
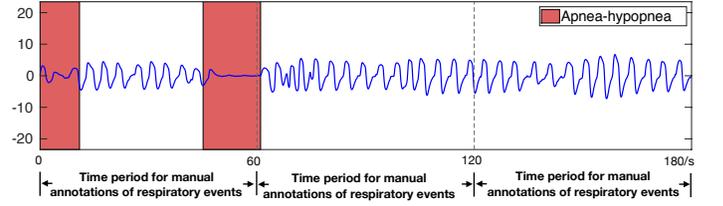
(a) Oronasal airflow signal and the segments for manual annotations of respiratory events.
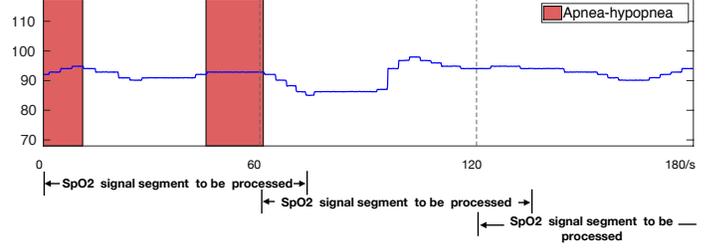
(b) SpO_2 signal and the segments to be processed and classified by the proposed framework.

Fig. 3: Diagrams of segments of signals for manual annotations by clinicians and being processed by the proposed framework automatically.

B. Data Preprocessing

The oximetry saturation signals are divided into overlapping segments of an equal size, instead of into non-overlapping ones which are commonly used. This choice is made due to the delay present between an SDB event and the corresponding changes of SpO_2 values. Figure 3 shows graphs of SpO_2, oronasal signals of three consecutive segments, and the ground truth of respiratory events. Figure 3a shows the oronasal airflow signals measured by thermistor based on which clinicians score SDB events according to the rules in [5], while Figure 3b provides the corresponding SpO_2 signals of Figure 3a. From Figure 3b we can see that the decreases of SpO_2 happen a few seconds after the SDB events which are marked in red shades in Figure 3. If the classification of segments is performed by detecting the changing patterns of SpO_2 signals, the second segment will be classified as an 'apnea' one instead of the first one, resulting in classification errors. To solve this problem, the proposed framework processes overlapping segments of signals. Specifically, the consecutive segments of signals have overlaps in the experiments, while the ground truth of SDB is manually marked by assigning a label to each minute based on the oronasal airflow signals as shown in Figure 3.

The annotations of the Apnea-ECG Database are provided by minutes, which can be used directly, while those of the St. Vincent's University Hospital Sleep Apnea Database need to be changed into annotations for each segment. The annotation transferring is conducted in a similar way as in [13]. The minutes in which SDB events happened are defined as 'apnea' while the ones without any disordered breathing are marked as 'normal'. If one apnea event happens across two successive minutes, the one with a consecutive disordered period shorter

than five seconds in it is annotated as 'normal'. Otherwise, the corresponding minute(s) will be marked as 'apnea'.

Poor contact of a nocturnal pulse oximetry caused by body movement may result in artefacts. SpO_2 values lower than 50 are treated as artefacts and the segments with such data points in them will not be used for training or testing. Besides, the signals of the Apnea-ECG dataset are downsampled to the sampling rate of 10Hz, to tackle the big difference in the sampling rates of the two datasets (Apnea-ECG 100Hz, St. Vincent's University Hospital Sleep Apnea Database 8Hz).

C. Performance Criteria

There are 2602 'apnea' segments and 7541 'normal' ones in the St. Vincent's University Hospital Sleep Apnea Database, while 1456 'apnea' segments and 2267 'normal' ones in the Apnea-ECG Database. The proposed framework is trained and tested on the two publicly available databases, independently with 10-fold cross-validation.

The performance of the proposed framework is measured by accuracy, sensitivity and specificity, defined as:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP},$$
$$Sensitivity = \frac{TP}{TP + FN},$$
$$Specificity = \frac{TN}{TN + FP},$$

where TP, FN, TN and FP denote the true positive, false negative, true negative and false positive numbers of segments, respectively [26].

D. Results and Discussion

The proposed framework achieves better performance compared with the results reported by the state-of-the-art methods tested on the same datasets, e.g. an approach combining conventional features [10] and an approach using a deep neural network [13]. The results on two datasets are given in Table I and Table II, respectively. We also present the results of the support vector machine (SVM) approach and the linear discriminant analysis (LDA) using the same features for comparison. Additionally, the proposed model is also trained and tested on features extracted from segments of different length (with the same minute-wise ground truth), which is shown in Table III and Table IV, to explore the influence of overlaps. All the results in Table I to Table IV are average scores across the testing folds of 10-fold cross-validation.

From the results in Table I and Table II, it can be seen that the best accuracy is achieved by the proposed framework with features extracted from 20 second overlapping segments. On the St. Vincent's University Hospital Sleep Apnea Database, the accuracy of the proposed framework has improvement of more than 1% and 2% over SVM and LDA, respectively, while on the Apnea-ECG dataset, all three methods achieve similar performance.

As patients suffer from sleep apnea and hypopnea syndrome of different severities, the features, such as decreasing levels

TABLE I: Results of St. Vincent's University Hospital Sleep Apnea Database

Method	Segment overlap(s)	Accuracy	Sensitivity	Specificity
LDA	20	82.68%	73.69%	85.79%
SVM	20	83.78%	50.46%	95.28%
Proposed framework	20	**84.92**%	62.65%	92.60%

TABLE II: Results of Apnea-ECG Database

Method	Segment overlap(s)	Accuracy	Sensitivity	Specificity
LDA	20	96.96%	97.04%	96.91%
SVM	20	**97.12**%	96.01%	97.84%
Proposed framework	20	97.01%	95.18%	98.19%

TABLE III: Results of the proposed framework with different segment length on St. Vincent's University Hospital Sleep Apnea Database

Segment overlap (seconds)	Accuracy	Sensitivity	Specificity
0	81.78%	63.27%	88.16%
10	84.09%	65.00%	90.68%
15	84.89%	62.11%	92.75%
20	**84.92**%	62.65%	92.60%
25	84.84%	66.74%	91.08%

TABLE IV: Results of the proposed framework with different segment length on Apnea-ECG dataset

Segment overlap (seconds)	Accuracy	Sensitivity	Specificity
0	96.32%	92.86%	98.55%
10	96.41%	93.08%	98.55%
15	96.45%	93.68%	98.24%
20	**97.01**%	95.18%	98.19%
25	96.85%	95.17%	97.92%

of SpO_2 and lasting time of events, vary with individuals, which results in misclassification errors. Besides, it is quite challenging to detect SDB events in signals from patients with mild symptoms. To solve the problems, the Dirichlet process mixture model can take into account the features reflecting different severity of SDB and individual differences when estimating the distributions, by 'learning' the Gaussian components from various data. Compared with the Apnea-ECG dataset, the individual differences are well obvious in the St. Vincent's University Hospital Sleep Apnea Database, resulted from a larger patient number and more patients of mild symptoms. Therefore, SDB detection on the St. Vincent's University Hospital Sleep Apnea Database is more challenging and the results of the three models have larger differences compared with those on Apnea-ECG dataset.

Additionally, the accuracy is improved with the increase of the overlap between segments and reaches the highest value at 20 seconds, as illustrated in Tables III and IV, since overlapping segments consider the delay of SpO_2 declines from SDB events and include more information as well.

VI. CONCLUSIONS

This paper proposes a framework for automatic SDB event detection using SpO_2 signals. It models the distributions of multiple features extracted from 'apnea' signal segments and

'normal' ones with two GMMs, respectively. All the parameters, including cluster numbers, can be learned from the training data by setting the prior of GMMs as DPs. In this way, the distributions are estimated accurately by avoiding the errors resulted from improperly set cluster numbers. Additionally, novel features based on Haar wavelet transform are proposed in this paper to describe the different changing patterns of 'apnea' and 'normal' segments. The experimental results of two publicly available databases show that the proposed framework achieves better performance than state-of-the-art methods.

Since some patients experience no or mild decreases of SpO_2 after SDB events, more accurate detection results are expected to be achieved by fusing other signals, including EEG and ECG, together with the oximetry data. While this work fuses features from the same type of signals - for oxygen saturation, our future work will focus on automatic detection of sleep disorders based on the fusion of SpO_2 and ECG data. Additionally, the proposed framework detects the existence of sleep disordered breathing without differentiating the central and obstructive apneas. Research on automatically examining the respiratory effort bands will be carried out in the future to provide more information for clinicians to determine the appropriate treatment for patients.

REFERENCES

[1] T. V. Steenkiste, W. Groenendaal, D. Deschrijver, and T. Dhaene, "Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2354–2364, 2018.

[2] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New England J. of Medicine*, vol. 328, no. 17, pp. 1230–1235, 1993.

[3] A. Jordan and R. D. McEvoy, "Gender differences in sleep apnea: epidemiology, clinical presentation and pathogenic mechanisms," *Sleep Medicine Reviews*, vol. 7, no. 5, pp. 377–389, 2003.

[4] F. Mendonca, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, "A review of obstructive sleep apnea detection approaches," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 825–837, 2018.

[5] R. B. Berry, C. L. Albertario, S. M. Harding, R. M. Lloyd, D. T. Plante, S. F. Quan *et al.*, *The AASM manual for the scoring of sleep and associated events : Rules, Terminology and Tech. Specifications*, Amer. Academy of Sleep Medicine, 2018.

[6] S. H. Hwang, J. G. Cho, B. H. Choi, H. J. Baek, Y. J. Lee, D. Jeong *et al.*, "Real-time automatic apneic event detection using nocturnal pulse oximetry," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 706–712, 2017.

[7] P. I. Terrill, "A review of approaches for analysing obstructive sleep apnoea-related patterns in pulse oximetry data," *Respirology*, 2019, early access.

[8] Y. Lee, M. Bister, P. Blanchfield, and Y. Salleh, "Automated detection of obstructive apnea and hypopnea events from oxygen saturation signal," in *Proc. IEEE Annu. Int. Conf. on Eng. in Medicine and Biology Society*, vol. 1, San Francisco, CA, USA, 2004, pp. 321–324.

[9] D. Alvarez, R. Hornero, J. V. Marcos, and F. d. Campo, "Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2816–2824, 2010.

[10] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," *IEEE Trans. Inform. Technol. in Biomedicine*, vol. 16, no. 3, pp. 469–477, 2012.

[11] C. E. Rasmussen, "Gaussian processes in machine learning," in *Proc. Summer School on Mach. Learning*, 2003, pp. 63–71.

[12] R. K. Pathinarupothi, E. S. Rangan, E. Gopalakrishnan, R. Vinaykumar, K. Soman, and K. P. Soman, "Single sensor techniques for sleep apnea diagnosis using deep learning," in *Proc. IEEE Int. Conf. on Healthcare Inform.*, Park City, UT, USA, 2017, pp. 524–529.

[13] S. S. Mostafa, F. Mendonça, F. Morgado-Dias, and A. Ravelo-García, "SpO2 based sleep apnea detection using deep learning," in *Proc. IEEE Int. Conf. on Intell. Eng. Syst.*, Larnaca, Cyprus, 2017, pp. 91–96.

[14] S. Gutta, Q. Cheng, H. D. Nguyen, and B. A. Benjamin, "Cardiorespiratory model-based data-driven approach for sleep apnea detection," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1036–1045, 2017.

[15] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge university press, 2000.

[16] V. I. Bogachev, *Measure Theory*. Springer Science & Business Media, 2007, vol. 1.

[17] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statistical Assoc.*, vol. 101, no. 476, pp. 1566–1581, 2006.

[18] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Anal.*, vol. 1, no. 1, pp. 121–143, 2006.

[19] Z. Li, L. S. Mihaylova, O. Isupova, and L. Rossi, "Autonomous flame detection in videos with a Dirichlet process Gaussian mixture color model," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1146–1154, 2017.

[20] S. J. Gershman and D. M. Blei, "A tutorial on Bayesian nonparametric models," *J. Math. Psychology*, vol. 56, no. 1, pp. 1–12, 2012.

[21] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. of the Amer. Statistical Assoc.*, vol. 112, no. 518, pp. 859–877, 2017.

[22] B. Efron, "The geometry of exponential families," *The Ann. of Statist.*, vol. 6, no. 2, pp. 362–376, 1978.

[23] K. Kurihara, M. Welling, and N. Vlassis, "Accelerated variational dirichlet process mixtures," in *Proc. Advances in Neural Inform. Process. Syst.*, Vancouver, BC, Canada, 2006, pp. 761–768.

[24] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[25] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The apnea-ECG database," in *Proc. IEEE Conf. on Comput. in Cardiology*, Boston, MA, USA, 2000, pp. 255–258.

[26] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Lett.*, vol. 27, no. 8, pp. 861–874, 2006.