IOP Publishing Physiol. Meas. 43 (2022) 015002 https://doi.org/10.1088/1361-6579/ac45e1

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
3 September 2021

REVISED
13 December 2021

ACCEPTED FOR PUBLICATION
22 December 2021

PUBLISHED
31 January 2022

PAPER

Automatic detection of respiratory events during sleep from Polysomnography data using Layered Hidden Markov Model

Azadeh Sadoughi[1], **Mohammad Bagher Shamsollahi**[2] **and Emad Fatemizadeh**[2]

[1] Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2] Biomedical Signal and Image Processing Laboratory (BiSIPL), School of Electrical Engineering, Sharif University of Technology, Tehran, Iran

E-mail: mbshams@sharif.edu

Keywords: Sleep apnea, Respiratory Event Related Arousals, Respiratory Disturbance Index, Hidden Markov Model, Event detection, Polysomnography

Abstract

Objective. Sleep apnea is a serious respiratory disorder, which is associated with increased risk factors for cardiovascular disease. Many studies in recent years have been focused on automatic detection of sleep apnea from polysomnography (PSG) recordings, however, detection of subtle respiratory events named Respiratory Event Related Arousals (RERAs) that do not meet the criteria for apnea or hypopnea is still challenging. The objective of this study was to develop automatic detection of sleep apnea based on Hidden Markov Models (HMMs) which are probabilistic models with the ability to learn different dynamics of the real time-series such as clinical recordings. *Approach*. In this study, a hierarchy of HMMs named Layered HMM was presented to detect respiratory events from PSG recordings. The recordings of 210 PSGs from Massachusetts General Hospital's database were used for this study. To develop detection algorithms, extracted feature signals from airflow, movements over the chest and abdomen, and oxygen saturation in blood (SaO_2) were chosen as observations. The respiratory disturbance index (RDI) was estimated as the number of apneas, hypopneas, and RERAs per hour of sleep. *Main results*. The best F1 score of the event by event detection algorithm was between 0.22 ± 0.16 and 0.70 ± 0.08 for different groups of sleep apnea severity. There was a strong correlation between the estimated and the PSG-derived RDI ($R^2 = 0.91, p < 0.0001$). The best recall of RERA detection was achieved 0.45 ± 0.27. *Significance*. The results showed that the layered structure can improve the performance of the detection of respiratory events during sleep.

1. Introduction

Sleep apnea is a life-long condition leading to an increased risk of hypertension (Calhoun and Harding 2010), obesity (Ogilvie and Patel 2017), depression (Nutt *et al* 2008), and cardiovascular diseases (Suzuki *et al* 2009). Sleep apnea is commonly associated with apneas, defined as >90% reduction in airflow for >10 s (Berry *et al* 2018), and hypopneas, defined as a reduction in airflow >30% for >10 s with >3% or 4% oxygen desaturation or cortical arousal (Berry *et al* 2018). The severity of sleep apnea is clinically quantified by Apnea/Hypopnea Index (AHI), which is calculated as the sum of the number of apneas and hypopneas per hour of sleep. AHI is known to predict hypertension, mortality, and low quality of life (Malhotra *et al* 2021). AHI only accounts for events with >30% reduction in airflow; however, there are respiratory events named Respiratory Event Related Arousal (RERA) with lower level of reduction or flattening in airflow that lasts >10 s leading to an arousal from sleep. The term Respiratory Disturbance Index (RDI) is another index similar to AHI that accounts for the number of apneas, hypopneas, and RERAs per hour of sleep (Berry *et al* 2018).

Since RERA ends with an arousal, it is associated with a marked surge in cardiac sympathetic modulation. A study showed individuals with high RERA index and even with low or normal AHI are still exposed to elevated sympathetic tone during sleep with significantly greater effect in females (Chandra *et al* 2013, Park *et al* 2020).

RERA can induce significant physiological changes (Guilleminault *et al* 1993, Calero *et al* 2006), alter the quality of life (Pépin *et al* 2012) that may progress to more severe respiratory events, and cardiovascular morbidity (Pépin *et al* 2012).

The clinical gold standard for identifying the severity of sleep apnea is Polysomnography (PSG), which requires visual scoring of sleep more than 20 recordings of a sleep test by technicians. Therefore, PSG is costly, inconvenient, and time-consuming with a long waiting list (Chesson *et al* 1997). To address the challenges of manual annotation of sleep tests, many studies have presented algorithms for the automatic detection of respiratory events using PSG recordings (Pombo *et al* 2017, Thorey *et al* 2019). For this purpose, previous studies have proposed various data processing techniques such as thresholding (Nakano *et al* 2007, Saha *et al* 2020), or developing mathematical detection models including K-nearest neighbor (Sharma and Sharma 2016, TİMUŞ and BOLAT 2017), support vector machines (Khandoker *et al* 2008, Almazaydeh *et al* 2012), deep neural network (Pourbabaee *et al* 2019, Hafezi *et al* 2020), Hidden Markov Model (HMM) (Travieso *et al* 2011, Song *et al* 2015). The majority of the proposed algorithms were validated through estimating AHI based on the detected respiratory events and comparing the estimated AHI to the PSG-derived AHI (Issa *et al* 1993, BaHammam *et al* 2011, Xie and Minn 2012, Pourbabaee *et al* 2019). A few studies have reported the accuracy of their algorithms in detecting every single apnea and hypopnea (Hafezi *et al* 2020, Saha *et al* 2020).

Despite the potential clinical outcomes of RERA, a limited number of studies have addressed the detection of RERA events (Ayappa *et al* 2000, Baisch *et al* 2007, Masa *et al* 2009, Nassi 2021), presumably due to the fact that in many sleep studies scoring RERAs are optional and most laboratories do not score them (Berry *et al* 2018). Among the studies that detected RERAs, Baisch *et al* extracted the shape and amplitude of airflow signal to detect RERAs and reported modest correlation ($r = 0.58$) between estimated and PSG generated RERA-indices (Baisch *et al* 2007). Ayappa *et al* used a nasal cannula/pressure transducer for recording airflow, which was analyzed to detect RERA, apnea, and hypopnea (Ayappa *et al* 2000). They reported a strong intra-class correlation coefficient of 0.96 between two scorers of the nasal cannula. Only in one study conducted by Nassi *et al*, the accuracy of detecting RERAs was reported (Nassi 2021). They analyzed the respiratory related movements over the chest and abdomen during sleep and proposed a multi-class stratification algorithm to detect apneas, hypopneas, and RERAs. They have specifically reported the accuracy of detecting each type of respiratory events. The reported accuracy of detecting RERA in this study was only 29%, indicating the gap in current techniques and algorithms for robust and accurate detection of RERAs.

Therefore, in this study, a hierarchical mathematical model was proposed for the detection of apneas, hypopneas, and RERAs using airflow, movements over the chest and abdomen, and oxygen saturation in blood (SaO_2) recorded as part of PSG. The proposed algorithm was validated through estimating RDI and detecting the respiratory events including RERAs. This paper is organized as follows: in section 2, the data used in this study and the details of the proposed method as well as the optimization and validation methodology are presented. The results obtained are exposed in section 3. Finally, the discussion and conclusion are outlined in section 4.

2. Method

2.1. Massachusetts General Hospital's (MGH) database

To conduct this study, we used the MGH database provided for the 2018 PhysioNet/Computing in Cardiology (CinC) Challenge. The MGH database included PSG recordings of 1983 adult individuals gathered at the MGH's sleep laboratory for the diagnosis of sleep disorders. The data were divided into training ($n = 994$), and test sets ($n = 989$) by CinC 2018.

An available public training set of the MGH database was included for this study. The Partners Institutional Review Board approved retrospective analysis of the MGH dataset without the need for additional consent. It has been reported that only one set of equipment at one site was used for collecting the whole database. Each PSG recording contained between 7 and 10 h of night sleep data of 13 physiological signals including electroencephalography (EEG), electrooculography (EOG), electromyography (EMG) (Chin, Chest, and Abdomen), electrocardiography (ECG), respiratory airflow, and SaO_2. All the signals were resampled to 200 Hz. The recordings were manually scored by certified sleep technicians at the MGH sleep laboratory according to the American Academy of Sleep Medicine (AASM) guidelines (Berry *et al* 2018). Different annotations for different sleep analysis purposes were provided in this dataset including apneas (central, obstructive and mixed), hypopneas, and RERAs. Obstructive events were defined as decreases in airflow with increased or continued movements over the chest and abdomen, whereas the central events were defined as reduced or no respiratory effort. Mixed apnea was characterized by reduced or no respiratory effort in the first section of the event and increased respiratory effort without airflow in the last section. When apnea/hypopnea events occur, SaO_2 decreases gradually until the subject breathes again and the start of the SaO_2 desaturation has a delay of about 5–50 s with respect to the start of the event (Kwon *et al* 2014). In this study, regions associated with apneas

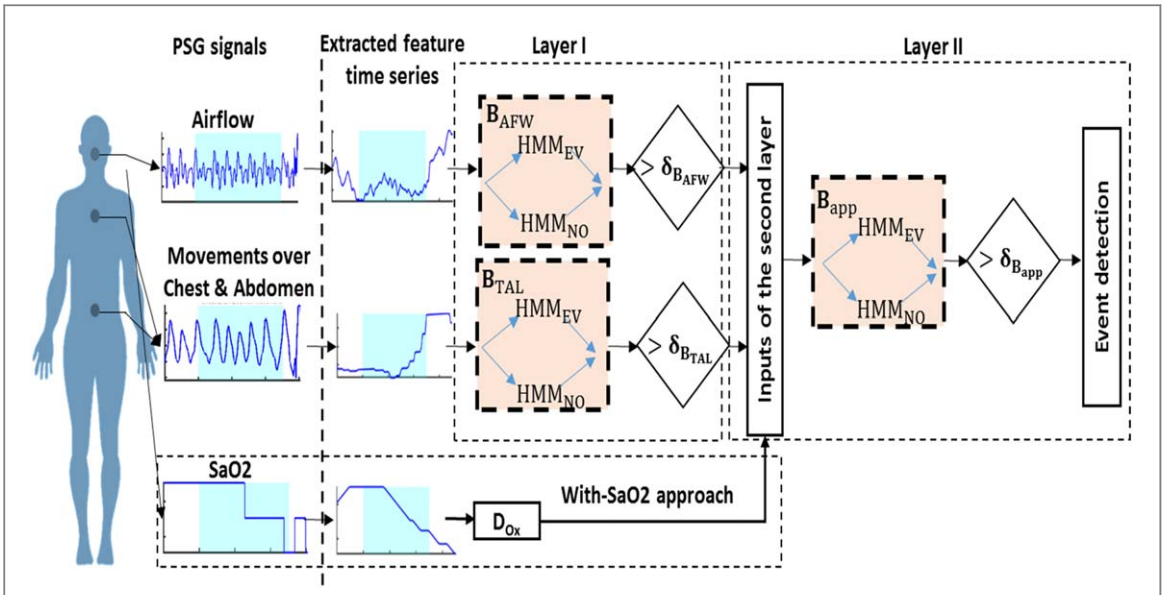

Figure 1. The proposed structure for event-by-event detection.

(central, obstructive, and mixed), hypopneas, and RERAs from the beginning to the end of the event were marked as 1 and otherwise set as 0. Furthermore, the EEG signals were scored in non-overlapping 30 s epochs according to the AASM standards as one of the five common sleep stages: wake, rapid eye movement (REM), non-REM stage 1, non-REM stage 2, and non-REM stage 3. The vector of the sleep-wake stage was formed in such a way that the times when a subject was asleep were zero and otherwise one. For more details see Ghassemi *et al* (2018).

In this study, the annotation of respiratory events and sleep-wake stages were used to develop the detection model and validate the detection performance during sleep in overnight data.

2.2. Data pre-processing

All the analyses were developed and implemented in Matlab (2018b, The MathWorks Inc., Natick, MA, USA) software. Two hundred and ten out of 994 PSG recordings were randomly selected for this study. From the selected recordings, movements over the chest and abdomen, airflow, and SaO_2 were extracted from PSG.

First, the airflow and movements over the chest and abdomen signals were filtered using a notch filter with bandwidth of 57–63 Hz to remove 60 Hz noise. Then, a spike removal algorithm was applied to remove noisy segments with amplitudes higher than the 99th percentile using moving windows with 10 s length with 50% overlap. Then, the signals were standardized with zero mean and unit variance. For SaO_2, only the algorithm of spike removal was applied.

2.3. Feature extraction

A moving window of 10 s with a stride of 0.5 s was used for segmenting the overnight data to obtain feature signals. The characteristics of the sliding window were selected as the average duration of each breath is 3 s, thus it is expected to find three breaths within the window during normal breathing. The average absolute of airflow (AFW) and the average value of SaO_2 (Ox) were calculated. For the movements over the chest and abdomen, local maxima and minima, which are associated with the end of inspiration and expirations, respectively, were detected. Then, the horizontal distance between the amplitude of each local minimum and its following local maximum was measured. The mean value of the measured differences within the segment was used as the range of respiratory movement feature (TAL). TAL was set as zero for the segments such as during apneas where no pair of minimum and maximum was detected (Ghahjaverestan *et al* 2021).

2.4. Detection model

To identify regions that are associated with one of the respiratory events including apneas, hypopneas, and RERAs, the extracted features were fed into the detection model categorized as Layered HMM (LHMM) (Oliver *et al* 2004). This model was consisted of a two-layer hierarchy of standard HMMs (figure 1). Standard HMM is a probabilistic model with finite number of unobservable or hidden states that produces a sequence of observations based on Markov process (a change in the current state depends on the previous state) (Rabiner 1989). Using different HMMs in a layered structure enables analyzing of feature vectors by different

time resolution, injecting signals with different manifestations of the dynamic caused by the event in different layers, and interpreting the effect of each layer separately (Oliver *et al* 2004). For example, airflow and movements over the chest and abdomen are associated with reductions during respiratory events; however, reduction in the SaO_2 signals happened with 5–50 s delay after the start of events due to the blood circulation (Kwon *et al* 2014).

To implement LHMM, the first layer included two HMM banks to separately analyze airflow (B_{AFW}) and movements over the chest and abdomen (B_{TAL}) feature signals. Each HMM bank consisted of two HMMs, one trained by segments associated with the events (EV) and the other one by segments selected from parts out of event segments (NO). At the first layer, each segment of data had $T_1 = 10$ s duration extracted by sliding windows and a stride of 0.5 s. The 10 s duration of segments was selected for the training of the models in the first layer based on the definition of the least length of respiratory events (Berry *et al* 2018).

Then, for each segment of a test data, $O_{t-T_1+1:t}$, the HMM of class k, $k \in \{\mathrm{EV},\ \mathrm{NO}\}$, in bank B, $B = B_{AFW},\ B_{TAL}$ generates a likelihood value as its output:

$$ll_k^B(t) = \log P(O_{t-T_1+1:t}|\lambda_k^B), \tag{1}$$

where λ_k^B is the set of parameters by which the related HMM is characterized. The likelihood of a segment generated by a model represents the chance that the segment belongs to the class/dynamic by which that model was trained. For each bank, the difference log-likelihood was calculated as:

$$ll_{diff}^B(t) = ll_{EV}^B(t) - ll_{NO}^B(t). \tag{2}$$

To detect the respiratory events at the first layer, a threshold was applied to each sample of $ll_{diff}^B(t)$, as:

$$ll_{diff}^B(t) \geqslant \delta_i,\ i = \{B_{AFW},\ B_{TAL}\}, \tag{3}$$

where $\delta_{B_{AFW}}$, $\delta_{B_{TAL}}$ were thresholds used for airflow (B_{AFW}) and movements over the chest and abdomen (B_{TAL}) banks, respectively. Difference log-likelihoods lower than the threshold indicates normal breathing. By applying the thresholds, two binary sequences of 0 (normal breathing) and 1 (event) were generated. Then, using the sleep-wake stage information, the classified events occurring when the subject was awake can be eliminated, hence the classification results and the sleep-wake stage vector were combined to remove the events mistakenly detected during the wake stage. Finally, the generated binary sequences of the first layer were used as the inputs of the second layer.

The purpose of the second layer is to analyze the inputs with usually longer analysis window to make the final decision of classifying each segment of the inputs into normal or within-event classes. In the second layer, each segment of the inputs was selected with the sliding window of T_2. To classify each segment, the two binary sequences generated by B_{AFW} and B_{TAL} in the first layer were segmented by a longer sliding window. Two approaches were implemented in the second layer:

a. Analysis of only binary sequences (without-SaO_2 approach): the two binary sequences were fed to HMMs with $T_{2\text{without-SaO2}}$.

b. Analysis of binary sequences and the feature Ox (with-SaO_2 approach): the two binary sequences along with Ox with a delay of D_{Ox} were used as the inputs of the models using $T_{2\text{with-SaO2}}$ sliding window.

For the two approaches, the second layer was designed by only one bank (B_{app}, app $\in$ {without-SaO_2, with-SaO_2}), which includes two HMMs (one for respiratory events class and another for normal class). To detect the segments associated with events, different thresholds $\delta_{B_{without-SaO_2}}$ and $\delta_{B_{with-SaO_2}}$ were applied to the difference log-likelihood of each approach (without-SaO_2, with-SaO_2). A window was labeled as respiratory event (1) if the following condition was met:

$$ll_{diff}^{B_{app}}(t) \geqslant \delta_{B_{app}},\ app \in \{without - SaO_2,\ with - SaO_2\} \tag{4}$$

otherwise, it was labeled as normal event (0).

2.5. Optimization process of parameters

The extracted recordings were randomly divided into two sets of 100 and 110 recordings, respectively for optimization and test. Two ways of validation, one sample-based for optimization and the other one event-based for event detection were applied. The optimization set was used to train the designed LHMM and find the best values for model parameters. The model parameters including HMM parameters (λ_k^B, $B \in \{B_{AFW},\ B_{TAL},\ B_{with-SaO_2}\},\ k \in \{NO,\ EV\}$), the detection thresholds ($\delta_{B_{AFW}}$, $\delta_{B_{TAL}}$, $\delta_{B_{without-SaO_2}}$, $\delta_{B_{with-SaO_2}}$), the amount delay for D_{Ox}, and $T_{2\text{without-SaO2}}$ and $T_{2\text{with-SaO2}}$ were optimally determined.

After optimizing the design of the LHMM by the determined parameters, the test set was used to evaluate the performance of the model in the event-by-event classification by five-fold cross-validation. In each fold, 20% of

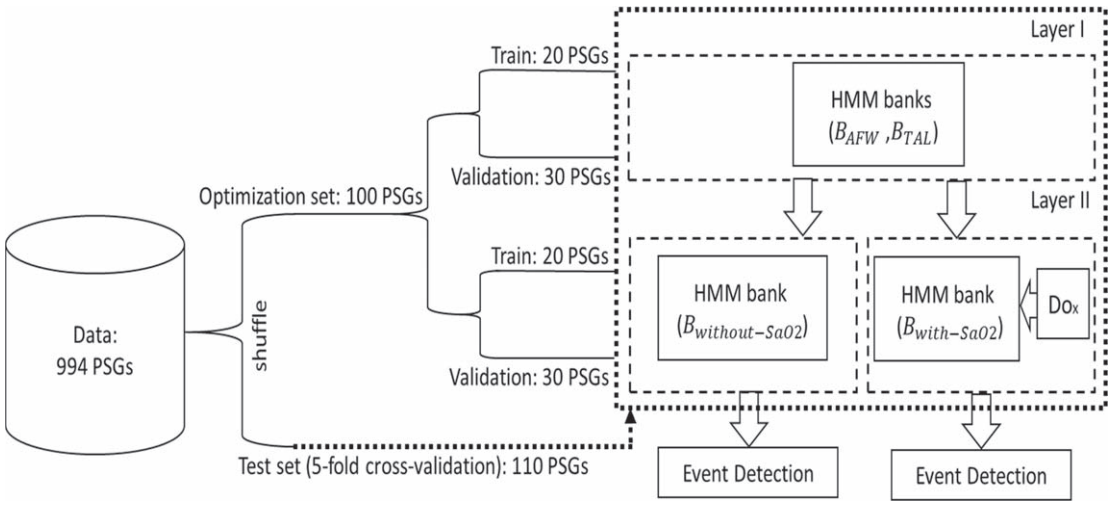

Figure 2. Flow diagram of data division.

the data was used for training and the rest remained for test data. The flow diagram of data division in this study is shown in figure 2.

Out of 100 recordings in the optimization set, 50 recordings were randomly selected and divided into two subsets 20 and 30 recordings for training and validation in the first layer, respectively. A training phase was applied to estimate the parameters of each HMM using training observation dataset.

To construct training data in the first layer, segments were synchronously selected from the extracted feature signals of airflow and movements over the chest and abdomen starting from the onset of the respiratory events with a duration of $T_1 = 10$ s for each event model in the two banks (B_{AFW}, B_{TAL}), and segments with a duration of $T_1 = 10$ s were randomly chosen out of event segments from the feature signals for normal models. The number of selected segments to train the normal model in both banks (B_{AFW}, B_{TAL}) was chosen equal to the number of data in the event model.

A set of number of states, {2,3,4,5}, were investigated to determine the optimal states of each HMM. Then, the selected segments were used to train event and normal models in relevant banks with different combinations of states [(2,2), …, (2,5), (3,2), …(5,5)]. A range of 301 values from −10 to 50 with interval 0.2 was investigated to optimize the log-likelihood thresholds ($\delta_{B_{AFW}}$, $\delta_{B_{TAL}}$). In each bank, the difference of the two log-likelihoods generated by the models was compared to the corresponding threshold (equation (3)), and metrics including recall (equation (5)), precision (equation (6)), and F1 score (equation (7)) were calculated as follows :

$$Recall = TP/(TP + FN), \tag{5}$$

$$Precision = TP/(TP + FP), \tag{6}$$

$$F1\ score = TP/(TP + 0.5 \times (FP + FN), \tag{7}$$

where TP, FP, and FN denote the number of true positives, false positives, and false negatives, respectively. Precision-recall curves (PRC) were used to calculate the optimal classification probability threshold and states. The PRC curves were created for models with different combinations of states and thresholds. For each bank, states and threshold that resulted in the highest F1 score on the validation dataset with PSG-derived RDI $\geqslant 5$ were chosen as the optimal parameters.

To determine the optimal parameters of HMMs in the second layer, the remaining 50 recordings were divided into two subsets 20 and 30 recordings for training and finding the optimal parameters (threshold values and states), respectively. According to each approach (without-SaO_2, with-SaO_2) in the second layer, selected segments relevant to normal and event models were constructed for the training phase. The number of selected segments for the normal model was chosen equal to the number of data in the event model. The algorithm to obtain the optimal parameters in the second layer for both approaches was similar to the first one except a range of 75 values from −0.4 to 7 with interval 0.1 was investigated to optimize the log-likelihood thresholds ($\delta_{B_{without-SaO2}}$, $\delta_{B_{with-SaO2}}$). In both approaches (without-SaO_2, with-SaO_2), to determine the sliding window length in the second layer by greedy search, values from 11 to 29 s with a 2 s step were investigated. To determine the amount of delay (D_{Ox}) in the with-SaO_2 approach, values from 0 to 25 s with a 5 s step were investigated by greedy search.

Table 1. PSG recordings characteristic in this study.

	RDI <5	$5 \leqslant$RDI <15	$15 \leqslant$RDI <30	RDI $\geqslant 30$
Sample size (male)	27 (14)	48 (26)	84 (63)	51 (43)
	Clinical features (Data is presented in mean (+/-standard deviation))			
Age (years)	50.85 (+/−17.43)	53.23 (+/−16.15)	56.88 (+/−13.78)	55.35 (+/−14.52)
Recording time (h)	7.71 (+/−0.47)	7.67 (+/−0.74)	7.69 (+/−0.66)	7.58 (+/−0.66)
Sleeping time (h)	6.51 (+/−1.20)	6.56 (+/−0.96)	6.28 (+/−1.13)	5.37 (+/−1.80)
PSG-derived RDI[a]	2.09 (+/−1.61)	10.03 (+/−3.0)	22.70 (+/−4.61)	44.21 (+/−14.12)
		Number of events		
Central apnea	136	762	2290	2580
Obstructive apnea	72	699	3442	3236
Mixed apnea	33	128	618	1570
Hypopnea	51	632	3771	2564
RERA[b]	81	954	1797	1800

[a] Respiratory Disturbance Index.

[b] Respiratory Event Related Arousal.

2.6. Validation and statistical analyses

2.6.1. Event detection

For event by event validation, PSG recordings were categorized into four groups based on PSG-derived RDI values; Normal: RDI <5; Mild: $5 \leqslant$ RDI <15; Moderate: $15 \leqslant$ RDI <30; Severe: RDI $\geqslant 30$. For each recording, the detected annotation (the output of each bank) was compared to the provided reference annotations, if there was an overlap between a detected event with an annotated event, the event was marked as true positive (TP), otherwise, it was false positive (FP). False negative (FN) happened when a reference event was not detected. Finally, precision, recall, and the F1 score were calculated in each group, in addition, the recall was reported for the apneas (central, obstructive, mixed), hypopneas, and RERAs separately.

2.6.2. Identifying people at risk

RDI was estimated as the number of detected events per hour of sleep. Based on RDI, each recording was categorized into five groups of different cutoffs (10, 15, 20, 25, and 30 events/hour). Metrics of the recall, specificity, precision, and accuracy to group recordings into different RDI cutoffs were calculated. Then, Bland–Altman plots were used to quantify the difference between estimated and PSG-derived RDI. To assess the agreement, Pearson and Spearman's correlation coefficient base on the normality of the data were calculated between the estimated and PSG-derived RDI.

2.6.3. Comparison of event detection approaches

To compare the performance of the two event detection approaches (without-SaO_2 approach and with-SaO_2 approach), T-test or Mann-Whitney test was applied based on the normality of the data distribution examined by Shapiro-Wilk test. The p-value <0.05 was considered as significant. Statistical analyses were conducted by R Statistical Software (version 3.6.2).

3. Results

3.1. Subject demographics

The selected population presented includes 210 participants (male = 146) with age: 54.9 $\pm$ 15.02 years old, sleep time: 6.14 $\pm$ 1.36 h, and RDI: 22.37 $\pm$ 16.18. Table 1 shows the demographic information of the included recordings. Figure 3 shows example traces of PSG recordings of airflow, movements over the chest and abdomen, and SaO_2 during various types of respiratory events.

3.2. Results for optimization step

The optimal values of the parameters in the first and the second layer, as well as calculated metrics in these values, are summarized in table 2. The maximum results of the F1 scores (without-SaO_2 approach: 0.56 $\pm$ 0.16, with-SaO_2 approach: 0.62 $\pm$ 0.14) in the second layer by varying the sliding window length on the optimization set of the approaches (without-SaO_2, with-SaO_2) were achieved by a 23 s and 21 s sliding window, respectively, for which we obtained recall and precision 0.64 $\pm$ 0.17 and 0.53 $\pm$ 0.20 for without-SaO_2 approach and 0.67 $\pm$ 0.13 and 0.61 $\pm$ 0.17 for with-SaO_2 approach, respectively. The best D_{Ox} in the with-SaO_2 approach was achieved by a 20 s delay. Figure 4 shows the F1 score measured in the second layer by varying the sliding window length and delay on the optimization set of the with-SaO_2 approach.

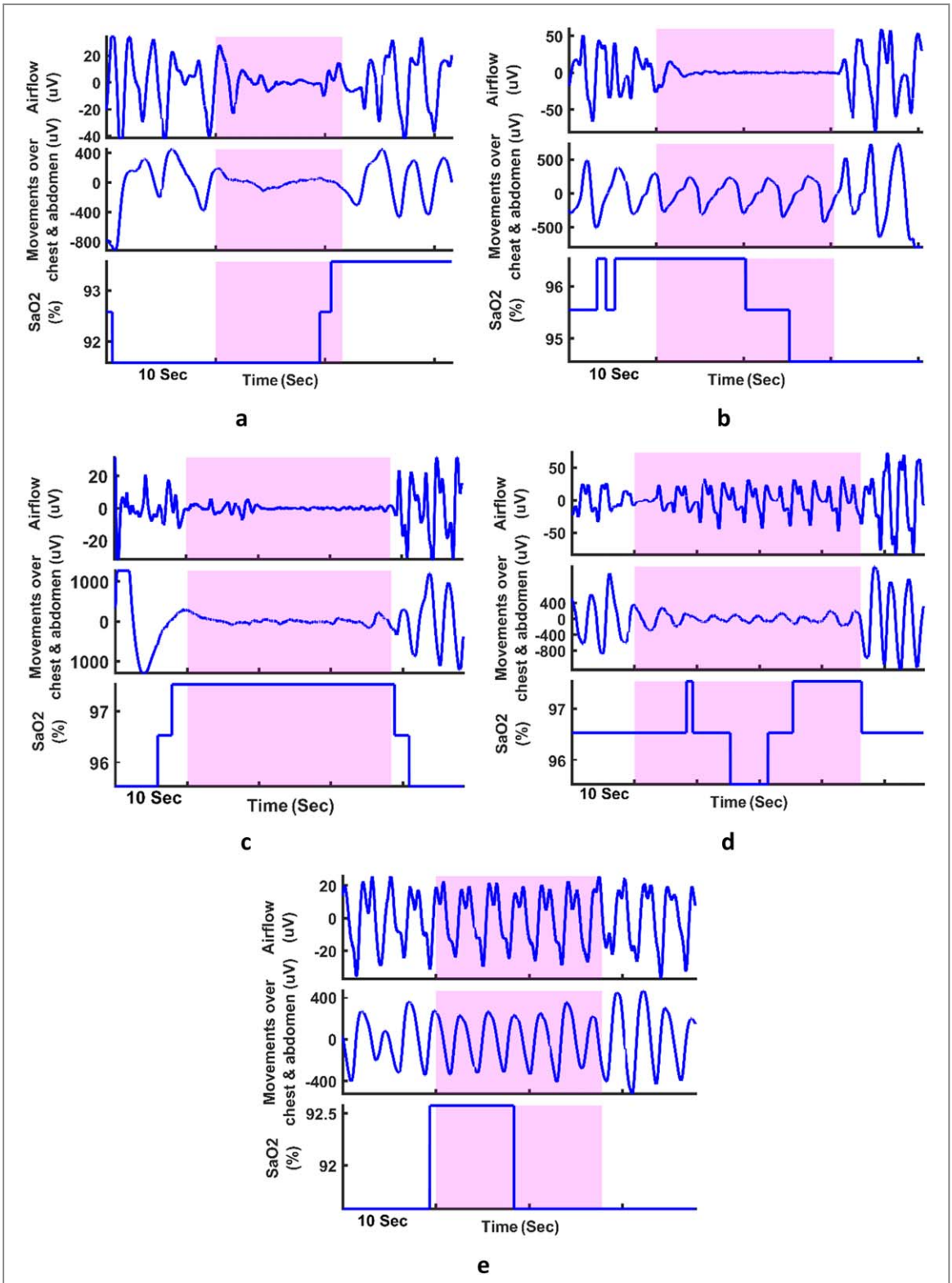

Figure 3. Example of (a) Central apnea; (b) Obstructive apnea; (c) Mixed apnea; (d) Hypopnea; (e) RERA marked in a red shaded area.

Table 3 summarizes the overall event-by-event detection performance metrics in each RDI group on the optimization set for the two approaches (without-SaO_2, with-SaO_2). The highest F1 score was obtained 0.72 ± 0.10 for RDI $\geqslant 30$ by without-SaO_2 approach, and 0.74 ± 0.06 for RDI $\geqslant 30$ by with-SaO_2 approach.

Furthermore, on the optimization set, the correlation between estimated RDI and PSG-derived RDI was calculated for two approaches according to sleep time (figure 5). The correlation between estimated RDI and PSG-derived RDI was ($R^2 = 0.80, p < 0.0001$) for the without-SaO_2 approach, and ($R^2 = 0.85, p < 0.0001$) for the with-SaO_2 approach. Figure 6 shows the Bland–Altman plots in both approaches, the mean and standard deviation are 4.6 and 12.63 for the without-SaO_2 approach, and 0.8 and 8.85 for the with-SaO_2 approach.

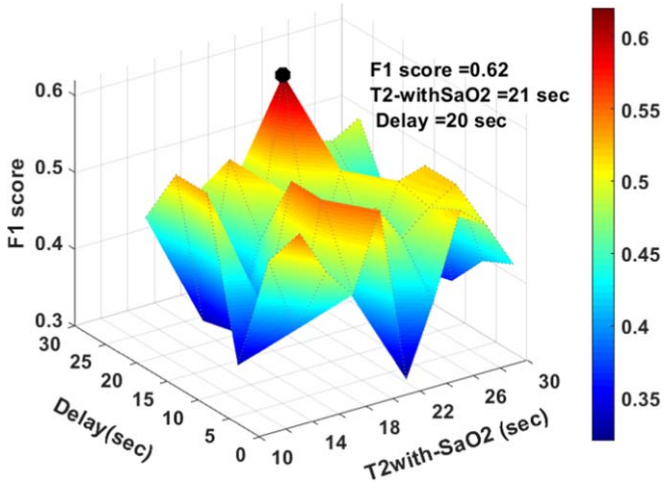

Figure 4. Determination of D_{Ox} in the second layer of the with-SaO_2 approach by varying the sliding window length ($T_{2\text{with-SaO2}}$) and delay.

Table 2. The optimal values of the parameters in the first layer and the second layer and related metrics.

Performance metrics	B_{AFW}	B_{TAL}	$B_{without-SaO2}$	$B_{with-SaO2}$
Optimal states [EV, NO]	[2,2]	[3, 2]	[2, 3]	[2, 4]
Recall	0.41 ± 0.15	0.36 ± 0.18	0.64 ± 0.17	0.67 ± 0.13
Precision	0.38 ± 0.20	0.32 ± 0.16	0.53 ± 0.20	0.61 ± 0.17
F1 score	0.37 ± 0.13	0.33 ± 0.15	0.56 ± 0.16	0.62 ± 0.14
D_{Ox} (s)	—	—	—	20
Optimal T_2 (s)	—	—	23	21

Table 3. Event detection metrics for different RDI groups on the optimization set.

RDI group	Number of subjects	Precision	Recall	F1 score
a) HMM bank related to airflow (B_{AFW}) in the first layer				
$5 \leqslant$ RDI $<$ 15	10	0.27 ± 0.13	0.33 ± 0.12	0.28 ± 0.10
$15 \leqslant$ RDI $<$ 30	10	0.40 ± 0.18	0.47 ± 0.12	0.41 ± 0.11
RDI $\geqslant$ 30	5	0.59 ± 0.15	0.48 ± 0.18	0.48 ± 0.11
b) HMM bank related to movements over the chest and abdomen (B_{TAL}) in the first layer				
$5 \leqslant$ RDI $<$ 15	10	0.21 ± 0.12	0.30 ± 0.20	0.23 ± 0.14
$15 \leqslant$ RDI $<$ 30	10	0.33 ± 0.11	0.35 ± 0.12	0.33 ± 0.10
RDI $\geqslant$ 30	5	0.50 ± 0.15	0.52 ± 0.13	0.50 ± 0.11
c) Without-SaO_2 approach in the second layer				
$5 \leqslant$ RDI $<$ 15	7	0.32 ± 0.09	0.58 ± 0.18	0.39 ± 0.10
$15 \leqslant$ RDI $<$ 30	11	0.54 ± 0.13	0.63 ± 0.15	0.56 ± 0.11
RDI $\geqslant$ 30	7	0.73 ± 0.12	0.72 ± 0.14	0.72 ± 0.10
d) With-SaO_2 approach in the second layer				
$5 \leqslant$ RDI $<$ 15	7	0.44 ± 0.17	0.63 ± 0.12	0.50 ± 0.14
$15 \leqslant$ RDI $<$ 30	11	0.62 ± 0.10	0.66 ± 0.15	0.63 ± 0.10
RDI $\geqslant$ 30	7	0.77 ± 0.07	0.72 ± 0.11	0.74 ± 0.06

Values are reported as mean ± standard deviation.

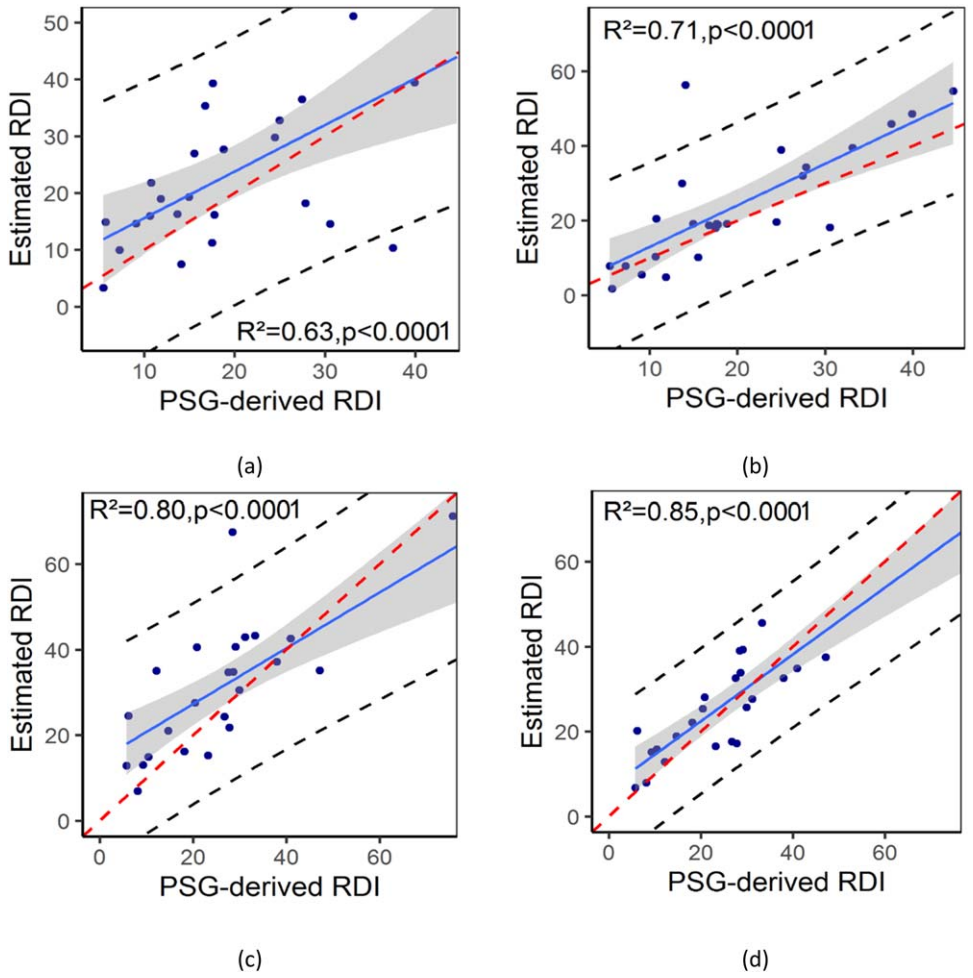

Figure 5. Correlation between estimated RDI and PSG-derived RDI on the optimization set. (a): HMM bank related to airflow (B_{AFW}); (b): HMM bank related to movements over the chest and abdomen (B_{TAL}); (c): without-SaO_2 approach; (d): with-SaO_2 approach. The blue and red dashed lines show the least-square regression and identity lines respectively. Dashed black lines indicate the 95% prediction limits and 95% confidence limits consist of the space between the two gray curves.

The accuracy of identifying individuals at risk of sleep apnea based on RDI on the optimization set were 0.88, 0.88, 0.84, 0.80, 0.72 for the without-SaO_2 approach, and 0.92, 0.84, 0.80, 0.84, and 0.80 for the with-SaO_2 approach, respectively (table 4).

3.3. Results of event-by-event detection

Figure 7 shows example traces of detected annotations compared to the reference ones using without-SaO_2 and with-SaO_2 approaches for RDI $<$ 15 (figure 7(a)) and RDI $\geqslant$ 30 (figure 7(b)).

Table 5 presents the performance of the proposed model in event detection for both approaches. The highest F1 score was obtained 0.58 $\pm$ 0.13 for RDI $\geqslant$ 30 by without-SaO_2 approach, and 0.70 $\pm$ 0.08 for RDI $\geqslant$ 30 by with-SaO_2 approach.

Figure 8 shows the correlation between the estimated RDI and the PSG-derived RDI for the two approaches over the recordings on the test set. Based on these results, strong correlation values were obtained for both approaches; ($R^2 = 0.85, p < 0.0001$) for the without-SaO_2 approach, and ($R^2 = 0.91, p < 0.0001$) for the with-SaO_2 approach. The Bland–Altman plots depicted in figure 9 for both approaches indicate the mean and standard deviation values of 1.29 and 8.9 for the without-SaO_2 approach and -2.98 and 7.25 for the with-SaO_2 approach.

For RDI cut-off thresholds of 10, 15, 20, 25, and 30, the accuracy of identifying individuals at risk of sleep apnea based on RDI were 0.87, 0.88, 0.83, 0.89, 0.89 for the without-SaO_2 approach, and 0.86, 0.82, 0.85, 0.86, and 0.88 for with-SaO_2 approach, respectively (table 6).

Table 7 shows the number of and classifier's recall of each event on the test set for apneas (central, obstructive, and mixed), hypopneas, and RERAs for both approaches. There were a total of 15 067 respiratory events (apneas = 5337, hypopneas = 4782, RERAs = 4948).

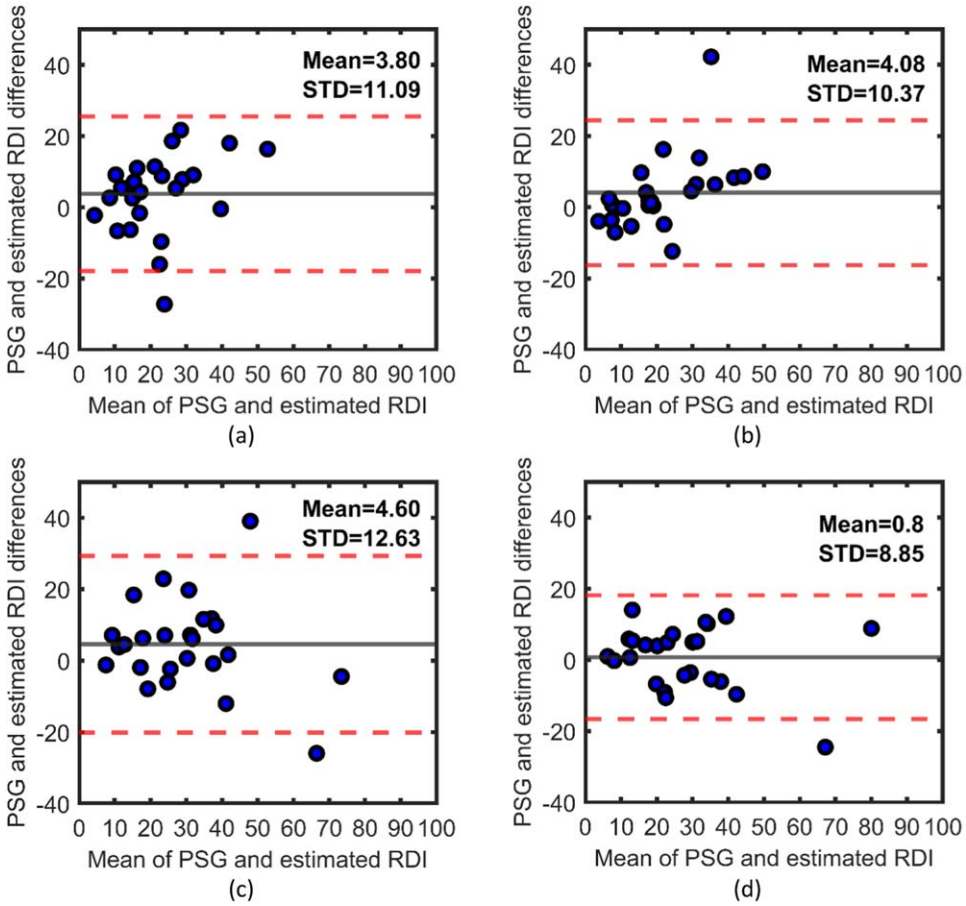

Figure 6. Bland–Altman limits of agreement between estimated RDI and PSG-derived RDI on the optimization set. (a): HMM bank related to airflow (B_{AFW}); (b): HMM bank related to movements over the chest and abdomen (B_{TAL}); (c): without-SaO_2 approach; (d): with-SaO_2 approach. The black line indicates the average difference and dashed red lines present 95% confidence interval (mean $\pm$ 1.96 standard deviation) of the difference between estimated RDI and PSG-derived RDI.

Table 8 presents the statistical comparison results of the metrics between the two event detection approaches. The results indicate that the performance metrics (precision, recall, F1 score) in $5 \leqslant$ RDI were significantly higher in the with-SaO_2 approach than in the without-SaO_2 approach ($p < 0.05$). In contrast, no significant differences were found between the performance metrics of the two event detection approaches in RDI $<$ 5 group ($p > 0.05$).

4. Discussion and conclusion

In this study, a probabilistic heretical model based on HMM was proposed to detect respiratory events including apnea, hypopnea, and RERA, estimate the severity of sleep apnea, and identify individuals at risk based on RDI. We were able to successfully: 1—implement the structure of hierarchy of HMMs to detect respiratory events from the features extracted from airflow, movements over the chest and abdomen, and SaO_2 channels of PSG, 2—validate the performance of each layer separately, 3—estimate RDI with high performance ($R^2 = 0.91$ when injecting the feature related to SaO_2), 4—identify the individuals at risk based on different RDI cutoffs.

The proposed method used three signals to monitor sleep apnea: airflow, movements over the chest and abdomen, and SaO_2. Reduction of airflow associated with sleep apnea impaired gas exchange in the lungs, which itself causes a decrease in SaO_2. SaO_2 signal has been used alone in some studies to estimate an event (Issa *et al* 1993, Chang *et al* 2020). This approach worked well as the goal of these studies was to estimate the severity of sleep apnea without clear validation about the event detection. However, SaO_2 is not immediately sensitive to the occurrence of sleep apnea due to the blood circulation delay. Therefore, we used two approaches to specifically address the event detection. At the first approach (without-SaO_2 approach), the event detection model was implemented by airflow, movements over the chest and abdomen, while at the second approach (with-SaO_2 approach), in addition to the signals used for the first approach, the feature extracted from SaO_2 with a delay was used.

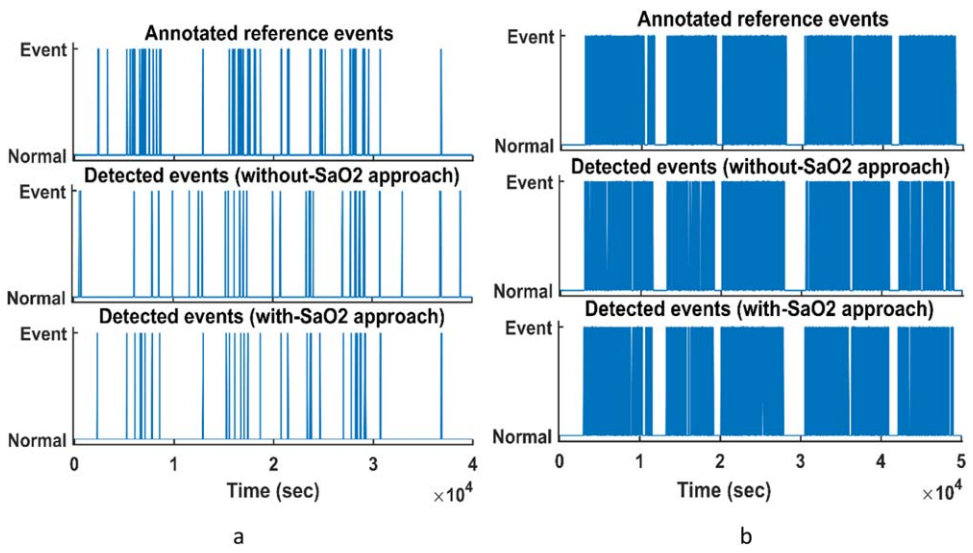

Figure 7. Examples of annotated and detected events by two approaches. (a) PSG with RDI $<$ 15, (b) PSG with RDI $\geqslant$ 30.

Table 4. Identifying individuals at risk of sleep apnea based on different RDI cutoffs on the optimization set.

Cutoff	Recall	Specificity	Precision	F1-score	Accuracy
a) HMM bank related to airflow (B_{AFW}) in the first layer					
10	0.50	0.95	0.67	0.57	0.88
15	0.50	0.80	0.63	0.56	0.68
20	0.69	0.67	0.79	0.73	0.68
25	0.67	0.57	0.80	0.73	0.64
30	0.80	0.60	0.89	0.84	0.76
b) HMM bank related to movements over the chest and abdomen (B_{TAL}) in the first layer					
10	1.00	0.95	0.80	0.89	0.96
15	0.60	0.93	0.86	0.71	0.80
20	0.81	0.78	0.87	0.84	0.80
25	0.83	0.86	0.94	0.88	0.84
30	0.80	0.80	0.94	0.86	0.80
c) Without-SaO_2 approach in the second layer					
10	0.25	1.00	1.00	0.40	0.88
15	0.57	1.00	1.00	0.73	0.88
20	0.63	0.94	0.83	0.71	0.84
25	0.73	0.86	0.80	0.76	0.80
30	0.61	1.00	1.00	0.76	0.72
d) With-SaO_2 approach in the second layer					
10	0.50	1.00	1.00	0.67	0.92
15	0.43	1.00	1.00	0.60	0.84
20	0.75	0.82	0.67	0.71	0.80
25	0.82	0.86	0.82	0.82	0.84
30	0.78	0.86	0.93	0.85	0.80

One of the important features of our detection approach was to detect the individual respiratory events and report the event-by-event detection results for each layer. Precision, recall, and F1 score were selected for assessing the performance of the algorithm to eliminate the effect of a high number of true negatives (normal segments).

Table 5. Overall event detection metrics for different RDI groups on the test set.

RDI group	Number of subjects	Precision	Recall	F1 score
a) HMM bank related to airflow (B_{AFW}) in the first layer				
RDI $<$ 5	14	0.06 $\pm$ 0.04	0.26 $\pm$ 0.26	0.09 $\pm$ 0.06
5 $\leqslant$ RDI $<$ 15	22	0.24 $\pm$ 0.07	0.22 $\pm$ 0.10	0.22 $\pm$ 0.08
15 $\leqslant$ RDI $<$ 30	45	0.40 $\pm$ 0.16	0.37 $\pm$ 0.16	0.36 $\pm$ 0.12
RDI $\geqslant$ 30	29	0.53 $\pm$ 0.18	0.33 $\pm$ 0.12	0.38 $\pm$ 0.11
b) HMM bank related to movements over the chest and abdomen (B_{TAL}) in the first layer				
RDI $<$ 5	14	0.04 $\pm$ 0.03	0.18 $\pm$ 0.09	0.06 $\pm$ 0.04
5 $\leqslant$ RDI $<$ 15	22	0.12 $\pm$ 0.05	0.19 $\pm$ 0.07	0.14 $\pm$ 0.06
15 $\leqslant$ RDI $<$ 30	45	0.28 $\pm$ 0.11	0.31 $\pm$ 0.12	0.29 $\pm$ 0.11
RDI $\geqslant$ 30	29	0.43 $\pm$ 0.12	0.40 $\pm$ 0.19	0.40 $\pm$ 0.13
c) Without-SaO_2 approach in the second layer				
RDI $<$ 5	14	0.09 $\pm$ 0.07	0.39 $\pm$ 0.21	0.13 $\pm$ 0.08
5 $\leqslant$ RDI $<$ 15	22	0.29 $\pm$ 0.12	0.39 $\pm$ 0.15	0.33 $\pm$ 0.13
15 $\leqslant$ RDI $<$ 30	45	0.45 $\pm$ 0.12	0.45 $\pm$ 0.15	0.44 $\pm$ 0.12
RDI $\geqslant$ 30	29	0.61 $\pm$ 0.14	0.57 $\pm$ 0.13	0.58 $\pm$ 0.13
d) With-SaO_2 approach in the second layer				
RDI $<$ 5	14	0.16 $\pm$ 0.13	0.49 $\pm$ 0.23	0.22 $\pm$ 0.16
5 $\leqslant$ RDI $<$ 15	22	0.42 $\pm$ 0.10	0.45 $\pm$ 0.17	0.43 $\pm$ 0.13
15 $\leqslant$ RDI $<$ 30	45	0.63 $\pm$ 0.10	0.52 $\pm$ 0.18	0.55 $\pm$ 0.13
RDI $\geqslant$ 30	29	0.79 $\pm$ 0.07	0.64 $\pm$ 0.10	0.70 $\pm$ 0.08

Values are reported as mean $\pm$ standard deviation.

Table 6. Identifying individuals at risk of sleep apnea based on different RDI cutoffs on the test set.

Cutoff	Recall	Specificity	Precision	F1-score	Accuracy
a) HMM bank related to airflow (B_{AFW}) in the first layer					
10	0.67	0.84	0.53	0.59	0.80
15	0.86	0.76	0.63	0.73	0.79
20	0.78	0.61	0.63	0.70	0.69
25	0.88	0.55	0.74	0.81	0.75
30	0.84	0.45	0.81	0.82	0.74
b) HMM bank related to movements over chest and abdomen (B_{TAL}) in the first layer					
10	0.33	0.99	0.89	0.48	0.85
15	0.44	0.99	0.94	0.60	0.81
20	0.63	0.88	0.82	0.71	0.76
25	0.83	0.77	0.85	0.84	0.81
30	0.90	0.66	0.88	0.89	0.84
c) Without-SaO_2 approach in the second layer					
10	0.42	1.00	1.00	0.59	0.87
15	0.81	0.92	0.83	0.82	0.88
20	0.78	0.86	0.83	0.81	0.83
25	0.91	0.86	0.91	0.91	0.89
30	0.90	0.86	0.95	0.92	0.89
d) With-SaO_2 approach in the second layer					
10	0.75	0.90	0.67	0.71	0.86
15	0.86	0.80	0.67	0.76	0.82
20	0.92	0.78	0.78	0.85	0.85
25	0.95	0.73	0.84	0.89	0.86
30	0.98	0.62	0.88	0.92	0.88

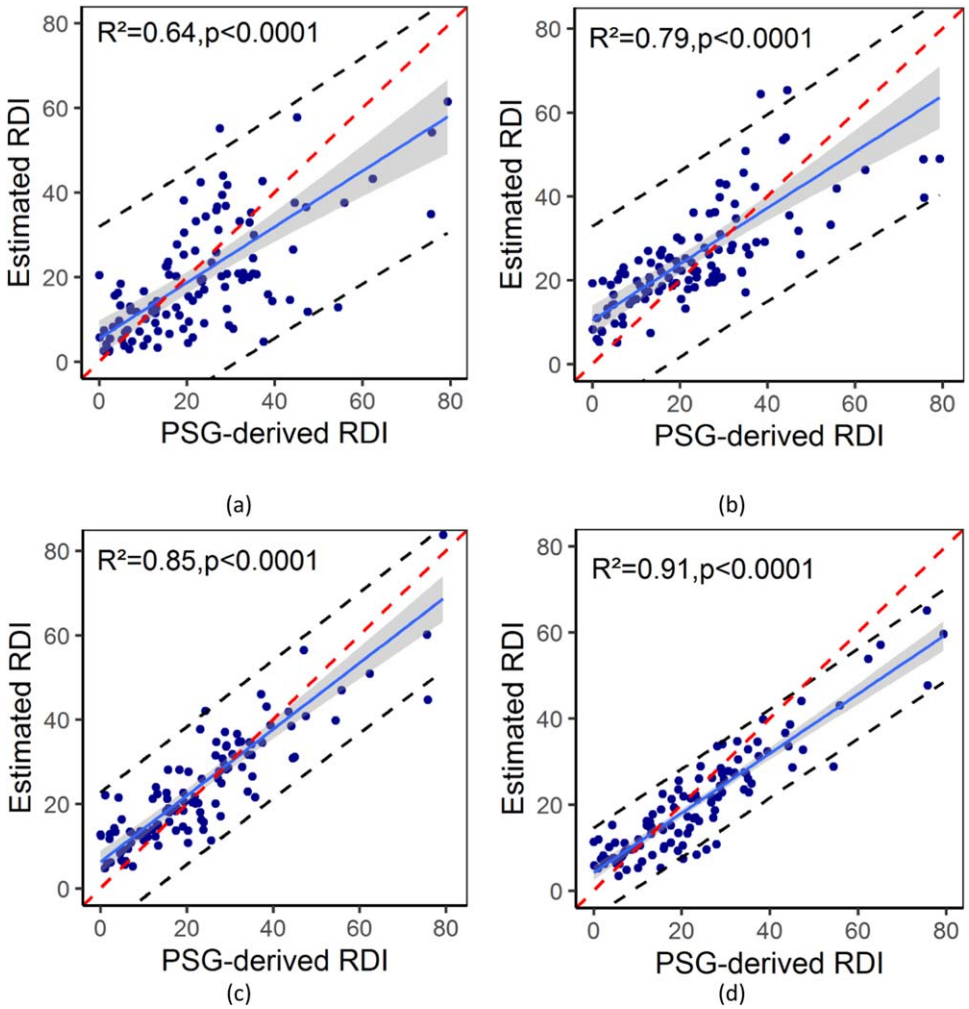

Figure 8. Correlation between estimated RDI and PSG-derived RDI in 110 PSGs on the test set. (a): HMM bank related to airflow (B_{AFW}), (b): HMM bank related to movements over the chest and abdomen (B_{TAL}), (c): without-SaO_2 approach; (d): with-SaO_2 approach. The blue and red dashed lines show the least-square regression and identity lines, respectively. Dashed black lines indicate the 95% prediction limits and 95% confidence limits consist of the space between the two gray curves.

Table 7. Recall values for detection of each respiratory event type.

Event type	Number of events	Recall (without-SaO_2 approach)	Recall (with-SaO_2 approach)
Central apnea	2092	0.42 ± 0.2	0.5 ± 0.29
Obstructive apnea	2964	0.55 ± 0.18	0.70 ± 0.24
Mixed apnea	281	0.61 ± 0.23	0.67 ± 0.31
Hypopnea	4782	0.53 ± 0.19	0.58 ± 0.21
RERA	4948	0.38 ± 0.19	0.45 ± 0.27

Values are reported as mean ± standard deviation.

Higher performance in the second layer indicates that combining the outputs of the first layer and analyzing them over a longer course of time can improve the event detection performance.

For recordings with RDI $\geqslant$ 30, the F1 score was the highest for both approaches (0.58 ± 0.13 for the without-SaO_2 approach, and 0.70 ± 0.08 for the with-SaO_2 approach). The results indicate that the two approaches can accurately detect the majority of the events in patients with severe sleep apnea. This is a population in which accurate diagnosis of events is very important to assess the pathophysiology of sleep apnea at different stages (Saha *et al* 2020). For recordings with low RDI, the F1 score was the least (0.13 ± 0.08 for without-SaO_2 approach, 0.22 ± 0.16 for with-SaO_2 approach), as there were a few respiratory events that make the data highly imbalanced. Therefore, small false negatives will largely affect precision and recall values.

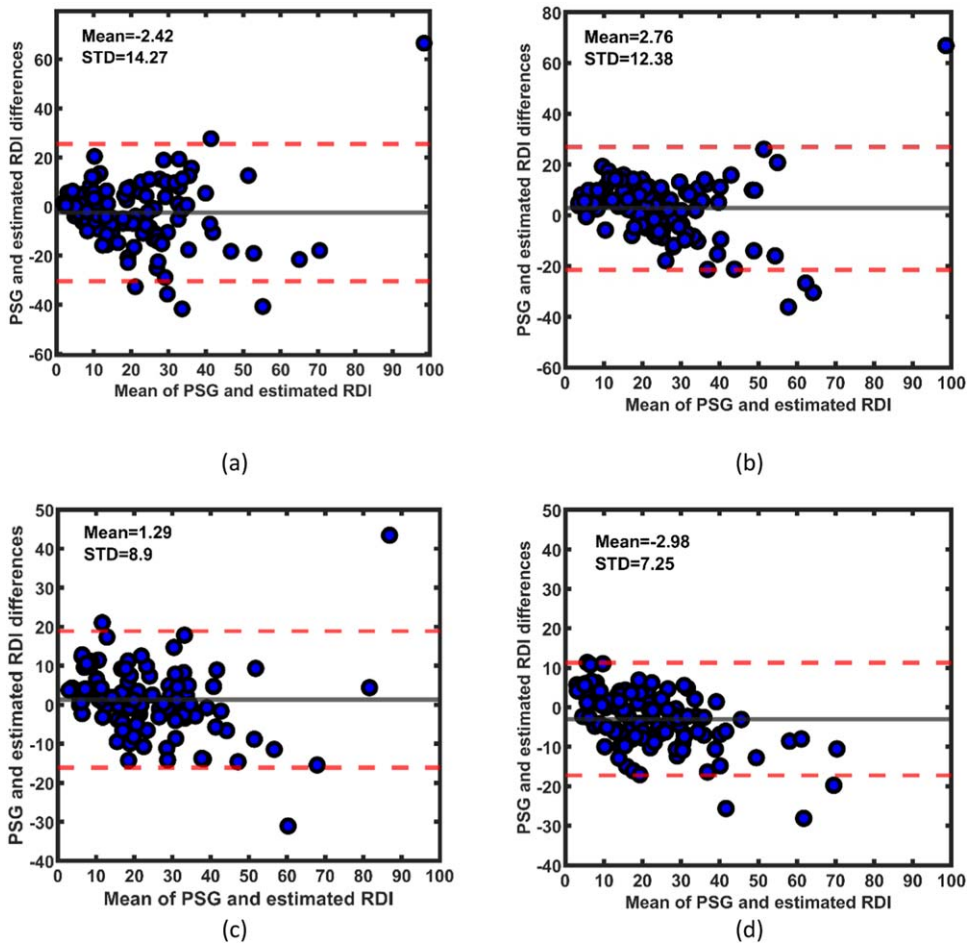

Figure 9. Bland–Altman limits of agreement between estimated RDI and PSG-derived RDI in 110 PSGs on the test set. (a): HMM bank related to airflow (B_{AFW}), (b): HMM bank related to movements over the chest and abdomen (B_{TAL}), (c): without-SaO_2 approach; (d): with-SaO_2 approach. The black line indicates the average difference and dashed red lines present 95% confidence interval (mean ± 1.96 standard deviation) of the difference between estimated RDI and PSG-derived RDI.

Table 8. Comparison statistical analyses of two event detection approaches.

RDI Group	Performance metrics	Without-SaO_2 approach	With-SaO_2 approach	p-value (cut off value=0.05)
RDI $<$ 5	Precision	0.09 ± 0.07	0.16 ± 0.13	0.1857
	Recall	0.39 ± 0.21	0.49 ± 0.23	0.1581
	F1-score	0.13 ± 0.08	0.22 ± 0.16	0.1053
5 $\leqslant$ RDI $<$ 15	Precision	0.29 ± 0.12	0.42 ± 0.10	0.000 9215
	Recall	0.39 ± 0.15	0.45 ± 0.17	0.020 31
	F1-score	0.33 ± 0.13	0.43 ± 0.13	0.016 14
15 $\leqslant$ RDI $<$ 30	Precision	0.45 ± 0.12	0.63 ± 0.10	1.067e-10
	Recall	0.45 ± 0.15	0.52 ± 0.18	0.04753
	F1-score	0.44 ± 0.12	0.55 ± 0.13	6.515e-05
RDI $\geqslant$ 30	Precision	0.61 ± 0.14	0.79 ± 0.07	5.931e-08
	Recall	0.57 ± 0.13	0.64 ± 0.10	0.01196
	F1-score	0.58 ± 0.13	0.70 ± 0.08	2.233e-05

Values are reported as mean ± standard deviation.

No significant differences were observed in the performance metrics (precision, recall, and F1 score) of the two event detection approaches for recordings with RDI $<$ 5 (normal). However, the with-SaO_2 approach provided higher performances than the other approach for recordings with RDI $\geqslant$ 5 (mild, moderate, and severe) presumably due to injecting the feature related to SaO_2 in the second layer.

By counting the detected events by the detection algorithms, we estimated the RDI, which was strongly correlated with the RDI reported from the PSG ($R^2 = 0.85$ for without-SaO_2 approach, and $R^2 = 0.91$ for

with-SaO_2 approach). In a previous study by Ayappa *et al* (2000), the intra-class correlation coefficient reported 0.96, however they didn't report the event by event detection results. More importantly, we found high performances for all the RDI cut-offs to diagnose sleep apnea, which indicates the robustness of the proposed algorithm for the clinical diagnosis of sleep apnea.

The recall of detecting RERA was 0.38 ± 0.19 for the without-SaO_2 approach, and 0.45 ± 0.27 for the with-SaO_2 approach, which is higher than the results reported by Nassi (2021) with recall of RERA; 29%.

One of the limitations of our work was that PSG channels available in this data were collected from one site and one equipment setup and the algorithm was validated on low sample size. For future work, a new model with the same architecture could be trained on more data from different sites and equipment to develop a model that can be generalized accordingly. Another limitation that can be explored in future work was to modify the model to distinguish the type of respiratory events, especially RERA's.

In conclusion, in this study, a hierarchical structure based on HMM was developed to detect respiratory events including RERAs, and to estimate RDI based on airflow, movements over chest and abdomen, and SaO_2. Two approaches were considered. Results showed that the first approach (without-SaO_2 approach), using features of airflow and movements over the chest and abdomen, was able to provide a satisfactory event detection performance, however, injecting the feature related SaO_2 in the second layer (with-SaO_2 approach) further improved the performance of the proposed algorithm in event detection. Automatic detection of RERAs together with other respiratory events (apneas and hypopneas) provide additional information on a patient's sleep quality and can also improve the quality of treatment.

Acknowledgments

The authors would like to thank Dr. Nasim Montazeri Ghahjaverestan for her insightful and valuable advice on this work.

References

Almazaydeh L, Elleithy K and Faezipour M 2012 Obstructive sleep apnea detection using SVM-based classification of ECG signal features *2012 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 4938–41

Ayappa I, Norman R G, Krieger A C, Rosen A, O'Malley R L and Rapoport D M 2000 Non-Invasive detection of respiratory effort-related arousals (REras) by a nasal cannula/pressure transducer system *Sleep* **23** 763–71

BaHammam A S, Sharif M, Gacuan D E and George S 2011 Evaluation of the accuracy of manual and automatic scoring of a single airflow channel in patients with a high probability of obstructive sleep apnea *Med. Sci. Monitor: Int. Med. J. Exp. Clin. Res.* **17** MT13–9

Baisch A, Afshar S, Hörmann K and Maurer J T 2007 detection of respiratory effort-related arousals using a single flow signal *Sleep Med.* O0056 **8** S62

Berry A C *et al* 2018 The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications *Darien, IL:, Am. Acad. Sleep Med.* **2** 5

Calero G, Farre R, Ballester E, Hernandez L, Daniel N and Montserrat J M 2006 Canal, physiological consequences of prolonged periods of flow limitation in patients with sleep apnea hypopnea syndrome *Respir. Med.* **100** 813–7

Calhoun D A and Harding S M 2010 Sleep and hypertension *Chest* **138** 434–43

Chandra S, Sica A L, Wang J, Lakticova V and Greenberg H E 2013 Respiratory effort-related arousals contribute to sympathetic modulation of heart rate variability *Sleep Breath* **17** 1193–200

Chang H-C, Wu H-T, Huang P-C, Ma H-P, Lo Y-L and Huang Y-H 2020 Portable sleep apnea syndrome screening and event detection using long short-term memory recurrent neural network *Sensors* **20** 6067

Chesson A L Jr *et al* 1997 The indications for polysomnography and related procedures *Sleep* **20** 423–87

Ghahjaverestan N M *et al* 2021 Relative tidal volume and respiratory airflow estimation using tracheal sound and movement during sleep *J. Sleep Res.* **30** e13279

Ghassemi M M *et al* 2018 You snooze, you win: the physionet/computing in cardiology challenge 2018 *2018 Computing in Cardiology Conf. (CinC)* pp 1–4

Guilleminault C, Stoohs R, Clerk A, Cetel M and Maistros P 1993 A cause of excessive daytime sleepiness. The upper airway resistance syndrome *Chest* **104** 781–7

Hafezi M, Montazeri N, Saha S, Zhu K, Gavrilovic B, Yadollahi A and Taati B 2020 Sleep apnea severity estimation from tracheal movements using a deep learning model *IEEE Access* **8** 22641–9

Issa F, Morrison D, Hadjuk E, Iyer A, Feroah T and Remmers J 1993 Digital monitoring of sleep-disordered breathing using snoring sound and arterial oxygen saturation *Am. Rev. Respiratory Dis.* **148** 1023–1023

Khandoker A H, Palaniswami M and Karmakar C K 2008 Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings *IEEE Trans. Inf. Technol. Biomed.* **13** 37–48

Kwon Y, Khan T, Pritzker M and Iber C 2014 Circulation time measurement from sleep studies in patients with obstructive sleep apnea *J. Clin. Sleep Med.* **10** 759–765A

Malhotra A *et al* 2021 Metrics of sleep apnea severity: beyond the apnea-hypopnea index *Sleep* **44** 1–16

Masa J F *et al* 2009 Apnoeic and obstructive nonapnoeic sleep respiratory events *Eur. Respiratory J.* **34** 156–61

Nakano H, Tanigawa T, Furukawa T and Nishima S 2007 Automatic detection of sleep-disordered breathing from a single-channel airflow record *Eur. Respiratory J.* **29** 728–36

Nassi T-E 2021 *Algorithms for Automated Scoring of Respiratory Events in Sleep* Master's thesis, University of Twente

Nutt D, Wilson S and Paterson L 2008 Sleep disorders as core symptoms of depression *Dialogues Clin. Neurosci.* **10** 329–36

Ogilvie R P and Patel S R 2017 The epidemiology of sleep and obesity *Sleep Health* **3** 383–8

Oliver N, Garg A and Horvitz E 2004 Layered representations for learning and inferring office activity from multiple sensory channels *Comput. Vision Image Understanding* **96** 163–80
Park S *et al* 2020 Polysomnographic phenotype as a risk factor for cardiovascular diseases in patients with obstructive sleep apnea syndrome: a retrospective cohort study *J. Thorac Dis.* **12** 907–15
Pépin J L, Guillot M, Tamisier R and Lévy P 2012 The upper airway resistance syndrome *Respiration* **83** 559–66
Pombo N, Garcia N and Bousson K 2017 Classification techniques on computerized systems to predict and/or to detect apnea: a systematic review *Comput. Methods Prog. Biomed.* **140** 265–74
Pourbabaee B, Patterson M H, Patterson M R and Benard F 2019 SleepNet: automated sleep analysis via dense convolutional neural network using physiological time series *Physiol. Meas.* **40** 084005
Rabiner L R 1989 A tutorial on hidden Markov models and selected applications in speech recognition *Proc. IEEE* **77** 257–86
Saha S *et al* 2020 Portable diagnosis of sleep apnea with the validation of individual event detection *Sleep Med.* **69** 51–7
Sharma H and Sharma K K 2016 An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions *Comput. Biol. Med.* **77** 116–24
Song C, Liu K, Zhang X, Chen L and Xian X 2015 An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals *IEEE Trans. Biomed. Eng.* **63** 1532–42
Suzuki E *et al* 2009 Sleep duration, sleep quality and cardiovascular disease mortality among the elderly: a population-based cohort study *Prev. Med.* **49** 135–41
Thorey V, Hernandez A B, Arnal P J and During E H 2019 AI vs Humans for the diagnosis of sleep apnea *2019 41st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* pp 1596–600
Timuş O H and Bolat E D 2017 k-NN-based classification of sleep apnea types using ECG *Turk. J. Electr. Eng. Comput. Sci.* **25** 3008–23
Travieso C M, Alonso J B, Ticay-Rivas J R and del Pozo-Baños M 2011 Apnea detection based on hidden Markov model Kernel *Int. Conf. on Nonlinear Speech Processing* pp 71–9
Xie B and Minn H 2012 Real-time sleep apnea detection by classifier combination *IEEE Trans. Inf. Technol. Biomed.* **16** 469–77