Unobtrusive Screening of Central Sleep Apnea From Pressure Sensors Measurements: A Patient-Specific Longitudinal Study

Hilda Azimi, *Student Member, IEEE*, Martin Bouchard, *Senior Member, IEEE*, Rafik Goubran, *Fellow, IEEE*, and Frank Knoefel

***Abstract*— Historically, the lack of patients' sleep histories has caused low identification of sleep apnea (SA) and referral rates. Moreover, the costly and time-consuming nature of polysomnography (PSG) as a standard clinical test for detecting SA and the lack of sleep clinics has created a demand for suitable home-based monitoring devices. Pressure measurement using a pressure sensitive mat (PSM) can address the challenges found in current sleep-monitoring solutions. The noncontact PSM has a potential to replace obtrusive breathing sensors in the sleep lab and to be used as a prescreening tool for patients suspected of having SA. Applying classical support vector machine (SVM), this article presents a personalized system based on the measurements of each patient to detect central SA (CSA) events and monitor sleep characteristics longitudinally. For this purpose, sensor set-ups were installed in nine seniors' homes to collect unsupervised pressure data in approximately one year ranging from 8 to 12 months. Cost-based and resampling-based approaches were examined to combat imbalanced data. The results showed that the cost-based method outperformed other methods. Next, the patient-specific system was used to determine the total number of CSA events, as well as their starting time and duration in each day. The SA severity was measured by the central apnea index (CAI). In addition, other sleep characteristics such as bed occupancy (BO), day clock, and night clock were extracted from the PSM measurements. The impact of longitudinal sleep monitoring could be in tracking SA treatment progression, and possibly providing information on the interaction between SA and other disease progressions.**

Manuscript received September 19, 2019; revised February 26, 2020; accepted March 3, 2020. Date of publication March 16, 2020; date of current version May 12, 2020. This work was supported in part by the AGE-WELL Government of Canada Networks of Centres of Excellence program and in part by a National Sciences and Engineering Research Council (NSERC) of Canada Discovery grant. The Associate Editor coordinating the review process was Bruno Ando. *(Corresponding author: Hilda Azimi.)*

Hilda Azimi and Martin Bouchard are with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: hazim009@uottawa.ca; bouchm@uottawa.ca).

Rafik Goubran is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada, also with the Bruyère Research Institute, Ottawa, ON K1R 6M1, Canada, and also with AGE-WELL National Innovation Hub on Sensors and Analytics for Monitoring Mobility and Memory (AGE-WELL NIH-SAM3), Ottawa, ON K1N 5C8, Canada (e-mail: goubran@sce.carleton.ca).

Frank Knoefel is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada, also with the Bruyère Research Institute, Ottawa, ON K1R 6M1, Canada, also with AGE-WELL NIHSAM3, Ottawa, ON K1N 5C8, Canada, and also with the Faculty of Medicine, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: fknoefel@sce.carleton.ca).

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

Digital Object Identifier 10.1109/TIM.2020.2981111

***Index Terms*— Biomedical measurement, data analysis, machine learning, patient monitoring, pressure measurement, sleep apnea (SA).**

I. Introduction

SLEEP is a dynamic process that varies from day-to-day [1]. Therefore, it is essential to measure multiple nights of sleep data for health, medical, and research reasons. In this regard, when compared with single-night polysomnography (PSG) as the current clinical test for diagnosing sleep disorders, home monitoring is preferred. It offers the potential to provide a more realistic platform, capturing many nights of sleep data. Longitudinal home monitoring allows tracking the sleep dynamics of the same patient at different points in time and reducing the between-subject variation of the measurements. One of the most notable examples of individual variability in sleep involves sleep apnea (SA) symptoms. In the field of sleep medicine, SA is described as the most irritating sleep disorder, causing sleep disturbance. It is characterized by repeated periods of reduction or complete cessation of airflow. Clinically, SA is defined when there is a drop in the peak signal excursion by at least 90% of the pre-event baseline for 10 s or more [2]. There are three forms of SA: central SA (CSA) characterized by a complete cessation of both respiratory movements and airflow, obstructive SA (OSA) characterized by the presence of abdominal and thoracic efforts for continuing breathing, while airflow completely stops, and mixed SA (MSA) defined by a CSA followed by an OSA. Given the absence of respiratory movement in CSA, our work using PSMs focuses on this SA subtype.

The structure of this article is as follows. Section II provides a summary of studies conducted to detect SA through unobtrusive home monitoring schemes. Section III covers the methodology for the design and optimization of patient-specific systems adopting PSM measurements. The outcome of different optimization processes and the obtained sleep measures resulting from the methodology applied are presented and discussed in Section IV. As a final point, Section V briefly restates the article and summarizes the main results and findings of this article.

II. Related Work

For home monitoring without any assisting personnel, especially for those without a high degree of technical knowledge,

TABLE I
SLEEPMINDER PERFORMANCE [3]

	Sensitivity	Specificity
AHI≥5	86.1	46.4
AHI>10	83.6	84.0
AHI≥15	88.7	92.1
AHI≥20	77.3	89.4

the recording of biosignals must be easy, without reducing the subject's comfort during sleep. Thus, a good sleep monitoring system: 1) should record data without restricting patient movements (e.g., no gauges or cables during the night); 2) Should detect SA, vital signs (e.g., breathing rate and heart rate), and movements during sleep even without attaching sensors directly to the body; and 3) Should be installed in the user's native sleep environment. For this purpose, several clinically useful sensor modalities and tools have been developed. These types of devices have the advantage of being unobtrusive and adequate for longer term monitoring without the user's intervention. For instance, in [3] a noncontact radio frequency sensor called SleepMinder was placed next to a bed and was adopted to continuously measure the biomotion due to breathing and body-movement during sleep. The device applied phase demodulation, amplitude, and correlation-based signal-processing methods to detect SA. Comparing with PSG recordings of 129 subjects, the apnea-hypopnea index (AHI) estimated by the device had a correlation of 91%. The sensitivity and specificity of SA detection for different AHI boundaries are presented in Table I.

Some articles have employed visual information in the form of photographs, film, or video signals to detect SA. In [4], it was shown that support vector machine (SVM) and neural network (NN) classifiers with depth video and audio signals recorded by a Microsoft Kinect camera could be used to detect and differentiate SA events, including OSA, CSA, and MSA. However, it was claimed that the proposed system required large storage for data recording, and it was relatively slow in the fitting of a respiratory model and training classifier for each patient. In [5], respiratory events were detected from a series of digital images provided by a digital video camera. The technology utilizes the circulation of air into the lungs, which is proportional to the patient's movement while breathing. The proposed apnea detection algorithm compared signal variations from the mean respiratory movement as a reference point for the magnitude, detected in data from the previous minutes. In the analysis of 50 patients, the system was found to have a sensitivity of 100% and a specificity of 83%. The positive predictive value (PPV) and negative predictive value (NPV) were 97% and 100%, respectively.

Force sensors placed on top or under the mattress have also been adopted for sleep monitoring, estimation of snoring periods, and vital signs such as heart rate and respiration rate. On this basis, a thin foam mattress with four polyvinylidene fluoride film-based sensors having dimensions that fit a standard single-sized bed has been used to diagnose OSA [6]. In addition to respiratory movements, the mattress was also able to record breathing sound. Apneic events captured by the mattress were identified visually and not by an automated algorithm, then scored and compared with the result of relative PSG recordings. There were strong correlations between the two devices for measures of sleep time, respiratory events, and the AHI (all correlations > 0.89). The device defined more than 93% of PSG defined respiratory events. The absence of respiratory events was correctly identified in 91% of occasions.

In comparison with [6], an air mattress with a balancing tube was presented in [7] to noninvasively monitor not just the respiratory signal but also heartbeat, as well as the events of snoring, SA, and body movement. The proposed system consisted of multiple cylindrical air cells, two sensor cells, and 18 support cells. The physiological signals were measured by the change in the pressure difference between the sensor cells. Healthy participants were asked to simulate CSA to investigate the ability of the system for the detection of apneic events through a simple threshold-based detection algorithm based on a variance analysis with a moving window technique. The algorithm achieved sensitivity and PPV of 93% and 88%, respectively. Based on similar functionality, in [8] a force-coupling pad equipped with a ballistocardiography (BCG) system was installed on top of the mattress of a bed to detect minute forces generated during cardiac contraction and relaxation, and body movement from the respiratory effort and postural changes. PSG of participants was recorded while the force-coupling pad was on the mattress. A sensitivity of 89.2% and specificity of 94.6% were attained in the detection of SA. The respiration analysis was partially automated. The manual portion of the algorithm involved applying a set of rules for marking apneas and arousals based on the scoring system and classification of breaths.

All these studies commonly attempt to examine the ability of their alternative system to detect SA unobtrusively applying different detection methods, sometimes for each patient individually. Mostly, the validation of the proposed methods was due to the paired comparison of the total number of respiratory events detected by an automated algorithm and those manually scored.

Pressure sensitive mats (PSM) adopted in this article can capture movement and breathing information while placed beneath a mattress [9], [10]. In [11], the validation and the assessment of PSM reliability for CSA detection were presented, with a comparison to PSG data by applying a simple decision threshold method. We recently designed a generalized detection system to detect CSA events from nocturnal data automatically [12]. The classifier achieved a better sensitivity of 96.3% in comparison to 87.6% in [11]. However, it is worthwhile to note that when compared with this article, all of the reviewed methods and studies have used different original sleep data. Therefore, a direct comparison of different results only has limited reference significance.

In this article, we expanded our previous algorithm for automatically detecting CSA events [12] for long-term sleep

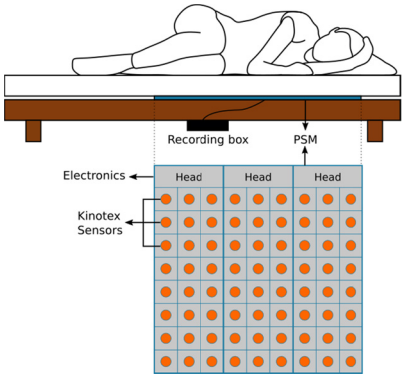

Fig. 1. PSM and measurement system.

monitoring. The focus of the previous study was to provide a generalized model for detecting CSA events, regardless of the sleep characteristics of individuals. While it could be represented as a primary diagnosis tool, leading to a significant reduction in the diagnostics cost and waiting time for a sleep study, considering subject-specific factors can greatly improve the accuracy of the system. Here, the goal is running longitudinal design for long-term investigation of intrasubject changes in sleep quality of older adults, extracted from unobtrusive pressure sensor array data (collected in approximately one year for each patient). The system is not only able to report the total number of extracted respiratory events but also indicate whether the events are detected at the same temporal position of manually labeled events or not, which may potentially prevent to mask inaccuracies in the detection of events by the device. With this scheme, the stability of sleep characteristics is to be determined within a subject. This approach has substantial clinical utility given the known night-to-night variability of SA [13]–[15].

III. Methodology

A. Database and Measuring System

The PSMs are the "bed occupancy sensors" (BOS) model from S4 Sensors Inc. (Victoria, B.C., Canada). Each set of PSM contains 72 Kinotex fiber optic sensors in total, distributed evenly, as shown in Fig. 1. The sensors are embedded in foam as an eight by nine matrix, with eight sensors lining the width and nine sensors lining the length of the bed, covering from head to hip of the patient. The head of each panel is a strip that contains the electronics for analog to digital conversion, sampling, and transmission of the sensor data. The device is connected to a recording box to save data for further processing. The recording box is placed under the bed and saves the data on a secure digital (SD) card.

Along with obesity, age is one of the most important risk factors for SA. The prevalence of OSA is two to three times higher in older persons ($\geq$65 years) when compared with those in middle age (30–64 years) [16]. Therefore, the main focus for data collection was to monitor adults aged 65 or older, living in affordable seniors housing.

To collect nonsupervised data, volunteers from different communities passed some inclusion criteria, which were as follows:

1) sixty-five years of age or more, female/male;
2) being community-dwelling older adults living in affordable seniors housing;
3) admitted to Geriatric Rehabilitation Unit at Élisabeth Bruyère Hospital.

Sensor setups were installed in the seniors' homes for long-term monitoring and collecting pressure data continuously, with no requirement from the patient for installation and data acquisition. The monitoring time was 8–12 months. Except for one visit every month, the rest of the data collection was not supervised. Recorded data were in ASCII format and autosaved as CSV files hourly. Files were concatenated on a day-by-day basis for 24 h, from noon to noon to have a daily observation, especially for sleep assessment. No other health information is known about the patients, as ethics clearance did not include access to patient health records. A week of data from each of the nine participants randomly selected for [12] is also employed for this article to optimize the individual CSA detection models.

B. Preprocessing

Preprocessing is a set of steps that are implemented and applied sequentially. In such a way, the output signal of each step is the input of the next step. The steps are as follows:

1) Occupancy Extraction: As a subject leaves the bed, a significant drop in PSM signal amplitude occurs. To decrease the computational complexity and mostly to discard irrelevant data, times when the bed was not occupied were detected and removed based on the algorithm in [17]. The basic idea behind the adopted algorithm was first introduced in [18]. However, the threshold selection in the algorithm depended on anthropometric information of participants. Therefore, for times when no prior anthropometric information was available (such as in this article), some modifications were applied to the original algorithm in [17] to select a proper threshold.

In the algorithm, a single output is first obtained from the summation of all 72 raw sensor signals, from noon to noon for each day, denoted as P. The occupancy is then determined when P exceeded a threshold, T, defined by

$$T = \beta + \tau \tag{1}$$

where β is the base value obtained as the minimum value of P. The value of τ is established based on using the maximum and minimum value of P to make it independent from the person who is lying in bed, respectively, denoted by $\mathrm{Max}(P)$ and $\mathrm{Min}(P)$, such that

$$\tau = \frac{\mathrm{Max}(P) - \mathrm{Min}(P)}{n} \tag{2}$$

where n can be adjusted depending on the sensitivity required in an application. Those parts of P where the difference between $\mathrm{Max}(P)$ and $\mathrm{Min}(P)$ is below the threshold are reflagged as not-occupied to avoid false detection of occupancy.

The occupation data obtained in this step are kept for further processing, and the rest of the data are discarded.

2) Bandpass Filtering: The average respiration rate for an adult at rest is 12 to 20 breaths per minute (bpm). However, the full range of possible respiration rates, including extreme conditions, can range from 4 to 48 (bpm), which is related to the frequency range of 0.07 to 0.8 Hz [9]. Therefore, for the extraction of the breathing signal, a 70th-order finite impulse response (FIR) linear phase bandpass filter with a passband corresponding to the frequency range of 0.07 to 0.8 Hz is employed to filter the signal from each sensor. The reason for choosing the linear-phase FIR filter is to avoid changing the time shape of the useful time signal components. After filtering, data at the edges (beginning and end) of the output signal were discarded due to the "edge effect."

3) Combining and Concatenating Signals: Signals of PSM placed below a mattress can be heavily attenuated and localized loading may be less distinct [19]. To overcome this issue, all 72 signals can be combined to generate a single output signal with better signal quality. In ambient systems, using the linear combining paradigm, sensor signals are conditioned by preprocessing, multiplied by a gain factor, and then summed. By adopting this procedure, in [9] two breathing signal fusion methods were proposed. They are Pearson's correlation coefficient (PCC)-based signal fusion and signal-to-noise ratio (SNR)-MAX-based signal fusion. Later in [20], for more complex data containing movement in bed and different types of breathing, such as shallow breathing, deep breathing, and periods of apnea, the SNR-MAX was found to be the best method of sensor signal combining, resulting in the highest Pearson Correlation Coefficient with the respiratory band signal as the gold standard. Therefore, in this section, the SNR-MAX method is applied to the 72 sensor signals to generate a single output signal with better signal quality. Signals are combined every 30 s, with a 50% overlap.

Different window sizes may be needed to detect CSA, especially since the minimum acceptable length for CSA events is 10 s. Therefore, after signal combining for every 30-s segments, all segments are concatenated to have one single signal. The following method is applied to achieve a smooth transition cross overlapping segments during concatenation [21], [22].

1) For every two consecutive segments, the last N overlapping samples of the first segment are gradually attenuated by multiplying with a linear cross-fading function changing from 1 to 0 ($f(n); n = 1, \ldots, N$), while the first N samples of the second segment are gradually amplified by multiplying with $1 - f(n)$.
2) All modified segments are aligned based on their starting point in the original signal and then added together to construct a concatenated signal.

From this stage onward, the rest of the algorithm uses this single signal and not an array of signals.

4) Signal Normalization: PSM does not directly measure the airflow or respiratory effort. It captures movement information transmitted through the mattress by a person lying on it. Changing the body position or posture on the bed is always accompanied by movements including bed entries/exits, rollovers, sleep starts, posture shifts, or small movements such as arm/leg twitches and gasps (deep breathing), which generate large fluctuations in the signal. Several studies have been done to detect movement [23]–[25]. This article adapted the previously published movement detection algorithm in [23] to detect movements, with a few modifications in terms of defining a threshold. Here, a movement is detected when the value reaches more than three-scaled median absolute deviations (MADs) away from the median of signal y. For y of length N, the MAD is defined as

$$\text{MAD} = c \times \text{median}(|y_i - \text{mean}(y)|); \quad i = 1, \ldots, N \tag{3}$$

where c is a constant scale factor.

The amplitude of the PSM signals and consequently y can vary based on body posture and position, respectively. The effect of body position on the output of PSMs has been investigated in a few article s such as [26]. Through the supervised data collection, a comparison of different body positions including side, supine, and prone showed that signals captured from the supine position have tended to have the lowest SNR [26]. Therefore, after detecting movements, each part of the signal Z between any two movements is normalized to a comparable range to reduce the number of false-positive events. For normalization, first the average of Z, Z_{Mean}, is removed. Then, Z is divided by Z_{Max}, which is calculated as the maximum absolute value of Z, excluding the highest 5% and lowest 5% of the values to ignore outliers. Sorting data to find outliers can be computationally expensive. However, dividing Z by Z_{Max} without ignoring outliers might result in a signal with a much smaller breathing range, which can potentially be confused with an apnea event. Z_{Max} and the subtracted mean Z_{Mean} are added as two features for each segment to avoid information loss, especially when Z is too short and mostly consists of noise.

C. Feature Extraction

For automatic detection of CSA, different algorithms are applied to extract suitable features. Some features are the result of time series data, directly fed to the classifier, whereas others first involve a transformation of time series data into the frequency domain. To extract features, the normalized signal is segmented by rectangular 9-s window functions with 50% overlap, whose length is chosen based on the best performance of the system on the training database. It was observed that by increasing the window duration beyond 10 s would reduce the performance of the proposed system as the CSA events can be of 10 s by definition. Increasing the window duration affects the performance of the event detector (which will be described later in the article) rather than the classifier. Contrarily, if the window length is too small and does not contain the major part of a respiration cycle, the performance of the classifier drops. However, the performance of the event detector to correct the misclassified windows and the detection of events improve. By considering all these factors and the number of calculations required, a window duration of 9 s is chosen. The next parameter which is adjusted is the amount of overlap. It is expected that a small amount of shift (large overlap) will improve the performance by increasing the resolution of the entire event detection process but will also impose a

computational burden. The amount of overlap was chosen to be 50%, since on average (at a rate of 15 bpm) it is long enough to provide a complete respiratory cycle to the next window.

The resulting segments are categorized into two classes: apneic event class, "A" or normal breathing class, "N." "A" labeled segments are those that have at least 50% of their length falling within the start-to-end time of apnea events. All the other remaining segments are labeled as "N." For the classification of these segments into two categories, 34 features from the time and frequency domain are extracted from each segment.

1) Time Domain-Based Features: Time series analysis includes methods for analyzing time-series signals directly in order to extract meaningful statistics and other characteristics of the signal. The brief descriptions of different time domain-based measures extracted in this article are as follows.

2) Statistical Measures: These first- to fourth-order statistical parameters, i.e., mean (M1), variance (M2), skewness (M3), and kurtosis (M4), are computed for each segment of data to quantify the central tendency, degree of dispersion, asymmetry, and peakedness, respectively. Other statistical measures extracted from time series segments are: the median as the measure of the "center" of a signal since it is less sensitive to outliers, the maximal (Max) and minimal (Min) value of the signal, the range as the difference between the Min and Max values of a signal, the root-mean-square (rms), and the standard deviation (SD).

3) Subsegment Ratio: Each segment is divided into five subsegments. Then, the maximum absolute value is obtained in each subsegment. Afterward, the SSR is calculated for each segment based on dividing the highest maximum absolute level by the lowest maximum absolute level from the five subsegments [27].

4) Hjorth Parameters: Hjorth parameters [28] are quite popular for the analysis of biomedical signals, especially EEG signals for sleep stage classification [29], [30]. For a signal x of length N, and x' as its first derivative, the parameters are calculated using the following equations:

$$\text{Mobility} = \sqrt{\frac{\text{Var}(x')}{\text{Var}(x)}} \tag{4}$$

$$\text{Complexity} = \frac{\text{Mobility}(x')}{\text{Mobility}(x)} \tag{5}$$

where $\text{Var}(x)$ and $\text{Var}(x')$ are the variance of x and x', respectively. x' is approximated by the difference between successive elements of x.

Nonlinear combination (i.e., multiplication and division) of existing features can be considered as new features and be helpful to optimize the linear classifier. A linear classifier cannot consider these complex features unless they are explicitly provided for the classifier. f1 to f4 listed in Table II are adopted based on this principle. A brief explanation of these features can be found in [31].

Time-domain

5) Frequency Domain-Based Features: based features are not always the best representation of signals for many signal-processing applications. In many cases, the most useful information is hidden in the frequency content of the signal. Herein, from the auto-power spectral density (PSD) of each segment, the following features were extracted. In this section, wherever it speaks of the frequency band of interest, it means 0.2–0.33 Hz, corresponding to normal breathing range (i.e., 12–20 bpm).

TABLE II
LIST OF THE NONLINEAR COMBINATION OF EXISTING FEATURES

Label	Formula/Description
f1	$\frac{Max^a(x)}{RMS^b(x)}$
f2	$\frac{RMS(x)}{Mean\left(Abs^c(x)\right)}$
f3	$\frac{Max(x)}{Mean\left(Abs(x)\right)}$
f4	$\frac{Max(x)}{Mean(Abs(x))^2}$

[a] Max: Maximum value.
[b] RMS: Root Mean Square value
[c] Abs: Absolute value.

6) Spectral Moments: Median and SD, as well as the first- to fourth-order spectral moments in the frequency domain, were calculated. These features characterize the shape of the spectral density of the signal in each segment. In [32], the spectral moments were applied to classify sleep stages from PSG recordings.

7) Peak Power Frequency: It is the frequency at which the highest power exists. Also, peak power frequency (PPF) corresponding to the frequency band of interest PPF_R and its proportion to PPF (PPF_R/ PPF) are calculated as two new features.

8) Spectral Flatness Measure: The ratio of the geometric mean of the signal PSD to the average of the signal PSD is a measure of the spectral flatness of the signal (SFM) also known as tonality coefficient. The meaning of tonal in this context is in the sense of the number of peaks or resonant structure in a power spectrum, as opposed to a flat spectrum of white noise. A high-spectral flatness represents a similar amount of power in all spectral bands, as in white noise. Therefore, the graph of the spectrum would appear relatively flat and smooth. A low-spectral flatness indicates that the spectral power is concentrated in a relatively small number of bands. Therefore, the spectrum would appear spiky, such as a mixture of sine waves [33]

$$\text{SFM} = \frac{\text{GeometricMean}(x)}{\text{Mean}(x)}. \tag{6}$$

9) Relative Spectral Powers (P_R): The percentage of the total area under the curve corresponding to the frequency band of interest.

The features listed in Table II are also calculated as frequency domain-based features where, in this case, x would be

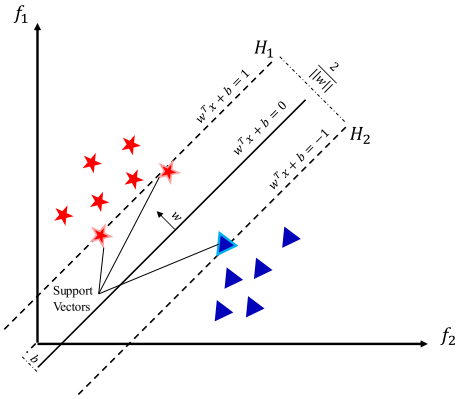

Fig. 2. Maximum margin hyperplane and margins for SVM trained with samples from two classes.

the power spectrum, corresponding to the frequency band of interest. Overall in total for each patient, the extracted features from time and frequency domains, in addition to Z_{Max} and Z_{Mean} from the normalization procedure, produce 34 features to be fed the patient-specific classifiers.

D. Classifier

The ultimate goal of the pattern recognition problem is to classify the objects into a number of categories or classes. For this purpose, the role of the classification models is to divide the feature space into disjoint regions assigned to class labels. The classification models are divided into different groups, including Naïve Bayes, linear discriminant analysis (LDA), NNs, kernel methods, similarity-based approaches, and decision trees. Recently, SVM as a kernel-based classifier was found to have much advanced theoretical background and often provide good performance, especially for binary classification in the field of biomedical pattern recognition [34]. Initially, SVM was designed to solve binary classification and regression problems. Nowadays, SVM has been successfully applied in many areas [35]–[37].

Finding an optimized hyperplane is the fundamental idea of SVM. Such a goal can be defined as the maximization of the minimum distance, i.e. margin, between the hyperplane, and the closest training data points, i.e., support vectors

$$\max_{w,b} \min\left\{||x - x_i|| : w^T x + b = 0, i = 1, \ldots, m\right\} \quad (7)$$

where w is the weight vector and b is the bias for the model. In linear binary SVM classification, w and b can be rescaled in such a way that the point closest to the hyperplane $w^T x+b = 0$ lies on hyperplanes $H_1 : w^T x+b = 1$ and $H_2 : w^T x+b = -1$ for either class, as shown in Fig. 2.

Suppose that all the training data satisfy the following constraints:

$$w^T x_i + b \geq 1 \text{ for } y_i = +1 \quad (8)$$

$$w^T x_i + b \leq -1 \text{ for } y_i = -1. \quad (9)$$

These can be combined into one set of inequalities

$$x_i : y_i\left[w^T x_i + b\right] - 1 \geq 0. \quad (10)$$

As it is shown in Fig. 2, the width of the margin is simply equal to $2/||w||$. Support vectors are the training points for which the equality in (10) holds.

The optimization problem in (7) can be restated as $\tau(w)$: $\min_{w,b} \tau(w) = (1/2)||w||^2$, subject to the constraint (10) to find the pair of hyperplanes, H_1 and H_2, which gives the maximum margin by minimizing $||w||$ [38].

In comparison with nonlinear kernels, a linear kernel is less prone to overfitting and faster to train. However, depending on data, a different type of kernel may be required. Comparing the cross-validation data, the results of various types of kernels including RBF, third-order polynomial, and linear kernel showed that the use of nonlinear kernels did not significantly improve the accuracy of the model in the data set. Therefore, a binary linear SVM classifier is considered to classify segments into two categories, i.e., "N" or "A." Those segments that overlap with the detected movement are considered as outliers, and they are eliminated from data before training the classifier.

E. System Optimization

1) Application of Learning Curve Concerning the Amount of Data: When it comes to building a personalized model, the first important question is how much data are needed to provide an optimized model for long-term monitoring. Herein, for each patient, a week of labeled data is split into different sizes of training and validation sets. Then, learning curves are plotted to show how the error changes as the size of the training data set increases. As a common issue among all patients, the distribution of the two classes, A and N, is imbalanced. Therefore, for every iteration, balancing the training data set, as well as the presence of apneic segments from class "A" in the training data set, is necessary to generate the validation curves. A random selection of segments from class "N," the majority class, has to be performed in each iteration to overcome the imbalanced data problem such that both classes have the same size. The procedure is well known in the literature as the "undersampling" method [39]. As a result of this technique, the maximum training data set is two times L, where L is the total number of class "A" instances in the main training data. The procedure of generating the learning curve is repeated seven times (as the total number of days of labeled data) to perform cross-validation. It means that in each fold, data of one day is put aside as a test data set, and the remaining data of six days are used to train the classifier.

Moreover, in each fold, to investigate how much the system performance depends on the characteristics of the selected training data set, the data are randomly shuffled 50 times before creating the learning curve. Once the data are shuffled, no further changes is made in the order of instances, so that one complete learning curve can be generated. The training is performed over iterations. In each iteration, based on the order of the instances in the data matrix, a new unseen patch of data including examples from class "A" and examples from class "N" is added to the existing training data set. Fig. 3 illustrates the summary of the generation of the learning curve for each patient.

2) Dealing With Imbalanced Classes: As it was discussed in the previous section, the CSA detection problem is faced

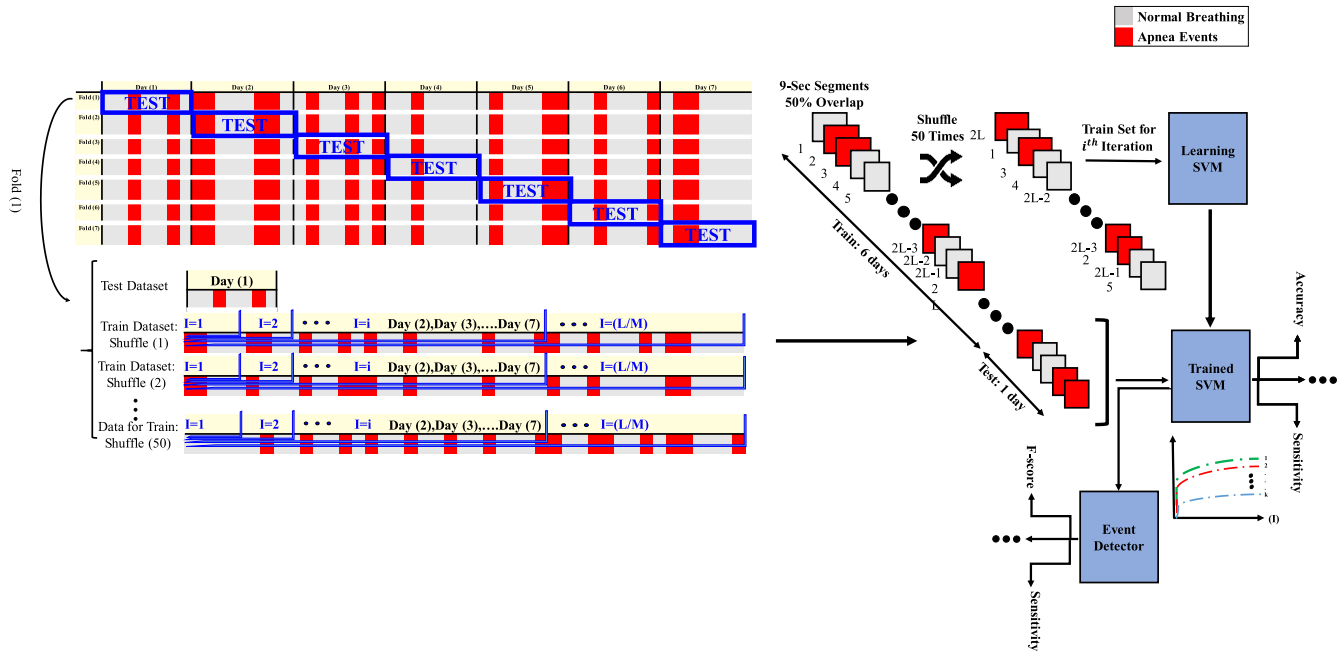

Fig. 3. Process of generating a learning curve for each patient.

with imbalanced data, where class "N" as a majority class dominates over class "A," the minority class. The problem causes the model to be more biased toward the majority class and to suffer from the accuracy paradox. There are many approaches to deal with this problem. They can be generally classified into two major categories of: 1) sampling-based and 2) cost-based. Sampling-based methods can be broken into three major categories: a) oversampling; b) undersampling; and c) hybrid of oversampling and undersampling. In [12], a combination of oversampling of class "A" instances and undersampling of class "N" examples was implemented. To avoid overfitting, the oversampling approach was not done by replicating class "A" instances but instead by constructing new minority class data instances via the synthetic minority oversampling technique (SMOTE) algorithm. SMOTE creates new instances of the minority class by forming convex combinations of neighboring instances, i.e., effectively drawing lines between minority points in the feature space, and samples along these lines [40].

In this article, the diversity of the data is narrowed down to only one patient's information, and for each patient a limited amount of data is available. Sampling-based methods change the original data distribution, causing loss of important information or the model overfitting [41]. Therefore, given these two issues instead, a cost-based method is used to address the class imbalance problem. A cost-sensitive classifier learns more characteristics of the minority-class instances by setting a high cost to the misclassification of a minority-class sample.

For cost-sensitive SVM in binary classification, the objective is to determine a hyperplane that maximizes the margin while minimizing a quantity proportional to the misclassification errors. This can be done by introducing positive slack variables ξ_i in (10), which then becomes

$$x_i : y_i\left[w^T x_i + b\right] \geq 1 - \xi_i. \tag{11}$$

Hence $\tau(w) = (1/2)||w||^2$ can be changed into (12) [34]

$$\tau(w, \xi) = \frac{||w||^2}{2} + \left(C \sum_{i=1}^{n} \xi_i\right). \tag{12}$$

The algorithm uses ξ_i to penalize the objective function for those observations that cross the margin boundary for their class; $\xi_i = 0$ for observations that do not cross the margin boundary, otherwise $\xi_i \geq 0$. C controls the trade-off between the margin and the misclassification errors. A larger C means that a higher penalty is assigned to misclassification errors.

F. Event Detector

The time sequence of consecutive feature vectors for each day of a subject data is fed one by one to the already trained and optimized classifier, to predict the possible classes. The output of the classifier is the series of class labels, separated by an interval equal to the window sliding duration. These time sequences of class labels are then fed to the event detector. During the process of event detection, first, the windows which contain movements are ignored by the classifier model; whether they are labeled as "A" or "N" since no information related to movements is presented to the classifier while it is being trained. A sequence of segments which are all flagged as events are merged as a single event. A common duration of an apneic event is about 20–40 s [42]. Therefore, detected events with a duration of more than 60 s are reflagged as "N" events.

G. Descriptive Statistics of the Sleep Measures

For each patient, after optimizing the cost matrix, the final classifier is trained on all data from 7 days. Then the data of the whole year are fed to the system day-by-day to extract sleep-related information. In this section, the main findings of the longitudinal study are described and broken down into the specific sleep measures.

1) Bed Occupancy: BO is defined as the total time spent in bed during the 24-h cycle. For every 24-h cycle from noon to noon the next day, BO is extracted by the occupancy extraction algorithm, as a part of the preprocessing procedure. Bed-occupied time can be described as "intention to sleep," and it includes nighttime sleep and daily naps. BO is also calculated separately for the day clock (defined as 10 A.M. to 10 P.M.) and the night clock (defined as 10 P.M. to 10 A.M.).

2) Central Apnea Index (CAI): CAI is defined by the number of CSA events detected per hour of BO. CAI can indicate the severity of the CSA, similar to the AHI.

3) CSA Duration per Night: As the output of the system, in addition to the number of extracted events and their temporal position for each night, the duration of each event is investigated.

H. System Evaluation

The event-based evaluation of the proposed system compares the proposed system outputs and the corresponding reference on an event-by-event basis [43]. The F-score judges the performance of event detection. The F-score is computed from the sensitivity and precision

$$\text{Fscore} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{13}$$

where sensitivity and precision are calculated by (14) and (15)

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \times 100\% \tag{14}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \times 100\%. \tag{15}$$

Herein, the event-based TP as true positive, FP as false positive, and FN as false negative represent the following conditions [44].

1) *TP:* An event in the output of the proposed method that has a temporal position overlapping with the temporal position of an event with the same label in the reference ground truth. Based on the definition of TP, multiple covering events are counted as a correct detection too.
2) *FP:* An event in the proposed method output that has no correspondence to an event with the same label in the reference ground truth.
3) *FN:* An event in the reference ground truth that has no correspondence to an event with the same label in the proposed method output.

It should be noted that the event-based metrics have no meaningful true negatives (TNs) [44].

Precision is a measure of exactness (i.e., of the events labeled as apnea, how many are labeled correctly), whereas sensitivity is a measure of completeness (i.e., how many apneic events are labeled correctly). These two performance metrics, much like accuracy and error, share an inverse relationship between each other. An attempt to maximize precision usually leads to lower sensitivity values and vice versa. However, unlike accuracy and error, precision and sensitivity are not both sensitive to changes in data distributions. Inspection on the precision and sensitivity formulas yields that precision is sensitive to data distributions, while sensitivity is not. An assertion based solely on sensitivity is misleading since sensitivity provides no insight into how many events are incorrectly labeled as apnea. On the other hand, precision cannot state how many apnea events are labeled incorrectly. The F-score metric combines precision and sensitivity as a measure of the effectiveness of the event detection system, in terms of a ratio of the weighted importance on either sensitivity or precision (here both are weighted equally). As a result, F-score provides more insight into the functionality of the system than the accuracy metric.

IV. Results and Discussion

Adopting PSM measurements, a personalized system was designed for longitudinal monitoring of sleep characteristics. There were some challenges in optimizing the patient-specific system, each of which has been met, and the results are further elaborated and interpreted in this section.

A. System Optimization

1) Application of Learning Curve Concerning the Amount of Data: One of the important points to be considered for designing personalized systems is the amount of data required. Fig. 4 shows the performance of the system on the test data set for each patient in terms of F-score. The plots are arranged based on the patients' average number of events per day (AED) in ascending order, from left to right and from top to bottom. Each plot contains seven curves, representing the results of different folds. Each curve denotes the average performance of the system in a 50-time shuffle. The SD of system performance in a 50-time shuffle is calculated to investigate the dependence of the system on the data characteristics, and it is displayed as shade around the average curve: the thinner the shade, the less dependent the characteristics of selected data. If the system performs properly, apart from individuals' differences and different fold of data set, as the training size increases the average performance of the classifier will increase because the model can generalize better from a higher amount of information. As the curve reaches a saturation point, it means that adding more training data points will not lead to significantly better models. Comparison of the different curves in each plot indicates how the model performs in general when used to make predictions on unseen data.

According to Fig. 4, interindividual and intraindividual differences show that the following.

1) Commonly among all patients, as the training size increases the average performance of the system increases since the system becomes more generalized.
2) As the AED among the patients increases, the overall average performance of the personalized system increases, from an F-score of 24% for the patient with AED of 6 to an F-score of 75% for the patient with AED of 210, which is the largest AED observed.
3) The system built for the patient with the largest AED (210) has the least dependence on the characteristics of the selected data and the best performance.

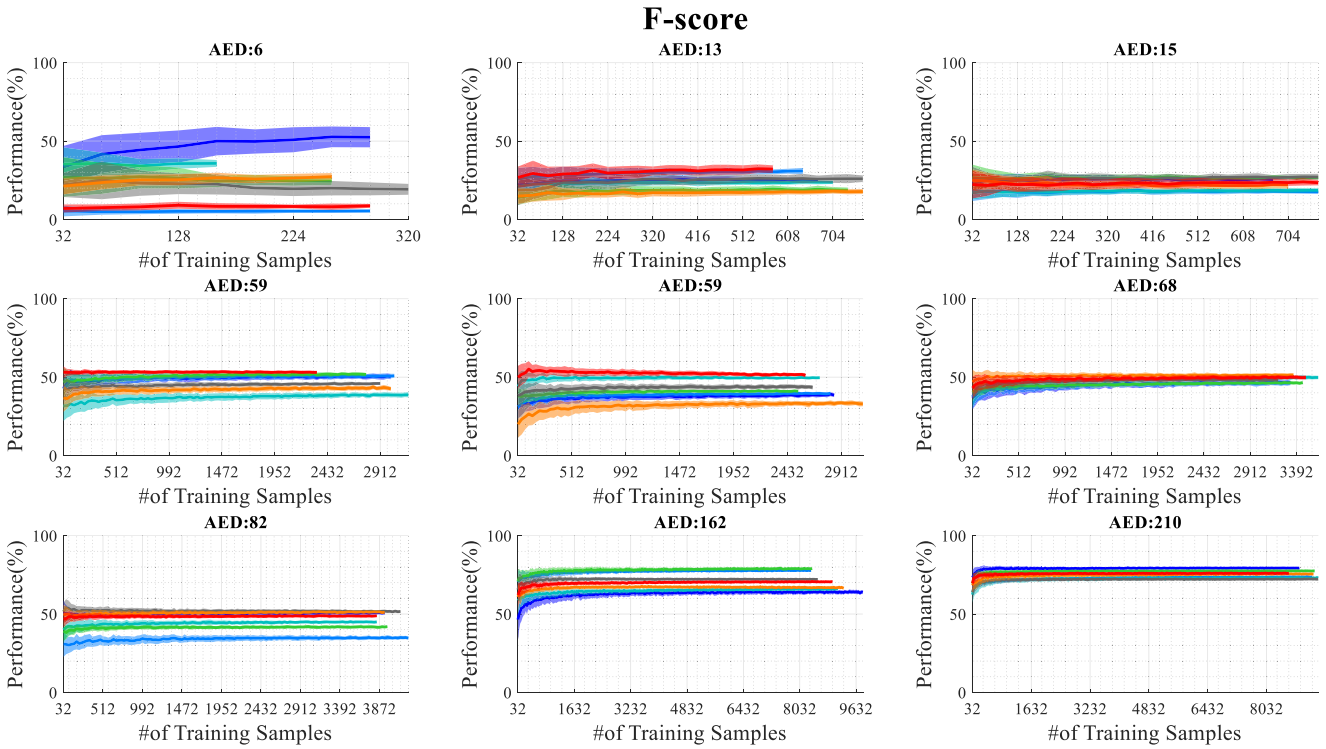

Fig. 4. System's performance in terms of F-score on the test data set for each patient. Each curve represents the result of the system for each fold.

4) The performance curves for all patients reach a saturation point, meaning that adding more training data points will not lead to a significantly better system and a week of labeled data is sufficient to build a generalized model for long-term monitoring.

Overall, the analysis shows that the more severe the disorder is, the fewer the number of days of data are needed from a patient to build a personalized model. In acute condition, i.e., for a patient with a great AED, even one day of data can be enough to optimize the model.

2) Dealing With Imbalanced Classes: After extracting the relevant features, the next challenge is to deal with imbalanced data. The cost for each patient is optimized based on the average F-score of the event detector, using a cross-validation scheme. Six out of seven days of data (six days, six folds) are used for training the classifier and consequently the system. One day of data is put aside as unseen data to evaluate the cost-optimized system. Given the cost structure array as

$$C = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix} \tag{16}$$

the cost ratio can then be defined as (b/a).

Fig. 5 depicts the cross-validation results as a function of the cost ratio. By changing the cost ratio, the F-score changes. The optimized cost ratio is the point on the curve that, on average, produces the best performance in all folds. For more investigation, the results of the optimized system using a cost-based approach are compared with a hybrid method of oversampling and undersampling applied in [12], i.e., a combination of oversampling of class "A" instances using SMOTE and random undersampling of class "N" instances. Class "N" instances are randomly undersampled by a factor of two. Then, Class "A" instances are oversampled to have the same size as the (undersampled) instances of the class "N." In addition, not only random but also two informed undersampling approaches using KNN classifier are considered [45]. They are:

1) *NearMiss-1 (NM-1):* Selecting those class "N" instances whose average distance to the three closest class "A" instances is the smallest.
2) *NearMiss-2 (NM-2):* Selecting the class "N" instances whose average distance to the three farthest class "A" instances is the smallest.

After resampling of both classes, the ratio of their size becomes (1:1).

The bar chart in Fig. 6 shows that for all patients, the optimized system has the best performance using the cost-based approach. In comparison with other approaches for patients with low AED, the superiority of this approach is more pronounced. Alongside the cost-based approach, the comparison of undersampling methods combined with SMOTE shows that superiority of method over others varies for each patient. However, by increasing AED on average, the performance of these methods improves.

The better performance of the cost-based approach in comparison with other implemented resampling methods is not the only advantage of this method. Systems with a cost-optimized SVM classifier are resistant to be over-fit. The metric C controls the tradeoff between achieving a low error on the training data and minimizing the norm of the features' weights. It can, in fact, be seen as a parameter controlling the number of features to be selected by the SVM, assigning more

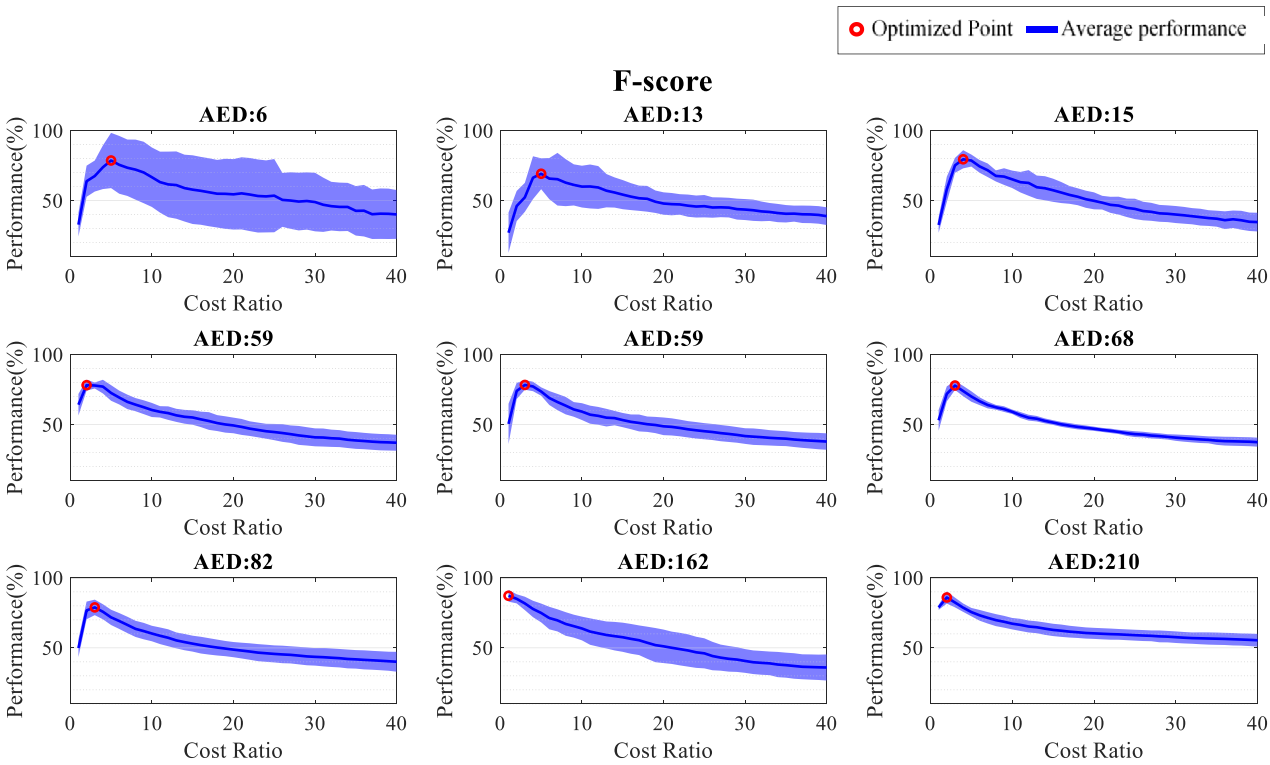

Fig. 5. Cross-validation results as a function of cost ratio.

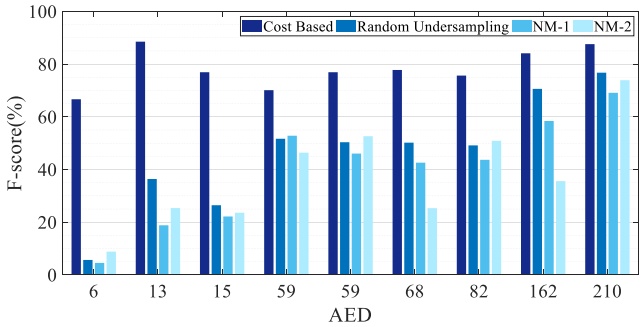

Fig. 6. System performances on unseen data for each patient in terms of F-score using different approaches for addressing imbalanced data: "Cost-Based" and "Hybrid of SMOTE and Undersampling" methods.

weights to more important features. To illustrate how one feature (variable) is affected by another, Fig. 7 contains scatter plots showing the data points of two classes, i.e., CSA class and normal breathing, for two of the nine patients. The selected features are R_1: Variance, R_2: f3 from Table II and R_3: f4 from Table II; which for the majority of the patients are the top three features with the largest absolute weights assigned by the SVM after optimization.

B. System Evaluation

Presently, AHI is known as the major and primary diagnostic and classification parameter for SA. However, it does not contain information on the morphology and duration of the breathing cessations and desaturations. Clearly, within the same severity of SA, shorter apnea-hypopnea duration and shallower desaturation may have different consequences than longer and deeper ones. The findings of [46] have supported that while the AHI reflected the frequency of respiratory events, parameters related to event duration such as mean total apnea duration (MTAD), mean central apnea duration (MCAD), mean obstructive apnea duration (MOAD), and mean mixed apnea duration (MMAD) can represent the severity of respiratory events in varying degrees.

Moreover, according to AASM, a device for home monitoring must allow sleep technician to look at the raw data and extracted events, and if it is necessary to edit the automated scoring [47]. As previously mentioned in Section II, in most of the reviewed studies, there is no discussion of these issues, and the evaluation of the previously proposed approaches has been done in terms of AHI or segment-by-segment basis through blind segmentation, where one apneic event can be covered by more than one segment or one segment may contain more than one event.

Event detector as a postprocessing step allows to detect and visualize the apneic events and compare them to a ground truth event-by-event. The system can provide the opportunity for a sleep technician to view events and perhaps edit the scoring if needed. Table III illustrates the performance of the system on unseen data. The results are arranged based on AED in ascending order. The event-by-event evaluation approach is extremely stringent in dealing with FP and FN events; especially if the patient is not severely ill, i.e., low AED. According to Table III, among all patients, the lowest F-score achieved

TABLE III
PERFORMANCE OF OPTIMIZED COST-SENSITIVE MODEL ON UNSEEN DATA

Patient No	AED	System Performance (%)		
		F-score	Sensitivity	Precision
1	6 (±6)	66.7	100	50
2	13 (±8)	88.5	100	79.4
3	15 (±6)	76.9	83.3	71.4
4	59 (±33)	70.1	73.1	67.4
5	59 (±24)	76.9	75.9	77.9
6	68 (±15)	77.8	81.7	74.2
7	82 (±19)	75.7	84.5	68.5
8	162 (±56)	84.1	80.4	88.2
9	210 (±40)	87.6	96	80.6

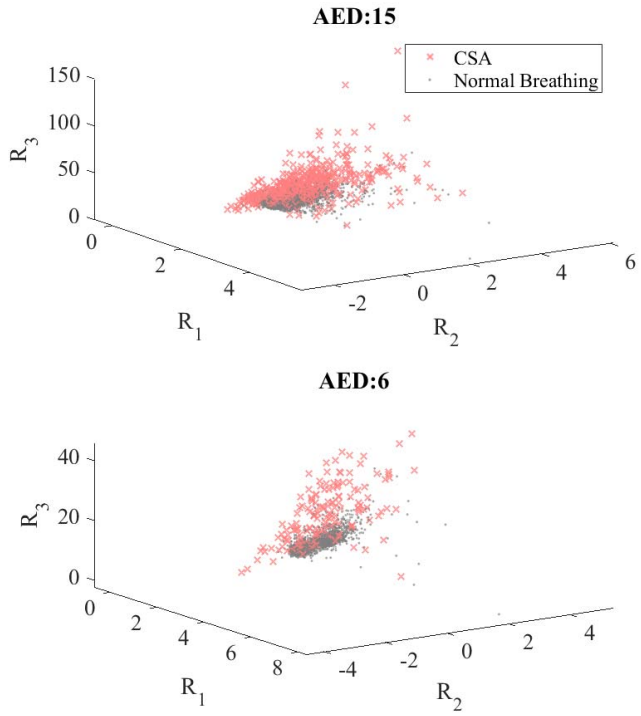

Fig. 7. Feature visualization for two of patients adopting 3 of 34 features: R_1: Variance, R_2: f3 from and R_3: f4 from Table II.

is 66.7% for the patient with AED of 6 (±6). Regarding this AED, in the worst scenario, i.e., a day with 12 events, a sensitivity of 100% means that all 12 events have been detected, and a precision of 50% means that the system only misclassified 12 events (FP). The same explanation applies to other patients. It should be noted that the severity of SA is based on AHI and therefore the number of events wrongly detected (FP) or ignored (FN) would not cause any significant misdiagnosis.

In current PSG scoring conventions, EEG recording is required to confirm that the patient is asleep so that respiratory events may be scored. However, it has been found that only 1% of a very large number of respiratory events were scored when PSG classified an awake patient [6]. This suggests that an EEG recording may not be essential in the diagnosis of SA and explains why many other diagnostic devices that do not quantify sleep are still diagnostically accurate. Also, a proof of associated oxyhemoglobin desaturation or EEG arousal may be needed when scoring SA, especially hypopneas where there is a decrease in the amplitude of the nasal flow signal in PSG. The same can be applied to OSA detection and consequently MSA detection through PSM. In most cases, there is a paradox breathing (180° difference in phase) between the abdomen and thorax signals while OSA happens. The pressure applied to an area of the rigid plastic covering the array can cause the area next to it to bend upward and subsequently create difficulty to detect the paradox breathing. Evaluating the reliably of PSM for detecting OSA and MSA needs more investigation and perhaps could lead to using an auxiliary sensor such as associated oxyhemoglobin desaturation simultaneously with PSM. It is technically possible to add the information in question which obviously would increase the overall performance and efficiency of the system, especially in terms of precision. However, that would lead to the drawback of the PSM measurement system ceasing to be noninvasive.

C. Descriptive Statistics of the Sleep Measures

Fig. 8 shows descriptive statistics of the sleep measures for one of the patients (chosen randomly), through his entire data. Each bar in the plots represents information on one specific day. For each day, CAI in the plot (c) is obtained by dividing a total number of extracted CSA events by a total number of hours spent in bed as "intention to sleep." Part of the plots in red represents the seven days of a week selected randomly for manual labeling. The usefulness of the CAI is that after an initial diagnosis, it can be monitored in the long-term to track a patient's treatment progress. In addition to CAI, the duration of CSA events are extracted to provide additional information for the assessment of CSA severity in Fig. 9. The bar chart in Fig. 9 groups total extracted CSA events into different time intervals from the data of the same patient in Fig. 8. The height of each bar shows how many events fall into each range. According to the bar chart, for this specific patient, most events have a duration of 13 to 14 s.

Table IV provides a summary of sleep characteristics for each patient, extracted from longitudinal data by applying day-by-day the personalized system (optimized on a week of data only), and averaged on the total number of occupied days. BO is separately calculated for the day clock and the night clock, to monitor daytime napping and nighttime sleep, respectively. Daytime napping may be a sign of excessive daytime somnolence. Subsequently, some have hypothesized that excessive daytime somnolence may be

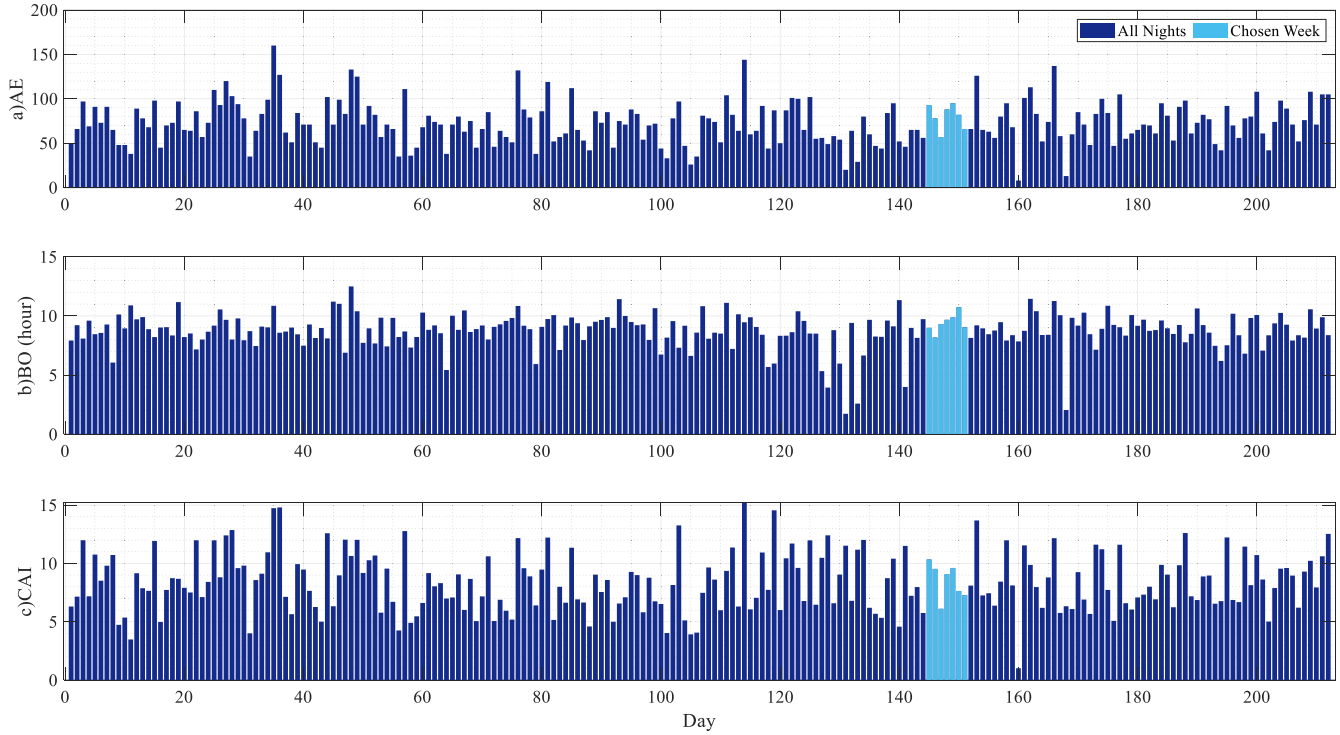

Fig. 8. Descriptive statistics of sleep measures for one of the patients. (a) Total number of CSA events extracted from each day. (b) Time spent in bed (in hours) during 24 h a day from noon to noon. (c) CAI per day. Bar charts in red represent the 7 days of a week selected randomly for manual labeling.

TABLE IV
SUMMARY OF SLEEP CHARACTERISTICS FROM THE LONGITUDINAL STUDY FOR EACH PATIENT AVERAGING THROUGH OCCUPIED DAYS

Patient No	**No of Occupied Days**	**BO (hour)**		**No of Detected CSA Events**	**Average CAI**
		Clock Day	Clock Night		
1	211	0.3 (±0.6)	2.5 (±1.6)	7 (±6)	2.6
2	150	0.1 (±0.3)	8.1 (±0.5)	15 (±7)	1.9
3	296	0.4 (±0.7)	7.5 (±1.5)	70 (±44)	8.9
4	200	0.8 (±1.2)	7.2 (±1.3)	57 (±22)	7.3
5	122	3.2 (±1.2)	7.3 (±1.5)	84 (±51)	8
6	212	0.9 (±1)	7.9 (±1.3)	73 (±24)	8.4
7	100	3.7 (±2.5)	8.2 (±2.3)	127 (±59)	10.7
8	393	0.3 (±0.7)	6.3 (±1.8)	80 (±63)	12.1
9	99	0.1 (±0.3)	8.2 (±1.1)	205 (±60)	24.6

due to the presence of an underlying and undetected sleep disorder such as SA, a condition that is strongly associated with increased cardiovascular risk [48]. Investigating the relationship between daily naps and SA severity over a long period can confirm the validity or incorrectness of this hypothesis. However, an investigation of this issue and interindividual sleep measures are beyond the scope of this article.

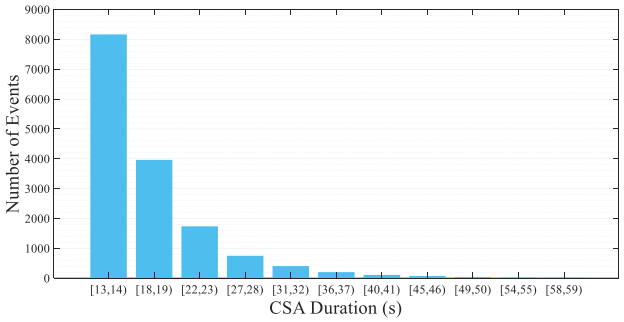

Fig. 9. Histogram of the CSA events' duration (in seconds).

V. Conclusion

This article explored the capability of PSM as an unobtrusive home monitoring sensor device. Adopting PSMs for measurements, a patient-specific system was designed and optimized for CSA detection and longitudinal monitoring of sleep characteristics in older adults. Data analysis showed that even for a healthy person with AHI < 5 and few numbers of CSA events, a week of data would be enough to design a personalized system. When there is a limited amount of labeled data available for each patient, optimization of the system especially in terms of data balancing should be done thoughtfully. Resampling the data may result in loss of information or changing its original distribution. To this end in this article, different methods were implemented and examined to deal with imbalanced data. As expected, the cost-based method conceded the best results without any data manipulation. Overall, systems designed for all patients achieved reasonable performances with sensitivity above 70%.

Although PSG is the most accurate clinical test to diagnose SA, unfamiliar surroundings in sleep laboratories and the knowledge of a clinical observer might affect the patients' sleep characteristics (i.e., "the first night effect" (FNE) [49]). Furthermore, a controlled environment like a sleep laboratory might not hold a home's sleep-disrupting factors, which can potentially result in an incorrect diagnosis. The extracted sleep measures such as BO and CAI are the results of PSM measurements obtained in a completely natural home environment, without the involvement of external factors such as discomfort caused by electrodes, limitation of movements by gauges and cables, potential psychological consequences of being under scrutiny, and a change of environment. However, despite all its benefits, unsupervised data collection faces some limitations. For instance, since no supplementary information was provided by the patient, it was challenging to draw a conclusion about abnormalities seen in the individual's data while reviewing PSM measurements (e.g., absence of the patient in bed for a long period of time). Therefore, for future studies, the patients should be asked to write sleep diaries.

This article proves that employing PSM as a home monitoring measurement device can be an effective approach for timely diagnosis of SA and presenting patients' sleep characteristics closer to real-life sleep in long-term monitoring. Given that access to information such as movements, respiration rate, and now CSA detection is possible from PSM measurements, future work will be to identify the sleep phases using PSM measurements.

Acknowledgment

The authors would like to thank Ms. Sareh Soleimani Gilakjani for sharing her code to perform formatting and preprocessing of data from patients.

References

[1] J. M. Kelly, R. E. Strecker, and M. T. Bianchi, "Recent developments in home sleep-monitoring devices," *ISRN Neurol.*, vol. 2012, pp. 1–10, Oct. 2012, doi: 10.5402/2012/768794.

[2] R. B. Berry *et al.*, "Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events: Deliberations of the sleep apnea definitions task force of the american academy of sleep medicine," *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597–619, Oct. 2012, doi: 10.5664/jcsm.2172.

[3] A. Zaffaroni, P. de Chazal, C. Heneghan, P. Boyle, P. R. Mppm, and W. T. McNicholas, "SleepMinder: An innovative contact-free device for the estimation of the apnoea-hypopnoea index," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 7091–9094, doi: 10.1109/IEMBS.2009.5332909.

[4] C. Yang, G. Cheung, V. Stankovic, K. Chan, and N. Ono, "Sleep apnea detection via depth video and audio feature learning," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 822–835, Apr. 2017, doi: 10.1109/TMM.2016.2626969.

[5] J. Abad *et al.*, "Automatic video analysis for obstructive sleep apnea diagnosis," *Sleep*, vol. 39, no. 8, pp. 1507–1515, Aug. 2016, doi: 10.5665/sleep.6008.

[6] M. B. Norman, S. Middleton, O. Erskine, P. G. Middleton, J. R. Wheatley, and C. E. Sullivan, "Validation of the sonomat: A contactless monitoring system used for the diagnosis of sleep disordered breathing," *Sleep*, vol. 37, no. 9, pp. 1477–1487, Sep. 2014, doi: 10.5665/sleep.3996.

[7] J. Hyuk Shin, Y. Joon Chee, D.-U. Jeong, and K. Suk Park, "Nonconstrained sleep monitoring system and algorithms using air-mattress with balancing tube method," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 147–156, Jan. 2010, doi: 10.1109/TITB.2009.2034011.

[8] D. C. Mack, M. Alwan, B. Turner, P. Suratt, and R. A. Felder, "A passive and portable system for monitoring heart rate and detecting sleep apnea and arousals: Preliminary validation," in *Proc. 1st Transdisciplinary Conf. Distrib. Diagnosis Home Healthcare*, Feb. 2006, pp. 51–54, Paper D2H2, doi: 10.1109/DDHH.2006.1624795.

[9] H. Azimi, S. Soleimani Gilakjani, M. Bouchard, S. Bennett, R. A. Goubran, and F. Knoefel, "Breathing signal combining for respiration rate estimation in smart beds," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, May 2017, pp. 303–307, doi: 10.1109/MeMeA.2017.7985893.

[10] S. Soleimani Gilakjani, S. Bennett, R. A. Goubran, H. Azimi, M. Bouchard, and F. Knoefel, "Movement detection with adaptive window length for unobtrusive bed-based pressure-sensor array," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, May 2017, pp. 355–360, doi: 10.1109/MeMeA.2017.7985902.

[11] D. Townsend, R. Goubran, F. Knoefel, and J. Leech, "Validation of unobtrusive pressure sensor array for central sleep apnea screening," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 7, pp. 1857–1865, Jul. 2012, doi: 10.1109/TIM.2012.2192342.

[12] H. Azimi, M. Bouchard, R. A. Goubran, and F. Knoefel, "Apnea event detection methodology using pressure sensors," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2019, pp. 1–6.

[13] O. Le Bon *et al.*, "Mild to moderate sleep respiratory events: One negative night may not be enough," *Chest*, vol. 118, no. 2, pp. 353–359, Aug. 2000.

[14] J. Newell, O. Mairesse, P. Verbanck, and D. Neu, "Is a one-night stay in the lab really enough to conclude? first-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples," *Psychiatry Res.*, vol. 200, nos. 2–3, pp. 795–801, Dec. 2012, doi: 10.1016/j.psychres.2012.07.045.

[15] D. J. Levendowski *et al.*, "Assessment of the test–retest reliability of laboratory polysomnography," *Sleep Breathing*, vol. 13, no. 2, pp. 163–167, May 2009, doi: 10.1007/s11325-008-0214-6.

[16] T. Young, J. Skatrud, and P. E. Peppard, "Risk factors for obstructive sleep apnea in adults," *J. Amer. Med. Assoc.*, vol. 291, no. 16, pp. 2013–2016, Apr. 2004, doi: 10.1001/jama.291.16.2013.

[17] S. S. Gilakjani, M. Bouchard, R. A. Goubran, and F. Knoefel, "Long-term sleep assessment by unobtrusive pressure sensor arrays," in *Proc. 3rd Int. Conf. Biomed. Imag., Signal Process. (ICBSP)*, Bari, Italy, 2018, pp. 23–35, doi: 10.1145/3288200.3288214.

[18] *Bed Occupancy Monitoring: Data Processing and Clinician User Interface Design—IEEE Conference Publication*. Accessed: Feb. 28, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/6347315

[19] M. Holtzman, D. Townsend, R. Goubran, and F. Knoefel, "Validation of pressure sensors for physiological monitoring in home environments," in *Proc. IEEE Int. Workshop Med. Meas. Appl.*, Apr. 2010, pp. 38–42, doi: 10.1109/MEMEA.2010.5480218.

[20] S. S. Gilakjani, H. Azimi, M. Bouchard, R. A. Goubran, and F. Knoefel, "Improved sensor selection method during movement for breathing rate estimation with unobtrusive pressure sensor arrays," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Mar. 2018, pp. 1–6, doi: 10.1109/SAS.2018.8336769.

[21] D. Dorran, R. Lawlor, and E. Coyle, "A comparison of time-domain time-scale modification algorithms," in *Proc. AES 120th Conv.*, Paris, France, 2006.

[22] C. A. Robles-Rubio, G. Bertolizio, K. A. Brown, and R. E. Kearney, "Scoring tools for the analysis of infant respiratory inductive plethysmography signals," *PLOS One*, vol. 10, no. 7, Jul. 2015, Art. no. e0134182, doi: 10.1371/journal.pone.0134182.

[23] M. H. Jones, R. Goubran, and F. Knoefel, "Identifying movement onset times for a bed-based pressure sensor array," in *Proc. IEEE Int. Workshop Med. Meas. Appl. (MeMea)*, Apr. 2006, pp. 111–114, doi: 10.1109/MEMEA.2006.1644474.

[24] K. Watanabe, T. Watanabe, H. Watanabe, H. Ando, T. Ishikawa, and K. Kobayashi, "Noninvasive measurement of heartbeat, respiration, snoring and body movements of a subject in bed via a pneumatic method," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 12, pp. 2100–2107, Dec. 2005, doi: 10.1109/TBME.2005.857637.

[25] M. Brink, C. H. Müller, and C. Schierz, "Contact-free measurement of heart rate, respiration rate, and body movements during sleep," *Behav. Res. Methods*, vol. 38, no. 3, pp. 511–521, Aug. 2006.

[26] D. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel, "Relative thresholding with under-mattress pressure sensors to detect central apnea," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3281–3289, Oct. 2011, doi: 10.1109/TIM.2011.2123250.

[27] S. Soleimani, "Long-term sleep assessment by unobtrusive pressure sensor arrays," M.S. thesis, School Elect. Eng. Comput. Sci., Univ. of Ottawa, Ottawa, ON, Canada, 2018.

[28] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, Sep. 1970, doi: 10.1016/0013-4694(70)90143-4.

[29] S. J. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 485–496, Mar. 2006, doi: 10.1109/TBME.2005.869773.

[30] S. Charbonnier, L. Zoubek, S. Lesecq, and F. Chapotot, "Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging," *Comput. Biol. Med.*, vol. 41, no. 6, pp. 380–389, Jun. 2011, doi: 10.1016/j.compbiomed.2011.04.001.

[31] J. Ben Ali, L. Saidi, S. Harrath, E. Bechhoefer, and M. Benbouzid, "Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning," *Appl. Acoust.*, vol. 132, pp. 167–181, Mar. 2018, doi: 10.1016/j.apacoust.2017.11.021.

[32] K. Šušmáková and A. Krakovská, "Discrimination ability of individual measures used in sleep stages classification," *Artif. Intell. Med.*, vol. 44, no. 3, pp. 261–277, Nov. 2008, doi: 10.1016/j.artmed.2008.07.005.

[33] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*. New York, NY, USA: Springer, 2003.

[34] H. Byun and S.-W. Lee, "A survey on pattern recognition applications of support vector machines," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 17, no. 03, pp. 459–486, May 2003, doi: 10.1142/S0218001403002460.

[35] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004, doi: 10.1109/TGRS.2004.831865.

[36] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, 2002, doi: 10.1023/A:1012487302797.

[37] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 37–48, Jan. 2009, doi: 10.1109/TITB.2008.2004495.

[38] H. Byun and S.-W. Lee, "Applications of support vector machines for pattern recognition: A survey," in *Pattern Recognition With Support Vector Machines*. Berlin, Germany: Springer-Verlag, 2002, pp. 213–236, doi: 10.1007/3-540-45665-1_17.

[39] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[40] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[41] P. Cao, D. Zhao, and O. Zaiane, "An optimized cost-sensitive SVM for imbalanced data learning," in *Advances in Knowledge Discovery and Data Mining*. Berlin, Germany: Springer-Verlag, 2013, pp. 280–292, doi: 10.1007/978-3-642-37456-2_24.

[42] D. Alvarez-Estevez and V. Moret-Bonillo, "Computer-assisted diagnosis of the sleep apnea-hypopnea syndrome: A review," *Sleep Disorders*, vol. 2015, Jul. 2015, Art. no. 237878, doi: 10.1155/2015/237878.

[43] B. L. Koley and D. Dey, "Real-time adaptive apnea and hypopnea event detection methodology for portable sleep apnea monitoring devices," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3354–3363, Dec. 2013, doi: 10.1109/TBME.2013.2282337.

[44] A. Mesaros, T. Heittola, and T. Virtanen, "Metrics for polyphonic sound event detection," *Appl. Sci.*, vol. 6, no. 6, p. 162, May 2016, doi: 10.3390/app6060162.

[45] *KNN Approach to Unbalanced Data Distributions: A Case Study Involving Information Extraction | BibSonomy*. Accessed: May 22, 2019. [Online]. Available: https://www.bibsonomy.org/bibtex/2cf4d2ac8bdac874b3d4841b4645a5a90/diana

[46] D. Y. Durmaz and A. Güneş, "Which is more important: The number or duration of respiratory events to determine the severity of obstructive sleep apnea?" *Aging Male*, pp. 1–6, Jun. 2019, doi: 10.1080/13685538.2019.1630062.

[47] N. A. Collop *et al.*, "Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients," *J. Clin. Sleep Med.*, vol. 3, no. 7, pp. 737–747, Dec. 2007.

[48] S. Ancoli-Israel and J. L. Martin, "Insomnia and daytime napping in older adults," *J. Clin. Sleep Med.*, vol. 2, no. 3, pp. 333–342, Jul. 2006.

[49] O. Le Bon *et al.*, "The first-night effect may last more than one night," *J. Psychiatric Res.*, vol. 35, no. 3, pp. 165–172, May 2001, doi: 10.1016/S0022-3956(01)00019-X.

Hilda Azimi (Student Member, IEEE) received the BA.Sc. and MA.Sc. degrees in biomedical engineering from Azad University, Science and Research Branch, Tehran, Iran, in 2011 and 2014, respectively. She is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Ottawa, Ottawa, ON, Canada.

The main statement of her project is unobtrusive elderly monitoring using sensor-based technologies to enhance the quality of life among older adults. She is currently working on longitudinal sleep disorders analysis especially sleep apnea. Her research interests include applied signal processing and machine learning, with an emphasis on physiological signal processing.

Martin Bouchard (Senior Member, IEEE) received the B.Ing., M.Sc.A., and Ph.D. degrees in electrical engineering from the Université de Sherbrooke, Sherbrooke, QC, Canada, in 1993, 1995, and 1997, respectively.

In January 1998, he joined the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada. In 1996, he co-founded SoftdB, Inc., Québec, QC, which is still active today. Over the years, he has conducted research activities and consulting activities with over 20 private sector and governmental partners, supervised over 50 graduate students and postdoctoral fellows, and authored or coauthored over 40 journal articles and 90 conference articles. His current research interests include applied signal processing and machine learning, with an emphasis on speech, audio, acoustics, hearing aids and biomedical engineering applications.

Dr. Bouchard served as a member of the Speech and Language Technical Committee of the IEEE Signal Processing Society from 2009 to 2011 and is a member of Ordre des Ingénieurs du Québec. He served as an Associate Editor of the *EURASIP Journal on Audio, Speech and Music Processing* from 2006 to 2011 and the IEEE TRANSACTIONS ON NEURAL NETWORKSfrom 2008 to 2009.

Rafik Goubran (Fellow, IEEE) received the B.Sc. degree in electronics and communications engineering and the M.Sc. degree in electrical engineering from Cairo University, Cairo, Egypt, in 1978 and 1981, respectively, and the Ph.D. degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 1987.

In 1987, he joined the Department of Systems and Computer Engineering, Carleton University. From 1997 to 2006, he was the Chair of the Department of Systems and Computer Engineering, Carleton University, where he was the Dean of the Faculty of Engineering and Design, from 2007 to 2016. He was involved in several research projects with industry and government organizations in the areas of digital signal processing, biomedical engineering, sensors, data analytics, microphone arrays, and the design of smart homes for the independent living of seniors. From 2006 to 2009, he was the Founding Director of the Ottawa-Carleton Institute for Biomedical Engineering, Ottawa. He is currently a Scientist with the Bruyère Research Institute, Ottawa. He is also a Professor and the Vice-President (Research and International), Carleton University.

Dr. Goubran is a fellow of the Canadian Academy of Engineering and a member of the Association of Professional Engineers of Ontario.

Frank Knoefel received the B.Sc. degree from McGill University, Montreal, QC, Canada, in 1985, the M.D. degree from the Université de Montréal, Montreal, in 1989, his certification with the College of Family Physicians through Université de Sherbrooke, Sherbrooke, QC, in 1991, the diploma in care of the elderly from the University of Ottawa, Ottawa, ON, Canada, in 1992, and the M.P.A. degree from Queen's University, Kingston, ON, in 1998.

He is currently a Physician with the Bruyère Memory Program, Elisabeth-Bruyère Hospital, Ottawa, and a Senior Investigator with the Bruyère Research Institute, Ottawa. He is also an Associate Professor with the Department of Family Medicine, University of Ottawa, and an Adjunct Research Professor with the Department of Systems and Computer Engineering, Carleton University, Ottawa. He is also the Co-Founder and the Co-Leader of the TAFETA: Smart Systems of Health. His current research interests include the use of technology in caring for older adults, and the remote monitoring of mobility and breathing, with current work, including the use of technology to monitor and minimize cognitive decline.