

Received August 20, 2020, accepted September 18, 2020, date of publication September 22, 2020, date of current version October 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3025808

Machine Learning-Based Automatic Detection of Central Sleep Apnea Events From a Pressure Sensitive Mat

HILDA AZIMI^{1,2,4}, (Member, IEEE), PENGCHENG XI^{2,6},
MARTIN BOUCHARD^{1,3}, (Senior Member, IEEE),
RAFIK GOUBRAN^{1,2,3,4}, (Life Member, IEEE),
AND FRANK KNOEFEL^{1,2,3,4,5}

¹School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada

²Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

³Bruyère Research Institute, Ottawa, ON K1R 6M1, Canada

⁴AGE-WELL National Innovation Hub on Sensors and Analytics for Monitoring Mobility and Memory (AGE-WELL NIH-SAM3), Ottawa, ON K1N 5C8, Canada

⁵Faculty of Medicine, University of Ottawa, Ottawa, ON K1N 6N5, Canada

⁶Digital Technologies, National Research Council Canada, Ottawa, ON K1A 0R6, Canada

Corresponding author: Hilda Azimi (hazim009@uottawa.ca)

This work was supported in part by the Centre for Aging + Brain Health Innovation (CABHI), in part by AGE-WELL NCE Inc. (Member of the Networks of Centres of Excellence Program, a Government of Canada Program that funds Partnerships Between Universities, Industry, Government and Not-for-Profit Organizations), and in part by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant.

ABSTRACT Polysomnography (PSG) is the standard test for diagnosing sleep apnea. However, the approach is obtrusive, time-consuming, and with limited access for patients in need of sleep apnea diagnosis. In recent years, there have been many attempts to search for an alternative device or approach that avoids the limitations of PSG. Pressure-sensitive mats (PSM) have proven to be able to detect central sleep apneas (CSA) and be a potential alternative for PSG. In the current study, we combine advanced machine learning approaches with a practical unobtrusive home monitoring device (PSM) to detect CSA events from data collected nocturnally and unattended. Two deep learning methods are implemented for the automatic detection of CSA events: a temporal convolutional network (TCN) and a bidirectional long short-term memory (BiLSTM) network. The deep learning models are compared to a classical machine learning approach (linear support vector machine, SVM) and a simple threshold-based algorithm. Considering the characteristics of each method, we choose strategies, including resampling and weighted cost-functions, to optimize the methods and to perform CSA detection as anomaly detection in an imbalanced data set. We evaluate the performance of all models on a database containing 7 days of data from 9 elderly patients. From the resulting 63 days, data from 7 patients (49 days) are devoted to training for optimizing hyperparameters, and data from 2 patients (14 days) are devoted to testing. Experimental results indicate that the best-performing model achieves an accuracy of 95.1% through training an BiLSTM network. Overall, the implemented deep learning methods achieve better performance than the conventional classification approach (SVM) and the simple threshold-based method, and show good potential for the use of PSM for practical unobtrusive monitoring of CSA.

INDEX TERMS Biomedical measurement, data analysis, deep learning, machine learning, patient monitoring, pressure measurement, central sleep apnea detection.

I. INTRODUCTION

Sleep apnea (SA) is a well-known sleep disorder. The three main types of SA events are central sleep apnea (CSA), obstructive sleep apnea (OSA), and mixed sleep apnea (MSA)

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano ¹.

which is a combination of the previous two (i.e., initiated by a CSA followed by an OSA event). The detection of SA events requires analyzing the physiological data collected during patients' sleep.

The conventional data collection approach for the diagnosis of SA is polysomnography (PSG), which is time-consuming and costly. Several techniques with fewer

sensors have been proposed to replace PSG. Unlike PSG, these approaches are mostly based on measuring physiological signals, such as airflow [1], [2], thoracic signal [3], abdominal signal [4], [5] or oxygen saturation [6]–[8]. However, they either induce discomfort caused by electrodes, enforce limited movements due to gauges and cables, or may have their results affected by the potential psychological consequences due to being “monitored”, possibly in an institutional environment or in home monitoring [9].

To address the limitations of the above techniques and to expedite and enhance SA diagnosis, environmental sensors have been taken into consideration in recent years. In contrast to commonly used sensors, they are not attached directly to the patients’ bodies, but are installed in their sleep environment. These devices are unobtrusive and suitable for monitoring patients longitudinally without intervention [10]. Examples of these alternative sensors include digital video cameras for measuring the volume of air circulating into the lungs [11], non-contact radio-frequency sensors for measuring the bio-motion caused by body movement and breathing [12], and pressure-sensitive mats (PSM) for measuring respiratory movements [13]. Among these sensors, the PSM sensing has advantages since it is capable of capturing body movements and breathing signals regardless of body position [14], and it does not compromise the privacy of patients. The device has proved to be a promising source of data collection for health monitoring [15], helping to manage ongoing illness and facilitating preventive care [16]. Therefore, optimizing approaches to extract information from PSM data is of great importance.

Approaches for processing collected data in the mentioned studies are very similar. Raw data need to be transformed before being fed to an SA event detector. In most cases, the system is composed of three successive steps: preprocessing, expert-driven feature extraction, and classification. However, the design or choice of features to be extracted in conventional classification requires expert knowledge. The process is time-consuming and domain-specific [17]. Manually designed feature extraction algorithms may fail to extract the most relevant information from the data. Furthermore, the amount of information fed to the traditional machine learning algorithms must be limited, since these algorithms would perform worse when dealing with high dimensional inputs. Therefore, the decision quality may suffer due to restricted information [18]. Consequently, automated design or selection of task-specific features is of value for solving complex real-world problems, such as the SA detection.

Deep learning (DL) approaches have revolutionized data modeling from “expert-driven” feature engineering to “data-driven” feature construction [19]. They have been favored in many data analytics applications, such as computer vision, natural language processing, speech recognition, and healthcare [18], [20]. The DL approaches that have been used previously for SA detection include long short-term memory (LSTM) [21], [22], convolutional neural network (CNN) model [22], [23], and pre-trained CNN models [24].

Despite their promising results, all these studies have used data collected via devices and sensors with the aforementioned limitations (obtrusive, uncomfortable, etc.).

To solve these issues, in this work we use unobtrusive PSMs to collect data and design optimized approaches to process the collected data. A PSM placed under a mattress therefore eliminates the disadvantages mentioned earlier. To choose the right approach for a classification task, one needs to take into consideration several factors, such as the properties and amount of data required and model selection. Therefore in this work, we perform a comprehensive comparison between support vector machine (SVM) (as a representative of traditional machine learning model [25]) and DL methods. In particular, we select the following DL models for designing an automatic CSA detector from PSM signals: temporal convolutional network (TCN) [26] and Bidirectional LSTM (BiLSTM) network. The former is a representative of CNN models and the latter is a representative of recurrent neural network (RNN) models. Both methods are used because of the temporal characteristics of the PSM signals. Due to its RNN-based structure, the BiLSTM model is used to detect the CSA events occurring successively and repetitively during sleep. Just like the BiLSTM, the TCN can be used to model sequence-to-sequence or sequence-to-one tasks. Furthermore, we adopt a basic threshold-based method [27] for comparison purposes. Overall, the objective is to find the best-performing method for the automatic detection of CSA from PSM signals.

The DL models are capable of learning representations of the key features and interactions from the data itself, through direct feature learning in a supervised manner [28]. We hypothesize that applying DL approaches may allow the unlocking of information in PSM signals that is key to the detection of CSAs. In comparison, an SVM classifier is used to deal with CSA detection from PSM data by choosing a limited number of variables in [25], which can be the reason for a noticeable amount of false-positive predictions. In the basic threshold-based method [27], this shortcoming becomes more problematic since the extracted signal information is summarized into only one variable.

Our contributions are three-fold: 1) the PSM data collection approach is convenient and unobtrusive to patients; 2) the two selected deep learning models are able to capture key features in the PSM data for conducting temporal event-by-event evaluations, and therefore indicating the severity of CSA patients; and 3) resampling approaches and weighted loss methods are implemented for addressing the imbalance challenge in the dataset.

In the following, section II introduces the nature of the PSM signal, the collected data, and the CSA detectors designed in the paper. The results of different methods are described and compared through training and testing in section III; strengths and weaknesses are discussed for all the constructed models in section IV; and section V concludes the paper.

II. MATERIAL AND METHOD

A. PSM

The PSMs used were manufactured by Tactex Control Inc. Each set of sensor array consists of 72 fiber optic pressure sensors. The sensors are evenly spaced 10 cm apart. As shown in Fig.1, a PSM covers the area from about a patient’s head to his/her hip, placed under a mattress.

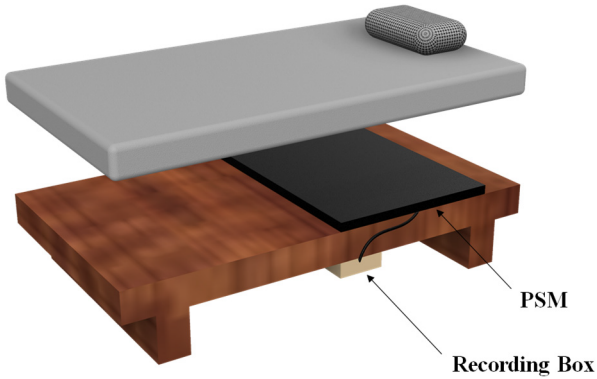


FIGURE 1. System setup.

B. SUBJECTS AND DATASETS

Clinically, the ability of PSMs to detect CSA events has been shown in [29], where the PSM signals were compared to three other scenarios and signal combinations including respiratory inductance plethysmography (RIP) bands combined with oxygen-saturation sensor or airflow, RIP bands alone, and finally expert PSG interpreters. The next step was to evaluate the performance of the device in real-life conditions. Sensor set-ups including PSMs and recording boxes were installed in elders’ homes to continuously collect data without the subjects wearing respiratory bands as a gold standard sensor. The monitoring period lasted 8 to 12 months. With the exception of one visit per month, the data was collected without supervision.

Nine volunteers from different communities participated in the study. They were community-dwelling older adults with age sixty-five years or more (female/male), who were living in affordable seniors housing, or discharged from the Geriatric Rehabilitation Unit at Élisabeth Bruyère Hospital, Ottawa, Canada [16].

Files were concatenated daily from noon to noon (24 hours) for sleep assessment. A week of data collected from subjects (7-day period, 9 subjects, 63 days of data) was randomly selected and used in this study. Start and end points of apneic events were manually marked by a trained person, following the American Academy of Sleep Medicine (AASM) rules, i.e., complete cessation of breathing movements captured by pressure sensors for a minimum length of 10 s.

C. DATA PRE-PROCESSING

The raw data collected by PSMs in non-laboratory environments are voluminous and noisy, and therefore require

preprocessing. We propose a pre-processing pipeline composed of the following steps:

1) OCCUPANCY EXTRACTION

The algorithm in [30] is adopted here to identify and remove those periods of time when the bed is not occupied. This discards irrelevant data and therefore decreases the computational complexity.

2) BANDPASS FILTERING

After removing the unoccupied parts of the signals, a finite impulse response (FIR) bandpass filter is applied to each of the 72 PSM signals, with a 0.07-0.8 Hz passband. The normal breathing range is 12 to 20 breath per minute (bpm). However, in order to prevent information loss and to be as general as possible, extreme conditions from 4 to 48 bpm are also considered in the selected passband frequency range.

3) SIGNAL COMBINATION & CONCATENATION

For every 30 s signal segment with 50% overlap, all 72 signals of the PSM are weighted by a SNR-maximizing sensor signal combination method, with unequal weights based on the quality of their information [13]. Therefore, sensors with stronger signals (and better SNR) contribute more heavily than others. Next, in each segment, the weighted signals are combined to produce a single signal that has higher signal quality compared to each of the 72 signals.

Defining $Z_i[n] = [z_{i,1}[n] \ z_{i,2}[n] \ \dots \ z_{i,N}[n]]^T$ as the current samples from each of the N signals in the i^{th} segment of data and $W_i = [w_{i,1} \ w_{i,2} \ \dots \ w_{i,N}]^T$ as the weights applied to the samples from each signal during the i^{th} segment, the output samples of the signal combination are:

$$y_i[n] = W_i^T Z_i[n] \tag{1}$$

where the weights W_i for each segment are computed from the cross-covariance between the reference signal (i.e., r^{th} signal with the maximum power in the breathing frequency band) and the other signals:

$$W_i = \begin{bmatrix} \sum_{n=1}^M (z_{i,r}[n] - \bar{z}_{i,r})(z_{i,1}[n] - \bar{z}_{i,1}) \\ \sum_{n=1}^M (z_{i,r}[n] - \bar{z}_{i,r})(z_{i,2}[n] - \bar{z}_{i,2}) \\ \vdots \\ \sum_{n=1}^M (z_{i,r}[n] - \bar{z}_{i,r})(z_{i,N}[n] - \bar{z}_{i,N}) \end{bmatrix} \tag{2}$$

where $\bar{z}_{i,j}$ is the sample mean from samples of the j^{th} signal during the i^{th} segment, and M is the number of samples within a segment (i.e., 300 samples for a PSM with a sampling rate of 10 Hz).

Different window sizes are needed in different parts of a CSA detection system, especially since the minimum duration of CSA events is 10 s. Thus, the output signals of the combining process for every 30 s segments are concatenated to produce a single signal.

4) SIGNAL NORMALIZATION

PSM signal amplitude can be affected by body movements and different body postures [14]. The signals must be normalized by the means of the body movement to estimate the respiration more accurately in terms of depth and volume. Body movements cause large fluctuations in the signal. These fluctuations are detected as a value that is more than three median absolute deviations (MAD) away from the median of the combined signal [25]. Next, each part of the signal between any two movements is normalized in two consecutive steps: first removing the average of the signal, next dividing the signal by the maximum absolute value of the signal, excluding the 5% largest negative and positive values.

The resulting output of this pre-processing pipeline, $q[n]$, is a single normalized signal per day. It is then fed to all the different systems designed in sections II.D, II.E and II.F. Depending on the method used, $q[n]$ is segmented into smaller sequences of 9 s with a 50% overlap, or it is injected into a detection system as a single observation for each day. Regardless of different data segmentation, the data of two patients (equivalent to 14 days of data) are randomly selected as test data. The test data are the same for all methods, to allow a meaningful evaluation and comparison.

D. POWER DETECTOR METHOD

The method presented in [27] is based on the AASM rules for detecting SA events by sleep technicians. [31]. As stated in the article, this method consists of two consecutive steps, which are:

1) POWER CALCULATOR

The breathing baseline in patients with unstable breathing patterns is specified based on the 3 largest breaths in the 2 minutes preceding the onset of an event. [31]. The breathing baseline determines a threshold for detecting an apneic event. For this purpose, the previous 2 minutes of data are divided into 9 s sub-segments with a 50% overlap, and for each sub-segment, the power of the signal is calculated.

Let $x_i[n]$, $n = 1, 2, \dots, M_p$ be the data samples of the i^{th} sub-segment taken from $q[n]$, the normalized signal previously described. Then the power of each sub-segment is calculated as:

$$P_i = \frac{1}{M_p} \sum_{n=1}^{M_p} |x_i[n]|^2 \quad (3)$$

where M_p is the number of samples within a sub-segment (i.e., 90 samples for a PSM with a sampling rate of 10 Hz).

2) DETECTOR

Considering the 2 minutes preceding the onset of an event, the threshold of the detector is calculated as a fraction (FC) of the sub-segment power with an 80-percentile position from the calculated sub-segment powers sorted in ascending order. Therefore, the threshold is less prone to cases where there are several events or movements (with extreme signal power) in

the 2 minutes. The value of the 80-percentile segment power is too large to be considered as the threshold by itself, and thus the FC helps to tune the threshold. Finally, the 9 s segment after 2 minutes of data is classified as an apneic event if its power is smaller or equal to the threshold.

E. CLASSIC SVM

Linear SVM has been applied to classify 9 s segments of PSM signals into two groups of apneic “A” and not-apneic “NA” [25]. Segments were considered to be class “A” instances if they had at least a 50% overlap with CSA events. In addition, segments that had overlap with detected movements from the preprocessing step were treated as outliers and they were removed from the training dataset.

As illustrated in Table 1, thirty-four (34) features from the time and frequency domains were extracted from the remaining segments and fed to the SVM [25]. For time-domain based features, x is a segment of the signal, whereas for frequency-domain based features, X represents the power spectrum of each segment. The average removed and the maximum absolute value of the signal between any two body movements, both used to produce the normalized signal $q[n]$, were also included as two features for each segment. This is implemented to prevent loss of information, especially when the part of the signal between two consecutive body movements is too short and mostly noisy.

When implementing the SVM model used in [25], we need to address the imbalanced data problem. The number of observations in the two classes “NA” and “A” is imbalanced. In comparison to class “A”, class “NA” is significantly over-represented in the dataset. A simple way to balance the datasets and consequently prevent the “accuracy paradox” (i.e., when the accuracy does not imply the actual performance of the classifier and only reflects the underlying class distribution) is to perform resampling of the classes, i.e., over-sampling the minority class or under-sampling the majority class.

The resampling approaches have drawbacks. Oversampling the minority class with a high factor introduces duplicated instances from a small pool of observations. Therefore, it can lead to model overfitting. On the other hand, under-sampling the majority class can result in eliminating important instances that provide important differences between the two classes. In this work, we follow an approach previously introduced in [25], where a combination of both strategies is used to reduce the negative impact of each one. Class “NA” was under-sampled randomly by a factor of two. Then, the synthetic minority oversampling technique (SMOTE, a well-known method for oversampling [32]) was used to oversample class “A” by a factor of 16. The approach balances the dataset without overfitting as much as with basic oversampling, since it creates new instances by forming convex combinations of neighboring instances rather than duplicate already-existing instances.

In this study, by adopting the same idea, we apply the combination of resampling approaches to balance classes and

TABLE 1. List of extracted features.

	Formula/Description
Time-domain based features	- Variance, Skewness, Kurtosis, Standard Deviation (SD), Root Mean Square (RMS), Mean, Maximum (Max), Median, Range, Minimum (Min), Hjorth Parameters (Complexity & Mobility)
	- Sub-Segment-Ratio [25]
	$\frac{Max(x)}{RMS(x)}$
	$\frac{RMS(x)}{Mean(Abs(x))}$
	$\frac{Max(x)}{Mean(Abs(x))}$
	$\frac{Max(x)}{Mean(Abs(x))^2}$
	Note: for time-domain based features, x is a segment of the signal.
Frequency-domain based features	- Spectral Variance, Spectral Skewness, Spectral Kurtosis, Spectral SD, Spectral Mean, Spectral Median, all corresponding to the selected frequency bands*.
	- Peak Power Frequency corresponding to the selected frequency bands (PPF_{FB}).
	- Peak Power Frequency (PPF)
	$\frac{PPF_{FB}}{PPF}$
	- Spectral Flatness Measure corresponding to the selected frequency bands.
	$\frac{Max(X)}{RMS(X)}$
	$\frac{RMS(X)}{Mean(X)}$
$\frac{Max(X)}{Mean(X)}$	
$\frac{Max(X)}{Mean(X)^2}$	
- The percentage of the total area under the curve corresponding to the selected frequency bands.	
Note: for frequency-domain based features, X represents the power spectrum of each segment.	
Normalization Procedure	- Average removed. - Normalization by maximum absolute value of the signal (excluding the 5% largest negative and positive values).

* Selected frequency bands correspond to normal breathing range (i.e. 12-20 bpm).

feed them to a linear SVM classifier. 5-fold cross-validation is performed on the training dataset to optimize the factors for under-sampling the majority and oversampling minority classes, respectively.

F. DEEP LEARNING

1) BILSTM

An LSTM is a type of RNN that can learn long-term dependencies between time steps of sequential data. Contrary to CNNs, an LSTM can remember the state of the network between predictions. The essential components of an LSTM network are a sequence input layer and an LSTM layer. A sequence input layer incorporates time-series data into the network. An LSTM layer learns long-term dependencies between time steps of sequence data. The layer contains hidden units providing inputs to memory cells and their corresponding gate units. All units (except for gate units) have connections to all units in the next layer [33].

As an extension of the traditional LSTM network, the bidirectional LSTM (BiLSTM) network can improve the performance of sequence classification problems. While the LSTM layer considers the time sequence in the forward direction, the BiLSTM layer considers the time sequence in both backward and forward directions [34]. In problems where the complete time steps of the input sequence are available, the BiLSTM network trains two LSTM networks on the input sequence. The process involves replicating the first recurrent layer in the network so that there are two layers side-by-side. While the input sequence is an input to the first layer, its reversed replica acts as an input to the second layer. This approach delivers additional context to the network and results in quicker and better learning of a model.

2) TCN

Sequence modeling for most deep learning practitioners is synonymous with recurrent networks. Yet recent results have shown that a simple convolutional architecture known as a temporal convolutional network (TCN) can outperform recurrent networks such as LSTMs across a wide range of datasets and tasks while demonstrating longer effective memory [26].

A general TCN architecture consists of multiple residual blocks. As shown in Fig. 2, each block comprises of two sets of dilated causal convolution layers with the same dilation factor, followed by normalization, rectified linear unit (ReLU) activation, and spatial dropout layers. The input to each block is added to the output of the block (including a 1-by-1 convolution on the input when the number of channels between the input and output does not match) and a final activation function is applied. TCN combines dilations and residual connections with the causal convolutions needed for autoregressive prediction. Weight normalization is applied to the convolutional filters, and a spatial dropout is added after each dilated convolution for regularization [26].

TCN is based on two principles:

- a. The convolutions in the architecture are causal, where an output at time t is convolved only with elements from time t and earlier in the previous layer. Therefore, there is no information “leakage” from the future to the past.
- b. The architecture produces an output of the same length as the input, just as with an RNN. It adopts a 1D fully convolutional network architecture, where

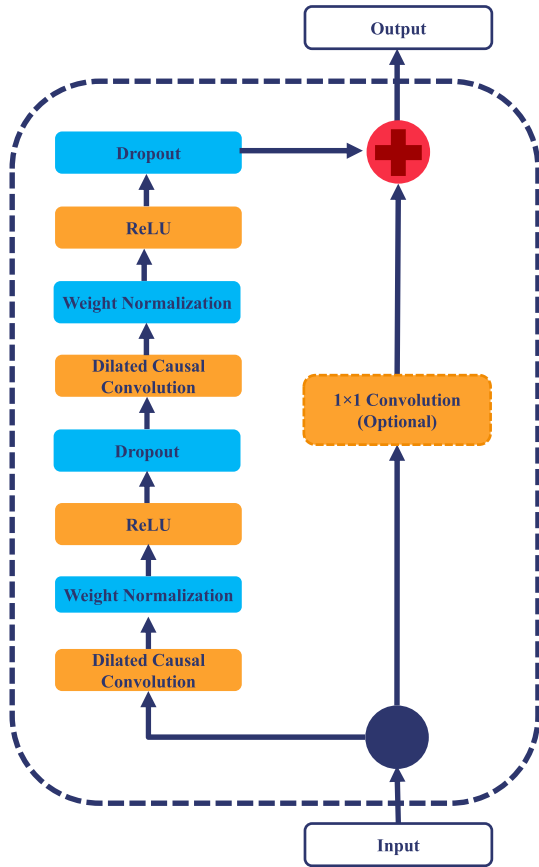


FIGURE 2. TCN residual block.

each hidden layer is the same length as the input layer.

Dealing with class imbalance is also necessary for DL approaches. Indeed, in the case of imbalanced classes, networks would learn that they can achieve a high accuracy simply by classifying all observations as a member of the majority class. To avoid this bias, for both DL approaches, i.e. BiLSTM and TCN, a weighted classification layer is adopted to compute the weighted cross-entropy loss.

Weighted cross-entropy is an error measure between two continuous random variables. For prediction scores S and training targets T , the weighted cross-entropy loss between S and T is given by:

$$L = -\frac{1}{N_{OB}} \sum_{n=1}^{N_{OB}} \sum_{i=1}^K w_i T_{ni} \log(S_{ni}) \quad (4)$$

where N_{OB} is the number of observations, K is the number of classes, and w is a vector of weights for each class. It should be noted that, as a vector of class weights in (4), w has different meaning and definition than (2) and it is inversely proportional to the number of training examples in each class, to give each class equal total weight in the loss [35]. Therefore, in comparison to majority class instances, each instance of the minority class contributes more to the final

loss and the majority class is prevented from over-exposure to the network representation.

G. PERFORMANCE EVALUATION

Regardless of the method used, the output of the classifier is a series of class labels for every time step (or separated by an interval equal to the 50% segment shift). The classification performances are assessed in terms of:

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \times 100 \quad (5)$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \times 100 \quad (6)$$

$$\text{Accuracy} = \frac{TP+TN}{(TP + TN + FP + FN)} \times 100 \quad (7)$$

where TP, TN, FP, and FN are abbreviations for true positives, true negatives, false positives, and false negatives respectively.

Moreover, a rule-based algorithm is adopted to have a representation of class labels similar to the annotations of the database [27] as follows:

- A sequence of instances, all classified as class “A”, are merged as a single event.
- A complete breathing cycle takes at least 3 s. Therefore, parts of the signal located between detected apneic events with a duration of less than 3 s are re-classified as part of apneic events.
- Detected events with a length of 60 s and more are re-classified as “NA” events since an apneic event usually lasts between 20 to 40 s [36].

An event-based evaluation is used to compare the detected events as the results of the rule-based algorithm with the reference events scored in data [27]. The performance of the rule-based algorithm is assessed by the F-score:

$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (8)$$

For event-based evaluation, TP, FP and FN are events defined as [37]:

- TP: An apneic event detected by the system that has a temporal position overlapping with an apneic event scored in the signal.
- FP: An apneic event detected by the system that has no overlap with any apneic event scored in the signal.
- FN: An apneic event scored in the signal that has no overlap with any apneic event detected by the system.

Fig. 3 summarizes the process of CSA detection by applying the proposed approaches. After preprocessing, for the first two methods (i.e., power detector method and SVM), $q[n]$ is a sequence of data fed to the feature extractor to drive the most informative features from the data to the classifiers. In this step, while 34 features are extracted for the SVM classifier, the number of features for the power detector method is reduced to one feature only. One of the advantages of SVM is that it performs feature reduction during the training process by assigning more weights to more important features, and

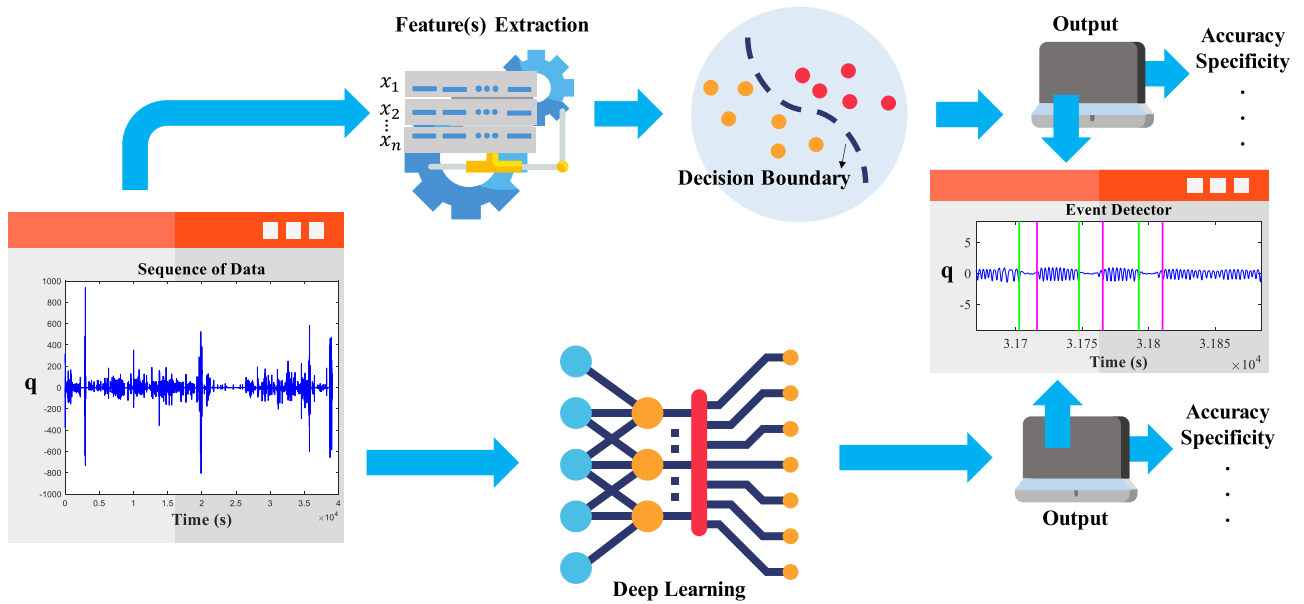


FIGURE 3. Flowchart of detection system. For the event detector, the green and magenta lines are the start and end of the CSA events, respectively.

therefore reduces the effect of possibly uninformative features on classification. Next, the features are used as inputs to the classifiers for training and optimizing the models. For DL approaches (i.e., BiLSTM and TCN) the processes of feature extraction and network optimization both occur during the training of the networks. All methods are evaluated on the same test data. Finally, a rule-based algorithm analyzes the results of classification as a series of class labels, to extract CSA events individually.

III. RESULTS

A total of four approaches are compared in this paper to detect CSA events. The model produced by each approach is tuned using a part of the training dataset as validation data. The results of optimization with each method are described next.

A. POWER DETECTOR METHOD

For the power detector method, 49 days of data are used to tune the FC parameter. FC is the only parameter that needs to be optimized in this method. As described earlier, FC is the fraction of the 80-percentile sub-segment power in the 2 minutes prior to the start of an event. According to Fig. 4, the F-score on the training data is gradually improved by increasing FC up to FC = 0.1, and further increasing FC reduces the F-score considerably. Therefore, FC = 0.1 is applied to the test data (14 days) to evaluate the method.

B. SVM

The ratio of under-sampled “NA” class instances to over-sampled “A” class instances is optimized by applying 5-fold cross-validation on the training dataset. Let α be the

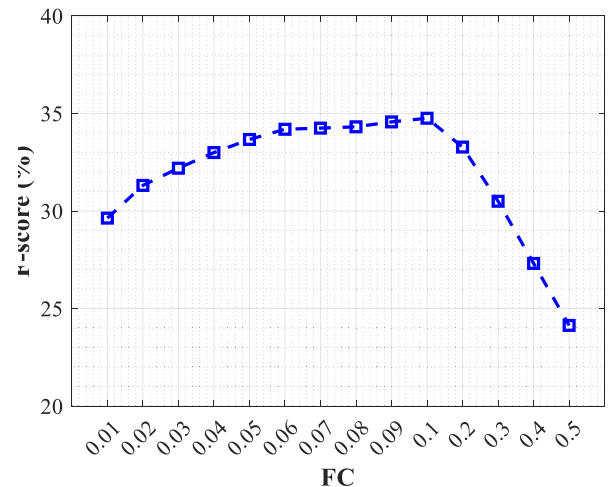


FIGURE 4. F-score of power detector method on training data as a function of FC.

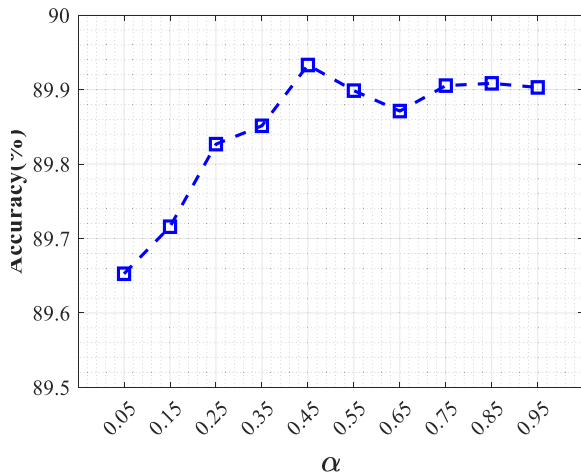
ratio of oversampled “A” class instances to under-sampled “NA” class instances, which is tested for different values from 0.05 to 0.95, with an increment of 0.1. Therefore, the range of α is determined by two extremes where:

- The smallest α represents an under-sampling of “N” class instances to reach the number of “A” class instances, without oversampling the “A” class instances.
- The largest α represents mostly an oversampling of “A” class instances to reach the number of “NA” class instances, without under-sampling the “NA” class instances.

All other values of α are between these two extremes. According to Fig. 5, the classifier achieved the highest F-score on validation data for $\alpha = 0.45$.

TABLE 2. Performance of the approaches considered.

	Model performance (%)			Event Detector Performance (%)		
	Accuracy	Specificity	Sensitivity	Sensitivity	Precision	F-score
Power Detector	---	---	---	86.0	39.2	53.9
SVM	89.1	91.9	67.2	97.7	55.6	70.8
BiLSTM	95.1	96.0	85.7	88.0	82.0	85.0
TCN	91.8	93.8	70.7	80.7	77.9	79.2

**FIGURE 5.** Accuracy of the SVM classifier on validation data as a function of the ratio of oversampled “A” class instances to under-sampled “N” class instances, with 0.1 increases.

C. DEEP LEARNING

For the BiLSTM and TCN models, each of the 63 days of training signals is presented to the network as an individual input sequence. 49 and 14 days of data are used for training and testing, respectively. The BiLSTM and TCN make predictions based on the individual time steps of the sequence data.

1) BiLSTM

The architecture of the network contains a BiLSTM layer with 64 hidden units, returning the hidden state output for each input time step, with a fully connected layer of size two, followed by a soft-max layer and a weighted classification layer. The number of epochs for training is set to 20 so that the network makes 20 passes through the training data. The training data is shuffled before each training epoch.

The batch size is set to 1 so that the network considers one day of data at a time. The learning rate is set to 10^{-3} . The model parameters are optimized by minimizing weighted cross-entropy loss functions, based on the update rule of the adaptive moment estimation (ADAM) solver [38]. The ADAM solver normally performs better with RNNs than the default stochastic gradient descent with momentum (SGDM) solver [39].

2) TCN

TCN architecture consists of four residual blocks. Each block contains dilated causal convolution layers, each with

175 filters of size 5. The weighted cross-entropy loss is calculated for every batch. The final network is trained via stochastic gradient descent by looping over the sequences in the training dataset computing parameter gradients and updating the network parameters via the ADAM update rule. The number of epochs is set to 10, with a batch size of 1. For each epoch, the training data is shuffled. The learning rate has an initial value of 10^{-5} , which is multiplied by 0.1 every four epochs.

For both DL approaches, i.e., BiLSTM and TCN, the number of epochs is determined by evaluating the performance of the networks on the training dataset and the validation dataset (i.e., data of 2 patients selected randomly). When using more than the final selected number of epochs in each method, the networks start to overfit and their generalization ability starts to degrade.

A synopsis of the results is shown in Table 2. The Table consists of two parts: Model performance, Event Detector performance. As mentioned in Section II, in addition to the performance of each classifier, the performance of each method is also assessed based on event-by-event evaluation as detected by a rule-based algorithm.

According to Table II, BiLSTM has the best performance in both “Model Performance” (accuracy of 95.1%) and “Event Detector Performance” (F-score of 85.0%) categories among all methods. Overall, the implemented DL methods have better performance than the classic SVM and the simple threshold-based algorithm.

In the following section, we discuss the performance of the methods in detail by identifying the strengths and weaknesses of each.

IV. DISCUSSION

In the previous section, we optimized DL and machine learning approaches to automatically detect CSA events from PSM data collected nocturnally. We demonstrated that the proposed DL models (i.e. TCN and BiLSTM) attained a high level of accuracy and outperformed the previously implemented methods (Table 2).

It is worth noting that attempts to automate the detection of SA events by adopting DL approaches have previously been made by others. In [21], a three-layer LSTM model was applied to photo-plethysmogram (PPG) signals for the classification of sleep apnea-hypopnea events [21]. In another study, the detection of sleep apnea was performed by a CNN model built from scratch based on oxygen saturation (SpO₂) signals. The CNN model outperformed other

models including linear discriminant analysis (LDA), SVM, bagging representation tree, and artificial neural networks (ANN) [23]. In [22], six DL methods were validated to find the optimal method for automatic detection of SA events from ECG signals. The methods included one-dimensional CNN, LSTM, two-dimensional CNN, deep neural network (DNN), and gated-recurrent unit (GRU) [22]. In [24], transfer learning was applied, using pre-trained CNN models for feature extraction (not as a classifier) from spectrogram images of pulse transition time (PTT) signals. With the features obtained, SVM and k-nearest neighbors (KNN) algorithms classified the participants to SA patients or healthy individuals. The paper did not address the separate classifications of each SA event.

All these studies employed inconvenient and obtrusive techniques to collect data. Moreover, the process of SA detection was either based on the whole signal sequence of interest or blind segmentation with no temporal event-by-event evaluations.

Inspired by the above, we applied DL approaches that are suitable for time series data to automatically detect CSA events from PSM signals. Moreover, in addition to evaluating the performance of the optimized models, we evaluated the implemented methods through a rule-based algorithm and event-by-event metrics. The number of SA events per hour of sleep indicates the severity of SA and therefore it is important for the clinical diagnosis of this disorder. However, developing and validating algorithms for SA monitoring only by comparing the number of estimated events with the one reported in the database is not sufficient. It is important to validate any detection method through a temporal event-by-event evaluation. Additionally, previous studies have suggested that respiratory event duration is an important physiological biomarker of SA and can be used for better management of the pathophysiology of this disorder [40]. Here, the rule-based algorithm can provide information about individual event durations since it is based on the temporal information of events in signal sequences.

PSMs measure the information of body movement through a mattress rather than measuring respiratory effort or the airflow signal directly. Patients are not restricted to a certain position on the mattress. Therefore, the position of the body parts, including the shoulders, torso, and limbs can vary with respect to the location of the sensors. Nevertheless, respiratory signals and therefore CSA events can still be captured by applying signal processing to PSM data. In [14], it was claimed that among different positions including prone, supine, and side, the supine position has the lowest signal-to-noise ratio, which affected the ability of a threshold-tuning algorithm to correctly classify CSA events. In this work, in order to mitigate the sensitivity of PSM to different body positions, the 72 measured PSM signals are combined based on their quality of capturing the breathing signal, and then the output is normalized to achieve consistency in the strength of the signal. However, according to the results in Table 2, given the sensitivity of PSM to body movements, the power detector

method as a simple method with an adaptive threshold is less reliable for CSA event detection, in comparison to more complex methods. The method has an event-based approach. It achieves the lowest precision (i.e., many FP events) on test data among all the implemented methods (Table 2).

In type-4 devices of sleep monitoring, signals such as oronasal thermal signals, positive airway pressure flow, or alternative signals such as RIP sum (sum of the thorax and abdomen belt signals) are usually used to score apneic events for adults. [31]. These signals are less sensitive to movement than PSM. Perhaps applying a simple threshold-based method to these signals could achieve better performance. Here, although the threshold in the power detector method is tuned adaptively, great signal amplitude variations due to body movements can cause the misdetection of normal breathing as an apnea, especially since normal breathing has very low variance in comparison to body movements. This can be seen as the weakness of a simple threshold-tuning method since it is incapable of reducing the sensitivity of the system to factors such as body movements.

Moreover, PSMs are used as unsupervised home-monitoring devices. Unlike supervised sensing with PSG, several unknown environmental parameters can affect the process of data acquisition using PSMs. Optimizing a threshold over all these parameters can be more challenging than for other approaches. Indeed, the presence of other physiological signals such as SpO2 would be helpful to reduce FP events and allow using a simpler method. However, this would be contrary to the unobtrusive use of PSM as a home monitoring device.

In comparison to the power detector method, SVM has better performance with an F-score of 70.8% (Table 2). SVM is a discriminative model that attempts to model the training data even if data is noisy. As mentioned in section III, in order to deal with imbalanced classes, the ratio of under-sampled “NA” class instances to over-sampled “A” class instances is optimized by applying 5-fold cross-validation on the training dataset. We found that for our setup and data, different values of α did not change the accuracy of the classifier significantly. At $\alpha = 0.45$ (Fig. 5), the classifier has a slightly better performance. Therefore, this value was used to generate the data. It should be noted that although the value of α , and consequently the scale of resampling the two classes “NA” and “A”, does not have much effect on the system optimization and associated metrics, it is a must-do step to prevent the “accuracy paradox” from imbalanced classes.

SVM treats the input data as a feature vector and therefore discards the temporal information of the signals. In contrast, BiLSTM is a static model that models temporal relations among time steps of sequence data or incorporates temporal consistency (by temporal pooling and/or regularization) [41]. For BiLSTM, the batch size is set to 1 to prevent the network from interpreting signals incorrectly due to padding (so that all sequences have the same length in the same batch of data).

Unlike RNNs (including BiLSTM) where the predictions for subsequent time steps must wait for their predecessors to

complete, in TCN networks convolutions can be performed in parallel because the same filter is applied in each layer. Therefore, in both testing and training, the input sequence can be processed as a whole in TCN networks, rather than sequentially in BiLSTM networks.

An epoch is one complete presentation of the training data to a DL network. The number of epochs used depends on a variety of factors, such as network architecture and solver, as well as the data available for the training procedures and the complexity of the problem. Due to the fundamental differences in the structure of TCN and BiLSTM networks, the number of epochs is individually optimized for each method.

For the BiLSTM and TCN approaches, the need for manual feature extraction is eliminated and discriminative features are directly learned from temporal PSM signals. These methods can be more flexible than SVM since each building block can be modular [42]. However, DL networks are computationally expensive, and they require high-end GPUs to train in a reasonable amount of time. Here, all approaches were implemented using MATLAB™ 9.6 with an NVIDIA GTX1050 Ti™ GPU on a Windows 10™ environment. Using the GPU, the TCN and BiLSTM networks needed respectively 182 min and 265 min to be trained, whereas with a single CPU the SVM method was trained in 118 min. These execution times are independent from the time taken to optimize the hyperparameters for each method. Table 3 summarizes the training time and hardware adopted for these three methods. Comparing the two DL approaches, training the TCN is found to be faster than BiLSTM.

TABLE 3. Speed comparison between the methods.

	Training Time (Min)	Hardware Component
SVM	118	Intel(R) Core™ i7-4790 CPU
BiLSTM	265	NVIDIA GTX1050 Ti™ GPU
TCN	182	NVIDIA GTX1050 Ti™ GPU

According to Table 2, the BiLSTM method outperforms all other methods with an F-score of 85%. Contrary to the results for generic data in [26], we observe that when dealing with PSM sequence data, the BiLSTM network leads to better performance than the TCN network. Overall, in comparison with the two other conventional classification methods, the implemented DL models had a greater ability to detect CSA events with the highest accuracy of 95.1% achieved by the BiLSTM. This is most likely due to exploring a broader range of features and finding a more suitable feature set by DL networks than those manually defined by the human operator.

V. CONCLUSION

In comparison with the manual operation of detecting and diagnosing SA by an expert, computer-assisted signal analysis systems can reduce errors due to inter- and intra- operation variability and fatigue caused by the tedious process of

annotation [35], [36]. In addition, most computer-based analyses can be performed in a more cost-effective and quicker way [18].

In this study, we combined the best modeling approach (DL models) and a practical unobtrusive data collection approach (PSM) to detect CSA events. To our best knowledge, this study is the first to use DL approaches for detecting CSA events on unattended data collection records from PSM as a home monitoring device.

The main contributions of our work can be summarized as follows: firstly, we employed PSM as an unobtrusive device to detect CSA events from data with no need for user intervention to collect data. In particular, PSM can be categorized as a type 4 sleep testing device to facilitate the identification of SA events in a variety of clinical scenarios such as home and intensive care units. Our approach can be utilized in conditions with personalized preferences on lighting, sound, and bed positions. Moreover, our work potentially has various applications for personalized health in smart hospitals, where automated detection of SA events can save time and effort and will result in more accurate values compared to manually going through PSG data.

Secondly, for automatic and computer-based detection of CSA from PSM signals, we implemented two DL approaches including BiLSTM (as a representative of RNNs) and TCN (as a representative of CNNs), and compared them to approaches used in previous studies: SVM (as representative of manual features combined with conventional classification approaches) [25] and a threshold-based method (implemented based on AASM rules for SA detection and diagnosis) [27]. We conducted various analyses to have the correct hyper-parameter setup for each of the four different approaches. In our experiments, the implemented DL models (with a highest accuracy of 95.1%, achieved by the BiLSTM model) outperformed classical machine learning approaches and have shown a greater ability to detect CSA events, because of their better capacity to automatically extract deep spatial and temporal features from the sequence.

Finally, we successfully solved the imbalanced data set problem. SA event detection is an anomaly detection where the data are highly imbalanced. Given the characteristics of each method, different approaches were implemented to handle the imbalance issue, including data resampling methods and a cost-sensitive approach (i.e. weighted cross-entropy).

We acknowledge that in supervised learning and especially in DL, a large amount of training data are required but are often not available in the medical domain. Obtaining datasets that are comprehensively labeled and annotated still remains a challenge in the biomedical field. [45]. Our study was affected by this limitation to some extent. Even with more than 400 hours of information, its variety was limited to only 9 patients. Nonetheless, we believe that the nocturnal data collected by PSM proved to contain valuable information and our study paves the way for the practical use of PSM as a home monitoring device on a larger scale.

ACKNOWLEDGMENT

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

REFERENCES

- [1] 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](https://doi.org/10.1109/TBME.2013.2282337).
- [2] H. Nazeran, A. Almas, K. Behbehani, J. Burk, and E. Lucas, "A fuzzy inference system for detection of obstructive sleep apnea," in *Proc. Conf. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Oct. 2001, vol. 2, pp. 1645–1648, doi: [10.1109/IEMBS.2001.1020530](https://doi.org/10.1109/IEMBS.2001.1020530).
- [3] A. Thommandram, J. M. Eklund, and C. McGregor, "Detection of apnoea from respiratory time series data using clinically recognizable features and kNN classification," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 5013–5016, doi: [10.1109/EMBC.2013.6610674](https://doi.org/10.1109/EMBC.2013.6610674).
- [4] M. E. Tagluk and N. Sezgin, "Classification of sleep apnea through sub-band energy of abdominal effort signal using wavelets + neural networks," *J. Med. Syst.*, vol. 34, no. 6, pp. 1111–1119, Dec. 2010, doi: [10.1007/s10916-009-9330-5](https://doi.org/10.1007/s10916-009-9330-5).
- [5] M. Emin Tagluk, M. Akin, and N. Sezgin, "Classification of sleep apnea by using wavelet transform and artificial neural networks," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1600–1607, Mar. 2010, doi: [10.1016/j.eswa.2009.06.049](https://doi.org/10.1016/j.eswa.2009.06.049).
- [6] J. V. Marcos, R. Hornero, D. Álvarez, F. D. Campo, and M. Aboy, "Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis," *Med. Biol. Comput.*, vol. 48, no. 9, pp. 895–902, Sep. 2010, doi: [10.1007/s11517-010-0646-6](https://doi.org/10.1007/s11517-010-0646-6).
- [7] J. V. Marcos, R. Hornero, D. Álvarez, F. del Campo, C. Zamarrón, and M. López, "Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry," *Comput. Methods Programs Biomed.*, vol. 92, no. 1, pp. 79–89, Oct. 2008, doi: [10.1016/j.cmpb.2008.05.006](https://doi.org/10.1016/j.cmpb.2008.05.006).
- [8] J. Victor Marcos, R. Hornero, D. Alvarez, F. Del Campo, C. Zamarron, and M. Lopez, "Single layer network classifiers to assist in the detection of obstructive sleep apnea syndrome from oximetry data," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 1651–1654, doi: [10.1109/IEMBS.2008.4649491](https://doi.org/10.1109/IEMBS.2008.4649491).
- [9] E. Dzięciolowska-Baran, A. Gawlikowska-Sroka, and J. Szczeniowski, "Diagnosis of sleep-disordered breathing in the home environment," *Adv. Exp. Med. Biol.*, vol. 1271, pp. 107–112, 2020, doi: [10.1007/5584_2020_497](https://doi.org/10.1007/5584_2020_497).
- [10] K. S. Park and S. H. Choi, "Smart technologies toward sleep monitoring at home," *Biomed. Eng. Lett.*, vol. 9, no. 1, pp. 73–85, Feb. 2019, doi: [10.1007/s13534-018-0091-2](https://doi.org/10.1007/s13534-018-0091-2).
- [11] J. Abad, A. Muñoz-Ferrer, M. Á. Cervantes, C. Esquinas, A. Marin, C. Martínez, J. Morera, and J. Ruiz, "Automatic video analysis for obstructive sleep apnea diagnosis," *Sleep*, vol. 39, no. 8, pp. 1507–1515, Aug. 2016, doi: [10.5665/sleep.6008](https://doi.org/10.5665/sleep.6008).
- [12] 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. 7091–7094, doi: [10.1109/IEMBS.2009.5332909](https://doi.org/10.1109/IEMBS.2009.5332909).
- [13] 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)*, Rochester, MN, USA, May 2017, pp. 303–307.
- [14] 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](https://doi.org/10.1109/TIM.2011.2123250).
- [15] H. Azimi, H. Liu, M. Bilodeau, B. Wallace, M. Bouchard, R. Goubran, and F. Knoefel, "Cloud processing of bed pressure sensor data to detect sleep apnea events," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Bari, Italy, Jun. 2020, pp. 1–5.
- [16] H. Azimi, M. Bouchard, R. Goubran, and F. Knoefel, "Unobtrusive screening of central sleep apnea from pressure sensors measurements: A patient-specific longitudinal study," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3282–3296, Jun. 2020.
- [17] N. Ganapathy, R. Swaminathan, and T. M. Deserno, "Deep learning on 1-D biosignals: A taxonomy-based survey," *Yearb. Med. Inform.*, vol. 27, no. 1, pp. 98–109, Aug. 2018, doi: [10.1055/s-0038-1667083](https://doi.org/10.1055/s-0038-1667083).
- [18] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Comput. Methods Programs Biomed.*, vol. 161, pp. 1–13, Jul. 2018, doi: [10.1016/j.cmpb.2018.04.005](https://doi.org/10.1016/j.cmpb.2018.04.005).
- [19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [20] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1419–1428, Oct. 2018, doi: [10.1093/jamia/ocy068](https://doi.org/10.1093/jamia/ocy068).
- [21] C.-H. Kang, U. Erdenebayar, J.-U. Park, and K.-J. Lee, "Multi-class classification of sleep apnea/hypopnea events based on long short-term memory using a photoplethysmography signal," *J. Med. Syst.*, vol. 44, no. 1, p. 14, Dec. 2019, doi: [10.1007/s10916-019-1485-0](https://doi.org/10.1007/s10916-019-1485-0).
- [22] U. Erdenebayar, Y. J. Kim, J.-U. Park, E. Y. Joo, and K.-J. Lee, "Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram," *Comput. Methods Programs Biomed.*, vol. 180, Oct. 2019, Art. no. 105001, doi: [10.1016/j.cmpb.2019.105001](https://doi.org/10.1016/j.cmpb.2019.105001).
- [23] H. T. Chaw, S. Kamolphiwong, and K. Wongsritrang, "Sleep apnea detection using deep learning," *Tehnički glasnik*, vol. 13, no. 4, pp. 261–266, Dec. 2019, doi: [10.31803/tg-20191104191722](https://doi.org/10.31803/tg-20191104191722).
- [24] S. Arslan Tuncer, B. Akılotu, and S. Toraman, "A deep learning-based decision support system for diagnosis of OSAS using PTT signals," *Med. Hypotheses*, vol. 127, pp. 15–22, Jun. 2019, doi: [10.1016/j.mehy.2019.03.026](https://doi.org/10.1016/j.mehy.2019.03.026).
- [25] 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.
- [26] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*. [Online]. Available: <http://arxiv.org/abs/1803.01271>
- [27] H. Azimi, S. S. Gilakjani, M. Bouchard, R. A. Goubran, and F. Knoefel, "Automatic apnea-hypopnea events detection using an alternative sensor," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Mar. 2018, pp. 1–5, doi: [10.1109/SAS.2018.8336762](https://doi.org/10.1109/SAS.2018.8336762).
- [28] B. Rim, N.-J. Sung, S. Min, and M. Hong, "Deep learning in physiological signal data: A survey," *Sensors*, vol. 20, no. 4, p. 969, Feb. 2020, doi: [10.3390/s20040969](https://doi.org/10.3390/s20040969).
- [29] 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](https://doi.org/10.1109/TIM.2012.2192342).
- [30] S. Soleimani, "Long-term sleep assessment by unobtrusive pressure sensor arrays," M.S. thesis, School Elect. Eng. Comput. Sci., Univ. Ottawa, Ottawa, ON, Canada, 2018.
- [31] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, S. Redline, K. P. Strohl, S. L. D. Ward, and M. M. Tangredi, "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. 05, pp. 597–619, Oct. 2012, doi: [10.5664/jcsm.2172](https://doi.org/10.5664/jcsm.2172).
- [32] 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, pp. 321–357, Jun. 2002, doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953).
- [33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [34] N. Wadhvani, N. Mehta, and N. Ruban, "IOT based biomedical wireless sensor networks and machine learning algorithms for detection of diseased conditions," in *Proc. Innov. Power Adv. Comput. Technol. (i-PACT)*, vol. 1, Mar. 2019, pp. 1–7, doi: [10.1109/i-PACT44901.2019.8960191](https://doi.org/10.1109/i-PACT44901.2019.8960191).
- [35] P. Badjatiya, L. J. Kurisinkel, M. Gupta, and V. Varma, "Attention-based neural text segmentation," in *Advances in Information Retrieval*. Cham, Switzerland: Springer, 2018, pp. 180–193, doi: [10.1007/978-3-319-76941-7_14](https://doi.org/10.1007/978-3-319-76941-7_14).

- [36] D. Alvarez-Estevéz and V. Moret-Bonillo, "Computer-assisted diagnosis of the sleep apnea-hypopnea syndrome: A review," *Sleep Disord.*, vol. 2015, Jul. 2015, Art. no. 237878, doi: [10.1155/2015/237878](https://doi.org/10.1155/2015/237878).
- [37] 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](https://doi.org/10.3390/app6060162).
- [38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [39] F. Ahamed, S. Shahrestani, and H. Cheung, "Intelligent fall detection with wearable IoT," in *Complex, Intelligent, and Software Intensive Systems*. Cham, Switzerland: Springer, 2020, pp. 391–401, doi: [10.1007/978-3-030-22354-0_35](https://doi.org/10.1007/978-3-030-22354-0_35).
- [40] S. Saha, M. Kabir, N. Montazeri Ghahjaverestan, M. Hafezi, B. Gavrilovic, K. Zhu, H. Alshaer, and A. Yadollahi, "Portable diagnosis of sleep apnea with the validation of individual event detection," *Sleep Med.*, vol. 69, pp. 51–57, May 2020, doi: [10.1016/j.sleep.2019.12.021](https://doi.org/10.1016/j.sleep.2019.12.021).
- [41] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014, doi: [10.1016/j.patrec.2014.01.008](https://doi.org/10.1016/j.patrec.2014.01.008).
- [42] P. M. Kulkarni, Z. Xiao, E. J. Robinson, A. S. Jami, J. Zhang, H. Zhou, S. E. Henin, A. A. Liu, R. S. Osorio, J. Wang, and Z. Chen, "A deep learning approach for real-time detection of sleep spindles," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 036004, doi: [10.1088/1741-2552/ab0933](https://doi.org/10.1088/1741-2552/ab0933).
- [43] J. B. Stephansen et al., "Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy," *Nature Commun.*, vol. 9, no. 1, pp. 1–15, Dec. 2018, doi: [10.1038/s41467-018-07229-3](https://doi.org/10.1038/s41467-018-07229-3).
- [44] L. Bote-Curiel, S. Muñoz-Romero, A. Gerrero-Curienes, and J. L. Rojo-Álvarez, "Deep learning and big data in healthcare: A double review for critical beginners," *Appl. Sci.*, vol. 9, no. 11, p. 2331, Jun. 2019, doi: [10.3390/app9112331](https://doi.org/10.3390/app9112331).
- [45] R. Zemouri, N. Zerhouni, and D. Racoceanu, "Deep learning in the biomedical applications: Recent and future status," *Appl. Sci.*, vol. 9, no. 8, p. 1526, Apr. 2019, doi: [10.3390/app9081526](https://doi.org/10.3390/app9081526).



MARTIN BOUCHARD (Senior Member, IEEE) received the B.Eng., M.Sc.A., and Ph.D. degrees in electrical engineering from the Université de Sherbrooke, Canada, in 1993, 1995, and 1997, respectively. In 1996, he co-founded SoftdB Inc., In January 1998, he joined the School of Electrical Engineering and Computer Science, University of Ottawa, Canada, where he is currently a Professor. He has conducted research activities and consulting activities with over 25 private sector and governmental partners. His current research interests include applied signal processing and machine learning, with an emphasis on speech, audio, acoustics, hearing aids, and biomedical engineering applications. He was a member of the Speech and Language Technical Committee of the IEEE Signal Processing Society from 2009 to 2011. He is a member of the Ordre des Ingénieurs du Québec. He serves as an Associate Editor for *EURASIP Journal on Audio, Speech, and Music Processing* from 2005 to 2011 and the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2009 to 2010.



RAFIK GOUBRAN (Life Member, 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. He joined the Department of Systems and Computer Engineering, Carleton University, in 1987. He was the Chair of the Department from 1997 to 2006 and the Dean of the Faculty of Engineering and Design, Carleton University, from 2006 to 2016. From 2006 to 2009, he was the Founding Director with the Ottawa-Carleton Institute for Biomedical Engineering, Ottawa. He is currently a Chancellor's Professor and the Vice-President (Research and International) with Carleton University. He is also an Affiliate Investigator with the Bruyère Research Institute, Ottawa. He was involved in many research projects with industry and government organizations in signal processing, biomedical engineering, sensors, data analytics, and design of smart homes for independent living. He has completed the co-supervision of 96 graduate students and coauthored over 300 journal and conference papers. He 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, the Diploma degree 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, Canada, in 1998. He is currently a Physician with the Bruyère Memory Program, Elisabeth-Bruyère Hospital, Ottawa, a Senior Investigator with the Bruyère Research Institute, Ottawa, 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 a Co-Founder and a Co-Leader with TAFETA: Smart Systems of Health. His current research interests include use of technology in caring for older adults and remote monitoring of mobility and breathing, including use of technology to monitor and minimize cognitive decline. He received the Certification from the College of Family Physicians, Université de Sherbrooke, Sherbrooke, QC, Canada, in 1991.



HILDA AZIMI (Member, IEEE) received the B.A.Sc. and M.A.Sc. degrees in biomedical engineering from the Science and Research Branch, Islamic Azad University, in 2011 and 2014, respectively, Tehran, Iran. She is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Ottawa, Canada. She is working on longitudinal sleep disorders analysis, especially sleep apnea. Her project is unobtrusive elderly monitoring using sensor-based technologies to enhance the quality of life amongst older adults. Her research interests include applied signal processing and machine learning, with an emphasis on physiological signal processing.



PENGCHENG XI received the B.E. and M.E. degrees in computer engineering, the M.S. degree in computer science from the University of Ottawa, Canada, and the Ph.D. degree in electrical and computer engineering from Carleton University, Canada. He is currently a Research Officer with the Computer Vision and Graphics Team, National Research Council Canada (NRC). His research interests include applied machine learning and deep learning, with applications to human modeling, computer vision, image processing, and signal analysis.