Portable Detection of Apnea and Hypopnea Events Using Bio-Impedance of the Chest and Deep Learning

Tom Van Steenkiste, Willemijn Groenendaal, Pauline Dreesen, Seulki Lee, Susie Klerkx, Ruben de Francisco, Dirk Deschrijver, and Tom Dhaene

***Abstract*—Sleep apnea is one of the most common sleep-related breathing disorders. It is diagnosed through an overnight sleep study in a specialized sleep clinic. This setup is expensive and the number of beds and staff are limited, leading to a long waiting time. To enable more patients to be tested, and repeated monitoring for diagnosed patients, portable sleep monitoring devices are being developed. These devices automatically detect sleep apnea events in one or more respiration-related signals. There are multiple methods to measure respiration, with varying levels of signal quality and comfort for the patient. In this study, the potential of using the bio-impedance (bioZ) of the chest as a respiratory surrogate is analyzed. A novel portable device is presented, combined with a two-phase Long Short-Term Memory (LSTM) deep learning algorithm for automated event detection. The setup is benchmarked using simultaneous recordings of the device and the traditional polysomnography in 25 patients. The results demonstrate that using only the bioZ, an area under the precision-recall curve of 46.9% can be achieved, which is on par with automatic scoring using a polysomnography respiration channel. The sensitivity, specificity and accuracy are 58.4%, 76.2% and 72.8% respectively. This confirms the potential of using the bioZ device and deep learning algorithm for automatically detecting sleep respiration events during the night, in a portable and comfortable setup.**

***Index Terms*—Sleep apnea, HSAT, bio-impedance, deep-learning.**

I. Introduction

SLEEP apnea is one of the most common sleep-related breathing disorders and consists of breathing pauses or shallow breathing during the night, known as apneic events [1]. These events can be categorized as either obstructive sleep apnea (OSA) when the airway is blocked by the throat muscles or central sleep apnea (CSA) when the signals to control the breathing are disturbed. When the breathing has become shallow but is not yet fully disturbed, it is classified as obstructive or central hypopnea. The consequences of undiagnosed sleep apnea can be severe, including hypertension, cardiac arrhythmia, heart attacks and strokes [2]–[4]. It has also been shown that sleep apnea patients have an increased chance of being involved in motor vehicle collisions [5].

Studies report that in some countries, over 50% of adults suffer from sleep-disordered breathing [6]. Yet, many cases remain undiagnosed as people are often unaware of their condition. For diagnosis, patients are admitted in a dedicated sleep clinic where they are monitored overnight using a polysomnograph (PSG) which measures a variety of signals pertaining to respiration, brain activity, sleep stages, heart rate, oxygen saturation and others. Afterwards, the recordings are analyzed by trained sleep technicians and annotated using a reference manual such as the American Association of Sleep Medicine (AASM) guidelines [7]. The condition of the patient is summarized into an apnea-hypopnea-index (AHI) representing the number of events per hour of sleep.

The complex PSG setup, and the limited number of beds and staff in sleep clinics are leading to large costs [8] and a long waiting time [9]. In addition, the PSG setup requires many sensors attached to the patient and a night of sleep in an unfamiliar hospital bed. This leads to an uncomfortable analysis and an inaccurate representation of an actual night of sleep in the patients home [10], [11]. To allow a more comfortable and representative analysis, while also enabling more patients to be tested and diagnosed patients to be continuously monitored, home sleep apnea tests (HSAT) are being developed. This is the result of recent hardware and algorithmic innovations leading to wearable devices for automated precision monitoring [12] and for continuous and longitudinal health monitoring [13].

As recommended by the AASM guidelines for portable devices [14], these HSAT typically include a respiration

Manuscript received August 21, 2019; revised December 8, 2019 and January 10, 2020; accepted January 14, 2020. Date of publication January 20, 2020; date of current version September 3, 2020. This work was supported by the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme. *(Corresponding author: Tom Van Steenkiste.)*

T. Van Steenkiste, D. Deschrijver, and T. Dhaene are with the Department of Information Technology, Ghent University - imec, IDLab, B-9052 Gent, Belgium (e-mail: tomd.vansteenkiste@ugent.be; dirk.deschrijver@ugent.be; tom.dhaene@ugent.be).

W. Groenendaal and S. Lee are with imec the Netherlands/Holst Centre, 5656 AE Eindhoven, The Netherlands (e-mail: willemijn.groenendaal@imec-nl.nl; seulki.lee@imec-nl.nl).

P. Dreesen is with Future Health, Ziekenhuis Oost-Limburg, Genk, Belgium and Mobile Health Unit, Faculty of Medicine and Life Sciences, Hasselt University, 3500 Hasselt, Belgium (e-mail: pauline.dreesen@zol.be).

S. Klerkx is with the Department of Pneumology, Ziekenhuis Oost-Limburg, 3600 Genk, Belgium (e-mail: susie.klerkx@zol.be).

R. de Francisco is with Onera Health, 5617 BD Eindhoven, The Netherlands (e-mail: ruben.defrancisco@oneramedical.com).

Digital Object Identifier 10.1109/JBHI.2020.2967872

measurement. This can be based on typical PSG sensors such as resistive bands around the chest or abdomen [15], and nasal or oral airflow [15]–[17]. However, many other types of sensor have been developed such as nearable sensors, which require proximity but do not require patient contact. Examples include load cells under the bed [18], film based sensors [19] or mobility monitors [20]. Respiration can also be measured directly through microelectronic systems (MEMS) [21] or wearable piezo-electric bands [22]. In addition, respiration can be extracted from other physiological measurements such as ECG [23], [24] but this approach is susceptible to noise [25]. Other popular HSAT methods include the use of pulse oximetry [26]–[29] or sound [30]–[32]. Lastly, there are several devices that combine multiple signals [15], [33], [34]. These devices can by categorized using the SCOPER system [35].

The recorded data from these HSAT is processed through automated methods ranging from threshold algorithms [36] and rule-based algorithms [16] to support vector machines [37] and artificial neural networks [38]. Recently, deep learning algorithms have also been proposed and have demonstrated good performance [39]. Commercial software for automated processing of respiratory signals can also be used [40].

In various medical domains, bioZ measurements have already been suggested for measuring, among others, respiration of the patient [41], [42]. However, the use of bioZ for sleep apnea detection has not yet been thoroughly analyzed or described. The key objective of this study is to analyze and discuss the potential of using a bioZ measurement of the chest as a respiratory surrogate for detecting sleep apnea events. A novel portable device is presented with a combined two-phase deep learning algorithm, and the performance is benchmarked on a clinically gathered dataset.

The outline of this paper is as follows. In Section II, a device is presented to directly measure the bioZ of the chest along with ECG and acceleration of the patient. The deep learning algorithm is presented in Section III. The experimental setup to analyze this approach is presented in Section IV. In Section V, the results are presented and in Section VI these results are discussed. Finally, conclusions are made and future work is considered in Section VII.

II. Wearable Bio-Impedance Device

To analyze the possibility of using the bioZ to detect sleep apnea events, a novel device, denoted as ROBIN, is proposed which measures and records the bioZ of the chest, in addition to ECG and acceleration. First, the hardware is discussed. Then, an algorithm is presented for aligning the recordings of the ROBIN with the PSG data to enable a comparison study.

A. ROBIN Device

The ROBIN was developed by imec (imec The Netherlands, Eindhoven). A detailed schematic of the hardware setup is shown in Fig. 1. The device itself is demonstrated in Fig. 2.

The device is able to measure and record several bio-signals. In this study, the signals of interest are the ECG, bioZ and motion of the patient. The ECG and bioZ are measured directly by MUSEIC, a system on a chip (SoC) designed by imec (imec The Netherlands, Eindhoven) [43]. Motion of the patient is measured via a separate accelerometer IC and data is transferred to MUSEIC for synchronization.

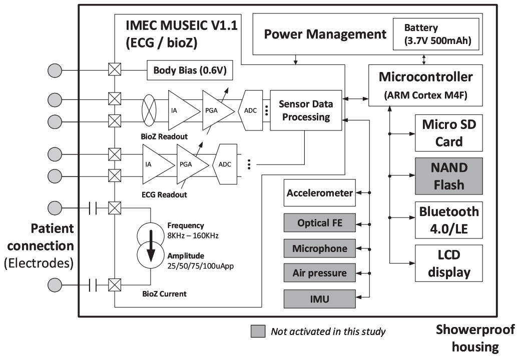


Fig. 1. Detailed hardware schematic of the ROBIN device used in this study.

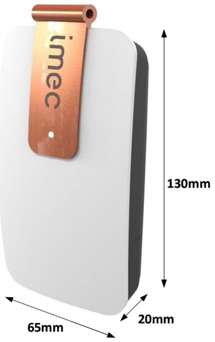


Fig. 2. ROBIN device used in this study. It is a compact wearable with minimal discomfort for the patient. The device is worn around the neck and is attached to the chest via adhesive stickers.

In order to measure bioZ, the tissue of the patient is stimulated with an AC current from the current generator of MUSEIC. The captured signal is demodulated to the baseband and amplified with the instrumentation amplifier (IA) and programmable gain amplifier (PGA) in the front-end readout circuits of MUSEIC. The frequency and amplitude of the current generator, and the gain of the amplifiers are configurable for the target application. Afterwards, the signal is digitized by an ADC, packetized in MUSEIC, and transmitted to the external microcontroller. This microcontroller, an ARM Cortex M4F, is responsible for processing the data. Once the microcontroller receives the data from the sensors, it stores the data into the on-board memory to enable offline retrieval for further processing and analysis.

In this study, the bioZ is measured at the chest to provide a measurement of respiration. An in-depth discussion on bioZ is provided in [41]. Details on the performance of the bioZ measurement chip used in this work are provided in [43]. Skin surface electrodes provide the current stimulation and voltage measurement. A tetrapolar electrode configuration is used to avoid the effect of electrode-tissue impedance. Two electrodes are dedicated to current stimulation, and the other two electrodes are adopted for the voltage measurement. An additional electrode is used to connect the patient to a fixed DC voltage (0.6 V) in order to keep the input bioZ signals within the dynamic range of the amplifiers.

The stimulation frequency of the ROBIN can be configured between 8 kHz and 160 kHz. For the sleep apnea study, the stimulating alternating current in the form of a pseudo sine wave is set to 160 kHz as this enables good sensitivity and linearity against impedance changes. The current amplitude can be configured at four different values (25, 50, 75, and 100uApp, corresponding to 8.8, 17.7, 26.5, and 35.4uArms). The international safety standard of medical electrical equipment (IEC 60601-1) states a maximum current of 50uArms at 160 kHz for stimulating tissues. In this study, 100uApp (35.4uArms) is used to maximize the signal-to-noise ratio. The stimulating current will generate a voltage potential in the region of interest, proportional to the impedance of the tissue. This induced voltage is captured by the amplifiers and demodulated to the baseband. In this way, only impedance changes without any stimulation current components remain in the final bioZ data.

With these configuration settings, the bioZ sensor of the ROBIN can detect impedance changes as small as 0.1 Ω. The measurable impedance range is from $-150\ \Omega$ to 120 Ω, where a negative sign of the impedance indicates the opposite polarity of stimulation current and measured voltage. Depending on the measurement subject, impedance changes due to respiration are 1–2 Ω, meaning that the bioZ sensor of the ROBIN device can measure the respiration effectively.

The bioZ measurements are recorded at 1024 Hz as complex-valued time-series data. For further analysis as a respiration surrogate, the magnitude is computed and the data is downsampled to 125 Hz for storage on the offline processing server of the study.

Two electrodes are used to collect single lead ECG data from the measurement subject with a sampling frequency of 512 Hz. Although limited respiration information can also be extracted from ECG data, the focus of this study is on the direct and full respiratory measurement using the bioZ sensor. The ECG data is used for data alignment. The ROBIN device also includes an accelerometer sensor which is used to measure dynamic movement of the patient to enable visual inspection of the data for verification of the recording.

B. PSG Alignment of ROBIN Data

To enable a comparison of the ROBIN data with the gold standard PSG data and the PSG annotations of the trained sleep technicians, the recordings of the ROBIN and PSG need to be synchronized. As there can be an offset on the starting time of the recording, an initial delay needs to be estimated. In addition, due to small differences in actual sampling frequency, the recordings of both devices can start drifting apart. Even a minimal difference in sampling frequency can lead to a large drift after a couple of hours of recording. For example, a difference of 0.1 Hz in a sampling frequency of 50 Hz can lead to a drift of 57.6 seconds after an eight hour recording.

The ECG of the patient is recorded by both devices. As QRS complexes in simultaneously recorded ECG measurements should occur in a synchronized way, the ECG signals provide a straightforward base to align the recordings. A possible approach for alignment is by consecutively resampling the signals and estimating the delay. In a first step, both ECG recordings are resampled to 50 Hz based on their listed sampling frequency. Then, an iterative process starts in which the starting delay and sampling frequency drift are computed and evaluated based on the cross-correlation between the two ECG signals. This process is repeated until convergence, which typically occurs after two iterations.

The alignment is only required for the analysis of the hardware and algorithms in this study. In practical settings, when the device is deployed, the PSG data is no longer recorded and hence, no alignment is needed.

III. Automated Sleep Apnea Detection Algorithm

In this section, an algorithm is presented to automatically detect sleep apnea events in respiratory data based on long short-term memory (LSTM) neural networks [44] which are trained in two separate phases: training phase 1 and training phase 2. First, the extraction and preprocessing of epochs of respiratory data is discussed. Next, the deep learning model is presented.

A. Epoch Creation

To filter noise and extract respiratory information, the recorded signal, either from the PSG or the ROBIN, is passed through a fourth order zero-phase-shift Butterworth low-pass filter with a cutoff frequency of 0.7 Hz. The resulting signal is filtered with a moving average filter of four seconds to reduce motion artifacts and to limit baseline wander. Finally, the signal is downsampled to 5 Hz for further processing. This final downsampling is done in two stages, each followed by an eight order Chebyshev type I zero-phase-shift filter to reduce resampling artifacts.

The filtered signal is split into epochs of 30 seconds each, with a stride of 1 second between consecutive epochs. The binary ground truth labels are provided by the annotations in the PSG data. If at the end of the epoch an apnea event was annotated, the entire epoch is labeled as positive.

The data of each epoch is scaled to the interval $[-1, 1]$ to increase the learning capacity of the LSTM nodes of the network. Instead of computing the normalization factors based on the minimum and maximum value of each epoch as in [39], the normalization factors are computed over the duration of the last M epochs, including the current epoch. This enables the inclusion of long-term contextual information into a single epoch. The full process is demonstrated in Fig. 3.

B. Deep Learning Model

The two-phase deep learning model is based on LSTM neural networks which are a type of recurrent neural network. Such models have been successfully applied across several

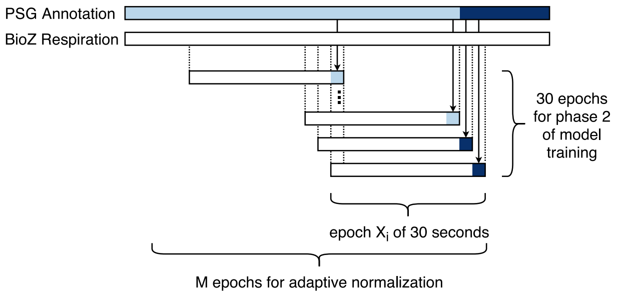

PSG Annotation
BioZ Respiration
30 epochs for phase 2 of model training
epoch X$_i$ of 30 seconds
M epochs for adaptive normalization

Fig. 3. Creating epochs from the recorded sleep apnea data. Labels are determined based on the annotation at the end of the epoch. The adaptive normalization procedure uses the last M epochs for computation of the scaling factors. In training phase 2 of the deep learning model, 30 epochs are used for the event detection in epoch X_i.

medical domains, including sleep apnea [45]–[48]. A theoretical discussion on LSTM neural networks can be found in [44].

In recent work [39], balanced bootstrapping [49] was proposed as a method to include all samples of an imbalanced dataset in a final model by training multiple models on the generated subdatasets. Each subset is used to train a separate LSTM model. During the prediction step, each separate model generates a prediction which is then combined through an aggregation metric such as averaging. In this work, this is denoted as training phase 1 and instead of averaging over the multiple predictions, they are used as inputs for training phase 2 LSTM models. This second training phase effectively replaces the initial averaging as aggregation method. By taking into account the variations in predictions across the different training phase 1 LSTM networks, and across time within a single epoch, more contextual information can be used in the final prediction. The complete two-phase approach is demonstrated in Fig. 4.

During training phase 1, N LSTM models are constructed using the N balanced bootstrap datasets (BB$_1$ - BB$_N$) as discussed in [39]. Each of the N models computes a series of outputs k_{i-29} up to k_i for epoch i. This results in N times 30 estimates, containing the prediction of each balanced bootstrap model at each point during the epoch, providing additional context over only a single epoch. This method uses data from the current epoch and the previous 29 overlapping epochs, leading to 59 seconds of data being used. The 30 estimations of the N balanced bootstrapped networks are combined into a single time-varying $N \times 30$ matrix which is used by the training phase 2 network.

All LSTM networks in the model follow a similar architecture as introduced in [39]. The LSTM layers are succeeded by a dropout layer to reduce overfitting and encourage generalization. The resulting network has several hyperparameters that need to be tuned. For this, Bayesian Optimization (BO) is used, which is a powerful method for tuning the hyperparameters of machine learning models [50], [51]. Each phase of the model has a separate set of parameters and the range of these hyperparameters is shown in Table I.

TABLE I
BOUNDARIES OF THE HYPERPARAMETERS FOR THE ENTIRE MODEL, USED DURING THE BAYESIAN OPTIMIZATION PROCEDURE

parameter	meaning	min	max
n_1 & n_2	number of LSTM nodes	10	200
p_1 & p_2	dropout probability	0.1	0.5
M	normalization range [seconds]	1	300

TABLE II
OVERVIEW OF PATIENT CHARACTERISTICS

	male	female
number	22	3
number$_{normal}$	7	1
number$_{mild}$	15	2
AHI	7.5 ± 3.3	5.4 ± 1.9
age	57.9 ± 12.7	56.7 ± 2.9
BMI	30.2 ± 3.7	28.1 ± 3.4

IV. EXPERIMENTAL SETUP

A. Data Acquisition

Data was gathered in a clinical setting from patients enrolled for an overnight PSG analysis at Ziekenhuis Oost-Limburg, a hospital in Belgium. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethical Committee before study onset (CME ZOL, reference: 16/042 U). All participants provided written informed consent before inclusion.

The PSG and ROBIN recorded data simultaneously. Afterwards, the PSG data was annotated by trained sleep technicians using AASM guidelines [7].

This study included all patients which had a successful ECG recording in both devices, as required for the alignment algorithm, and for which there was a successful bioZ recording, i.e. there was no saturation in the bioZ signal. A saturation in the bioZ signal occurs when the device is not properly tuned to a specific patient's characteristics. This can easily be circumvented in a following recording by changing the ROBIN settings.

In total, 25 patients were used in this study. Table II shows the patient characteristics. The patient was allowed to sleep in a hospital bed as desired and was able to change position during the night.

The PSG device recorded the typical signals used for a full sleep study in a sleep clinic. The recorded PSG respiration measurements included in this study are:

- PSG$_{\text{abd. belt}}$: Abdominal respiratory belt below the lower edge of the left ribcage.
- PSG$_{\text{thor. belt}}$: Thoracic respiratory belt below left armpit.

The ROBIN recorded chest bioZ, ECG and acceleration as discussed in Section II-A. The bioZ respiration signal is denoted as ROBIN$_{\text{bioZ}}$.

B. Device and Model Performance Evaluation

All recorded data was visually inspected in multiple stages. First, the PSG data was analyzed by the trained staff to provide annotations. Then, ROBIN recordings were analyzed to check

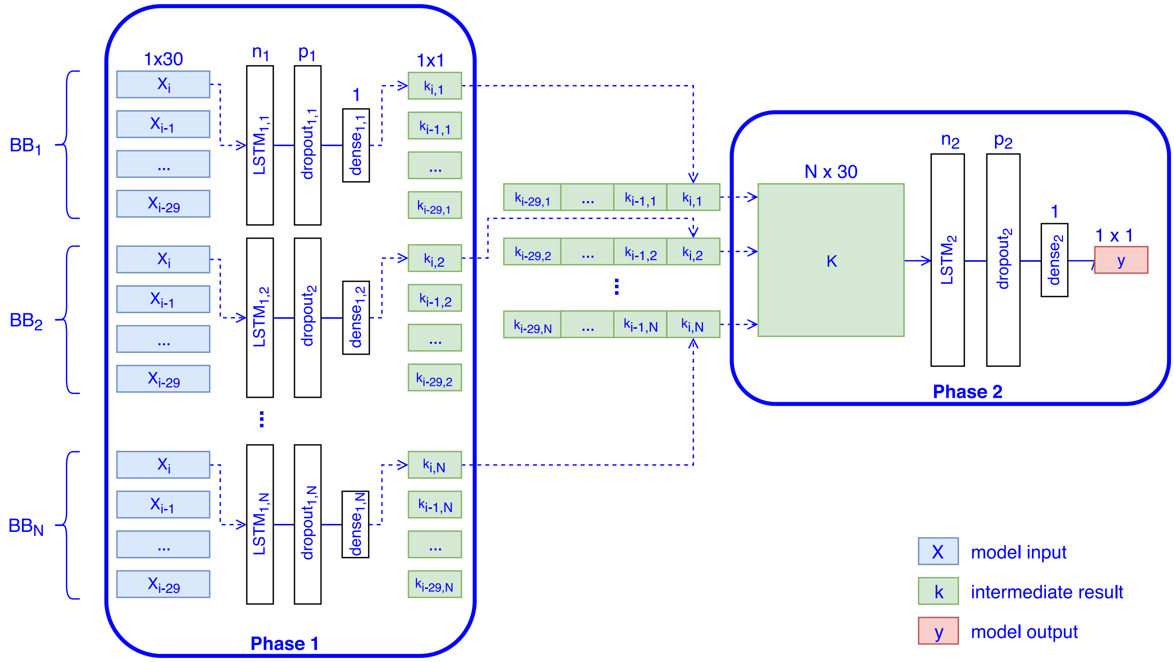


Fig. 4. Two-phase deep learning model for automated detection of sleep apnea events. The input X is passed through each of the N balanced bootstrapped networks of *training phase 1*, resulting in an intermediate result k. These results are then used in the *training phase 2* model to refine and enhance the prediction performance.

recording success. Finally, the quality of the ECG recording of both devices was inspected for subsequent alignment.

The PSG data with annotations from trained staff and the ROBIN data were aligned as discussed in Section II and preprocessed as discussed in Section III. In total, there were 4647 OSA epochs, 2851 CSA epochs, 19469 hypopnea epochs and 92096 epochs without sleep apnea.

The gold-standard PSG respiratory recordings were used to construct two models: $\mathrm{PSG}_{\mathrm{abd.\ belt}}$ and $\mathrm{PSG}_{\mathrm{thor.\ belt}}$. The bioZ data from the ROBIN was used to construct $\mathrm{ROBIN}_{\mathrm{bioZ}}$. This enables a comparison of the automated detection algorithm versus the annotations of trained staff as well as a comparison of the bioZ measurement versus PSG recordings.

For each model, the data was split across a five-fold per-patient cross-validation setup. Within each iteration, 15 patients were used for training the two-phase deep learning model, five patients were used for validation of the model and five patients were used as a separate test set for final performance evaluation. This was repeated for each fold, which enables a statistical analysis of the results. Patients were randomized but the order is static across the three models to ensure a fair comparison.

Analysis of sleep apnea algorithms is done using several metrics for binary classification problems. The performance is analyzed using the sensitivity, specificity and accuracy metrics which are based on the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN).

$$\mathrm{sensitivity} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$$

$$\mathrm{specificity} = \mathrm{TN}/(\mathrm{TN} + \mathrm{FP})$$

$$\mathrm{accuracy} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{FP} + \mathrm{TN} + \mathrm{TP} + \mathrm{FN})$$

The goal of the algorithm is to achieve a high score across all these metrics. However, the binary classifier outputs a probability of an event occurring, and as such, the final values of these metrics can easily be influenced by changing a decision threshold $0 \leq \tau \leq 1$. To analyze the performance of the model across all possible thresholds, the Receiver Operator Characteristic (ROC) curve, demonstrating the balance between sensitivity and specificity, and the associated area under it (AUROC) are computed.

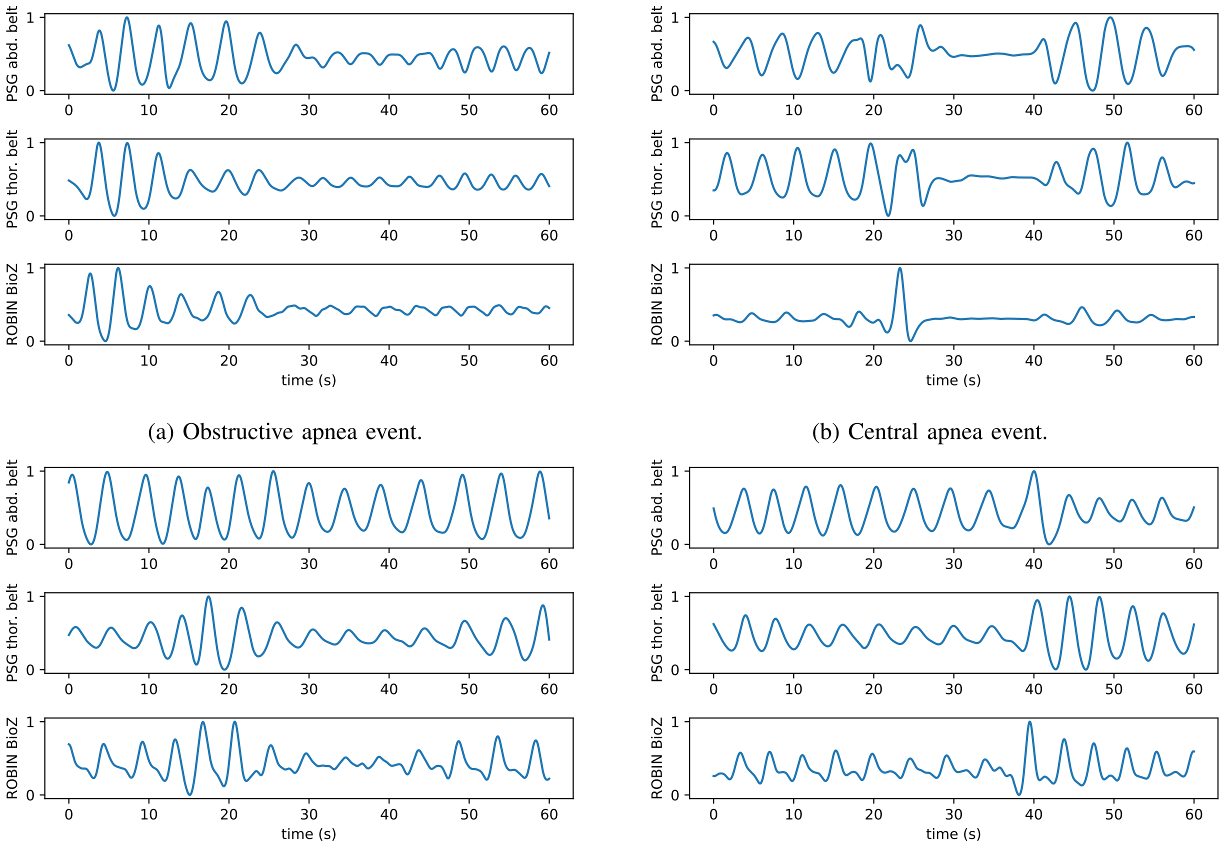
Another factor influencing the analysis of the model is the imbalance in the dataset. When working with highly imbalanced datasets, sensitivity and specificity scores can provide misleading insights [52]. As the number of apnea epochs in the data is much less than the number of non-apnea epochs, this imbalance has to be taken into account. For this, the precision, also known as the positive-predictive-value (PPV) and the negative-predictive-value (NPV) metrics are used.

$$\mathrm{precision} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$$

$$\mathrm{NPV} = \mathrm{TN}/(\mathrm{TN} + \mathrm{FN})$$

To analyze the performance of the imbalanced dataset across a range of decision thresholds, the precision-recall (PR) curve and associated area under the curve (AUPRC) are computed. This AUPRC metric is the main analysis point of this study, and is used as a driving metric for the BO of the model hyperparameters.

Statistical analysis for comparing the $\mathrm{ROBIN}_{\mathrm{bioZ}}$ against the baseline $\mathrm{PSG}_{\mathrm{abd.\ belt}}$ and $\mathrm{PSG}_{\mathrm{thor.\ belt}}$ is done using a paired two-tailed T-test across the different fold results with target significance level of $p < 0.05$. In addition, the statitical significance of



(a) Obstructive apnea event.

(b) Central apnea event.

(c) Hypopnea event.

(d) No apnea event.

Fig. 5. Example of respiration traces recorded with $\text{PSG}_{\text{abd. belt}}$, $\text{PSG}_{\text{thor. belt}}$ and $\text{ROBIN}_{\text{bioZ}}$ sensors for the different event types. Each event occurs at the 30 seconds mark. Note that in Fig. 5(d), there is no sleep apnea event even though there is a visual disturbance in the respiration. All three signals capture the respiration pattern of the patient and there is an observable similarity between the different signals.

any improvement due to the second phase model is also analyzed using a paired two-tailed T-test with target significance level of $p < 0.05$. All tests are based on the AUPRC metric.

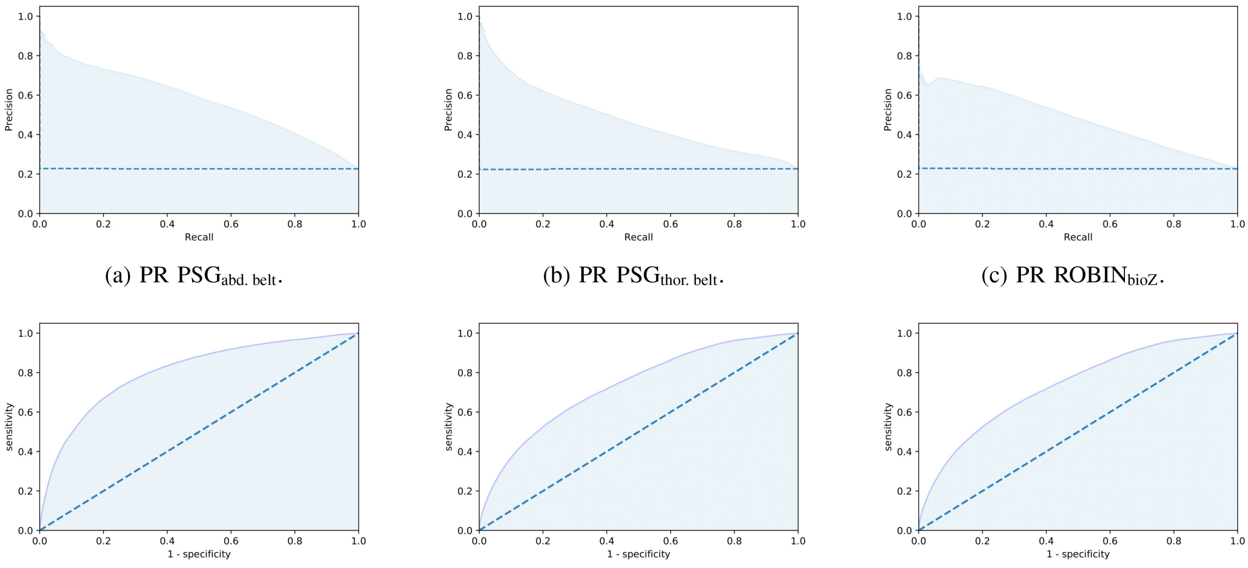
Unless otherwise mentioned, all metrics are computed using the predictions of the L2 models and binary classification metrics are reported at the decision threshold $p_{\text{event}} > 0.5$ which was learned during training.

V. Results

Fig. 5 presents a snapshot comparison between the different PSG respiratory signals ($\text{PSG}_{\text{abd. belt}}$ and $\text{PSG}_{\text{thor. belt}}$) and the ROBIN bioZ signal ($\text{ROBIN}_{\text{bioZ.}}$). There is a clear visual similarity between the three different signals. Note that in Fig. 5(d), there is no sleep apnea event even though there is a disturbance or change in respiration across the three signals at the 40 second mark.

The main analysis point of this study is the PR curve, which is shown in Fig. 6 in addition to the ROC curve for the three different models. There is a similarity across all models. The $\text{PSG}_{\text{abd. belt}}$ leads to the largest area under the curve. This summary of the PR curve is also visualized in Fig. 7 and Fig. 8 where the AUPRC metric is shown for the train, validation and test dataset of the first and second training phase which resulted in $L1$ models and $L2$ models respectively. The analysis across the different datasets demonstrates the stability and generalizability of the results. After the second training phase, the $\text{ROBIN}_{\text{bioZ}}$ achieves similar performance as the $\text{PSG}_{\text{thor. belt}}$ model. The relative improvement due to the second training phase is shown in Fig. 9. The $\text{ROBIN}_{\text{bioZ}}$ model benefits the most from this addition. The improvement is not substantial for the $\text{PSG}_{\text{abd. belt}}$ and the $\text{PSG}_{\text{thor. belt}}$ models. The performance on the training dataset increases but the performance of the test dataset remains the same, indicating that the models struggle to generalize the additional knowledge in the models. However, the $\text{ROBIN}_{\text{bioZ}}$ model is able to effectively generalize to new patients, as reflected by the large improvement on the test dataset.

In addition to the AUPRC, other typical metrics were computed. A full overview of all recorded metrics, evaluated on the test set as per-epoch classification performance, is provided in Table III. All metrics are reported at the decision point ($p_{\text{event}} > 0.5$) learned during training of the model. The statistical significance of these results is analyzed based on the AUPRC as described in Section IV. Comparing $\text{ROBIN}_{\text{bioZ}}$ versus $\text{PSG}_{\text{abd. belt}}$ and $\text{ROBIN}_{\text{bioZ}}$ versus $\text{PSG}_{\text{thor. belt}}$ results in p-values of 0.03 and 0.71 respectively. These results show that



(a) PR PSG$_{\text{abd. belt}}$.

(b) PR PSG$_{\text{thor. belt}}$.

(c) PR ROBIN$_{\text{bioZ}}$.

(d) ROC PSG$_{\text{abd. belt}}$.

(e) ROC PSG$_{\text{thor. belt}}$.

(f) ROC ROBIN$_{\text{bioZ}}$.

Fig. 6. ROC and PR curves for the three different models. The dashed line indicates the performance of a model with random predictions based on the imbalance in the training data.

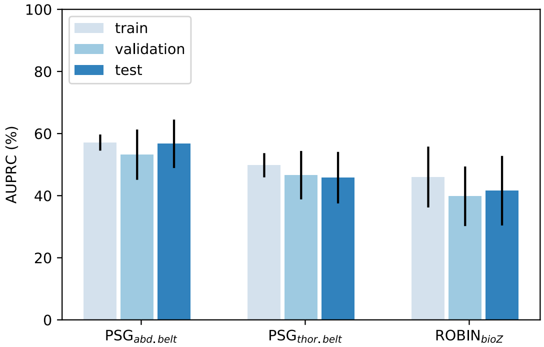


Fig. 7. AUPRC scores for the three different sensors, evaluated at the first layer (L1) of the model. All three models are capable of automatically detecting sleep apnea events.

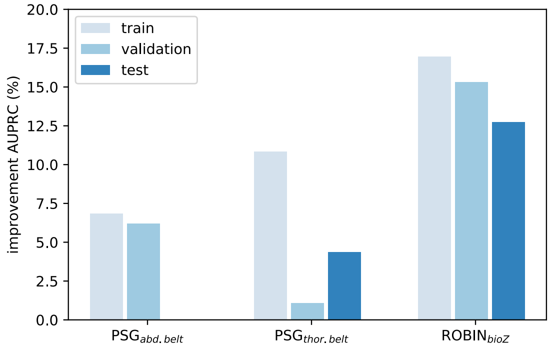


Fig. 9. Relative improvement of AUPRC score through the introduction of the second training phase. The ROBIN$_{\text{bioZ}}$ scores demonstrate a significant improvement through the introduction of the second model layer.

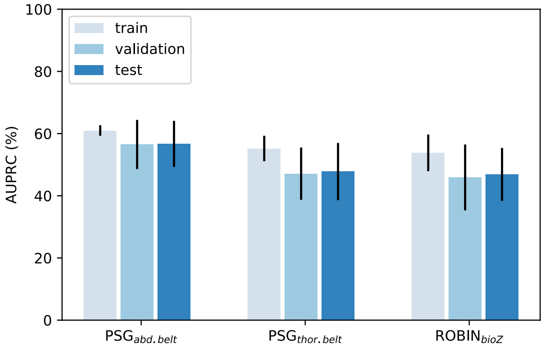


Fig. 8. AUPRC scores for the three different sensors, evaluated at the second layer (L2) of the model. With the addition of the second training phase, the ROBIN$_{\text{bioZ}}$ model performance is comparable to that of the PSG$_{\text{thor. belt}}$ model.

there is a statistically significant difference between ROBIN$_{\text{bioZ}}$ and PSG$_{\text{abd. belt}}$ but that no statistically significant difference between ROBIN$_{\text{bioZ}}$ and PSG$_{\text{thor. belt}}$ can be proven. Statistical significance of using the two-phased approach is analyzed by comparing the L1 and L2 models for ROBIN$_{\text{bioZ}}$ which results in a p-value of 0.03, demonstrating a statistically significant improvement in performance.

To accurately interpret these results, a detailed analysis of the model errors is required. The various types of model error are shown in Fig. 10. When the respiratory disturbance of the apnea event is of short duration (type A error), the model will return a false positive result as only apnea events longer than 10 seconds are scored in polysomnography annotations [53]. In practical settings, type A errors can be reduced by only taking into account model activations that are longer than 10 seconds. Sometimes,

TABLE III

OVERVIEW OF ALL RECORDED MEASURES, EXPRESSED AS PERCENTAGES, FOR ALL THREE SIGNALS, EVALUATED ON THE SEPARATE TEST DATASET. THE PSG$_{\text{ABD. BELT}}$ MEASURED USING THE GOLD STANDARD PSG DEVICE RESULTS IN THE BEST OVERALL PERFORMANCE. THE COMPACT WEARABLE ROBIN$_{\text{BIOZ.}}$ IS ALSO ABLE TO ACCURATELY DETECT SLEEP APNEA EVENTS IN RESPIRATORY DATA AND HAS A PERFORMANCE COMPARABLE TO PSG$_{\text{THOR. BELT}}$. THE BINARY CLASSIFICATION METRICS ARE REPORTED AT THE DECISION THRESHOLD $p_{\text{EVENT}} > 0.5$ WHICH WAS LEARNED DURING TRAINING

	PSG$_{\text{abd. belt}}$		**PSG$_{\text{thor. belt}}$**		**ROBIN$_{\text{bioZ.}}$**	
	L1	**L2**	**L1**	**L2**	**L1**	**L2**
AUPRC	53.7 ± 7.8	56.7 ± 7.4	45.8 ± 8.3	47.8 ± 9.2	41.6 ± 11.2	46.9 ± 8.5
AUROC	80.3 ± 2.9	80.1 ± 2.2	74.1 ± 4.3	74.0 ± 4.7	72.5 ± 7.5	73.6 ± 3.7
sensitivity	71.5 ± 7.9	66.5 ± 8.6	64.5 ± 8.2	56.4 ± 8.6	67.5 ± 8.9	58.4 ± 12.7
specificity	74.5 ± 4.4	79.0 ± 6.7	70.0 ± 5.1	77.5 ± 5.0	66.2 ± 11.4	76.2 ± 8.7
precision	45.1 ± 6.3	48.6 ± 4.1	38.7 ± 8.7	42.5 ± 9.7	38.1 ± 8.9	42.2 ± 6.6
accuracy	74.1 ± 2.4	76.6 ± 3.8	68.5 ± 3.4	72.3 ± 3.3	66.8 ± 7.8	72.8 ± 4.4
NPV	90.1 ± 3.2	89.3 ± 1.8	87.1 ± 3.9	85.7 ± 4.6	87.9 ± 2.4	86.7 ± 2.2

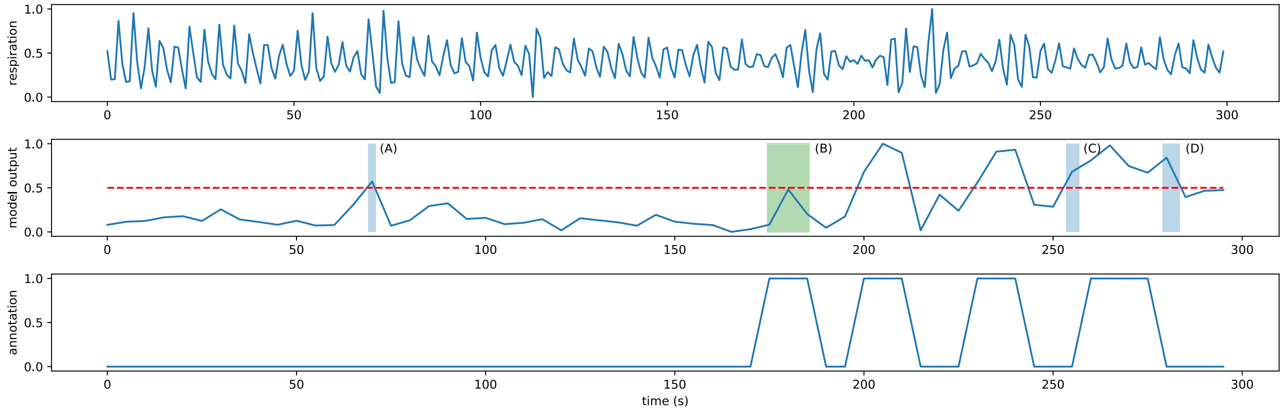


Fig. 10. Example of the model prediction output for a given respiration sequence. The input signal is the ROBIN$_{\text{bioZ}}$. The red dotted line indicates the decision boundary on which the model was trained ($p_{\text{event}} > 0.5$). The model output mimics the annotations of the trained sleep technicians which are based on the full PSG. However, there are some erroneous outputs as indicated by A, B, C and D. (A) false-positive event due to short-term respiratory disturbance, (B) false negative event due to insufficient activation of model output, (c) false positive event due to early detection, (D) false positive event due to overshoot.

TABLE IV

OVERVIEW OF CLASSIFICATION ACCURACY FOR THE DIFFERENT SLEEP APNEA TYPES AND DIFFERENT TESTED MODELS

	PSG$_{\text{abd. belt}}$	**PSG$_{\text{thor. belt}}$**	**ROBIN$_{\text{bioZ.}}$**
obstructive	80.7 ± 3.4	72.5 ± 4.7	74.8 ± 1.8
central	94.2 ± 1.2	88.5 ± 2.4	64.6 ± 8.7
hypopnea	59.2 ± 2.3	51.6 ± 2.9	56.3 ± 2.1
non-event	80.8 ± 2.1	76.8 ± 1.7	77.5 ± 1.5

the model output does not activate sufficiently (type B error). This happens when the event is not clear in the respiratory data, for example when the breathing has been shallow for a longer period than the model window. The annotations in the polysomnography data can not be interpreted as an exact boundary. However, the analysis metrics require an exact event cut-off time. This leads to false positive errors due to early detection (type C error) or due to the model overshooting the annotation boundary (type D error). Although type C and type D errors negatively impact the model scores, they are not relevant in practical settings as the sleep apnea event was correctly detected and hence such errors do not influence the diagnostic AHI score.

Finally, the performance of the models across the different types of apnea is analyzed by computing the accuracy of the models for each type separately. These results are shown in Table IV. For all models, OSA classification has the best performance.

VI. DISCUSSION

The aim of this study was to thoroughly analyze and investigate the use of bioZ as a respiratory surrogate for the automated detection of sleep apnea events. The results show that this approach can reach a similar performance as an automated model based on respiratory data from the PSG. In addition, the model performance is stable across the various folds of the dataset, demonstrating generalizability. The main advantage of the bioZ compared to the conventional PSG methods is the smaller form factor and increase in patient comfort. In addition, the use of the bioZ does not require a respiration derivation algorithm as is necessary for ECG derived respiration. These derivation algorithms struggle when patients suffer additional conditions such as arrhythmia. The main disadvantage of the bioZ is the reduced accuracy for the detection of CSA events as shown in Table IV.

Other methods for HSAT respiration measurements have also been proposed. Table V provides an overview of the recorded metrics for some reference works. None of these studies reports the AUPRC metric. The results of these devices and algorithms

TABLE V
COMPARISON OF METRICS FOR DIFFERENT HSAT DEVICES REPORTED IN LITERATURE. NONE OF THE STUDIES REPORT THE AUPRC METRIC

Study	sensitivity	specificity	accuracy	AUROC
[18]	77	94	/	/
[19]	72.9	90.6	85.5	/
[20]	/	/	72.1	/
[38]	100	85.9	/	/
[22]	/	/	81.8	/
[36]	92.4	88.3	/	/
[60]	90	96	/	/
[16]	80	54.5	/	/
[17]	/	/	/	71

need to be carefully interpreted. In some cases, the presented metrics are based on a per-patient classification whereas in others these are presented as per-epoch scores. In addition, the studies are performed with different patients, different setups and different scoring and analysis criteria. There are several works in which these devices and algorithms are compared with each other and analyzed [54]–[59]. To fully compare the bioZ approach with these devices, a large further clinical study is required.

There are several limitations to this study which warrant careful interpretation of the results. First, the patients that were included in the study had, in general, low AHI scores. A more relevant set of patients would include several high AHI scores. However, as this study includes patients that come to the sleep clinic for testing, the AHI of the patients to be included cannot be determined upfront. The low AHI represents a challenge for the algorithms as the data imbalance is further enlarged, impacting the metrics. Second, the presented device is not compatible with AASM guidelines as several other physiological signals need to be measured as well. Oxygen desaturations and arousals linked to apnea events are important for relevant insights. In this initial study, the aim was to analyze and discuss the use of bioZ as a respiratory surrogate. Further research should focus on combining this bioZ signal with the other required signals. The next generation of the ROBIN device will be equipped with PPG functionality. As significant oxygen desaturation mostly occurs in long ($\geq$ 10 seconds) apnea events, the amount of type A errors would reduce. Third, there are other possible respiratory disturbances, such as movements or coughs, which are not necessarily sleep apnea and which can impact the results. However, the model is trained to detect sleep apnea patterns in the respiration by using the clinical annotations provided by the nurse. As the analysis is also based on these annotations, the performance for detecting such sleep apnea events is measured.

This study confirms that the use of bioZ is a promising option for automated sleep apnea detection using a comfortable wearable and that it could potentially serve as a basis for future home monitoring devices.

VII. CONCLUSION

In this work, the use of bioZ as a respiratory surrogate for detecting sleep apnea was analysed and discussed. A novel method was proposed consisting of a wearable device, combined with a deep learning algorithm. The compact wearable is capable of capturing a respiration surrogate by use of a direct bio impedance measurement of the chest. The recorded data is then processed by an extended two-phase deep learning algorithm to automatically detect the sleep apnea events. The results demonstrate that the performance of the setup is comparable to automated detection using gold-standard PSG respiration data. This setup offers a powerful basis for the future development of portable home sleep apnea monitors. In future work, this approach should be extended with additional signals such as oxygen saturation to improve the robustness of the detection.

REFERENCES

[1] C. Guilleminault, A. Tilkian, and W. C. Dement, “The sleep apnea syndromes,” *Ann. Rev. Med.*, vol. 27, no. 1, pp. 465–484, 1976.

[2] H. K. Yaggi, J. Concato, W. N. Kernan, J. H. Lichtman, L. M. Brass, and V. Mohsenin, “Obstructive sleep apnea as a risk factor for stroke and death,” *New England J. Med.*, vol. 353, no. 19, pp. 2034–2041, 2005.

[3] V. K. Somers *et al.*, “Sleep apnea and cardiovascular disease,” *Circulation*, vol. 118, no. 10, pp. 1080–1111, 2008.

[4] J. S. Floras, “Sleep apnea and cardiovascular disease: An enigmatic risk factor,” *Circulation Res.*, vol. 122, no. 12, pp. 1741–1764, 2018.

[5] A. Sassani, L. J. Findley, M. Kryger, E. Goldlust, C. George, and T. M. Davidson, “Reducing motor-vehicle collisions, costs, and fatalities by treating obstructive sleep apnea syndrome,” *Sleep*, vol. 27, no. 3, pp. 453–458, 2004.

[6] A. Benjafield *et al.*, “An estimate of the global prevalence and burden of obstructive sleep apnoea,” *The Lancet*, vol. 7, no. 8, pp. 687–698, Aug. 2019.

[7] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. L. Marcus, and B. V. Vaughn, “The AASM manual for the scoring of sleep and associated events,” in *Rules, Terminology and Technical Specifications*. Darien, Illinois: American Academy of Sleep Medicine, 2012.

[8] R. D. Kim *et al.*, “An economic evaluation of home versus laboratory-based diagnosis of obstructive sleep apnea,” *Sleep*, vol. 38, no. 7, pp. 1027–1037, 2015.

[9] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, “Access to diagnosis and treatment of patients with suspected sleep apnea,” *Amer. J. Respiratory Crit. Care Med.*, vol. 169, no. 6, pp. 668–672, 2004.

[10] S. Su, F. M. Baroody, M. Kohrman, and D. Suskind, “A comparison of polysomnography and a portable home sleep study in the diagnosis of obstructive sleep apnea syndrome,” *Otolaryngology—Head Neck Surgery*, vol. 131, no. 6, pp. 844–850, 2004.

[11] M. Madani, M. Frank, R. Lloyd, D. I. Dimitrova, and F. Madani, “Polysomnography versus home sleep study: Overview and clinical application,” *Atlas Oral Maxillofacial Surgery Clin. North Amer.*, vol. 15, no. 2, pp. 101–109, 2007.

[12] I. C. Jeong, D. Bychkov, and P. Searson, “Wearable devices for precision medicine and health state monitoring,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1242–1258, May 2019.

[13] J. Dunn, R. Runge, and M. Snyder, “Wearables and the medical revolution,” *Personalized Medicine*, vol. 15, no. 5, pp. 429–448, 2018.

[14] 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, 2007.

[15] M. Hug *et al.*, “Screening for obstructive sleep apnea among hospital outpatients,” *PloS One*, vol. 13, no. 5, 2018, Art. no. e0198315.

[16] S. Z. Assefa, M. Diaz-Abad, A. Korotinsky, S. E. Tom, and S. M. Scharf, “Comparison of a simple obstructive sleep apnea screening device with standard in-laboratory polysomnography,” *Sleep Breathing*, vol. 20, no. 2, pp. 537–541, 2016.

[17] J.-I. Byun, S. J. Song, H.-K. Cha, and W. C. Shin, “Reliability of manual and automatic scoring of single channel nasal airflow device (apnealink) in determining moderate or severe obstructive sleep apnea syndrome,” *Sleep Med. Res.*, vol. 7, no. 2, pp. 68–73, 2016.

[18] Z. T. Beattie, C. C. Hagen, M. Pavel, and T. L. Hayes, “Classification of breathing events using load cells under the bed,” in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 3921–3924.

[19] S. H. Hwang *et al.*, “Unconstrained sleep apnea monitoring using polyvinylidene fluoride film-based sensor,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2125–2134, Jul. 2014.

[20] D. Waltisberg, O. Amft, D. P. Brunner, and G. Tröster, “Detecting disordered breathing and limb movement using in-bed force sensors,” *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 930–938, Jul. 2017.

[21] J. Jin and E. Sánchez-Sinencio, “A home sleep apnea screening device with time-domain signal processing and autonomous scoring capability,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 1, pp. 96–104, Feb. 2015.

[22] Y.-Y. Lin, H.-T. Wu, C.-A. Hsu, P.-C. Huang, Y.-H. Huang, and Y.-L. Lo, “Sleep apnea detection based on thoracic and abdominal movement signals of wearable piezoelectric bands,” *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1533–1545, Nov. 2017.

[23] T. Penzel, J. McNames, P. De Chazal, B. Raymond, A. Murray, and G. Moody, “Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings,” *Med. Biol. Eng. Comput.*, vol. 40, no. 4, pp. 402–407, 2002.

[24] G. B. Moody *et al.*, “Clinical validation of the ECG-derived respiration (EDR) technique,” *Comput. Cardiol.*, vol. 13, no. 1, pp. 507–510, 1986.

[25] T. Van Steenkiste *et al.*, “Systematic comparison of respiratory signals for the automated detection of sleep apnea,” in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 449–452.

[26] J.-C. Vázquez *et al.*, “Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea,” *Thorax*, vol. 55, no. 4, pp. 302–307, 2000.

[27] N. Netzer, A. H. Eliasson, C. Netzer, and D. A. Kristo, “Overnight pulse oximetry for sleep-disordered breathing in adults: A review,” *Chest*, vol. 120, no. 2, pp. 625–633, 2001.

[28] T. Gumb *et al.*, “Comparison of two home sleep testing devices with different strategies for diagnosis of OSA,” *Sleep Breathing*, vol. 22, no. 1, pp. 139–147, 2018.

[29] G. C. Gutiérrez-Tobal, D. Álvarez, A. Crespo, F. del Campo, and R. Hornero, “Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 882–892, Mar. 2019.

[30] H. Alshaer, G. R. Fernie, E. Maki, and T. D. Bradley, “Validation of an automated algorithm for detecting apneas and hypopneas by acoustic analysis of breath sounds,” *Sleep Med.*, vol. 14, no. 6, pp. 562–571, 2013.

[31] D. J. Levendowski, B. Veljkovic, S. Seagraves, and P. R. Westbrook, “Capability of a neck worn device to measure sleep/wake, airway position, and differentiate benign snoring from obstructive sleep apnea,” *J. Clin. Monitoring Comput.*, vol. 29, no. 1, pp. 53–64, 2015.

[32] J. Kim, “0493 a predictive model of sleep disordered breathing based on deep neural network,” *Sleep*, vol. 41, no. Suppl. 1, pp. A185–A186, Apr. 2018.

[33] R. S. Puri *et al.*, “Design and preliminary evaluation of a wearable device for mass-screening of sleep apnea,” in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 1870–1873.

[34] J. Durán-Cantolla *et al.*, “Validation of a new domiciliary diagnosis device for automatic diagnosis of patients with clinical suspicion of OSA,” *Respirology*, vol. 22, no. 2, pp. 378–385, 2017.

[35] N. A. Collop *et al.*, “Obstructive sleep apnea devices for out-of-center (OOC) testing: Technology evaluation,” *J. Clin. Sleep Med.*, vol. 7, no. 5, pp. 531–548, 2011.

[36] J. Han, H.-B. Shin, D.-U. Jeong, and K. S. Park, “Detection of apneic events from single channel nasal airflow using 2nd derivative method,” *Comput. Methods Programs BioMed.*, vol. 91, no. 3, pp. 199–207, 2008.

[37] 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.

[38] P. Várady, T. Micsik, S. Benedek, and Z. Benyó, “A novel method for the detection of apnea and hypopnea events in respiration signals,” *IEEE Trans. Biomed. Eng.*, vol. 49, no. 9, pp. 936–942, Sep. 2002.

[39] T. Van Steenkiste, W. Groenendaal, D. Deschrijver, and T. Dhaene, “Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2354–2364, Nov. 2019.

[40] U. J. Magalang *et al.*, “Home sleep apnea testing: Comparison of manual and automated scoring across international sleep centers,” *Sleep Breathing*, vol. 23, no. 1, pp. 25–31, 2019.

[41] O. G. Martinsen and S. Grimnes, *Bioimpedance and Bioelectricity Basics*. New York, NY, USA: Academic, 2011.

[42] D. Blanco-Almazán, W. Groenendaal, F. Catthoor, and R. Jané, “Wearable bioimpedance measurement for respiratory monitoring during inspiratory loading,” *IEEE Access*, vol. 7, pp. 89 487–89 496, 2019.

[43] N. Van Helleputte *et al.*, “A 345 μW multi-sensor biomedical SOC with bio-impedance, 3-channel ECG, motion artifact reduction, and integrated DSP,” *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 230–244, Jan. 2015.

[44] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[45] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, “Learning to diagnose with lstm recurrent neural networks,” 2015, *arXiv:1511.03677*.

[46] S. Biswal *et al.*, “SLEEPNET: Automated sleep staging system via deep learning,” 2017, *arXiv:1707.08262*.

[47] A. Sathyanarayana, J. Srivastava, and L. Fernandez-Luque, “The science of Sweet Dreams: Predicting sleep efficiency from wearable device data,” *Comput.*, vol. 50, no. 3, pp. 30–38, 2017.

[48] T. Van Steenkiste *et al.*, “Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks,” *Artif. Intell. Med.*, 2018.

[49] B. C. Wallace, K. Small, C. E. Brodley, and T. A. Trikalinos, “Class imbalance, redux,” in *Proc. IEEE 11th Int. Conf. Data Mining*, 2011, pp. 754–763.

[50] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient global optimization of expensive black-box functions,” *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.

[51] J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian optimization of machine learning algorithms,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2951–2959.

[52] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[53] C. A. Robles-Rubio, K. A. Brown, and R. E. Kearney, “Optimal classification of respiratory patterns from manual analyses using expectation-maximization,” *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1026–1035, Jul. 2018.

[54] W. W. Flemons *et al.*, “Home diagnosis of sleep apnea: A systematic review of the literature: An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society,” *CHEST J.*, vol. 124, no. 4, pp. 1543–1579, 2003.

[55] J. M. Kelly, R. E. Strecker, and M. T. Bianchi, “Recent developments in home sleep-monitoring devices,” *ISRN Neurol.*, vol. 2012, pp. 768–794, 2012.

[56] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, “Devices for home detection of obstructive sleep apnea: A review,” *Sleep Med. Rev.*, vol. 41, pp. 149–160, 2018.

[57] R. N. Aurora, R. Swartz, and N. M. Punjabi, “Misclassification of OSA severity with automated scoring of home sleep recordings,” *Chest*, vol. 147, no. 3, pp. 719–727, 2015.

[58] N. Scalzitti, S. Hansen, S. Maturo, J. Lospinoso, and P. O’Connor, “Comparison of home sleep apnea testing versus laboratory polysomnography for the diagnosis of obstructive sleep apnea in children,” *Int. J. Pediatric Otorhinolaryngol.*, vol. 100, pp. 44–51, 2017.

[59] J. Arguelles *et al.*, “Comparing three home sleep apnea testing devices to polysomnography: Evaluating diagnostic rates for obstructive sleep apnea,” *Sleep*, vol. 42, no. Suppl.1, Apr. 2019, Art. no. A183.

[60] M. Ciołek, M. Niedźwiecki, S. Sieklicki, J. Drozdowski, and J. Siebert, “Automated detection of sleep apnea and hypopnea events based on robust airflow envelope tracking in the presence of breathing artifacts,” *IEEE J. Biomed. Health Inform.*, vol. 19, no. 2, pp. 418–429, Mar. 2015.