

Analysis and Classification of Oximetry Recordings to Predict Obstructive Sleep Apnea Severity in Children

Gonzalo C. Gutiérrez-Tobal, *Student Member, IEEE*, Leila Kheirandish-Gozal, Daniel Álvarez, *Member, IEEE*, Andrea Crespo, Mona F. Philby, Meelad Mohammadi, Félix del Campo, David Gozal, and Roberto Hornero, *Senior Member, IEEE*

Abstract— Current study is focused around the potential use of oximetry to determine the obstructive sleep apnea-hypopnea syndrome (OSAHS) severity in children. Single-channel SpO₂ recordings from 176 children were divided into three severity groups according to the apnea-hypopnea index (AHI): AHI < 1 events per hour (e/h), 1 ≤ AHI < 5 e/h, and AHI ≥ 5 e/h. Spectral analysis was conducted to define and characterize a frequency band of interest in SpO₂. Then we combined the spectral data with the 3% oxygen desaturation index (ODI₃) by means of a multi-layer perceptron (MLP) neural network, in order to classify children into one of the three OSAHS severity groups. Following our MLP multiclass approach, a diagnostic protocol with capability to reduce the need of polysomnography tests by 46% could be derived. Moreover, our proposal can be also evaluated, in a binary classification task for two common AHI diagnostic cutoffs (AHI = 1 e/h and AHI = 5 e/h). High diagnostic ability was reached in both cases (84.7% and 85.8% accuracy, respectively) outperforming the clinical variable ODI₃ as well as other measures reported in recent studies. These results suggest that the information contained in SpO₂ could be helpful in pediatric OSAHS severity detection.

I. INTRODUCTION

Pediatric obstructive sleep apnea-hypopnea syndrome (OSAHS) has emerged as a frequent and concerning medical condition in the past 2-3 decades. It is characterized by an abnormal breathing pattern during sleep that includes the recurrence of apneas (complete airflow cessation) and hypopneas (airflow limitation), caused by total or partial upper airway obstruction, respectively [1]. Inadequate gas exchange characterized by repetitive hypoxia, hypercapnia, and accompanied by arousal episodes during the night has been suggested as the cause for serious comorbidities related to central nervous system and cardiovascular and metabolic system [2]. Consequently, several daytime symptoms related to OSAHS, such as cognitive and behavioral irregularities as well as atypical growth are frequently present and reported by parents [3, 4]. Furthermore, the prevalence of OSAHS in

children is high, with studies reporting up to 5.7% among general pediatric population [4].

The “gold standard” approach to diagnose OSAHS in children is overnight polysomnography (PSG) [4]. However, PSG has several limitations since it is both complex and costly due to the high number of physiological signals that need to be recorded [5]. Additionally, all these signals need offline inspection in order to derive the apnea-hypopnea index (AHI), which is used to establish whether OSAHS is present and its severity. Consequently, PSG is also time-consuming [6]. Moreover, children often do not tolerate well the equipment involved in PSG [7].

These limitations have led to the search for simpler diagnostic alternatives. Thus, one common approach has been the study of the diagnostic ability of reduced sets of signals derived from those involved in PSG, such as electrocardiography (ECG), photoplethysmography (PPG), airflow (AF), or blood oxygen saturation (SpO₂) [8-11]. Particularly, frequency and time domain analyses of ECG-derived signals showed utility in pediatric OSAHS diagnosis [8]. Moreover, the analysis of pulse transit time variability from PPG was successfully used to classify time segments into apneic or non-apneic, as well as children into normal subjects and OSAHS patients [9]. Additionally, a recent study from our research group reported high diagnostic ability when combining the oxygen desaturation index (ODI) from SpO₂ with spectral information from AF [10]. Finally, spectral, nonlinear, and statistical features from SpO₂ and pulse rate variability (PRV) recordings were obtained and successfully combined to establish OSAHS in children [11].

In this study, however, we focus on the use of the information contained in a single-channel SpO₂ to help in OSAHS severity detection. Our hypothesis is that exclusive utilization of data from the SpO₂ channel could simplify the OSAHS diagnosis and severity assessment in children. Hence, the main objective of this preliminary work is to evaluate the diagnostic ability of the information contained in the SpO₂ signal. Specifically, we analyze the spectrum of SpO₂ recordings from children, divided into three groups according to their corresponding AHI. There exists a lack of consistency in the literature as to the optimal AHI cutoff to determine OSAHS in children, with most of the studies applying 1, 3, or 5 events per hour (e/h) [4]. We chose AHI < 1 e/h as the most restrictive cutoff to discard OSAHS and AHI ≥ 5 e/h to define a group with the highest OSAHS severity. Additionally, we formed another group with those patients in the range 1 ≤ AHI < 5 e/h, which is recognized as the most challenging concerning the decision to implement treatment, usually consisting of surgical removal of tonsils

*Research supported by the project VA059U13 (Consejería de Educación de la Junta de Castilla y León). G. C. Gutiérrez-Tobal was in receipt of a PIRTU grant (Consejería de Educación de la Junta de Castilla y León and the European Social Fund).

G. C. Gutiérrez-Tobal, D. Álvarez, and R. Hornero, are with the Biomedical Engineering Group (University of Valladolid, Spain) (e-mail: Gonzalo.gutierrez@gib.tel.uva.es)

L. Kheirandish-Gozal, M. F. Philby, M. Mohammadi, and D. Gozal are with the Section of Sleep Medicine, Department of Pediatrics, Pritzker School of Medicine, Biological Sciences Division (The University of Chicago, Chicago, IL, USA) (e-mail: dgozal@peds.bsd.uchicago.edu).

A. Crespo and F. del Campo are with the sleep unit of Hospital Universitario Río Hortega (Valladolid, Spain) (e-mail: fsas@telefonica.net).

and adenoids [2]. Therefore, we here evaluate the spectrum of the SpO₂ recordings from children in the three groups looking for discriminative features. Additionally, we use 3% ODI (ODI₃) for comparison purposes. Finally, we combine the spectral information and ODI₃ by means of an artificial neural network, a multi-layer perceptron (MLP) [12], in order to classify children into one of the three groups. This multiclass approach lets us define a protocol which includes doubtful subjects, as well as allows the evaluation, at the same time, of both AHI=1 e/h and AHI=5 e/h cutoffs from a binary classification point of view.

II. SUBJECTS AND SIGNALS UNDER STUDY

The study involved SpO₂ recordings from 176 children (97 males and 79 females). All of them were clinically referred to the Pediatric Sleep Unit at the University of Chicago Medicine Comer Children's Hospital (Chicago, IL, USA) due to suspicion of OSAHS. The Ethical Committee approved the protocol and an informed consent to participate in the study was obtained for each child. Overnight PSGs were conducted from 20:00 to 08:00. Recordings were acquired by means of a digital polysomnography system (Polysmith; Nihon Kohden America Inc., CA, USA). Detection and quantification of sleep and cardiorespiratory events were carried out according to the rules of the American Academy of Sleep Medicine [13]. Thus, apnea was defined as the absence of oronasal airflow during at least 2 respiratory cycles. Accordingly, hypopnea was defined as a decrease $\geq 30\%$ in the nasal pressure airflow signal lasting at least 2 respiratory cycles, leading to a desaturation $\geq 3\%$ and/or an arousal [13]. As previously stated, children were divided into three groups according to their corresponding AHI: AHI under 1 e/h (AHI_{<1}), AHI in the range [1, 5) e/h (AHI_{[1,5)}), and AHI equal or above 5 e/h (AHI _{≥ 5}). Table I summarizes demographic and clinical data from subjects according to this division. No statistical significant differences (p -value <0.01) were found in age, gender, and body mass index (BMI) when applying the non-parametric Kruskal-Wallis test to compare the three groups.

The SpO₂ recordings were acquired during PSG at a sampling rate of $f_s = 25$ Hz. Artifacts due to children movements were automatically removed during preprocessing. Thus, SpO₂ values equal to zero as well as differences between consecutive SpO₂ samples $\geq 4\%$ were considered artifacts [14]. Removed samples were substituted by interpolated data. ODI₃ was estimated as the number of desaturations (at least 3%) per hour of sleep time.

III. METHODOLOGY

The methodology was divided into three steps. First, a spectral analysis of the SpO₂ recordings was conducted to look for differences among the three groups. Then we extracted several spectral features according to this analysis. Finally, we combined the spectral data and ODI₃ through MLP to classify the children into one of the three classes.

A. Spectral analysis and feature extraction

Power spectral density (PSD) was estimated for each SpO₂ recording by means of the Welch's method [15]. We used a Hamming window of 2^{13} samples (5.5 minutes), 50% overlap, and a discrete Fourier transform of 2^{14} samples. Fig. 1 shows the median PSD for each group of OSAHS severity.

TABLE I. DEMOGRAPHIC AND CLINICAL DATA

	All	AHI _{<1}	AHI _{[1,5)}	AHI _{≥ 5}
# Subjects	176	30	75	71
Age [*] (years)	7.0 \pm 3.6	8.2 \pm 3.3	7.3 \pm 3.5	6.1 \pm 3.6
Male (%)	55.1	56.7	54.7	54.9
BMI [*] (kg/m ²)	20.6 \pm 7.3	20.5 \pm 6.8	20.6 \pm 6.7	20.7 \pm 8.2
AHI (e/h)	-	0.5 \pm 0.3	2.6 \pm 1.1	19.3 \pm 23.1

BMI: Body Mass Index; AHI: Apnea Hypopnea Index; ^{*} p -value=0.016; ^{*} p -value=0.816

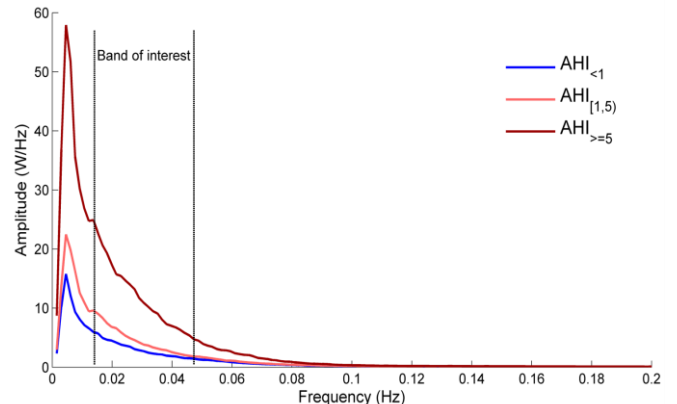


Figure 1. Median PSD for each of the OSAHS severity groups

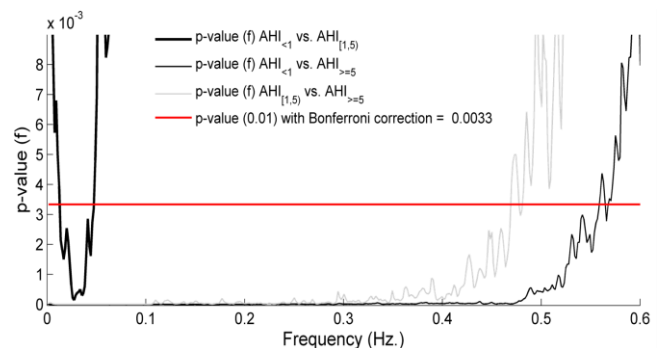


Figure 2. p -value vs. frequency for each of the three possible comparisons between groups.

Higher PSDs can be observed as the severity increases. A band of interest (BW) is also shown in the range 0.0137-0.0473 Hz. This corresponds to the spectral bandwidth in which the three groups showed statistical significant differences (Mann-Whitney U test) in their PSD amplitude values (p -value = 0.01, p -value = 0.0033 after Bonferroni correction). In this case, BW is equivalent to the bandwidth in which AHI_{<1} and AHI_{[1,5)} showed significant differences. Fig. 2 displays the p -value vs. frequency plots for each of the three possible comparisons. The limits of BW are easily located as the crosspoints between the AHI_{<1} vs. AHI_{[1,5)} p -value (f) curve and the p -value significance level line (in red).

We extracted the following features from the BW of each PSD: maximum PSD value (MA), minimum PSD value (mA), spectral power (P_S , as the area under the PSD at BW), and standard deviation of the PSD values (SD_f). According to Fig. 1, higher values were expected in these features as the OSAHS severity increases. After feature extraction, each subject under study is characterized by a vector \mathbf{x}_i ($i = 1, 2, \dots$

M , $M = 176$) whose 5 components are the corresponding values of the four spectral features and ODI_3 .

B. Multi-layer perceptron

MLP is a supervised learning algorithm whose architecture is arranged in several interconnected layers (input, hidden, and output). These are composed of units known as neurons or perceptrons [12]. Each neuron is characterized by an activation function $g(\cdot)$ and their connections to neurons from other layers ($w_{i,j}$). In our case, the input layer has five units, corresponding to the number of spectral features obtained for each subject (MA , mA , P_S , SD_f) and ODI_3 . Moreover, since the purpose is to carry out a three-class classification, three output units with a logistic activation function were used. We also implemented a single hidden layer, composed of neurons with non-linear activation functions. This configuration is known to be able to provide a universal function approximation [12]. Since the number of neurons in the hidden layer (N_H) controls the effective complexity of the network [12], we chose a small number, $N_H = 5$, to prevent network from overfitting. Thus, the final input-layer: hidden-layer: output-layer architecture was 5:5:3 neurons. The weights $w_{i,j}$ were optimized using the sum of squares error function minimization criterion by means of the scaled conjugate gradient algorithm [12]. For each subject under study, the final classification task was performed by assigning the corresponding \mathbf{x}_i ($i = 1, 2 \dots M$, $M = 176$) to the class with the highest probability in the output layer.

C. Statistical analysis

The non-parametric Kruskal-Wallis test was used to assess statistical differences in the spectral features from the OSAHS severity groups. A confusion matrix was used to evaluate the performance of multiclass MLP. Also, to assess the output of MLP from a binary classification point of view, sensitivity (Se, percentage of OSAHS-positive subjects rightly classified), specificity (Sp, percentage of OSAHS-negative subjects rightly classified), accuracy (Acc, overall percentage of subjects rightly classified), positive predictive value (PPV, proportion of positive test results which are true positives), negative predictive value (NPV, proportion of negative test results which are true negatives), positive likelihood ration (LR+, $Se/(1-Sp)$), and negative likelihood ratio (LR-, $(1-Se)/Sp$) measured the diagnostic ability for both $AHI=1$ e/h and $AHI=5$ e/h cutoffs. All these statistics were obtained after leave-one-out cross-validation (loo-cv).

IV. RESULTS

Table II displays the values of the spectral features and ODI_3 for each of the three OSAHS severity groups (mean \pm standard deviation). All of them showed large statistical significant differences when comparing the three groups by means of Kruskal-Wallis test. As expected, the five features are higher as the OSAHS severity increases.

Table III shows the confusion matrix resulting from the diagnostic ability assessment of the MLP network for the three-class classification task (results after loo-cv). A total of 125 out of 176 subjects were rightly classified in their actual class (71.0%). Per classes, 50.0% (15 out of 30) of the subjects in $AHI_{<1}$, 80.0% (60 out of 75) in $AHI_{1,5}$, and 70.4% (50 out of 71) in $AHI_{\geq 5}$ were rightly classified.

TABLE II. VALUES OF THE SPECTRAL FEATURES AND ODI_3 (MEAN \pm STANDARD DEVIATION)

Features	$AHI_{<1}$	$AHI_{1,5}$	$AHI_{\geq 5}$	p -value
ODI_3 (e/h)	1.01 \pm 1.10	3.21 \pm 2.80	16.18 \pm 20.78	$<10^{-17}$
MA (W/Hz)	7.76 \pm 7.41	13.58 \pm 16.98	71.61 \pm 149.83	$<10^{-12}$
mA (W/Hz)	1.44 \pm 0.59	2.91 \pm 5.23	19.51 \pm 70.43	$<10^{-11}$
P_S (W) (10^{-1})	7.80 \pm 4.32	15.15 \pm 21.02	97.74 \pm 248.19	$<10^{-13}$
SD_f (W/Hz)	1.84 \pm 1.81	3.23 \pm 4.12	16.31 \pm 27.29	$<10^{-11}$

TABLE III. CONFUSION MATRIX FOR THE MLP MULTICLASS TASK (AFTER LOO-CV)

		estimated		
		$AHI_{<1}$	$AHI_{1,5}$	$AHI_{\geq 5}$
actual	$AHI_{<1}$	15	15	0
	$AHI_{1,5}$	11	60	4
	$AHI_{\geq 5}$	1	20	50

TABLE IV. DIAGNOSTIC ABILITY OF MLP AND ODI_3 (BINARY CLASSIFICATION AFTER LOO-CV)

	Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	LR+	LR-
ODI_3 (AHI=1)	78.1	80.0	78.4	95.0	48.9	3.91	0.27
ODI_3 (AHI=5)	69.0	81.9	76.7	79.6	72.1	3.81	0.38
MLP (AHI=1)	91.8	50.0	84.7	89.9	55.6	1.84	0.16
MLP (AHI=5)	70.4	96.2	85.8	92.6	82.8	18.5	0.31

Table IV shows the diagnostic ability of MLP and ODI_3 when assessing both the $AHI = 1$ e/h and $AHI = 5$ e/h cutoffs (results after loo-cv). MLP results are directly derived from the confusion matrix. For both cutoffs the global Acc of MLP is higher than the corresponding ODI_3 (84.7% vs. 78.4% and 85.8% vs. 76.7%, respectively). In the case of $AHI=1$ e/h, ODI_3 is much more specific than MLP, leading to higher PPV and LR+. In the case of $AHI=5$ e/h, however, MLP outperforms ODI_3 at each statistic.

V. DISCUSSION AND CONCLUSIONS

In this work, we have developed an automatic diagnostic methodology for pediatric OSAHS severity based on the information contained in single-channel SpO_2 . Features from a spectral band of interest and the clinical variable ODI_3 were combined by means of MLP to classify subjects into one out of the three OSAHS severity levels.

The spectral analysis of the SpO_2 signal revealed a band of interest (BW=0.0137-0.0473 Hz) in which statistically significant differences were found for the three classes. The lower limit of BW is consistent with the corresponding band of interest in adults (0.014-0.033 Hz, i.e., events lasting from 30 to 71 s) [16]. Conversely, a higher upper limit was found in children, suggesting shorter events as also significant for them. This agrees with the higher respiratory rate reported in children [17]. However, further analysis is required regarding the causes of the differences in both bands.

The spectral features extracted from BW showed statistically significant differences when comparing the three

classes. All of them reached higher values as the OSAHS severity increases. Since the ideal SpO₂ time series is a constant, close to 100%, the higher PSD values in the frequencies correspond to more desaturations and recoveries to the baseline. Consequently, higher MA , mA , P_s , and SD_f suggest more desaturation events both in discrete frequencies (MA , mA) and in the whole band (P_s , SD_f), which is consistent with the clinically used severity classification of OSAHS.

Our multiclass MLP proposal correctly classified 71% of the subjects. Although this overall accuracy is not high enough, a deeper study of the subjects wrongly classified reveals that the 11 children who belong to $AHI_{[1,5]}$, and were assigned to $AHI_{<1}$, present an AHI of 1.65 ± 0.42 e/h. This means that 96.3% of subjects predicted as $AHI_{<1}$ have no OSAHS or a low severity degree. Additionally, the 4 children from $AHI_{[1,5]}$ assigned to $AHI_{\geq 5}$ present an AHI of 3.0 ± 1.7 e/h, i.e., 100% of children predicted as $AHI_{\geq 5}$ have severe OSAHS or a higher severity degree comparing with the mean of the $AHI_{[1,5]}$ class. Finally, children assigned to $AHI_{[1,5]}$ come from the three classes: $AHI_{<1}$ (15.8%), $AHI_{[1,5]}$ (63.2%), and $AHI_{\geq 5}$ (21.0%). Consequently, subjects assigned to this class should be regarded as inconclusive. A screening protocol could be generated from these results as follows: *i*) if MLP predicts $AHI_{<1}$, discard OSAHS; *ii*) if MLP predicts $AHI_{\geq 5}$, consider treatment; *iii*) if MLP predicts $AHI_{[1,5]}$, send to overnight PSG. Since the SpO₂ signal is easily acquired from an oximeter, such a protocol would reduce the need by 46% (81/176) of overnight PSGs.

Other studies analyzed physiological signals to help in pediatric OSAHS diagnosis. All of them reported results from a binary classification point of view. Shouldice *et al.* analyzed 50 ECG recordings, reaching 85.7% Se, 81.8% Sp, and 84.0% Acc using a quadratic linear discriminant applied to 23 features (AHI cutoff = 1 e/h) [8]. Gil *et al.* assessed the diagnostic ability of information contained in 21 PPG time series, reporting 75.0 % Se, 85.7% Sp, and 80.0% Acc (AHI cutoff = 5 e/h) [9]. Gutiérrez-Tobal *et al.* combined spectral features from 50 AF recordings with ODI₃ from SpO₂ to achieve 85.9% Se, 87.4% Sp, and 86.3% Acc with a logistic regression methodology (AHI cutoff = 3 e/h) [10]. Finally, Garde *et al.* reported 83.6 % Se, 88.4% Sp, and 85.0% Acc in a 146 subject database by combining 8 features from SpO₂ and PRV in a linear discriminant [11]. Our MLP methodology can be assessed for AHI = 1 e/h and AHI = 5 e/h at the same time (84.7% and 85.8% Acc, respectively). Only Gutiérrez-Tobal *et al.* reported higher Acc. However, they used two channels and the AHI cutoff was 3 e/h.

This work has some limitations that are worthy of mention. First, although the number of subjects is not small when comparing to other similar studies, more children, particularly those with AHI < 1 e/h, would be necessary for the sake of a proper MLP training. This would include a training-test strategy as well as the evaluation of a range of neurons in the hidden layer, which was arbitrary set to a low value in order to decrease the chances for overfitting.

Additionally, more subjects would also let us use a training set from which we could independently obtain the spectral band of interest. However, a loo-cv methodology was used to validate our results. Finally, the use of features from time domain could complement our findings. One future goal is the assessment of features and classification models other than those presented in this work using a larger dataset.

In summary, we have developed a multiclass MLP methodology with capability to help in pediatric OSAHS severity screening. The SpO₂ features obtained from a frequency band of interest, combined with ODI₃ through MLP, outperform the single diagnostic yield of this clinical variable. Our proposal can be also evaluated for binary classification purposes, reaching high diagnostic ability comparing with recent state-of-the-art studies. Thus, our results suggest that the information contained in single-channel SpO₂ is helpful to detect severity categories among children with OSAHS that are worthy of mention.

REFERENCES

- [1] C. Guilleminault, et al., "A review of 50 children with obstructive sleep apnea syndrome," *Lung*, vol. 159(1), pp. 275-287, 1981.
- [2] O. S. Capdevila, et al., "Pediatric obstructive sleep apnea: complications, management, and long-term outcomes," *Proceedings of the American Thoracic Society*, vol. 5(2), pp. 274-282, 2008.
- [3] C. Guilleminault, et al., "Pediatric obstructive sleep apnea syndrome," *Arch Pediatr Adolesc Med*, vol. 159, pp. 775-85, 2005.
- [4] C. L. Marcus, et al., "Diagnosis and management of childhood obstructive sleep apnea syndrome," *Pediatrics*, vol. 130(3), pp.e714-e755, 2012.
- [5] H. Reuveni, et al., "Health care services utilization in children with obstructive sleep apnea syndrome," *Pediatrics*, vol. 110, pp. 68-72, 2012.
- [6] J. A. Bennet and W. J. M. Kinnear, "Sleep on the cheap: the role of overnight oximetry in the diagnosis of sleep apnoea hypopnoea syndrome," *Thorax* vol. 54, pp. 958-959, 1999.
- [7] E. S. Katz, et al., "Obstructive sleep apnea in infants," *Am J Respir Crit Care Med*, vol. 185(8), pp. 805-816, 2012.
- [8] R. E. Shouldice, et al., "Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features," *Sleep*, vol. 27(4), pp. 784-92, 2004.
- [9] E. Gil, et al., "PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children," *IEEE Trans Biomed Eng*, vol. 57(5), pp. 1079-1088, 2010.
- [10] G. C. Gutiérrez-Tobal, et al. "Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home." *Biomed Signal Process Control*, vol. 18, pp. 401-407, 2015.
- [11] A. Garde, et al., "Development of a Screening Tool for Sleep Disordered Breathing in Children Using the Phone Oximeter™," *PLoS one*, vol. 9(11), pp. e112959, 2014.
- [12] C. M. Bishop, *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press, 1996.
- [13] R. B. Berry, et al., "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events," *J Clin Sleep Med*, vol. 8(5), pp. 597-619, 2012.
- [14] U. J. Magalang, et al., "Prediction of the apnea-hypopnea index from overnight pulse oximetry," *Chest*, vol. 124, pp. 1694-1701, 2003.
- [15] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on time Averaging Over Short, Modified Periodograms," *IEEE Trans on Audio and Electroacoustics*, vol. AU-15, pp. 70-73, 1967.
- [16] C. Zamarrón, et al., "Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea," *Clin Sci*, vol. 97, pp. 467-73, 1999.
- [17] S. Fleming, et al., "Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies," *The Lancet*, vol. 377, pp. 1011-1018, 2011.