

# Statistical and Nonlinear Analysis of Oximetry from Respiratory Polygraphy to Assist in the Diagnosis of Sleep Apnea in Children

Daniel Álvarez, *IEEE Member*, Gonzalo C. Gutiérrez-Tobal, *IEEE Student Member*, María L. Alonso, Joaquín Terán, Félix del Campo, and Roberto Hornero, *IEEE Senior Member*

**Abstract**— Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS) is a sleep related breathing disorder that has important consequences in the health and development of infants and young children. To enhance the early detection of OSAHS, we propose a methodology based on automated analysis of nocturnal blood oxygen saturation ( $SpO_2$ ) from respiratory polygraphy (RP) at home. A database composed of 50  $SpO_2$  recordings was analyzed. Three signal processing stages were carried out: (i) feature extraction, where statistical features and nonlinear measures were computed and combined with conventional oximetric indexes, (ii) feature selection using genetic algorithms (GAs), and (iii) feature classification through logistic regression (LR). Leave-one-out cross-validation (loo-cv) was applied to assess diagnostic performance. The proposed method reached 80.8% sensitivity, 79.2% specificity, 80.0% accuracy and 0.93 area under the ROC curve (AROC), which improved the performance of single conventional indexes. Our results suggest that automated analysis of  $SpO_2$  recordings from at-home RP provides essential and complementary information to assist in OSAHS diagnosis in children.

## I. INTRODUCTION

Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS) is characterized by recurrent episodes of partial or complete collapse of the upper airway during sleep [1]. Untreated OSAHS leads to several negative consequences in the health and development of infants and young children, such as neuropsychological and cognitive deficits, cardiovascular dysfunction, and/or growth impairment [2]. A recent report of the American Academy of Pediatrics suggests a prevalence of OSAHS in the range of 1% to 5% [2].

The gold standard test for assessing children with suspected OSAHS is in-hospital overnight polysomnography (PSG) [3, 2], in which children's sleep is supervised and monitored by the use of multiple sensors. Therefore, PSG is costly due to the need of expensive equipment and trained

This work has been partially supported by the Ministerio de Economía y Competitividad and FEDER under project TEC2011-22987 and by Project Cero 2011 on Ageing from Fundación General CSIC, Obra Social La Caixa and CSIC and a grant by Consejería de Educación de la Junta de Castilla y León under project VA059U13. Gonzalo C. Gutiérrez-Tobal was in receipt of a PIRTU grant from the Consejería de Educación de la Junta de Castilla y León and the European Social Fund.

D. Álvarez, G. C. Gutiérrez-Tobal, and R. Hornero, are with the Biomedical Engineering Group, E.T.S.I. Telecomunicación, Universidad de Valladolid, Paseo de Belén 15, 47011, Valladolid, Spain (phone: +34 983 423000, ext. 4716, fax: +34 983 423667; e-mails: dalygon@ribera.tel.uva.es, gcuttob@ribera.tel.uva.es, robhor@tel.uva.es).

F. del Campo is with the Hospital Universitario Río Hortega of Valladolid, Spain (e-mail: fsas@telefonica.net).

M. L. Alonso and J. Terán are with the University Hospital of Burgos (Spain) (mlalonso@hubu.es, jteran@hubu.es).

staff, and its availability is limited, resulting in long waiting lists [3, 4]. Furthermore, the use of multiple sensors makes PSG highly intrusive and limits the effectiveness of this methodology, often leading to poor results when used on young children and infants [5].

There is an increasing research on novel methodologies in the context of sleep apnea diagnosis in children, including history and physical examination, respiratory polygraphy (RP), daytime (nap) PSG, and ambulatory PSG [2, 6, 7]. The American Academy of Pediatrics reported that these methods tend to be helpful if patients test positive but have a poor predictive value if results are negative [2]. Therefore, further research is needed. In this regard, automated signal processing could improve the diagnostic performance of screening tests for OSAHS detection in children. Electrocardiogram (ECG) [8, 9], photoplethysmography (PPG) [10, 11] and blood oxygen saturation ( $SpO_2$ ) [1, 3, 4, 12, 13] are commonly used in this context.

In this study,  $SpO_2$  recordings from at-home RP were analyzed. We used  $SpO_2$  due to its reliability, simplicity and suitability for children. Previous studies in the context of OSAHS diagnosis in children by means of  $SpO_2$  assessed conventional oximetric indexes [1, 3, 4, 12, 13], common statistics [4, 12] and conventional spectral features [12]. In the present research, first to fourth statistical moments and three nonlinear methods were applied: mean ( $M1$ ), variance ( $M2$ ), skewness ( $M3$ ), and kurtosis ( $M4$ ), as well as nonlinear measures of irregularity, variability and complexity by means of sample entropy (*SampEn*), central tendency measure (*CTM*) and Lempel-Ziv complexity (*LZC*), respectively. These methods previously achieved high performance in the context of OSAHS diagnosis in adults [14-16]. We hypothesized that these measures could provide useful and complementary information to conventional oximetric indexes in children. Genetic algorithms (GAs) and logistic regression (LR) are proposed for feature selection and classification. LR is a suitable alternative to the reference linear discriminant analysis (LDA) in binary classification problems [14, 15] but its performance has been weakly assessed as a diagnostic tool for OSAHS in children [1]. Similarly, GAs are optimization methods that could improve performance in the subsequent classification stage [14]. However, few studies applied feature selection [12].

## II. SUBJECTS AND SIGNALS

A total of 50 children (23 boys and 27 girls) were included in the study. All children were suspected of suffering from OSAHS and derived to the Sleep Breathing

Disorders Unit of the Hospital Universitario de Burgos (Spain). Informed consents to participate in the study were obtained and the Ethical Committee approved the protocol.

In-hospital attended PSGs were carried out from 22:00 to 08:00. Children were continuously monitored using a polysomnograph Deltamed Coherence® 3NT version 3.0 (Diagniscan, SA, ACH-Werfen Company, Paris, France). As stated by the American Academy of Sleep Medicine, apnea was defined as the absence of oronasal airflow lasting at least 2 respiratory cycles, whereas hypopnea was defined as a decrease in airflow greater than or equal to 50% for more than 2 respiratory cycles, accompanied by a desaturation greater than or equal to 3% and/or an EEG arousal. An apnea hypopnea index (AHI)  $\geq 3$  events per hour (e/h) from PSG was considered OSAHS-positive [17]. A positive diagnosis was confirmed in 26 children (52%). The remaining 24 subjects (48%) composed the OSAHS-negative group. Table I displays demographic and clinical features for both groups.

At-home 6-channel RP was carried out previously to in-hospital PSG using a polygraph eXim Apnea Polygraph (Bitmed®, Sibel S.A., Barcelona, Spain). SpO<sub>2</sub> recordings from RP (sampling frequency  $f_s = 100$  Hz) were saved to separate files and processed offline. A preprocessing stage was applied to remove artifacts linked with bad contact with the finger probe due to children movements. Fig. 1 depicts representative SpO<sub>2</sub> recordings in our dataset.

### III. METHODOLOGY

#### A. Feature Extraction

Oximetric recordings were parameterized by means of 13 features. Three complementary feature subsets were composed: first to fourth order statistical moments, nonlinear measures and common oximetric indexes. Statistical moments and nonlinear measures were computed offline using epochs of  $N=6000$  samples (1 min) and subsequently averaging to obtain a single value per recording.

1) *Statistical moments*. First to fourth order statistical moments ( $M1-M4$ ) were applied to characterize the data histogram in the time domain. Mean ( $M1$ ), variance ( $M2$ ), skewness ( $M3$ ), and kurtosis ( $M4$ ) were computed to estimate central tendency, dispersion, symmetry and peakedness, respectively [18].

2) *Nonlinear features*. *SampEn*, *CTM* and *LZC* were applied to quantify irregularity, variability and complexity, respectively [14-16, 19]:

- *SampEn*( $m, r, N$ ) quantifies irregularity in time series, with larger values corresponding to more irregular data. It is defined as the negative logarithm of the conditional probability that two sequences that are similar (within the tolerance width  $r$ ) for  $m$  contiguous points remain similar when the run length increases to  $m+1$  [14]. In this study, the widely applied  $m = 1$  and  $r=0.25$  times SD were used [14-16].
- *CTM*( $r$ ) provides a variability measure from second order difference plots. It is computed selecting a

TABLE I. DEMOGRAPHIC AND CLINICAL FEATURES OF THE POPULATION SET

	All Children	OSAHS-negative	OSAHS-positive	<i>p</i> -value
<b>Subjects (n)</b>	50	24	26	–
<b>Age (years)</b>	5.30 $\pm$ 2.55	5.17 $\pm$ 2.44	5.42 $\pm$ 2.69	NS
<b>Males (n)</b>	23 (46%)	13 (54%)	10 (38%)	NS
<b>BMI (kg/m<sup>2</sup>)</b>	16.51 $\pm$ 2.47	16.13 $\pm$ 1.74	16.87 $\pm$ 2.99	NS
<b>Records (h)</b>	8.86 $\pm$ 0.83	8.99 $\pm$ 0.48	8.74 $\pm$ 1.05	NS
<b>AHI (e/h)</b>		1.30 $\pm$ 0.79	17.89 $\pm$ 15.41	< 0.01

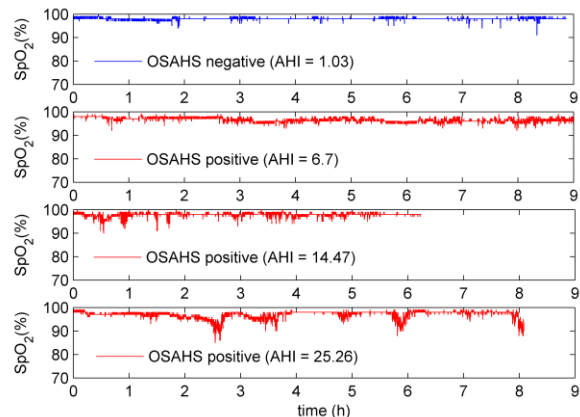


Figure 1. Representative SpO<sub>2</sub> recordings of children in the population under study showing different OSAHS severity.

circular region of radius  $r$  around the origin, counting the number of points that fall inside, and dividing by the total number of points [19]. In this study, the optimum radius  $r=1$  in the context of OSAHS diagnosis from SpO<sub>2</sub> were used [14-16].

- *LZC* is a nonparametric measure of complexity linked with the rate of new subsequences and their repetition. To compute *LZC*, each recording is first converted into a binary sequence [19].

3) *Conventional oximetric indexes*. Average saturation during the recording time (*Avg*), global minimum saturation (*Min*), oxygen desaturation index of 3% (*ODI3*) and cumulative time spent below a saturation of 85% (*CT85*), 90% (*CT90*) and 95% (*CT95*) were provided by the polygraph and included in the study.

#### C. Feature Selection

GAs are optimization algorithms commonly used to exhaustively inspect the search space of variables that govern a model [20]. A particular group of solutions (parents) are selected from an initial population to generate the offspring by means of crossover and mutation operations. The offspring replaces parents in the current population based on a replacement strategy. The optimization process is carried out in cycles called generations [20]. In this study, GAs were applied to obtain the optimum input feature subset to a LR classifier in terms of classification performance. Roulette strategy was used for parent selection, one-point crossover ( $P_c=0.5$  to 0.9) was applied to produce offspring and uniform mutation ( $P_m = 0.01$  to 0.09) was applied to introduce variations into the offspring [20, 21]. The

percentage of the best individuals preserved after each generation were varied between 0 and 25% [15].

### C. Feature Classification

LR relates a categorical dependent variable with a set of independent explanatory variables. For binary classification problems, input patterns are classified into one of two mutually exclusive classes and the probability density for the response variable can be modeled by a Bernoulli distribution [18]. LR classifiers assign an input vector to the class with the maximum a posteriori probability value. The maximum likelihood criterion is used to optimize coefficients of the independent input features in the logistic model [18].

### D. Statistical Analysis

The non-parametric Mann-Whitney U test was applied to search for statistical significant differences. A  $p$ -value < 0.01 was considered significant. Matlab R2012a (7.14.0.739) was used to implement feature extraction, selection and classification stages. Diagnostic performance was assessed by means of sensitivity (true positive rate), specificity (true negative rate), positive predictive value (proportion of positive tests that are true positive patients), negative predictive value (proportion of negative tests that are true negative subjects), accuracy (percentage of children correctly classified) and area under the receiver operating characteristics (ROC) curve (AROC). Leave-one-out cross-validation (loo-cv) was applied to properly assess our results.

## IV. RESULTS

Fig. 2 shows the envelopes of the average histograms of SpO<sub>2</sub> recordings for the whole OSAHS-negative and OSAHS-positive groups. As can be observed, OSAHS-positive children show lower mean, higher dispersion (width) and lower symmetry (skewed to the left) and peakedness (height) than OSAHS-negative children. This agrees with the effects of recurrent apneas: desaturations due to respiratory events lead to lower values in the SpO<sub>2</sub> overnight profile.

Table II summarizes the diagnostic performance of every single variable under study after a loo-cv procedure. Accuracy ranges from 56.0% to 76.0% in the statistical moments feature subset, from 58.0% to 68.0% in the nonlinear feature subset, and from 56.0% to 70.0% in the conventional oximetric indexes subset. *M2* achieved the highest accuracy in the whole dataset, reaching 76.0% sensitivity, 75.0% specificity, 76.0% accuracy and 0.81 AROC. Table III summarizes the performance of optimum LR models from GAs using loo-cv. Feature selection was applied looking for the highest accuracy of a LR classifier with different number of input explanatory variables from the initial feature set from oximetry. After applying GAs, 5 LR models achieved a maximum accuracy of 90.0% in the whole dataset. Input feature subsets are listed in Table III. Next, LR models were tested using loo-cv. Models with 6 (*M1*, *CTM*, *LZC*, *Min*, *ODI3*, *CT95*) and 7 (*M1*, *CTM*, *LZC*, *Avg*, *Min*, *ODI3*, *CT95*) variables achieved the highest performance: 80.8% sensitivity, 79.2% specificity, 80.0% accuracy and 0.93 AROC.

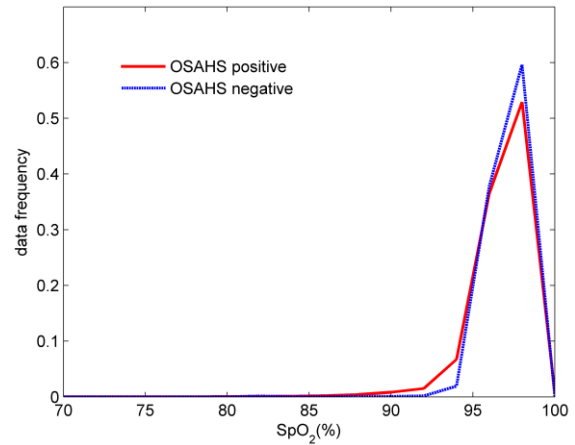


Figure 2. Envelopes of the average histograms for SpO<sub>2</sub> data from the whole OSAHS-negative and OSAHS-positive groups.

TABLE II. DIAGNOSTIC ASSESSMENT OF EACH SINGLE FEATURE UNDER STUDY BY MEANS OF LOO-CV

Feature	TP	TN	FP	FN	Se	Sp	PPV	NPV	Acc	AROC
<i>M1</i>	15	15	9	11	57.7	62.5	62.5	57.7	60.0	0.61
<i>M2</i>	20	18	6	6	76.9	75.0	76.9	75.0	76.0	0.81
<i>M3</i>	14	14	10	12	53.8	58.3	58.3	53.8	56.0	0.53
<i>M4</i>	17	16	8	9	65.4	66.7	68.0	64.0	66.0	0.68
<i>SampEn</i>	17	17	7	9	65.4	70.8	70.8	65.4	68.0	0.77
<i>CTM</i>	17	16	8	9	65.4	66.7	68.0	64.0	66.0	0.77
<i>LZC</i>	16	13	11	10	61.5	54.2	59.3	56.5	58.0	0.70
<i>Avg</i>	17	17	7	9	65.4	70.8	70.8	65.4	68.0	0.85
<i>Min</i>	20	15	9	6	76.9	62.5	69.0	71.4	70.0	0.70
<i>ODI3</i>	10	20	4	16	38.5	83.3	71.4	55.6	60.0	0.60
<i>CT85</i>	5	23	1	21	19.2	95.8	83.3	52.3	56.0	0.57
<i>CT90</i>	14	19	5	12	53.8	79.2	73.7	61.3	66.0	0.67
<i>CT95</i>	18	16	8	8	69.2	66.7	69.2	66.7	68.0	0.75

TP: true positives (n); TN: true negatives (n); FP: false positives (n); FN: false negatives (n); Se: sensitivity (%); Sp: specificity (%); PPV: positive predictive value (%); NPV: negative predictive value (%); Acc: accuracy (%); AROC: area under the ROC curve

## V. DISCUSSION AND CONCLUSIONS

In this study, statistical and nonlinear measures from at-home RP-derived SpO<sub>2</sub> recordings were combined with conventional oximetric indexes to help in OSAHS diagnosis in children. Automated feature extraction, selection (GAs) and classification (LR) were applied. The assessment of every single feature did not reach substantially high diagnostic performance. Nevertheless, the joint analysis of properly selected features by means of GAs achieved significantly higher accuracy. Optimum LR models showed a balanced number of input variables from both approaches: in the LR model with 6 input variables there were 3 features from the automated signal processing stage and 3 from conventional oximetric indexes, whereas in the LR model with 7 variables *Avg* was just added to the previous model. Moreover, all three feature subsets, i.e. statistical, nonlinear and conventional indexes, were represented in all LR models from feature selection. *M1*, *LZC*, *Min* and *ODI3* were selected in all models from GAs, suggesting these features enclose essential information (average/basal saturation, degree of changes, depth/severity of desaturations and rate of

TABLE III. DIAGNOSTIC ASSESSMENT OF OPTIMUM LR MODELS FROM GAS BY MEANS OF LOO-CV

N	Variables in the LR model	TP	TN	FP	FN	Se	Sp	PPV	NPV	Acc	AROC
6	MI, CTM, LZC, Min, ODI3, CT95	21	19	5	5	80.8	79.2	80.8	79.2	80.0	0.93
7	MI, SampEn, LZC, Avg, Min, ODI3, CT90	20	18	6	6	76.9	75.0	76.9	75.0	76.0	0.92
7	MI, M4, CTM, LZC, Min, ODI3, CT95	21	18	6	5	80.8	75.0	77.8	78.3	78.0	0.93
7	MI, CTM, LZC, Avg, Min, ODI3, CT95	21	19	5	5	80.8	79.2	80.8	79.2	80.0	0.93
8	MI, M4, CTM, LZC, Avg, Min, ODI3, CT95	21	18	6	5	80.8	75.0	77.8	78.3	78.0	0.93

N: number of independent explanatory variables in the LR model

desaturations, respectively) on changes in SpO<sub>2</sub> recordings due to apneic events typical of OSAHS in children.

Our results agreed with recent studies focusing on the SpO<sub>2</sub> signal. The proposed methodology outperformed results from Chang *et al.* [1] and Cohen & de Chazal [4], where symptoms and conventional oximetric indexes were used as input variables to LR (60.0% Se, 86.0% Sp) and LDA (55.9% Se, 68.6% Sp) classifiers, respectively. Therefore, statistical and nonlinear measures proposed in our study seem to provide additional information to conventional measures. The study by Tsai *et al.* [13] reached sensitivity-specificity pairs of 77.7-88.9, 83.8-86.5 and 89.1-86.0 when specific groups of mild, moderate and severe OSAHS were analyzed, respectively. Nevertheless, no cross-validation was applied. Finally, the study by Garde *et al.* [12] achieved 80.0% sensitivity and 92.1% specificity using time, spectral and oximetric features as input variables to a LDA classifier with feature selection. Although a portable device was used, all recordings were attended and carried out in the hospital facilities. In the present research, children really benefited from the advantages of unattended at-home RP (simplicity and suitability).

Some limitations should be taken into account. First, the dataset should be larger. Second, some input parameters were set using information from sleep apnea studies in adults. Thus, they should be optimized using SpO<sub>2</sub> recordings from children.

In summary, our results suggested that statistical and nonlinear features provide useful and complementary information to conventional oximetric indexes, improving the performance of at-home SpO<sub>2</sub> in the context of OSAHS diagnosis in children.

#### REFERENCES

- [1] L. Chang, J. Wu, and L. Cao, "Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea," *Int. J. Pediatr. Otorhinolaryngol.*, vol. 77, pp. 365–371, 2013.
- [2] C. L. Marcus, L. J. Brooks, S. D. Ward, *et al.*, "Diagnosis and Management of Childhood Obstructive Sleep Apnea Syndrome," *Pediatrics*, vol. 130, pp. e714–e755, 2012.
- [3] V. G. Kirk, S. G. Bohn, W. W. Flemons, and J. E. Remmers, "Comparison of home oximetry monitoring with laboratory polysomnography in children," *Chest*, vol. 124, pp. 1702–1708, 2003.
- [4] G. Cohen and P. de Chazal, "Automated detection of sleep Apnea in infants using minimally invasive Sensors," in *Proc. 35th Annu. Int. Conf. IEEE-EMBS*, Osaka, 2013, pp. 1652–1655.
- [5] E. S. Katz, B. M. Ron, and C. M. D'Ambrosio, "Obstructive Sleep Apnea in Infants," *Am. J. Respir. Crit. Care Med.*, vol. 185, pp. 805–816, 2012.
- [6] P. J. P. Poels, A. G. M. Schilder, S. van den Berg, A. W. Hoes, and K. F. M. Joosten, "Evaluation of a new device for home cardiorespiratory recording in children," *Arch. Otolaryngol. Head Neck Surg.*, vol. 129, pp. 1281–1284, 2003.
- [7] M. L. Alonso, J. Terán, J. A. Cordero, A. I. Navazo, E. Ordax, J. F. Masa, and R. Pelayo, "Reliability of respiratory polygraphy for the diagnosis of sleep apnea-hypopnea syndrome in children," *Arch. Bronconeumol.*, vol. 44, pp. 318–323, 2008.
- [8] Y. Pu, R. P. Patterson, M. A. CramerBornemann, "Nocturnal cardiorespiratory indices - A novel screening tool for pediatric obstructive sleep disordered breathing," in *Proc. 27th Annu. Int. Conf. IEEE-EMBS*, Shanghai, 2005, pp. 2575–2578.
- [9] Z. D. Deng, C-S. Poon, N. M. Arzeno, and E. S. Katz, "Heart rate variability in pediatric obstructive sleep apnea," in *Proc. 28th Annu. Int. Conf. IEEE-EMBS*, New York, 2006, pp. 3565–3568.
- [10] E. Gil, R. Bailon, J. M. Vergara, and P. Laguna, "PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 1079–1088, 2010.
- [11] J. Lázaro, E. Gil, J. M. Vergara, and P. Laguna, "Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children," *IEEE Biomed. Health Inf.*, vol. 18, pp. 240–246, 2014.
- [12] A. Garde, W. Karlen, P. Dehkordi, D. Wensley, J. M. Ansermino, and G. A. Dumont, "Oxygen saturation in children with and without obstructive sleep apnea using the phone-oximeter," in *Proc. 35th Annu. Int. Conf. IEEE-EMBS*, Osaka, 2013, pp. 2531–2534.
- [13] C-M. Tsai, C-H. Kang, M.-C. Su, *et al.*, "Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children," *Int. J. Pediatr. Otorhinolaryngol.*, vol. 77, pp. 1286–1290, 2013.
- [14] D. Álvarez, R. Hornero, J. V. Marcos, and F. del Campo, "Multivariate Analysis of Blood Oxygen Saturation Recordings in Obstructive Sleep Apnea Diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 2816–2824, 2010.
- [15] D. Álvarez, R. Hornero, J. V. Marcos, and F. del Campo, "Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis," *Med. Eng. Phys.*, vol. 34, pp. 1049–1057, 2012.
- [16] D. Alvarez, R. Hornero, J. V. Marcos, N. Wessel, T. Penzel, M. Glos, and Félix del Campo, "Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of sleep apnea diagnosis," *Int. J. Neural Syst.*, vol. 23, pp. 1–18, 2013.
- [17] M. L. Alonso-Álvarez, T. Canet, M. Cubel- Alarco, E. Estivill, E. Fernandez-Julian, D. Gozal, *et al.*, "Consensus document on sleep apnea-hypopnea syndrome in children," *Arch Bronconeumol*, vol. 47 (Suppl 5), pp. 1–18, 2011.
- [18] J. D. Jobson, *Applied multivariate data analysis*. New York, Springer-Verlag, New York, 1991.
- [19] D. Álvarez, R. Hornero, D. Abásolo, F. del Campo, and C. Zamarrón, "Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection," *Physiol. Meas.*, vol. 27, pp. 399–412, 2006.
- [20] K. S. Tang, K. F. Man, S. Kwong, and Q. He, "Genetic algorithms and their applications," *IEEE Signal Proc. Mag.*, vol. 13 pp. 22–37, 1996.
- [21] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large scale feature selection," *Pattern Recogn. Lett.*, vol. 10, pp. 335–347 1989.