Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

Detrended fluctuation analysis of the oximetry signal to assist in paediatric sleep apnoea–hypopnoea syndrome diagnosis

To cite this article: Fernando Vaquerizo-Villar *et al* 2018 *Physiol. Meas.* **39** 114006

View the article online for updates and enhancements.

This content was downloaded by fervaquerizo from IP address 157.88.129.84 on 14/11/2018 at 12:16

IOP Publishing Physiol. Meas. 39 (2018) 114006 (13pp) https://doi.org/10.1088/1361-6579/aae66a

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

Detrended fluctuation analysis of the oximetry signal to assist in paediatric sleep apnoea–hypopnoea syndrome diagnosis

Fernando Vaquerizo-Villar[1,5], Daniel Álvarez[1,2], Leila Kheirandish-Gozal[3], Gonzalo C Gutiérrez-Tobal[1], Verónica Barroso-García[1], Andrea Crespo[1,2], Félix del Campo[1,2], David Gozal[3] and Roberto Hornero[1,4]

[1] Biomedical Engineering Group, Universidad de Valladolid, Valladolid, Spain
[2] Servicio de Neumología, Hospital Universitario Rio Hortega, Valladolid, Spain
[3] Department of Child Health, The University of Missouri School of Medicine, Columbia, MO, United States of America
[4] Instituto de Investigación en Matemáticas, Universidad de Valladolid, Valladolid, Spain
[5] Author to whom any correspondence should be addressed.

E-mail: fervaquerizo@gmail.com

RECEIVED 20 July 2018

REVISED 20 September 2018

ACCEPTED FOR PUBLICATION 5 October 2018

PUBLISHED 13 November 2018

Keywords: blood oxygen saturation (SpO_2), detrended fluctuation analysis (DFA), feature selection, apnoea–hypopnoea index (AHI) estimation, oximetry, paediatric sleep apnoea–hypopnoea syndrome (SAHS)

Abstract

Objective: To evaluate whether detrended fluctuation analysis (DFA) provides information that improves the diagnostic ability of the oximetry signal in the diagnosis of paediatric sleep apnoea–hypopnoea syndrome (SAHS). *Approach*: A database composed of 981 blood oxygen saturation (SpO_2) recordings in children was used to extract DFA-derived features in order to quantify the scaling behaviour and the fluctuations of the SpO_2 signal. The 3% oxygen desaturation index (*ODI*3) was also computed for each subject. Fast correlation-based filter (FCBF) was then applied to select an optimum subset of relevant and non-redundant features. This subset fed a multi-layer perceptron (MLP) neural network to estimate the apnoea–hypopnoea index (AHI). *Main results*: *ODI*3 and four features from the DFA reached significant differences associated with the severity of SAHS. An optimum subset composed of the slope in the first scaling region of the DFA profile and the *ODI*3 was selected using FCBF applied to the training set (60% of samples). The MLP model trained with this feature subset showed good agreement with the actual AHI, reaching an intra-class correlation coefficient of 0.891 in the test set (40% of samples). Furthermore, the estimated AHI showed high diagnostic ability, reaching an accuracy of 82.7%, 81.9%, and 91.1% using three common AHI cut-offs of 1, 5, and 10 events per hour (e h^{-1}), respectively. These results outperformed the overall performance of *ODI*3. *Significance*: DFA may serve as a reliable tool to improve the diagnostic performance of oximetry recordings in the evaluation of paediatric patients with symptoms suggestive of SAHS.

1. Introduction

Childhood sleep apnoea–hypopnoea syndrome (SAHS) is a breathing disorder whereby paediatric subjects manifest recurrent episodes of either complete cessation (apnoea) or significant reductions (hypopnoea) of airflow while sleeping (Marcus *et al* 2012). Paediatric SAHS has become a major health problem due to its high prevalence and negative effects. SAHS has an estimated prevalence in the range of 1%–5% in the general paediatric population (Marcus *et al* 2012). In addition, cognitive deficits, behavioural abnormalities, daytime sleepiness, cardiac and metabolic derangements, and systemic inflammation are all morbid consequences that adversely affect the optimal development of children affected by SAHS (Marcus *et al* 2012).

Based on the aforementioned considerations, an early diagnosis of paediatric SAHS is vital. The gold standard diagnostic approach to childhood SAHS is overnight polysomnography (PSG) (Marcus *et al* 2012). It requires patients to spend the night in a specialised sleep laboratory while being recorded for a wide range of biomedical signals, including electrocardiogram, electroencephalogram, electrooculogram, submental and leg electromyo-

© 2018 Institute of Physics and Engineering in Medicine

gram, oronasal airflow, and blood oxygen saturation (SpO_2) by pulse oximetry (Alonso-Álvarez *et al* 2011, Kaditis *et al* 2016). However, PSG is a complex text which is also quite costly due to the necessary equipment and specialized medical personnel that is required to supervise the PSG and to score the recordings. PSG is also intrusive, especially for children, due to the use of multiple sensors. Additionally, PSG shows limited availability in many, if not most, places around the world, which results in long waiting lists, thus delaying the diagnosis and treatment of the affected children (Nixon *et al* 2004, Katz *et al* 2012).

Considering the inherent disadvantages and limitations of PSG, along with the need for an early and timely diagnosis of SAHS, the search for simplified alternative techniques has emerged in recent years. In this regard, one common approach consists of the automated analysis of a reduced subset of cardiorespiratory signals that is normally included in the overnight PSG. One of these alternatives is nocturnal pulse oximetry (NPO), which records the blood oxygen saturation signal (SpO_2) with a pulse oximeter probe, usually placed on a finger (Netzar *et al* 2001). NPO can be readily performed without the need for professional supervision in the patient's home and is widely available, as reflected by the large number of commercially available portable pulse oximeters (Nixon *et al* 2004, Garde *et al* 2014). Thus, NPO is a technically simple test for children, and the SpO_2 signal from NPO provides moment-to-moment oxygen content in haemoglobin (McClatchey 2002), a signal that contains essential information about the apnoeic events from SAHS, since these events induce recurrent decreases in blood oxygen levels, otherwise termed oxygen desaturations (Berry *et al* 2012).

Previous studies have examined the SpO_2 signal as a potential alternative to PSG in the screening of paediatric SAHS. These studies employed different signal processing techniques (Kirk *et al* 2003, Tsai *et al* 2013, Garde *et al* 2014, Eyck *et al* 2015, Álvarez *et al* 2017, Crespo *et al* 2017, Hornero *et al* 2017, Vaquerizo-Villar *et al* 2018), and more specifically, conventional oximetry indices, common statistics, frequency domain analysis techniques, and nonlinear methods. Among these approaches, nonlinear parameters proved useful to characterise the oxygen desaturations caused by apnoeic events in adults and children. However, a recent study using a very large database of 4191 paediatric subject recordings showed that traditional nonlinear metrics (central tendency measure, Lempel–Ziv complexity and sample entropy) were redundant with respect to the 3% oxygen desaturation index (*ODI*3) (Hornero *et al* 2017), an oximetry index commonly used in the clinical practice for simplified screening purposes. Therefore, additional research is needed to find alternative and better performant nonlinear methods that may provide further insights into the properties of the oximetry signal and allow for the extraction of additional information to that provided by *ODI*3. In this regard, detrended fluctuation analysis (DFA) is a nonlinear analysis technique widely used to detect the correlation properties of a non-stationary signal (Peng *et al* 1994, 1995). DFA computes the logarithm of the fluctuation function of a time series versus the logarithm of a window time length (scale). DFA provides a quantitative parameter, the scaling exponent (α), which measures the linear relationship between the fluctuation function and the scale (Peng *et al* 1994). The variation of α value for different ranges of scales (different window time lengths) identifies regions with different correlations (Peng *et al* 1995). In this sense, the scaling behaviour of a signal is given by the different regions observed in the DFA profile and the value of α in these regions (Peng *et al* 1995). Thus, DFA is a useful tool to analyse signals with segments that modify its scaling behaviour, such as random spikes or segments which have a different local behaviour (Chen *et al* 2002, Hua and Yu 2017). Apnoeic events produce random spikes and/or irregular fluctuations in the SpO_2 signal. Hence, DFA could be useful to analyse the oximetry signal in the context of SAHS.

Previous work has suggested the ability of DFA to analyse the correlation properties of physiological signals in the context of both adult and paediatric SAHS (Lee *et al* 2002, Penzel *et al* 2003, Dehkordi *et al* 2016, Kaimakamis *et al* 2016, Hua and Yu 2017). Hua and Yu (2017) applied DFA to SpO_2 signals in the context of diagnosing adult SAHS. However, no studies have focused on applying DFA to SpO_2 recordings in the context of paediatric SAHS. SpO_2 signal properties in children differ from those of adults. Furthermore, the frequency of events that are required to define abnormality or severity markedly differ between adults and children. In addition, scoring rules for apnoeas and hypopnoeas are also more restrictive in the case of paediatric SAHS (Berry *et al* 2012). Thus, the diagnosis of SAHS in children is vastly more challenging than in adults.

Thus, we hypothesised that DFA could extract additional information from the oximetry signal, which could be associated with the presence and severity of SAHS in children and could therefore assist in the diagnostic accuracy of overnight oximetry. Accordingly, the aim of this study was to assess the usefulness of DFA-derived features obtained from the oximetry signal to simplify the diagnosis of paediatric SAHS.

2. Methods

2.1. Subjects and signals under study

The dataset included 981 children (602 boys and 379 girls) ranging from 2–13 years of age. All children were consecutively and prospectively referred to the Pediatric Sleep Unit at the University of Chicago Medicine Comer Children's Hospital (Chicago, IL, USA) due to clinical suspicion of SAHS. Their legal caretakers gave their informed consent as a prerequisite to participate in the study. The Ethical Committee of the University

of Chicago Medicine approved the research protocols (#11-0268-AM017, # 09-115-B-AM031, and # IRB14-1241).

A digital polysomnography system (Polysmith; Nihon Kohden America Inc., CA, USA) was used to monitor the childrens' sleep. SpO_2 recordings were obtained during overnight PSG at sampling rates of 25 Hz, 200 Hz, or 500 Hz. They were exported and processed offline. Artefacts were rejected from oximetric recordings by removing those SpO_2 values below 50% and sudden changes between consecutive SpO_2 samples faster than 4%/second (Magalang *et al* 2003). Then, a non-overlapping averaging-window of 1 s was applied (effective sampling rate = 1 Hz) to speed up the signal processing stage, which has been found to be appropriate to perform a multiscale analysis of the oximetry signal (Crespo *et al* 2017, Hua and Yu 2017). This window size is lower than 3 s, which is the maximum averaging-time recommended by the American Academy of Sleep Medicine (AASM) (Berry *et al* 2012). The resolution of the SpO_2 signals was set to two decimal points to ensure the resolution was the same (Hornero *et al* 2017).

Sleep and cardiorespiratory events were scored and quantified by specialised technologists and further confirmed by paediatric sleep medicine specialists who were unaware of the study purpose. The AHI was estimated according to the AASM guidelines (Berry *et al* 2012). In this sense, there is no consensus regarding the AHI cut-off used to determine SAHS and its severity (Alonso-Álvarez *et al* 2011, Church 2012, Marcus *et al* 2012, Tan *et al* 2014). However, a wide range of studies typically classify children into four SAHS severity degrees: no-SAHS (AHI < 1 e h^{-1}), mild SAHS ($1 \leqslant$ AHI < 5 e h^{-1}), moderate SAHS ($5 \leqslant$ AHI < 10 e h^{-1}), and severe SAHS (AHI $\geqslant 10$ e h^{-1}) (Alonso-Álvarez *et al* 2011, Church 2012, Tan *et al* 2014, Hornero *et al* 2017). Thus, the AHI cut-offs of 1, 5, and 10 e h^{-1} were adopted in this study.

The dataset was randomly divided into a training set (60%) and a test set (40%). Table 1 shows the clinical and demographic data of the population under study. No statistically significant differences (p-value < 0.01) were found in either age or body mass index.

2.2. Automated signal processing

Our approach consisted of three sequential stages. First, features derived from DFA and *ODI3* were obtained from the SpO_2 recording of each subject. Then, a smaller subset of relevant and non-redundant features was selected using the fast correlation-based filter (FCBF) method (Yu and Liu 2004). Finally, a multi-layer perceptron (MLP) neural network (Bishop 1995) was applied to this optimum subset in order to estimate the AHI of each patient.

2.2.1. *Detrended fluctuation analysis*

DFA performs a multiscale analysis of a time series to study its correlation properties (Peng *et al* 1994). The DFA profile shows changes in the correlation properties for different ranges of scales, termed 'crossovers', which may be caused by different non-stationarities in the signal such as (Chen *et al* 2002): (i) segments removed from the signal; (ii) random spikes with variable amplitude; (iii) segments with different local behaviour. Segments of the SpO_2 signal associated with apnoeic events typically have different statistical properties, presenting fluctuations and spikes (Crespo *et al* 2017, Hua and Yu 2017). Thus, these properties of the SpO_2 signal may be reflected in the DFA profile.

Given a signal $x(t)$, the DFA method consists of the following steps (Peng *et al* 1994):

1. The time series $x(t)$ is integrated:

$$y(i) = \sum_{j=1}^{i} \left[x(j) - x_{avg}\right], i = 1, \ldots, N, \tag{1}$$

   where x_{avg} is the average of the whole signal $x(t)$, and N is the length of the SpO_2 signal.
2. The integrated signal $y(i)$ is divided into B non-overlapping windows of equal size. In the case of SpO_2 recordings, the minimum length of the signal is 3 h (10 800 samples) to ensure there were enough sleep cycles (Berry *et al* 2012). Thus, the length of each window (i.e. the scale), k, is between 3 and 1080, since the maximum box size in DFA must be one-tenth of the signal length (Chen *et al* 2002).
3. For each window b ($b = 1, \ldots, B$), the local trend was obtained as a straight line, y^b, estimated by applying a least squares fitting to y_i.
4. The variance of the fluctuation in each window, $F_b^2(k)$, is defined as follows:

$$F_b^2(k) = \frac{1}{k} \sum_{j=(b-1)k+1}^{bk} \left(y(j) - y^b(j)\right)^2. \tag{2}$$

Table 1. Clinical and demographic data of the population under study.

Characteristics	All	AHI < 1	$1 \leqslant$ AHI < 5	$5 \leqslant$ AHI < 10	AHI $\geqslant 10$
All subjects					
Subjects (n)	981	175	401	176	229
Age (years)	6 [3,9]	7 [4,10]	6 [4,9]	5 [2,8]	4 [2,8]
Males (%)	602 (61.4%)	109 (62.3%)	247 (61.6%)	107 (60.8%)	139 (60.7%)
BMI (kg m^{-2})	17.9 [15.8,21.9]	17.4 [15.5,20.9]	17.7 [15.9,21.2]	18.6 [16.2,24.0]	18.3 [16.0,23.2]
AHI (e h^{-1})	3.8 [1.5,9.3]	0.5 [0.1,0.8]	2.5 [1.7,3.5]	6.8 [5.8,8.3]	19.1 [13.9,31.1]
Training set (60%)					
Subjects (n)	589	98	232	113	146
Age (years)	6 [3,8]	6 [4,8]	7 [4,9]	5 [2,8]	5 [3,8]
Males (%)	348 (59.1%)	61 (62.2%)	140 (60.3%)	72 (63.7%)	75 (51.4%)
BMI (kg m^{-2})	17.6 [15.9,22.0]	17.0 [15.4,19.9]	17.5 [15.9,21.6]	18.6 [16.2,23.7]	18.1 [15.9,23.6]
AHI (e h^{-1})	4.1 [1.7,9.9]	0.4 [0.0,0.8]	2.5 [1.8,3.6]	6.9 [5.8,8.5]	18.9 [13.8,33.5]
Test set (40%)					
Subjects (n)	392	77	169	63	83
Age (years)	6 [3,9]	8 [5,10]	5 [2,9]	6 [4,9]	4 [2,8]
Males (%)	254 (64.8%)	48 (62.3%)	107 (63.3%)	35 (55.6%)	64 (77.1%)
BMI (kg m^{-2})	18.1 [15.8,21.7]	18.0 [15.6,21.7]	18.0 [15.8,20.7]	18.9 [15.7,26.3]	18.3 [16.0,22.1]
AHI (e h^{-1})	3.3 [1.4,7.8]	0.5 [0.3,0.8]	2.5 [1.7,3.4]	6.8 [5.8,7.8]	19.2 [15.1,28.2]

BMI: body mass index; AHI: apnoea hypopnoea index. Data are presented as median [interquartile range], n or %.

5. The fluctuation function, $F(k)$, is obtained as the square root of the average of $F_b^2(k)$ over all windows:

$$F(k) = \sqrt{\frac{1}{B}\sum_{b=1}^{B} F_b^2(k)}. \tag{3}$$

Steps 2–5 are iterated until the highest scale is used.

A double logarithmic plot was used to analyse the evolution of the DFA plot along scales: $\log(F(k))$ versus $\log(k)$ (Penzel *et al* 2003, Dehkordi *et al* 2016, Hua and Yu 2017). Figure 1 shows the averaged DFA plot for the four SAHS severity groups (AHI < 1 e h^{-1}, $1 \leqslant$ AHI < 5 e h^{-1}, $5 \leqslant$ AHI < 10 e h^{-1}, and AHI $\geqslant 10$ e h^{-1}) in the training set. It can be shown that higher fluctuations are observed as the SAHS severity increases. Additionally, two scaling regions can be observed in the DFA plot:

- Region 1 for scales in the range $0.48 \leqslant \log(k) \leqslant 1.3$ ($3 \leqslant k \leqslant 20$).
- Region 2 for scales in the range $1.60 \leqslant \log(k) \leqslant 3.03$ ($40 \leqslant k \leqslant 1080$).

A crossover is produced in the space between these two regions of the DFA profile. Robust linear regression (Hua and Yu 2017) was applied to estimate the line that fits both regions for each SpO_2 recording.

Figure 2 shows the lines fitted in both regions in an illustrative example of a patient from the training set. We characterised the DFA plot by extracting the following features, as can be seen in figure 2 (Penzel *et al* 2003, Hua and Yu 2017):

- Slopes (scaling exponents) in the line that fits the DFA profile in both regions ($slope_1$ and $slope_2$), as well as their ratio ($slope_{1/2}$). These parameters measure the scaling behaviour of the oximetry signal in each region ($slope_1$ and $slope_2$) and the relative differences in this behaviour between both regions ($slope_{1/2}$).
- Coordinates (k_{12} and $F(k_{12})$) of the intersection formed by the lines fitted in regions 1 and 2. These parameters are intended to characterise the crossover point of the DFA plot.
- Fluctuation function in the scale with a maximum correlation with the severity of SAHS ($F(k_x)$). This parameter was extracted to quantify the fluctuations of the oximetry signal. In order to obtain the optimum value of k_x, Spearman's correlation was computed for each scale between $F(k)$ and the AHI. $K_x = 21$ was therefore obtained as the scale with a maximum Spearman's correlation with the AHI.

It is expected that these parameters allow for the quantification of the differences in the scaling behaviour and the fluctuations of the SpO_2 signal are associated with the severity of SAHS, as shown in figure 1.

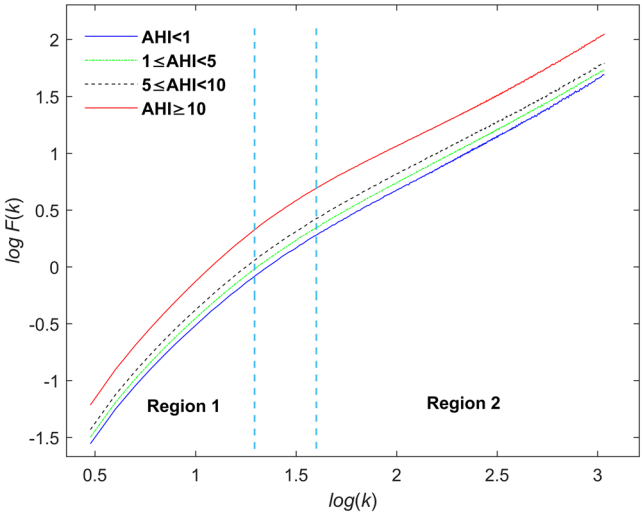

Figure 1. Averaged DFA profile for the four SAHS severity groups: (a) AHI < 1 e h^{-1}, (b) $1 \leqslant$ AHI < 5 e h^{-1}, (c) $5 \leqslant$ AHI < 10 e h^{-1}, and (d) AHI $\geqslant 10$ e h^{-1} in the training set.

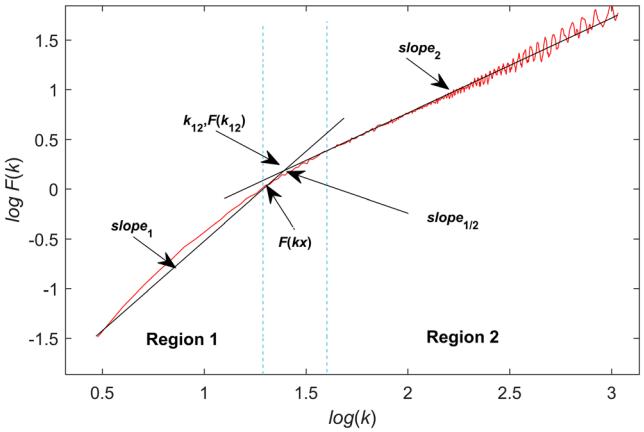

Figure 2. Illustrative example of the DFA plot of the SpO_2 signal of a patient from the training set.

2.2.2. *Oxygen desaturation index*

*ODI*3 was computed as the number of oxygen desaturations from the preceding baseline greater than or equal to 3% per hour of recording (Taha *et al* 1997). This clinical parameter has commonly been used in the SAHS context (Kirk *et al* 2003, Chang *et al* 2013, Tsai *et al* 2013). Higher values of *ODI*3 are expected in patients with a higher severity of SAHS, since oxygen desaturations are associated with apnoea or hypopnoea events (Berry *et al* 2012).

2.2.3. *Feature selection: FCBF*

FCBF was applied to evaluate the relevance of the extracted features and their redundancy within them (Yu and Liu 2004). FCBF has proven its utility in the context of paediatric SAHS diagnosis to obtain subsets of relevant and non-redundant features (Hornero *et al* 2017, Vaquerizo-Villar *et al* 2018). First, FCBF computes the symmetrical uncertainty (*SU*) between each feature x_i and the dependent variable y in order to assess its relevancy (Yu and Liu 2004):

$$SU(x_i, y) = 2\left(\frac{IG(x_i|y)}{H(x_i) + H(y)}\right), i = 1, 2, \ldots, N, \tag{4}$$

where $IG(x_i|y) = H(x_i) - H(x_i|y)$, N is the total number of features extracted ($N = 7$), y is the AHI value of each subject, and H refers to Shannon's entropy (Yu and Liu 2004). *SU* values vary between 0 and 1. $SU = 1$ means that one variable is completely predictable from the other, whereas $SU = 0$ indicates that the two variables are independent.

According to their *SU* value, features are ranked from the most relevant (highest *SU* with the AHI) to the least relevant one (lowest *SU* with the AHI). Different *SU*-based thresholds can be used to discard non-relevant features. Nevertheless, the number of features comprising our original feature set is not high. Therefore, as proposed

by Yu and Liu (2004), no relevance threshold was applied to discard non-relevant features in order to maximize the relevancy of information derived from oximetry (Gutierrez-Tobal *et al* 2018, Hornero *et al* 2017). In this regard, a feature that is useless by itself still may provide useful information when being selected with others (Guyon 2003). A redundancy analysis of each feature is then performed. The *SU* value between each pair of features (x_i, x_j) is computed, beginning with the most relevant one (Yu and Liu 2004). When $SU(x_i, x_j) \geqslant SU(x_i, y)$, the feature x_j is considered redundant with respect to the feature x_i and discarded. In this way, an optimum subset composed of the most relevant and non-redundant features is obtained (Yu and Liu 2004).

A bootstrap methodology was used in order to compose a stable optimum feature subset independent of a particular dataset. FCBF was applied to 1000 bootstrap replicates built from our training data (Efron and Tibshirani 1994, Guyon 2003). Those variables which were selected for at least half of the runs (500) formed the optimum subset (Hornero *et al* 2017, Vaquerizo-Villar *et al* 2018).

2.2.4. AHI estimation: MLP neural network

MLP was applied to estimate the AHI of the subjects under study using the optimum feature subset obtained with FCBF. MLP is one of the most widely used artificial neural networks (ANNs). This ANN has already demonstrated its usefulness in the screening of paediatric SAHS diagnosis using SpO_2 recordings (Hornero *et al* 2017). MLP is arranged in several interconnected layers (input, hidden layers, and output) composed of simple units called perceptrons or neurons (Bishop 1995). Each neuron consists of an activation function g_i and adaptive weights w_{jk} representing connections with neurons from the following layer. In our case, the output layer has one neuron y, which represents the estimated AHI. Additionally, a single hidden layer configuration was implemented, since it is able to provide universal approximation to any function (Bishop 1995). Thus, the output unit in our MLP architecture is calculated as follows:

$$y = g_l\left\{\sum_{j=1}^{N_H} w_{jk} g_t\left\{\sum_{i=1}^{d} w_{ij} x_i + b_j\right\} + b_k\right\}, \tag{5}$$

where g_l and g_t are the activation functions of the output and hidden layer, respectively, w_{jk} are the weights connecting the hidden layer to the output layer, w_{ij} are the weights connecting the input layer to the hidden layer, x_i is the input feature i, b_j and b_k are the biases associated with the hidden and the output units respectively, N_H is the number of units in the hidden layer, and d is the number of input features (Bishop 1995). Weights of the network were randomly initialised. Then, the scaled conjugate gradient with weight-decay regularisation was applied to optimise these weights. This optimisation algorithm minimises the cross-entropy error function and achieves good generalisation, as recommended for pattern recognition tasks (Bishop 1995).

Our MLP network was implemented using the Netlab toolbox (Nabney 2002). The design parameters of the MLP network (the regularisation parameter (α) and N_H) were optimised by means of 10-fold cross-validation using the training set. This optimisation allows us to control the complexity of the MLP network, thus minimising under-fitting and overfitting. Once these parameters were optimised, the MLP model was built using the whole training dataset.

2.2.5. Statistical analysis and diagnostic performance

Matlab R2016a (The MathWorks Inc., Natick, MA, USA) was used to implement automated signal processing algorithms, as well as to perform statistical analyses. The Kruskal–Wallis test was used to assess the statistical differences (p-value < 0.01) between groups, since the extracted features did not pass the Lilliefors normality test. The Bonferroni correction was applied to deal with multiple comparisons. Both agreement between estimated AHI (AHI_{MLP}) and actual AHI (AHI_{PSG}), as well as agreement between *ODI*3 and AHI_{PSG} were assessed by means of Bland–Altman plots and the intra-class correlation coefficient (ICC). Cohen's kappa index (kappa) was used to measure the agreement between AHI_{MLP} and AHI_{PSG}, the agreement between *ODI*3 and AHI_{PSG} to estimate the severity of SAHS (Cohen 1960). The diagnostic ability of *ODI*3 and AHI_{MLP} was assessed in terms of sensitivity (Se, percentage of SAHS positive patients correctly classified), specificity (Sp, percentage of SAHS negative children correctly classified), positive predictive value (PPV, proportion of subjects classified as positive that are true positives), negative predictive value (NPV, proportion of subjects classified as negative that are true negatives), positive likelihood ratio (LR+, likelihood ratio for subjects classified as positive), negative likelihood ratio (LR−, likelihood ratio for subjects classified as negative), accuracy (Acc, percentage of subjects correctly classified), and area under the ROC curve (AUC). A bootstrapping approach was employed in order to compare the ICC, kappa, overall Acc (four classes), and AUC values between *ODI*3 and AHI_{MLP}. The number of bootstrap replicates built from the test data was set to 1000, since it ensures a proper estimation of the 95% confidence interval (Efron and Tibshirani 1994). ICC, kappa, overall Acc, and AUC values were obtained for *ODI*3 and AHI_{MLP} from each of these replicates. Then, the p-value between *ODI*3 and AHI_{MLP} was computed for each of these metrics according to the Mann–Whitney U test.

Table 2. Feature values for the SAHS severity groups (median [interquartile range]) in the training set.

Features	AHI < 1	$1 \leqslant$ AHI < 5	$5 \leqslant$ AHI < 10	AHI $\geqslant 10$	p-value
*ODI*3	1.04 [0.52,2.47]	2.03 [0.93,3.89]	3.69 [1.94,7.23]	12.35 [6.65,24.49]	<0.01
$slope_1$	1.63 [1.58,1.68]	1.64 [1.58,1.70]	1.67 [1.60,1.71]	1.74 [1.66,1.79]	<0.01
$slope_2$	0.96 [0.90,1.05]	0.95 [0.87,1.03]	0.92 [0.85,1.02]	0.94 [0.88,1.01]	0.18[a]
$slope_{12}$	1.66 [1.53,1.82]	1.69 [1.55,1.87]	1.77 [1.60,1.94]	1.82 [1.68,1.95]	<0.01
k_{12}	1.33 [1.23,1.42]	1.36 [1.26,1.44]	1.38 [1.29,1.45]	1.34 [1.23,1.42]	0.04[a]
$F(k_{12})$	0.01 [−0.18,0.18]	0.12 [−0.12,0.26]	0.22 [0.04,0.38]	0.42 [0.16,0.61]	<0.01
$F(k_x)$	−0.05 [−0.13,0.04]	0.02 [−0.07,0.11]	0.10 [0.00,0.20]	0.31 [0.18,0.52]	<0.01

[a] Not lower after Bonferroni correction (p-value $= 0.01/6$).

3. Results

3.1. Training set

3.1.1. Features separability

A total of seven features were obtained for each subject: *ODI*3, and six DFA-derived features. Table 2 shows the median and interquartile range of these features in the training set for each SAHS severity group, along with their corresponding p-values. *ODI*3 and four out of six DFA-derived features ($slope_1$, $slope_{1/2}$, $F(k_{12})$, and $F(k_x)$) showed statistically significant differences (p-value < 0.01 after Bonferroni correction).

3.1.2. Optimum feature subset

Figure 3 displays the histogram with the number of times that each feature was selected over the 1000 bootstrap replicates. *ODI*3 was selected all the time, which agrees with previous studies (Hornero *et al* 2017, Vaquerizo-Villar *et al* 2018). Additionally, $slope_1$ was selected more than half the time (535). Thus, *ODI*3 and $slope_1$ were chosen as the optimum subset.

3.1.3. MLP model optimisation and training

The MLP network was designed and trained using this optimum feature subset (*ODI*3 and $slope_1$). In order to find the optimum values, N_H and α were varied from $N_H = 2$ up to $N_H = 30$ and $\alpha = 0$ up to $\alpha = 10$, respectively. For each N_H-α pair, kappa was obtained through ten-fold cross validation. Since the network is sensitive to the initial random values of the weights, kappa was computed on the cross validation set and averaged for a total of ten runs for each N_H-α pair. Figure 4 shows the kappa value obtained for each N_H-α pair. According to this figure, the optimum values $N_H = 5$ and $\alpha = 6$ were obtained as those for which kappa was higher. Finally, the optimum feature subset (*ODI*3 and $slope_1$) from the entire training set was used to train the MLP model (AHI_{MLP}) with these optimum user-dependent network parameters.

3.2. Test set

Figures 5(a) and (b) show the Bland–Altman plots of *ODI*3 and AHI_{MLP} compared with AHI_{PSG}, respectively, in the test set. ICC between *ODI*3 and AHI_{MLP} with AHI_{PSG} is also shown. AHI_{MLP} reached a lower mean difference (bias) with AHI_{PSG} than *ODI*3 (0.75 versus −1.65), whereas *ODI*3 achieved a slightly lower confidence interval than AHI_{MLP} (23.2 versus 24.3). Notice that *ODI*3 underestimates AHI, whereas AHI_{MLP} corrects this behaviour by showing a slight overestimation. In addition, AHI_{MLP} achieved better agreement with AHI_{PSG} (ICC = 0.891) than *ODI*3 (ICC = 0.866). Regarding the diagnostic performance, table 3 shows the confusion matrices of *ODI*3 and AHI_{MLP} in the test group. These matrices show the class predicted by both original *ODI*3 and AHI_{MLP} for each subject versus the actual SAHS severity group, according to AHI_{PSG}. Using *ODI*3, 55.4% of the subjects (217/392) were correctly assigned to their actual group of SAHS severity (sum of the main diagonal elements of the matrix). Conversely, AHI_{MLP} rightly assigned 60.0% (235/392) of the subjects to their SAHS severity group. Kappa values were 0.355 (*ODI*3) and 0.412 (AHI_{MLP}). Table 4 shows the diagnostic ability of both *ODI*3 and AHI_{MLP} for the AHI_{PSG}-based cut-offs of 1, 5, and 10 e h^{-1}. AHI_{MLP} outperformed single *ODI*3 in terms of ICC, overall Acc and kappa. Additionally, our AHI_{MLP} reached higher Acc for the single AHI cut-offs of 1 and 10 e h^{-1}. With respect to the comparison of the results of *ODI*3 and AHI_{MLP}, statistically significant higher values (p-value < 0.01) were obtained using AHI_{MLP} in the case of ICC, kappa, and overall Acc. In addition, statistically significant differences were found for the AHI cut-offs of 5 and 10 e h^{-1} between AUC of *ODI*3 and AHI_{MLP}.

4. Discussion

This study evaluated the usefulness of DFA to provide additional information from oximetry dynamics in order to assist with the screening of children at risk for paediatric SAHS. To our knowledge, the application of DFA

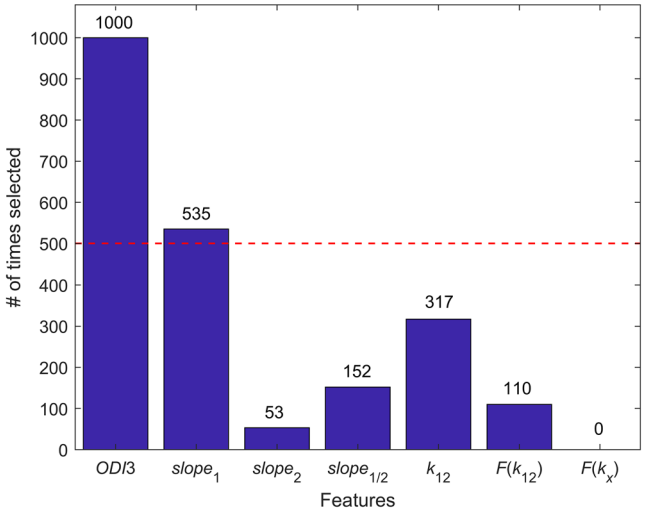

Figure 3. Histogram with the number of times each feature is selected over the 1000 bootstrap iterations.

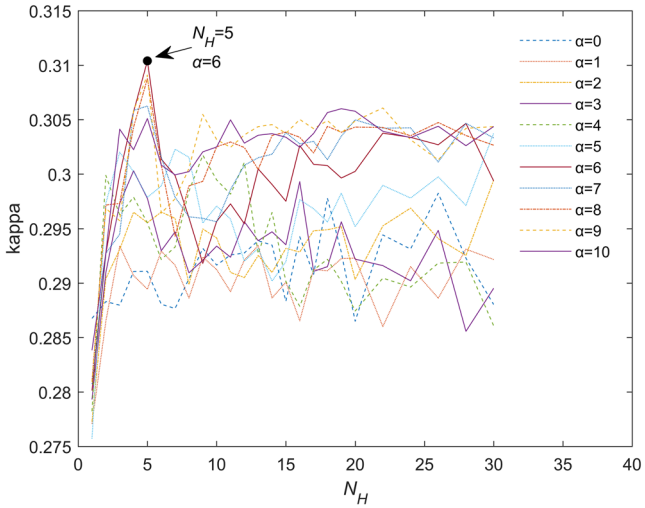

Figure 4. Averaged kappa for each N_H-α pair.

to SpO_2 recordings is novel in the context of paediatric SAHS. Our proposed approach shows a high diagnostic ability which outperforms the conventional oximetric index *ODI*3.

*ODI*3 and four out of six features from DFA ($slope_1$, $slope_{1/2}$, $F(k_{12})$, and $F(k_x)$) reached significantly higher values that were associated with increased severity of SAHS. The statistical differences shown by these DFA-derived parameters indicate that the scaling behaviour of the DFA profile of the SpO_2 signal is affected in the presence of SAHS, as illustrated by figure 1. This change in the correlation properties of the SpO_2 signal along time scales may be caused by the presence of spikes or segments with different statistical properties (Chen *et al* 2002, Hua and Yu 2017). Figure 1 shows two regions with different scaling exponents (correlation)—one region for short-time scales (region 1) and another region for long-time scales (region 2). Two scaling regions were also obtained in the studies developed by Dehkordi *et al* (2016) and Penzel *et al* (2003). Dehkordi *et al* and Penzel *et al* applied DFA to analyse the scaling behaviour of the pulse rate variability (PRV) and heart rate variability (HRV) signals in the context of SAHS, respectively. According to these studies (Dehkordi *et al* 2016, Penzel *et al* 2003), the time scales of these regions may be related to the duration of apnoeic events. In these studies, short-time scales relate to the effects of respiration on the heart rate, whereas long time scales relate to the effects of sleep stages and circadian rhythm (Penzel *et al* 2003).

According to the physiological interpretation of both regions in the DFA profile, the higher values shown by $slope_1$ and $slope_{1/2}$ that are associated with the severity of SAHS may be related to the variations in the SpO_2 signals caused by respiratory events (Hua and Yu 2017, Peng *et al* 1995, Penzel *et al* 2003) which directly affect the oximetry dynamics. On the contrary, $slope_2$ did not show statistically significant differences. According to Penzel *et al* (2003), $slope_2$ is related to the effects of slower brain functions on the HRV signal. Nevertheless, slower brain functions may not be related to the effects of SAHS in the oximetry signal. That is one possible reason why the

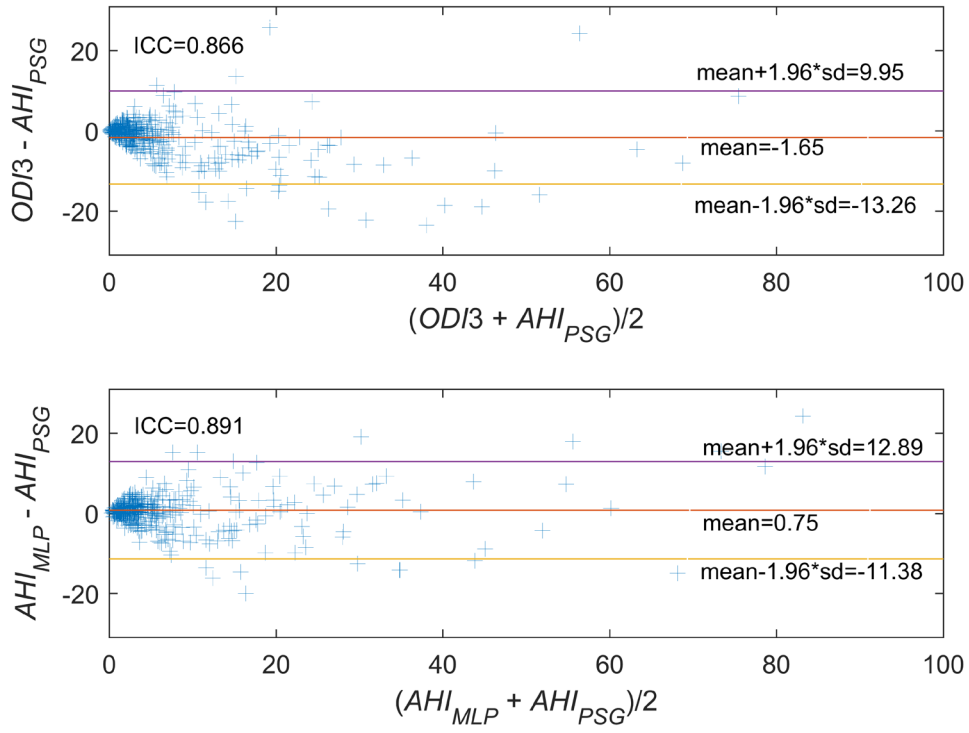

Figure 5. Bland–Altman plots comparing (a) *ODI3* with AHI_{PSG} and (b) AHI_{MLP} with AHI_{PSG}.

Table 3. Confusion matrices of *ODI3* and AHI_{MLP} in the test set. Regarding *ODI3*, average Acc = 55.4% and kappa = 0.355, whereas for AHI_{MLP}, average Acc = 60.0% and kappa = 0.412.

		ODI3				AHI_{MLP}			
		AHI < 1	1 ⩽ AHI < 5	5 ⩽ AHI < 10	AHI ⩾ 10	AHI < 1	1 ⩽ AHI < 5	5 ⩽ AHI < 10	AHI ⩾ 10
AHI_{PSG}	AHI < 1	**39**	36	1	1	**18**	55	3	1
	1 ⩽ AHI < 5	47	**107**	12	3	8	**125**	33	3
	5 ⩽ AHI < 10	5	33	**17**	8	1	22	**28**	12
	AHI ⩾ 10	0	13	16	**54**	0	8	11	**64**

Table 4. Diagnostic ability of *ODI3* and AHI_{MLP} in the test set for AHI cut-offs = 1, 5, and 10 e h^{-1}.

AHI cut-off = 1 e h^{-1}								
Features	Se	Sp	PPV	NPV	LR+	LR−	Acc	AUC
ODI3	83.5	50.6	87.4	42.9	1.7	0.33	77.0	0.811
AHI_{MLP}	97.1	23.3	83.9	66.7	1.3	0.12	82.7	0.813
AHI cut-off = 5 e h^{-1}								
Features	Se	Sp	PPV	NPV	LR+	LR−	Acc	AUC
ODI3	65.1	93.1	84.8	81.8	9.4	0.37	82.7	0.883
AHI_{MLP}	78.8	83.7	74.2	86.9	4.8	0.25	81.9	0.888
AHI cut-off = 10 e h^{-1}								
Features	Se	Sp	PPV	NPV	LR+	LR−	Acc	AUC
ODI3	65.1	96.1	81.8	91.1	16.7	0.36	89.5	0.921
AHI_{MLP}	77.1	94.8	80.0	93.9	14.9	0.24	91.1	0.930

value of $slope_2$ does not increase with the severity of SAHS. Figure 1 also shows higher values of $F(k)$ in the SpO_2 signal as the severity of SAHS increases. These differences may be due to the fluctuations produced in the SpO_2 signal by apnoeic events (Hua and Yu 2017). These fluctuations are reflected in the significantly higher values of $F(k_{12})$ and $F(k_x)$ associated with a higher SAHS severity. Finally, it can be seen in figure 1 that the crossover point between the two regions of the DFA profile occurs at similar time scales for the different SAHS severity groups. The scale value k of the crossover point is related to the duration of apnoeic events (Penzel *et al* 2003), which does

Table 5. Summary of state-of-the-art in the context of the analysis of SpO_2 recordings to assist in the diagnosis of paediatric SAHS.

Studies (year)	Subjects (*n*)	AHI cut-off	Methods	Validation	Se (%)	Sp (%)	Acc (%)
Kirk *et al* (2003)	58	5	*ODI3*	Direct validation[b]	67	60	64[a]
Tsai *et al* (2013)	148	1	*ODI4*	No	77.7	88.9	79.0[a]
		5			83.8	86.5	85.1[a]
		10			89.1	86.0	87.1[a]
Garde *et al* (2014)	146	5	Statistical, nonlinear features, classical indices, and PSD	Four-fold cross validation	80.0	83.9	78.5
Van Eyck *et al* (2015)	130	2	*ODI3* and clusters of desaturations	Train-test for *ODI3*	57	73	68[a]
					58	88	78[a]
					66	69	68[a]
Álvarez *et al* (2017)	50	1	Statistical, nonlinear features, PSD, and classical indices	Bootstrap 0.632	89.6	71.5	85.5
		3			82.9	84.4	83.4
		5			82.2	83.6	82.8
Crespo *et al* (2017)	50	3	Multiscale entropy and classical indices	Bootstrap 0.632	84.5	83.0	83.5
Hornero *et al* (2017)	4191	1	Statistical, nonlinear features, PSD, and *ODI3*	Training-test	84.0	53.2	75.2
		5			68.2	87.2	81.7
		10			68.7	94.1	90.2
Vaquerizo *et al* (2018)	298	5	Bispectrum, PSD, *ODI3*, anthropometric variables	Feature optimisation-training-test	61.8	97.6	81.3
		10			60.0	94.5	85.3
Our proposal	981	1	DFA and *ODI3*	Training-test	97.1	23.3	82.7
		5			78.8	83.7	81.9
		10			77.1	94.8	91.1

[a] Computed from reported data.
[b] Direct validation of a scoring criteria against AHI from PSG.

not depend on the severity of SAHS. This may be the reason why k_{12} did not show statistically significant differences.

Regarding the results of the feature selection stage, figure 3 shows that only *ODI3* and $slope_1$ were selected more than 500 times after the bootstrapping approach. The remaining features showed high redundancy. *ODI3* and $slope_1$ come from different methodological approaches. Therefore, this suggests that information from DFA is complementary to that obtained from the conventional *ODI3*. As aforementioned, an MLP neural network fed with this optimum subset outperformed the *ODI3* (tables 3 and 4). A better agreement with the AHI_{PSG} was achieved with our AHI_{MLP}, as well as a higher diagnostic ability to predict SAHS severity. This highlights the usefulness of FCBF, the feature selection method employed in our proposal. According to our results, $slope_1$, which was involved in the optimum subset, quantifies changes in the scaling behaviour of the DFA profile that provides additional information regarding oximetry dynamics able to enhance its diagnostic ability.

Previous studies also evaluated the usefulness of DFA to characterise SAHS in both adults (Lee *et al* 2002, Penzel *et al* 2003, Kaimakamis *et al* 2016, Hua and Yu 2017) and children (Dehkordi *et al* 2016). Penzel *et al* (2003) and Dehkordi *et al* (2016) extracted the slopes in the scaling regions of the DFA profile from the HRV and PRV signals in order to discriminate sleep stages and detect the presence of SAHS in adults and paediatric patients, respectively. Their findings indicate that the scaling analysis provided by DFA is suitable to quantify the changes of the cardiac signals during sleep stages, as well as the properties of these signals associated with apnoeic events. These results agree with Lee *et al* (2002), who also reported that the scaling exponents of the DFA of the electroencephalogram signal are useful to discriminate between sleep stages in adult patients. Kaimakanis *et al* (2016) reported a 0.77 correlation coefficient in predicting AHI with a linear regression model fed with DFA and other nonlinear methods applied to airflow and thoracic signals from adult patients. Finally, Hua and Yu (2017) evaluated the diagnostic ability of the slopes of four different scaling regions and the coordinates and angles of the intersections of these regions in the DFA plot of the SpO_2 signal in the context of adult SAHS. A high diagnostic performance was achieved with these features, with an accuracy of 90.8%, 80.1%, and 87.4% for the common adult SAHS cut-offs of 5, 15, and 30 e h^{-1}, respectively. Importantly, our research is not limited to the analysis of individual features from DFA, and it assesses the capability of DFA to provide additional and relevant information complementary to conventional approaches (i.e. *ODI3*) to simplify the diagnosis of paediatric SAHS.

Table 5 summarises the performance of previous studies focused on the analysis of SpO_2 as a simplified technique in the screening of paediatric SAHS (Kirk *et al* 2003, Tsai *et al* 2013, Garde *et al* 2014, Eyck *et al* 2015, Álvarez *et al* 2017, Crespo *et al* 2017, Hornero *et al* 2017, Vaquerizo-Villar *et al* 2018). Some of them have applied the ODI and clusters of desaturations (Kirk *et al* 2003, Tsai *et al* 2013, Eyck *et al* 2015). However, only Tsai *et al* (2013) reached accuracies higher than 80%. Notwithstanding, further validation was felt to be still necessary in order to independently assess the proposed ODI-based cut-offs.

Recent studies have focused on the application of automated signal processing approaches to enhance the diagnostic ability of the SpO_2 signal (Garde *et al* 2014, Álvarez *et al* 2017, Crespo *et al* 2017, Hornero *et al* 2017, Vaquerizo-Villar *et al* 2018). From these studies, only Hornero *et al* (2017) assessed an AHI estimation model. Hornero *et al* (2017) built an MLP regression model with *ODI*3 and the skewness of the PSD extracted from 4191 SpO_2 recordings from 13 sleep laboratories worldwide. Our study outperformed the state-of-the-art approaches except the performance reported by the study of Álvarez *et al* (2017), which achieved higher accuracies for the AHI cut-offs of 1 and 5 e h^{-1}. However, the database used by Álvarez *et al* (2017) had only 50 subjects. As a consequence, their results are less generalizable, and they performed binary classification instead of estimating the AHI of each patient.

In spite of the promising results of our proposed approach, several limitations must be taken into account. First, the number of subjects belonging to the no-SAHS (AHI $<$ 1 e h^{-1}) group is low when compared to the other severity groups. This issue likely contributes to the slight trend of the MLP model to overestimate the AHI of the subjects belonging to this group, thus resulting in a low specificity for an AHI-threshold of 1 e h^{-1}. However, this is likely the situation in clinical settings when only symptomatic children would be referred for evaluation. Nonetheless, a more balanced proportion of subjects among SAHS severity groups would likely minimise this effect. Another limitation concerns the use of the SpO_2 signal alone to detect SAHS, since some physiological perturbations of SAHS may not be detected by the oximetry signal, such as airflow reductions, electroencephalographic arousals, or increased intrathoracic pressure swings (Marcus *et al* 2012). The use of SpO_2 together with other biomedical signals could detect these perturbations and, consequently, enhance the detection of SAHS. However, this would increase the complexity of the screening method. Additionally, the application of more advanced machine learning algorithms could be potentially useful to improve the diagnostic ability of our proposal. It would also be appropriate to evaluate our methodology in a database of oximetry recordings obtained with the patients being evaluated at home. Finally, the implementation of our proposal in a portable oximeter could facilitate its use in ambulatory settings.

5. Conclusion

In summary, we investigated the usefulness of DFA to obtain additional information from SpO_2 recordings in order to simplify the detection of paediatric SAHS. Four features extracted from DFA showed significant differences between the SAHS severity groups. An optimum subset composed of *ODI*3 and $slope_1$ was obtained with FCBF, which suggests that these features are complementary and non-overlapping. An MLP model fed with this optimum subset achieved a good agreement with the AHI from PSG, obtaining 0.891 ICC and 0.412 kappa, as well as high diagnostic ability. This MLP model achieved better agreement (ICC and kappa) than *ODI*3, as well as higher accuracies for the cut-offs of 1 and 10 e h^{-1}. Our methodology achieved a high diagnostic performance in comparison with state-of-the-art techniques. This suggests that the changes in the scaling behaviour of the DFA profile quantified by $slope_1$ can provide additional information to enhance the diagnostic ability of the oximetry signal in the context of paediatric SAHS.

Acknowledgments

This work was supported by the 'Ministerio de Ciencia, Innovación y Universidades' and 'European Regional Development Fund (FEDER)' under projects DPI2017-84280-R and RTC-2015-3446-1, and by the 'European Commission' and 'FEDER' under project 'Análisis y correlación entre el genoma completo y la actividad cerebral para la ayuda en el diagnóstico de la enfermedad de Alzheimer' ('Cooperation Pro- gramme Interreg V-A Spain-Portugal POCTEP 2014–2020').

F Vaquerizo-Villar was in receipt of a 'Ayuda para contratos predoctorales para la Formación de Profesorado Universitario (FPU)' grant from the Ministerio de Educación, Cultura y Deporte (FPU16/02938). V Barroso-García was in a receipt of a 'Ayuda para financiar la contratación predoctoral de personal investigador' grant from the Consejería de Educación de la Junta de Castilla y León and the European Social Fund. D Álvarez was in receipt of a Juan de la Cierva grant from MINECO (IJCI-2014-22664). L Kheirandish-Gozal and D Gozal were supported by the National Institutes of Health (NIH) grant HL130984.

Conflict of interest

There are no conflicts of interest that could inappropriately influence this research work.

Ethical approval

In all participants, the informed consent to be included in the research was obtained and the Ethical Committee of the University of Chicago Medicine approved the protocols (#11-0268-AM017, # 09-115-B-AM031, and # IRB14-1241).

Authorship responsibility

- The material in this manuscript is original and contains no libelous or otherwise unlawful matter.
- The manuscript represents valid work and neither this manuscript or any other with substantially similar content under our authorship has been published or is being considered for publication elsewhere.
- We have participated sufficiently in the work to take public responsibility for all its content.

ORCID iDs

Fernando Vaquerizo-Villar https://orcid.org/0000-0002-5898-2006
Daniel Álvarez https://orcid.org/0000-0003-1027-2395
Gonzalo C Gutiérrez-Tobal https://orcid.org/0000-0002-1237-3424
Verónica Barroso-García https://orcid.org/0000-0001-5648-206X
Félix del Campo https://orcid.org/0000-0002-4554-2167
David Gozal https://orcid.org/0000-0001-8195-6036
Roberto Hornero https://orcid.org/0000-0001-9915-2570

References

Alonso-Álvarez M L *et al* 2011 Documento de consenso del síndrome de apneas-hipopneas durante el sueño en niños *Arch. Bronconeumol.* **47** 2–18

Álvarez D, Alonso-Álvarez M L, Gutiérrez-Tobal G C, Crespo A, Kheirandish-Gozal L, Gozal D, Terán-Santos J and Del Campo F 2017 Automated screening of children with obstructive sleep apnea using nocturnal oximetry: an alternative to respiratory polygraphy in unattended settings *J. Clin. Sleep Med.* **13** 7–11

Berry R B *et al* 2012 Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine *J. Clin. Sleep Med.* **8** 597

Bishop C M 1995 *Neural Networks for Pattern Recognition* (New York: Oxford University Press)

Chang L, Wu J and Cao L 2013 Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea *Int. J. Pediatr. Otorhinolaryngol.* **77** 365–71

Chen Z, Ivanov P C, Hu K and Stanley H E 2002 Effect of nonstationarities on detrended fluctuation analysis *Phys. Rev.* E **65** 15

Church G D 2012 The role of polysomnography in diagnosing and treating obstructive sleep apnea in pediatric patients *Curr. Problems Pediatr. Adolesc. Health Care* **42** 22–5

Cohen J 1960 A coefficient of agreement for nominal scales *Educ. Psychol. Meas.* **20** 37–46

Crespo A, Álvarez D, Gutiérrez-Tobal G C, Vaquerizo-Villar F, Barroso-García V, Alonso-Álvarez M L, Terán-Santos J, Hornero R and del Campo F 2017 Multiscale entropy analysis of unattended oximetric recordings to assist in the screening of paediatric sleep apnoea at home *Entropy* **19** 284

Dehkordi P, Garde A, Karlen W, Petersen C C, Wensley D, Dumont G A and Ansermino J M 2016 Evaluation of cardiac modulation in children in response to apnea/hypopnea using the phone oximeter TM *Physiol. Meas.* **37** 187–202

Efron B and Tibshirani R J 1994 *Introduction to the Bootstrap*

Van Eyck A, Lambrechts C and Vanheeswijck L 2015 The role of nocturnal pulse oximetry in the screening for obstructive sleep apnea in obese children and adolescents *Sleep Med.* **16** 1409–12

Garde A, Dehkordi P, Karlen W, Wensley D, Ansermino J M and Dumont G A 2014 Development of a screening tool for sleep disordered breathing in children using the phone Oximeter™ *PLoS One* **9** e112959

Gutierrez-Tobal G C, Alvarez D, Crespo A, Del Campo F and Hornero R 2018 Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings *IEEE J. Biomed. Heal. Informatics* (https://doi.org/10.1109/JBHI.2018.2823384)

Guyon I 2003 An introduction to variable and feature selection *J. Mach. Learn. Res.* **3** 1157–82

Hornero R *et al* 2017 Nocturnal oximetry-based evaluation of habitually snoring children *Am. J. Respir. Crit. Care Med.* **196** 1591–8

Hua C and Yu C-C 2017 Detrended fluctuation analysis of oxyhemoglobin saturation by pulse oximetry in sleep apnea syndrome *J. Med. Biol. Eng.* **37** 791–9

Kaditis A, Kheirandish-Gozal L and Gozal D 2016 Pediatric OSAS: oximetry can provide answers when polysomnography is not available *Sleep Med. Rev.* **27** 96–105

Kaimakamis E, Tsara V, Bratsas C, Sichletidis L, Karvounis C and Maglaveras N 2016 Evaluation of a decision support system for obstructive sleep apnea with nonlinear analysis of respiratory signals *PLoS One* **11** 1–16

Katz E S, Mitchell R B and Ambrosio C M D 2012 Obstructive sleep apnea in infants *Am. J. Respir. Crit. Care Med.* **185** 805–16

Kirk V G, Bohn S G, Flemons W W and Remmers J E 2003 Comparison of home oximetry monitoring with laboratory polysomnography in children *Chest J.* **124** 1702–8

Lee J M, Kim D J, Kim I Y, Park K S and Kim S I 2002 Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data *Comput. Biol. Med.* **32** 37–47

Magalang U J, Dmochowski J, Veeramachaneni S, Draw A, Mador J, El-Solh A and Grant B J B 2003 Prediction of the apnea–hypopnea index from overnight pulse oximetry *Chest J.* **124** 1694–701

Marcus C L *et al* 2012 Diagnosis and management of childhood obstructive sleep apnea syndrome *Pediatrics* **130** 576–84

McClatchey K D 2002 *Clinical Laboratory Medicine* (Baltimore, MD: Williams & Wilkins)

Nabney I 2002 *NETLAB: Algorithms for Pattern Recognition* (Berlin: Springer)

Netzar N, Eliasson A H, Netzar C and Kristo D A 2001 Overnight pulse oximetry for sleep- disordered breathing in adults *Chest J.* **120** 625–33

Nixon G M, Kermack A S, Davis G M, Manoukian J J, Brown A and Brouillette R T 2004 Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry *Pediatrics* **113** e19–25

Peng C-K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 Mosaic organization of DNA nucleotides *Phys. Rev.* E **49** 1685

Peng C K, Havlin S, Stanley H E and Goldberger A L 1995 Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series *Chaos* **5** 82–7

Penzel T, Kantelhardt J W, Grote L, Peter J and Bunde A 2003 Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea *IEEE Trans. Biomed. Eng.* **50** 1143–51

Taha B H, Dempsey J A, Weber S M, Badr M S, Skatrud J B, Young T B, Jacques A J and Seow K C 1997 Automated detection and classification of sleep-disordered breathing from conventional polysomnography data *Sleep* **20** 991–1001

Tan H-L, Gozal D, Ramirez H M, Bandla H P R and Kheirandish-Gozal L 2014 Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea *Sleep* **37** 255–60

Tsai C, Kang C, Su M, Lin H, Huang E, Chen C, Hung J, Niu C, Liao D and Yu H 2013 Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children *Int. J. Pediatr. Otorhinolaryngol.* **77** 1286–90

Vaquerizo-Villar F, Álvarez D, Kheirandish-Gozal L, Gutiérrez-Tobal G C, Barroso-García V, Crespo A, del Campo F, Gozal D and Hornero R 2018 Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings *Comput. Methods Programs Biomed.* **156** 141–9

Yu L and Liu H 2004 Efficient feature selection via analysis of relevance and redundancy *J. Mach. Learn. Res.* **5** 1205–24