Cite this document as:

Physiological Measurement

Symbolic dynamics to enhance diagnostic ability of portable oximetry from the Phone Oximeter in the detection of paediatric sleep apnoea

Daniel Álvarez[1,2], Andrea Crespo[1,2], Fernando Vaquerizo-Villar[2], Gonzalo C Gutiérrez-Tobal[2], Ana Cerezo-Hernández[1], Verónica Barroso-García[2], J Mark Ansermino[3], Guy A Dumont[3], Roberto Hornero[2], Félix del Campo[1,2], Ainara Garde[4]

[1]Servicio de Neumología, Hospital Universitario Río Hortega, c/ Dulzaina 2, 47012, Valladolid, España
[2]Biomedical Engineering Group, University of Valladolid, Paseo de Belén 15, 47011, Valladolid, España
[3]The University of British Columbia and British Columbia Children's Hospital, 4480 Oak St., Vancouver, Canada
[4]Biomedical Signals and Systems, Faculty EEMCS, University of Twente, 7500 AE, Enschede, The Netherlands

E-mail: dalvgon@gmail.com

Keywords: Paediatric obstructive sleep apnoea-hypopnoea syndrome, nocturnal oximetry, portable, signal processing, symbolic dynamics, pattern recognition

Abstract

Objective. This study is aimed at assessing symbolic dynamics as a reliable technique to characterise complex fluctuations of portable oximetry in the context of automated detection of childhood obstructive sleep apnoea-hypopnoea syndrome (OSAHS). *Approach*. Nocturnal oximetry signals from 142 children with suspected OSAHS were acquired using the Phone Oximeter: a portable device that integrates a pulse oximeter with a smartphone. An apnoea-hypopnoea index (AHI) ≥5 events/h from simultaneous in-lab polysomnography was used to confirm moderate-to-severe childhood OSAHS. Symbolic dynamics was used to parameterise non-linear changes in the overnight oximetry profile. Conventional indices, anthropometric measures, and time-domain linear statistics were also considered. Forward stepwise logistic regression was used to obtain an optimum feature subset. Logistic regression (LR) was used to identify children with moderate-to-severe OSAHS. *Main results*. The histogram of 3-symbol words from symbolic dynamics showed significant differences ($p < 0.01$) between children with AHI <5 events/h and moderate-to-severe patients (AHI ≥5 events/h). Words representing increasing oximetry values after apnoeic events (re-saturations) showed relevant diagnostic information. Regarding the performance of individual characterization approaches, the LR model composed of features from symbolic dynamics alone reached a maximum performance of 78.4% accuracy (65.2% sensitivity; 86.8% specificity) and 0.83 area under the ROC curve (AUC). The classification performance improved combining all features. The optimum model from feature selection achieved 83.3% accuracy (73.5% sensitivity; 89.5% specificity) and 0.89 AUC, significantly (p-value <0.01) outperforming the other models. *Significance*. Symbolic dynamics provides complementary information to conventional oximetry analysis enabling reliable detection of moderate-to-severe paediatric OSAHS from portable oximetry.

1. Introduction

Obstructive sleep apnoea syndrome (OSAHS) is a prevalent condition in children (Marcus *et al* 2012). Untreated paediatric OSAHS is associated with significant negative consequences for children's health and quality of life, including impairment of neuropsychological and cognitive performance, metabolic dysfunction, behavioural abnormalities, and long-term cardiovascular effects (Marcus *et al* 2012).

Childhood OSAHS is characterised by recurrent episodes of prolonged partial and/or intermittent complete upper airway obstruction during sleep, leading to disruption of normal oxygenation and sleep patterns (Marcus *et al* 2012, Kaditis *et al* 2016a). In-laboratory nocturnal polysomnography (PSG) is considered the gold standard for an objective diagnosis of this disease (Marcus *et al* 2012, DeHaan *et al* 2015). PSG allows for quantitative evaluation of cardiorespiratory events so that children can be stratified in terms of OSAHS severity (Marcus *et al* 2012). However, several studies pointed out the drawbacks of in-lab PSG concerning lack of availability and intrusiveness (Kheirandish-Gozal 2010, Lesser *et al* 2012, Katz *et al* 2012). In order to overcome these limitations, abbreviated portable monitoring has been proposed to increase accessibility to diagnostic resources while decreasing intrusiveness, particularly pertinent for children (Marcus *et al* 2012, Kaditis *et al* 2016a).

Simplified portable monitors focus on a relevant subset of physiological sleep-related recordings while allowing the attachment of sensors by caretakers or patients themselves. In the most common procedure, the device is returned to the sleep laboratory next morning in order to be scored by a qualified sleep technician. Although this protocol has enabled abbreviated portable monitoring to be a first-line screening method for OSAHS (Penzel *et al* 2018), telemedicine-based technologies are able to enhance diagnostic methodologies and overall health management. In this regard, the use of smartphones has gained increasing popularity due to their acquisition, storage, processing, and transmission capabilities. In the framework of sleep apnoea, smartphone-based medical applications cover both diagnostic and therapeutic purposes, although treatment compliance monitoring in adult patients is the most popular (Penzel *et al* 2018). In addition, there are several popular apps for mobile devices aimed at tracking and assessing sleep quality. However, there are major concerns on their clinical effectiveness due to the lack of scientific evidence and regulatory approval (Behar *et al* 2013, Ko *et al* 2015). In the context of childhood OSAHS, the Phone Oximeter is a major exception, achieving a robust design and independent clinical validation. The Phone Oximeter integrates a low-cost oximetry probe in a smartphone, showing high diagnostic performance in the detection of moderate-to-severe paediatric OSAHS and has been extensively validated against the standard PSG (Garde *et al* 2014).

Oximetry is considered an appropriate screening tool for sleep-related breathing disorders due to its simplicity and readiness (Tsai *et al* 2013, Kaditis *et al* 2016b). Conventional indexes such as the oxygen desaturation index (ODI) and the presence of clusters of desaturations (Velasco *et al* 2013, Van Eyck *et al* 2015, Villa *et al* 2015, Chang *et al* 2013, Tsai *et al* 2013) have been found to provide significant information for OSAHS detection. Recently, the application of novel signal processing and automated pattern recognition techniques have increased the diagnostic ability of overnight oximetry (Garde *et al* 2014, Álvarez *et al* 2017, Hornero *et al* 2017). However, there are still some discrepancies on the efficacy of abbreviated methods as a single tool for childhood OSAHS detection, particularly oximetry alone. Recent reports have highlighted that nocturnal pulse oximetry is far from a perfect screening tool for mild cases and specific subgroups of children (Kirk *et al* 2017, Van Eyck and Verhulst 2018). Therefore, further research is demanded in order to provide additional evidence of the effectiveness of single-channel overnight oximetry for paediatric OSAHS diagnosis.

In this regard, recent studies have focused on obtaining relevant as well as complementary information to conventional signal processing approaches due to the presence of non-stationarities and nonlinearities inherent in biological systems. Crespo *et al* (2017) used the multiscale sample entropy to characterise non-linear patters present in the overnight profile of oximetry linked with apnoeic events. Recently, Vaquerizo-Villar *et al* (2018) assessed the bispectrum as an alternative to conventional PSD in order to characterise deviations from linearity and stationarity of SpO_2 recordings. Previously, sample entropy, central tendency measure, and Lempel-Ziv complexity were predominantly used in the context of paediatric OSAHS detection from oximetry due to their proven performance in adult cases (Garde et al 2014, Álvarez *et al* 2017, Hornero *et al* 2017).

In the present study, symbolic dynamics is proposed to analyse changes in overnight oximetry recordings from children suspected of suffering from OSAHS. Symbolic dynamics provides an

alternative approach to investigate complex non-linear systems (Voss *et al* 1996). Data is transformed into a small set of symbols so that the study of the dynamics of the system is accomplished by describing symbol sequences (Kurths *et al* 1995). In the framework of biomedical signals and systems, it has been mainly used to analyse non-linear characteristics of heart rate modulation (Kurths *et al* 1995, Voss *et al* 1996, Yeragani *et al* 2000, Guzzetti *et al* 2005). Similarly, it has been also applied to characterise cardiovascular regulation and cardio-respiratory coupling during sleep (Suhrbier et al 2010, Penzel *et al* 2016). In the field of automated OSAHS detection, previous studies assessed its ability to increase the performance of single-lead ECG as a simplified screening test for adult OSAHS (Ravelo-García *et al* 2014, 2015). Recent preliminary studies also investigated the usefulness of symbolic analysis to characterise sleep disordered breathing in children. Immanuel *et al* (2014) analysed the EEG complexity throughout the respiratory cycle to detect electroencephalographic alterations. Similarly, Baumert *et al* (2015) applied joint symbolic dynamics to study temporal interactions between heart period and pulse transit time in children with and without sleep disordered breathing while sleeping. However, symbolic dynamics has not been previously used to thoroughly analyse the overnight oximetry tracing from paediatric OSAHS patients.

In this research, we hypothesised that symbolic dynamics is able to provide significant new information on the non-linear behaviour of overnight oximetric recordings from children with suspected OSAHS resulting in improved diagnostic performance. Accordingly, the aim of our research was to assess the usefulness of symbolic dynamics for detecting childhood OSAHS from portable nocturnal oximetry.

2. Materials and methods

2.1. Population under study and biomedical recordings

A total of 142 children (85 boys and 57 girls) with median age of 9 years old and median body mass index (BMI) of 18.4 kg/m^2 composed the population under study. Table 1 summarises the socio-demographic and clinical data of the children involved in the study. All children were referred to the British Columbia Children's Hospital of Vancouver (Canada) showing clinical suspicion of suffering from OSAHS due to the following symptoms reported by their parents or caretakers: snoring, daytime sleepiness, behavioural problems and/or clinically large tonsils. Children suffering from cardiac arrhythmia or abnormal haemoglobin were excluded from the study. The Research Ethics Board of the University of British Columbia and the Children's and Women's Health Centre approved the protocol (H11-01769). Written informed consent to participate in the study was obtained from children's parents/guardians prior to the enrolment. In addition, all patients over 11 years of age were asked to provide a written assent.

Table 1. Demographic and clinical data of the children involved in the study. Data are presented as median [interquartile range] or n (%). The *p*-value shown in the last column was computed using the Chi2 test for categorical variables and the Mann-Whitney test for continuous ones. A level of 0.01 was considered for statistical significance (N.S.: non-significant).

Characteristics	All children	AHI<5 group	AHI≥5 group	*p*-value
Nº of children (%)	142	87 (61.3%)	55 (38.7%)	-
Age (years)	9 [7]	10 [6]	8 [8]	N.S.
Nº of boys (%)	85 (59.9%)	47 (54.0%)	38 (69.1%)	N.S.
BMI (Kg/m^2)	18.4 [7.99]	17.57 [6.12]	20.07 [11.69]	N.S.
AHI (events/h)	2.65 [8.30]	1.20 [1.58]	13.10 [17.85]	*p* <0.01

All the children underwent in-laboratory nocturnal PSG, which was used as gold standard for objective OSAHS diagnosis. In-lab PSG was carried out using a polysomnograph Embla Sandman S4500 (Natus Medical Inc., Pleasanton, CA, USA). The following signals were recorded: electrocardiogram, electroencephalogram, peripheral blood oxygen saturation (SpO_2) and heart rate by means of oximetry, chest and abdominal effort, nasal and oral airflow (thermistor and nasal cannula), and video recordings. All PSG studies included in our dataset showed more than 3 hours of total sleep time.

The same sleep technician visually scored cardiorespiratory and neurophysiological events according to the American Academy of Sleep Medicine criteria (Iber *et al* 2007). Obstructive apnoeas were defined as the complete cessation of oronasal airflow during at least two respiratory cycles. Similarly, a decrease ≥50% in the amplitude of the nasal pressure signal lasting two respiratory cycles or more, accompanied by a desaturation ≥3% or an arousal, was scored as a hypopnoea. The apnoea-hypopnoea index (AHI) was defined as the number of apnoeas and hypopnoeas per hour of sleep, which is commonly used by physicians to diagnose or discard the disease and categorise its severity. In this regard, a cut-off of 5 events per hour (events/h) is commonly used to recommend adenotonsillectomy due to moderate-to-severe OSAHS is less likely to resolve spontaneously (Marcus *et al* 2012, Kaditis *et al* 2016b). In addition, an AHI ≥5 events/h is linked with an increased risk for cardiovascular negative effects in children (Kaditis *et al* 2016b). Accordingly, a cut-off of 5 events/h was used to split the cohort into children with moderate-to-severe OSAHS (AHI≥5 group) and children without moderate-to-severe OSAHS (AHI<5 group).

Simultaneously to standard in-lab PSG, an additional SpO_2 signal was acquired using the Phone Oximeter (Hudson *et al* 2012, Petersen *et al* 2013). The Phone Oximeter (prototype version 1.0) is a portable device that integrates a commercially available and Federal Drug Administration (FDA) approved microcontroller-based pulse oximetry sensor (Masimo SET® uSpO$_2$ Pulse Oximetry Cable, Masimo Corporation, Irvine, CA, USA) with a smartphone (iPhone 4S or later). The sensor is directly connected to the smartphone enabling portable acquisition, monitoring, and storage of SpO_2 recordings both in hospital and at home. The main purpose of the Phone Oximeter is to increase the portability of conventional oximeters by attaching the sensor with a mobile smartphone, leading to improved availability and accessibility to diagnostic resources. Previous works exhaustively validated technical features and usability of SpO_2 acquisition systems based on oximetry sensors attached to mobile phones via standard communication ports (Karlen *et al* 2011, Hudson et al 2012, Petersen et al 2013). Particularly, Garde *et al* (2014) found that the Phone Oximeter is able to accurately measure the overnight SpO_2 profile of children with significant sleep disturbed breathing.

The finger probe of the Phone Oximeter was applied to the finger adjacent to the one used during simultaneous complete PSG. Using this portable device, the SpO_2 signal was recorded at a sampling rate of 1 Hz and 0.1% resolution. Recordings were downloaded and an automated pre-processing stage was carried out before pattern recognition. Samples showing oxygen saturation values below 50% as well as changes with slope >4 %/s were consider to be artefacts and removed. In addition, SpO_2 recordings with total recording time <3 hours after pre-processing were discarded.

2.2. Automated pattern recognition

Portable SpO_2 recordings from the Phone Oximeter were processed off-line. The aim of the automated pattern recognition procedure was to perform binary classification of children into two categories: "AHI<5 group" *versus* "AHI≥5 group". In order to achieve this goal, feature extraction, selection, and classification stages were implemented.

2.2.1. Feature extraction

Every child in our dataset was represented using 24 variables from four *a priori* complementary feature subsets: 6 conventional oximetry indices; 2 anthropometric variables; 4 common statistical moments; and 12 novel non-linear measures from symbolic dynamics.

Age and BMI were considered potentially discriminant features and they were both included in our initial feature set (Marcus *et al* 2012). Similarly, conventional oximetry indices commonly used in clinical practice were used to characterise portable oximetry recordings from the Phone Oximeter. Particularly, the number of desaturations greater than or equal to 2% (*ODI2*), 3% (*ODI3*), and 4% (*ODI4*) from baseline per hour of recording were computed (Chang *et al* 2013, Tsai *et al* 2013). The ratio of the cumulative time spent below a saturation of 95% to the total recording time (*CT95*) as well as the average (*Sat$_{AVG}$*) and minimum (*Sat$_{MIN}$*) saturation were also quantified (Álvarez *et al* 2017, Crespo *et al* 2017).

First-to-fourth order statistical moments of the SpO$_2$ amplitude distribution have also demonstrated to provide discriminant information between OSAHS patients and children without the disease (Álvarez *et al* 2017, Crespo *et al* 2017, 2018). Accordingly, mean (*M1t*), variance (*M2t*), skewness (*M3t*) and kurtosis (*M4t*) were computed to characterise changes in central tendency, dispersion, asymmetry, and peakedness of the data histogram related to desaturations, respectively.

A symbolic dynamics-based non-linear approach was applied to obtain additional as well as complementary information to that provided by traditional linear methods in the time domain. Symbolic dynamics is based on a coarse-graining procedure, which involves partitioning the range of original observations (SpO$_2$ amplitudes) into a finite number of non-overlapping regions (Daw *et al* 2003). Accordingly, each sample from the original time series is mapped into the corresponding symbol so that the time series is transformed into a symbol sequence (Kurths *et al* 1995, Voss *et al* 1996). In order to capture the dynamics of the system successfully, a suitable number of symbols (alphabet size) and the partition of the range of amplitudes of the time series have to be set. In the present study, we defined an alphabet composed of p=4 symbols {1, 2, 3, 4}, which has been found to be the most appropriate for quantifying cardiovascular dynamics (Kurths *et al* 1995, Voss *et al* 1996) and, particularly, in the framework of adult OSAHS detection from HRV recordings (Ravelo-García *et al* 2014, 2015).

The partition generating the coding scheme is usually defined in terms of two parameters (Kurths *et al* 1995, Voss *et al* 1996):

- m, which is an overall measure of central tendency of the time series (usually the mean).
- a, which is a tolerance or measure of dispersion.

Nevertheless, values assigned to parameters m and a are context dependent. Figure 1 illustrates the symbolization process proposed in the present study for mapping the original SpO$_2$ signal into the predefined four symbols. Firstly, a region of signal stability was defined using the median of the recording: threshold T_m = median(SpO$_2$). Then, additional thresholds were set both below (threshold $T_{m-3} = T_m$-3%) and above (threshold $T_{m+3} = T_m$+3%) the median in order to capture significant changes in SpO$_2$ dynamics related to apnoeic events, i.e. desaturations and re-saturations, respectively. As the number of oxygen desaturations ≥3%,(i.e., ODI3), have been demonstrated to capture discriminant information between children with and without moderate-to-severe OSAHS (Chang *et al* 2013), we used this "step" below and above the median in order to codify relevant changes in SpO$_2$ recordings. Accordingly, the coding scheme is defined as

$$s[n] = \begin{cases} \text{Symbol '4' if} & T_{m+3} < SpO_2[n] \leq 100\% \\ \text{Symbol '3' if} & T_m \quad < SpO_2[n] \leq T_{m+3} \\ \text{Symbol '2' if} & T_{m-3} < SpO_2[n] \leq T_m \\ \text{Symbol '1' if} & 50\% < SpO_2[n] \leq T_{m-3} \end{cases} . \tag{1}$$
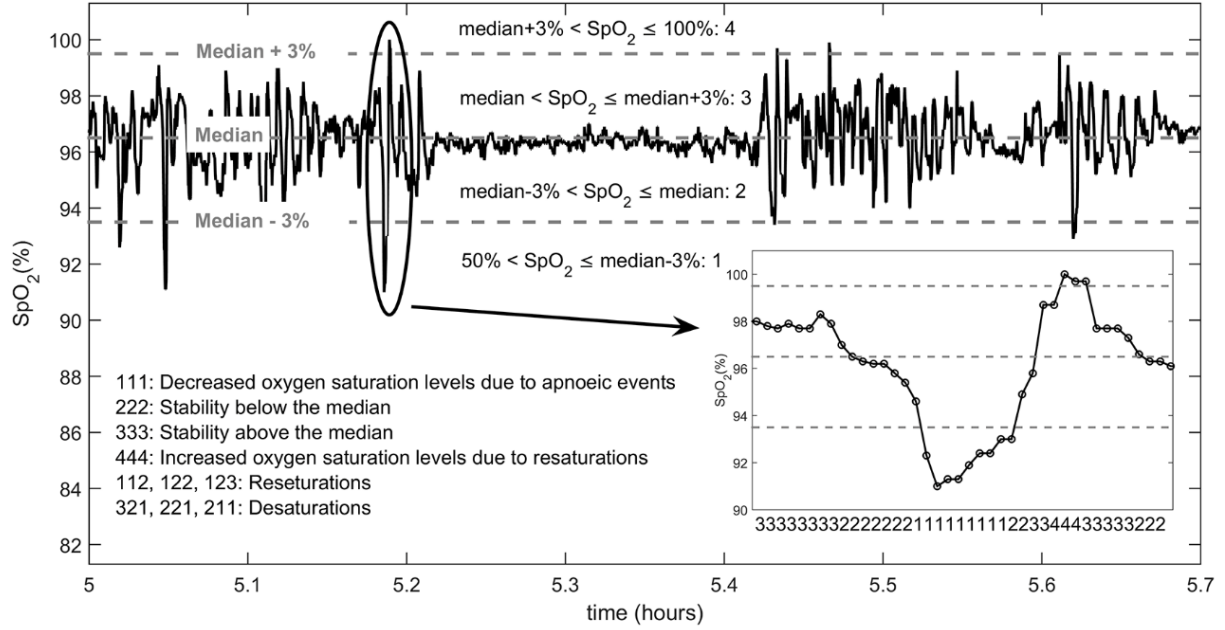
Figure 1. Symbolization process within the framework of symbolic dynamics for a segment of an overnight SpO₂ recording from a common moderate-to-severe (AHI ≥5 events/h) OSAHS patient. The continuous black line represents the overnight SpO₂ recording whereas the dashed grey lines represent the thresholds used to quantise the range of input values in order to obtain symbols.

Once the original time series is transformed into a symbol sequence, symbols are typically grouped to compose words of a fixed length in order to compute the frequencies of occurring words. In this regard, both the number of symbols per word (k) and the length of the sequence (N) must be taken into account to reach a compromise between capturing short-term dynamics and the reliability in estimating the word frequency (Kurths *et al* 1995). Previous studies have reported that words composed of $k=3$ symbols are appropriate to analyse symbolic dynamics of cardiorespiratory signals (Suhrbier *et al* 2010, Kabir *et al* 2011, Ravelo-García *et al* 2014, 2015). Using 3-symbol words and an alphabet of 4 symbols, we obtained a vocabulary W^k composed of $M=64$ ($p^k = 4^3$) unique words, which lead to each single bin in the histogram of words. In the present study, each SpO₂ recording was divided into non-overlapping segments of equal length (N) prior to symbolic dynamic analyses in order to update the thresholds systematically, i.e., the median, throughout the symbolization process. According to Voss *et al* (1996), 20 should be the average minimal number of words per bin (ratio N:M) to accurately estimate the word distribution of a sequence. In order to fit this requirement, we divided each overnight oximetric recording into 25-min length segments, i.e., $N = 1500$ samples ($f_s = 1$ Hz) and hence about 23 words per bin ($\approx 1500/64$) on average.

In order to parameterise the dynamics of the symbol sequence, the following features were computed:

- The percentage or probability of words ($PW_{\{sss\}}$) representative of different states and changes in the signal (Voss *et al* 1996). In this study, we defined the relative frequency of appearance of the word {111} as a measure of decreased oxygen saturation due to desaturations, whereas the word {444} is representative of increased SpO₂ due to re-saturations. Similarly, the probability of the words {222} and {333} estimates the stability of the SpO₂ signal below and above the median value, respectively. Finally, the relative frequency of the appearance of words {112}, {122} and {123} quantifies the presence of re-saturations, while the probability of words {321}, {221} and {211} estimates the presence of desaturations. As mentioned, the probability of each word is computed as the relative frequency of appearance:

$$PW_{\{sss\}} = \frac{\sum_{i=1}^{N-k+1} I_{\{sss\}}}{N-k+1}, \tag{2}$$

where N is the sequence length (1500 symbols), k is the word length, and

$$I_{\{sss\}} = \begin{cases} 1 \text{ if word } \{s[i]\, s[i+1]\, s[i+2]\} = \{sss\} \\ 0 \text{ otherwise} \end{cases} . \tag{3}$$

It is expected that the probability of words characteristic of acute changes in the SpO_2 signal linked with desaturations and subsequent re-saturations, as well as those words representing decreased and increased saturation levels, would be higher in children with moderate-to-severe OSAHS (AHI≥5 group) due to frequent apnoeic events. On the other hand, the probability of words representative of stability around the median is expected to be higher in the AHI<5 group.

- Forbidden words (*FW*), which is the number of words in the vocabulary appearing with a probability of less than 0.001 in the symbol sequence (Kurths *et al* 1995, Voss *et al* 1996). The number of forbidden words is a measure of a stability: the higher the number of forbidden words the higher the stability, i.e., the lower the complexity. In the context of OSAHS, children without the disease or with mild OSAHS are characterised by an overnight oximetry profile showing an almost stable behaviour with small changes around the baseline saturation (lower complexity), i.e., there is a small region of predominant amplitude values. On the other hand, children showing moderate and severe OSAHS are characterised by lower periods of stability and more changes (higher complexity), i.e., the overnight oximetry profile shows a wider range of saturation values (new significant words) due to desaturations and re-saturations. Therefore, *FW* is expected to be higher in children without moderate-to-severe OSAHS.

- Symbolic entropy (*SymbEn*), which is computed as the normalised corrected Shannon's entropy of the symbol sequence (Aziz and Arif 2006). The Shannon's entropy of *k*-th order is computed as follows:

$$H^k = -\sum_{\substack{i=1 \\ w_i^k \in W^k, \\ p(w)>0}}^{M} p(w_i^k) log_2(p(w_i^k)), \tag{4}$$

where $p(w_i^k)$ is the probability density function of the words belonging to the vocabulary W^k composed of k-symbol length words and M is the total number of single words in the vocabulary. Then, the *SymbEn* is computed according to the following equation (Aziz and Arif 2006):

$$SymbEn = \frac{H^k + \frac{C_R - 1}{2MLn2}}{-log_2\left(\frac{1}{M}\right) + \frac{M-1}{2MLn2}}, \tag{5}$$

where C_R is the number of words occurring from the possible M words composing the whole vocabulary. Regarding equation (5), the term added to H^k is a correction to avoid random as well as systematic error or bias in the estimation of the Shannon's entropy. In addition, the normalization makes *SymbEn* vary from 0 to 1 regardless the word length and symbolization scheme. Higher values of *SymbEn* account for higher complexity in the word distribution, i.e., there are more words with significant probability of appearance instead of a dominant word. Accordingly, lower *SymbEn* values are expected for SpO_2 recordings from children without moderate-to-severe OSAHS due to a higher dominance of words representative of a stable behaviour, while higher *SymbEn* is expected in children with AHI ≥5 events/h due to the appearance of more words indicative of desaturations and re-saturations.

The proposed features from symbolic dynamics were computed for each 25-min segment. For each single feature, all segment-based values were averaged to obtain a single measure per SpO_2 recording.

2.2.2. Feature selection and classification

The widely used binary logistic regression (LR) model was applied both for feature selection and for pattern recognition. This conventional statistical classifier has demonstrated good performance in the context of paediatric OSAHS (Gutiérrez-Tobal *et al* 2015a, Álvarez *et al* 2017, Crespo *et al* 2017, 2018). Therefore, we considered LR as an appropriate reference modelling approach to assess the usefulness of symbolic dynamics to characterise overnight SpO_2 recordings.

Regarding variable selection, bidirectional forward stepwise logistic regression (FSLR) is a well-known procedure for LR-based model optimization (Álvarez *et al* 2010, 2013, Gutiérrez-Tobal *et al* 2012, 2015b). As proposed by Hosmer and Lemeshow (2000), FSLR explores the original feature space looking for a reduced as well as representative feature subset. In order to achieve this goal, FSLR iteratively assesses statistical differences between the current model and a candidate one in terms of the likelihood ratio test. Both models differ in just one degree of freedom, i.e., one candidate feature (added or removed in the current iteration). This procedure yields several nested feature subsets, where the most relevant variables are progressively added to the current feature subset (forward selection) while the redundant ones are removed (backward elimination). In addition, a bootstrapping procedure was applied during the feature selection stage in order to obtain a reduced feature subset independent of a particular dataset. According to the bootstrap approach (Witten et al 2011), 1000 bootstrap replicates were derived from the whole dataset by means of resampling with replacement. Then, the FSLR algorithm was applied to each replicate so that 1000 potentially different optimum subsets were obtained. Finally, only variables automatically selected at least 50% of the runs composed our optimum feature space (Hornero *et al* 2017, Vaquerizo-Villar *et al* 2018). In order to avoid overfitting, the optimization of the input feature subset was carried out using only the training bootstrap replicates. The subsequent assessment of the proposed optimum model took into account the remaining instances not involved in the tuning of the algorithm, according to bootstrap 0.632 (Witten *et al* 2011).

Binary LR was used to classify oximetry-derived patterns into the mutually exclusive categories under study: children with AHI <5 *versus* children with AHI ≥5. Every input pattern representing the nocturnal SpO_2 recording of each child consisted of the optimum independent variables from the previous feature selection stage. LR models the probability density of the dependent variable as a Bernoulli distribution, applying the well-known logit function and finally assigning each pattern to the class with the maximum *posterior* probability (Bishop 2006). In the present study, several models were investigated. Firstly, every feature subset (conventional oximetry indices, anthropometric measures, statistical moments, and symbolic dynamics features) was individually assessed. Then, starting from a baseline model composed of conventional oximetry indices, the remaining feature subsets were added in increasing order of individual diagnostic performance. Finally, the "optimum" LR model was built using the features automatically selected by means of FSLR.

2.3. Statistical analysis

Matlab R2015a (The MathWorks Inc., Natick, MA, USA) was used to implement signal processing and pattern recognition algorithms, as well as to perform classification performance analyses. IBM SPSS Statistics 20 (IBM Corp., Armonk, NY, USA) was used for descriptive and statistical analyses. Statistical differences between AHI<5 and AHI≥5 patient groups were assessed by means of the non-parametric Mann-Whitney *U* test. A *p*-value <0.01 was considered statistically significant.

Regarding classification performance assessment, standard metrics derived from binary confusion matrices and receiver operating characteristics (ROC) curves were computed: sensitivity (Se), specificity (Sp), positive (PPV) and negative (NPV) predictive values, positive (LR+) and negative (LR-) likelihood ratios, accuracy (Acc), and area under the ROC curve (AUC). The 95% confidence interval (95% CI) was provided per each performance metric. In order to obtain a proper estimation of each metric, the common bootstrap 0.632 was applied (Witten et al 2011, Gutiérrez-Tobal *et al* 2015a, Álvarez *et al* 2017). Briefly,

given an original dataset of size N, M new datasets of equal size (N) are composed using resampling with replacement with uniform probability distribution, the so-called bootstrap replicates. Accordingly, each bootstrap replicate m_i ($1 \leq i \leq M$) will most likely contain repeated instances, whereas a number of cases from the original dataset are not selected. At each iteration i ($1 \leq i \leq M$), the current replicate m_i is considered the training set in order to fit coefficients of a LR model, whereas cases from the original dataset not included in m_i are used for independent assessment. According to bootstrap 0.632, each performance metric is obtained as a weighted contribution of both the training and the test components, in order to avoid a downward estimation:

$$\text{metric}^{(m_i)} = \left(0.632 \times \text{metric}_{\text{TEST}}^{(m_i)} \right) + \left(0.368 \times \text{metric}_{\text{TRAIN}}^{(m_i)} \right). \quad (6)$$

Finally, each metric is obtained as the average of the M bootstrap estimates:

$$\text{metric} = \frac{1}{M} \sum_{i=1}^{M} \text{metric}^{(m_i)}. \quad (7)$$

The user-dependent parameter M was set to 1000, which is considered appropriate to estimate the 95% CI accurately (Witten et al 2011). The default output cut-off of 0.5 commonly used in LR-based modelling was applied to compute performance metrics in both the training and test stages.

4. Results

Portable SpO$_2$ recordings from the Phone Oximeter were automatically analysed by means of the proposed techniques. Tables 2 – 4 show the median values of the features composing the initial feature set for the children groups under study.

Table 2. Descriptive analysis of conventional oximetry indexes from overnight SpO$_2$ recordings acquired using the Phone Oximeter. Data are presented as median [interquartile range]. The p-value shown in the last column was computed using the non-parametric Mann-Whitney U test. A level of 0.01 was considered for statistical significance (N.S.: non-significant).

Characteristics	AHI<5 group	AHI≥5 group	p-value
Sat_{AVG}	97.78 [1.19]	97.26 [1.49]	N.S.
Sat_{MIN}	89.70 [7.50]	85.90 [9.73]	$p<0.01$
CT95	0.25 [1.14]	2.34 [8.63]	$p<0.01$
ODI2	7.25 [6.72]	17.59 [17.60]	$p<0.01$
ODI3	1.85 [2.09]	6.29 [9.30]	$p<0.01$
ODI4	0.73 [1.00]	2.70 [5.52]	$p<0.01$

Table 3. Descriptive analysis of common statistical moments in the time domain from overnight SpO$_2$ recordings acquired using the Phone Oximeter. Data are presented as median [interquartile range]. The p-value shown in the last column was computed using the non-parametric Mann-Whitney U test. A level of 0.01 was considered for statistical significance (N.S.: non-significant).

Characteristics	AHI<5 group	AHI≥5 group	p-value
M1t	97.78 [1.19]	97.26 [1.49]	N.S.
M2t	0.29 [0.24]	0.61 [0.53]	$p<0.01$
M3t	-0.71 [0.63]	-0.70 [0.60]	N.S.
M4t	4.06 [2.97]	2.84 [2.33]	$p<0.01$

Table 4. Descriptive analysis of all the non-linear features derived from the histogram of words built in the framework of symbolic dynamics analysis of overnight SpO$_2$ recordings acquired using the Phone Oximeter. Data are presented as median [interquartile range]. The *p*-value shown in the last column was computed using the non-parametric Mann-Whitney U test. A level of 0.01 was considered for statistical significance (N.S.: non-significant).

Characteristics	AHI<5 group	AHI≥5 group	*p*-value
FW	56 [1]	51 [4]	*p*<0.01
PW$_{\{111\}}$ (x10^{-3})	0.83 [2.60]	4.67 [10.89]	*p*<0.01
PW$_{\{222\}}$	0.34 [0.040]	0.36 [0.047]	N.S.
PW$_{\{333\}}$	0.47 [0.041]	0.46 [0.036]	*p*<0.01
PW$_{\{444\}}$ (x10^{-3})	0 [0.0]	0 [0.53]	N.S.
PW$_{\{112\}}$ (x10^{-3})	0.28 [0.38]	1.11 [1.75]	*p*<0.01
PW$_{\{122\}}$ (x10^{-3})	0.35 [0.42]	1.24 [1.47]	*p*<0.01
PW$_{\{123\}}$ (x10^{-4})	0.42 [1.23]	1.85 [3.57]	*p*<0.01
PW$_{\{321\}}$ (x10^{-5})	3.51 [7.42]	6.36 [14.84]	N.S.
PW$_{\{221\}}$ (x10^{-3})	0.37 [0.44]	1.34 [1.70]	*p*<0.01
PW$_{\{211\}}$ (x10^{-3})	0.28 [0.36]	1.11 [1.75]	*p*<0.01
SymbEn	0.30 [0.019]	0.31 [0.043]	*p*<0.01

Regarding conventional oximetric indexes (table 2), no significant statistical differences between groups were found for the average saturation, which highlight the challenge of detecting OSAHS from oximetry in paediatric patients. Similarly (table 3), *M1t* (mean, i.e. central tendency) and *M3t* (skewness, i.e., asymmetry) did not reach statistical significance. On the other hand, common ODIs and *M2t* (variance, i.e., dispersion) reached the highest differences (p <0.01) between moderate-to-severe OSAHS patients and children with AHI <5 events/h.

Figure 2 shows the normalised histogram of words composing the vocabulary from symbolic dynamics analysis. A total of 9 out of 12 features derived from the symbolic dynamics approach reached significant statistical differences (p <0.01) between groups (table 4). The number of forbidden words was significantly higher in children with an AHI <5 events/h. In the same way, the normalised corrected *SymbEn* was significantly higher in moderate-to-severe OSAHS patients. The probability of appearance of single words, *PW* of word {111}, which represents a decreased saturation level, and words {211, 221}, indicative of smooth oxygen desaturations, was significantly higher in OSAHS patients due to the presence of apnoeic events. On the contrary, *PW*$_{\{321\}}$, which accounts for the deepest desaturations, did not reach statistical significant differences between groups. Similarly, *PW* of words {112, 122, 123} indicative of oxygen re-saturations aimed at restoring the normal saturation level after apnoeic events was significantly higher in children with OSAHS. *PW*$_{\{444\}}$, which accounts for increased saturation levels due to these re-saturations, was slightly higher in the moderate-to-severe OSAHS group though no statistically significant for the level of 0.01 (p=0.039). Regarding stability around the median, *PW*$_{\{222\}}$ (stability below the median) was not significantly different between groups, whereas *PW*$_{\{333\}}$ (stability above the median) was significantly higher in patients with an AHI <5 events/h.

Table 5 summarises the results from the feature selection stage. A total of 9 out to 24 (37.5%) features were automatically selected. *Sat$_{MIN}$*, *ODI3*, *BMI*, *M2t*, *M3t*, *M4t*, *PW*$_{\{333\}}$, *PW*$_{\{444\}}$ and *PW*$_{\{112\}}$ composed the optimum feature subset. All the approaches proposed to characterise the population under study were represented: conventional oximetric indices (*Sat$_{MIN}$*, *ODI3*); anthropometric measures (*BMI*); statistical moments (*M2t*, *M3t*, *M4t*); and parameters from symbolic dynamics (*PW*$_{\{333\}}$, *PW*$_{\{444\}}$, *PW*$_{\{112\}}$); which suggests their relevancy as well as complementarity in the characterization of children with and without moderate-to-severe OSAHS.
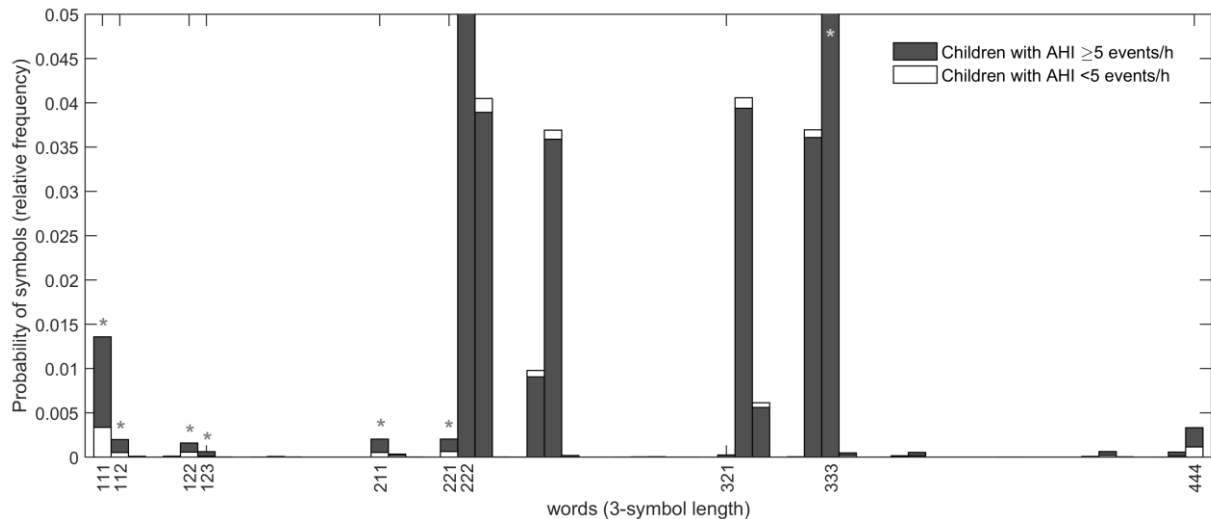
Figure 2. Normalised histograms (probability density function) of 3-symbol length words from symbolic dynamics analysis for children with AHI <5 events/h (white) and patients with moderate-to-severe OSAHS (dark grey). The words proposed to characterise changes in the oximetric pattern due to apnoeic events are labelled and those showing significant statistical differences between groups are highlighted (grey asterisk).

Table 5. Optimum feature subset from FSLR using a bootstrapping approach (1000 repetitions). Features automatically selected 50% of runs composed the proposed final feature subset.

Feature subset	Feature	Nº of times selected
Conventional oximetry indices (Oxi)	Sat_{MIN}	802
	ODI3	575
Anthropometric (Anthr)	BMI	734
First-to-fourth statistical moments (Stats)	M2t	584
	M3t	809
	M4t	926
Symbolic dynamics (Symb)	$PW_{\{333\}}$	581
	$PW_{\{444\}}$	875
	$PW_{\{112\}}$	562

Tables 6 and 7 summarise the performance assessment of different LR models composed of the proposed features. Individually, anthropometric (Anth) and statistical (Stats) feature subsets reached the lowest diagnostic performance, achieving 65.1% Acc (0.65 AUC) and 69.7% Acc (0.75 AUC), respectively (table 6). Conventional oximetric indices (Oxi) achieved moderate accuracy (73.2% Acc and 0.76 AUC) with a highly unbalanced sensitivity-specificity pair (49.5% Se *vs.* 88.2% Sp), whereas the feature subset from symbolic dynamics (Symb) reached the highest performance individually, with 78.4% Acc (65.2% Se and 86.8% Sp) and 0.83 AUC. Regarding the combination of the proposed feature subsets (table 7), the LR model composed of the 5 common indices from oximetry (LR_{Oxi}) was the baseline (reference) model for comparison purposes. When age and BMI were added as input features to the baseline model ($LR_{Oxi+Anthr}$), a non-significant (*p*-value >0.01) performance improvement was obtained: 74.0% Acc and 0.77 AUC was reached. On the contrary, a statistically significant (*p*-value <0.01) diagnostic performance improvement was obtained by adding features from automated signal processing approaches, particularly from symbolic dynamics. Using common statistical moments, the accuracy of the model $LR_{Oxi+Anthr+Stats}$ increased to 77.9% Acc (0.83 AUC). In the same way, adding features from symbolic dynamics, the model $LR_{Oxi+Anthr+Stats+Symb}$ reached higher accuracy (80.3% Acc and 0.85 AUC) and more balanced sensitivity-specificity pair (75.0% Se *vs.* 83.9% Sp). Finally, the highest performance was obtained using the optimum feature subset (LR_{OPT}), leading to 83.3% Acc and 0.89 AUC.

Table 6. Diagnostic ability of each individual feature subset under study using binary LR modelling for moderate-to-severe paediatric OSAS detection and a bootstrap approach.

Feature subset	Se (%)	Sp (%)	PPV (%)	NPV (%)	LR+	LR-	Acc (%)	AUC
Oxi	49.5	88.2	73.4	73.6	5.20	0.57	73.2	0.76
	(28.9, 71.8)	(76.3, 98.4)	(53.2, 95.9)	(62.6, 85.3)	(2.21, 12.61)	(0.34, 0.79)	(64.0, 81.7)	(0.63, 0.86)
Anthr	31.2	86.8	65.7	66.8	3.02	0.79	65.1	0.65
	(4.7, 59.7)	(63.5, 99.6)	(37.3, 96.6)	(56.1, 78.4)	(1.05, 11.36)	(0.55, 1.01)	(55.0, 75.1)	(0.54, 0.77)
Stats	50.4	82.0	64.8	72.6	3.29	0.60	69.7	0.75
	(22.6, 76.2)	(64.8, 95.9)	(45.3, 86.0)	(60.2, 85.5)	(1.53, 7.79)	(0.31, 0.88)	(59.4, 79.6)	(0.64, 0.85)
Symb	65.2	86.8	76.2	79.9	6.88	0.41	78.4	0.83
	(46.2, 84.5)	(73.9, 96.7)	(59.2, 92.8)	(69.1, 90.8)	(2.86, 17.39)	(0.19, 0.62)	(68.3, 87.1)	(0.73, 0.92)

Table 7. Diagnostic performance of different LR models from joint feature subsets from oximetry.

Feature subset	Se (%)	Sp (%)	PPV (%)	NPV (%)	LR+	LR-	Acc (%)	AUC
Baseline (Oxi)	49.5	88.2	73.4	73.6	5.20	0.57	73.2	0.76
	(28.9, 71.8)	(76.3, 98.4)	(53.2, 95.9)	(62.6, 85.3)	(2.21, 12.61)	(0.34, 0.79)	(64.0, 81.7)	(0.63, 0.86)
Oxi + Anthr	53.9	86.9	72.9	75.0	5.34	0.53	74.0	0.77
	(33.8, 73.2)	(74.0, 96.9)	(52.6, 91.5)	(64.3, 85.8)	(2.14, 14.76)	(0.31, 0.77)	(64.0, 83.2)	(0.66, 0.87)
Oxi + Anthr + Stats	64.6	86.5	75.7	79.5	6.81	0.41	77.9	0.83
	(45.1, 83.6)	(72.4, 96.8)	(57.4, 92.8)	(68.6, 90.7)	(2.67, 18.31)	(0.20, 0.62)	(67.9, 86.5)	(0.74, 0.92)
Oxi + Anthr + Stats + Symb	75.0	83.9	75.5	84.0	12.17	0.32	80.3	0.85
	(56.4, 91.9)	(68.8, 95.9)	(58.0, 92.6)	(72.7, 94.8)	(3.23, 37.57)	(0.12, 0.55)	(69.9, 89.9)	(0.71, 0.95)
Optimum subset (OPT)	73.5	89.5	82.0	84.3	10.40	0.30	83.3	0.89
	(56.1, 89.7)	(77.7, 99.2)	(65.8, 98.7)	(73.9, 93.9)	(3.78, 27.77)	(0.12, 0.50)	(74.4, 91.0)	(0.79, 0.96)

Figure 3 shows the ROC curves of the two components of the proposed bootstrap 0.632 approach (training and test contributions based on resampling with replacement) for every model. Both the "complete" model (composed of all 24 variables) and the optimum model from FSLR reached an AUC >0.90 in the bootstrap training group. In the bootstrap test group, the proposed optimum model showed the highest AUC as well as the lower performance decrease, suggesting no influence of overfitting and higher generalization ability.

5. Discussion

In this study, advanced signal processing algorithms and portable technologies, by means of symbolic dynamics and smartphones, are combined to develop a reliable screening tool for paediatric OSAHS. The main novelty of this study is that symbolic dynamics was applied to analyse complex non-linear changes due to intermittent desaturations present in the oximetry signal of children obtained with the Phone Oximeter. The normalised histogram of words has been found to provide discriminant variables able to enhance the detection of moderate-to-severe paediatric OSAHS from oximetry. Particularly, words representing increased values of oximetry and re-saturations after apnoeic events were automatically selected to compose the optimum feature subset. The proposed optimum LR model reached 83.3% Acc (73.5% Se and 89.5% Sp) and 0.89 AUC, significantly improving the diagnostic performance of models composed of conventional variables.

Regarding the features derived from symbolic dynamics (table 4), there were five words with low probability of appearance (<0.001, i.e., forbidden words) within SpO_2 recordings from children with AHI <5 events/h whose probability of appearance increased in patients with AHI ≥5 events/h, ceasing to be forbidden words in this group (median *FW*: 56 *versus* 51, respectively): {211}, {221}, {111}, {112}, {122}. These words are representative of desaturations and re-saturations due to apnoeic events and reached significant differences between both patient groups, which suggest the appropriateness of the proposed features to parameterise the histogram of words.
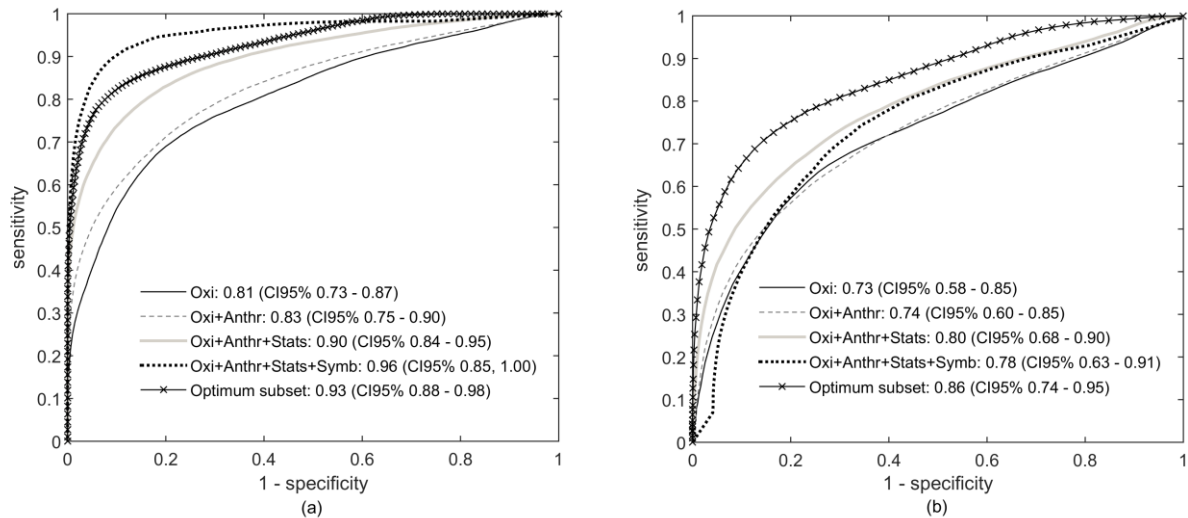
Figure 3. ROC curves for the models composed of conventional variables and for the proposed optimum model from FSLR feature selection for (a) training and (b) test contributions from the bootstrap 0.632 procedure.

All the words proposed to parameterise re-saturations in the framework of symbolic dynamics achieved significant statistical differences ($p < 0.01$) between children with an AHI <5 events/h and moderate-to-severe OSAHS patients, from the slowest {112, 122} to the fastest re-saturation {123}. This would suggest that the recovery process towards a normal saturation plays a major role in children having moderate-to-severe OSAHS. Conversely, the word representing the deepest desaturation {321} did not reach statistical significant differences, showing that the magnitude of sudden desaturations is not as discriminant as the re-saturations in the detection of the disease in children.

Variables composing the optimum feature subset provide more insight into the relevance of re-saturations and symbolic dynamics. A total of 3 variables from symbolic dynamics were automatically selected: $PW_{\{112\}}$, $PW_{\{333\}}$ and $PW_{\{444\}}$. All the words were indicative of re-saturations or increased saturation values: the word {112} is a small slow re-saturation while the word {333} represents saturation values above the median and the word {444} notably higher saturation levels (close to 99%).

In the optimum feature subset, measures from all the approaches proposed to characterise children suspected of suffering from OSAHS were selected, which highlights the complimentary nature of symbolic dynamics and conventional features. Conventional oximetry indices Sat_{MIN} and $ODI3$ account for the magnitude and the number of desaturations while novel $PW_{\{112\}}$, $PW_{\{333\}}$, and $PW_{\{444\}}$ from non-linear symbolic dynamics quantify the relative number of words representing increasing slope and higher saturation levels. In a complementary way, $M2t$, $M3t$, and $M4t$ provided overall measures of dispersion (variability, asymmetry, and concentration, respectively) of saturation values due to recurrent apnoeic events. In addition, the BMI was also automatically selected, which provides information on the physical status of individual children.

Previous studies applied symbolic dynamics in the context of adult OSAHS from HRV analysis. Ravelo-García *et al* (2014) first proposed the use of features from symbolic dynamics in order to enhance automated classification of adult patients. An alphabet of four symbols was used for quantization of the HRV time series and symbols were subsequently grouped to compose 3-symbol length words. According to previous studies (Kurths *et al* 1995), the percentage of words that contain symbols '1' and '3' (*WPSUM13*), which is indicative of increased HRV, was proposed to parameterise the histogram of words. The authors reported that the classification performance of a LR model composed of both clinical (Epworth sleepiness score and intensity of snoring) and physical (age and neck circumference) variables significantly increased from 0.907 AUC (87.1% Se and 80.0% Sp) to 0.941 AUC (88.7% Se and 82.9% Sp) when adding *WPSUM13* to the model. In a recent study by the same group (Ravelo *et al* 2015), the

authors assessed a linear discriminant analysis (LDA) classifier composed of common time- and frequency-domain oximetry indices and both linear and non-linear features from HRV. Likewise, words of length 3 symbols within a sequence derived from a 4-level coding scheme were used to investigate non-linear dynamics of HRV time series. Then, the probability of words indicative of increased and decreased complexity, symbolic entropy, symbol variability, and the number of forbidden words, were used to parameterise the histogram of words. The LDA model reached 86.5% Acc (75.6% Se and 91.0% Sp) using only oximetric indices in an epoch-based binary classification task, whereas the performance slightly increased to 86.9% (73.4% Se and 92.3% Sp) when linear and non-linear features from HRV were added to the model. In the present research, our results suggest that symbolic dynamics is also a reliable methodology able to increase the diagnostic ability of portable overnight oximetry in the context of childhood OSAHS detection.

Tables 8 and 9 summarise the state-of-the-art in the framework of childhood OSAHS detection from nocturnal oximetry. Previous studies mainly focused on the screening ability of ODIs (Kirk *et al* 2003, Tsai *et al* 2013), sometimes including additional data from the clinical history to increase diagnostic performance (Chang *et al* 2013). Using a common cut-off of 5 events/h for OSAHS, in these studies, accuracy ranged between 64.0% (automated scoring) and 85.1% (manual scoring). Similarly, the number and severity of clusters of desaturations (visual inspection) was assessed as abbreviated screening tool (Brouillette et al 2000, Velasco *et al* 2013, Van Eyck *et al* 2015, Villa *et al* 2015). Using a conservative cut-off of 1 or 2 events/h for positive OSAHS, accuracy ranged 64.7% to 93.4%, whereas 69.4% Acc was reached using an AHI ≥5 events/h.

Table 8. Summary of the state-of-the-art in the context of paediatric OSAHS diagnosis using conventional desaturation indices from oximetry.

Authors (year)	Dataset	Gold standard (cut-off)	Aim and setting	Oximetry inspection approach	Classification approach	Se (%)	Sp (%)	Acc (%)
Brouillette *et al* (2000)	349 children with suspected OSAHS	In-lab PSG (AHI ≥1)	Binary classif. / In-lab oximetry	Nº of clusters of desaturations >3 + Nº drops <90% ≥3	Visual inspection	42.9	97.8	64.7
Kirk *et al* (2003)	58 children with suspected OSAHS	In-lab PSG (AHI ≥5)	Binary classif. / At-home oximetry	Automated ODI3	ODI3 ≥ 5	66.7	60.0	64.0
Chang *et al* (2013)	141 children with suspected OSAHS	In-lab PSG (AHI ≥5)	Binary classif. / Questionnaires and oximetry	Presence of mouth breathing, restless sleep, ODI4	LR	60.0	86.0	71.6
Velasco-Suárez *et al* (2013)	167 children with suspected OSAHS	In-lab PSG (AHI ≥1)	Binary classif. / In-lab oximetry	Nº of clusters of desaturations >2 + Nº drops <90% >1	Visual inspection	86.6	98.9	93.4
Tsai *et al* (2013)	148 children with suspected OSAHS	In-lab PSG (AHI ≥1, 5, 10)	Binary classif. / In-lab oximetry	Manual ODI4	ODI4 >2.05 (AHI ≥1) ODI4 >3.50 (AHI ≥5) ODI4 >4.15 (AHI ≥10)	77.7 83.8 89.1	88.9 86.5 86.0	79.0 85.1 87.1
Van Eyck *et al* (2015)	130 obese children with suspected OSAHS	In-lab PSG (AHI ≥2)	Binary classif. / In-lab oximetry	Brouillette criteria Velasco criteria	Manual scoring of desaturations	58 66	88 69	78 68
Villa *et al* (2015)	268 children with suspected OSAHS	In-lab PSG (AHI ≥1, 5)	Binary classif. / In-lab oximetry	Clusters of desaturations and clinical history	Semi-automatic - AHI ≥1 - AHI ≥5	 91.6 40.6	 40.6 97.9	 85.8 69.4

Recent studies focused on increasing the diagnostic capability of oximetry using automated signal processing and machine learning techniques (Garde *et al* 2014, Álvarez *et al* 2017, Hornero *et al* 2017, Crespo *et al* 2017, 2018, Vaquerizo-Villar *et al* 2018). Time- and frequency-domain statistics, spectral features, and non-linear measures (approximate or sample entropy, central tendency measure, and Lempel-Ziv complexity) usually compose a wide initial feature set from oximetry, which is subsequently optimised using automated feature selection techniques. In these studies, diagnostic accuracies ranged between 78.5% and 82.5% (sensitivity: 68.2%-82.2%; specificity: 83.6%-91.4%) when a cut-off of 5 events/h was set to confirm OSAHS and using information from the SpO_2 signal alone.

Although the widely known non-linear behaviour of biological systems induces a relevant component in biomedical signals, no features from non-linear analysis composed the optimum feature subset in the works by Garde *et al* (2014) and Hornero *et al* (2017). Similarly, only sample entropy was selected in the studies by Álvarez *et al* (2017) and Crespo *et al* (2018), suggesting that novel non-linear techniques could be applied to properly parameterise non-linear dynamics of the oximetry signal. In this regard, Crespo *et al* (2017) demonstrated the reliability of multiscale entropy to characterise non-linear patterns present in the nocturnal oximetry signal. Similarly, Vaquerizo-Villar *et al* (2018) used the bispectrum to quantify deviations from linearity in the SpO_2 signal linked with apnoeic events. This approach has been found to provide relevant and non-redundant information to conventional time- and frequency-domain methods.

Table 9. Summary of the state-of-the-art in the context of paediatric OSAHS diagnosis applying advanced signal processing and pattern recognition techniques to overnight oximetry recordings.

Authors (year)	Dataset	Gold standard (cut-off)	Aim and setting	Signal processing approaches	Pattern recognition approach	Se (%)	Sp (%)	Acc (%)
Garde *et al* (2014)	146 children with suspected OSAHS	In-lab PSG (AHI ≥5)	Binary classif. / Portable oximetry (attended)	Time and spectral: - SpO2 - SpO2 + PR	LDA LDA	80.0 88.4	83.9 83.6	78.5 84.9
Álvarez *et al* (2017)	50 children with suspected OSAHS	In-lab PSG (AHI ≥1, 3, 5)	Binary classif. / Port. oximetry from at-home RP	Statistical, spectral and nonlinear features	LR (AHI ≥1) LR (AHI ≥3) LR (AHI ≥5) (bootstrapping)	89.6 82.9 82.2	71.5 84.4 83.6	85.5 83.4 82.8
Crespo *et al* (2017)	50 children with suspected OSAHS	In-lab PSG (AHI ≥3)	Binary classif. / Port. oximetry from at-home RP	Nonlinear features and conventional oximetric indices	LR (bootstrapping)	84.5	83.0	83.5
Hornero *et al* (2017)	4191 habitually snoring children	In-lab PSG (AHI ≥1, 5, 10)	AHI Estimation / In-lab oximetry	Statistical, spectral, nonlinear features, and ODI3	MLP ANN: - AHI ≥1 - AHI ≥5 - AHI ≥10	84.0 68.2 68.7	53.2 87.2 94.1	75.2 81.7 90.2
Vaquerizo-Villar *et al* (2018)	298 habitually snoring children	In-lab PSG (AHI ≥1, 5, 10)	Multiclass classif / In-lab oximetry	Bispectrum, PSD, ODI3, age, sex, BMI	3-class MLP: - AHI ≥5 - AHI ≥10	61.8 60.0	97.6 94.5	81.3 85.3
Crespo *et al* (2018)	176 children with suspected OSAHS	In-lab PSG (AHI ≥1, 3, 5)	Binary classif. / In-lab oximetry	Statistical, spectral and nonlinear features	LR (AHI ≥1) LR (AHI ≥3) LR (AHI ≥5) (bootstrapping)	93.9 74.6 70.0	37.8 81.7 91.4	84.3 77.7 82.7
Our proposal	142 children with suspected OSAHS	In-lab PSG (AHI ≥5)	Binary classif. / Portable oximetry (attended)	BMI, Age, statistical moments, desaturation indices, symbolic dynamics	LR (AHI ≥5) (bootstrapping)	**73.5**	**89.5**	**83.3**

Our optimum model involving non-linear features from symbolic dynamics outperformed several previous approaches for automated detection of paediatric OSAHS. However, major methodological differences between studies are present and need to be discussed. In the study by Hornero *et al* (2017), PSG-derived SpO_2 recordings from a large cohort of habitually snoring children were processed. On the other hand, Álvarez *et al* (2017) and Crespo *et al* (2017) analysed oximetry recordings from portable unattended respiratory polygraphy at children's home. Regarding automated pattern recognition, Hornero *et al* (2017) used a regression artificial neural network to estimate the AHI and subsequently set a cut-off of 5 events/h, while Crespo *et al* (2017) used a clinical threshold of 3 events/h to diagnose OSAHS. Similarly, a multi-class artificial neural network was proposed by Vaquerizo *et al* (2018) to classify children into non-OSAHS, moderate, and severe OSAHS. In the study by Garde et al (2014), the diagnostic accuracy increased up to 84.9% (88.4% sensitivity and 83.6% specificity) when features from SpO_2 and pulse rate variability were used jointly.

Concerning the signal acquisition system, in the present study we used the Phone Oximeter in order to increase accessibility to diagnostic resources while decreasing intrusiveness for children. The use of portable technologies is a major novelty in the framework of paediatric sleep medicine. Particularly, the use of smartphones and telemedicine applications aimed at providing unattended testing at home and therapy monitoring is gaining popularity, owing the increasing recognition of both the prevalence and the impact of OSAHS (Singh *et al* 2015). However, the vast majority of smartphone-based tools are mobile apps oriented to adult users in the area of wellness and lifestyle with inadequate scientific validation against standardised methodologies (Behar *et al* 2013, Ko *et al* 2015, Penzel *et al* 2018). On the contrary, in this proposal we developed and properly validated a novel signal-processing module for the Phone Oximeter able to enhance its screening functionalities for paediatric sleep apnoea detection.

Some limitations should be considered in order to generalise our conclusions. A larger dataset would be needed for external validation and extensive and universal optimization of the proposed methodology. Nonetheless, a bootstrapping procedure was applied both for feature selection and for pattern recognition in order to minimise the effect of a limited sample size. Similarly, a larger and independent dataset would allow for a better characterisation of changes in the overnight oximetry profile by means of symbolic dynamics. Nevertheless, our results revealed consistent and significant differences between histograms of symbol sequences of children with AHI <5 events/h and moderate-to-severe OSAHS patients. Concerning the proposed pattern recognition approach, this study focused on binary classification, which is able to derive accurate screening protocols for the disease. In this regard, a cut-off of 5 events/h was used for positive OSAHS diagnosis, i.e., we focused on moderate-to-severe cases. This is a clinically relevant threshold because it is used by paediatricians to recommend surgical treatment routinely. Moderate-to-severe OSAHS is less likely to resolve spontaneously. Furthermore, children showing an AHI ≥5 events/h suffer from the most negative consequences, including an increased cardiovascular risk (Marcus *et al* 2012, Kaditis *et al* 2016b). Notwithstanding, it would be interesting to detect additional categories of severity, i.e., non-OSAHS, mild, moderate, and severe.

Regarding context-dependent parameter tuning in the symbolic dynamics framework, we took into account both the nature and the sampling frequency of oximetry during the symbolization process, which led to a histogram of words showing significant differences between the groups under study. Based on previous evidence (Ravelo-García *et al* 2014, 2015), we adopted the values of the parameters used in similar studies as appropriate for assessing non-linear components of oximetry by means of symbolic dynamics in the same context. As previously reported by Voss *et al* (1996), it is important to point out that small changes of the threshold values (*m* and *a*) do not influence the results considerably. Although *m* is usually set to the mean of the time series, we selected the median because it is more robust against artefacts. Additionally, the same optimum features were systematically selected by FSLR using the proposed bootstrapping approach despite slight changes in *a* (2.8 to 3.2 in steps of 0.1), which shows the robustness of the methodology against small changes in the tuning parameters. Nevertheless, in this study we investigated a particular scheme: the probability distribution of words of length 3 from an alphabet of 4 symbols. Therefore, future work could be aimed at optimizing the tuning parameters (*m*, *a*, *p*, *k*, and *N*)

and searching for novel discriminant words to be able to maximise the performance of symbolic dynamics in the context of SpO_2 recordings from children showing OSAHS symptoms. In addition, it would be interesting to assess the dependence of the proposed algorithm on some technical features of the recording, such as the sampling frequency, the averaging time, the threshold for artefact removal used in the pre-processing stage, or the minimum number of technically adequate recording hours.

6. Conclusion

Using the Phone Oximeter to accomplish portable nocturnal oximetry, we investigated the ability of symbolic dynamics to discriminate between children with and without moderate-to-severe OSAHS. The histogram of words from the symbolization process showed significant differences between children with AHI <5 events/h and moderate-to-severe OSAHS patients. We found that changes in the oximetry dynamics associated with increased saturation levels and re-saturations after apnoeic events may be important in the detection of OSAHS. Our results suggest that features from the histogram of symbols complement oximetry indices commonly focused on the number and severity of desaturations. An optimum feature subset composed of conventional linear measures in the time domain and novel non-linear features from symbolic dynamics significantly improved the diagnostic performance of overnight oximetry. Therefore, automated analysis of portable nocturnal oximetry by means of symbolic dynamics is able to increase the diagnostic capability of smartphone-based screening tests in order to provide available, as well as accurate detection of moderate-to-severe OSAHS in children.

Acknowledgements

References

Álvarez D *et al* 2010 Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis *IEEE Trans. Biomed. Eng.* **57** 2816–24

Álvarez D *et al* 2013 Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of sleep apnea diagnosis *Int. J. Neural Syst.* **23** 1–18

Álvarez D *et al* 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** 693–702

Aziz W and Arif M 2006 Complexity analysis of stride interval time series by threshold dependent symbolic entropy *Eur. J. Appl. Physiol.* **98** 30–40

Baumert M *et al* 2015 Symbolic dynamics of pulse transit time and heart period in children with upper airway obstruction *37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 1801–04

Behar J *et al* 2013 A review of current sleep screening applications for smartphones *Physiol. Meas.* **34** R29–6

Bishop C M 2006 *Pattern recognition and machine learning* (New York, USA: Springer-Verlag)

Brouillette R T *et al* 2000 Nocturnal pulse oximetry as an abbreviated testing modal- ity for pediatric obstructive sleep apnea *Pediatrics* **105**

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

Crespo A *et al* 2017 Multiscale Entropy Analysis of Unattended Oximetric Recordings to Assist in the Screening of Paediatric Sleep Apnoea at Home *Entropy* **19** 284

Crespo A *et al* 2018 Assessment of oximetry-based statistical classifiers as simplified screening tools in the management of childhood obstructive sleep apnea *Sleep Breath.* In Press

Daw C S, Finney C E A and Tracy E R 2003 A review of symbolic analysis of experimental data. *Review of Scientific Instruments* **74** 915

DeHaan K L *et al* 2015 Polysomnography for the Diagnosis of Sleep Disordered Breathing in Children Under 2 Years of Age *Pediatr. Pulmonol.* **50** 1346–53

Garde A *et al* 2014 Development of a screening tool for sleep disordered breathing in children using the Phone OximeterTM *PLoS One* **9** e112959

Gutiérrez-Tobal G C *et al* 2015a Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home *Biomed. Signal Process. Control* **18** 401–7

Gutiérrez-Tobal GC *et al* 2015b Assessment of time and frequency domain entropies to detect sleep apnoea in heart rate variability recordings from men and women *Entropy* **17** 123–41

Gutiérrez-Tobal G C *et al* 2012 Linear and nonlinear analysis of airflow recordings to help in sleep apnoea–hypopnoea syndrome diagnosis *Physiol. Meas.* **33** 1261–75

Guzzetti S *et al* 2005 Symbolic Dynamics of Heart Rate Variability A Probe to Investigate Cardiac Autonomic Modulation *Circulation* **112** 465–70

Hornero R *et al* 2017 Nocturnal oximetry-based evaluation of habitually snoring children Am. J. Respir. Crit. Care. Med. **196** 1591–98

Hosmer D W and Lemeshow S 2000 *Applied Logistic Regression* (London, UK: John Wiley & Sons)

Hudson J *et al* 2012 Usability testing of a prototype phone oximeter with healthcare providers in high- and low-medical resource environments. *Anaesthesia* **67** 957–67

Iber C *et al* 2007 *The AASM Manual for the scoring of sleep and associated events. Rules, terminology and technical specifications* 1st Ed (Westchester, Illinois: American Academy of Sleep Medicine)

Immanuel S A *et al* 2014 Symbolic dynamics of respiratory cycle related sleep EEG in children with sleep disordered breathing *36th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 6016-6019

Kabir M M *et al* 2011 Quantification of cardiorespiratory interactions based on joint symbolic dynamics *Ann. Biomed. Eng.* **39** 2604–14

Kaditis A G *et al* 2016a Obstructive sleep disordered breathing in 2- to 18-year-old children: diagnosis and management *Eur. Respir. J.* **47** 69–94

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

Karlen W *et al* 2011 Human-centered phone oximeter interface design for the operating room. In *Proc. Int. Conf. Health Inf. Traver V, Fred A, Filipe J, Gamboa H, editors* **2011** 433–8

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

Kheirandish-Gozal L 2010 What is "abnormal" in pediatric sleep? *Respir. Care* **55** 1366–76

Kirk V G *et al* 2003 Comparison of home oximetry monitoring with laboratory polysomnography in children *Chest* **124** 1702–08

Kirk V *et al* 2017 American Academy of Sleep Medicine position paper for the use of a home sleep apnea test for the diagnosis of OSA in children **J. Clin. Sleep Med. 13** 1199–203

Ko P-R T *et al* 2015 Consumer Sleep Technologies: A Review of the Landscape *Clin. Sleep Med.* **11** 1455–61

Kurths J *et al* 1995 Quantitative analysis of heart rate variability *Chaos* **5** 88–94

Lesser D J 2012 The utility of a portable recording device for screening of obstructive sleep apnea in obese adolescents *J. Clin. Sleep Med.* **8** 271–7

Marcus C L *et al* 2012 Diagnosis and management of childhood obstructive sleep apnea syndrome *Pediatrics* 2012 **130** e714–55

Penzel T *et al* 2016 Modulations of heart rate, ECG, and cardiorespiratory coupling observed in polysomnography *Front. Physiol.* **7** 460

Penzel T, Schöbel C and Fietze I 2018 New technology to assess sleep apnea: wearables, smartphones, and accessories *F1000 Res.* **7** 413

Petersen C L *et al* 2013 Design and evaluation of a low-cost smartphone pulse oximeter *Sensors* **13** 16882–93

Ravelo-García A G *et al* 2014 Symbolic dynamics marker of heart rate variability combined with clinical variables enhance obstructive sleep apnea screening *Chaos* **24**, 024404

Ravelo-García A G *et al* 2015 Oxygen saturation and RR intervals feature selection for sleep apnea detection *Entropy* **17** 2932–57

Singh J *et al* 2015 American Academy of Sleep Medicine (AASM) position paper for the use of telemedicine for the diagnosis and treatment of sleep disorders *J. Clin. Sleep Med.* **11** 1187 –98

Suhrbier A *et al* 2010 Cardiovascular regulation during sleep quantified by symbolic coupling traces *Chaos* **20** 045124

Tsai C M *et al* 2013 Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children *Int. J. Pediatr. Otorhinolaryngol.* **77** 1286–90

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

Van Eyck A and Verhulst S L 2018 Improving the diagnosis of obstructive sleep apnea in children with nocturnal oximetry-based evaluations *Expert Rev. Respir. Med.* **12** 165–7

Vaquerizo-Villar F et al 2018 Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings *Comput. Meth. Prog. Biomed.* **156** 141–9

Velasco C T *et al* 2013 Pulse oximetry recording in children with adenotonsillar hypertrophy: usefulness in the diagnosis of obstructive sleep apnea syndrome *Arch. Argent. Pediatr.* **111** 196–201

Verbraecken J 2016 Telemedicine applications in sleep disordered breathing. Thinking out of the box *Sleep Med. Clin.* **11** 445–59

Villa M P *et al* 2015 Diagnosis of pediatric obstructive sleep apnea syn- drome in settings with limited resources *JAMA Otolaryngol. Neck Surg.* **141** 990–6

Voss A *et al* 1996 The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death *Cardiovasc. Res.* **31** 419–33

Witten I H, Frank E and Hall M A 2011 *Data Mining Practical Machine Learning Tools and Techniques* (Burlington, MA, USA: Morgan Kaufmann, Elsevier)

Yeragani V K *et al* 2000 Nonlinear measures of heart period variability: decreased measures of symbolic dynamics in patients with panic disorder *Depress. Anxiety* **12** 67–77