Multiscale Entropy Analysis of Unattended Oximetric Recordings to Assist in the Screening of Paediatric Sleep Apnoea at Home

Andrea Crespo [1,2,*], Daniel Álvarez [1,2], Gonzalo C. Gutiérrez-Tobal [2],
Fernando Vaquerizo-Villar [2], Verónica Barroso-García [2], María L. Alonso-Álvarez [3],
Joaquín Terán-Santos [3], Roberto Hornero [2] and Félix del Campo [1,2]

[1]   Sleep-Ventilation Unit, Pneumology Service, Río Hortega University Hospital, c/Dulzaina 2,
      Valladolid 47012, Spain; dalvgon@gmail.com (D.A.); fsas@telefonica.net (F.d.C.)
[2]   Biomedical Engineering Group, University of Valladolid, Paseo de Belén 15, Valladolid 47011, Spain;
      gguttob@gmail.com (G.C.G.-T.); fernando.vaquerizo@gib.tel.uva.es (F.V.-V.);
      veronica.barroso@gib.tel.uva.es (V.B.-G.); robhor@tel.uva.es (R.H.)
[3]   Unidad Multidisciplinar de Sueño, CIBER Respiratorio, Hospital Universitario de Burgos, Burgos 09006,
      Spain; mlalonso@hubu.es (M.L.A.-A.); jteran@hubu.es (J.T.-S.)
*   Correspondence: andreacresposedano@gmail.com; Tel.: +34-983-420-400 (ext. 85776)

Abstract: Untreated paediatric obstructive sleep apnoea syndrome (OSAS) can severely affect the development and quality of life of children. In-hospital polysomnography (PSG) is the gold standard for a definitive diagnosis though it is relatively unavailable and particularly intrusive. Nocturnal portable oximetry has emerged as a reliable technique for OSAS screening. Nevertheless, additional evidences are demanded. Our study is aimed at assessing the usefulness of multiscale entropy (MSE) to characterise oximetric recordings. We hypothesise that MSE could provide relevant information of blood oxygen saturation (SpO_2) dynamics in the detection of childhood OSAS. In order to achieve this goal, a dataset composed of unattended SpO_2 recordings from 50 children showing clinical suspicion of OSAS was analysed. SpO_2 was parameterised by means of MSE and conventional oximetric indices. An optimum feature subset composed of five MSE-derived features and four conventional clinical indices were obtained using automated bidirectional stepwise feature selection. Logistic regression (LR) was used for classification. Our optimum LR model reached 83.5% accuracy (84.5% sensitivity and 83.0% specificity). Our results suggest that MSE provides relevant information from oximetry that is complementary to conventional approaches. Therefore, MSE may be useful to improve the diagnostic ability of unattended oximetry as a simplified screening test for childhood OSAS.

Keywords: paediatric obstructive sleep apnoea syndrome; unattended oximetry; multiscale entropy; automated pattern recognition

1. Introduction

Paediatric obstructive sleep apnoea syndrome (OSAS) is a sleep-related breathing disorder characterised by intermittent and repetitive episodes of partial or complete collapse of the child's upper airway while sleeping [1]. Recurrent apnoeic events lead to gas exchange abnormalities and sleep disruption [2], which may cause major long-term adverse consequences in several body systems, such as neuropsychological and cognitive deficits, cardiovascular and metabolic dysfunction, and growth impairment [1–3]. Consequently, this condition severely affects health, development and quality of life of infants and young children [4]. In addition, untreated OSAS increases healthcare utilization and associated costs [5]. Therefore, early detection is essential in order to initiate treatment.

In this regard, a recent report of the American Academy of Paediatrics re-emphasised the need for OSAS screening in every habitually snoring child [2].

The prevalence of OSAS is estimated to range 1% to 5% of children in the general paediatric population [2]. Despite its major negative consequences, childhood OSAS is considered a relatively under-diagnosed condition [6]. Overnight polysomnography (PSG) in a supervised sleep laboratory is the gold standard technique for a definitive diagnosis [2,4]. One of the most important factors responsible for this under-diagnosis is the limited availability of paediatric sleep units in most countries [4,7]. An additional major limitation is the intrusiveness of PSG for children, who showed high aversion to spend the whole night in the sleep unit with several sensors attached [4,8]. These drawbacks limit the effectiveness of conventional PSG as a screening technique for OSAS in every symptomatic child as suggested by the international medical community. Therefore, during the last decade, it has emerged a great demand for novel and simplified screening tools for the disease [9–11].

In the context of simplified alternatives to PSG, attended respiratory polygraphy (RP) has become a reliable method for OSAS detection in clinical settings [11–13]. In addition, unattended RP at home has been recently proposed as a feasible approach in low resource settings when in-lab PSG is not available [10,11]. Nevertheless, RP, which measures airflow (thermistor and/or nasal pressure), respiratory movements (chest and abdominal effort), body position, pulse rate and blood oxygen saturation (SpO_2), also manage several sensors, being still potentially intrusive for infants and young children. In this regard, recording of single-channel SpO_2 from overnight oximetry has been also proposed as a highly simple as well as effective screening technique for paediatric OSAS due to its suitability for children [10,14,15]. Moreover, automated processing of oximetric recordings has been proposed to enhance the diagnostic performance of overnight oximetry as a single screening test for childhood OSAS [16–19].

Several automated signal processing methods have been applied during the last years to parameterise changes in the overnight SpO_2 profile due to apnoeic events. Previous studies in the framework of paediatric OSAS detection by means of oximetry assessed conventional desaturation indexes [16–21], common statistics in the time domain [16,17,19], spectral features in the frequency domain [16,19] and nonlinear measures [16,19]. Among these complementary approaches, nonlinear methods have been marginally explored. Approximate entropy (ApEn) [22], sample entropy (SampEn) [23], central tendency measure [24], and Lempel–Ziv complexity [25] have demonstrated their usefulness to characterise desaturations linked to apnoeic events both in adults [26–32] and children [16,19]. Nevertheless, we hypothesise that different nonlinear metrics could gain insight into the dynamics of oximetry leading to additional and essential information. Furthermore, common apnoeic events in children with OSAS lead to slight fluctuations in SpO_2 recordings compared with deeper desaturations commonly present in adult patients. Consequently, screening for paediatric OSAS using only information from nocturnal oximetry is more challenging and thus more powerful methods are needed to thoroughly characterise all the changes linked with the disease. In the present paper, we propose the multiscale entropy (MSE) as a method able to exhaustively inspect nonlinear dynamics of SpO_2 recordings.

MSE is a nonlinear measure of complexity previously applied in different medical frameworks to quantify entropy changes in biomedical recordings over time scales [33]. In this regard, MSE has demonstrated to be useful to characterise differences in the heart rate dynamics due to age [33], obesity [34] and cardiac disease [35] or to analyse human gait [36], as well as to quantify changes in the complexity of the electroencephalogram (EEG) background activity in Alzheimer's disease patients [37] and EEG changes due to pharmacological intervention in schizophrenia [38]. Similarly, MSE has been applied to cerebral oxygenation signals from infrared spectroscopy in order to study mortality and brain injury in preterm infants [39]. In the context of OSAS, MSE has been recently used to analyse heart rate dynamics in adult patients. Particularly, in the study by Pan et al. [40], MSE was applied to estimate the deterioration in autonomic and vascular regulatory function linked with increasing OSAS severity and the subsequent improvement after continuous positive airway pressure

treatment. Similarly, MSE has demonstrated to be useful in the analysis of speech signals in order to quantify disorderliness in vocal patterns indicative of sleep apnoea [41]. In a previous study by our group [42], MSE was also applied to characterise the dynamics of heart rate variability time series in order to derive new patterns able to detect adult OSAS.

The aim of this study was two-fold: (i) firstly, to accomplish a comprehensive analysis of oximetry dynamics by means of MSE in order to characterise differences between non-OSAS children and paediatric patients suffering from the disease; (ii) and second, to assess the usefulness of MSE-derived features in order to compose an optimum model from unattended oximetry able to accurately screen for paediatric OSAS at home.

2. Methodology

2.1. Dataset and Sleep Studies

The population under study was composed of 50 children referred to the Respiratory Sleep Disorders Unit of the University Hospital of Burgos (Spain). All children showed common symptoms linked with clinical suspicion of OSAS, i.e., habitual snoring and/or witnessed breathing pauses during sleep reported by their parents or caretakers. According to our recruitment protocol, children referred to the sleep unit were randomly selected to participate in the study in order to avoid potential bias linked with the inclusion process. In addition, children suffering from serious chronic medical and/or psychiatric additional conditions, those showing symptoms indicative of sleep disorders other than OSAS, and those children who required urgent interventions were excluded. Sleep studies consisted of unsupervised RP at children's home and a subsequent in-hospital PSG. Table 1 summarises the socio-demographic and clinical features of the dataset. The Ethical Review Committee of the hospital approved the protocol (#CEIC 936) and informed consent to participate in the study was obtained from all caretakers prior to the enrolment.

Table 1. Demographic, anthropometric, polysomnographic, and oximetric characteristics of the paediatric population under study using a cut-off of 3 events/h for positive OSAS.

	All Children	OSAS-Negative	OSAS-Positive	*p*-Value
No. Subjects (%)	50	24 (48.0%)	26 (52.0%)	-
Age (years)	4 [4, 6]	4.5 [4, 6]	4 [4, 6]	N.S.
No. Males (n)	27 (54.0%)	11 (45.8%)	16 (61.5%)	N.S.
BMI (kg/m^2)	16.42 [15.00, 17.53]	16.42 [15.61, 17.42]	16.38 [14.57, 17.70]	N.S.
OAHI (events/h)	3.56 [1.21, 17.28]	1.18 [0.54, 1.87]	15.88 [6.72, 23.49]	<<0.05
Recording time (h)	9.05 [8.40, 9.27]	9.0 [8.74, 9.22]	9.08 [8.28, 9.56]	N.S.
ODI3 (events/h)	1.89 [5.19]	0.87 [1.48]	5.90 [7.26]	<<0.05
Sat$_{MIN}$ (%)	90 [4]	91 [2]	89 [3]	<0.05
Sat$_{AVG}$ (%)	97 [2]	97 [1]	97 [2]	N.S.
CT95 (%)	0.82 [3.00]	0.36 [1.10]	1.62 [6.61]	<0.05

OSAS: obstructive sleep apnoea syndrome; BMI: body mass index; OAHI: obstructive apnoea–hypopnoea index; ODI3: oxygen desaturation index of 3%; Sat$_{MIN}$: minimum SpO$_2$ during the whole recording; Sat$_{AVG}$: average SpO$_2$ during the whole recording; CT95: percentage of time with a SpO$_2$ below 95%; N.S.: non-significant statistical differences (*p* > 0.05)

According to the American Academy of Sleep Medicine (AASM), in-lab PSG was used as gold standard for a definitive diagnosis of paediatric OSAS [43]. In-lab supervised PSG was conducted from 22:00 to 08:00 using a digital polysomnograph Deltamed Coherence® 3NT version 3.0 (Diagniscan, S.A.U., Group Werfen, Paris, France). The following signals were recorded and stored for subsequent manual scoring: EEG, right and left electrooculogram, tibia and submental electromyogram, electrocardiogram, airflow (thermistor and nasal cannula), chest and abdominal movements (effort bands), oximetry, continuous transcutaneous carbon dioxide (PtcCO2), snoring and body position. The 2012 AASM rules for children were used to perform sleep staging and score apnoeic events [43]: an obstructive apnoea was quantified when a drop in the peak signal excursion ≥90% from pre-event baseline of oronasal thermal sensor occurred during at least the duration

of two breaths while maintaining the presence of respiratory effort throughout the entire period of airflow cessation. On the other hand, hypopnoea was quantified when peak signal excursions in the nasal pressure recording drop by $\geq$30% of pre-events baseline lasting at least two breaths, accompanied by a desaturation $\geq$3% or an electroencephalographic arousal. After manual scoring, the standard obstructive apnoea–hypopnoea index (OAHI) was computed, which measures the number of obstructive apnoeas and hypopnoeas per hour of sleep. There is a great controversy regarding the clinical cut-off used to confirm childhood OSAS [2,4,10,15]. In order to address this issue, a common OAHI cut-off point of 3 events/h was used in the present study [11,18,19]. According to this clinical threshold, a positive diagnosis was confirmed in 26 children (OSAS prevalence 52%). Table 1 shows the clinical and oximetric characteristics for the OSAS-negative and the OSAS-positive groups.

Unattended RP was carried out at children's home by means of a portable polygraphy equipment (eXim Apnea Polygraph by Bitmed, Sibel S.A., Barcelona, Spain). Unsupervised SpO_2 recordings from RP were acquired using a high sampling rate of 100 Hz in order to assist with artefact rejection. As suggested by the AASM, a pre-processing stage was implemented to remove artefacts due to patient's movements. Then, a non-overlapping time-averaging moving window of 1 s was applied, which is lower than the maximum acceptable signal averaging time of 3 s recommended by the AASM [44]. Every oximetric recording was downloaded as a single European Data Format (EDF) file for subsequent automated processing by means of MSE. Figure 1 depicts representative at-home SpO_2 portable recordings from our dataset. It is important to note that even the SpO_2 overnight profile of a child with moderate OSAS (OAHI $\geq$ 5 events/h) showed small fluctuations. Furthermore, all the desaturations of the oximetric recording from a severe OSAS-positive patient (OAHI $\geq$ 10 events/h) are comprised in the range 90%–100%, making it difficult to search for differences between OSAS-negative and OSAS-positive children just using the nocturnal oximetry profile.
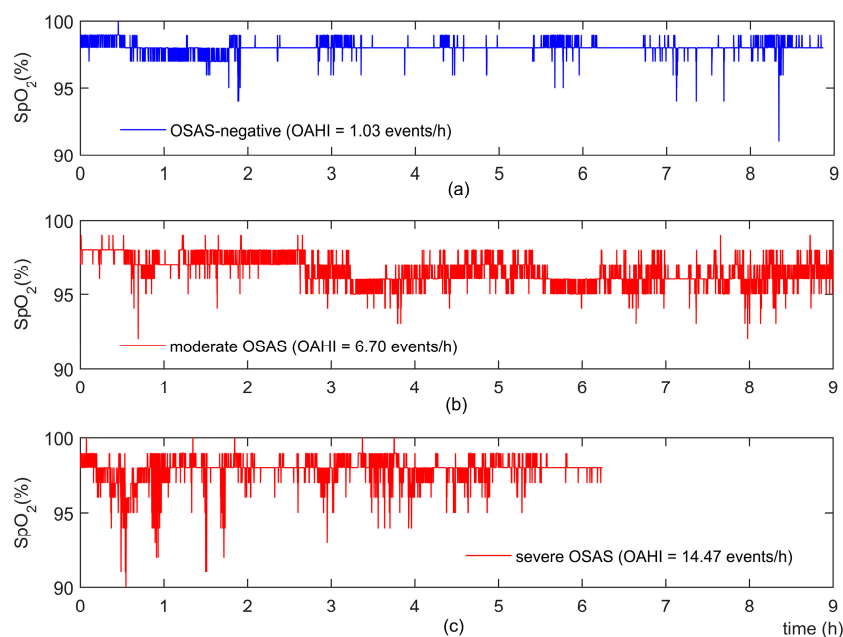
Figure 1. Representative nocturnal oximetric profiles for: (**a**) a children suspected of suffering from OSAS with OAHI < 5 events/h; (**b**) an OSAS-positive patient with 5 $\leq$ OAHI < 10 events/h (moderate OSAS); and (**c**) an OSAS-positive patient with OAHI $\geq$ 10 events/h (severe OSAS).

2.2. Automated Signal Processing

An automated signal processing scheme composed of three stages was accomplished. Firstly, every oximetric recording was analysed by means of MSE. MSE curves were parameterised to thoroughly characterise the complexity of overnight desaturations. In addition, conventional oximetric

indexes were used to account for additional information linked with the number and the severity of desaturations. After the feature extraction stage, an initial feature set composed of 21 variables was built. Secondly, an automated feature selection stage was conducted to identify the optimum feature subset composed of the most relevant as well as complementary variables. The widely known forward stepwise logistic regression (FSLR) algorithm was used to accomplish feature selection [45]. Finally, a binary logistic regression (LR) model aimed at discerning among OSAS-negative and OSAS-positive children was composed using the optimum features.

2.2.1. Multiscale Entropy

MSE is a nonlinear method proposed by Costa et al. [46] aimed at quantifying the complexity of a time series taking into account entropy changes along multiple time scales. As complex temporal fluctuations are inherent to physiological dynamics, MSE is able to provide additional and useful information that conventional entropy-based measures cannot. Despite its validated usefulness, traditional single-scale entropies do not account for the a priori relevant information linked with the dynamical structure of the signal on scales other than the shortest one [33,46]. In order to overcome this issue, MSE estimates the entropy of consecutive coarse-grained versions of the original time series so that each coarse-grained sequence characterises the system state on increasing time scales.

Regarding the algorithm proposed by Costa et al. [46], given a one-dimensional discrete time series $x(i)$ of length N ($i = 1, \ldots, N$), the coarse-grained versions for a time scale factor τ are computed as follows:

$$y^{(\tau)}(j) = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x(i), 1 \leq j \leq \frac{N}{\tau} \tag{1}$$

For $\tau = 1$, the sequence $y^{(1)}$ is really the original time series, whereas elements of each coarse-grained sequence are the average of the original samples within non-overlapping segments of length τ. Therefore, the length of each coarse-grained sequence is the original length N divided by the scale factor τ. Accordingly, MSE analysis is performed by computing the single-scale entropy measure for each coarse-grained time series plotted as a function of τ, from the original signal ($\tau = 1$) to the highest time scale [46].

Approximate entropy (ApEn) and sample entropy (SampEn) are commonly used as single-scale entropy measures to compute MSE [36,42,46]. Similarly, both the Tsallis and the Rényi entropies have been also proposed [34]. Nevertheless, the use of SampEn has several major advantages [23,33]: it is less dependent of the sequence length so it can be applied to relatively short and noisy biomedical recordings, and it shows relative consistency over a broader range of input model-dependent parameters. In addition, SampEn reduces the bias caused by self-matching inherent to the ApEn algorithm [23]. Therefore, SampEn was used in the present research.

Briefly, SampEn (m, r, N) is aimed at quantifying irregularity of one-dimensional time series, assigning larger values to sequences showing larger degree of disorder, i.e., higher entropy [23]. SampEn is computed as follows [23]:

$$SampEn(m, r, N) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right] \tag{2}$$

where A^m and B^m are the average number of segments $X_m(i)$ ($1 \leq i \leq N - m + 1$) of length m and $m+1$, respectively, such that the distance between every pair of segments $X_m(i)$ and $X_m(j)$ is less than or equal to a tolerance r according to the following equation:

$$d[X_m(i), X_m(j)] = \max_{k=0,\ldots,m-1} (|x(i+k) - x(j+k)|), \text{with } 1 \leq j \leq N - m \text{ and } j \neq i \tag{3}$$

There is not a consensus to set the highest time scale in MSE analyses though it depends on the problem under study, as well as on the characteristics of the signal and the single-scale measure

of entropy [37,42]. In the present study, overnight oximetric recordings had a median duration of 9.05 h, i.e., $\approx 2^{15}$ samples after pre-processing and time-averaging. As a proper estimation of SampEn requires at least 10^m samples [23], we set a conservative maximum scale factor of $\tau = 50$ in order to inspect oximetry dynamics. Regarding the input parameters of SampEn, the values of m and r are critical in the estimation of entropy and thus in the performance of MSE analyses. In the present study, we used $m = 1$ and $r = 0.25$ times the standard deviation (SD) of the original recording, which have demonstrated to be optimal in previous analysis of oximetry by means of SampEn [19,27,31]. As recommended by Costa et al. [33], r was not normalised for time scales $\tau > 1$ because changes of variance in the coarse-grained versions of the signal have information about the whole original time series. As in similar studies [37,42], MSE curves were parameterised by means of slopes and single-entropy values for the time scales showing the most significant visual differences among the groups under study. Similarly, the area under the MSE profile for these scales and the time scale where the MSE function is maximum were used to characterise each curve.

2.2.2. Conventional Oximetric Indexes

Information linked with the number and severity of desaturations is commonly used in clinical settings in the context of childhood OSAS due to its readiness and easy interpretation. In fact, despite evidences showing their inherent underestimation, conventional oximetric indexes have demonstrated to be useful in OSAS screening [1,20,21]. Therefore, the following indices were included in our initial feature space to account for this relevant data: the oxygen desaturation index $\geq 3\%$ (ODI3), which measures the number of desaturations greater than or equal to 3% from baseline per hour of recording; the minimum (Sat_{MIN}) and the average (Sat_{AVG}) saturation values along the whole recording; and the cumulative time spent with a saturation below 95% (CT95) as a percentage of the total recording time.

2.2.3. Feature Selection and Classification

The well-known binary logistic regression (LR) algorithm was involved both in feature selection and classification stages. Regarding dimensionality reduction, bidirectional forward stepwise logistic regression (FSLR) is a widely applied method for LR model optimization [29,31,42,47]. FSLR is able to find the simplest as well as still representative feature subset conducting an efficient and robust iterative process [45]. Briefly, bidirectional FSLR selects the most relevant variables (forward selection) and simultaneously removes the redundant ones (backward elimination) in terms of statistically significant differences between the current model and a candidate one. In the present study, a bootstrapping approach was implemented to obtain an optimal feature space independent of a particular dataset. Bootstrapping is particularly useful to estimate statistics in small-sized datasets [48]. Hence, the FSLR feature selection algorithm was applied 1000 times to different bootstrap replicates derived from the original dataset. In order to gather as much relevant information as possible, as well as maintain a moderate number of variables, a conservative threshold for feature selection was set: all variables automatically selected at least 10% of the runs composed the optimum feature space. Finally, a LR model aimed at classifying OSAS-negative and OSAS-positive children was built using our optimum feature subset.

2.2.4. Statistical Analyses

Matlab R2015a (The MathWorks Inc., Natick, Massachusetts) was used to perform statistical analyses as well as to implement automated pattern recognition stages. The Kolmogorov-Smirnoff's normality test and the Levene's homoscedasticity test revealed that the oximetric features involved in our study were not normally distributed and variances were unequal. Accordingly, a descriptive analysis of every feature was carried out by means of the median and interquartile range. Similarly, significant statistical differences ($p < 0.05$) between the groups under study (OSAS-negative vs. OSAS-positive) were assessed using the non-parametric Mann-Whitney U test.

The common bootstrap 0.632 approach was applied in order to validate our proposal since it is particularly useful to estimate performance metrics in small-sized datasets [18,19,48]. Given an original dataset of size *N*, bootstrap 0.632 applies resampling with replacement to build *M* new datasets, the so-called bootstrap replicates, each one composed of *N* instances. For each replicate m_i ($1 \leq i \leq M$), every instance from the original dataset can be selected several times with equal (uniform) probability, i.e., all replicates will contain repeated instances. Consequently, for each bootstrap replicate, a number of instances from the original dataset are not selected. At each iteration, a dataset m_i is used for training purposes, whereas instances not involved in the replicate are used for validation. In order to obtain a proper estimation of the 95% confidence interval (CI95%), the number of bootstrap replicates was set to *M* = 1000 [48]. According to bootstrap 0.632, every statistic or performance metric must be computed as a contribution of both the training replicate and its corresponding test dataset as follows [48]:

$$metric^{(m_i)} = 0.632 \cdot metric_{TEST}^{(m_i)} + 0.368 \cdot metric_{TRAIN}^{(m_i)} \tag{4}$$

Finally, the estimation of each performance metric is obtained as the average across all the *M* bootstrap replicates.

The performance of the LR classifier was assessed by means of sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR−), accuracy (Acc), and area under the receiver operating characteristics (ROC) curve (AUC).

3. Results

Figure 2 plots individual SampEn values as a function of τ for every overnight oximetric recording in the population under study, as well as the average for the whole OSAS-positive and OSAS-negative groups. Despite the inherent variance, OSAS-positive patients showed greater averaged entropy, i.e., irregularity, than OSAS-negative children due to desaturations caused by apnoeic events for every time scale. We performed a visual inspection of the averaged MSE profiles to properly parameterise each curve. Regarding the smaller time scales, it is important to note that SampEn values of OSAS-positive patients increased with a substantially higher slope than for the OSAS-negative group from scales $\tau = 1$ to $\tau = 6$. Then, SampEn values increased monotonically for both groups until reaching a similar slope. In addition, we can observe that there was a maximum difference between both MSE averaged profiles for time scale $\tau = 14$. In order to gather this information, the following parameters were derived from the MSE profile of each oximetric recording:

i.   Slope of the MSE curve between scale $\tau = 1$ and scales $\tau = 2$ (Slp_{1-2}), $\tau = 3$ (Slp_{1-3}), $\tau = 4$ (Slp_{1-4}), $\tau = 5$ (Slp_{1-5}) and $\tau = 6$ (Slp_{1-6}). It is estimated as the slope of the straight-line connecting the MSE values of the time scales under study. Higher slope accounts for a larger entropy increase between the original signal ($\tau = 1$) and coarse-grained versions in consecutive short time scales ($\tau = 2$ to 6), i.e., the control mechanisms regulating peripheral blood oxygen saturation on such short time scales are the most affected by recurrent apnoeic events.

ii.   Individual SampEn values from scale $\tau = 1$ to scale $\tau = 6$ (SE_1 to SE_6). Single-scale SampEn is a measure of entropy or disorderliness and thus larger individual values are linked with more complex underlying mechanisms governing the dynamics of the oximetric signal for these time scales.

iii.   SampEn single value in the scale reaching the maximum margin between MSE curves of the groups under study, i.e., $\tau = 14$ (SE_{max}). This feature quantifies the irregularity of the oximetric recording for the time scale where the maximum difference between the classes under study (OSAS-negative vs. OSAS-positive) is expected.

iv.   Area enclosed under the MSE curve between scale $\tau = 1$ and scales $\tau = 2$ (Ar_{1-2}), $\tau = 4$ (Ar_{1-4}) and $\tau = 6$ (Ar_{1-6}). MSE curves allow us to compare the relative complexity of time series [33].

Higher area is achieved when SampEn values are higher for the majority of the time scales, suggesting that the time series is more complex.

v. Area enclosed under the MSE curve between scale $\tau = 1$ and the scale reaching the maximum margin ($\tau = 14$) between the averaged MSE curves (Ar_{1-max}). After time scale $\tau = 14$, the MSE curves of OSAS-negative and OSAS-positive groups monotonically increase with a similar slope, showing almost equal behaviour. From short time scales to scale $\tau = 14$, the MSE curves of both groups show the greatest differences regarding shape and individual entropy values. Thus, this feature gathers the contribution of the time scales showing the maximum differences in the dynamics of nocturnal oximetry between the groups under study.

vi. Time scale where the maximum SampEn value is reached (τ_{max}). This feature is related to the level of depth of changes in the underlying complexity of the signal, i.e., it shows the time scale up to which entropy increases.
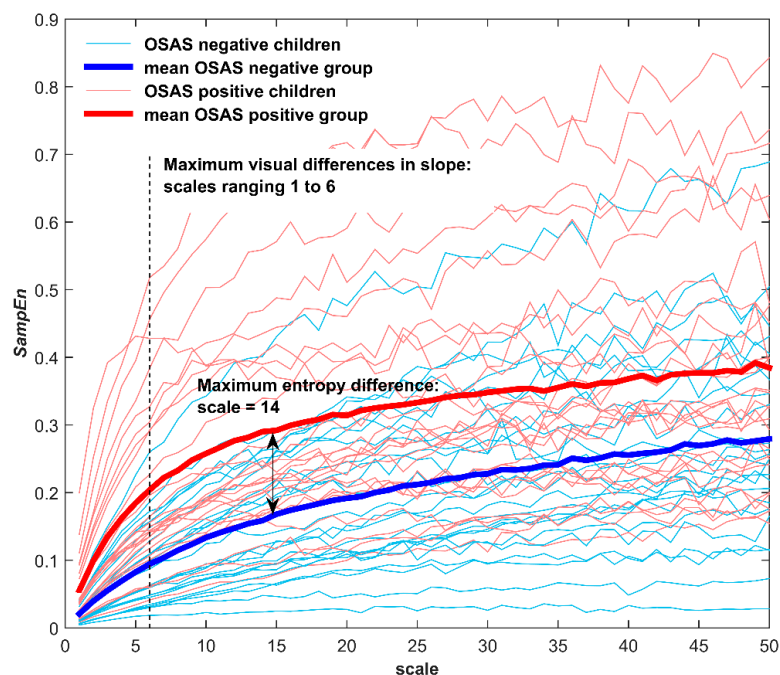
Figure 2. MSE curves for every overnight oximetric recording in the population set. Averaged MSE profiles along each time scale are also plotted for the whole OSAS-negative (blue) and OSAS-positive (red) groups.

Table 2 summarises the median and interquartile range (IQR) of all these MSE-derived parameters for the OSAS-negative and the OSAS-positive groups. It is noticeable that almost all features achieved statistically significant differences between groups ($p < 0.05$). On average, OSAS-positive patients showed significantly higher slopes (Slp_{1-2} to Slp_{1-6}), higher irregularity (SE_1 to SE_6), and higher area under the MSE curve (Ar_{1-2} to Ar_{1-6}) in the smaller time scales than OSAS-negative children. Similarly, OSAS-positive patients also showed significantly higher area under the MSE curve between time scales 1 and the maximum-margin scale ($\tau = 14$) as well as higher entropy in such a maximum-margin scale than OSAS-negative children.

Table 3 summarises the diagnostic performance of every entropy-based parameter derived from MSE analysis. Almost all features under study showed balanced sensitivity and specificity values, as well as moderate diagnostic accuracy. Regarding slope-based MSE features, accuracy ranged 71.4% to 73.4% and both Slp_{1-3} and Slp_{1-4} reached the maximum AUC (0.80). Similarly, the accuracy of area-based MSE features ranged 71.3% to 72.3% and Ar_{1-2}, Ar_{1-4} and Ar_{1-6} reached 0.80 AUC. In regard to SampEn values at individual time scales, accuracy ranged 68.1% to 73.4% and both SE_1 and SE_3

achieved 0.81 AUC. Finally, τ_{max} reached poor accuracy (Acc = 55.8%) and poor area under the ROC curve (AUC = 0.60). Table 4 shows the performance of every conventional oximetric index involved in the study. Accuracy ranged 58.8% to 74.5% and all indices showed balanced sensitivity and specificity. It is important to highlight that ODI3 performed notably higher than the remaining conventional features, reaching 74.5% Acc (71.9% Se and 77.6% Sp) and 0.85 AUC.

Table 2. Descriptive analysis (median and interquartile range) of each MSE-derived feature for each patient group under study.

MSE Features	OSAS-Negative	OSAS-Positive	*p*-Value
Slp_{1-2}	0.015 [0.017]	0.031 [0.041]	<<0.05
Slp_{1-3}	0.028 [0.031]	0.057 [0.070]	<<0.05
Slp_{1-4}	0.038 [0.044]	0.079 [0.094]	<<0.05
Slp_{1-5}	0.049 [0.059]	0.096 [0.111]	<<0.05
Slp_{1-6}	0.058 [0.073]	0.110 [0.128]	<0.05
SE_1	0.016 [0.019]	0.039 [0.055]	<<0.05
SE_2	0.032 [0.036]	0.071 [0.097]	<<0.05
SE_3	0.045 [0.050]	0.097 [0.126]	<<0.05
SE_4	0.055 [0.063]	0.118 [0.149]	<<0.05
SE_5	0.067 [0.078]	0.135 [0.166]	<<0.05
SE_6	0.075 [0.092]	0.147 [0.175]	<<0.05
SE_{max}	0.129 [0.147]	0.229 [0.216]	<0.05
Ar_{1-2}	0.048 [0.055]	0.110 [0.152]	<<0.05
Ar_{1-4}	0.151 [0.168]	0.324 [0.427]	<<0.05
Ar_{1-6}	0.292 [0.338]	0.599 [0.767]	<<0.05
Ar_{1-max}	1.164 [1.382]	2.114 [2.463]	<<0.05
τ_{max}	48.000 [3.500]	48.000 [5.000]	N.S.

MSE: multiscale entropy; Slp_{1-x}: slope of the MSE curve between scale $\tau = 1$ and scale $\tau = x$; SE_x: Sample entropy value in the scale $\tau = x$; SE_{max}: Sample entropy value in the scale reaching the maximum margin between MSE curves of the groups under study; Ar_{1-x}: Area enclosed under the MSE curve between scale $\tau = 1$ and scale $\tau = x$; Ar_{1-max}: Area enclosed under the MSE curve between scale $\tau = 1$ and the scale reaching the maximum margin between MSE curves; τ_{max}: Scale where the maximum sample entropy value is reached; N.S.: non-significant statistical differences ($p > 0.05$).

Figure 3 shows the number of times each variable was selected after FSLR feature selection using the bootstrapping approach proposed to improve model generalization. A total of nine features (five from MSE analysis and four conventional indices) were above the specified threshold (10% of total runs): Slp_{1-2}, Slp_{1-6}, SE_1, SE_{max}, τ_{max}, ODI3, Sat_{MIN}, Sat_{AVG} and CT95. Using these features, an optimum LR model was composed. Figure 4 shows the ROC curves during the training stage of the bootstrapping procedure (average across all bootstrap replicates) for each feature subset under study, i.e., MSE-derived features (MSE), conventional oximetric indices (OX), all features together without feature selection (MSE-OX) and the optimum feature subset from FSLR (OPT). We can observe that the feature subsets using MSE-derived features and oximetric indices jointly achieved notably higher AUC than each single approach individually, which supports our initial hypothesis of complementarity between both approaches.

Table 5 summarises the diagnostic performance of all approaches involved in the study. A LR model composed of all the MSE-derived features reached 75.2% Acc (75.7% Se, 75.3% Sp) and 0.79 AUC, whereas a LR model built with all the oximetric indices reached 76.0% Acc (74.7% Se, 77.7% Sp) and 0.82 AUC. Similarly, a LR model composed of all the MSE and oximetric variables under study reached comparable performance, achieving 79.0% Acc (79.4% Se, 79.3% Se) and 0.80 AUC. Interestingly, our optimum LR model composed of features automatically selected by FSLR performed significantly better, reaching 83.5% Acc (84.5% Se, 83.0% Sp) and 0.86 AUC.

Table 3. Diagnostic performance using a bootstrap approach for each feature derived from the MSE profile of nocturnal oximetric recordings.

	Se (%)	Sp (%)	PPV (%)	NPV (%)	LR+	LR−	Acc (%)	AUC
Slp_{1-2}	75.6 (47.7, 98.3)	67.8 (36.6, 96.3)	72.5 (50.5, 96.8)	72.6 (48.3, 97.7)	2.83 (1.30, 6.71)	0.37 (0.03, 0.77)	71.8 (56.8, 86.5)	0.79 (0.62, 0.94)
Slp_{1-3}	78.1 (49.4, 98.5)	68.1 (38.4, 95.8)	73.1 (51.2, 96.4)	74.9 (49.7, 98.0)	2.89 (1.34, 6.71)	0.33 (0.02, 0.72)	73.2 (58.8, 88.2)	0.80 (0.63, 0.94)
Slp_{1-4}	78.6 (50.6, 98.5)	68.0 (37.6, 96.2)	73.2 (51.4, 96.5)	75.1 (49.6, 98.1)	2.94 (1.36, 7.07)	0.32 (0.02, 0.72)	73.4 (57.7, 88.6)	0.80 (0.64, 0.95)
Slp_{1-5}	76.1 (48.7, 98.4)	66.6 (35.7, 95.7)	71.8 (48.7, 96.2)	72.6 (46.9, 97.6)	2.71 (1.26, 6.10)	0.37 (0.03, 0.81)	71.4 (55.2, 86.4)	0.78 (0.61, 0.94)
Slp_{1-6}	76.6 (45.8, 97.6)	66.5 (35.7, 95.6)	71.8 (48.8, 95.9)	72.8 (47.6, 97.1)	2.72 (1.23, 6.33)	0.36 (0.03, 0.81)	71.6 (55.2, 87.4)	0.77 (0.60, 0.92)
SE_1	75.2 (47.4, 98.6)	70.2 (36.0, 96.9)	74.0 (51.4, 97.3)	73.0 (48.3, 98.0)	3.10 (1.35, 7.43)	0.36 (0.02, 0.76)	72.7 (57.6, 88.6)	0.81 (0.65, 0.95)
SE_2	75.0 (48.4, 98.7)	68.7 (36.6, 96.6)	73.0 (50.9, 96.8)	72.6 (48.3, 98.1)	2.92 (1.34, 6.99)	0.37 (0.02, 0.77)	71.9 (56.8, 87.2)	0.80 (0.64, 0.95)
SE_3	76.9 (49.1, 98.3)	68.9 (39.1, 96.3)	73.3 (51.4, 96.7)	74.0 (48.6, 97.8)	2.96 (1.34, 6.92)	0.34 (0.02, 0.74)	72.9 (58.0, 88.2)	0.81 (0.64, 0.94)
SE_4	77.3 (50.1, 98.5)	69.4 (36.6, 96.5)	73.9 (51.7, 96.8)	74.6 (48.7, 97.9)	3.06 (1.35, 7.14)	0.33 (0.02, 0.74)	73.4 (57.6, 88.2)	0.80 (0.64, 0.95)
SE_5	76.4 (49.1, 98.7)	67.9 (36.1, 96.1)	72.8 (50.5, 96.6)	73.3 (48.3, 98.1)	2.88 (1.30, 6.81)	0.35 (0.02, 0.76)	72.2 (56.8, 87.0)	0.79 (0.63, 0.94)
SE_6	75.8 (49.9, 97.7)	67.1 (36.2, 96.1)	72.0 (48.6, 96.5)	72.3 (47.9, 97.2)	2.79 (1.25, 6.63)	0.37 (0.03, 0.82)	71.5 (55.2, 87.0)	0.78 (0.61, 0.93)
SE_{max}	74.3 (39.4, 98.7)	61.6 (31.7, 96.0)	68.3 (46.0, 95.7)	70.7 (43.4, 97.8)	2.26 (1.13, 5.49)	0.42 (0.02, 0.95)	68.1 (52.3, 83.6)	0.75 (0.58, 0.91)
Ar_{1-2}	74.8 (48.4, 98.7)	68.8 (34.8, 96.7)	73.1 (50.7, 96.8)	72.5 (48.3, 98.1)	2.94 (1.31, 7.03)	0.37 (0.02, 0.78)	71.8 (55.2, 87.2)	0.80 (0.64, 0.95)
Ar_{1-4}	76.1 (49.1, 98.6)	68.5 (36.6, 96.5)	73.0 (51.3, 96.8)	73.4 (48.3, 98.0)	2.92 (1.33, 7.07)	0.35 (0.02, 0.76)	72.3 (57.8, 87.2)	0.80 (0.64, 0.95)
Ar_{1-6}	75.1 (49.2, 98.6)	68.7 (36.9, 96.5)	73.0 (50.5, 96.8)	72.5 (47.9, 98.0)	2.94 (1.30, 7.05)	0.37 (0.02, 0.78)	71.9 (55.5, 87.2)	0.80 (0.64, 0.94)
Ar_{1-max}	75.5 (47.4, 98.3)	67.1 (35.0, 96.2)	72.0 (49.1, 96.3)	72.3 (47.8, 97.3)	2.75 (1.26, 6.47)	0.37 (0.03, 0.82)	71.3 (55.2, 86.4)	0.78 (0.61, 0.93)
τ_{max}	57.0 (21.2, 98.3)	55.0 (20.3, 89.2)	58.3 (31.9, 84.0)	56.6 (29.0, 96.8)	1.47 (0.64, 3.59)	0.81 (0.04, 1.94)	55.8 (39.6, 71.8)	0.60 (0.51, 0.76)

Se: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR- negative likelihood ratio; Acc: accuracy; AUC: area under the receiver operating characteristic curve; Slp_{1-x}: slope of the MSE curve between scale $\tau = 1$ and scale $\tau = x$; SE_x: SampEn in the scale $\tau = x$; SE_{max}: SampEn in the scale reaching the maximum margin between MSE curves of the groups under study; Ar_{1-x}: Area enclosed under the MSE curve between scale $\tau = 1$ and scale $\tau = x$; Ar_{1-max}: Area enclosed under the MSE curve between scale $\tau = 1$ and the scale reaching the maximum margin between MSE curves; τ_{max}: Scale where the maximum sample entropy value is reached.
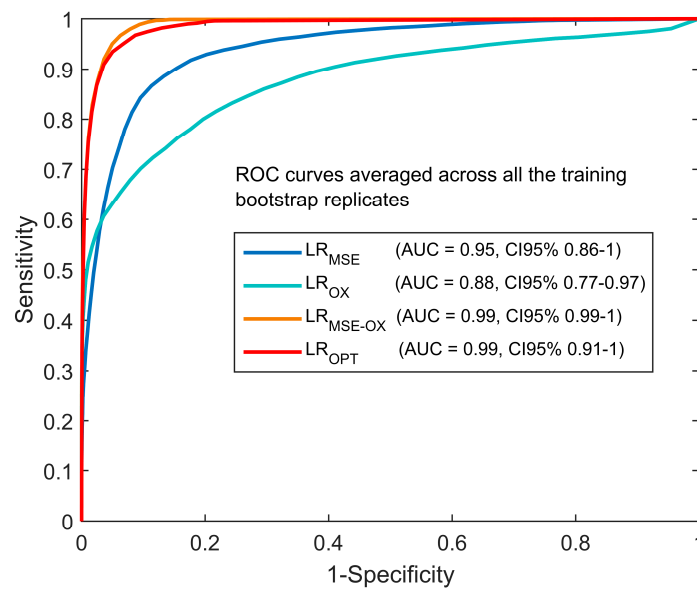
Table 4. Diagnostic performance using a bootstrap approach for each oximetric variable involved in the model development.

	Se (%)	Sp (%)	PPV (%)	NPV (%)	LR+	LR-	Acc (%)	AUC
$ODI3$	71.9 (44.9, 97.6)	77.6 (40.7, 100)	79.7 (55.1, 100)	72.6 (49.7, 97.3)	3.97 (1.50, 11.20)	0.36 (0.03, 0.70)	74.5 (58.5, 88.9)	0.85 (0.70, 0.97)
Sat_{MIN}	54.2 (20.3, 97.8)	64.2 (21.2, 98.2)	64.8 (34.7, 97.3)	58.4 (30.8, 96.1)	1.96 (0.72, 5.62)	0.72 (0.05, 1.44)	58.8 (42.9, 73.6)	0.62 (0.51, 0.79)
Sat_{AVG}	69.4 (34.2, 96.7)	66.8 (39.5, 95.9)	69.8 (45.1, 95.6)	68.1 (42.1, 96.0)	2.41 (1.07, 5.40)	0.46 (0.05, 0.94)	68.1 (49.9, 82.9)	0.70 (0.52, 0.88)
$CT95$	66.8 (34.2, 96.3)	69.3 (35.9, 98.6)	72.0 (48.3, 98.2)	66.4 (42.3, 95.4)	2.63 (1.14, 8.02)	0.49 (0.06, 0.97)	67.8 (52.3, 83.0)	0.75 (0.57, 0.92)

Se: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR- negative likelihood ratio; Acc: accuracy; AUC: area under the receiver operating characteristic curve; $ODI3$: oxygen desaturation index of 3%; Sat_{MIN}: minimum SpO_2 during the whole recording; Sat_{AVG}: average SpO_2 during the whole recording; $CT95$: percentage of time with a SpO_2 below 95%.
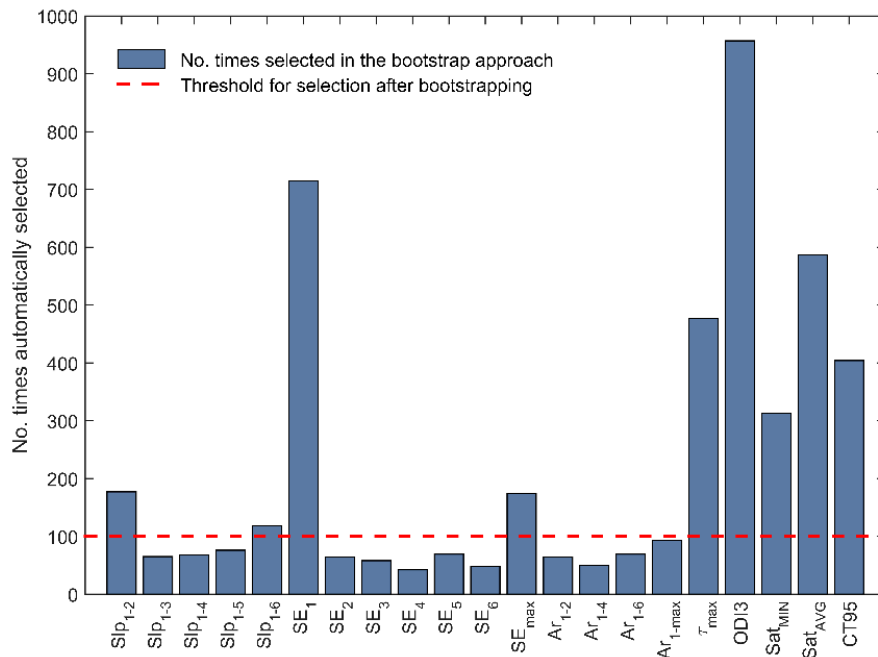
Figure 3. Histogram summarising the number of times each variable from the initial feature space is selected after forward stepwise logistic regression (FSLR) feature selection using the proposed bootstrapping approach with 1000 bootstrap replicates.

Figure 4. Receiver operating characteristics (ROC) curves of the proposed logistic regression (LR) models from training. Each curve is the average of all the bootstrap training replicates.

Table 5. Diagnostic performance using a bootstrap approach for all the binary LR models under study using a cut-off of 3 events/h for positive OSAS.

	Se (%)	Sp (%)	PPV (%)	NPV (%)	LR+	LR−	Acc (%)	AUC
LR_{MSE}	75.7	75.3	77.5	74.4	5.95	0.38	75.2	0.79
	(49.0, 100)	(43.4, 100)	(53.6, 100)	(48.7, 100)	(1.78, 13.41)	(0, 0.93)	(57.0, 90.0)	(0.58, 0.95)
LR_{OX}	74.7	77.7	79.7	74.2	4.76	0.35	76.0	0.82
	(47.3, 98.8)	(44.4, 1)	(55.9, 1)	(49.9, 98.7)	(1.62, 12.92)	(0.01, 0.75)	(58.8, 90.4)	(0.64, 0.97)
LR_{MSE-OX}	79.4	79.3	80.8	78.4	6.65	0.34	79.0	0.80
	(54.9, 1)	(55.3, 1)	(62.1, 1)	(58.5, 1)	(2.56, 14.39)	(0, 0.92)	(63.1, 92.8)	(0.62, 0.95)
LR_{OPT}	84.5	83.0	84.7	83.3	7.81	0.23	83.5	0.86
	(60.1, 1)	(54.2, 1)	(61.6, 1)	(58.7, 1)	(1.58, 15.05)	(0, 0.70)	(63.2, 96.1)	(0.65, 0.99)

Se: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR− negative likelihood ratio; Acc: accuracy; AUC: area under the receiver operating characteristic curve; LR_{MSE}: LR model composed of all the MSE-derived features; LR_{OX}: LR model composed of all the oximetric features; LR_{MSE-OX}: LR model composed of all the MSE-derived and oximetric features; LR_{OPT}: LR model composed of the optimum features derived from FSLR.

4. Discussion

In this study, we analysed the usefulness of MSE in the context of screening for paediatric OSAS by means of overnight unattended oximetry at home. Nocturnal oximetry has emerged as a reliable technique for simplified OSAS detection in children. The present research gains additional insight into the diagnostic capability of portable oximetry. Our results showed that MSE is able to provide relevant and complementary information to conventional oximetric indices commonly used by physicians. MSE profiles of oximetric recordings showed a consistent behaviour, with significant differences between OSAS-negative and OSAS-positive children, particularly for lower time scales. In addition to single-scale entropy measures from scales $\tau = 1$ to $\tau = 6$, where maximum visual differences were found, all the proposed slope- and area-based parameters derived from the MSE curve achieved statistically significant differences between groups. According to these MSE-measures, OSAS-positive patients showed significantly higher average irregularity than non-OSAS children due to desaturations linked with OSAS along all the time scales. Regarding the diagnostic performance, individual MSE-based parameters reached moderate accuracy (maximum Acc of 73.4%), similar to that achieved by conventional ODI3 (74.5% Acc). A multivariate analysis showed that putting all this data together led to a notably overall performance increase, which demonstrated that MSE is able to provide additional as well as complementary information. A LR model composed of all features under study reached 79.0% Acc (79.4% Se, 79.3% Sp) and 0.80 AUC. Furthermore, our novel LR model fed with an optimum feature subset derived from FSLR feature selection reached 83.5% Acc (84.5% Se, 83.0% Sp) and 0.86 AUC, which significantly outperformed all the approaches under study. Comparing the ROC curves from training shown in Figure 4 and the final performance of each LR model in terms of the AUC values shown in Table 5, we can observe that LR_{OX} and LR_{OPT} showed the smallest decrease in performance. On the other hand, the model composed of all the variables from the original feature space (LR_{MSE-OX}) showed a significant performance decrease, which supports the convenience of our optimum feature subset derived from FSLR. Therefore, LR_{OPT} reached higher diagnostic accuracy as well as higher generalization ability.

It is essential to highlight that the optimum feature subset was composed of variables from both MSE analysis (five features: Slp_{1-2}, Slp_{1-6}, SE_1, SE_{max}, τ_{max}) and conventional indices (four features: $ODI3$, Sat_{MIN}, Sat_{AVG} and $CT95$), which points out the complementarity of the proposed MSE approach. Regarding the physiological interpretation of the optimum features, it is widely known that $ODI3$ quantifies the number of desaturations per hour of recording, whereas the remaining conventional indices account for the overall (Sat_{AVG} and $CT95$) and maximum (Sat_{MIN}) severity of the desaturations. On the other hand, it is unknown how the proposed MSE-derived measures are related to particular changes in the dynamics of oximetry. As proposed by Costa et al. [33], we used MSE curves to compare the relative complexity of normalised time series, i.e., oximetric recordings belonging to

OSAS-negative and OSAS-positive children. Accordingly, Slp_{1-2} and Slp_{1-6} reflect that the degree of change in the complexity of overnight oximetry (computed as the slope of the MSE curve) due to apnoeic events is more relevant in smaller scales (scales 1-2 and 1-6). The single-scale entropy measure SE_1 shows that original ($\tau = 1$) oximetric recordings from OSAS-positive children have significantly higher irregularity than OSAS-negative children. Moreover, the influence of apnoeic events is still relevant in moderate time scales since SE_{max} reflects significantly higher irregularity in oximetric recordings from OSAS-positive patients for time scale $\tau = 14$. Finally, according to Table 2, τ_{max} did not show statistically significant differences between groups. Nevertheless, it was automatically selected by FSLR, suggesting that the level up to which entropy increases due to the influence of apnoeic events provides complementary information to direct measures of entropy.

In order to gain insight into the performance of our binary classifier, we analysed polysomnographic and polygraphic features of misclassified children. Notice that a bootstrapping technique was carried out to validate our approach and thus all performance metrics were computed as the average across all the bootstrap replicates. Therefore, we analysed those patients misclassified in a significant number of repetitions of the algorithm. Accordingly, there were five false positive children. Two of them were borderline, showing an OAHI from PSG equal to 2.5 and 2.6 events/h, respectively. It is important to highlight that children were diagnosed according to in-lab PSG, which is the gold standard, whereas our screening method is based on the oximetry signal recorded at-home in a different night. Hence, it is essential to consider common night-to-night variability inherent to OSAS when analysing misclassifications. In this regard, three out of five false positive patients showed an at-home OAHI > 3 events/h, ranging 3.6 to 7.7 events/h. Regarding false negative patients, there were five OSAS-positive children incorrectly classified as no-OSAS using our screening oximetry-based tool. It is important to note that they were all moderate-to-severe OSAS patients (in-lab OAHI > 5 events/h), showing a similar behaviour at home (unsupervised OAHI ranging 4.0 to 11.1 events/h). Nevertheless, four out of five false negative children showed an at-home ODI3 significantly lower, ranging 0.5 to 2.3 events/h. This suggests that apnoeic events did not lead to a matching desaturation in the oximetric profile, which is probably the main limitation of oximetry as a single screening tool for the disease.

Previous studies used MSE in the context of automated characterisation of OSAS in adults. In a recent study by Roebuck and Clifford [41], MSE was applied to characterise irregularity of speech patterns from subjects suspected of suffering from sleep apnoea. MSE coefficients of speech signals for small ($\tau = 1, 2, 4, 8$) and large ($\tau = 16, 32, 65, 130, 180$) scales were used to detect moderate-to-severe OSAS (AHI $\geq$ 15 events/h) using a random forest classification paradigm. An overall accuracy of 79.9% (66.0% Se, 88.8% Sp) was obtained, whereas the performance increased up to 80.5% Acc (69.2% Se, 87.9% Sp) when demographic variables were added to the model. Pan et al. [40] analysed heart rate variability (HRV) time series of snoring patients with and without OSAS by means of MSE in order to assess changes in autonomic and vascular regulatory function. A significant irregularity decrease in HRV dynamics both for small ($\tau < 6$) and large ($\tau > 6$) time scales were found for moderate-to-severe OSAS patients (AHI $\geq$ 15 events/h), whereas non-OSAS subjects and patients with continuous positive airway pressure therapy showed a similar increase in MSE features in the larger scales. Similarly, Gutiérrez-Tobal et al. [42] applied MSE to HRV recordings in order to model adult OSAS in independent populations separated by gender. Using together MSE coefficients and spectral entropy measures automatically selected by means of FSLR, a LR model achieved 85.2% Acc (80.8% Se, 89.3% Sp) in the classification of women with OSAS (AHI $\geq$ 10 events/h). The performance decreased to 77.6% Acc (87.1% Se, 56.1% Sp) when modelling OSAS in men.

To our knowledge, this is the first study assessing MSE in the context of childhood OSAS. Moreover, in the present research, we focused on the diagnostic ability of unattended oximetry at home as a single screening tool for the disease, which is a major novelty in the framework of paediatric sleep apnoea. Previously, few studies assessed the usefulness of automated analysis of overnight portable oximetry [16–20,49]. Kirk et al. [20] analysed a dataset composed of 57 children suspected of suffering from OSAS. The oxygen desaturation index > 4% (ODI4) from unattended portable oximetry was used

to characterise OSAS (AHI $\geq$ 5 events/h), reaching 66.7% sensitivity and 60.0% specificity when a cut-off of ODI4 $\geq$ 5 events/h was used. In the study by Garde et al. [16], time and spectral features from overnight pulse oximetry were used to assist in paediatric OSAS detection (AHI $\geq$ 5 events/h). The Authors analysed 146 SpO_2 recordings acquired by means of a portable device, although all sleep studies were carried out in a supervised hospital setting. Stepwise linear discriminant analysis (LDA) reached 78.5% accuracy (80.0% Se, 83.9% Sp) using only features from SpO_2, whereas the performance increased up to 84.9% accuracy (88.4% Se, 83.6% Sp) when features from SpO_2 and pulse rate were used jointly. In the study by Sahadan et al. [49], pulse rate time series from unattended pulse oximetry was analysed to assist in the management of childhood OSAS (AHI $\geq$ 1 event/h). The quantification of pulse rate increases of 15 bpm (PRI-15) reached the highest performance, achieving 18.0% sensitivity and 97.0% specificity when a cut-off of PRI-15 > 35/h was used. In a previous study by our group [18], ODI3 from nocturnal unsupervised SpO_2 was combined with spectral measures from at-home airflow recordings to characterise children with suspected OSAS (OAHI $\geq$ 3 events/h). A LR model from stepwise feature selection reached 86.3% accuracy (85.9% Se, 87.4% Sp). Cohen and De Chazal [17] analysed a large dataset composed of 288 children showing suspicion of suffering from OSAS. ECG and SpO_2 from unattended PSG at home were automatically processed in order to detect every individual apnoeic event. An accuracy of 74.7% (39.6% Se, 76.4% Sp) was reached under an epoch-based classification approach using a LDA model only composed of features from time-frequency analysis of ECG. Conversely, the accuracy decreases up to 66.7% (58.1% Se, 67.0% Sp) when statistics from SpO_2 were added to the model. In a recent study by our group [19], single-scale non-linear measures of entropy, complexity and variability were combined with conventional statistics, spectral features and oximetric indices to parameterise unattended SpO_2 recordings. Optimum LR models were composed from stepwise feature selection for different cut-offs for the disease, reaching 83.4% accuracy (82.9%, 84.4%) for a clinical threshold of OAHI $\geq$ 3 events/h. In the present research, we reached similar diagnostic performance using MSE as unique complement of conventional oximetric indices, suggesting that different nonlinear methods other than SampEn, central tendency measure (CTM) or Lempel-Ziv complexity (LZC) are also able to provide relevant information in the context of paediatric OSAS detection from oximetry.

A trade-off between the reduction of complexity of the diagnostic methodology by means of simplified techniques and the diagnostic accuracy have to be taken into account. Previous studies reported contradictory data regarding the complementarity of information from different signals in the context of paediatric OSAS. Some studies [16,18] showed a slight-to-moderate performance increase when using jointly features from oximetry and other cardiorespiratory signals such as pulse rate or airflow, whereas other studies [17] reported a significant decrease in performance when oximetry and ECG recordings are combined. Similarly, a recent study by Álvarez et al. suggested that automated analysis of unattended oximetry at home might be as accurate as manual scoring of at-home respiratory polygraphy, particularly when using low OAHI cut-off points for a positive diagnosis of the disease [19]. Therefore, additional robust evidences are still needed to define the best combination of cardiorespiratory signals in order to design an accurate as well as simplified screening tool for paediatric OSAS.

Some limitations should be taken into account to be able to properly generalise our conclusions. Firstly, a larger population would allow for optimal design and assessment of the proposed LR models. Notwithstanding, a bootstrapping approach was conducted both for feature selection and classification in order to overcome this drawback. In the same way, a wider dataset would let us a better characterisation of overnight oximetry dynamics by means of MSE for childhood OSAS. Nevertheless, our results revealed a quite consistent trend of average MSE curves for OSAS-negative and OSAS-positive groups, as well as statistically significant differences between groups for almost all MSE-based parameters. Regarding feature extraction from overnight oximetry, our findings suggest that multiscale processing methods provide relevant information from oximetric recordings. In this regard, additional time-scale techniques, such as the wavelet transform, could provide useful as well

as complementary features to MSE and conventional clinical indices in the framework of childhood OSAS detection from oximetry. Finally, the proposed methodology focused on binary classification. Although this is a very useful approach in order to implement automated screening tools for the disease, it would be very interesting to develop a pattern recognition scheme aimed at classifying patients into the four common categories of severity, i.e., non-OSAS, mild, moderate and severe.

LR could be considered the reference classifier in the context of automated pattern recognition to assist in childhood OSAS. Linear discriminant analysis (LDA) [16,17,50,51] and LR [1,18,19,52] have been predominantly used for binary classification of children suspected of suffering from the disease. LDA assumes that all the input variables show normal distribution and equal variances, assumptions that are not always consistent in real-world pattern classification tasks. LR provides a more general approach that fits better to the characteristics of the problem under study. Nevertheless, additional automated pattern recognition techniques such as decision trees, artificial neural networks or support vector machines, which have demonstrated its usefulness in the context of adult OSAS [31,53–56], need to be assessed in the context of paediatric OSAS.

5. Conclusions

A comprehensive analysis of overnight oximetry dynamics along increasing time scales demonstrated an ability to provide additional information about the influence of desaturations in the characterisation of paediatric OSAS. SampEn values from the MSE profiles of OSAS-positive children were consistently higher than the entropy values obtained for non-OSAS patients for all scales. Particularly, MSE-derived parameters reached significant statistical differences between both groups for small time scales ($\tau \leq 6$). An exhaustive feature selection methodology confirmed that MSE analysis provided relevant as well as complementary information to conventional oximetric indices. A LR model composed of optimum features properly selected from both MSE and conventional analyses outperformed each of these approaches taken individually. Therefore, MSE may be useful to improve the diagnostic ability of unattended oximetry as a simplified screening test for childhood OSAS.

Author Contributions: Andrea Crespo performed statistical analyses, interpreted the results, and wrote the manuscript. Daniel Álvarez and Félix del Campo designed the study, performed automated analysis of oximetry and interpreted the results. Gonzalo C. Gutiérrez-Tobal, Fernando Vaquerizo-Villar, Verónica Barroso-García and Roberto Hornero pre-processed the recordings and interpreted the results. María Luz Alonso-Álvarez and Joaquín Terán-Santos recruited and diagnosed all patients involved in the study. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chang, L.; Wu, J.; Cao, L. Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea. *Int. J. Pediatr. Otorhinolaryngol.* **2013**, *77*, 365–371. [CrossRef] [PubMed]

2. Marcus, C.L.; Brooks, L.J.; Ward, S.D. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics* **2012**, *130*, 714–755. [CrossRef] [PubMed]

3. Capdevila, O.S.; Kheirandish-Gozal, L.; Dayyat, E.; Gozal, D. Pediatric obstructive sleep apnea: Complications, management, and long-term outcomes. *Proc. Am. Thorac. Soc.* **2008**, *5*, 274–282. [CrossRef] [PubMed]

4. Kheirandish-Gozal, L. What is "abnormal" in pediatric sleep? *Respir. Care* **2010**, *55*, 1366–1376. [PubMed]

5. Tarasiuk, A.; Simon, T.; Tal, A.; Reuveni, H. Adenotonsillectomy in children with obstructive sleep apnea syndrome reduces health care utilization. *Pediatrics* **2004**, *113*, 351–356. [CrossRef] [PubMed]

6. Kadmon, G.; Shapiro, C.M.; Chung, S.A.; Gozal, D. Validation of a pediatric obstructive sleep apnea screening tool. *Int. J. Pediatr. Otorhinolaryngol.* **2013**, *77*, 1461–1464. [CrossRef] [PubMed]

7. Lesser, D.J.; Haddad, G.G.; Bush, R.A.; Pian, M.S. The utility of a portable recording device for screening of obstructive sleep apnea in obese adolescents. *J. Clin. Sleep Med.* **2012**, *8*, 271–277. [CrossRef] [PubMed]

8. Katz, E.S.; Ron, B.M.; D'ambrosio, C.M. Obstructive sleep apnea in infants. *Am. J. Respir. Crit. Care Med.* **2012**, *185*, 805–816. [CrossRef] [PubMed]

9. Alonso-Álvarez, M.L.; Terán-Santos, J.; Cordero-Guevara, J.A. Reliability of respiratory polygraphy for the diagnosis of sleep apnea-hypopnea syndrome in children. *Arch. Bronconeumol.* **2008**, *44*, 318–323. [CrossRef]

10. Kaditis, A.G.; Alonso-Alvarez, M.L.; Boudewyns, A. Obstructive sleep disordered breathing in 2–18 year-old children: Diagnosis and management. *Eur. Respir. J.* **2016**, *47*, 69–94. [CrossRef] [PubMed]

11. Alonso-Álvarez, M.L.; Terán-Santos, J.; Ordax-Carbajo, E. Reliability of home respiratory polygraphy for the diagnosis of sleep apnea in children. *Chest* **2015**, *147*, 1020–1028. [CrossRef] [PubMed]

12. Brockmann, P.E.; Schaefer, C.; Poets, A.; Poets, C.F.; Urschitz, M.S. Diagnosis of obstructive sleep apnea in children: A systematic review. *Sleep Med. Rev.* **2013**, *17*, 331–340. [CrossRef] [PubMed]

13. Alonso-Álvarez, M.L.; Navazo-Egüia, A.I.; Cordero-Guevara, J.A.; Ordax-Carbajo, E.; de La Mata, G.; Barba-Cermeño, J.L.; Terán-Santos, J. Respiratory polygraphy for follow-up of obstructive sleep apnea in children. *Sleep Med.* **2012**, *13*, 611–615. [CrossRef] [PubMed]

14. Brouillette, R.T.; Morielli, A.; Leimanis, A.; Waters, K.A.; Luciano, R.; Ducharme, F.M. Nocturnal pulse oximetry as an abbreviated testing modality for pediatric obstructive sleep apnea. *Pediatrics* **2000**, *105*, 405–412. [CrossRef] [PubMed]

15. Kaditis, A.; Kheirandish-Gozal, L.; Gozal, D. Pediatric OSAS: Oximetry can provide answers when polysomnography is not available. *Sleep Med. Rev.* **2016**, *27*, 96–105. [CrossRef] [PubMed]

16. Garde, A.; Dehkordi, P.; Karlen, W.; Wensley, D.; Ansermino, J.M.; Dumont, G.A. Development of a screening tool for sleep disordered breathing in children using the Phone OximeterTM. *PLoS ONE* **2014**, *9*, e112959. [CrossRef] [PubMed]

17. Cohen, G.; de Chazal, P. Automated detection of sleep apnea in infants: A multi-modal approach. *Comput. Biol. Med.* **2015**, *63*, 118–123. [CrossRef] [PubMed]

18. Gutiérrez-Tobal, G.C.; Alonso-Álvarez, M.L.; Álvarez, D.; del Campo, F.; Terán-Santos, J.; Hornero, R. Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home. *Biomed. Signal Process Control* **2015**, *18*, 401–407. [CrossRef]

19. Álvarez, D.; Alonso-Álvarez, M.L.; Gutiérrez-Tobal, G.C.; Crespo, A.; Kheirandish-Gozal, L.; Hornero, R.; Gozal, D.; Terán-Santos, J.; Del Campo, F. Automated screening of children with obstructive sleep apnea using nocturnal oximetry: An alternative to respiratory polygraphy in unattended settings. *J. Clin. Sleep Med.* **2017**, *13*, 693–702. [CrossRef] [PubMed]

20. Kirk, V.G.; Bohn, S.G.; Flemons, W.W.; Remmers, J.E. Comparison of home oximetry monitoring with laboratory polysomnography in children. *Chest* **2003**, *124*, 1702–1708. [CrossRef] [PubMed]

21. Tsai, C.; Kang, C.; Su, M. Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. *Int. J. Pediatr. Otorhinolaryngol.* **2013**, *77*, 1286–1290. [CrossRef] [PubMed]

22. Pincus, S.M. Assessing serial irregularity and its implications for health. *Ann. N. Y. Acad. Sci.* **2001**, *954*, 245–267. [CrossRef]

23. Richman, J.S.; Moorman, J.R. Physiological time series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, 2039–2049.

24. Cohen, M.E.; Hudson, D.L.; Deedwania, P.C. Applying continuous chaotic modeling to cardiac signals analysis. *IEEE Eng. Med. Biol.* **1996**, *15*, 97–102. [CrossRef]

25. Stam, C.J. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clin. Neurophysiol.* **2005**, *116*, 2266–2301. [CrossRef] [PubMed]

26. Álvarez, D.; Hornero, R.; Abásolo, D.; del Campo, F.; Zamarrón, C. Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection. *Physiol. Meas.* **2006**, *27*, 399–412. [CrossRef] [PubMed]

27. Hornero, R.; Álvarez, D.; Abásolo, D.; del Campo, F.; Zamarrón, C. Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 107–113. [CrossRef] [PubMed]

28. Del Campo, F.; Hornero, R.; Zamarrón, C.; Abasolo, D.E.; Álvarez, D. Oxygen saturation regularity analysis in the diagnosis of obstructive sleep apnea. *Artif. Intell. Med.* **2006**, *37*, 111–118. [CrossRef] [PubMed]

29. Álvarez, D.; Hornero, R.; Marcos, J.V.; del Campo, F. Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2816–2824. [CrossRef] [PubMed]

30. Álvarez, D.; Hornero, R.; Marcos, J.V.; del Campo, F. Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis. *Med. Eng. Phys.* **2012**, *34*, 1049–1057. [CrossRef] [PubMed]

31. Álvarez, D.; Hornero, R.; Marcos, J.V.; Wessel, N.; Penzel, T.; Glos, M.; del Campo, F. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of sleep apnea diagnosis. *Int. J. Neural. Syst.* **2013**, *23*, 1–18. [CrossRef] [PubMed]

32. Marcos, J.V.; Hornero, R.; Nabney, I.T.; Álvarez, D.; Gutiérrez-Tobal, G.C.; del Campo, F. Regularity analysis of nocturnal oximetry recordings to assist in the diagnosis of sleep apnoea syndrome. *Med. Eng. Phys.* **2016**, *38*, 216–224. [CrossRef] [PubMed]

33. Costa, M.D.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, 021906. [CrossRef] [PubMed]

34. Vanderlei, F.M.; Vanderlei, L.C.M.; de Abreu, L.C.; Garner, D.M. Entropic analysis of HRV in obese children. *Int. Arch. Med.* **2015**, *8*, 1–9. [CrossRef]

35. Costa, M.D.; Peng, C.K.; Goldberger, A.L. Multiscale analysis of heart rate dynamics: Entropy and time irreversibility measures. *Cardiovasc. Eng.* **2008**, *8*, 88–93. [CrossRef] [PubMed]

36. Costa, M.D.; Peng, C.K.; Goldberger, A.L.; Hausdorff, J.M. Multiscale entropy analysis of human gait dynamics. *Phys. A* **2003**, *330*, 53–60. [CrossRef]

37. Escudero, J.; Abásolo, D.; Hornero, R.; Espino, P.; López, M. Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy. *Physiol. Meas.* **2006**, *27*, 1091–1106. [CrossRef] [PubMed]

38. Takahashi, T.; Cho, R.Y.; Mizuno, T.; Kikuchi, M.; Murata, T.; Takahashi, K.; Wada, Y. Antipsychotics reverse abnormal EEG complexity in drug-naive schizophrenia: A multiscale entropy analysis. *Neuroimage* **2010**, *51*, 173–182. [CrossRef] [PubMed]

39. Da Costa, C.S.; Placek, M.M.; Czosnyka, M.; Cabella, B.; Kasprowicz, M.; Austin, T. Complexity of brain signals is associated with outcome in preterm infants. *J. Cereb. Blood Flow Metab.* **2017**, in press.

40. Pan, W.; Su, M.; Wu, H.; Su, T.; Lin, M.; Sun, C. Multiscale entropic assessment of autonomic dysfunction in patients with obstructive sleep apnea and therapeutic impact of continuous positive airway pressure treatment. *Sleep Med.* **2016**, *20*, 12–17. [CrossRef] [PubMed]

41. Roebuck, A.; Clifford, G.D. Comparison of standard and novel signal analysis approaches to obstructive sleep apnea classification. *Front. Bioeng. Biotech.* **2015**, *3*. [CrossRef] [PubMed]

42. Gutiérrez-Tobal, G.C.; Álvarez, D.; Gomez-Pilar, J.; del Campo, F.; Hornero, R. Assessment of time and frequency domain entropies to detect sleep apnea in heart rate variability recordings from men and women. *Entropy* **2015**, *17*, 123–141. [CrossRef]

43. Berry, R.B.; Budhiraja, R.; Gottlieb, D.J. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. *J. Clin. Sleep Med.* **2012**, *8*, 597–619. [CrossRef] [PubMed]

44. Iber, C.; Ancoli-Israel, S.; Chesson, A.L.; Quan, S.F. *The AASM Manual for the Scoring of Sleep and Associated Events. Rules, Terminology and Technical Specifications*; American Academy of Sleep Medicine: Westchester, IL, USA, 2007.

45. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*; Wiley: London, UK, 2000.

46. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [CrossRef] [PubMed]

47. Gutiérrez-Tobal, G.C.; Hornero, R.; Álvarez, D.; Marcos, J.V.; del Campo, F. Linear and nonlinear analysis of airflow recordings to help in sleep apnoea–hypopnoea syndrome diagnosis. *Physiol. Meas.* **2012**, *33*, 1261–1275. [CrossRef] [PubMed]

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

49. Sahadan, D.Z.; Davey, M.J.; Horne, R.S.; Nixon, G.M. Improving detection of obstructive sleep apnoea by overnight oximetry in children using pulse rate parameters. *Sleep Breath* **2015**, *19*, 1409–1414. [CrossRef] [PubMed]

50. Gil, E.; Bailón, R.; Vergara, J.M.; Laguna, P. PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1079–1088. [CrossRef] [PubMed]

51. Lázaro, J.; Gil, E.; Vergara, J.M.; Laguna, P. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 240–246. [CrossRef] [PubMed]

52. Wu, D.; Li, X.; Guo, X.; Qin, J.; Li, S. A simple diagnostic scale based on the analysis and screening of clinical parameters in paediatric obstructive sleep apnoea hypopnea syndrome. *J. Laryngol. Otol.* **2017**, *131*, 363–367. [CrossRef] [PubMed]

53. Sánchez-Morillo, D.; López-Gordo, M.A.; León, A. Novel multiclass classification for home-based diagnosis of sleep apnea hypopnea syndrome. *Expert Syst. Appl.* **2014**, *41*, 1654–1662. [CrossRef]

54. Marcos, J.V.; Hornero, R.; Álvarez, D.; del Campo, F.; Zamarrón, C.; López, M. Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnea syndrome from nocturnal oximetry. *Comput. Meth. Programs Biomed.* **2008**, *92*, 79–89. [CrossRef] [PubMed]

55. Marcos, J.V.; Hornero, R.; Álvarez, D.; Nabney, I.T.; del Campo, F.; Zamarrón, C. The classification of oximetry signals using Bayesian neural networks to assist in the detection of obstructive sleep apnoea syndrome. *Physiol. Meas.* **2010**, *31*, 375–394. [CrossRef] [PubMed]

56. Karandikar, K.; Le, T.Q.; Sa-ngasoongsong, A.; Wongdhamma, W.; Bukkapatnam, S.T. Detection of sleep apnea events via tracking nonlinear dynamic cardio-respiratory coupling from electrocardiogram signals. In Proceedings of the 6th International IEEE/EMBS Conference on Neural Engineering, San Diego, CA, USA, 6–8 November 2013; pp. 1358–1361.