

# Automated analysis of the oximetry signal to simplify the diagnosis of pediatric sleep apnea

From feature-engineering to deep-learning approaches

Fernando Vaquerizo Villar



# Doctoral Thesis





---

# Universidad de Valladolid

DOCTORAL PROGRAM OF  
INFORMATION AND TELECOMMUNICATION TECHNOLOGIES

DOCTORAL THESIS

**AUTOMATED ANALYSIS OF THE OXIMETRY SIGNAL  
TO SIMPLIFY THE DIAGNOSIS OF PEDIATRIC SLEEP  
APNEA: FROM FEATURE-ENGINEERING TO  
DEEP-LEARNING APPROACHES**

THESIS PRESENTED BY **D. FERNANDO VAQUERIZO VILLAR**  
TO APPLY FOR THE *Ph.D. degree*  
FROM THE *University of Valladolid*

DIRECTED BY:

**DR. ROBERTO HORNERO SÁNCHEZ AND DR. DANIEL ÁLVAREZ GONZÁLEZ**

2021

VALLADOLID, SPAIN



*A todos los que me habéis apoyado estos años*



# Defense

|              |  |
|--------------|--|
| TÍTULO       | Automated analysis of the oximetry signal to simplify the diagnosis of pediatric sleep apnea: from feature-engineering to deep-learning approaches |
| AUTOR        | D. Fernando Vaquerizo Villar   |
| DIRECTORES   | Dr. D. Roberto Hornero Sánchez y Dr. D. Daniel Álvarez González  |
| DEPARTAMENTO | Teoría de la Señal y Comunicaciones e Ingeniería Telemática  |

---

## Tribunal

|            |                            |
|------------|----------------------------|
| PRESIDENTE | Dr. D.                     |
| VOCALES    | Dr. D.<br>Dr. D.<br>Dr. D. |
| SECRETARIO | Dr. D.                     |

acuerda otorgarle la calificación de

En Valladolid, a      de      del





---

## Universidad de Valladolid

---

Escuela Técnica Superior de Ingenieros de Telecomunicación  
Dpto. de Teoría de la Señal y Comunicaciones e Ingeniería Telemática

### **Research Stay for the International Mention**

City: Berlin (Alemania)

Institution: *Charité UniversitätsMedizin* Berlin

Center: Interdisciplinary Center of Sleep Medicine

Dates: 01/05/2019-31/07/2019

Duration: 92 days (3 months)

Supervisor: Prof. Dr. Thomas Penzel





# Agradecimientos

En primer lugar, deseo agradecer a mis directores de Tesis, los Doctores Roberto Hornero Sánchez y Daniel Álvarez González, por dirigirme durante todos estos años en mi actividad investigadora. Desde que hace siete años empezaron a supervisarme el Trabajo Fin de Grado han confiado en mí, mostrando un trato humano excelente en todo momento. Su apoyo, implicación y, sobretodo, su paciencia, han sido muy importantes para poder concluir esta Tesis Doctoral.

También me gustaría dar las gracias a todos los compañeros del Grupo de Ingeniería Biomédica (GIB), a los que están y a los que se fueron, a Jesús, Carlos, María, Gonzalo, Javier G., Rebeca, Alejandro, Luis F., Víctor M., Pablo, Saúl, Verónica, Roberto O., Eduardo, Jorge, Adrián, Víctor R., Víctor G., Aarón, Javier O., Marcos y Celia. Además de ser un grupo de investigación excelente para realizar la tesis doctoral, el GIB ofrece un ambiente de trabajo excelente, desde las discusiones metodológicas a las bromas y anécdotas de los cafés y los *'happy meals'*, que solían acabar en las portadas de las mejores revistas. Estoy especialmente agradecido a Gonzalo y Verónica, que les ha tocado sufrir leyendo cada uno de mis artículos y comunicaciones científicas, contribuyendo en gran medida a mi formación. Tampoco puedo olvidarme de Jesús, Carlos y María, que me han ayudado enormemente en las tareas docentes que he realizado estos años.

Quiero extender también mi gratitud al Dr. David Gozal y a la Dra. Leila Kheirandish Gozal, del *University of Missouri School of Medicine*, así como al Dr. Félix del Campo del Hospital Universitario Río Hortega. Sus conocimientos médicos y colaboración han sido imprescindibles para llevar a cabo esta tesis.

I would also like to thank Dr. Thomas Penzel and the members from the Interdisciplinary Center of Sleep Medicine from the *Charité UniversitätsMedizin Berlin* and the Cardiovascular Physics Group from the Humboldt University in Berlin for accepting me in one of the best sleep research centers in the world. My research stay in Berlin was an enriching experience, which helped me to grow

both personally and professionally. Thanks Thomas for giving me the opportunity to collaborate in international research projects. I am also very grateful to Niels Wessel, Jan Kraemer, and Hua Qin for their help, advice, and feedback with my work, as well as to all the nice people in Berlin who made me feel like home.

Finalmente, quiero dar las gracias a mi familia, que fueron los primeros que me animaron a seguir este camino. He tenido los mejores padres, hermanos, abuelos, tíos, primos y demás familia, que se han esforzado desde el día que nací para que pudiera llegar hasta aquí, aunque algunos no vayáis a poder verlo. Tampoco me puedo olvidar de mi novia, ni de mis amigos, que se han convertido en la segunda familia por su preocupación, apoyo y ayuda a que lo difícil fuera fácil. En este sentido, estoy especialmente agradecido a Alex B. por su ayuda con las ilustraciones del documento.

# Abstract

Obstructive sleep apnea (OSA) is a high prevalent respiratory disorder in the pediatric population (1%-5%). Untreated pediatric OSA is associated with significant adverse consequences affecting metabolic, cardiovascular, neurocognitive, and behavioral systems, thus resulting in a decline of overall health and quality of life. Consequently, it is of paramount importance to accelerate the diagnosis and treatment in these children.

Overnight polysomnography (PSG) is the gold standard to diagnose OSA in children. This test requires an overnight stay of pediatric subjects in a specialized sleep laboratory, as well as the recording of up to 32 biomedical signals. These recordings are used to quantify respiratory events in order to obtain the apnea-hyponea index (AHI), which is used to establish pediatric OSA severity. Nonetheless, PSG is technically complex, time-consuming, costly, highly intrusive for the children, and relatively unavailable, thus delaying the access for both the diagnosis and treatment. Consequently, simplified diagnostic techniques become necessary.

In an effort to overcome these drawbacks and increase the accessibility of pediatric OSA diagnosis, many simplified alternative procedures have been developed. Among these, a common approach is the analysis of the blood oxygen saturation ( $\text{SpO}_2$ ) signal from overnight oximetry due to its easy acquisition and interpretation, as well as its suitability for children. Many studies have demonstrated the utility of the automated analysis of  $\text{SpO}_2$  recordings to help in adult OSA diagnosis. Conversely, the preceding studies focused on pediatric patients reported lower accuracies than those reached in the case of adults, suggesting the need to seek novel signal processing algorithms that provide additional information from the  $\text{SpO}_2$  signal for the particularities of childhood OSA.

In the present Doctoral Thesis, we hypothesize that the application of novel feature extraction and deep-learning algorithms could increase the diagnostic

ability of the oximetry signal in the context of pediatric OSA. Consequently, the general objective of this Doctoral Thesis is to design, develop, and assess novel clinical decision-support models in the context of childhood OSA based on the automated analysis of the SpO<sub>2</sub> signal.

To achieve this goal, 3196 SpO<sub>2</sub> recordings from three different databases of children were involved: (i) the Childhood Adenotonsillectomy Trial (CHAT) database, (ii) the University of Chicago (UofC) database, and (iii) the Burgos University Hospital (BUH) database. These recordings were automatically analyzed using feature-engineering and deep-learning methodologies. On one hand, feature-engineering methodologies were conducted in three phases. First, a set of OSA-related features were extracted from the SpO<sub>2</sub> signal using different analytical approaches: statistical parameters, conventional oximetric indices, frequency domain methods, and nonlinear analysis. Particularly, we have evaluated the usefulness of bispectrum, wavelet, and detrended fluctuation analysis (DFA) to provide additional and complementary information to conventional approaches linked to pediatric OSA and its severity. As a second step, the fast correlation-based filter algorithm was applied to select optimum subsets of features that provide relevant and non-redundant information related to pediatric OSA and its severity. Finally, pattern recognition algorithms were applied to these optimum subsets of features in order to estimate pediatric OSA and its severity. To this effect, different approaches were explored: binary (OSA negative vs. OSA positive) and multi-class (OSA severity degrees) classification and regression (estimation of the AHI). On the other hand, a deep-learning methodology based on convolutional neural networks (CNN) was employed to automatically estimate pediatric OSA severity from raw oximetry data.

A high performance was obtained with both the proposed feature-engineering and deep-learning approaches. Thus, in the case of feature-engineering, our results showed that the application of bispectrum, wavelet, and DFA allowed to obtain features that provide relevant and complimentary information to conventional methods regarding OSA-related changes in the oximetry signal. Specifically, a multiclass multi-layer perceptron (MLP) neural network was fed with an optimum subset composed of the mean amplitude of the bispectrum, the mean of the bispectrum invariant, variables from the power spectral density (PSD), the 3% oxygen desaturation index (ODI3), and anthropometric variables. This MLP model reached 81.3% and 85.3% accuracy (Acc) in the AHI cutoffs of  $\geq 5$  (moderate OSA) and  $\geq 10$  (severe OSA) events per hour (e/h), respectively, outperforming a MLP model trained without bispectral features. In

addition, a binary support vector machines (SVM) model was trained with an optimum subset composed of the skewness and energy of the wavelet coefficients in the 9th detail level and the wavelet entropy, together with ODI3, statistical moments in the time domain and PSD-derived parameters. This optimum SVM model showed a high capability as a screening tool to detect moderate-to-severe pediatric OSA ( $AHI \geq 5$  e/h), with 84.0% Acc and a positive likelihood ratio of 14.6, which are higher than the obtained with every single feature. Finally, a regression MLP model trained with a subset of features composed of the ODI3 and the slope in the first scaling region of the DFA obtained 82.7%, 81.9%, and 91.1% Acc for the AHI cutoffs of 1 e/h, 5 e/h, and 10 e/h, respectively. This regression MLP model outperformed the conventional ODI3, commonly used in clinical settings.

On the other hand, it was found that deep-learning approaches can automatically learn additional information from the oximetry signal linked to apneic events. A CNN-based deep-learning architecture trained to estimate the AHI from raw SpO<sub>2</sub> segments reached 0.515, 0.422, and 0.423 Cohen's kappa in three independent datasets (CHAT, UofC, and BUH). In addition, the proposed CNN-based model reached high accuracies for the AHI severity cutoffs of 1 e/h (77.6%, 80.1%, and 79.2%), 5 e/h (97.4%, 83.9%, and 83.5%), and 10 e/h (97.8%, 92.3%, and 91.3%) in the CHAT, UofC, and BUH datasets. This CNN-based model achieved a higher overall performance than feature-engineering approaches. The application of this deep-learning model as a screening protocol would avoid the need for 73.7% (CHAT), 50.0% (UofC), and 45.9% (BUH) of full PSGs in pediatric subjects.

Our proposed methodologies also achieved a higher overall performance than state-of-the-art studies, especially for moderate-to-severely affected pediatric subjects. Therefore, the results obtained in this Doctoral Thesis suggest that bispectrum, wavelet, and DFA are able to further characterize changes in the SpO<sub>2</sub> signal caused by apneic events in pediatric subjects. Furthermore, it is also concluded that deep-learning algorithms can learn complex features from oximetry dynamics that allow to enhance the diagnostic capability of nocturnal oximetry in the context of childhood OSA. We feel that these studies could contribute to the use of clinical screening tools to diagnose pediatric OSA based on the automated analysis of the oximetry signal, aiming at providing an early and timely diagnosis and treatment of the affected children.



# Acronyms

|        |   |
|--------|---|
| AASM   | American Academy of Sleep Medicine                          |
| AAP    | American Academy of Pediatrics                              |
| Acc    | Accuracy  |
| AF     | Airflow   |
| AHI    | Apnea-hypopnea index  |
| ApEn   | Approximate entropy   |
| AUC    | Area under the ROC curve                                    |
| BI     | Band of interest  |
| BMI    | Body mass index   |
| BN     | Batch normalization   |
| BO-TPE | Bayesian optimization with tree-structured Parzen estimator |
| BUH    | Burgos University Hospital                                  |
| CHAT   | Childhood Adenotonsillectomy Trial                          |
| CNN    | Convolutional neural networks                               |
| CTM    | Central tendency measure                                    |
| CWT    | Continuous wavelet transform                                |
| DFA    | Detrended fluctuation analysis                              |
| DWT    | Discrete wavelet transform                                  |
| ECG    | Electrocardiogram   |
| EEG    | Electroencephalogram  |
| EMG    | Electromyogram  |
| EOG    | Electrooculogram  |
| FCBF   | Fast correlation-based filter                               |
| FN     | False negatives   |
| FP     | False positives   |
| HHb    | Deoxymehoglobin   |
| HR     | Heart rate  |

---

|                   |                                     |
|-------------------|-------------------------------------|
| ICC               | Intra-class correlation coefficient |
| JCR               | Journal Citation Reports            |
| Kappa             | Cohen's kappa                       |
| LDA               | Linear discriminant analysis        |
| LR                | Logistic regression                 |
| LR+               | Positive likelihood ratio           |
| LR-               | Negative likelihood ratio           |
| LZC               | Lempel-Ziv complexity               |
| MLP               | Multi-layer perceptron              |
| MOS               | McGill oximetry score               |
| MSE               | Multiscale entropy                  |
| NPV               | Negative predictive value           |
| O <sub>2</sub> Hb | Oxyhemoglobin                       |
| ODI               | Oxygen desaturation index           |
| ODI3              | 3% Oxygen desaturation index        |
| ODI4              | 4% Oxygen desaturation index        |
| OSA               | Obstructive sleep apnea             |
| POW               | Pulse oximetry watch                |
| PPG               | Photoplethysmography                |
| PPV               | Positive predictive value           |
| PSD               | Power spectral density              |
| PSG               | Polysomnography                     |
| QDA               | Quadratic discriminant analysis     |
| ReLU              | Rectified linear unit               |
| RMSE              | Root mean square error              |
| ROC               | Receiving-operating characteristics |
| RP                | Respiratory polygraphy              |
| SampEn            | Sample entropy                      |
| Se                | Sensitivity                         |
| Sp                | Specificity                         |
| SpO <sub>2</sub>  | Blood oxygen saturation             |
| SAHS              | Sleep apnea-hypopnea syndrome       |
| STFT              | Short time Fourier transform        |
| SU                | Symmetrical uncertainty             |
| SVM               | Support vector machine              |
| TN                | True negatives                      |
| TP                | True positives                      |

UofC  
WT

University of Chicago  
Wavelet transform



# Contents

|  |           |
|--|-----------|
| <b>Abstract</b>  | <b>I</b>  |
| <b>Acronyms</b>  | <b>V</b>  |
| <b>1 Introduction</b>  | <b>1</b>  |
| 1.1 Compendium of publications: thematic consistency . . . . .   | 2         |
| 1.2 The biomedical signal processing framework: feature-engineering<br>and deep learning . . . . .       | 9         |
| 1.3 Pediatric Obstructive Sleep Apnea (OSA) . . . . .  | 10        |
| 1.4 Pediatric OSA diagnosis: Polysomnography (PSG) . . . . .   | 11        |
| 1.5 Alternatives to PSG . . . . .  | 12        |
| 1.5.1 Overnight oximetry . . . . .   | 13        |
| 1.6 State-of-the-art: Automated analysis of the oximetry signal to di-<br>agnose pediatric OSA . . . . . | 14        |
| 1.6.1 Conventional oximetric indices . . . . .   | 14        |
| 1.6.2 Automated signal processing methods . . . . .  | 16        |
| <b>2 Hypotheses and objectives</b>   | <b>19</b> |
| 2.1 Hypotheses . . . . .   | 19        |
| 2.2 Objectives . . . . .   | 20        |
| <b>3 Subjects and signals under study</b>  | <b>21</b> |
| 3.1 Childhood Adenotonsillectomy Trial (CHAT) database . . . . .   | 22        |
| 3.2 University of Chicago (UofC) database . . . . .  | 23        |
| 3.3 Burgos University Hospital (BUH) database . . . . .  | 24        |
| <b>4 Methods</b>   | <b>27</b> |
| 4.1 Pre-processing . . . . .   | 27        |
| 4.2 Feature engineering . . . . .  | 29        |
| 4.2.1 Feature extraction . . . . .   | 29        |

|          |   |           |
|----------|---|-----------|
| 4.2.2    | Feature selection . . . . .   | 39        |
| 4.2.3    | Pattern recognition . . . . .   | 40        |
| 4.3      | Deep learning . . . . .   | 43        |
| 4.3.1    | Proposed convolutional neural network (CNN) model . . .   | 44        |
| 4.3.2    | CNN training and optimization process . . . . .   | 46        |
| 4.4      | Statistical analysis . . . . .  | 47        |
| 4.4.1    | Statistical hypothesis tests . . . . .  | 47        |
| 4.4.2    | Diagnostic performance metrics . . . . .  | 48        |
| 4.4.3    | Measures of agreement . . . . .   | 50        |
| 4.4.4    | Validation strategies . . . . .   | 51        |
| <b>5</b> | <b>Results</b>  | <b>53</b> |
| 5.1      | Application of novel feature-extraction algorithms . . . . .  | 53        |
| 5.1.1    | Bispectral analysis . . . . .   | 53        |
| 5.1.2    | Wavelet analysis . . . . .  | 56        |
| 5.1.3    | Detrended fluctuation analysis . . . . .  | 60        |
| 5.2      | Application of deep-learning techniques . . . . .   | 62        |
| <b>6</b> | <b>Discussion</b>   | <b>69</b> |
| 6.1      | Novel features to provide relevant and complementary informa-<br>tion from oximetry recordings . . . . .                      | 69        |
| 6.1.1    | Bispectral analysis . . . . .   | 70        |
| 6.1.2    | Wavelet analysis . . . . .  | 71        |
| 6.1.3    | Detrended fluctuation analysis . . . . .  | 73        |
| 6.2      | A deep-learning based methodology to automatically extract the<br>relevant information from raw oximetry recordings . . . . . | 74        |
| 6.3      | Comparison of performance: feature-engineering, deep-learning,<br>and state-of-the-art . . . . .                              | 78        |
| 6.3.1    | Comparison between feature-engineering and deep-<br>learning approaches . . . . .   | 78        |
| 6.3.2    | Comparison with state-of-the-art studies . . . . .  | 80        |
| 6.4      | Limitations of the study . . . . .  | 83        |
| <b>7</b> | <b>Conclusions</b>  | <b>87</b> |
| 7.1      | Contributions . . . . .   | 87        |
| 7.2      | Main conclusions . . . . .  | 89        |
| 7.3      | Future research lines . . . . .   | 90        |

---

|          |  |            |
|----------|--|------------|
| <b>A</b> | <b>Papers included in this Doctoral Thesis</b>                 | <b>93</b>  |
| A.1      | Vaquerizo-Villar <i>et al.</i> (2018a) . . . . .               | 93         |
| A.2      | Vaquerizo-Villar <i>et al.</i> (2018b) . . . . .               | 105        |
| A.3      | Vaquerizo-Villar <i>et al.</i> (2018c) . . . . .               | 118        |
| A.4      | Vaquerizo-Villar <i>et al.</i> (2021) . . . . .                | 130        |
| <b>B</b> | <b>Scientific achievements</b>                                 | <b>143</b> |
| B.1      | Publications . . . . .   | 143        |
| B.1.1    | Papers indexed in the Journal Citation Reports (JCR) . . . . . | 143        |
| B.1.2    | International conferences . . . . .                            | 147        |
| B.1.3    | National conferences . . . . .                                 | 151        |
| B.2      | International internship . . . . .                             | 156        |
| B.3      | Grants . . . . .   | 157        |
| B.4      | Awards and honors . . . . .                                    | 158        |
| <b>C</b> | <b>Resumen en castellano</b>                                   | <b>161</b> |
| C.1      | Introducción . . . . .   | 161        |
| C.2      | Hipótesis y objetivos . . . . .                                | 163        |
| C.3      | Sujetos y señales . . . . .                                    | 164        |
| C.4      | Métodos . . . . .  | 165        |
| C.5      | Resultados y discusión . . . . .                               | 168        |
| C.6      | Conclusiones . . . . .   | 171        |
|          | <b>Bibliography</b>  | <b>175</b> |
|          | <b>Index</b>   | <b>185</b> |



# List of Figures

|            |   |    |
|------------|---|----|
| Figure 1.1 | Main contributions of the papers included in the compendium of publications, arranged along the automated signal processing methodologies developed. CMPB: Computer Methods and Programs in Biomedicine, IEEE JBHI: IEEE Journal of Biomedical and Health Informatics, PM: Physiological Measurement. . . . .   | 3  |
| Figure 1.2 | Examples of SpO <sub>2</sub> corresponding to: (a) a no OSA subject, (b) doubtful no OSA subject, (c) a mild OSA subject, (d) a doubtful mild OSA subject, (e) a moderate OSA subject, (f) a doubtful moderate OSA subject, (g) a severe OSA subject, (h) a doubtful severe OSA subject. In doubtful subjects, it is difficult to visually discriminate the OSA severity group from the overnight SpO <sub>2</sub> profile, leading to the need of non-subjective automated analyses. . . . . | 15 |
| Figure 4.1 | Scheme of the general signal processing methodology conducted in the study. . . . .   | 28 |
| Figure 4.2 | Example of the DFA profile of the SpO <sub>2</sub> recording of a subject. This figure has been derived from <a href="#">Vaquerizo-Villar et al. (2018c)</a> . . . . .  | 32 |
| Figure 4.3 | DWT computation. (a) Decomposition process of a signal using DWT. (b) Original SpO <sub>2</sub> signal, detail signals at each decomposition level and approximation signal at the maximum level of the decomposition. This figure has been taken from <a href="#">Vaquerizo-Villar et al. (2018c)</a> . . . . .  | 38 |
| Figure 4.4 | Overview of the proposed CNN-based deep-learning methodology. (a) Signal segmentation, (b) CNN architecture, and (c) AHI estimation. This figure has been taken from <a href="#">Vaquerizo-Villar et al. (2021)</a> . . . . .   | 44 |

|             |   |    |
|-------------|---|----|
| Figure 5.1  | Averaged amplitude in absolute value of the bispectrum for the three OSA severity groups. (a) negative OSA, (b) moderate OSA, and (c) severe OSA. This figure has been taken from <a href="#">Vaquerizo-Villar et al. (2018b)</a> . . . . . | 54 |
| Figure 5.2  | Confusion matrices of $MLP_{nobis}$ and $MLP_{bis}$ in the test set. 1: negative OSA, 2: moderate OSA; 3: severe OSA. This figure has been derived from <a href="#">Vaquerizo-Villar et al. (2018b)</a> . . . . .                           | 56 |
| Figure 5.3  | Histogram of the $D_9$ coefficients for each group in the optimization set. This figure has been derived from <a href="#">Vaquerizo-Villar et al. (2018c)</a> . . . . .   | 58 |
| Figure 5.4  | Averaged DFA profile by OSA severity group in the training dataset. This figure has been derived from <a href="#">Vaquerizo-Villar et al. (2018a)</a> . . . . .   | 61 |
| Figure 5.5  | Bland-Altman plots comparing (a) ODI3 and (b) $AHI_{MLP-DFA}$ with $AHI_{PSG}$ in the test set. This figure has been derived from <a href="#">Vaquerizo-Villar et al. (2018a)</a> . . . . .   | 63 |
| Figure 5.6  | Confusion matrices of ODI3 and $AHI_{MLP-DFA}$ in the test set. 1: no OSA; 2: mild OSA; 3: moderate OSA; 4: severe OSA. This figure has been derived from <a href="#">Vaquerizo-Villar et al. (2018a)</a> . . . . .                         | 63 |
| Figure 5.7  | Results of the hyperparameter optimization in the validation set. This figure has been taken from <a href="#">Vaquerizo-Villar et al. (2021)</a> . . . . .  | 65 |
| Figure 5.8  | Scatter plot comparing $AHI_{CNN}$ with $AHI_{PSG}$ in the CHAT, UofC, and BUH test databases. This figure has been taken from <a href="#">Vaquerizo-Villar et al. (2021)</a> . . . . .   | 65 |
| Figure 5.9  | Error distribution plot of $AHI_{CNN}$ in the CHAT, UofC, and BUH test databases. This figure has been taken from <a href="#">Vaquerizo-Villar et al. (2021)</a> . . . . .  | 66 |
| Figure 5.10 | Confusion matrices of $AHI_{CNN}$ in the CHAT, UofC, and BUH test datasets. 1: no OSA; 2: mild OSA; 3: moderate OSA; 4: severe OSA. This figure has been taken from <a href="#">Vaquerizo-Villar et al. (2021)</a> . . . . .                | 66 |
| Figure 6.1  | Screening protocol of the proposed CNN-based deep-learning model . . . . .  | 77 |

# List of Tables

|           |   |    |
|-----------|---|----|
| Table 3.1 | Clinical and sociodemographic data of the CHAT database. . . .  | 23 |
| Table 3.2 | Clinical and sociodemographic data of the complete UofC database (981 subjects). . . . .  | 24 |
| Table 3.3 | Clinical and sociodemographic data of the initial version of the UofC database (298 subjects). . . . .  | 24 |
| Table 3.4 | Clinical and sociodemographic data of the BUH database. . . .   | 25 |
| Table 5.1 | Feature values for the OSA severity groups (median [interquartile range]) in the feature optimization set. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2018b)</a> . . . . .           | 55 |
| Table 5.2 | Diagnostic ability of the $MLP_{nobis}$ and $MLP_{bis}$ models in the test set for the AHI cutoffs=5 e/h and 10 e/h. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2018b)</a> . . . . . | 57 |
| Table 5.3 | DWT-derived feature values for each group (median [interquartile range]) in the optimization set. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2018c)</a> . . . . .                    | 58 |
| Table 5.4 | Diagnostic ability of all the extracted features in the cross-validation set. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2018c)</a> . . . . .  | 59 |
| Table 5.5 | Diagnostic ability of the LR, SVM, and MLP classifiers in the cross-validation set. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2018c)</a> . . . . .                                  | 60 |
| Table 5.6 | Feature values for the OSA severity groups (median [interquartile range]) in the training set. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2018a)</a> . . . . .                       | 61 |
| Table 5.7 | Diagnostic ability of ODI3 and $AHI_{MLP-DEA}$ in the test set for the AHI cutoffs= 1 e/h, 5 e/h, and 10 e/h. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2018a)</a> . . . . .        | 62 |

|           |   |     |
|-----------|---|-----|
| Table 5.8 | Diagnostic ability of $AHI_{CNN}$ for the AHI cutoffs=1 e/h, 5 e/h, and 10 e/h in the CHAT, UofC, and BUH test databases. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2021)</a> . . . . . | 67  |
| Table 5.9 | Diagnostic performance of $AHI_{CNN}$ vs. ODI3 and $AHI_{MLP}$ in the CHAT, UofC, and BUH test databases. This table has been derived from <a href="#">Vaquerizo-Villar et al. (2021)</a> . . . . .                 | 67  |
| Table 6.1 | Summary of the comparison of $AHI_{CNN}$ with ODI3 and $AHI_{MLP}$ in the CHAT, UofC, and BUH test databases. . . . .   | 75  |
| Table 6.2 | Summary of the diagnostic performance of the methods developed during the compendium of publications . . . . .  | 79  |
| Table 6.3 | Summary of state-of-the-art studies based on conventional oximetric indices . . . . .   | 81  |
| Table 6.4 | Summary of state-of-the-art studies based on automated signal processing approaches . . . . .   | 82  |
| Tabla C.1 | Datos clínicos y sociodemográficos de la base de datos CHAT. . .  | 165 |
| Tabla C.2 | Datos clínicos y sociodemográficos de la base de datos UofC. . .  | 166 |
| Tabla C.3 | Datos clínicos y sociodemográficos de la base de datos BUH. . .   | 166 |
| Tabla C.4 | Resumen del rendimiento diagnóstico de los métodos desarrollados en el compendio de publicaciones. . . . .  | 169 |
| Tabla C.5 | Resumen del estado del arte de los estudios basados en enfoques de procesado automático de la señal de oximetría. . . . .   | 171 |

# Chapter 1

## Introduction

The present Doctoral Thesis aims to design, develop, and assess new automated algorithms to improve the diagnosis ability of the oximetry signal in the context of pediatric obstructive sleep apnea (OSA). During the study, several feature-engineering methodologies, as well as deep-learning approaches, have been evaluated. This investigation has led to four scientific articles that have been accepted or published in journals indexed in the Journal Citation Reports (JCR) from the Web of Science™. Specifically, three articles have been published in 2018. Additionally, a fourth article will be published in August 2021. This scientific production has led to write this study as a compendium of publications.

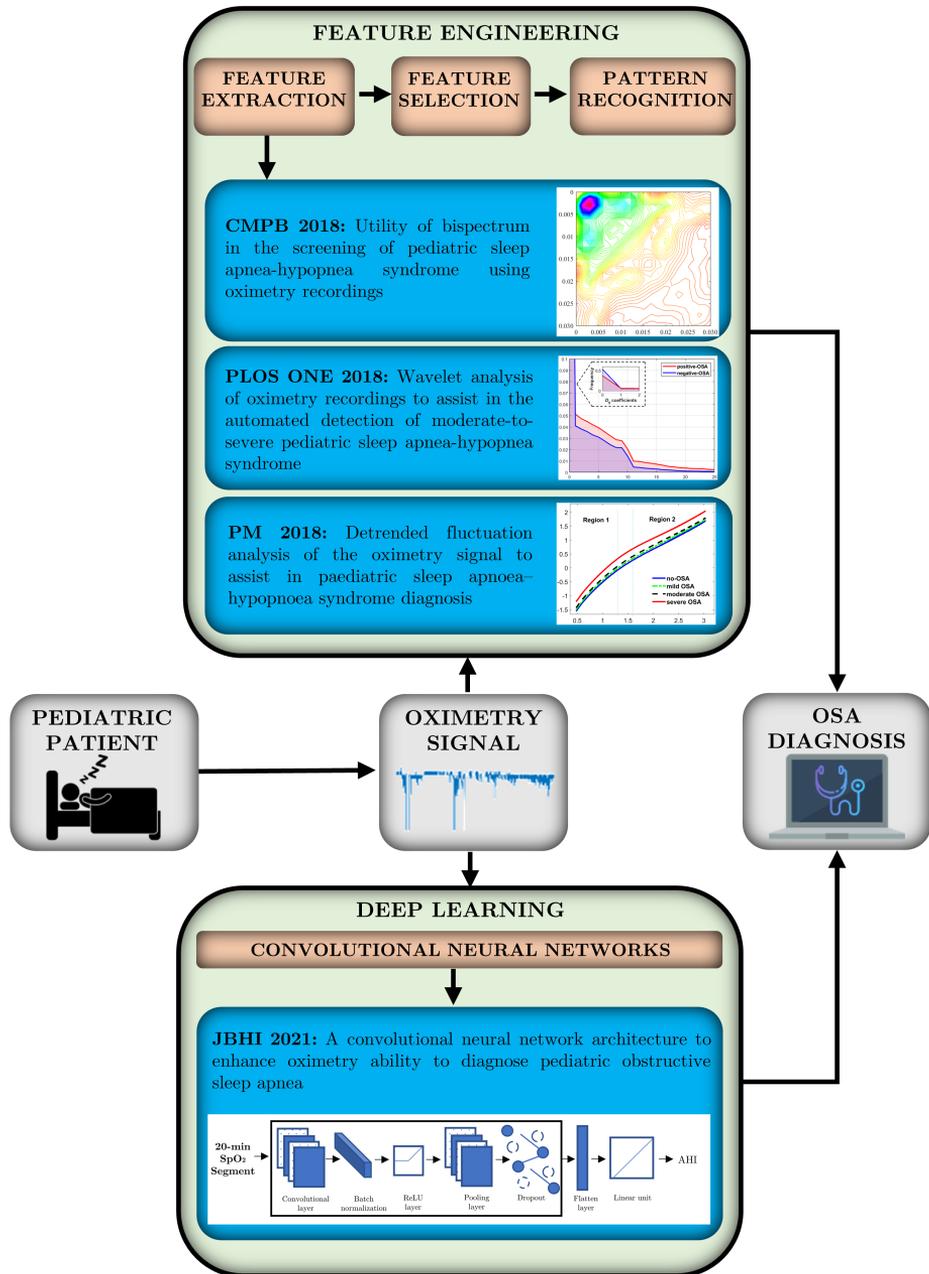
The thematic consistency of the articles included in this Doctoral Thesis is explained in Section 1.1. The general context of biomedical signal processing and deep learning is concisely described in Section 1.2. Section 1.3 provides a description of pediatric OSA, including its risks and adverse consequences. Section 1.4 focuses on the gold standard for pediatric OSA diagnosis, polysomnography (PSG), and its limitations. Finally, Section 1.5 is devoted to explain oximetry as an alternative to PSG for the diagnosis of pediatric OSA and Section 1.6 provides a description of state-of-the-art studies focused on the analysis of the oximetry signal as a simplified tool in the diagnosis of pediatric OSA.

## 1.1 Compendium of publications: thematic consistency

Pediatric OSA is a high prevalent disease (1%-5%) (Marcus et al., 2012). It is associated with many negative effects on the overall health and life quality of the affected children when it is untreated, including cardiometabolic malfunctioning and neurobehavioral abnormalities (Capdevila et al., 2008). Overnight PSG is the gold standard for pediatric OSA diagnosis (Marcus et al., 2012). Despite its effectiveness, PSG is costly, complex, highly intrusive, and lacks availability (Katz et al., 2012; Tan et al., 2015). This has prompted the search for simplified screening tests (Kaditis et al., 2016b; Marcus et al., 2012). One of these alternative tests is overnight oximetry, which measures the blood oxygen saturation ( $\text{SpO}_2$ ) signal with a pulse-oximeter, typically placed onto the end of a finger, thus being especially suitable for children (del Campo et al., 2018). A wide range of investigations have shown the utility of the automated analysis of the  $\text{SpO}_2$  signal as a clinically beneficial tool for the screening of OSA in adult patients (del Campo et al., 2018). Nonetheless, state-of-the-art studies focused on the automated analysis of the oximetry signal in the context of pediatric OSA followed a similar signal-processing methodology (del Campo et al., 2018), but achieving a inferior diagnostic performance than in the case of adult patients (del Campo et al., 2018).

In this context, the present Doctoral Thesis focuses on applying novel signal processing algorithms in order to enhance the diagnostic ability of the oximetry signal in the framework of pediatric OSA. All the papers that comprise the compendium of publications share this common thread. Figure 1.1 shows the thematic consistency and the main contributions of the papers included in this Thesis.

Earlier studies focused on the automated diagnosis of childhood OSA from the oximetry signal employed a three-stage feature-engineering methodology (del Campo et al., 2018). First, statistical analysis, conventional indices, frequency domain and nonlinear analysis methods were used to extract features from the  $\text{SpO}_2$  signal. Then, feature selection methods were applied to obtain optimum subsets of relevant and non-redundant features. Finally, pattern recognition algorithms were trained with the optimum subsets of features to detect pediatric OSA and its severity. Nonetheless, a previous study using a very large multicenter database of 4191 PSG sleep studies showed significant redundancy between the information extracted by conventional feature extraction methods and the



**Figure 1.1.** Main contributions of the papers included in the compendium of publications, arranged along the automated signal processing methodologies developed. CMPB: Computer Methods and Programs in Biomedicine, IEEE JBHI: IEEE Journal of Biomedical and Health Informatics, PM: Physiological Measurement.

3% oxygen desaturation index (ODI3), an oximetric variable commonly used for abbreviated screening purposes (Hornero et al., 2017). In this respect, the first three papers in chronological order (Vaquerizo-Villar et al., 2018a,b,c) were focused on the application of novel feature extraction algorithms to provide additional information from the oximetry signal. Given that the information extracted by conventional spectral analysis techniques is highly redundant, the first two papers (Vaquerizo-Villar et al., 2018b,c) were aimed at assessing the usefulness of two frequency domain techniques, bispectrum (Vaquerizo-Villar et al., 2018b) and wavelet analysis (Vaquerizo-Villar et al., 2018c), to provide discriminative frequency domain features from the oximetry signal. In the third paper (Vaquerizo-Villar et al., 2018a), we investigated if detrended fluctuation analysis (DFA), a nonlinear analysis method in the time domain, can extract complementary information from the oximetry signal linked to apneic events. As we will see, these feature extraction methods, novel in the context of pediatric OSA, have been found to provide complementary information to improve the diagnosis ability of the SpO<sub>2</sub> signal.

In contrast to the first three papers, which followed a feature-engineering methodology, the last paper of the Doctoral Thesis employed a deep-learning methodology to analyze the oximetry signal. Deep-learning approaches have emerged in the last years as a suitable tool to learn complex features from raw data using architectures with multiple layers of representation (LeCun et al., 2015). These algorithms have outperformed traditional feature-engineering approaches in many fields, including image recognition, natural language processing, and time series analysis (LeCun et al., 2015). Accordingly, in the last paper (Vaquerizo-Villar et al., 2021) we investigated the ability of convolutional neural networks (CNN), the most widely-used deep-learning technique, to automatically extract all the relevant information from the oximetry signal. The proposed CNN architecture was validated in a multicenter database of 3196 SpO<sub>2</sub> recordings, showing a high diagnostic ability, which outperformed conventional feature-engineering approaches.

The papers that compose the compendium of publications of the present Doctoral Thesis are included in the Appendix A. Titles, authors, and abstracts of each one, as well as the indexed journals in which they were accepted and published are shown below:

**Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings (Vaquerizo-Villar et al., 2018b).**

**Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-Garía, Andrea Crespo, Félix del Campo, David Gozal, and Roberto Hornero. *Computer Methods and Programs in Biomedicine*, vol. 156, p. 141-149, 2018. Impact factor in 2018: 3.424, Q1 in "COMPUTER SCIENCE, THEORY & METHODS" (JCR-WOS).

*Abstract:* Background and objective: The aim of this study was to assess the utility of bispectrum-based oximetry approaches as a complementary tool to traditional techniques in the screening of pediatric sleep apnea-hypopnea syndrome (SAHS). Methods: 298 blood oxygen saturation (SpO<sub>2</sub>) signals from children ranging 0–13 years of age were recorded during overnight polysomnography (PSG). These recordings were divided into three severity groups according to the PSG-derived apnea hypopnea index (AHI): AHI < 5 events per hour (e/h), 5 ≤ AHI < 10 e/h, AHI ≥ 10 e/h. For each pediatric subject, anthropometric variables, 3% oxygen desaturation index (ODI3) and spectral features from power spectral density (PSD) and bispectrum were obtained. Then, the fast correlation-based filter (FCBF) was applied to select a subset of relevant features that may be complementary, excluding those that are redundant. The selected features fed a multiclass multi-layer perceptron (MLP) neural network to build a model to estimate the SAHS severity degrees. Results: An optimum subset with features from all the proposed methodological approaches was obtained: variables from bispectrum, as well as PSD, ODI3, Age, and Sex. In the 3-class classification task, the MLP model trained with these features achieved an accuracy of 76.0% and a Cohen's kappa of 0.56 in an independent test set. Additionally, high accuracies were reached using the AHI cutoffs for diagnosis of moderate (AHI = 5 e/h) and severe (AHI = 10 e/h) SAHS: 81.3% and 85.3%, respectively. These results outperformed the diagnostic ability of a MLP model built without using bispectral features. Conclusions: Our results suggest that bispectrum provides additional information to anthropometric variables, ODI3 and PSD regarding characterization of changes in the SpO<sub>2</sub> signal caused by respiratory events. Thus, oximetry bispectrum can be a useful tool to provide complementary information for screening of moderate-to-severe pediatric SAHS.

**Wavelet analysis of oximetry recordings to assist in the automated detection of moderate-to-severe pediatric sleep apnea-hypopnea syndrome (Vaquerizo-Villar et al., 2018c).**

**Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-Garía, Andrea Crespo, Félix del Campo, David Gozal, and Roberto Hornero. *PLOS One*, vol. 13 (12), p.e0208502, 2018. Impact factor in 2018: 2.776, Q2 in “MULTIDISCIPLINARY SCIENCES” (JCR-WOS).

*Abstract:* Background: The gold standard for pediatric sleep apnea hypopnea syndrome (SAHS) is overnight polysomnography, which has several limitations. Thus, simplified diagnosis techniques become necessary. Objective: The aim of this study is twofold: (i) to analyze the blood oxygen saturation (SpO<sub>2</sub>) signal from nocturnal oximetry by means of features from the wavelet transform in order to characterize pediatric SAHS; (ii) to evaluate the usefulness of the extracted features to assist in the detection of pediatric SAHS. Methods: 981 SpO<sub>2</sub> signals from children ranging 2–13 years of age were used. Discrete wavelet transform (DWT) was employed due to its suitability to deal with non-stationary signals as well as the ability to analyze the SAHS-related low frequency components of the SpO<sub>2</sub> signal with high resolution. In addition, 3% oxygen desaturation index (ODI3), statistical moments and power spectral density (PSD) features were computed. Fast correlation-based filter was applied to select a feature subset. This subset fed three classifiers (logistic regression, support vector machines (SVM), and multilayer perceptron) trained to determine the presence of moderate-to-severe pediatric SAHS (apnea-hypopnea index cutoff  $\geq 5$  events per hour). Results: The wavelet entropy and features computed in the  $D_9$  detail level of the DWT reached significant differences associated with the presence of SAHS. All the proposed classifiers fed with a selected feature subset composed of ODI3, statistical moments, PSD, and DWT features outperformed every single feature. SVM reached the highest performance. It achieved 84.0% accuracy (71.9% sensitivity, 91.1% specificity), outperforming state-of-the-art studies in the detection of moderate-to-severe SAHS using the SpO<sub>2</sub> signal alone. Conclusion: Wavelet analysis could be a reliable tool to analyze the oximetry signal in order to assist in the automated detection of moderate-to-severe pediatric SAHS. Hence, pediatric subjects suffering from moderate-to-severe SAHS could benefit from an accurate simplified screening test only using the SpO<sub>2</sub> signal.

**Detrended fluctuation analysis of the oximetry signal to assist in paediatric sleep apnoea-hypopnoea syndrome diagnosis (Vaquerizo-Villar et al., 2018a).**

**Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-Garía, Andrea Crespo, Félix del Campo, David Gozal, and Roberto Hornero. *Physiological Measurement*, vol. 39 (11), p. 114006, 2018. Impact factor in 2018: 2.246, Q3 in “ENGINEERING, BIOMEDICAL” (JCR-WOS).

*Abstract:* Objective: To evaluate whether detrended fluctuation analysis (DFA) provides information that improves the diagnostic ability of the oximetry signal in the diagnosis of paediatric sleep apnoea-hypopnoea syndrome (SAHS). Approach: A database composed of 981 blood oxygen saturation (SpO<sub>2</sub>) recordings in children was used to extract DFA-derived features in order to quantify the scaling behaviour and the fluctuations of the SpO<sub>2</sub> signal. The 3% oxygen desaturation index (ODI3) was also computed for each subject. Fast correlation-based filter (FCBF) was then applied to select an optimum subset of relevant and non-redundant features. This subset fed a multi-layer perceptron (MLP) neural network to estimate the apnoea-hypopnoea index (AHI). Main results: ODI3 and four features from the DFA reached significant differences associated with the severity of SAHS. An optimum subset composed of the slope in the first scaling region of the DFA profile and the ODI3 was selected using FCBF applied to the training set (60% of samples). The MLP model trained with this feature subset showed good agreement with the actual AHI, reaching an intra-class correlation coefficient of 0.891 in the test set (40% of samples). Furthermore, the estimated AHI showed high diagnostic ability, reaching an accuracy of 82.7%, 81.9%, and 91.1% using three common AHI cut-offs of 1, 5, and 10 events per hour (e/h), respectively. These results outperformed the overall performance of ODI3. Significance: DFA may serve as a reliable tool to improve the diagnostic performance of oximetry recordings in the evaluation of paediatric patients with symptoms suggestive of SAHS.

**A convolutional neural network architecture to enhance oximetry ability to diagnose pediatric obstructive sleep apnea (Vaquerizo-Villar et al., 2021).**

Fernando Vaquerizo-Villar, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-Garía, Eduardo Santamaría-Vázquez, Félix del Campo, David Gozal, and Roberto Hornero. *IEEE Journal of Biomedical and Health Informatics*, vol. 25 (8), p. 2906-2916, 2021. Impact factor in 2020 (last year available): 5.772, D1 in "MATHEMATICAL & COMPUTATIONAL BIOLOGY" (JCR-WOS).

*Abstract:* This study aims at assessing the usefulness of deep learning to enhance the diagnostic ability of oximetry in the context of automated detection of pediatric obstructive sleep apnea (OSA). A total of 3196 blood oxygen saturation (SpO<sub>2</sub>) signals from children were used for this purpose. A convolutional neural network (CNN) architecture was trained using 20-min SpO<sub>2</sub> segments from the training set (859 subjects) to estimate the number of apneic events. CNN hyperparameters were tuned using Bayesian optimization in the validation set (1402 subjects). This model was applied to three test sets composed of 312, 392, and 231 subjects from three independent databases, in which the apnea-hypopnea index (AHI) estimated for each subject (AHI<sub>CNN</sub>) was obtained by aggregating the output of the CNN for each 20-min SpO<sub>2</sub> segment. AHI<sub>CNN</sub> outperformed the 3% oxygen desaturation index (ODI3), a clinical approach, as well as the AHI estimated by a conventional feature-engineering approach based on multi-layer perceptron (AHI<sub>MLP</sub>). Specifically, AHI<sub>CNN</sub> reached higher four-class Cohen's kappa in the three test databases than ODI3 (0.515 vs 0.417, 0.422 vs 0.372, and 0.423 vs 0.369) and AHI<sub>MLP</sub> (0.515 vs 0.377, 0.422 vs 0.381, and 0.423 vs 0.306). In addition, our proposal outperformed state-of-the-art studies, particularly for the AHI severity cutoffs of 5 e/h and 10 e/h. This suggests that the information automatically learned from the SpO<sub>2</sub> signal by deep-learning techniques helps to enhance the diagnostic ability of oximetry in the context of pediatric OSA.

## 1.2 The biomedical signal processing framework: feature-engineering and deep learning

Biomedical signals convey information on the functioning of the human body (Bronzino, 2000). The study of these signals allows to analyze the properties of the underlying biological systems (Bronzino, 2000), which makes possible to identify several pathological conditions (Sörnmo and Laguna, 2005). Nonetheless, the physiological information contained in these signals cannot be typically extracted in a visual way (Sörnmo and Laguna, 2005). In this respect, biomedical signal processing provides methods that help to understand and characterize the hidden information from these signals that can not be obtained through visual assessment (Sörnmo and Laguna, 2005). It also allows to develop automated systems for the diagnosis, treatment, and/or monitoring of a wide range of pathologies (Sörnmo and Laguna, 2005).

In this type of systems (e.g., the screening of pediatric OSA), the automated analysis of biomedical signals has been traditionally performed following a feature-engineering methodology (Najarian and Splinter, 2012), which consists of three main stages. In the first stage, known as feature extraction, the hidden characteristic information (features) about the biomedical signals is obtained (Krishnan and Athavale, 2018; Najarian and Splinter, 2012). To extract these features, different algorithms based on mathematical methods are used such as statistical, morphological, frequency domain, time-frequency, or nonlinear analysis (Krishnan and Athavale, 2018; Najarian and Splinter, 2012; Rangayyan, 2015). The second stage is the use of automatic feature selection methods to find the relevant and non-redundant information among that extracted in the previous step (Rangayyan, 2015). Finally, in the third stage, the selected information is used to train pattern recognition algorithms in order to obtain predicted models aimed at providing a diagnostic decision (Najarian and Splinter, 2012; Rangayyan, 2015). The range of pattern-recognition methods include from weak classifiers like logistic regression (LR) and Fisher linear discriminant analysis (LDA) to more complex algorithms such as support vector machines (SVM), Bayesian classifiers, and multi-layer perceptron (MLP) neural networks (Najarian and Splinter, 2012; Rangayyan, 2015).

Deep learning has emerged in recent year as a novel methodological approach aimed at changing the paradigm of data processing (LeCun et al., 2015). Conventional feature-engineering approaches have two main disadvantages: (i) a hu-

man expert must determine which relevant features obtain from the input data, and (ii) these methods provide a low level of abstraction that limits their capability to learn complex features from the data. These issues may result in missing important information from the data. In contrast to conventional approaches, a deep-learning model automatically discovers the intricate information in the data (LeCun et al., 2015). In this regard, deep-learning methods automatically learn complex patterns and extract features from raw data by the use of various processing layers with multiple levels of representation (LeCun et al., 2015). Starting from the representation at the lowest level, the raw input data, simple non-linear modules transform them into representations at a higher abstract level (LeCun et al., 2015). With an architecture composed of enough transformations, deep-learning algorithms can learn very complex features from the data. As aforementioned, these algorithms have improved predictive performances in a broad range of traditionally challenging domains, such as image, genomics, and signal processing (LeCun et al., 2015). Specifically, in the biomedical signal processing field, these algorithms have beaten conventional methods in many relevant domains, including sleep stage scoring (Faust et al., 2019), congestive heart failure diagnosis (Jahmunah et al., 2019), epileptic seizure detection (Roy et al., 2019), and brain-machine interfaces (Roy et al., 2019).

This Doctoral Thesis is aimed at enhancing the diagnosis ability of the oximetry signal in the context of childhood OSA. For this purpose, novel feature-engineering and deep-learning methodologies have been developed and assessed.

### 1.3 Pediatric Obstructive Sleep Apnea (OSA)

Although originally described for adults, OSA has been recognized in recent years as a high prevalent condition among children (1.2% - 5.7%) (Marcus et al., 2012), with etiological, diagnostical, and therapeutical considerations that are different for the pediatric population (Capdevila et al., 2008; Marcus et al., 2012). According to the American Academy of Pediatrics (AAP), childhood OSA is a respiratory disorder marked by repetitive episodes of complete absence (apnea) and/or considerable reduction (hypopnea) of airflow during sleep (Marcus et al., 2012). It is associated with the presence of nocturnal symptoms that cause disturbed sleep. Thus, apneic (apneas and hypopneas) events derive in inadequate gas exchange, leading to hypercapnia and hypoxia states, which induce oxygen desaturations, arousals, and sleep fragmentation. Gasping and snoring also occur

frequently (Loughlin et al., 1996).

As a consequence of these symptoms, OSA have many negative effects that reduce health and quality of life of the children (Capdevila et al., 2008; Marcus et al., 2012). In this respect, children suffering from OSA are at an increased risk for developing cardiovascular morbidities, such as systemic hypertension, changes in blood pressure regulation, and altered left ventricular geometry (Capdevila et al., 2008). OSA during childhood may also lead to neurobehavioral abnormalities, such as cognitive deficits, reduced academic achievements, hyperactivity, aggressive behavior, and excessive daytime sleepiness (Hunter et al., 2016). Moreover, it is also related to the metabolic syndrome, which includes hypertension, insulin resistance, dyslipidemia, and obesity (Capdevila et al., 2008). Finally, somatic growth impairment has been related to pediatric OSA as well (Alonso-Álvarez et al., 2011).

Despite its high prevalence, pediatric OSA is an underdiagnosed condition (Kheirandish-Gozal, 2010). Estimations indicate that approximately 90% of the affected children have not been diagnosed yet (Kheirandish-Gozal, 2010). Treatment interventions for pediatric OSA have led to a reduction in neurocognitive, cardiometabolic, and growth stunting risks (Tan et al., 2017). Nonetheless, the low percentage of diagnosis, together with the high prevalence, result in a high number of children being exposed to its adverse consequences.

## 1.4 Pediatric OSA diagnosis: Polysomnography (PSG)

OSA is diagnosed by means of the overnight PSG test, which acts as "gold standard" (Marcus et al., 2012). During PSG, multiple neurophysiological and cardiorespiratory signals from patients are monitored and recorded: electroencephalogram (EEG), electromyogram (EMG), electrooculogram (EOG), electrocardiogram (ECG), oronasal airflow (AF), abdominal and chest wall movements respiratory effort, SpO<sub>2</sub>, and photoplethysmography (PPG), among others (Tan et al., 2014). Thus, patients need to stay a complete night in a sleep laboratory, where skilled staff care for them as well as monitor the course of the test. After the PSG, the sleep recordings need an offline inspection to annotate apneas and hypopneas in order to compute the apnea-hypopnea index (AHI), which is the clinical variable employed to establish a diagnosis (Tan et al., 2014).

Despite the well-known effectiveness of PSG, it presents several limitations.

PSG is a complex test due to the necessity to record a high number of signals, which requires that patients spend at least one night in a sleep laboratory (Tan et al., 2015). In addition, trained personnel is needed to be responsible for the children and a proper development of the test, resulting in high hospital expenses (Tan et al., 2015). Similarly, apneic events are manually scored by trained specialists, which is labor intensive and may result in subjective diagnoses (Tan et al., 2015). Finally, the nature of the PSG requires to perform the test out of the sleep environment of the patients and with the use of multiple sensors placed on their bodies, which results highly uncomfortable and intrusive for children (Katz et al., 2012). This may derive in obtaining sleep recordings which are not representative of natural sleep, thus resulting in the need to repeat the diagnostic test (Katz et al., 2012).

Due to the complexity, cost, and time needed to analyze the sleep signals, available resources are not enough to cope with the high demand of OSA diagnosis (Tan et al., 2015). This results in long waiting lists, thus hindering the diagnosis and treatment of the affected children (Nixon et al., 2004). These drawbacks, together with the high prevalence rate of pediatric OSA, have led the scientific community to explore the use of simplified screening tests (Kaditis et al., 2016b; Marcus et al., 2012). In this sense, the guidelines of the AAP recommend conducting alternative tests to address PSG unavailability, while still requiring more conclusive evidences about the efficacy of these tests (Marcus et al., 2012).

## 1.5 Alternatives to PSG

In order to address the above-mentioned limitations, the use of portable monitoring equipment has been suggested as the main alternative to PSG in the diagnosis of pediatric OSA (Kaditis et al., 2016b; Marcus et al., 2012). According to the Portable Monitoring Task Force of the American Academy of Sleep Medicine (AASM), the equipment used in sleep studies can be classified into four types, depending on the number and type of the recorded signals (Standards of Practice Committee of the American Sleep Disorders Association, 1994):

- I. **Type I: Standard PSG.** This first type consists of the conventional PSG equipment, which requires the supervision of trained personnel in the hospital facilities. These devices are considered the gold standard to which the remaining types must be compared.
- II. **Type II: Comprehensive portable PSG.** These devices record a minimum

of seven channels, including EOG, chin EMG, EEG, airflow, ECG or heart rate (HR), respiratory effort, and SpO<sub>2</sub>. These studies do not require the presence of trained personnel. These devices allow to identify sleep stages and calculate the AHI.

- III. **Type III: Modified portable sleep apnea testing.** These studies, also called respiratory polygraphy (RP) studies, include the recording of ventilation (a minimum of two respiratory movement signals or one respiratory movement signal and airflow), ECG or HR, and SpO<sub>2</sub>.
- IV. **Type IV: Continuous single-bioparameter or dual-bioparameter recording.** These devices only record one or two physiological signals, being oximetry traditionally one of these measurements. In addition, all the equipment that does not meet Type III criteria is included in this group.

### 1.5.1 Overnight oximetry

In recent years, there has been an increased interest in overnight oximetry as the main alternative to PSG in the context of pediatric OSA diagnosis due to its simplicity, reliability, and suitability for children ([del Campo et al., 2018](#)). Overnight oximetry records the SpO<sub>2</sub> signal in a non-invasive way with a pulse oximeter, usually located on the finger, toe or earlobe of the patient ([Netzer et al., 2001](#)).

SpO<sub>2</sub> expresses the amount of oxygen combined with the hemoglobin with respect to the total hemoglobin in the blood, the oxyhemoglobin (O<sub>2</sub>Hb), which is responsible for transporting the blood oxygen to the tissues. The operating principle of the SpO<sub>2</sub> sensors is based on the optical properties of the hemoglobin ([Chan et al., 2013](#)). O<sub>2</sub>Hb absorbs more infrared light, acquiring a red hue. On the contrary, deoxyhemoglobin (HHb) absorbs a higher amount of red light, thus having a more bluish hue. Pulse oximeters exploit this difference in the light absorption of O<sub>2</sub>Hb and HHb to obtain SpO<sub>2</sub>. To achieve this, pulse oximeters contain two light-emitting diodes on one side of the finger that transmit at red (around 660 nm) and near infrared wavelengths (around 940 nm). On the opposite side of the finger, a photodiode is placed to measure the amount of red and infrared light absorbed by the tissues, which allows to determine SpO<sub>2</sub> ([Chan et al., 2013](#)).

Due to these easy acquisition of the SpO<sub>2</sub> signal, commercial pulse oximetry devices have been developed, which facilitate to perform the test in an unsupervised way at children's home ([Garde et al., 2014a](#); [Nixon et al., 2004](#)). Apneic

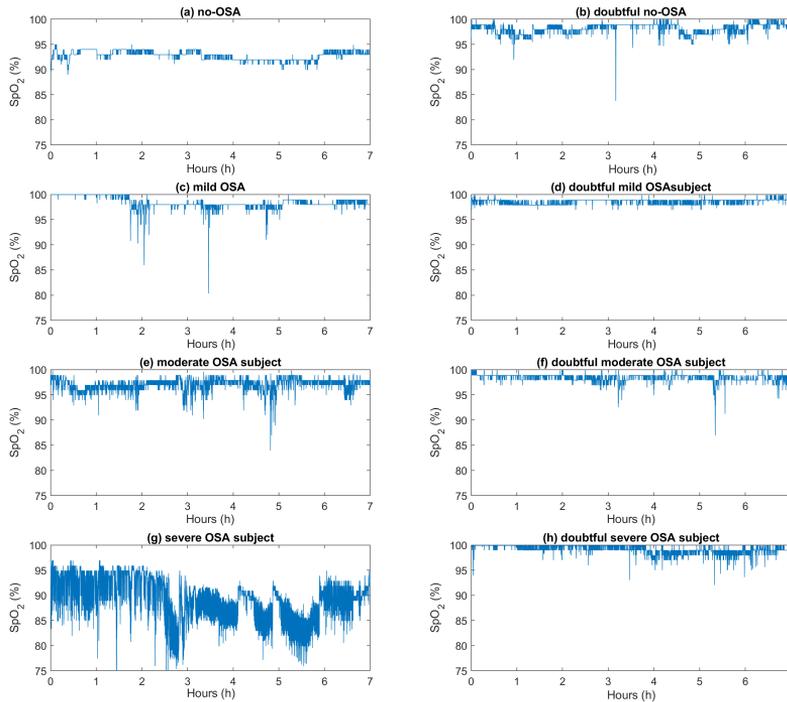
events from OSA induce recurrent decreases in the SpO<sub>2</sub> (Berry et al., 2012), also called oxygen desaturations, so that this signal contains important OSA-related information. Figure 1.2 shows the SpO<sub>2</sub> signal corresponding to (a) a no OSA pediatric subject (AHI < 1 events per hour), (b) doubtful no OSA pediatric subject (AHI < 1 events per hour), (c) a mild OSA pediatric subject (1 ≤ AHI < 5 events per hour), (d) a doubtful mild OSA pediatric subject (1 ≤ AHI < 5 events per hour), (e) a moderate OSA pediatric subject (5 ≤ AHI < 10 events per hour), (f) a doubtful moderate OSA pediatric subject (5 ≤ AHI < 10 events per hour), (g) a severe OSA pediatric subject (AHI ≥ 10 events per hour), and (h) a doubtful severe OSA pediatric subject (AHI ≥ 10 events per hour). It can be observed that there are more oxygen desaturations as the severity of OSA increases. However, it is difficult to visually discriminate the SpO<sub>2</sub> signal from doubtful subjects.

## 1.6 State-of-the-art: Automated analysis of the oximetry signal to diagnose pediatric OSA

The usefulness of SpO<sub>2</sub> recordings from nocturnal oximetry to help in the screening of pediatric OSA has been widely assessed, especially in the last years (del Campo et al., 2018). The analysis of this signal has been addressed by using conventional oximetric indices, as well as automated signal processing methodologies.

### 1.6.1 Conventional oximetric indices

Previous studies predominantly used conventional oximetric indices for this task (Brouillette et al., 2000; Chang et al., 2013; Kirk et al., 2003; Ma et al., 2018; Nixon et al., 2004; Tsai et al., 2013; Van Eyck et al., 2015; Velasco et al., 2013; Villa et al., 2015). In this respect, researchers mainly assessed the screening ability of oxygen desaturation index (ODI), which accounts for the number of drops of the SpO<sub>2</sub> signal larger than a defined threshold (Kirk et al., 2003; Ma et al., 2018; Tsai et al., 2013). Kirk et al. (2003) assessed, in a population of 58 pediatric subjects, the agreement between the AHI from PSG and the ODI3 from the SpO<sub>2</sub> signal simultaneously recorded with a portable monitoring device (SnoreSat, SagaTech Electronics, Calgary, AB, Canada). Similarly, Tsai et al. (2013) evaluated the yield of several cut-offs for the 4% ODI (ODI4) to establish pediatric OSA severity in a database of 146 pediatric PSGs. Recently, Ma et al. (2018) studied the clinical applicability of a pulse oximetry watch (POW) for pediatric OSA diagnosis. To



**Figure 1.2.** Examples of  $SpO_2$  corresponding to: (a) a no OSA subject, (b) doubtful no OSA subject, (c) a mild OSA subject, (d) a doubtful mild OSA subject, (e) a moderate OSA subject, (f) a doubtful moderate OSA subject, (g) a severe OSA subject, (h) a doubtful severe OSA subject. In doubtful subjects, it is difficult to visually discriminate the OSA severity group from the overnight  $SpO_2$  profile, leading to the need of non-subjective automated analyses.

this effect, they measured the concordance between the AHI from PSG and the ODI4 derived from the  $SpO_2$  signal simultaneously recorded with the POW in 32 children.

Likewise, the number and depth of clusters of desaturations in the  $SpO_2$  recordings have been quantified by means of a visual inspection in order to develop OSA screening protocols (Brouillette et al., 2000; Nixon et al., 2004; Van Eyck et al., 2015; Velasco et al., 2013). A cluster of desaturations was defined by Brouillette et al. (2000) as 5 or more oxygen desaturations of at least 4% occurring in a 10-30 minute window. Based on the number of clusters of desaturations and the number of drops of the  $SpO_2$  signal below 90%, Brouillette et al. (2000) defined a positive, negative, or inconclusive score for pediatric OSA and compared

it with the standard diagnosis from PSG in a dataset of 349 children. Similarly, [Nixon et al. \(2004\)](#) developed a severity score for pediatric OSA, named McGill oximetry score (MOS), which is also based on the number of clusters of desaturations and the number of drops of the oximetry signal below a defined threshold. Furthermore, [Velasco et al. \(2013\)](#) assessed the diagnostic ability of a positive OSA score defined as 2 or more clusters of desaturations, one of them with an oxygen drop below 90%, in a sample of 167 children with adenotonsillar hypertrophy. Finally, [Van Eyck et al. \(2015\)](#) assessed the diagnostic ability of the methodologies proposed by [Brouillette et al. \(2000\)](#) and [Velasco et al. \(2013\)](#), as well as the ODI3, in a population of 130 obese patients.

These oximetric indices have also been combined with common symptoms ([Chang et al., 2013](#)) and clinical history ([Villa et al., 2015](#)) in order to enhance their diagnostic ability. [Chang et al. \(2013\)](#) used common symptoms (witnessed apneas, mouth breathing, and restless sleep) and the ODI3 to evaluate both a LR classifier and a discriminative score to diagnose pediatric OSA in a sample of 141 children. Conversely, [Villa et al. \(2015\)](#) combined the MOS and a sleep clinical record that includes physical examination and children's history to detect pediatric OSA in a database of 236 pediatric subjects.

Differing from these studies, the research conducted in the present Doctoral Thesis has included the use of automated signal processing algorithms to further characterize the oximetry recordings.

## 1.6.2 Automated signal processing methods

Recent works have focused on enhancing the diagnostic capability of the oximetry signal by means of the application of automated signal processing algorithms ([Álvarez et al., 2017](#); [Álvarez et al., 2018](#); [Crespo et al., 2017, 2018](#); [Garde et al., 2014a](#); [Hornero et al., 2017](#); [Xu et al., 2019](#)). As mentioned in Section 1.1, these studies typically follow a three-stage feature-engineering methodology: (i), feature extraction; (ii) feature selection; and (iii) pattern recognition.

Importantly, in the feature extraction stage, the majority of studies ([Álvarez et al., 2017](#); [Crespo et al., 2018](#); [Garde et al., 2014a](#); [Hornero et al., 2017](#); [Xu et al., 2019](#)) employed signal processing techniques that had already shown its usefulness to characterize the changes in oximetry dynamics associated to apneic events in adult patients both in the time and frequency domains ([del Campo et al., 2018](#)). In the time domain, SpO<sub>2</sub> recordings were characterized using statistical moments, oximetric indices, and several non-linear measures: approximate en-

tropy (Pincus, 1991), sample entropy (Richman and Moorman, 2000), Lempel-Ziv complexity (Lempel and Ziv, 1976), and central tendency measure (Cohen and Hudson, 2000). Conversely, the power spectral density (PSD) (Welch, 1967) was used to characterize the oximetry signal in the frequency domain (Álvarez et al., 2017; Crespo et al., 2018; Garde et al., 2014a; Hornero et al., 2017; Xu et al., 2019).

First, Garde et al. (2014a) developed and validated the Phone Oximeter, a portable monitoring device consisting of a pulse oximetry sensor connected to a mobile phone, as a diagnostic tool for childhood OSA. To this effect, they assessed a LDA model fed with statistical parameters, classical indices, nonlinear features, and PSD-derived features from 146 SpO<sub>2</sub> recordings of pediatric patients. Similarly, Álvarez et al. (2017) evaluated at-home unsupervised oximetry in pediatric OSA using a LR model trained with statistical moments, PSD-derived parameters, nonlinear features, and classical indices from 50 patients that underwent RP. A thorough comparative analysis of statistical binary classifiers for the diagnosis of childhood OSA was performed by Crespo et al. (2018). Specifically, they assessed LDA, LR, and quadratic discriminant analysis (QDA) pattern recognition models trained with statistical moments, PSD variables, nonlinear features, and the ODI3 extracted from a database of 176 children (Crespo et al., 2018). The usefulness of automated processing of oximetric recordings as a screening tool for pediatric OSA was also examined in a multicenter international study developed by Hornero et al. (2017), which involved 4191 pediatric subjects from 13 sleep centers. Particularly, Hornero et al. (2017) assessed a MLP neural network trained to estimate the AHI with the ODI3 and the third-order moment of the PSD. This MLP model was further validated in 432 children along with a remote cloud system (Xu et al., 2019). Nonetheless, Hornero et al. (2017) reported a high redundancy in the information extracted from the oximetry recordings. As mentioned in section 1.1, these studies used the same methods employed in adult OSA patients, but reaching a lower performance (Álvarez et al., 2017; Crespo et al., 2018; Garde et al., 2014a; Hornero et al., 2017; Xu et al., 2019). This highlights the need for novel signal processing methods able to provide specific features for the particularities of pediatric OSA.

In order to address this issue, two recent studies conducted by Crespo et al. (2017) and Álvarez et al. (2018) applied two novel nonlinear analysis methods in the context of pediatric OSA, multiscale entropy (Costa et al., 2005) and symbolic dynamics (Daw et al., 2003), respectively. Crespo et al. (2017) investigated the ability of multiscale entropy (MSE) to further characterize the dynamics of unattended oximetry using the database employed in Álvarez et al. (2017). For this

purpose, a LR model was trained to automatically detect childhood OSA with conventional oximetric indices and nonlinear MSE-derived parameters (Crespo et al., 2017). Additionally, Álvarez et al. (2018) evaluated the usefulness of symbolic dynamics to increase the diagnostic capability of portable oximetry recordings from the Phone Oximeter. Specifically, a LR model was designed using conventional oximetric indices, anthropometric variable, statistical parameters, and nonlinear features from symbolic dynamics (Álvarez et al., 2018). In this Doctoral Thesis, we have assessed the usefulness of DFA (Peng et al., 1994), a time domain nonlinear analysis method, and wavelet (Rioul and Vetterli, 1991) and bispectral analysis (Chua et al., 2010), two frequency domain techniques, to provide complementary features that help to further characterize apneic events from pediatric OSA (Vaquerizo-Villar et al., 2018a,b,c).

As aforementioned, there has been a breakthrough in the last years in the data science field thanks to the emergence of deep learning approaches (LeCun et al., 2015). Due to its multilayer architecture with multiple levels of representation, deep-learning methods are suitable to learn very complex patterns from the raw data, which has led them to outperform conventional approaches in many fields (LeCun et al., 2015). In this respect, previous studies have applied deep-learning methods to analyze polysomnographical signals in adult OSA patients (Faust et al., 2019; Mostafa et al., 2019). These works have focused on the detection of sleep stages (Faust et al., 2019), apneic events (Mostafa et al., 2019), and/or direct AHI estimation (Mostafa et al., 2019). To our knowledge, Vaquerizo-Villar et al. (2021), the last article of the present Doctoral Thesis, is the first study that applies deep learning techniques in the context of childhood OSA. In Vaquerizo-Villar et al. (2021), a CNN-based deep-learning architecture was fed with raw SpO<sub>2</sub> data (Vaquerizo-Villar et al., 2021), which has outperformed conventional approaches in the framework of pediatric OSA.

In this chapter, we have introduced the topic of this Doctoral Thesis. In the next Chapter (see Section 2), the hypotheses and objectives of this research will be stated.

## Chapter 2

# Hypotheses and objectives

The automated analysis of the oximetry signal has become as the main alternative to PSG in the screening of pediatric OSA. Accordingly, the present Doctoral Thesis focuses on the use of novel signal processing algorithms, intended to increase the diagnosis ability of the SpO<sub>2</sub> signal from overnight oximetry. Therefore, Section 2.1 describes the different hypothesis that have been formulated throughout the present Doctoral Thesis, as well as the global hypothesis that raised from these statements. Similarly, the main objective of this Thesis is stated in Section 2.2, as well as the specific objectives that have been accomplished to achieve it.

### 2.1 Hypotheses

Simplification of OSA diagnosis has become a main research topic in past years (Bertoni and Isaiah, 2019; Kaditis et al., 2016b; Marcus et al., 2012). As previously explained, the SpO<sub>2</sub> signal allows to detect oxygen desaturations associated to apneic events (Berry et al., 2012), which has led to its use in the screening of pediatric OSA (del Campo et al., 2018). Thus, at the beginning of this Doctoral Thesis, the following hypothesis was formulated: *the oximetry signal on its own may contain enough information for the screening of pediatric OSA*. Nonetheless, this statement does not completely describe the starting point of the different investigations conducted in this study.

As stated in Subsection 1.6, a recent study reported a high redundancy in the conventional features extracted from the oximetry signal (Hornero et al., 2017), suggesting the need to apply novel signal processing algorithms. Thereby, it has been hypothesized that *novel feature extraction methods could further charac-*

terize OSA-related changes in the oximetry signal. Similarly, we wonder whether the features extracted by these methods provide complimentary information to improve the diagnostic capability of the oximetry signal in the context of pediatric OSA. Feature selection and pattern recognition algorithms have been used for this purpose.

In spite of the usefulness of conventional feature-engineering approaches, they are limited to learn all the relevant information from the data because of: (i) the domain expert that designs the feature extractor decides which features are relevant *a priori*; and (ii) their reduced level of abstraction. Owing to their great capability to automatically learn very complex features from raw data, it has been hypothesized that *deep-learning algorithms could extract all the OSA-related information from the SpO<sub>2</sub> signal, thus enhancing its diagnostic ability.*

These statements are the main hypotheses that form the core of the current Doctoral Thesis, which can be joint into the following global hypothesis:

*“The application of novel feature extraction and deep-learning algorithms allows to capture hidden patterns of desaturations linked to apneic events, enhancing the diagnostic ability of the single-channel oximetry in the context of pediatric OSA.”*

## 2.2 Objectives

The main goal of the present Doctoral Thesis is to design, develop, and evaluate novel clinical decision-support models in the context of pediatric OSA based on the automated analysis of the oximetry signal. To reach this general goal, the following specific objectives arise:

- I. To further characterize changes in the oximetry signal caused by apneic events linked to pediatric OSA both in the time and frequency domains.
- II. To identify novel features from the oximetry signal able to provide relevant and complimentary information to conventional oximetry variables.
- III. To design and optimize high-performance pattern recognition models aimed at the automated detection of pediatric OSA and its severity using optimum subsets of features from the SpO<sub>2</sub> signal.
- IV. To explore and develop novel deep-learning based architectures capable to automatically learn all the OSA-related information from raw oximetry data.

## Chapter 3

# Subjects and signals under study

During this research, three different databases of pediatric subjects were analyzed: (i) the Childhood Adenotonsillectomy Trial (CHAT) database, (ii) the University of Chicago (UofC) database, and (iii) the Burgos University Hospital (BUH) database. All of them contained SpO<sub>2</sub> recordings of children ranging from 0 to 18 years of age. These pediatric subjects were referred to nocturnal PSG showing clinical suspicion from OSA due to one or several of the following criteria: snoring, apneas, arousals, excessive daytime sleepiness, restless sleep, hyperactivity, tonsillar hypertrophy, increase in neck circumference, developmental disorder depression and low self-esteem, enuresis, obesity, attention deficit, behavioral problems, and cephalas. The CHAT dataset, a public multicenter dataset, was integrated by 1638 sleep studies. The UofC dataset consisted of 980 children and the BUH dataset was composed of 578 pediatric subjects.

SpO<sub>2</sub> recordings were acquired from pediatric subjects during their standard diagnostic PSG test using sampling frequencies ranging from 1 to 512 Hz. Following the rules of the AASM (Berry et al., 2012; Iber et al., 2007), all the sleep recordings were manually inspected by trained staff to quantify sleep and annotate apneas and hypopneas. These annotations were used to obtain the AHI, which is used to establish a diagnosis (Marcus et al., 2012). Common AHI cutoffs used to establish pediatric OSA severity are 1, 5, and 10 e/h (Alonso-Álvarez et al., 2011; Church, 2012; Tan et al., 2014). In this respect, AHI = 5 e/h is commonly employed as a cutoff to recommend surgical treatment (Tan et al., 2014),

as children with an AHI  $\geq 5$  e/h are at a higher risk of developing comorbidities (Church, 2012; Hunter et al., 2016; Kaditis et al., 2016a). Accordingly, under a binary classification approach, pediatric subjects were distinguished into negative OSA (AHI  $< 5$  e/h) and positive OSA (AHI  $\geq 5$  e/h). In addition, children can also be classified into four pediatric OSA severity degrees: no OSA (AHI  $< 1$  e/h), mild OSA ( $1 \leq$  AHI  $< 5$  e/h), moderate OSA ( $5 \leq$  AHI  $< 10$  e/h), and severe OSA (AHI  $\geq 10$  e/h). Details of each database, such as number of subjects, sex (male percentage), age, body mass index (BMI), and number of patients of each OSA severity group are provided in Tables 3.1, 3.2, 3.3, and 3.4.

### 3.1 Childhood Adenotonsillectomy Trial (CHAT) database

CHAT database is composed of 1639 sleep studies from children ranging 5 to 10 years old with clinical suspicion of OSA. Sleep studies were obtained in 6 pediatric sleep centers of the United States of America (Children's Hospital of Pennsylvania, Philadelphia, PA; Cincinnati Children's Medical Center, Cincinnati, OH; Rainbow Babies and Children's Hospital, Cleveland, OH; Boston Children's Hospital, Boston, MA; Cardinal Glennon Children's Hospital, St. Louis, MI; Montefiore Medical Center, Bronx, NY) of which 1638 contained SpO<sub>2</sub> recording (Marcus et al., 2013; Redline et al., 2011). This database is divided into three groups:

- Baseline, composed of 453 SpO<sub>2</sub> recordings from children that met the criteria defined in Marcus et al. (2013) and Redline et al. (2011) to be randomized to early adenotonsillectomy or a strategy of watchful waiting.
- Follow-up, composed of SpO<sub>2</sub> recordings from 406 pediatric subjects of the baseline group performed after a 7-month observation period.
- Nonrandomized, composed of 779 SpO<sub>2</sub> recordings from children who did not meet the criteria defined in Marcus et al. (2013) and Redline et al. (2011) to be included in the follow-up study.

Overnight PSG was performed following a strict standardized procedure (Redline et al., 2011), which includes the acquisition of SpO<sub>2</sub> recordings with a Nonin 8000J or comparable sensor at sampling rates ranging from 1 to 512 Hz. The clinical trial identifier of the CHAT database is available in NCT00560859 and its full

protocol is provided in the supplementary material of [Marcus et al. \(2013\)](#). This database was used in [Vaquerizo-Villar et al. \(2021\)](#). Table 3.1 shows the sociodemographic and clinical data from this database.

## 3.2 University of Chicago (UofC) database

UofC database is composed of 981 SpO<sub>2</sub> recordings from children aged 0 to 13 years of age. All children were referred to the pediatric sleep unit at the University of Chicago Medicine Comer Children’s Hospital (Chicago, IL, USA) due to clinical suspicion of OSA. The legal guardians of all the children signed the informed consent and the Ethics Committee of the Comer Children’s Hospital approved the protocols (#11-0268-AM017, #09-115-B-AM031, and #IRB14-1241). Overnight PSGs were performed using a digital polysomnography system (Polysmith; Nihon Kohden America Inc., CA, USA), which includes an internal pulse oximetry sensor. In this way, SpO<sub>2</sub> recordings were obtained from PSG at sampling rates of 25 Hz, 200 Hz, or 500 Hz.

This dataset was used in all the studies of this Doctoral Thesis ([Vaquerizo-Villar et al., 2018a,b,c, 2021](#)). In the first manuscript of this doctoral thesis ([Vaquerizo-Villar et al., 2018b](#)), a subset of the whole UofC database composed of 298 SpO<sub>2</sub> recordings, those sampled at 25 Hz, was analyzed. In [Vaquerizo-Villar et al. \(2018c\)](#) and [Vaquerizo-Villar et al. \(2018a\)](#), the complete database, 981 subjects, was employed. Finally, one subject was removed from the complete dataset for the last manuscript of this Thesis, as the total sleep time was less than 3 hours after signal preprocessing, thus analyzing 980 SpO<sub>2</sub> recordings in [Vaquerizo-Villar et al. \(2021\)](#). Table 3.2 and Table 3.3 show the clinical and

**Table 3.1.** Clinical and sociodemographic data of the CHAT database.

|                                 | All                  | no OSA               | mild OSA             | moderate OSA         | severe OSA           |
|---------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| SpO <sub>2</sub> recordings (%) | 1638<br>(100%)       | 637<br>(38.9%)       | 609<br>(37.2%)       | 205<br>(12.5%)       | 187<br>(11.4%)       |
| Age (years)                     | 7<br>[6, 8]          |
| Males (%)                       | 602<br>(47.4%)       | 297<br>(46.6%)       | 287<br>(47.1%)       | 101<br>(49.3%)       | 92<br>(49.2%)        |
| BMI (kg/m <sup>2</sup> )        | 17.3<br>[15.8, 21.7] | 17.0<br>[15.5, 19.6] | 17.4<br>[15.6, 21.7] | 18.6<br>[15.4, 23.3] | 18.9<br>[16.0, 24.3] |
| AHI (e/h)                       | 1.6<br>[0.6, 4.7]    | 0.4<br>[0.2, 0.7]    | 2.2<br>[1.5, 3.2]    | 7.1<br>[5.9, 8.4]    | 17.9<br>[12.8, 26.9] |

Data are presented as median [interquartile range], *n*, or *n* (%). OSA = Obstructive sleep apnea.

sociodemographic data from the complete dataset (981 subjects) and the initial dataset of 298 SpO<sub>2</sub> recordings used in [Vaquerizo-Villar et al. \(2018b\)](#), respectively.

### 3.3 Burgos University Hospital (BUH) database

The BUH database included 578 children ranging 0 to 18 years of age who were referred to the pediatric sleep laboratory at Burgos University Hospital with high suspicion from OSA. All legal guardians of the pediatric subjects involved gave their informed consent and the Ethics Committee of the BUH approved the protocol. Children's sleep was monitored using the Deltamed Coherence 3NT Polysomnograph, version 3.0 system (Diagniscan S.A.U., Group Werfen, Paris, France), which includes a Nellcor Puritan Bennett, NPB-290 pulse oximeter. In

**Table 3.2.** Clinical and sociodemographic data of the complete UofC database (981 subjects).

|                                 | All                  | no OSA               | mild OSA             | moderate OSA         | severe OSA           |
|---------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| SpO <sub>2</sub> recordings (%) | 981<br>(100%)        | 175<br>(17.8%)       | 401<br>(40.9%)       | 176<br>(17.9%)       | 229<br>(23.4%)       |
| Age (years)                     | 6<br>[3, 9]          | 7<br>[4, 10]         | 6<br>[4, 9]          | 5<br>[2, 8]          | 4<br>[2, 8]          |
| Males (%)                       | 602<br>(61.4%)       | 109<br>(62.3%)       | 247<br>(61.6%)       | 107<br>(60.8%)       | 139<br>(60.7%)       |
| BMI (kg/m <sup>2</sup> )        | 17.9<br>[15.8, 21.9] | 17.7<br>[15.5, 20.9] | 17.7<br>[15.9, 21.2] | 18.6<br>[16.2, 24.0] | 18.3<br>[16.0, 23.2] |
| AHI (e/h)                       | 3.8<br>[1.5, 9.3]    | 0.5<br>[0.1, 0.8]    | 2.5<br>[1.7, 3.5]    | 6.8<br>[5.8, 8.3]    | 19.1<br>[13.9, 31.1] |

Data are presented as median [interquartile range], *n*, or *n* (%). OSA = Obstructive sleep apnea.

**Table 3.3.** Clinical and sociodemographic data of the initial version of the UofC database (298 subjects).

|                                 | All                  | negative OSA         | moderate OSA         | severe OSA           |
|---------------------------------|----------------------|----------------------|----------------------|----------------------|
| SpO <sub>2</sub> recordings (%) | 298<br>(100%)        | 164<br>(55.0%)       | 56<br>(18.8%)        | 78<br>(26.2%)        |
| Age (years)                     | 6<br>[4, 9]          | 7<br>[5, 10]         | 5<br>[3, 8]          | 6<br>[3, 9]          |
| Males (%)                       | 166<br>(55.7%)       | 91<br>(55.5%)        | 32<br>(57.1%)        | 43<br>(55.1%)        |
| BMI (kg/m <sup>2</sup> )        | 18.4<br>[16.3, 23.0] | 18.2<br>[16.3, 22.3] | 18.1<br>[16.3, 22.6] | 19.1<br>[16.5, 25.7] |
| AHI (e/h)                       | 4.2<br>[1.8, 10.4]   | 1.9<br>[1.0, 3.5]    | 7.0<br>[5.9, 8.5]    | 17.7<br>[11.7, 27.3] |

Data are presented as median [interquartile range], *n*, or *n* (%). OSA = Obstructive sleep apnea.

this way, SpO<sub>2</sub> recordings were obtained during nocturnal PSG at a sampling rate of 200 Hz. This database was used in [Vaquerizo-Villar et al. \(2021\)](#). Table 3.4 shows the sociodemographic and clinical data from this sample.

**Table 3.4.** Clinical and sociodemographic data of the BUH database.

|                                 | All                  | no OSA               | mild OSA             | moderate OSA         | severe OSA           |
|---------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| SpO <sub>2</sub> recordings (%) | 578<br>(100%)        | 205<br>(35.5%)       | 220<br>(38.1%)       | 65<br>(11.3%)        | 88<br>(15.2%)        |
| Age (years)                     | 5<br>[4, 7]          | 6<br>[4, 8]          | 5<br>[3, 6]          | 5<br>[3, 6]          | 4<br>[3, 5]          |
| Males (%)                       | 356<br>(61.6%)       | 127<br>(62.0%)       | 129<br>(58.7%)       | 38<br>(58.5%)        | 62<br>(70.5%)        |
| BMI (kg/m <sup>2</sup> )        | 16.0<br>[14.6, 18.2] | 16.1<br>[14.5, 18.8] | 16.0<br>[14.7, 17.7] | 15.4<br>[14.6, 18.1] | 16.1<br>[14.7, 17.3] |
| AHI (e/h)                       | 1.8<br>[0.6, 5.3]    | 0.4<br>[0.0, 0.6]    | 2.1<br>[1.5, 3.4]    | 6.9<br>[5.8, 8.1]    | 24.3<br>[14.8, 34.9] |

Data are presented as median [interquartile range], *n*, or *n* (%). OSA = Obstructive sleep apnea.



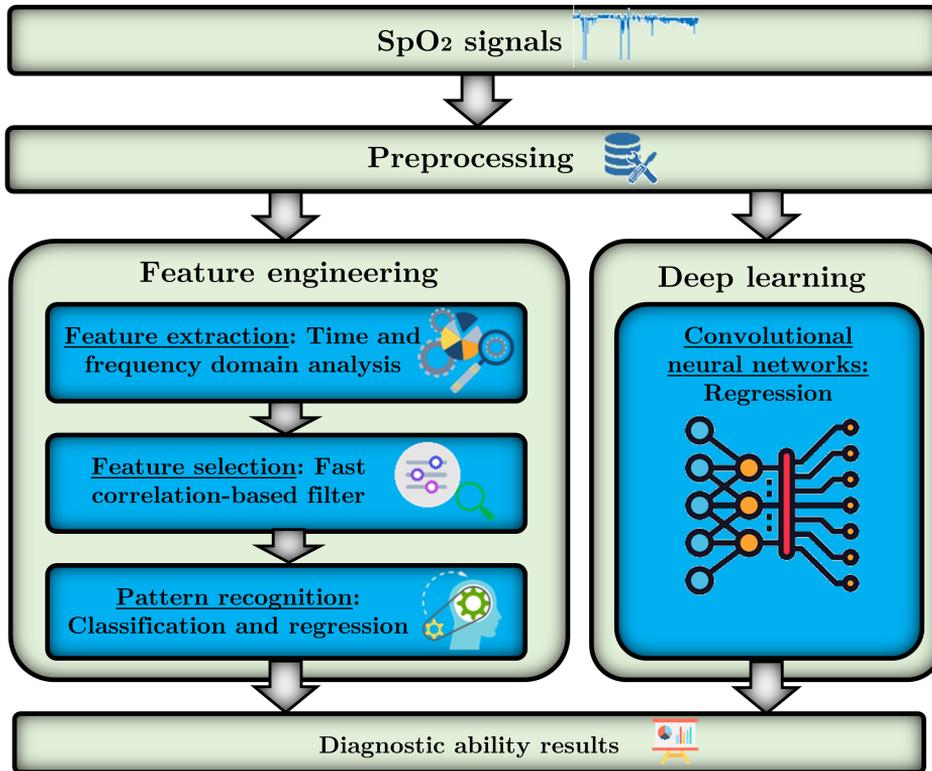
# Chapter 4

## Methods

This chapter describes the general signal processing methodology that has been conducted through the compendium of publications (see Figure 4.1). It starts with a signal pre-processing stage (Section 4.1), which adapts the oximetry data to the requirements of the different signal processing algorithms. Following the feature-engineering branch (Section 4.2), Sections 4.2.1 and 4.2.2 describe the signal processing methods applied to extract features from the SpO<sub>2</sub> recordings in the time and frequency domains and to select optimum subsets of optimum OSA-related features, respectively. Afterward, Section 4.2.3 is devoted to describe the pattern recognition algorithms employed to establish pediatric OSA and its severity. As the last stage of the feature-engineering methodology, pattern recognition algorithms are fed with the optimum features obtained in the feature selection stage. In the deep-learning branch (Section 4.3), CNNs are trained using raw oximetry signal to estimate the AHI and thereby the severity of pediatric OSA. Finally, Section 4.4 describes the statistical analysis techniques employed in this research.

### 4.1 Pre-processing

As seen in Sections 3.1 3.1 3.1, SpO<sub>2</sub> recordings were acquired during PSG using different pulse oximetry sensors and recording systems. Heterogeneity in the devices could lead to slight differences in the overnight oximetric profile able to influence the diagnostic ability of the signal. Additionally, the SpO<sub>2</sub> signal contains artifacts caused by loss of contact of the pulse oximeter probe due to subject movements. Therefore, a signal pre-processing step was included to standardize



**Figure 4.1.** Scheme of the general signal processing methodology conducted in the study.

the signals obtained from the different recording devices, as well as to remove motion artifacts. Signal pre-processing was different for the two main methodological approaches conducted in this research: (i) feature-engineering and (ii) deep-learning approaches.

Following a feature-engineering methodology, artifacts were first discarded from SpO<sub>2</sub> recordings by removing samples with SpO<sub>2</sub> values below 50% as well as sudden changes with a slope  $\geq 4\%$  per second (Magalang et al., 2003). As the sampling frequency among the different recording devices ranged from 1 to 512 Hz, oximetry signals were re-sampled to a common sample rate of: (i) 25 Hz in Vaquerizo-Villar et al. (2018b) and Vaquerizo-Villar et al. (2018c), as advocated by the AASM (Berry et al., 2012); and (ii) 1 Hz for the multiscale analysis of the oximetry signal proposed in Vaquerizo-Villar et al. (2018a), which has been considered more appropriate for multiscale analysis approaches in previous studies (Crespo et al., 2017; Hua and Yu, 2017). Finally, the resolution of the SpO<sub>2</sub> recordings was set to two decimal points (resolution of 0.01%) in order to homogenize

the resolution of signals from different recording devices (Garde et al., 2014b).

In the deep-learning branch, the pre-processing stage was simpler, as deep-learning approaches are able to automatically process raw data (Mostafa et al., 2019). In this respect, SpO<sub>2</sub> signal pre-processing consisted on: (i) down-sampling of the SpO<sub>2</sub> recordings to 1 Hz in order to homogenize the frequency (Mostafa et al., 2017); and (ii) segmentation of the SpO<sub>2</sub> recordings into 20-min segments (1200 samples) prior to train the CNN-based deep-learning architecture (Vaquerizo-Villar et al., 2021).

## 4.2 Feature engineering

Feature engineering is the conventional methodology employed to analyze biomedical signals. This methodology consists of the three following stages: (i) feature extraction; (ii) feature selection; and (iii) pattern recognition.

### 4.2.1 Feature extraction

As aforementioned, SpO<sub>2</sub> signals were characterized both in time and frequency domains. In this respect, different signal processing methods were applied to obtain OSA-related features from the oximetry signal (Vaquerizo-Villar et al., 2018a,b,c, 2021).

#### 4.2.1.1 Conventional measures in the time domain

In this research, conventional features were obtained from the oximetry signal using different time-domain analysis methods: oxygen desaturation index, statistical moments, and nonlinear parameters (Vaquerizo-Villar et al., 2018a,b,c, 2021). A description of these methods is found below.

**Oxygen desaturation index.** As mentioned in Section 1.5.1, oxygen desaturations are related to apneic episodes (Berry et al., 2012). In this research, the number of oxygen desaturations of at least 3% (ODI3) from prior SpO<sub>2</sub> baseline was computed (Taha et al., 1997).

**Statistical moments.** Common first-to-fourth order statistical moments ( $M1_t$  –  $M4_t$ ) were calculated to characterize the SpO<sub>2</sub> signal amplitude distribution. Accordingly, mean ( $M1_t$ ), variance ( $M2_t$ ), skewness ( $M3_t$ ), and kurtosis ( $M4_t$ ) allow

to measure the central tendency, dispersion, asymmetry, and peakedness of the data, respectively.

**Nonlinear parameters.** In recent years, nonlinear methods derived from the chaos theory have shown its usefulness to extract additional information of oximetry dynamics in both adults and pediatric OSA patients (Álvarez et al., 2017; Alvarez et al., 2013; Crespo et al., 2018; Garde et al., 2014a; Hornero et al., 2017). In this respect, the following common nonlinear features have been obtained from each SpO<sub>2</sub> recording:

- Central tendency measure (*CTM*). Using second-order differences plots (Cohen et al., 1996), *CTM* allows to quantify the variability of the oximetry signal associated to apneic events.
- Lempel-Ziv complexity (*LZC*). *LZC* measures changes in the complexity of the oximetry signal related to pediatric OSA severity by transforming the SpO<sub>2</sub> recordings into a two-symbol sequence and quantifying the number of different substrings in this transformed sequence (Lempel and Ziv, 1976).
- Sample entropy (*SampEn*). *SampEn* allows to quantify the irregularity of the oximetry signal by the evaluation of both prevailing and secondary patterns in the the SpO<sub>2</sub> recordings (Richman and Moorman, 2000).

#### 4.2.1.2 Novel oximetric indices in the time domain: detrended fluctuation analysis

DFA is an important tool to analyze the correlation properties of a non-stationary time series (i.e., the oximetry signal) through its multiscale analysis (Peng et al., 1994). In this respect, DFA allows to detect changes in the correlation properties of a signal along temporal scales caused by random spikes and/or segments with a distinct local behavior (Chen et al., 2002; Hua and Yu, 2017). In Vaquerizo-Villar et al. (2018a), we propose DFA as a novel nonlinear analysis method to analyze the irregular fluctuations and random spikes of the SpO<sub>2</sub> signal related to apneic events. Given a time series  $x(t)$ , the DFA procedure comprises the following four steps (Peng et al., 1994):

- 1) Integration of the input signal.  $x(t)$  is integrated using the following expression:

$$y(i) = \sum_{j=1}^i [x(j) - x_{mean}], i = 1, \dots, N, \quad (4.1)$$

where  $x_{mean}$  and  $N$  are the mean and length of the whole time series, respectively.

- 2) Window-segmentation.  $y(i)$  is divided into  $B$  non-overlapping time windows of equal size. The length of each window (i.e., the scale) ranges from  $k = 3$  to  $k = 1080$ , being the maximum scale (1080) one-tenth of the minimum signal length (10800 samples= 3 hours with a sampling of 1 Hz), which ensures that the recording contains an adequate number of sleep cycles (Davis et al., 2004; Kapur et al., 2017).
- 3) Obtaining of the local trend. A straight line least-squares fit is applied to  $y(i)$  in order to obtain the local trend  $y^b$  for each window ( $b = 1, \dots, B$ ).
- 4) Obtaining of the fluctuation function. The fluctuation function,  $F(k)$ , is obtained for each scale using the following expression:

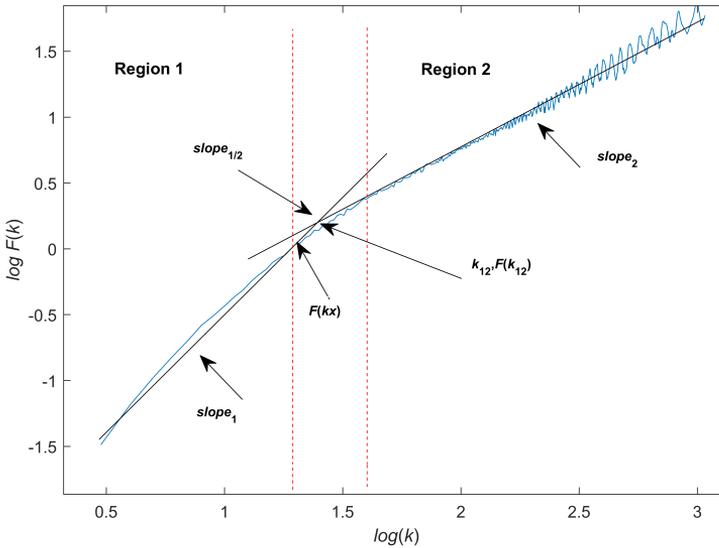
$$F(k) = \sqrt{\frac{1}{B} \cdot \sum_{b=1}^B F_b^2(k)}, \quad (4.2)$$

where  $F_b^2(k)$  is the variance of the fluctuation function in each window, defined as follows:

$$F_b^2(k) = \frac{1}{k} \sum_{j=(b-1)k+1}^{bk} [(y(j) - y^b(j))^2]. \quad (4.3)$$

Steps 2-4 are iterated until  $F(k)$  is obtained for each scale ( $k$ ) of the DFA profile. Figure 4.2 represents, in a logarithmic-scale plot, the evolution of the DFA profile of a subject along scales:  $\log_{10}F(k)$  versus  $\log_{10}(k)$ . It can be observed that there are two different scaling regions according to the linear relationship between  $\log_{10}F(k)$  and  $\log_{10}(k)$ . In Vaquerizo-Villar et al. (2018a), region 1 was obtained in the range of scales  $0.48 \leq \log_{10}(k) \leq 1.3$  ( $3 \leq k \leq 20$ ), whereas region 2 was obtained in the range  $1.60 \leq \log_{10}(k) \leq 3.03$  ( $40 \leq k \leq 1080$ ). In order to quantify the differences in the DFA plot associated to pediatric OSA and its severity, the following features were extracted, as shown in Figure 4.2 (Hua and Yu, 2017; Penzel et al., 2003; Vaquerizo-Villar et al., 2018a):

- Scaling exponents (slopes) in the straight line that fits both regions of the DFA plot ( $slope_1$  and  $slope_2$ ), which are intended to characterize the scaling behavior of the SpO<sub>2</sub> signal in each region.



**Figure 4.2.** Example of the DFA profile of the SpO<sub>2</sub> recording of a subject. This figure has been derived from [Vaquerizo-Villar et al. \(2018c\)](#)

- Ratio between  $slope_1$  and  $slope_2$  ( $slope_{12}$ ), which characterizes the distinct scaling behavior observed in both regions.
- Coordinates of the intersection between the straight lines adjusted in regions 1 and 2 of the DFA plot ( $k_{12}$  and  $F(k_{12})$ ), which characterize the crossover point of the DFA profile ([Vaquerizo-Villar et al., 2018a](#)).
- Value of the fluctuation function ( $F(k_x)$ ) in the scale  $k_x = 21$ , which has the highest correlation with the AHL.  $F(k_x)$  quantifies the fluctuations of the SpO<sub>2</sub> recording associated to apneic events.

#### 4.2.1.3 Conventional frequency domain analysis: power spectral density

Frequency domain analysis allows to measure the recurrence and duration of apneic events from OSA in children ([del Campo et al., 2018](#)). In this respect, PSD is the most used frequency domain technique to analyze the spectral components of the oximetric dynamics in the framework of pediatric OSA ([Álvarez et al., 2017](#); [Crespo et al., 2018](#); [del Campo et al., 2018](#); [Garde et al., 2014a](#); [Hornero et al., 2017](#)). PSD was computed using the Welch's approach ([Welch, 1967](#)) to look for differences in the oximetric recordings related to apneic events due to OSA. In order

to characterize the effects of OSA in the PSDs of the SpO<sub>2</sub> recordings, [Vaquerizo-Villar et al. \(2018b\)](#) and [Hornero et al. \(2017\)](#) determined two frequency bands of interest (BI): 0.018-0.050 Hz (*BI1*) and 0.021-0.044 Hz (*BI2*), respectively. In these bands, the highest statistically significant differences were obtained in the PSD amplitude among OSA severity groups. Once spectral bands were established, the following features were computed from the PSDs of each SpO<sub>2</sub> recording ([Vaquerizo-Villar et al., 2018a,b,c, 2021](#)):

- First-to-fourth order statistical moments, extracted from the full PSD ( $M1_{PSD} - M4_{PSD}$ ) and the two bands of interest: *BI1* ( $M1_{BI1} - M4_{BI1}$ ) and *BI2* ( $M1_{BI2} - M4_{BI2}$ ). These parameters measure the central tendency, dispersion, asymmetry and peakedness of the PSD in each region, respectively.
- Relative power ( $RP_{BI1}$ ), defined as the power ratio between *BI1* and the whole PSD.  $RP_{PSDBI1}$  is intended to reflect the effects in the band of interest caused by a higher occurrence of apneic events at these frequencies.
- Maximum amplitude ( $MA_{PSD}$ ) and minimum amplitude ( $mA_{PSD}$ ) of the PSD in the full spectrum ( $MA_{PSD}$  and  $mA_{PSD}$ ), *BI1* ( $MA_{BI1}$  and  $mA_{BI1}$ ), and *BI2* ( $MA_{BI2}$  and  $mA_{BI2}$ ). These features allow to quantify the highest and the lowest values of the PSD in these regions related to the occurrence of apneic events.
- Spectral entropy of the full PSD ( $SE_{PSD}$ ) and *BI1* ( $SE_{BI1}$ ), which are irregularity parameters that measure the flatness of the PSD and its band of interest, respectively ([Poza et al., 2007](#)).
- Mobility of the PSD in the band *BI1* ( $Mb_{BI1}$ ), defined as the squared root of the ratio between the variance and the power of the PSD in this band.  $Mb_{BI1}$  quantifies the concentration of the signal power ([Blanco-Velasco et al., 2010](#)).
- Median frequency of the full PSD ( $MF_{PSD}$ ), defined as the spectral component that separates the PSD into two regions, each of them with 50% of the total power.  $MF_{PSD}$  characterizes the distribution of the PSD content ([Poza et al., 2007](#)).
- Wootten's distance of the full PSD ( $WD_{PSD}$ ), which is a disequilibrium parameter that measures the distance between the probability density function of the PSD and an uniform distribution ([Martin et al., 2003](#)).

#### 4.2.1.4 Novel frequency domain features from the oximetry signal: bispectrum

Bispectrum can be described as the spectral representation of the third-order cumulant (skewness) of a time series (Chua et al., 2010). In contrast to conventional PSD, bispectrum preserves both amplitude and phase information of the spectral components of a signal. This allows bispectrum to detect phase relationships and deviations from linearity and gaussianity in a signal (Chua et al., 2010), such as those produced in physiological signals by respiratory events (Atri and Mohebbi, 2015; Tagluk and Sezgin, 2011). For these reasons, bispectral analysis is employed for the first time in Vaquerizo-Villar et al. (2018b) to characterize changes produced in the oximetry signal by respiratory events.

Given a deterministic and zero-mean time series  $x(t)$ , bispectrum can be non-parametrically estimated as follows:

$$B(f_1, f_2) = X(f_1) \cdot X(f_2) \cdot X(f_1 + f_2), f_1, f_2 = 0, \dots, N, \quad (4.4)$$

where  $X(f)$  is the discrete Fourier transform of  $x(t)$  and  $f_1$  and  $f_2$  are the frequencies of the two axes of the bispectrum. According to its symmetry conditions, bispectrum was computed in the non-redundant region ( $\Omega$ ), which satisfies  $f_1 \geq 0$ ,  $f_2 \geq f_1$ , and  $f_1 + f_2 \leq f_s/2$ , being  $f_s$  the sample rate of  $x(t)$ .

Once the bispectrum was computed in this region, the following parameters were computed:

- Mean amplitude of the bispectrum ( $M1_{BISP}$ ) (Chua et al., 2008), which allows to detect deviations of gaussianity in a signal (Ning and Bronzino, 1990):

$$M1_{BISP} = \frac{1}{L} \sum_{f_1, f_2 \in \Omega} |B(f_1, f_2)|, \quad (4.5)$$

where  $L$  is the number of points of the bispectrum in  $\Omega$ .

- Sum of logarithmic amplitudes of the bispectrum ( $H1_{BISP}$ ), sum of logarithmic amplitudes of diagonal elements of the bispectrum ( $H2_{BISP}$ ), and spectral moment of first order of diagonal elements of the bispectrum ( $H3_{BISP}$ ) (Zhou et al., 2008), which are related to the moments of the bispectrum.  $H1_{BISP}$ ,  $H2_{BISP}$ , and  $H3_{BISP}$  quantify the non-linearity of a signal and are calculated based on the bispectral amplitudes contained in  $\Omega$  and its main

diagonal (Chua et al., 2010; Zhou et al., 2008):

$$H1_{BISP} = \sum_{f_1, f_2 \in \Omega} \log(|B(f_1, f_2)|), \quad (4.6)$$

$$H2_{BISP} = \sum_{f_k \in \Omega_{diagonal}} \log(|B(f_k, f_k)|), \quad (4.7)$$

$$H3_{BISP} = \sum_{f_k \in \Omega_{diagonal}} k \cdot \log(|B(f_k, f_k)|), \quad (4.8)$$

where  $\Omega_{diagonal}$  is the main diagonal of the bispectrum.

- Bispectral amplitude entropies of first ( $BE1_{BISP}$ ) and second order ( $BE2_{BISP}$ ), which quantify the irregularity of the bispectral amplitude of a signal (Chua et al., 2008):

$$BE1_{BISP} = - \sum_{j \in \Omega} p_j \cdot \log(p_j), \quad (4.9)$$

where

$$p_j = \frac{|B(f_1, f_2)|}{\sum_{f_1, f_2 \in \Omega} |B(f_1, f_2)|}. \quad (4.10)$$

$$BE2_{BISP} = - \sum_{j \in \Omega} q_j \cdot \log(q_j), \quad (4.11)$$

where

$$q_j = \frac{|B(f_1, f_2)|^2}{\sum_{f_1, f_2 \in \Omega} |B(f_1, f_2)|^2}. \quad (4.12)$$

- Phase entropy ( $PE_{BISP}$ ), which measures irregularity in the bispectral phase (Chua et al., 2008):

$$PE_{BISP} = - \sum_n p(\Psi_n) \cdot \log(p(\Psi_n)), n = 0, 1, \dots, N - 1, \quad (4.13)$$

where  $N$  is the number of bins of the histogram and  $p(\Psi_n)$  is the distribution of the phase angles (Chua et al., 2010; Doane, 1976):

$$p(\Psi_n) = \frac{1}{L} \sum_{f_1, f_2 \in \Omega} \text{Ind}(\phi(B(f_1, f_2)) \in \Psi_n), \quad (4.14)$$

where  $\phi$  is the phase angle of the bispectrum and  $\text{Ind}(\cdot)$  is an indicator

function, whose value is 1 if  $\phi$  is within the range of the bin  $\Psi_n$ , being  $\Psi_n$  the range of histogram bins (Chua et al., 2010):

$$\Psi_n = \left\{ \phi \mid -\pi + \frac{2\pi n}{N} \leq \phi < -\pi + \frac{2\pi(n+1)}{N} \right\}. \quad (4.15)$$

- Mean ( $meanPa_{BISP}$ ) and variance ( $varPa_{BISP}$ ) of the bispectral invariant ( $P(a)$ ), which detect third-order time correlations or phase coupling between spectral components of a chaotic process (Chua et al., 2008).  $P(a)$  is defined as the phase of the integrated bispectrum along the straight line passing through the origin with slope  $a$  (Chua et al., 2008):

$$P(a) = \arctan\left(\frac{I_i(a)}{I_r(a)}\right), \quad (4.16)$$

where  $I_r(a)$  and  $I_i(a)$  are the real and imaginary components of the integrated bispectrum ( $I(a)$ ) (Chua et al., 2008):

$$I(a) = \int_{f_1=0^+}^{1/1+a} B(f_1, af_1) df_1 = I_r(a) + j \cdot I_i(a), 0 \leq a \leq 1. \quad (4.17)$$

#### 4.2.1.5 Novel frequency domain features from the oximetry signal: wavelet transform

Wavelet transform (WT) is a suitable method to analyze the spectral content of non-stationary signals (Rioul and Vetterli, 1991). In contrast to the fixed analysis window used by conventional frequency domain analysis techniques, which are based on the Short Time Fourier Transform (STFT), WT employs long windows at low frequencies and short windows at high frequencies. These multiscale analysis approach provide high frequency resolution at low frequencies and high temporal resolution at high frequencies, whereas the single scale analysis of STFT-based techniques is limited by its fixed time-frequency resolution (Rioul and Vetterli, 1991). Due to this multiresolution property, WT is applied in Vaquerizo-Villar et al. (2018c) as a novel technique especially designed to accurately detect SpO<sub>2</sub> desaturations elicited by apneic events, which have very low frequency components due to the slow variation nature of the oximetry signal.

WT decomposes a signal  $x(t)$  onto a set of basis functions, known as wavelets (Rioul and Vetterli, 1991). Wavelets are obtained by means of the scaling and time translation of a basic wavelet prototype, the mother wavelet (Rioul and Vetterli, 1991). According to the scale and translation values, there are two main types of

WT (Rioul and Vetterli, 1991): (i) continuous wavelet transform (CWT), where these time-scale parameters are continue; and (ii) discrete wavelet transform (DWT), where wavelets are computed only for power of 2 scales. In Vaquerizo-Villar et al. (2018c), DWT was chosen to analyze SpO<sub>2</sub> recordings, as it is less complex and computationally less expensive than CWT (Rioul and Vetterli, 1991).

Figure 4.3 shows the computation process of DWT. Given a SpO<sub>2</sub> recording  $x[n]$  with length  $N$ , DWT decomposes it using a filter-bank tree with  $L = \log_2(N)$  steps (Rioul and Vetterli, 1991), as shown in Figure 4.3a. Each step consists of a half-band high-pass filter  $g[n]$ , the mother wavelet, and a half-band low-pass filter  $h[n]$ , the mirror version of  $g[n]$ , each of them followed by a subsampling operation with a factor two (dyadic sampling). Thereby, in the first step ( $i = 1$ ), the SpO<sub>2</sub> signal  $x[n]$  is decomposed into an approximation  $A_1$  (low-pass) and a detail  $D_1$  (high-pass) signal. Next, the approximation signal  $A_1$  is further decomposed, thus generating another approximation signal  $A_2$  and another detail  $D_2$  signal. This decomposition iterates on the approximation signal until the maximum level of decomposition  $i = L$ . The frequency resolution of the approximation  $A_i$  and detail signals  $D_i$  is increased at each iteration by a factor two, while reducing their time resolution due to the dyadic sampling. At each step  $i = 1, 2, \dots, L$ ,  $A_i$  and  $D_i$  are obtained as follows:

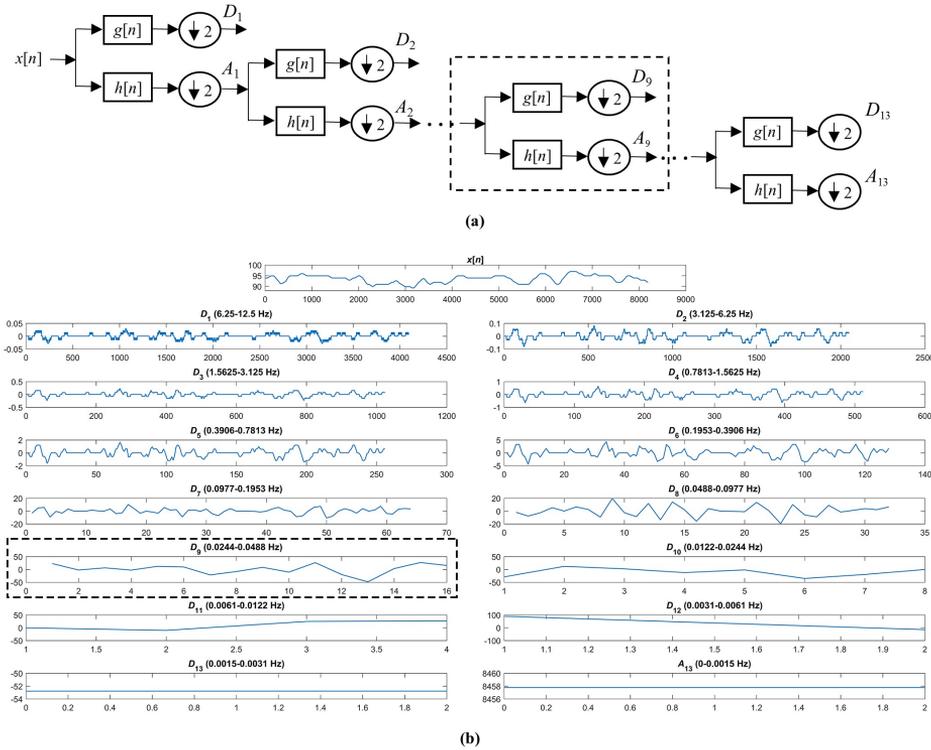
$$A_i[k] = \sum_n A_{i-1}[n] \cdot h[2k - n], \quad (4.18)$$

$$D_i[k] = \sum_n A_{i-1}[n] \cdot g[2k - n], \quad (4.19)$$

where  $A_{i-1}$  is the approximation signal in the step  $i - 1$ , being  $A_0$  the time series  $x[n]$ .

In Vaquerizo-Villar et al. (2018c), DWT was applied to  $N = 2^{13}$  (5.5 min) sample segments of the SpO<sub>2</sub> signal. Haar was chosen as the mother wavelet due to: (i) its stepped shape, which allows to detect oxygen desaturations elicited by apneic episodes; (ii) its smoothing property, which does not cause edge effects in the SpO<sub>2</sub> signal waveform.

Figure 4.3b shows an example of DWT decomposition of a SpO<sub>2</sub> segment,  $x[n]$ , using the Haar wavelet. It can be seen that detail coefficients from each decomposition level ( $D_i$ ) provide information about a specific frequency range. Our DWT analysis mainly focused on  $D_9$  (0.0244-0.0488 Hz), which is comprised in the frequency range of interest *BI1* (0.018-0.050 Hz) related to the changes caused by pediatric OSA in nocturnal oximetry (Vaquerizo-Villar et al., 2018c). As it can



**Figure 4.3.** DWT computation. (a) Decomposition process of a signal using DWT. (b) Original SpO<sub>2</sub> signal, detail signals at each decomposition level and approximation signal at the maximum level of the decomposition. This figure has been taken from [Vaquerizo-Villar et al. \(2018c\)](#)

be seen in Figure 4.3B,  $D_9$  coefficients show decreased and increased values associated to oxygen desaturations. The presence of both negative and positive values may hinder the information contained in the DWT coefficients, such as the mean or skewness. Thereby, the following DWT-derived features were extracted from the  $D_i$  coefficients in absolute value ([Vaquerizo-Villar et al., 2018c](#)):

- First-to-fourth order statistical moments of the  $D_9$  coefficients ( $M1_{D_9}$  –  $M4_{D_9}$ ), which are the mean ( $M1_{D_9}$ ), variance ( $M2_{D_9}$ ), skewness ( $M3_{D_9}$ ), and kurtosis ( $M4_{D_9}$ ). They measure the central tendency, dispersion, asymmetry and peakedness of the data, respectively.
- Maximum amplitude of the  $D_9$  coefficients ( $MA_{D_9}$ ), which allows to quantify the maximum amplitude reached in this frequency range.
- Energy of the  $D_9$  coefficients ( $En_{D_9}$ ), which allows to quantify the averaged

quadratic amplitude of the detail signal  $D_9$ . It is calculated using the following expression:

$$En_{D_9} = \sum_k |D_9[k]|^2. \quad (4.20)$$

- Wavelet entropy (WE), which quantifies the energy distribution changes elicited in the detail levels of the oximetry signal by apneic events (Rosso et al., 2001):

$$WE = - \sum_{i=1}^L p_i \log(p_i), \quad (4.21)$$

where  $p_i$  is the normalized wavelet energy at the detail level  $i$ :

$$p_i = \frac{En_{D_i}}{L} \sum_{i=1}^L En_{D_i}. \quad (4.22)$$

## 4.2.2 Feature selection

Once features are extracted, the information contained in a SpO<sub>2</sub> recording is synthesized in a wide set of variables. Nonetheless, there may be some features that provide irrelevant and/or redundant information that may cause overfitting in the pattern recognition stage (Guyon, 2003). Therefore, a feature selection stage was included to prevent from this undesired effect. Specifically, the fast correlation-based filter (FCBF) algorithm (Yu and Liu, 2004) was used in Vaquerizo-Villar et al. (2018b), Vaquerizo-Villar et al. (2018c), and Vaquerizo-Villar et al. (2018a) to obtain an optimum subset of OSA-related features (Yu and Liu, 2004).

### 4.2.2.1 Fast Correlation-Based Filter (FCBF)

FCBF is an automated feature selection algorithm that has been widely used in the pediatric OSA context (Barroso-García et al., 2020, 2021; Hornero et al., 2017; Jiménez-García et al., 2020; Vaquerizo-Villar et al., 2018a,b,c). FCBF follows a two-stage procedure to evaluate the relevance and redundancy of the variables under study. Given a feature vector  $x_i$  and a dependent variable  $y$ , FCBF first calculates the symmetrical uncertainty (SU) between each single feature  $x_i$  and  $y$  in order to evaluate its relevancy (Yu and Liu, 2004):

$$SU(x_i, y) = 2 \cdot \left( \frac{IG(x_i|y)}{H(x_i) + H(y)} \right), i = 1, 2, \dots, N, \quad (4.23)$$

where  $IG(x_i|y) = H(x_i) - H(x_i|y)$ ,  $H$  refers to the well-know Shannon's entropy,  $N$  is the number of input variables (Yu and Liu, 2004), and  $y$  is a context-dependent characteristic. Particularly, in the framework of pediatric OSA,  $y$  is a vector composed of the AHI values for each pediatric subject.  $SU$  can take values in the range  $0 - 1$ .  $SU = 0$  indicates that the two variables are totally independent, whereas  $SU = 1$  indicates that it is completely possible to forecast one feature from the other. Therefore, features having a higher value of  $SU$  are more relevant, as they share more information with the dependent variable (Yu and Liu, 2004).

Once  $SU$  is obtained for all the input variables, features are sorted from the most relevant (highest  $SU$ ) to the least relevant one (lowest  $SU$ ). Based on this ranking, the second step consists on the computation of  $SU$  between each pair of features ( $SU(x_i, x_j)$ ),  $j = 1, 2, \dots, N, j \neq i$  to assess its redundancy, beginning from the most relevant ones (Yu and Liu, 2004). If  $SU(x_i, x_j) \geq SU(x_i, y)$ , being  $x_i$  more relevant than  $x_j$ ,  $x_j$  is removed from the selection process due to redundancy with  $x_i$ . Accordingly, an optimum subset of features is obtained with those not discarded during the redundancy analysis.

In the present Doctoral Thesis, the FCBF algorithm was applied in the three studies that followed a feature-engineering methodology (Vaquerizo-Villar et al., 2018a,b,c). In these papers, FCBF was applied to 1000 bootstrap-derived samples from the corresponding training set, as it contributes to select a more stable subset of features (Witten et al., 2011). Then, the features selected at least 500 times (50% of the iterations) constituted the optimum subset.

### 4.2.3 Pattern recognition

Pattern recognition consists on the application of automated algorithms to identify the underlying behavior in the data (Bishop, 2006). Taking as input data the optimum subset of features from each subject obtained with FCBF, three pattern recognition algorithms have been applied in the present Doctoral Thesis to automatically detect pediatric OSA and its severity: (i) LR; (ii) SVM; and (iii) MLP.

#### 4.2.3.1 Logistic regression (LR)

LR is a common method for two-class (binary) classification. It is a supervised algorithm which assigns an input feature vector  $x_i$  (with  $i = 1, 2, S$ , being  $S$  the number of instances) into one out of two mutually exclusive groups ( $C_j = 1, 2$ ) (Hosmer and Lemeshow, 2004). In this Thesis, LR was used to estimate the pos-

terior probability of a given instance  $x_i$  (subject) belonging to the negative OSA ( $C_1$ ) and positive OSA ( $C_2$ ) groups (Vaquerizo-Villar et al., 2018c). This was carried out through the logistic function (Hosmer and Lemeshow, 2004):

$$p(C_j|x_i) = \frac{\exp(\beta_0 + x_i\beta)}{1 + \exp(\beta_0 + x_i\beta)}, \quad (4.24)$$

where  $\beta_0$  and  $\beta = \beta_1, \beta_2, \dots, \beta_N$  are the coefficients of the LR model, and  $N$  is the number of input features that compose each vector  $x_i$ .  $\beta_0$  and  $\beta$  coefficients are optimized using the maximum likelihood estimator (Hosmer and Lemeshow, 2004).

#### 4.2.3.2 Support vector machines (SVM)

SVM is a supervised algorithm for binary classification that finds a separating hyperplane with a decision boundary that maximizes the distance between instances belonging to different classes. The hyperplane is expressed as follows (Bishop, 2006):

$$y(x, w) = w^T \cdot \phi(x) + w_0, \quad (4.25)$$

where  $x \in R_N$  is the input feature vector,  $N$  is the number of features,  $\phi(x) \in R_D$  is the feature vector in the high-dimensional transformed space ( $D > N$ ), and  $w$  denotes the weight vector, which is optimized in order to obtain a maximal margin hyperplane (Bishop, 2006). This optimization can be formulated using Lagrange multipliers:

$$y(x, w) = - \sum_{j \in V} \eta^j t^j K(x^j, x) + w_0, \quad (4.26)$$

where  $V$  is a subset of indices  $1, \dots, S$  corresponding to the Lagrange multipliers  $\eta^j$  related to the supported vectors,  $S$  is the number of examples in the training group,  $K(\cdot, \cdot)$  is the Kernel function, and  $t^j$  are the output labels. The Lagrange multipliers ( $\eta^1, \dots, \eta^S$ ) are subjected to the following constraints:

$$\sum_{j=1}^S \eta^j t^j = 0 \quad (4.27)$$

and

$$0 \leq \eta^j \leq C, \quad (4.28)$$

where  $C$  is the regularization parameter, which controls the balance between maximizing the margin of separation between classes and minimizing the classification error.

In [Vaquerizo-Villar et al. \(2018c\)](#), SVM was applied using a linear kernel to assign every input feature vector to the groups negative OSA ( $t^j = -1$ ) and positive OSA ( $t^j = 1$ ), whereas the optimum value for the regularization parameter  $C$  was obtained using a 10-fold cross validation procedure during the training stage.

#### 4.2.3.3 Multilayer perceptron neural network (MLP)

Artificial neural networks (ANN) arised from the need to model information processing in biological systems using mathematical representations ([Bishop, 2006](#)). MLP is the ANN-based pattern recognition algorithm most widely used in the pediatric OSA context ([Gutiérrez-Tobal et al., 2021](#)). MLP is a feed-forward neural network with an architecture consisting on several fully-connected layers (input, hidden, and output layers) composed of basic mathematical units that imitate biological neurons, called perceptrons ([Bishop, 2006](#)). These units are described by a differentiable activation function  $g(\cdot)$  that performs a nonlinear transformation of the data, as well as by adaptive weights  $w$  that connect each unit with every neuron from the subsequent layer. In this Thesis, a configuration with a single hidden layer was used. Given an input feature vector  $x_i, i = 1, \dots, N$ , being  $N$  the number of features, the values of the output units  $y_k$  of the MLP architecture are computed as follows:

$$y_k = g_o \left\{ \sum_{j=1}^{N_H} w_{jk} g_h \left\{ \sum_{i=1}^N w_{ij} x_i + b_j \right\} + b_k \right\}, k = 1, \dots, K, \quad (4.29)$$

where  $g_o(\cdot)$  and  $g_h(\cdot)$  are the activation functions for units in the output and hidden layer, respectively,  $w_{jk}$  is the weight connecting the neuron  $j$  of the hidden layer with the output unit  $y_k$ ,  $w_{ij}$  is the weight connecting the input feature  $i$  with the hidden neuron  $j$ ,  $b_k$  and  $b_j$  are the biases of the output and hidden layers, respectively,  $K$  is the number of output units, and  $N_H$  is the number of neurons in the hidden layer. In this Thesis, MLP was used in [Vaquerizo-Villar et al. \(2018b\)](#) for multiclass classification (negative OSA, moderate OSA, and severe OSA), in [Vaquerizo-Villar et al. \(2018c\)](#) for binary classification (negative OSA and positive OSA), and in [Vaquerizo-Villar et al. \(2018a\)](#) for regression (AHI estimation). Accordingly, the number of output units  $K$  was different in each article:  $K = 3$  for three-class classification ([Vaquerizo-Villar et al., 2018b](#)),  $K = 2$  for binary classifi-

cation (Vaquerizo-Villar et al., 2018c), and  $K = 1$  for AHI estimation (Vaquerizo-Villar et al., 2018a).

Network weights were adjusted during training using the scale conjugate gradient algorithm with weigh-decay regularization, which allows to minimize the loss function and achieve a good generalization ability (Bishop, 2006). During the training process, two hyperparameters were optimized through 10-fold cross validation: the regularization parameter ( $\alpha$ ) and  $N_H$ .

### 4.3 Deep learning

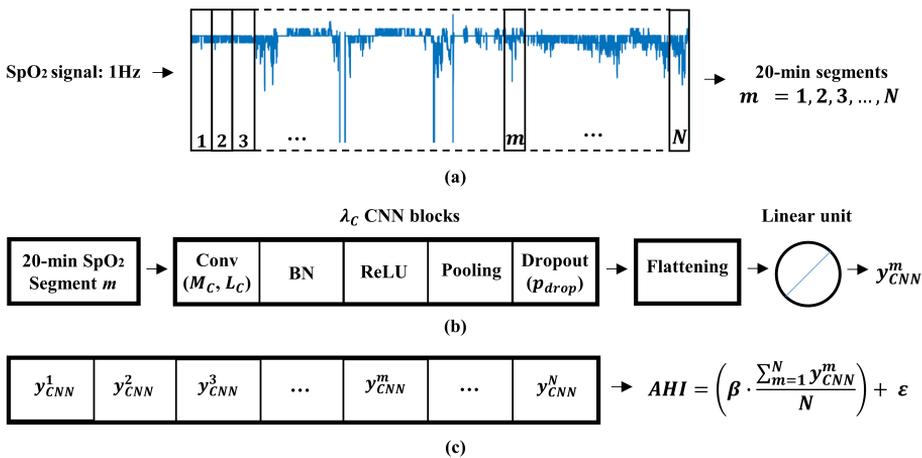
Conventional feature-engineering approaches have two main limitations (Goodfellow et al., 2016): (i) a substantial knowledge in the specific field is required to determine, *a priori*, a set of relevant features that must be obtained from the input data; and (ii) these methods are limited by their low level of abstraction, which limits their capability to find relevant features in the raw data. These limitations may lead to the omission of specific features from the oximetry signal linked to the physiological perturbations of pediatric OSA. These shortcomings can be minimized by means of the application of deep-learning algorithms. As aforementioned, deep-learning methods have beaten conventional approaches in many fields (LeCun et al., 2015), including the OSA context (Faust et al., 2019; Mostafa et al., 2019), primarily due to their capability to automatically discover intricate patterns from the raw data using ANNs with a high number of hidden layers (Faust et al., 2019). In this research, we have evaluated a new deep-learning model based on CNNs (Vaquerizo-Villar et al., 2021), the most widely-used deep-learning architecture in the OSA context (Faust et al., 2019; Mostafa et al., 2019). Despite the fact that CNNs were originally inspired to process image data, these architectures have proven to be the most appropriate for time series classification in many domains (Ismail Fawaz et al., 2019), including biomedical signal analysis (Ebrahimi et al., 2020; Faust et al., 2018; Murat et al., 2020; Roy et al., 2019). CNNs have a multi-layer architecture whose main properties are: shared weights, sparse connections, and pooling operations (Goodfellow et al., 2016). The first layers of the CNN architecture identify local motifs, whereas the deeper layers detect long-term patterns occurring in different parts of the array (Ebrahimi et al., 2020). Accordingly, a CNN architecture is applied in Vaquerizo-Villar et al. (2021) to identify oxygen desaturations (simple patterns) and clusters of desaturations (complex patterns) linked to pediatric OSA and its severity occurring in different time locations of the SpO<sub>2</sub> recording.

### 4.3.1 Proposed convolutional neural network (CNN) model

Figure 4.4 depicts a visual overview of the CNN-based deep-learning model, which consists of three stages: (i) signal segmentation; (ii) CNN architecture; and (iii) AHI estimation.

#### 4.3.1.1 Signal segmentation

SpO<sub>2</sub> recordings were first decimated to a common sample frequency of 1 Hz. As shown in Figure 4.4a, SpO<sub>2</sub> signals were divided into 20-min non-overlapping segments (1200 samples). This segment size was chosen as it allows to detect clusters of desaturations linked to apneic events from pediatric OSA, whose duration is of at least 10-min (Brouillette et al., 2000). Then, each 20-min SpO<sub>2</sub> segment used to train the CNN architecture was labelled with the respiratory event annotations scored by the technicians. In this respect, only the CHAT dataset was used for training purposes, as the UofC and BUH databases do not contain files with annotations of time location of apnea and hypopnea events. Accordingly, the output label for each 20-min segment of the CHAT dataset was obtained as the total number of apneic events (apnea plus hypopnea) associated to 3% oxygen desaturations occurring in this time window, according to the annotations provided by sleep technicians (Marcus et al., 2013).



**Figure 4.4.** Overview of the proposed CNN-based deep-learning methodology. (a) Signal segmentation, (b) CNN architecture, and (c) AHI estimation. This figure has been taken from Vaquerizo-Villar et al. (2021)

### 4.3.1.2 CNN architecture

Figure 4.4b shows the architecture of the proposed CNN. The input section of the CNN consists of the 20-min SpO<sub>2</sub> segments. Each segment is processed using  $\lambda_C$  convolutional blocks, each one ( $i = 1, \dots, \lambda_C$ ) consisting of:

- I. 1-D convolution. This layer extracts feature maps from the input  $a_i[n]$  using  $M_C$  convolutional filters, so-called kernels (Goodfellow et al., 2016):

$$x_i^l[n] = \sum_{k=1}^{L_C} w_k^l * a_i[n - k + 1] + b_k^l, \quad (4.30)$$

where  $x_i^l$  is the  $l$ th feature map ( $l = 1, \dots, M_C$ ) in the  $i$ th convolutional block,  $L_C$  is the filter (kernel) size,  $w_k^l$  and  $b_k^l$  are the kernel weights and biases, respectively, and  $a_0[n]$  is the input 20-min SpO<sub>2</sub> segment.

- II. Batch normalization (BN). BN is used to normalize the amplitude of each feature map  $x_i^l$  obtained after the 1-D convolution (Goodfellow et al., 2016).
- III. Rectified Linear Unit (ReLU). ReLU is the most common activation function for CNNs. It performs a thresholding operation in order to decide which normalized feature maps are relevant (Goodfellow et al., 2016):

$$f(x) = \max(0, x). \quad (4.31)$$

- IV. Pooling. After the activation function, a max-pooling operation with a pool factor  $K = 2$ , the standard choice for CNNs, is applied to reduce dimensionality, while retaining the most significant features (Goodfellow et al., 2016).
- V. Dropout. As a final step in each convolutional block, dropout operation was included to minimize overfitting. It randomly remove connections between network elements with a probability  $p_{drop}$  during the training process (Goodfellow et al., 2016).

Following the last convolutional block ( $i = \lambda_C$ ), a flattening layer is applied to convert the 2-D feature maps into 1-D data (Goodfellow et al., 2016). Finally, a linear activation unit estimates the number of apneic episodes associated to desaturations,  $y_{CNN}^m$ , which is the output of the network.

### 4.3.1.3 AHI estimation

Once the output  $y_{CNN}^m$  of the CNN is obtained for each segment  $m = 1, 2, \dots, N$ , the AHI of each pediatric subject can be computed. First, the mean value of all the outputs of the CNN of each patient is obtained using the following expression:

$$y_{CNN}^{avg} = \frac{\sum_{m=1}^N y_{CNN}^m}{N}, \quad (4.32)$$

where  $N$  is the number of segments in the SpO<sub>2</sub> signal, which is different for each patient. Then, the final AHI of the patient is calculated using a linear regression model, as depicted in Figure 4.4c:

$$AHI = (\beta \cdot y_{CNN}^{avg}) + \epsilon, \quad (4.33)$$

where  $\epsilon$  and  $\beta$  are the disturbance and interception terms of a linear regression model, which was adjusted during the optimization stage. This procedure allows to counteract the underestimation of the AHI due to (Deviaene et al., 2018): (i) not all apneic events result in an oxygen desaturation, so that the CNN can not detect them; (ii) AHI estimation is performed using the total recording time, as the total sleep time is unknown using only the oximetry signal.

## 4.3.2 CNN training and optimization process

The network training was performed using the following configuration: He-normal method for weights initialization (He et al., 2015); adaptive moment estimation (Adam) algorithm with an initial learning rate of 0.001 for the optimization of weights and biases (Kingma and Ba, 2014); Huber loss with a tunable hyperparameter  $\delta$  as the loss function to minimize during training (Huber, 1964); batch size of 100 with a data shuffling strategy to accelerate the convergence of Adam method (Goodfellow et al., 2016); and 500 epochs. To speed up training and obtain a final stable set of network weights, the learning rate was decreased by a factor of 2 when the loss in the validation set did not improve for 10 consecutive epochs, and early stopping was applied after 30 epochs of non-improvement, restoring the weights that minimized the validation loss (Goodfellow et al., 2016).

The hyperparameter optimization plays an essential role in the design of a suitable deep-learning model (Goodfellow et al., 2016). In this research, the following hyperparameters of the deep-learning model were optimized: the number of CNN blocks ( $\lambda_C$ ), the number of filters ( $M_C$ ) and the filter size ( $L_C$ ) in each

1-D convolution, the dropout probability ( $p_{drop}$ ), and the delta parameter of the Huber loss function ( $\delta$ ). In order to reach an optimal solution, these hyperparameters were automatically optimized using Bayesian optimization with tree-structured Parzen estimator (BO-TPE) (Bergstra et al., 2011). BO-TPE is considered a suitable strategy for optimization purposes, since it forms a probabilistic model that tries to approximate the objective function iteratively, based on past evaluation results (Snoek et al., 2012).

## 4.4 Statistical analysis

The following techniques have been employed to interpret and evaluate the results obtained with the signal processing methods developed in this Doctoral Thesis: (i) statistical hypothesis tests; (ii) diagnosis performance metrics; (iii) measures of agreement; and (iv) validation strategies.

### 4.4.1 Statistical hypothesis tests

Hypothesis tests are methods of statistical inference that assess whether it is possible to infer properties of a population from the results observed in a given data sample (Jobson, 2012). In this research, statistical hypothesis testing was first employed to evaluate the normality (Lilliefors test) and homocedasticity (Levene test) of the demographic variables and the features extracted from the SpO<sub>2</sub> recordings. As not all the demographic and oximetric variables passed normality and homocedasticity tests, non-parametric tests were applied to search for statistical significant differences among the different groups under study (negative OSA/positive OSA, OSA-severity degrees, and validation groups) (Jobson, 2012). The Mann-Whitney  $U$  test was used to evaluate statistically significant differences between two groups (negative OSA/positive OSA and two validation groups), whereas the Kruskal-Wallis test was employed for comparisons in those cases with more than two groups (OSA-severity degrees and three validation groups). Depending on the number of subjects, two different  $p$ -values were used in this Thesis to search for statistically significant differences. In Vaquerizo-Villar et al. (2018b), where the sample size was 298 pediatric subjects, a  $p$ -value  $< 0.05$  was employed to evaluate statistical differences. Conversely, a  $p$ -value  $< 0.01$  was employed in the remaining studies (Vaquerizo-Villar et al., 2018a,c, 2021), where a larger sample size was used. In the case of multiple comparisons, the Bonferroni correction was used.

### 4.4.2 Diagnostic performance metrics

The usefulness of a diagnostic test can be expressed by the use of different statistical measures. The definition of these statistics is based on the number of subjects correctly and wrongly classified. In the problem of statistical classification, these measures are derived from the confusion matrix, which compares the class predicted by our test with the actual class obtained with the reference test. When the confusion matrix is intended to discern between two population groups (i.e. negative OSA and positive OSA), a binary confusion matrix is obtained, whose main elements are:

- True positives (*TP*). Number of subjects with the disease (positive subjects, according to the gold standard) that have been rightly classified by the test under study.
- False negatives (*FN*). Number of positive subjects (according to the gold standard) that have been wrongly classified as healthy or negative by the test under study.
- True negatives (*TN*). Number of negative subjects (according to the gold standard) that have been correctly classified by the test under study.
- False positives (*FP*). Number of negative subjects (according to the gold standard) that have been wrongly classified as positive by the test under study.

Based on the aforementioned elements (Flemons and Littner, 2003), the following statistics have been calculated in this Doctoral Thesis (Vaquerizo-Villar et al., 2018a,b,c, 2021):

- Sensitivity (*Se*). Proportion of positive subjects correctly classified:

$$Se = \frac{TP}{TP + FN} \cdot 100 \quad (4.34)$$

- Specificity (*Sp*). Proportion of negative subjects correctly classified:

$$Sp = \frac{TN}{TN + FP} \cdot 100 \quad (4.35)$$

- Accuracy (*Acc*). Proportion of subjects correctly classified. This definition can also be extended for confusion matrices with more than two groups. In

the case of binary classification, it is obtained as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100 \quad (4.36)$$

- Positive predictive value (*PPV*). Proportion of subjects rightly classified among all the subjects that the test under study has assigned to the positive class:

$$PPV = \frac{TP}{TP + FP} \cdot 100 \quad (4.37)$$

- Negative predictive value (*NPV*). Proportion of subjects rightly classified among all the subjects that the test under study has assigned to the negative class:

$$NPV = \frac{TN}{TN + FN} \cdot 100 \quad (4.38)$$

- Positive likelihood ratio (*LR+*). Proportion of positive subjects rightly classified with respect to the proportion of negative subjects wrongly classified:

$$LR+ = \frac{Se}{1 - Sp} \quad (4.39)$$

*LR+* varies between 1 and  $+\infty$ , being desired values close to  $+\infty$ .

- Negative likelihood ratio (*LR-*). Proportion of positive subjects wrongly classified with respect to the proportion of negative subjects right classified:

$$LR- = \frac{1 - Se}{Sp} \quad (4.40)$$

*LR-* varies between 0 and 1, being desired values close to 0.

- Area under the Receiver-Operating Characteristics (ROC) curve. The ROC curve is commonly employed to compare the yield of different diagnostic tests. A ROC plot represents the *Se* vs.  $1 - Sp$  curve, where *Se* and *Sp* are obtained varying the decision threshold of the test under study (Zweig and Campbell, 1993). Once the plot has been constructed, the area under the ROC curve (AUC) can be explained as a method to assess the overall performance of a test (Zweig and Campbell, 1993). The values of AUC are constrained to the interval between 0.5 and 1, as values below 0.5 indicate that the positiveness of the test must be changed. In this respect, a high discriminative performance is obtained when the AUC value is close to 1.

### 4.4.3 Measures of agreement

The agreement between the gold standard and the simplified diagnostic alternatives under study was assessed by means of the following measures:

- Cohen's kappa ( $\kappa$ ).  $\kappa$  measures the agreement between observed and predicted classes, i.e., when considering binary or multiclass classification, without considering the agreement that may occur by chance (Cohen, 1960). It is computed as:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (4.41)$$

where  $p_o$  is the observed agreement between predicted and observed classes and  $p_e$  is the probability of agreement by chance.  $\kappa$  varies from -1 (total disagreement) to 1 (perfect agreement), with  $\kappa=0$  indicating that the agreement is due completely to chance (Cohen, 1960).  $\kappa$  was used to assess the overall agreement of our signal processing algorithms to establish pediatric OSA and its severity in Vaquerizo-Villar et al. (2018b), Vaquerizo-Villar et al. (2018a), and Vaquerizo-Villar et al. (2021).

- Root mean square error (RMSE). RMSE is an estimate of concordance between predicted and observed continuous variables, thus being useful to assess the performance of regression methods. In this Thesis, RMSE was used to measure the agreement between the AHI predicted by our signal processing algorithms ( $AHI_{pred}$ ) and the actual AHI from PSG ( $AHI_{PSG}$ ) (Vaquerizo-Villar et al., 2021). Given  $AHI_{pred}$  and  $AHI_{PSG}$ , RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (AHI_{pred}(n) - AHI_{PSG}(n))^2}{N}} \quad (4.42)$$

where  $N$  is the number of instances (subjects).

- Intra-class correlation coefficient (ICC). ICC is another index commonly used to evaluate the performance of a regression algorithm. In contrast to conventional correlation coefficients, ICC considers systematic error when assessing agreement. Depending on its specific purpose and the statistical model assumed, there exist several definitions of ICC (Chen and Barnhart, 2008). In this research (Vaquerizo-Villar et al., 2018a, 2021), the purpose is to measure agreement between  $AHI_{pred}$  and  $AHI_{PSG}$ , without making any ANOVA assumptions or using replicated measurements. For this case, the

following definition is recommended (Chen and Barnhart, 2008):

$$ICC = \frac{MS_S - MS_E}{MS_S - (N_O - 1) \cdot MS_E + N_O \cdot (MS_T - MS_E) / N} \quad (4.43)$$

where  $N_O$  is the number of observers,  $N$  is the number of subjects considered,  $MS_S$  is the subjects mean square,  $MS_E$  is the error mean square, and  $MS_T$  is the observers mean square. The values of ICC ranges from 0 (no agreement) to 1 (total agreement) (Weir, 2005).

#### 4.4.4 Validation strategies

Several validation methodologies have been applied with the aim to maximize the generalization of the results obtained during the compendium of publications. For the purpose of minimizing potential overfitting, a dataset must be divided in as many subsets as the number of degrees of freedom (number of optimization steps) needed to be fitted for a specific problem (Witten et al., 2011). In this respect, a hold-out strategy (training-test, training-validation-test, etc.) was employed in all the articles of this compendium to properly validate their results (Vaquerizo-Villar et al., 2018a,b,c, 2021). When the sample size was not large enough to handle all the optimizations required, two additional validation techniques were used: stratified K-fold cross-validation and bootstrapping.

**Hold-out validation** The natural way to correctly validate the results of a given model is to use a different set to optimize each stage of the proposed methodology (Witten et al., 2011). In the most simple case, a training group is employed to fit the model parameters and an independent test group, also called hold-out set, is used to estimate its performance (Bishop, 2006; Witten et al., 2011). As previously mentioned, the dataset should be divided into more than two sets when model fitting requires more than one optimization step (Witten et al., 2011). In this respect, a common strategy consists of using three subsets: training, used to adjust model parameters; validation, used to adjust model hyperparameters; and test, used for independent diagnostic performance assessment. When it is not possible to use a different set to optimize each stage because of the available amount of data, hold-out must be combined with additional validation strategies, such as K-fold cross-validation and bootstrapping. In this research, hold-out was used with two subsets in Vaquerizo-Villar et al. (2018c) and Vaquerizo-Villar et al. (2018a), and with three subsets in Vaquerizo-Villar et al. (2018b) and Vaquerizo-

Villar et al. (2021).

**Stratified K-fold cross-validation** Stratified K-fold cross-validation is another common validation approach. This technique randomly divides the data into K folds, keeping the proportion of instances pertaining to the groups under study. Then, K-1 folds are employed for model fitting (training), whereas the remaining fold is used for model evaluation (test). This procedure is repeated K times, so that each subset was considered once as the test group (Steyerberg and Vergouwe, 2014). Once the model has been tested using all the instances, diagnostic ability statistics can be averaged across the different folds. This may lead to more generalizable models, at the cost of increasing the computational cost (Witten et al., 2011). This validation approach was used for hyperparameters optimization in Vaquerizo-Villar et al. (2018b), Vaquerizo-Villar et al. (2018c), and Vaquerizo-Villar et al. (2018a), and for diagnostic performance assessment in Vaquerizo-Villar et al. (2018c).

**Bootstrapping** A bootstrap procedure can also be used for validation purposes when the data set is small (Witten et al., 2011). As stated in Section 4.2.2, FCBF was applied along with a bootstrap procedure to find a more generalizable optimum set of features (Guyon, 2003). Given a feature set  $x$  of  $S$  instances (subjects),  $x = x_1, x_2, \dots, x_S$ ,  $x_b (b = 1, 2, \dots, B)$  new sets (bootstrap replicates) of size  $S$  are built by resampling with replacement from the initial set, following a uniform distribution. In this Doctoral Thesis (Vaquerizo-Villar et al., 2018a,b,c), the FCBF algorithm was applied to each of these  $B = 1000$  bootstrap samples, obtaining optimum subsets composed of those features selected in more than 500 replicates (see section 4.2.2).

# Chapter 5

## Results

This chapter presents the main outcomes obtained in this Doctoral Thesis. These results are organized according to the statements that compose the global hypothesis of Section 2.1, thus having almost a directly connection with the papers included in the compendium of publications (see Appendix A).

### 5.1 Application of novel feature-extraction algorithms

As explained in Section 1.1, three novel feature-extraction algorithms were applied in this research to obtain features able to provide additional information from the oximetry signal linked to apneic events due to pediatric OSA: bispectrum, wavelet analysis, and DFA.

#### 5.1.1 Bispectral analysis

The complementarity of bispectrum to conventional approaches in the screening of childhood OSA using SpO<sub>2</sub> recordings was evaluated in [Vaquerizo-Villar et al. \(2018b\)](#). For this purpose, up to 22 features were obtained from a database of 298 pediatric patients (i.e., the initial version of the UofC database): ODI3, 3 anthropometric variables, 9 features from the PSD, and 9 bispectral parameters. Then, the FCBF algorithm was applied to select an optimum subset of OSA-related features. Finally, a MLP neural network was trained with the selected variables to detect pediatric OSA severity degrees. According to this feature-engineering methodology, the database was split into three sets: (i) feature opti-

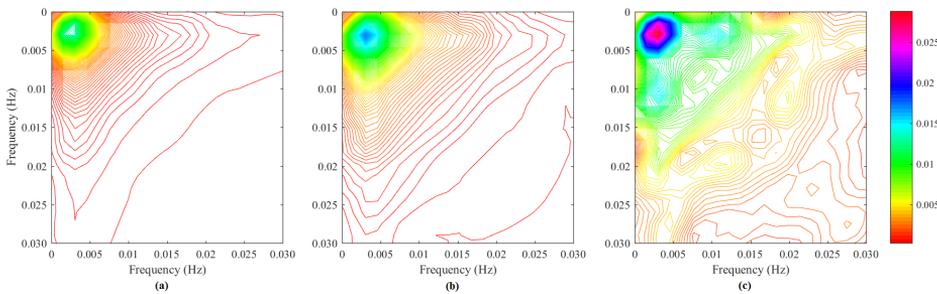
mization set (74 subjects, 25%), employed to optimize the feature extraction and selection stages; (ii) training set (149 subjects, 50%), employed to optimize the hyperparameters of the MLP classifier as well as to train the MLP model; (iii) test set (75 subjects, 25%), used to examine the diagnostic ability of the proposed methodology in an independent set.

Figure 5.1 shows the averaged amplitude in absolute value of the bispectrum in the feature optimization set for the following OSA severity groups: negative OSA (AHI  $< 5$  e/h), moderate OSA ( $5 \leq$  AHI  $< 10$  e/h), and severe OSA (AHI  $\geq 10$  e/h). As shown, a higher bispectral amplitude is appreciated in the frequency range 0-0.03 Hz as the OSA severity increases. Table 5.1 shows the median and interquartile range of the extracted features in the feature optimization set for negative OSA, moderate OSA, and severe OSA groups, along with their corresponding  $p$ -values, obtained by means of the Kruskal-Wallis test. ODI3, 6 out of 9 PSD-derived parameters ( $M1_{B1L}$ ,  $M2_{B1L}$ ,  $MA_{B1L}$ ,  $mA_{B1L}$ ,  $SE_{B1L}$ , and  $Mb_{B1L}$ ), and 4 out of 9 bispectrum-derived features ( $M1_{BISP}$ ,  $H1_{BISP}$ ,  $H2_{BISP}$ , and  $H3_{BISP}$ ) showed statistically significant differences ( $p$ -value  $< 0.05$  after Bonferroni correction). Specifically, higher values were obtained in these variables as pediatric OSA severity increased.

As the goal is to assess the complementarity of bispectrum with respect to conventional approaches, two different feature sets were composed:

- $set_{nobis}$ : composed of all but bispectrum features (ODI3, anthropometric variables, and PSD-derived features).
- $set_{bis}$ : consisting of all the extracted features.

FCBF was applied to these two feature sets ( $set_{nobis}$  and  $set_{bis}$ ) in the fea-



**Figure 5.1.** Averaged amplitude in absolute value of the bispectrum for the three OSA severity groups. (a) negative OSA, (b) moderate OSA, and (c) severe OSA. This figure has been taken from [Vaquerizo-Villar et al. \(2018b\)](#).

**Table 5.1.** Feature values for the OSA severity groups (median [interquartile range]) in the feature optimization set. This table has been derived from [Vaquerizo-Villar et al. \(2018b\)](#)

| Features                      | negative OSA        | moderate OSA        | severe OSA          | <i>p</i> -value* |
|-------------------------------|---------------------|---------------------|---------------------|------------------|
| Age                           | 7.0 [4.8,11.0]      | 4.0 [2.0,8.0]       | 6.0 [3.0,9.0]       | 0.24             |
| Sex                           | -                   | -                   | -                   | 0.52             |
| BMI ( $10^1$ )                | 1.82 [1.68,2.26]    | 1.9 [1.62,2.30]     | 1.77 [1.54,2.23]    | 0.76             |
| ODI3                          | 2.08 [0.77,3.93]    | 5.82 [3.79,9.28]    | 8.72 [7.23,19.65]   | <0.05            |
| $M1_{BI1}$                    | 2.95 [1.93,4.23]    | 5.67 [5.10,8.58]    | 14.73 [7.51,28.45]  | <0.05            |
| $M2_{BI1}$ ( $10^1$ )         | 0.18 [0.05,0.32]    | 1.11 [0.56,1.54]    | 4.07 [1.34,14.39]   | <0.05            |
| $M3_{BI1}$                    | 0.48 [0.29,0.75]    | 0.68 [0.33,1.04]    | 0.54 [0.32,0.83]    | 0.54             |
| $M4_{BI1}$                    | 2.10 [1.79,2.73]    | 2.33 [1.87,3.06]    | 2.20 [1.60,2.70]    | 0.65             |
| $RP_{BI1}$                    | 0.31 [0.25,0.35]    | 0.24 [0.21,0.34]    | 0.34 [0.30,0.39]    | 0.05             |
| $MA_{BI1}$                    | 0.60 [0.33,0.87]    | 1.27 [1.00,1.54]    | 2.92 [1.68,4.59]    | <0.05            |
| $mA_{BI1}$                    | 1.21 [0.84,2.12]    | 2.30 [1.91,2.88]    | 3.46 [2.56,8.59]    | <0.05            |
| $SE_{BI1}$                    | 4.33 [4.26,4.37]    | 4.30 [4.24,4.32]    | 4.20 [4.12,4.30]    | <0.05            |
| $Mb_{BI1}$                    | 0.17 [0.11,0.20]    | 0.28 [0.23,0.32]    | 0.42 [0.25,0.56]    | <0.05            |
| $M1_{BISP}$ ( $10^{-1}$ )     | 0.05 [0.02,0.21]    | 0.23 [0.09,0.63]    | 0.49 [0.17,1.22]    | <0.05            |
| $H1_{BISP}$ ( $10^8$ )        | -5.78 [-5.90,-5.57] | -5.59 [-5.73,-5.45] | -5.40 [-5.58,-5.28] | <0.05            |
| $H2_{BISP}$ ( $10^5$ )        | -1.38 [-1.41,-1.33] | -1.32 [-1.36,-1.29] | -1.28 [-1.32,-1.25] | <0.05            |
| $H3_{BISP}$ ( $10^8$ )        | -2.96 [-3.02,-2.86] | -2.87 [-2.93,-2.79] | -2.77 [-2.86,-2.71] | <0.05            |
| $BE1_{BISP}$                  | 8.51 [7.88,9.43]    | 8.17 [7.71,8.71]    | 8.60 [8.11,8.79]    | 0.34             |
| $BE2_{BISP}$                  | 6.08 [4.88,6.62]    | 5.26 [4.43,5.58]    | 6.70 [5.68,7.07]    | 0.13             |
| $PE_{BISP}$                   | 2.14 [2.08,2.15]    | 2.12 [2.05,2.14]    | 2.11 [2.08,2.13]    | 0.64             |
| $meanPa_{BISP}$ ( $10^{-2}$ ) | -1.60 [-3.19,0.81]  | 0.89 [-0.47,2.21]   | 0.92 [-1.96,3.20]   | 0.29             |
| $varPa_{BISP}$                | 0.38 [0.26,0.43]    | 0.34 [0.22,0.37]    | 0.33 [0.22,0.44]    | 0.87             |

\**p*-values obtained from the Kruskal-Wallis test after Bonferroni correction, OSA = obstructive sleep apnea.

ture optimization set, obtaining two optimum feature subsets ( $subset_{nobis}$  and  $subset_{bis}$ ), each one composed of those features selected in at least 50% of the 1000 bootstrap replicates (500 times) :

- $subset_{nobis}$  was composed of ODI3, the three anthropometric features (Age, Sex, and BMI), and five features from the PSD ( $M1_{BI1}$ ,  $M2_{BI1}$ ,  $MA_{BI1}$ ,  $RP_{BI1}$ , and  $Mb_{BI1}$ ).
- $subset_{bis}$  was composed of ODI3, two anthropometric features (Age, and Sex), five PSD-derived features ( $M1_{BI1}$ ,  $M2_{BI1}$ ,  $MA_{BI1}$ ,  $RP_{BI1}$ , and  $Mb_{BI1}$ ), and two bispectral features ( $M1_{BISP}$ , and  $meanPa_{BISP}$ ).

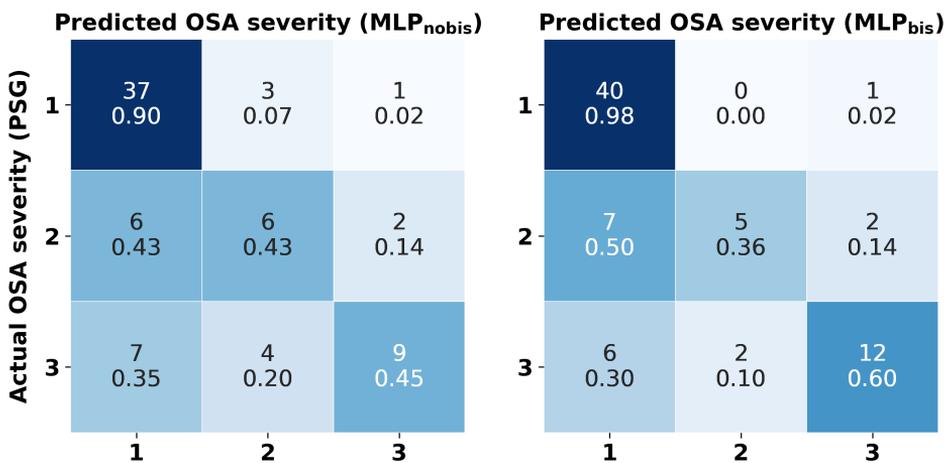
Taking as input these optimum subsets of features ( $subset_{nobis}$  and  $subset_{bis}$ ), two MLP networks were trained and optimized using the training set ( $MLP_{nobis}$  and  $MLP_{bis}$ ). Network hyperparameters, the regularization parameter ( $\alpha$ ) and the number of hidden units ( $N_H$ ), were varied from  $\alpha = 0$  up to  $\alpha = 5$  (step of 1) and from  $N_H = 2$  up to  $N_H = 50$  (step of 1), respectively. The maximum value of kappa, obtained using ten-fold cross-validation, determined the optimum values of  $\alpha$  and  $N_H$  in each case. Thereby,  $\alpha = 1$  and  $N_H = 3$  were chosen for  $MLP_{nobis}$

and  $\alpha = 2$  and  $N_H = 4$  were chosen for  $MLP_{bis}$ , as the pairs that reached the highest kappa. Then,  $MLP_{nobis}$  and  $MLP_{bis}$ , configured with the optimized hyperparameters, were fitted using the entire training set.

Figure 5.2 displays the confusion matrices of  $MLP_{nobis}$  and  $MLP_{bis}$ , evaluated in the test set. These matrices show the OSA severity group predicted by both  $MLP_{nobis}$  and  $MLP_{bis}$  versus the actual OSA severity group from PSG. Using  $MLP_{nobis}$ , 69.3% of the test patients (52/75) were correctly assigned to their actual group of OSA severity (sum of the main diagonal elements of the confusion matrix). Conversely,  $MLP_{bis}$  correctly assigned 76.0% (57/75) of the children to their OSA severity group. Table 5.2 shows diagnostic ability statistics of these models for the AHI severity cutoffs of 5 and 10 e/h. Notice that  $MLP_{bis}$  outperformed  $MLP_{nobis}$  in terms of Se, Sp, PPV, NPV, LR+, LR-, Acc, and kappa for both cutoffs.

### 5.1.2 Wavelet analysis

In [Vaquerizo-Villar et al. \(2018c\)](#), the utility of wavelet analysis to provide discriminative features from overnight oximetry associated to pediatric OSA was examined. To this effect, seven DWT-derived features were extracted for each oximetry signal of a database of 981 children with suspected OSA (i.e the complete UofC database). In order to assess complementarity with conventional ap-



**Figure 5.2.** Confusion matrices of  $MLP_{nobis}$  and  $MLP_{bis}$  in the test set. 1: negative OSA, 2: moderate OSA; 3: severe OSA. This figure has been derived from [Vaquerizo-Villar et al. \(2018b\)](#).

**Table 5.2.** Diagnostic ability of the MLP<sub>nobis</sub> and MLP<sub>bis</sub> models in the test set for the AHI cutoffs=5 e/h and 10 e/h. This table has been derived from [Vaquerizo-Villar et al. \(2018b\)](#).

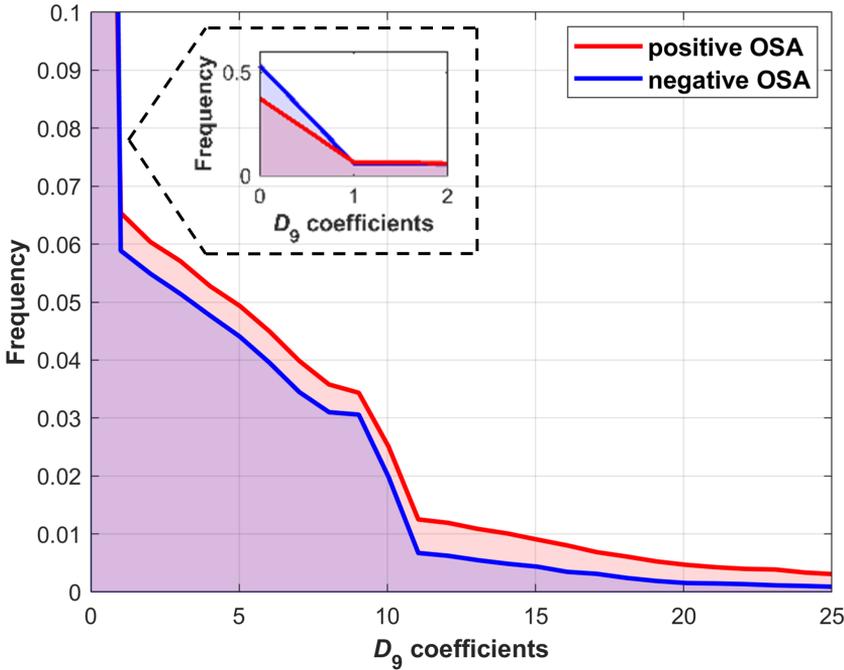
|                      | AHI cutoff | Se   | Sp   | PPV  | NPV  | LR+   | LR-  | Acc  | kappa |
|----------------------|------------|------|------|------|------|-------|------|------|-------|
| MLP <sub>nobis</sub> | 5 e/h      | 61.8 | 90.2 | 84.0 | 74.0 | 6.33  | 0.42 | 77.3 | 0.45  |
|                      | 10 e/h     | 45.0 | 94.5 | 75.0 | 82.5 | 8.25  | 0.58 | 81.3 |       |
| MLP <sub>bis</sub>   | 5 e/h      | 61.8 | 97.6 | 95.5 | 75.5 | 25.32 | 0.39 | 81.3 | 0.56  |
|                      | 10 e/h     | 60.0 | 94.5 | 80.0 | 86.7 | 11.00 | 0.42 | 85.3 |       |

AHI = apnea-hypopnea index, Se = sensitivity (%), Sp = specificity (%), PPV = positive predictive value (%), NPV = negative predictive value (%), LR+ = positive likelihood ratio, LR- = negative likelihood ratio, Acc = accuracy (%), kappa=Cohen's kappa index.

proaches, ODI3, statistical moments in the time domain, and PSD features were also extracted to obtain a wide initial feature set. FCBF was then employed to select an optimum feature subset. Finally, LR, SVM, and MLP binary classifiers were fitted with the optimum feature subset to detect moderate-to-severe pediatric OSA (AHI  $\geq 5$  e/h). In order to assure a proper validation of the proposed methodology, the database was divided into two sets: optimization set (589 subjects, 60%) and cross-validation set (392 subjects, 40%). The optimization group was used to: (i) perform descriptive analysis of the DWT-derived parameters; (ii) obtain an optimum subset of features with FCBF; (iii) optimize the hyperparameters of the SVM and MLP classifiers. Conversely, the cross-validation group was employed to assess the diagnostic performances of the extracted features and the LR, SVM, and MLP classifiers using stratified 5-fold cross-validation.

Figure 5.3 shows the histogram of the  $D_9$  coefficients for the negative OSA (AHI  $< 5$  e/h) and positive OSA (AHI  $\geq 5$  e/h) groups in the optimization set. According to this figure,  $D_9$  coefficients are more concentrated near zero in the negative OSA group, whereas in the positive OSA group the coefficients are more disperse. Table 5.3 shows the median and interquartile range of the DWT-derived features in the optimization set for negative OSA and positive OSA groups, along with their corresponding  $p$ -values, computed according to the Mann-Whitney  $U$  test. Noteworthy is the fact that all DWT-derived features showed statistically significant differences ( $p$ -value  $< 0.01$ ). The values of  $M1_{D_9}$ ,  $M2_{D_9}$ ,  $Max_{D_9}$ ,  $En_{D_9}$ , and  $WE$  were significantly higher in the positive OSA group, whereas  $M3_{D_9}$  and  $M4_{D_9}$  showed higher values in the negative OSA group.

Taking as input all the extracted features (ODI3, statistical moments, PSD, and DWT features), FCBF was applied to 1000 bootstrap replicates derived from the optimization set. In this way, ODI3, one statistical moment ( $M2_t$ ), three features from PSD ( $M2_{B11}$ ,  $M3_{B11}$ , and  $Max_{B11}$ ), and three DWT-derived features ( $M3_{D_9}$ ,



**Figure 5.3.** Histogram of the  $D_9$  coefficients for each group in the optimization set. This figure has been derived from [Vaquerizo-Villar et al. \(2018c\)](#).

**Table 5.3.** DWT-derived feature values for each group (median [interquartile range]) in the optimization set. This table has been derived from [Vaquerizo-Villar et al. \(2018c\)](#).

| Features           | negative OSA      | positive OSA      | $p$ -value |
|--------------------|-------------------|-------------------|------------|
| $M1_{D_9}$         | 3.04 [2.26, 3.92] | 5.36 [3.77, 7.70] | <0.01      |
| $M2_{D_9}$         | 3.78 [3.23, 4.63] | 5.73 [4.30, 7.57] | <0.01      |
| $M3_{D_9}$         | 1.31 [1.20, 1.44] | 1.19 [1.06, 1.32] | <0.01      |
| $M4_{D_9} (10^2)$  | 3.58 [1.03, 7.69] | 0.06 [0.04, 2.69] | <0.01      |
| $Max_{D_9} (10^1)$ | 1.23 [1.04, 1.55] | 1.96 [1.42, 2.62] | <0.01      |
| $En_{D_9} (10^3)$  | 0.54 [0.37, 0.89] | 1.54 [0.78, 2.96] | <0.01      |
| $WE (10^{-4})$     | 1.83 [1.18, 2.86] | 4.27 [2.52, 9.41] | <0.01      |

OSA = obstructive sleep apnea.

$En_{D_9}$ , and  $WE$ ), which were selected more than 500 times, formed the optimum feature subset. LR, SVM, and MLP binary classifiers were configured using this feature subset. Regarding SVM, we conducted trials varying the regularization parameter ( $C$ ) from  $C = 10^{-5}$  to  $C = 10^5$  (step of 1 in base-10 logarithmic scale), obtaining  $C = 10^3$  as the optimum value for which the accuracy was higher in the optimization set. With respect to MLP, the values of  $\alpha$  and  $N_H$  were from 0

up to 10 (step of 1) and from 2 up to 50 (step of 1), respectively, obtaining  $\alpha = 1$  and  $N_H = 5$  as the optimum  $N_H - \alpha$  pair that maximized the accuracy in the optimization set.

Table 5.4 displays the diagnostic ability of all the extracted features in the cross-validation set. Notice that 5 out of 7 DWT-derived features reached accuracies near 80%, being  $Max_{D_9}$  the DWT-derived feature that achieved the highest performance (81.7±5.6% Acc, with 75.4±7.1% Se and 85.4±6.8% Sp). In this respect,  $Max_{D_9}$  showed a similar Acc to ODI3 (81.9±7.2% Acc, with 78.1±7.3% Se and 84.2±8.1% Sp) and outperformed the remaining features. Table 5.5 shows the diagnostic performance metrics of LR, SVM, and MLP models, evaluated in the cross-validation set. Notice that these classifiers showed a high diagnostic ability for the diagnosis of moderate-to-severe OSA ( $AHI \geq 5$  e/h), outperforming all the extracted features individually. Specifically, the SVM binary classifier reached the highest diagnostic performance (84.0±5.2% Acc, with 71.9±4.4% Se and 91.1±7.2% Sp).

**Table 5.4.** Diagnostic ability of all the extracted features in the cross-validation set. This table has been derived from [Vaquerizo-Villar et al. \(2018c\)](#).

| Feature                  | Se        | Sp        | PPV       | NPV      | LR+     | LR-       | Acc      |
|--------------------------|-----------|-----------|-----------|----------|---------|-----------|----------|
| <i>ODI3</i>              | 78.1±7.3  | 84.2±8.1  | 75.2±10.2 | 86.5±5.0 | 6.1±2.9 | 0.27±0.11 | 81.9±7.2 |
| <i>M1<sub>T</sub></i>    | 62.3±6.8  | 65.0±2.6  | 51.4±2.1  | 74.6±3.6 | 1.8±0.2 | 0.58±0.10 | 64.0±2.3 |
| <i>M2<sub>T</sub></i>    | 72.6±13.6 | 67.1±6.6  | 56.7±2.8  | 81.2±6.6 | 2.2±0.3 | 0.40±0.17 | 69.2±3.1 |
| <i>M3<sub>T</sub></i>    | 65.0±8.5  | 61.4±6.8  | 50.1±2.8  | 74.9±2.8 | 1.7±0.2 | 0.57±0.09 | 62.7±2.7 |
| <i>M4<sub>T</sub></i>    | 60.9±15.6 | 49.9±8.4  | 41.6±5.0  | 69.0±7.5 | 1.2±0.3 | 0.78±0.26 | 54.0±5.2 |
| <i>M1<sub>BI1</sub></i>  | 75.3±7.9  | 82.5±7.4  | 73.0±8.5  | 85.1±3.5 | 5.3±3.1 | 0.30±0.08 | 79.9±3.8 |
| <i>M2<sub>BI1</sub></i>  | 69.8±7.3  | 83.4±5.2  | 71.8±6.2  | 82.5±3.0 | 4.5±1.4 | 0.36±0.08 | 78.3±3.2 |
| <i>M3<sub>BI1</sub></i>  | 47.2±11.7 | 58.1±11.9 | 40.4±4.1  | 65.0±2.8 | 1.2±0.2 | 0.91±0.12 | 54.1±4.5 |
| <i>M4<sub>BI1</sub></i>  | 63.6±8.3  | 47.1±6.2  | 41.7±4.2  | 68.7±6.1 | 1.2±0.2 | 0.79±0.23 | 53.3±5.0 |
| <i>Max<sub>BI1</sub></i> | 78.1±8.8  | 75.2±9.9  | 66.2±6.9  | 85.6±3.6 | 3.5±1.1 | 0.29±0.09 | 76.3±4.3 |
| <i>SE<sub>BI1</sub></i>  | 48.6±14.4 | 61.8±11.8 | 43.0±4.8  | 67.3±3.3 | 1.3±0.3 | 0.82±0.12 | 56.9±4.2 |
| <i>M1<sub>D9</sub></i>   | 73.4±9.1  | 82.6±7.8  | 72.2±10.2 | 84.0±5.1 | 5.2±2.7 | 0.32±0.12 | 79.1±6.2 |
| <i>M2<sub>D9</sub></i>   | 74.7±6.1  | 81.7±6.5  | 71.5±6.9  | 84.6±3.0 | 4.6±1.7 | 0.31±0.07 | 79.1±3.3 |
| <i>M3<sub>D9</sub></i>   | 58.3±9.2  | 63.4±6.5  | 48.7±3.1  | 72.1±3.3 | 1.6±0.2 | 0.66±0.10 | 61.5±3.2 |
| <i>M4<sub>D9</sub></i>   | 71.2±6.7  | 64.6±5.7  | 54.6±3.3  | 79.2±4.0 | 2.0±0.3 | 0.45±0.10 | 67.1±3.5 |
| <i>Max<sub>D9</sub></i>  | 75.4±7.1  | 85.4±6.8  | 76.0±9.0  | 85.4±4.3 | 6.2±2.8 | 0.29±0.10 | 81.7±5.6 |
| <i>En<sub>D9</sub></i>   | 78.8±4.4  | 81.7±5.2  | 72.2±5.5  | 86.7±2.4 | 4.6±1.4 | 0.26±0.05 | 80.6±3.4 |
| <i>WE</i>                | 76.0±8.2  | 78.4±5.6  | 68.0±3.8  | 84.9±3.5 | 3.6±0.7 | 0.30±0.09 | 77.6±2.5 |

Se = sensitivity (%), Sp = specificity (%), PPV = positive predictive value (%), NPV = negative predictive value (%), LR+ = positive likelihood ratio, LR- = negative likelihood ratio, Acc = accuracy (%).

**Table 5.5.** Diagnostic ability of the LR, SVM, and MLP classifiers in the cross-validation set. This table has been derived from [Vaquerizo-Villar et al. \(2018c\)](#).

| Feature | Se       | Sp       | PPV       | NPV      | LR+       | LR-       | Acc      |
|---------|----------|----------|-----------|----------|-----------|-----------|----------|
| LR      | 72.6±4.7 | 90.2±6.2 | 82.3±8.8  | 84.7±2.8 | 9.8±5.5   | 0.31±0.06 | 83.7±4.9 |
| SVM     | 71.9±4.4 | 91.1±7.2 | 83.8±10.8 | 84.5±2.6 | 14.6±12.9 | 0.31±0.06 | 84.0±5.2 |
| MLP     | 73.3±6.6 | 89.0±6.9 | 80.7±9.2  | 84.9±3.3 | 9.0±5.8   | 0.30±0.08 | 83.2±5.2 |

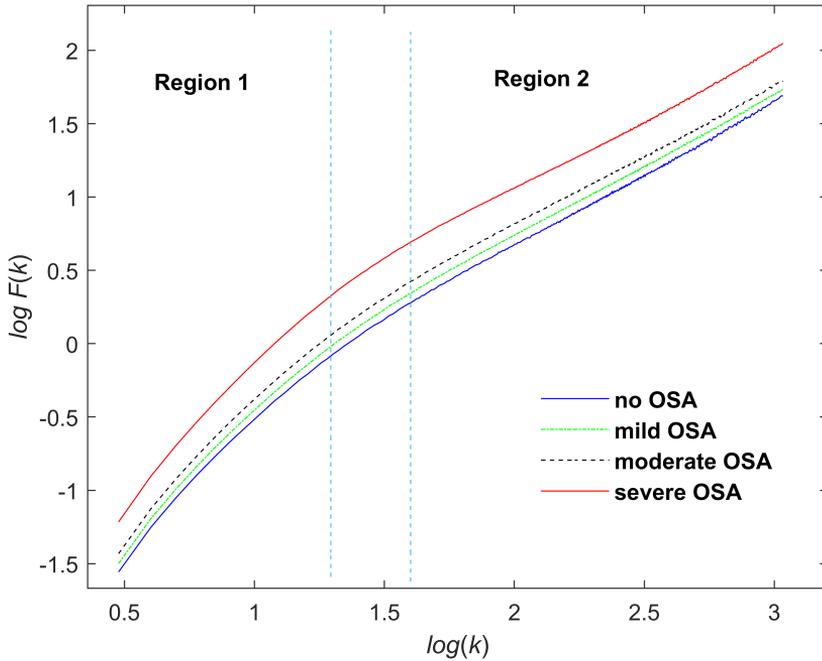
Se = sensitivity (%), Sp = specificity (%), PPV = positive predictive value (%), NPV = negative predictive value (%), LR+ = positive likelihood ratio, LR- = negative likelihood ratio, Acc = accuracy (%).

### 5.1.3 Detrended fluctuation analysis

The usefulness of DFA-derived features to supply further information from the oximetry signal linked to pediatric OSA was explored in [Vaquerizo-Villar et al. \(2018a\)](#), employing a conventional three-stage feature-engineering methodology. ODI3 and six DFA-derived parameters were first extracted from the SpO<sub>2</sub> recording of 981 children (i.e., the complete UofC database). A reduced subset of relevant and non-redundant features was then obtained using the FCBF algorithm. Finally, a MLP regression neural network was trained and optimized using this feature subset to estimate the AHI and hence pediatric OSA severity degrees. According to this feature-engineering methodology, the database was split into two sets: (i) training set (589 subjects, 60%), employed for optimization purposes; and (ii) test set (392 subjects, 40%), used to evaluate the diagnostic performance of the developed approach.

Figure 5.4 displays the averaged DFA profile in the training set for the four OSA severity groups: no OSA ( $AHI < 1$  e/h), mild OSA ( $1 \leq AHI < 5$  e/h), moderate OSA ( $5 \leq AHI < 10$  e/h), and severe OSA ( $AHI \geq 10$  e/h). In this figure, higher fluctuations are observed as the OSA severity increases. In addition, the two scaling regions (region 1 and region 2) are observed for the four OSA severity groups, as mentioned in Section 4.2.1.2. Table 5.6 shows the median and interquartile range values by OSA severity group of each extracted feature, as well as the  $p$ -value, obtained by means of the Kruskal-Wallis test. ODI3, and 4 out of 6 DFA-derived parameters ( $slope_1$ ,  $slope_{12}$ ,  $F(k_{12})$ , and  $F(k_x)$ ) showed statistically significant differences ( $p$ -value  $< 0.01$  after Bonferroni correction). Specifically, these features experienced an increasing trend as OSA severity increases.

FCBF was applied to 1000 bootstrap replicates obtained from all the extracted features (ODI3 and DFA-derived parameters) in the training group. The optimum subset was composed of ODI3 and  $slope_1$ , as these features were selected more than 500 times. Using this feature subset, the MLP network was trained and



**Figure 5.4.** Averaged DFA profile by OSA severity group in the training dataset. This figure has been derived from [Vaquerizo-Villar et al. \(2018a\)](#).

**Table 5.6.** Feature values for the OSA severity groups (median [interquartile range]) in the training set. This table has been derived from [Vaquerizo-Villar et al. \(2018a\)](#).

| Features     | no OSA             | mild OSA          | moderate OSA     | severe OSA         | <i>p</i> -value* |
|--------------|--------------------|-------------------|------------------|--------------------|------------------|
| ODI3         | 1.04 [0.52,2.47]   | 2.03 [0.93,3.89]  | 3.69 [1.94,7.23] | 12.35 [6.65,24.49] | <0.01            |
| $slope_1$    | 1.63 [1.58,1.68]   | 1.64 [1.58,1.70]  | 1.67 [1.60,1.71] | 1.74 [1.66,1.79]   | <0.01            |
| $slope_1$    | 0.96 [0.90,1.05]   | 0.95 [0.87,1.03]  | 0.92 [0.85,1.02] | 0.94 [0.88,1.01]   | 0.18             |
| $slope_{12}$ | 1.66 [1.53,1.82]   | 1.69 [1.55,1.87]  | 1.77 [1.60,1.94] | 1.82 [1.68,1.95]   | <0.01            |
| $k_{12}$     | 1.33 [1.23,1.42]   | 1.36 [1.26,1.44]  | 1.38 [1.29,1.45] | 1.34 [1.23,1.42]   | 0.04             |
| $F(k_{12})$  | 0.01 [-0.18,0.18]  | 0.12 [-0.12,0.26] | 0.22 [0.04,0.38] | 0.42 [0.16,0.61]   | <0.01            |
| $F(k_x)$     | -0.05 [-0.13,0.04] | 0.02 [-0.07,0.11] | 0.10 [0.00,0.20] | 0.31 [0.18,0.52]   | <0.01            |

\**p*-values obtained from the Kruskal-Wallis test after Bonferroni correction, OSA = obstructive sleep apnea.

optimized. Network hyper-parameters,  $\alpha$  and  $N_H$ , were varied from  $\alpha = 0$  up to  $\alpha = 10$  (step of 1) and from  $N_H = 2$  up to  $N_H = 30$  (step of 1), obtaining  $\alpha = 1$  and  $N_H = 5$  as the optimum  $N_H - \alpha$  pair that maximized kappa in the training set.

Figure 5.5 displays the Bland-Altman graphs of ODI3 and the AHI estimated by the MLP network ( $AHI_{MLP-DFA}$ ) compared to  $AHI_{PSG}$ . ICC is also shown.

The confidence interval was slightly lower in ODI3 than in  $AHI_{MLP-DFA}$  plot (23.3 versus 24.3). However,  $AHI_{MLP-DFA}$  reached a lower absolute mean error than ODI3 (0.75 vs -1.65), as well as a higher ICC (0.891 versus 0.866). Figure 5.6 shows the confusion matrices of ODI3 and  $AHI_{MLP-DFA}$  in the test group. Using ODI3, 55.4% of the children (217/392) were rightly assigned to their corresponding OSA severity group. By contrast, 60.0% of subjects (235/392) were rightly assigned by  $AHI_{MLP-DFA}$ . Table 5.7 shows diagnostic performance metrics of ODI3 and  $AHI_{MLP-DFA}$  for the AHI-based cutoffs of 1, 5, and 10 e/h. Notice that  $AHI_{MLP-DFA}$  outperformed ODI3 in terms of kappa, Acc for the severity cutoffs of 1 and 10 e/h, and AUC for the three cutoffs. In order to provide a more comprehensive comparison between ODI3 and  $AHI_{MLP-DFA}$ , ICC, kappa, overall Acc (four classes), and AUC values were obtained from 1000 bootstrap replicates and the  $p$ -value between ODI3 and  $AHI_{MLP-DFA}$  was calculated for each of these performance metrics by means of the Mann–Whitney  $U$  test. In this way, statistically significant higher values ( $p$ -value < 0.01) were obtained using  $AHI_{MLP-DFA}$  in the case of ICC, kappa, overall Acc, and AUC for the AHI-based cutoffs of 5 e/h and 10 e/h.

## 5.2 Application of deep-learning techniques

The previously mentioned results (see Sections 5.1.1, 5.1.2, and 5.1.3) were obtained following a feature-engineering methodology. As stated in Section 4.3, this approach may lead to the omission of OSA-related information from the oximetry signal. In [Vaquerizo-Villar et al. \(2021\)](#), we evaluate a novel methodology based on deep learning to automatically find the relevant information of the oximetry signal linked to pediatric OSA. This methodology consisted of two steps. First,

**Table 5.7.** Diagnostic ability of ODI3 and  $AHI_{MLP-DFA}$  in the test set for the AHI cutoffs= 1 e/h, 5 e/h, and 10 e/h. This table has been derived from [Vaquerizo-Villar et al. \(2018a\)](#).

|                 | AHI cutoff | Se   | Sp   | PPV  | NPV  | LR+  | LR-  | Acc  | AUC   | kappa |
|-----------------|------------|------|------|------|------|------|------|------|-------|-------|
| ODI3            | 1 e/h      | 83.5 | 50.6 | 87.4 | 42.9 | 1.7  | 0.33 | 77.0 | 0.811 | 0.355 |
|                 | 5 e/h      | 65.1 | 93.1 | 84.8 | 81.8 | 9.4  | 0.37 | 82.7 | 0.883 |       |
|                 | 10 e/h     | 65.1 | 96.1 | 81.8 | 91.1 | 16.7 | 0.36 | 89.5 | 0.921 |       |
| $AHI_{MLP-DFA}$ | 1 e/h      | 97.1 | 23.3 | 83.9 | 66.7 | 1.3  | 0.12 | 82.7 | 0.813 | 0.412 |
|                 | 5 e/h      | 78.8 | 83.7 | 74.2 | 86.9 | 4.8  | 0.25 | 81.9 | 0.888 |       |
|                 | 10 e/h     | 77.1 | 94.8 | 80.0 | 93.9 | 14.9 | 0.24 | 91.1 | 0.930 |       |

AHI = apnea-hypopnea index, 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 ROC curve, kappa=Cohen's kappa index.

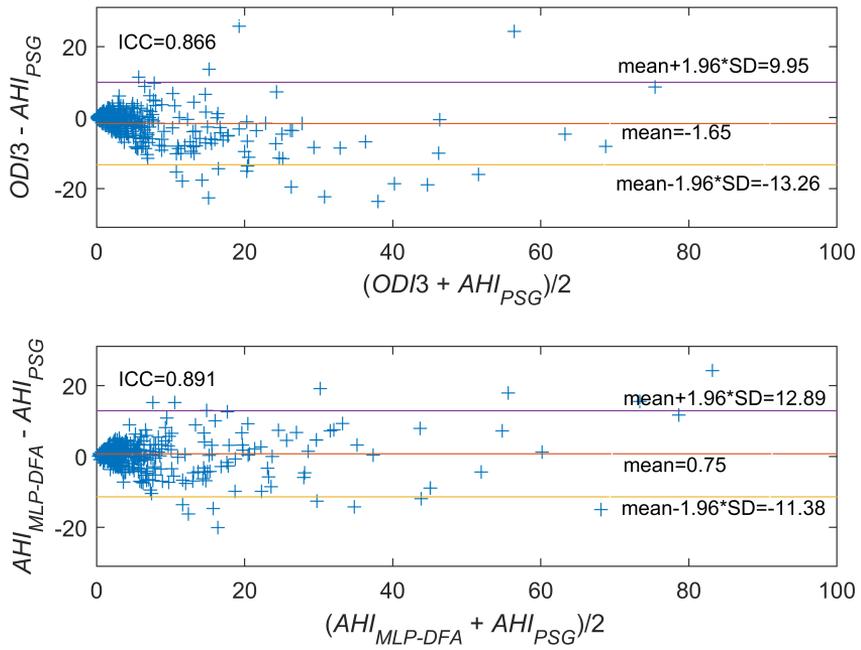


Figure 5.5. Bland-Altman plots comparing (a) ODI3 and (b)  $AHI_{MLP-DFA}$  with  $AHI_{PSG}$  in the test set. This figure has been derived from [Vaquerizo-Villar et al. \(2018a\)](#).

| Actual OSA severity (PSG) | Predicted OSA severity (ODI3) |             |            |            | Predicted OSA severity ( $AHI_{MLP-DFA}$ ) |             |            |            |
|---------------------------|-------------------------------|-------------|------------|------------|--|-------------|------------|------------|
|                           | 1                             | 2           | 3          | 4          | 1  | 2           | 3          | 4          |
| 1                         | 39<br>0.51                    | 36<br>0.47  | 1<br>0.01  | 1<br>0.01  | 18<br>0.23                                 | 55<br>0.71  | 3<br>0.04  | 1<br>0.01  |
| 2                         | 47<br>0.28                    | 107<br>0.63 | 12<br>0.07 | 3<br>0.02  | 8<br>0.05                                  | 125<br>0.74 | 33<br>0.20 | 3<br>0.02  |
| 3                         | 5<br>0.08                     | 33<br>0.52  | 17<br>0.27 | 8<br>0.13  | 1<br>0.02                                  | 22<br>0.51  | 8<br>0.19  | 12<br>0.28 |
| 4                         | 0<br>0.00                     | 13<br>0.16  | 16<br>0.19 | 54<br>0.65 | 0<br>0.00                                  | 8<br>0.10   | 11<br>0.13 | 64<br>0.77 |

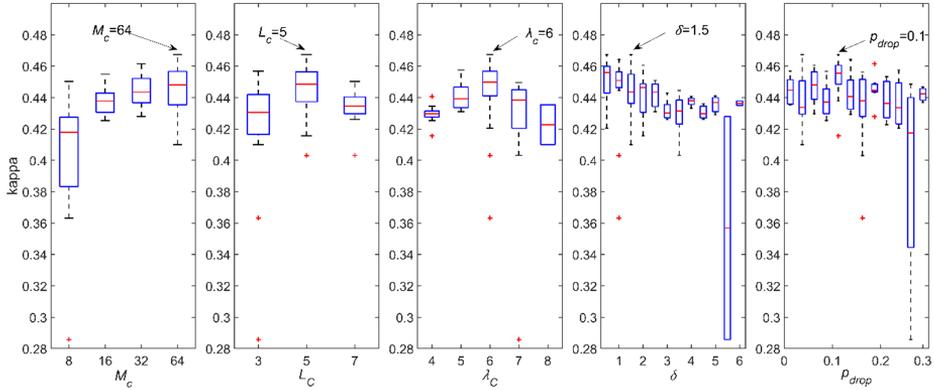
Figure 5.6. Confusion matrices of ODI3 and  $AHI_{MLP-DFA}$  in the test set. 1: no OSA; 2: mild OSA; 3: moderate OSA; 4: severe OSA. This figure has been derived from [Vaquerizo-Villar et al. \(2018a\)](#).

a CNN architecture was trained to estimate the number of apneic events in each 20-min non-overlapping SpO<sub>2</sub> segment. Then, the outputs of the CNN for each segment are aggregated to estimate the AHI for each subject using a database of 3196 SpO<sub>2</sub> recordings of children from three independent datasets: the CHAT dataset (see Section 3.1), the UofC dataset (see Section 3.2), and the BUH dataset (see Section 3.3).

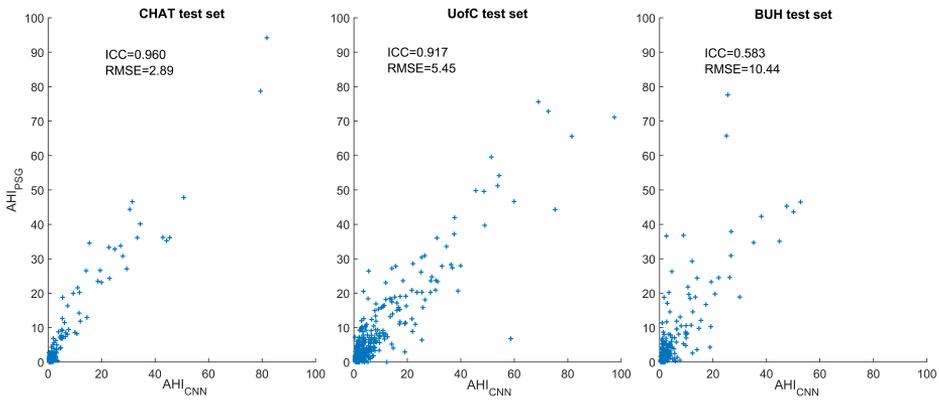
In order to assure a proper generalization of the proposed approach, the whole population under study was divided into three sets: (i) training set, used to train the CNN architecture; (ii) validation set, used for hyperparameter optimization; (iii) test set, used to assess the diagnostic performance of the deep-learning model. The training group was composed of 859 SpO<sub>2</sub> recordings from the CHAT dataset, as the UofC and BUH databases do not contain respiratory events annotations, which are needed in the CNN architecture to compose the output labels of each 20-min non-overlapping segment from the training set (see Section 4.3.1). The remaining subjects from the CHAT dataset, as well as the subjects of the UofC and BUH sets, were randomly divided into a validation set (1402 subjects, 60%) and a test set (40%), composed of 312 children from the CHAT dataset, 392 children from the UofC dataset, and 231 children from the BUH dataset.

Figure 5.7 shows the results of the BO-TPE algorithm for the hyperparameters of the CNN architecture: the number of filters ( $M_C$ ) and the filter size ( $L_C$ ) in each 1-D convolution, the number of CNN blocks ( $\lambda_C$ ), the delta parameter of the Huber loss function ( $\delta$ ), and the dropout probability ( $p_{drop}$ ). For each hyperparameter, the values of kappa are displayed in a boxplot. It can be observed that there is not a high confidence (interquartile range) of kappa on the values of the hyperparameter. Slightly higher overall kappa values are achieved with an increasing tendency of  $M_C$  and a decreasing trend of  $\delta$ , as well as when  $\lambda_C = 6$  and  $L_C = 5$ , whereas the value of  $p_{drop}$  had little effect on kappa. In this way,  $M_C = 64$ ,  $L_C = 5$ ,  $\lambda_C = 6$ ,  $\delta = 1.5$ , and  $p_{drop} = 0.1$  were obtained as the optimum values that maximized kappa in the validation set.

Figure 5.8 shows the scatter plots of the AHI estimated by the CNN model ( $AHI_{CNN}$ ) compared to  $AHI_{PSG}$  in the CHAT, UofC, and BUH test sets. ICC and RMSE between  $AHI_{CNN}$  and  $AHI_{PSG}$  are also shown. It can be observed that points of the scatter plot are more concentrated near the diagonal line in the CHAT test set, which results in a higher agreement (ICC=0.960 and RMSE=2.89) than in the UofC (ICC=0.917 and RMSE=5.45) and BUH test sets (ICC=0.583 and RMSE=10.44). Figure 5.9 displays the error distribution plots of  $AHI_{CNN}$  in the three test sets. A low mean error was obtained in the three test datasets. However,



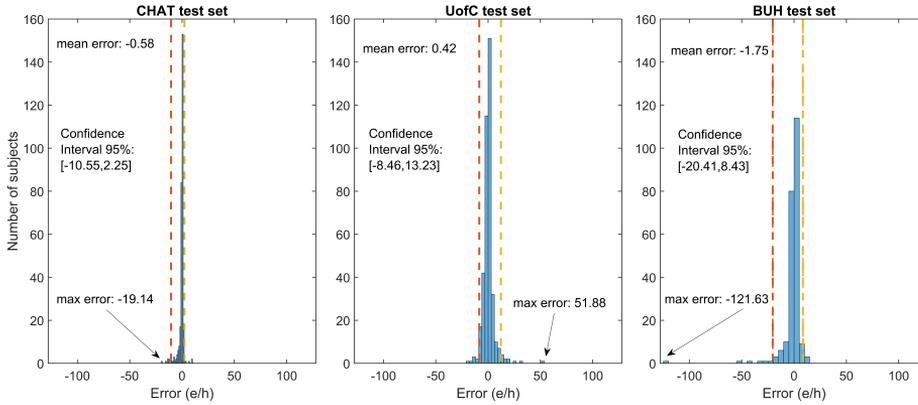
**Figure 5.7.** Results of the hyperparameter optimization in the validation set. This figure has been taken from [Vaquerizo-Villar et al. \(2021\)](#).



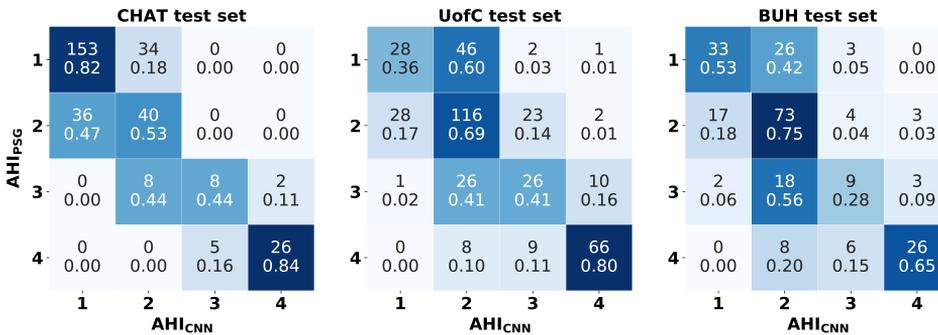
**Figure 5.8.** Scatter plot comparing  $AHI_{CNN}$  with  $AHI_{PSG}$  in the CHAT, UofC, and BUH test databases. This figure has been taken from [Vaquerizo-Villar et al. \(2021\)](#).

95% confidence interval of  $AHI_{CNN}$  was lower in the CHAT test set (12.80 e/h) than in the UofC (21.69 e/h) and BUH (28.84 e/h) test sets. In this respect, some outliers were observed in the UofC and BUH sets, as indicated by the maximum error.

Figure 5.10 shows the confusion matrices that face the pediatric OSA severity degrees established by the  $AHI_{PSG}$  and the corresponding assignation using  $AHI_{CNN}$  in the three test sets. Notice that a higher four-class overall accuracy was reached by  $AHI_{CNN}$  in the CHAT test set (72.8%, 227/312), than in the UofC (60.2%, 236/392) and BUH test sets (61.0%, 141/231), as anticipated by the scatter and error distribution plots. Table 5.8 shows the diagnostic performance statistics of  $AHI_{CNN}$  for each AHI threshold that establishes the pediatric OSA severity



**Figure 5.9.** Error distribution plot of  $AHI_{CNN}$  in the CHAT, UofC, and BUH test databases. This figure has been taken from [Vaquerizo-Villar et al. \(2021\)](#).



**Figure 5.10.** Confusion matrices of  $AHI_{CNN}$  in the CHAT, UofC, and BUH test datasets. 1: no OSA; 2: mild OSA; 3: moderate OSA; 4: severe OSA. This figure has been taken from [Vaquerizo-Villar et al. \(2021\)](#).

degrees (1 e/h, 5 e/h, and 10 e/h), which are derived from the confusion matrix. The value of kappa was remarkably higher in the CHAT test set (0.515) than in the UofC (0.422) and BUH test sets (0.423). A higher diagnostic ability is also observed in the CHAT test set for the AHI-based cutoffs of 5 and 10 e/h.

In order to provide a thorough comparison between our proposal and conventional approaches, we have compared the results of  $AHI_{CNN}$  with ODI3, a clinical parameter, as well as the AHI obtained with a classical feature-engineering approach based on MLP ( $AHI_{MLP}$ ) ([Vaquerizo-Villar et al., 2021](#)). Table 5.9 shows the comparison of the performance of  $AHI_{CNN}$  with ODI3 and  $AHI_{MLP}$  in the three test sets. Notice that  $AHI_{CNN}$  outperformed ODI3 and  $AHI_{MLP}$  in terms of overall accuracy, kappa, RMSE, and ICC in the three test sets.

**Table 5.8.** Diagnostic ability of  $AHI_{CNN}$  for the AHI cutoffs=1 e/h, 5 e/h, and 10 e/h in the CHAT, UofC, and BUH test databases. This table has been derived from [Vaquerizo-Villar et al. \(2021\)](#).

| Test set | AHI cutoff | Se   | Sp   | PPV  | NPV  | LR+    | LR-  | Acc  | kappa |
|----------|------------|------|------|------|------|--------|------|------|-------|
| CHAT     | 1 e/h      | 71.2 | 81.8 | 72.4 | 81.0 | 3.92   | 0.35 | 77.6 | 0.515 |
|          | 5 e/h      | 83.7 | 100  | 100  | 97.0 | N.D.   | 0.16 | 97.4 |       |
|          | 10 e/h     | 83.9 | 99.3 | 92.9 | 98.2 | 117.84 | 0.16 | 97.8 |       |
| UofC     | 1 e/h      | 90.8 | 36.4 | 85.4 | 49.1 | 1.43   | 0.25 | 80.1 | 0.421 |
|          | 5 e/h      | 76.0 | 88.6 | 79.8 | 86.2 | 6.68   | 0.27 | 83.9 |       |
|          | 10 e/h     | 79.5 | 95.8 | 83.5 | 94.6 | 18.90  | 0.21 | 92.3 |       |
| BUH      | 1 e/h      | 88.8 | 53.2 | 83.8 | 63.5 | 1.90   | 0.21 | 79.2 | 0.423 |
|          | 5 e/h      | 61.1 | 93.7 | 81.5 | 84.2 | 9.72   | 0.42 | 83.5 |       |
|          | 10 e/h     | 65.0 | 96.9 | 81.3 | 93.0 | 20.69  | 0.36 | 91.3 |       |

AHI = apnea-hypopnea index, Se = sensitivity (%), Sp = specificity (%), PPV = positive predictive value (%), NPV = negative predictive value (%), LR+ = positive likelihood ratio, LR- = negative likelihood ratio, Acc = accuracy (%), kappa = Cohen's kappa index, N.D. = not defined

**Table 5.9.** Diagnostic performance of  $AHI_{CNN}$  vs.  $ODI3$  and  $AHI_{MLP}$  in the CHAT, UofC, and BUH test databases. This table has been derived from [Vaquerizo-Villar et al. \(2021\)](#).

| Test set | Method      | ICC   | RMSE  | 4-class kappa | 4-class Acc |
|----------|-------------|-------|-------|---------------|-------------|
| CHAT     | $AHI_{CNN}$ | 0.960 | 2.89  | 0.515         | 72.8        |
|          | $ODI3$      | 0.871 | 4.63  | 0.417         | 65.1        |
|          | $AHI_{MLP}$ | 0.832 | 5.51  | 0.377         | 63.3        |
| UofC     | $AHI_{CNN}$ | 0.917 | 5.45  | 0.422         | 60.2        |
|          | $ODI3$      | 0.861 | 6.21  | 0.372         | 56.6        |
|          | $AHI_{MLP}$ | 0.890 | 6.02  | 0.381         | 56.9        |
| BUH      | $AHI_{CNN}$ | 0.583 | 10.44 | 0.423         | 61.0        |
|          | $ODI3$      | 0.520 | 10.64 | 0.369         | 57.6        |
|          | $AHI_{MLP}$ | 0.500 | 11.05 | 0.306         | 52.4        |

ICC = intra-class correlation coefficient, RMSE = root mean squared error, kappa = Cohen's kappa index, Acc = accuracy (%)

In this chapter, the most relevant results obtained during this Doctoral Thesis have been presented. In the next Chapter (see Section 6), these results will be discussed, as well as compared with state-of-the-art studies.



# Chapter 6

## Discussion

This Doctoral Thesis addresses the simplification of pediatric OSA diagnosis. For this purpose, novel signal processing algorithms have been applied to improve the diagnostic ability of the oximetry signal. In this regard, feature-engineering and deep-learning methodologies have been approached. On the one hand, we have gone further in SpO<sub>2</sub> characterization in the time and frequency domain using bispectrum, wavelet, and DFA, obtaining new features that provide additional information from the oximetry signal related to pediatric OSA and its severity. On the other hand, a deep-learning model based on CNNs was able to automatically learn discriminative features from raw SpO<sub>2</sub> data linked to apneic events, outperforming conventional approaches. In this chapter, the main outcomes obtained during this Thesis are discussed. Moreover, a comparison of the proposed methodologies in terms of diagnostic performance is provided, as well as a comparison with state-of-the-art works. Finally, the main limitations of this Thesis are stated.

### **6.1 Novel features to provide relevant and complementary information from oximetry recordings**

As aforementioned, bispectrum ([Vaquerizo-Villar et al., 2018b](#)), wavelet analysis ([Vaquerizo-Villar et al., 2018c](#)), and DFA ([Vaquerizo-Villar et al., 2018a](#)) were applied to identify features able to provide additional information regarding OSA-related changes in the oximetry dynamics.

### 6.1.1 Bispectral analysis

Spectral analysis has been widely used to analyze SpO<sub>2</sub> in both adult and pediatric OSA contexts (Álvarez et al., 2017; Alvarez et al., 2013; Crespo et al., 2018; Garde et al., 2014a; Hornero et al., 2017), as it reflects the changes in the SpO<sub>2</sub> spectrum elicited by the recurrence of respiratory events while sleeping. This analysis has been commonly accomplished using PSD (Álvarez et al., 2017; Crespo et al., 2018; Garde et al., 2014a; Hornero et al., 2017). However, PSD cannot characterize changes of linearity and gaussianity in a signal, as the phase relationship among spectral components is lost (Chua et al., 2010). In contrast, bispectral analysis preserves both amplitude and phase information of the Fourier transform, which enables the detection of phase relationships and deviations from gaussianity and linearity of a signal (Chua et al., 2010), such as those that may be elicited in SpO<sub>2</sub> recordings by physiological perturbations of OSA.

In the present Doctoral Thesis, bispectrum has been applied as a complementary tool to conventional spectral analysis (Vaquerizo-Villar et al., 2018b). To our knowledge, this is the first time that bispectral analysis is used in the framework of pediatric OSA. A MLP model fed with a feature subset composed of bispectrum-derived parameters, together with anthropometric variables, ODI3, and PSD-derived parameters (MLP<sub>bis</sub>) reached a high diagnostic performance, with a 3-class Acc of 76%, a kappa value of 0.56, and 81.3% Acc and 85.3% Acc for the AHI cutoffs of 5 e/h and 10 e/h, respectively. These results outperformed a MLP neural network trained without information from bispectrum (MLP<sub>nobis</sub>). It is worthy to note that MLP<sub>bis</sub> reached a PPV of 95.5% for 5 e/h and a NPV of 86.7% for 10 e/h. As stated in Section 3, these cutoffs are commonly use in the clinical practice to detect moderate ( $5 \leq \text{AHI} < 10$  e/h) and severe OSA ( $\text{AHI} \geq 10$  e/h). In this respect, adenotonsillectomy treatment is recommended in children with an  $\text{AHI} \geq 5$  e/h, as they have an increased chance of suffering adverse health consequences and comorbidities (Church, 2012; Hunter et al., 2016; Kaditis et al., 2016a). Furthermore, children with an  $\text{AHI} \geq 10$  e/h can present persistent risk factors and residual OSA after treatment (Alonso-Álvarez et al., 2011).

It is also important to highlight that two bispectral-derived features,  $M1_{BISP}$  and  $meanPa_{BISP}$ , were involved in the optimum subset.  $M1_{BISP}$  contains information about changes in the amplitude of the bispectrum related to deviations of gaussianity in the SpO<sub>2</sub> signal, whereas  $meanPa_{BISP}$  measures changes in the bispectral phase associated to a phase coupling between spectral components of the SpO<sub>2</sub> signal related to pediatric OSA severity. Furthermore, bispectral moments

( $H1_{BISP}$ ,  $H2_{BISP}$ , and  $H3_{BISP}$ ), which measure nonlinear relationships between the frequency components of the oximetry signal, showed significantly higher values as the severity of OSA increased.

These findings agree with previous works that also examined the usefulness of bispectrum to characterize OSA-related changes in adults (Atri and Mohebbi, 2015; Tagluk and Sezgin, 2011) and children (Barroso-García et al., 2021). Tagluk and Sezgin (2011) reported changes in the quadratic phase coupling of the EEG signal, whereas Atri and Mohebbi (2015) reported changes in the non-Gaussian and nonlinear dynamical information of heart rate variability and ECG-derived respiratory signals during apneic episodes by means of bispectrum. Finally, in a recent work developed by Barroso-García et al. (2021), it was found that bispectrum provides information regarding changes in the gaussianity, linearity, and regularity of the AF signal elicited by apneic events. In the present research, it has been demonstrated that bispectral analysis can identify phase relationships and deviations from linearity and gaussianity of the SpO<sub>2</sub> signal that provide additional and complementary information to conventional approaches in the automated detection of childhood OSA. This study was the starting point for the second study of the Doctoral Thesis, in which we apply the wavelet transform to further characterize the SpO<sub>2</sub> signal in the frequency domain.

### 6.1.2 Wavelet analysis

PSD and bispectrum are frequency domain analysis techniques based on the STFT (Chua et al., 2010). STFT uses a fixed length window to analyze each segment of the signal, assuming that it is stationary (Rioul and Vetterli, 1991). Nonetheless, non-stationary changes occur in the oximetry signal during sleep, mainly due to apneic events (Berry et al., 2012). This limitation is overcome by the WT, which does not make assumptions about the stationarity of the signal (Rioul and Vetterli, 1991). Wavelet analysis provides an optimal time-frequency resolution (high frequency resolution at low frequencies and high temporal resolution at high frequencies) (Rioul and Vetterli, 1991), which is useful to analyze OSA-related information at the low frequency components of the SpO<sub>2</sub> recordings.

In this research, wavelet analysis has been employed to further characterize the oximetry dynamics related to the presence of moderate-to-severe pediatric OSA (Vaquerizo-Villar et al., 2018c). Previous studies have shown the usefulness of the WT to characterize the changes in physiological signals related to ap-

neic episodes in adult patients (Fontenla-Romero et al., 2005; Khandoker et al., 2008; Lee et al., 2004; Lin et al., 2006; Mendez et al., 2010; Tagluk and Sezgin, 2011), though it had not been previously applied in the screening of childhood OSA. Our results revealed that all the DWT-derived features,  $WE$  and features computed from the coefficients in  $D_9$  ( $M1_{D_9}$ ,  $M2_{D_9}$ ,  $M3_{D_9}$ ,  $M4_{D_9}$ ,  $Max_{D_9}$ , and  $En_{D_9}$ ), showed statistically significant differences between negative OSA ( $AHI < 5$  e/h) and positive OSA ( $AHI \geq 5$  e/h) groups. Furthermore, these features reached a overall higher performance than statistical moments in the time domain and PSD-derived parameters, thus suggesting that DWT is a suitable tool to identify OSA-related changes occurring in the oximetry signal. Finally, MLP, LR, and SVM binary classifiers fed with an optimum subset composed of features from these complementary approaches (DWT, statistical moments, ODI3, and PSD) reached a high diagnostic performance, improving the diagnostic ability of all the extracted features. Noteworthy, the SVM model reached the highest Acc (84.0%), Sp (91.1%), PPV (83.8%), and LR+ (14.6%) among the individual features and binary classification algorithms. A high LR+ ( $LR > 10$ ) is considered to present solid evidence to determine the presence of a disease, which indicates that this model is especially useful as a screening method to confirm the presence of moderate-to-severe pediatric OSA ( $AHI \geq 5$  e/h). Accordingly, our DWT-based SVM model could be used to automatically detect moderate-to-severe pediatric OSA at patient's home, thus reducing associated healthcare costs and intrusiveness of overnight PSG.

Importantly, three DWT-derived parameters,  $M3_{D_9}$ ,  $En_{D_9}$ , and  $WE$ , were automatically selected with FCBF. As shown,  $M3_{D_9}$  was significantly lower in the positive-OSA group, which indicates that apneas and hypopneas modify the frequency distribution of  $SpO_2$  signal and increase its frequency components in the  $D_9$  band (0.0244–0.0488 Hz), thus resulting in values less proximal to zero. Regarding  $En_{D_9}$ , it was higher in the positive-OSA group, which agrees with a higher occurrence of respiratory events that increase the amplitude of the  $D_9$  coefficients. These changes of the  $SpO_2$  signal in the  $D_9$  band are linked to the recurrence and duration of the oxygen desaturation associated to apneic episodes. In addition,  $WE$  revealed a higher irregularity in the positive-OSA group, which indicates that pediatric OSA disturbs the energy distribution of the DWT decomposition of the  $SpO_2$  signal. According to our results, the information about the occurrence of apneic events provided by DWT through the amplitude ( $En_{D_9}$ ) and the concentration of the  $D_9$  coefficients near zero ( $M3_{D_9}$ ), and the irregularity of the energy distribution ( $WE$ ) of the  $SpO_2$  is complementary (non-redundant) to

the information provided by conventional approaches. Hence, this study confirms that, the great resolution provided by DWT in the low frequency range, as well as its suitability to analyze non-stationary signals, make DWT an appropriate tool to further characterize the changes occurring in the oximetry signal associated with pediatric OSA. These findings, together with those obtained in the Section 6.1.1, led us to the third study of this Thesis, in which we apply the detrended fluctuation analysis method to gain insight into the nonlinear and non-stationary properties of the oximetry signal in the time domain.

### 6.1.3 Detrended fluctuation analysis

As aforementioned, biomedical signals typically present non-stationarities and non-linearities, since biological systems have a stochastic behavior. In this respect, nonlinear methods derived from the chaos theory have proved high capability to characterize changes in SpO<sub>2</sub> dynamics related to physiological perturbations of OSA, both in adult and in pediatric patients (del Campo et al., 2018). Nonetheless, Hornero et al. (2017) and Garde et al. (2014a) reported that conventional nonlinear metrics (SampEn, CTM, and LZC) were redundant with regard to common statistical moments, conventional oximetric indices, and frequency domain features. In order to provide further insights into its nonlinear properties, we have applied DFA to characterize changes in the scaling behavior (i.e., irregular fluctuations and random spikes) of the oximetry signal related to pediatric OSA and its severity (Vaquerizo-Villar et al., 2018a). Previous studies have assessed the capability of DFA to characterize OSA in adults (Hua and Yu, 2017; Kaimakamis et al., 2016; Lee et al., 2002; Penzel et al., 2003) and children (Dehkordi et al., 2016). Nonetheless, no studies have applied DFA to analyze SpO<sub>2</sub> recordings in the context of childhood OSA.

Our results revealed that the scaling behavior of the SpO<sub>2</sub> recordings is affected by pediatric OSA. This agrees with Dehkordi et al. (2016) and Penzel et al. (2003), who also obtained two scaling regions with different correlation properties in adult OSA patients, one region for short-time scales related to respiratory events and another region for long-time scales associated to the effects of circadian rhythm and sleep stages. As shown,  $slope_1$  and  $slope_{12}$  showed significantly higher values as the AHI increased. This can be explained by the higher occurrence of respiratory events that changes the oximetry dynamics in the short-time scales. In addition,  $F(k_{12})$  and  $F(k_x)$  showed significantly higher values as the OSA severity increased. This fact is consistent with the higher fluctuations ob-

served in the DFA profile as AHI increased, which indicates that apneic events cause irregular fluctuations in the oximetry signal, as also reported by [Hua and Yu \(2017\)](#).

The information provided by DFA was also complimentary to ODI3. Specifically, FCBF automatically selected  $slope_1$ , together with ODI3. As shown, a MLP neural network trained to estimate the AHI using this optimum subset reached an ICC of 0.891, a 4-class Acc of 60%, a kappa value of 0.41, and 82.7%, 81.9%, and 91.1% Acc for the AHI cutoffs of 1 e/h, 5 e/h, and 10 e/h, respectively. This MLP model showed an overall higher diagnostic performance than the conventional ODI3. These results suggest that the changes in the scaling behavior of the DFA profile quantified by  $slope_1$  provide relevant and additional information that contributes to improve the diagnostic ability of the SpO<sub>2</sub> signal in the framework of pediatric OSA.

## 6.2 A deep-learning based methodology to automatically extract the relevant information from raw oximetry recordings

The feature-engineering approach has shown its usefulness to characterize pediatric OSA and its severity ([Vaquerizo-Villar et al., 2018a,b,c](#)). It has been demonstrated that the application of novel signal processing algorithms from different analytical approaches provide relevant features that parameterize OSA-related oximetric changes. Furthermore, this feature-engineering approach has allowed us to identify which features provide additional information to classical methods regarding oximetric changes related to childhood OSA and its severity ([Vaquerizo-Villar et al., 2018a,b,c](#)). Nonetheless, the feature-engineering approach is limited to the existing human knowledge, which may lead to the omission of relevant information concerning pediatric OSA that is still undiscovered ([Dehlink and Tan, 2016](#)).

The aforementioned limitation of the feature-engineering methodology is overcome by the deep-learning approach, which are based on an automatic identification of the important information that is not controlled by human experts ([LeCun et al., 2015](#)). In [Vaquerizo-Villar et al. \(2021\)](#), we have proven that a CNN-based deep-learning method can automatically learn discriminative information from the raw oximetry data linked to apneic events. These findings are consistent with recent studies that also demonstrated the utility of deep-learning ap-

proaches to automatically identify OSA-related changes in physiological signals from adult subjects (Biswal et al., 2018; Choi et al., 2018; Nikkonen et al., 2019; Van Steenkiste et al., 2018). These studies analyzed raw signals from PSG using deep-learning models based on recurrent neural networks (RNN) (Biswal et al., 2018; Van Steenkiste et al., 2018), MLP (Nikkonen et al., 2019), and CNN (Choi et al., 2018). In this regard, CNN has a lower computational cost than RNN and MLP, which makes it more suitable for screening purposes using wearable and portable pulse oximetry devices.

Table 6.1 shows a summary of the comparison of the estimated AHI from the proposed CNN architecture ( $AHI_{CNN}$ ) with ODI3, a clinical approach, as well as the AHI estimated by a classical feature-engineering approach ( $AHI_{MLP}$ ). As expected,  $AHI_{CNN}$  outperformed ODI3 and  $AHI_{MLP}$ , showing a high diagnostic ability in a large sample of 3196  $SpO_2$  signals from three independent datasets. Specifically,  $AHI_{CNN}$  reached a high agreement in the CHAT (ICC=0.960), UofC (ICC=0.917), and BUH (ICC=0.583) test sets. In addition, high 4-class accuracies (72.8%, 60.2%, and 61.0%), high kappa values (0.515, 0.422, and 0.423), and high accuracies for the AHI severity cutoffs of 1 e/h (77.6%, 80.1%, and 79.2%), 5 e/h (97.4%, 83.9%, and 83.5%), and 10 e/h (97.8%, 92.3%, and 91.3%) were obtained in the CHAT, UofC, and BUH test sets, respectively.

Figure 6.1 displays a possible screening protocol that shows the clinical applicability of  $AHI_{CNN}$ . This screening protocol, which is derived from the confusion matrices of  $AHI_{CNN}$  (see Figure 5.10), would act as follows: (i) If  $AHI_{CNN} < 1$  e/h (no OSA), clinicians could discard OSA as 96.2% (BUH), 98.2% (UofC), and 100% (CHAT) of these subjects will have an  $AHI_{CNN} < 5$  e/h. These patients might be eventually referred to PSG on the persistence of symptoms (Alonso-Álvarez

**Table 6.1.** Summary of the comparison of  $AHI_{CNN}$  with ODI3 and  $AHI_{MLP}$  in the CHAT, UofC, and BUH test databases.

| Test set | Method      | kappa |
|----------|-------------|-------|
| CHAT     | $AHI_{CNN}$ | 0.515 |
|          | ODI3        | 0.417 |
|          | $AHI_{MLP}$ | 0.377 |
| UofC     | $AHI_{CNN}$ | 0.422 |
|          | ODI3        | 0.372 |
|          | $AHI_{MLP}$ | 0.381 |
| BUH      | $AHI_{CNN}$ | 0.423 |
|          | ODI3        | 0.369 |
|          | $AHI_{MLP}$ | 0.306 |

kappa = Cohen's kappa index.

et al., 2011); (ii) If  $1 \leq \text{AHI}_{\text{CNN}} < 5$  e/h, the clinicians could suggest to conduct overnight PSG as doubts exist about the true diagnosis of these subjects; (iii) If  $5 \leq \text{AHI}_{\text{CNN}} < 10$  e/h, the clinicians could consider treatment as 86.4% (BUH), 96.7% (UofC), and 100% (CHAT) of these children are at least mild OSA ( $\text{AHI}_{\text{CNN}} \geq 1$  e/h) and they were referred to the sleep laboratory showing symptoms; (iv) If  $\text{AHI}_{\text{CNN}} \geq 10$  e/h, the clinicians could suggest treatment, as most probably (90.6% in BUH, 96.2% in UofC, and 100% in CHAT) these children are at least moderate OSA ( $\text{AHI}_{\text{CNN}} \geq 5$  e/h), and also consider a follow-up of these children, as they are prone to present persistent risk factors and residual OSA after being treated. The implementation of this screening protocol in a pediatric sleep unit could lead to a 45.9% (BUH), 50.0% (UofC), and 73.7% (CHAT) reduction in full PSGs, thus reducing health costs and waiting lists. In addition, these children would benefit from a more comfortable diagnostic test that could be easily performed at home.

Despite the fact that the CNN model outperformed conventional methods in the three datasets, it is noteworthy that our proposal reached a higher performance in the CHAT dataset than in the UofC and BUH datasets. This is consistent with the fact that the weights of the CNN model were obtained using only the CHAT dataset. As there is a large variability in the annotation of cardiorespiratory events and sleep stages among different sleep technologists (Collop, 2002), this variance may influence the external assessment of our proposed CNN model in two external datasets. Nonetheless, we tried to reduce this variability by using a validation group formed by children from the CHAT, UofC, and BUH datasets. There are also differences in the following clinical characteristics that could help to explain the varying diagnostic performance among datasets:

- AHI distribution. The median [interquartile range] of the AHI values vary among datasets: 0.8 [0.4-1.7] in the CHAT test set, 3.3 [1.4-7.8] in the UofC test set, and 2.3 [0.9-6.4] in the BUH test set.
- Age. There are also differences in the age range among datasets: 5-10 in the CHAT dataset, 0-13 in the UofC dataset, and 0-18 in the BUH dataset.
- Sampling rate. The sampling rate of the  $\text{SpO}_2$  recordings were: 1, 2, 10, 12, 16, 200, 256, and 512 Hz in the CHAT dataset, 25, 200, and 500 Hz in the UofC dataset, and 200 Hz in the BUH dataset.
- Population group. CHAT and UofC datasets are composed of pediatric subjects from the United States of America (USA), whereas BUH dataset is com-

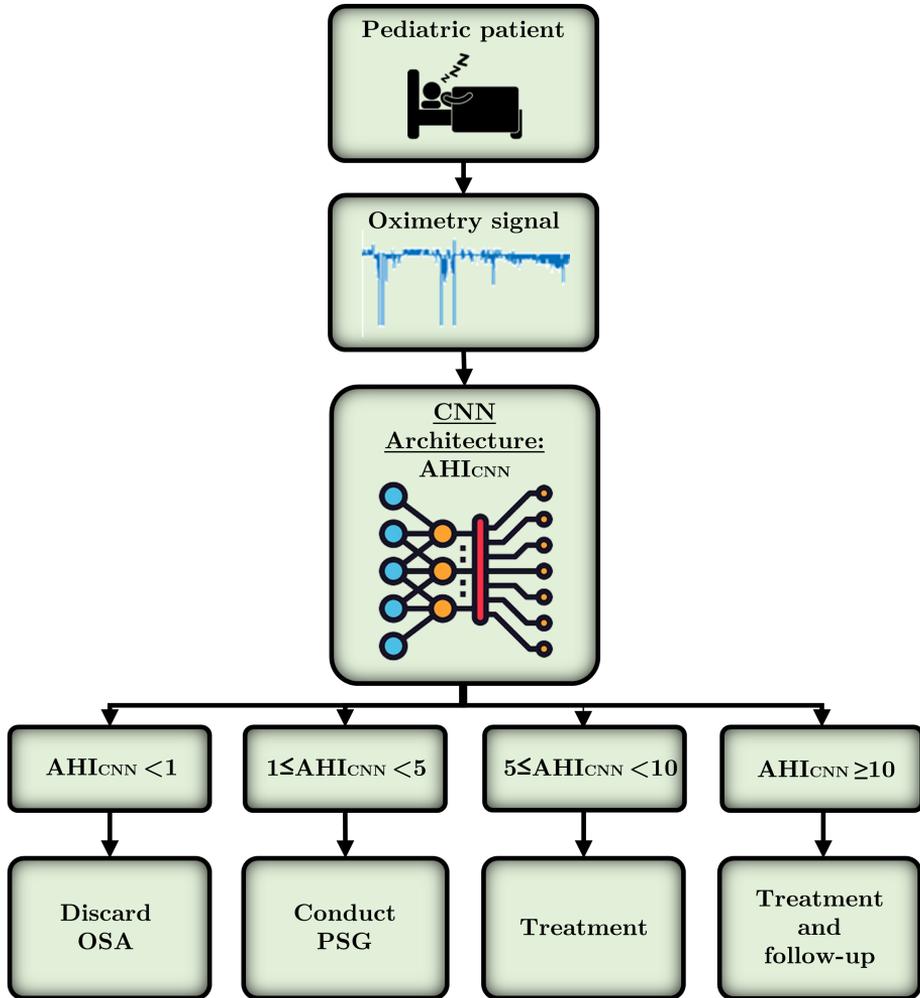


Figure 6.1. Screening protocol of the proposed CNN-based deep-learning model

posed of children from Spain. In this respect, obesity prevalence and race are different in these countries, as well as the health system, which is mainly public in Spain and mostly private in the USA. This influences the socio-economic condition of the subjects under study, thus having a substantial consequence on the health status.

These findings are consistent with previous works (Alvarez et al., 2013; Biswal et al., 2018; Nikkonen et al., 2019), which also reported a varying performance among sleep databases with different clinical features.

To sum up, it was found that deep-learning allows to automatically iden-

tify discriminative information from oximetry dynamics associated to apneic episodes. In addition, the proposed CNN-based deep-learning model showed a high diagnostic ability, outperforming a classical clinical parameter, ODI3, as well as a conventional feature-engineering approach based on MLP ( $AHI_{MLP}$ ). The applicability of our results was also highlighted by the validation of our proposal in 3196  $SpO_2$  recordings from three different datasets, as well as by the design of a screening protocol. These findings suggest that the use of automated methodologies based on deep learning contributes to further improving the diagnostic ability of overnight oximetry in the screening of childhood OSA.

### 6.3 Comparison of performance: feature-engineering, deep-learning, and state-of-the-art

In order to further discuss the most relevant findings of this Thesis, we have compared the diagnostic performance obtained with the different feature-engineering and deep-learning methodologies proposed in the papers included in the compendium of publications (Section 6.3.1). In addition, we have compared these results with those achieved in the state-of-the-art, considering the methodological differences among studies that limit generalization. (Section 6.3.2).

#### 6.3.1 Comparison between feature-engineering and deep-learning approaches

Table 6.2 displays the overall diagnostic performance of the automated feature-engineering and deep-learning models obtained in the compendium of publications. Notice that a high performance was achieved in all the publications, with accuracies ranging between 77.6%-82.7% Acc for an AHI cutoff of 1 e/h, 81.3%-97.4% Acc using an AHI cutoff of 5 e/h, and 85.3%-97.8% using an AHI cutoff of 10 e/h. It can also be observed that, in [Vaquerizo-Villar et al. \(2018b\)](#), the performance was lower than in the remaining studies. Nonetheless, an initial version of the UofC database (298 children) was used in this study, so that a thorough comparison of the results cannot be performed. Conversely, the test set of the UofC dataset employed in [Vaquerizo-Villar et al. \(2018c\)](#), [Vaquerizo-Villar et al. \(2018a\)](#), and [Vaquerizo-Villar et al. \(2021\)](#) was composed of the same subjects (392 pediatric patients). This enables a direct comparison of their results.

Regarding this comparison, it is important to note that a high diagnostic performance was obtained in [Vaquerizo-Villar et al. \(2021\)](#) in the UofC dataset.

**Table 6.2.** Summary of the diagnostic performance of the methods developed during the compendium of publications

| Study                           | Dataset | N<br>(Total/test) | AHI | Methods<br>(Feature/classification)   | Se   | Sp   | Acc  | kappa | ICC   |
|---------------------------------|---------|-------------------|-----|---|------|------|------|-------|-------|
| Vaquerizo-Villar et al. (2018b) | UofC    | 298/75            | 5   | Bispectrum, PSD features, ODI3, and anthropometric variables / Multiclass MLP | 61.8 | 97.6 | 81.3 | 0.56  | -     |
|                                 |         |                   | 10  |   | 60.0 | 94.5 | 85.3 |       |       |
|                                 |         |                   |     |   |      |      |      |       |       |
| Vaquerizo-Villar et al. (2018c) | UofC    | 981/392           | 5   | ODI3, statistical moments, PSD, and DWT features / Binary SVM                 | 71.9 | 91.1 | 84.0 | -     | -     |
| Vaquerizo-Villar et al. (2018a) | UofC    | 981/392           | 1   | DFA and ODI3 / Regression MLP   | 97.1 | 23.3 | 82.7 | 0.41  | 0.891 |
|                                 |         |                   | 5   |   | 78.8 | 83.7 | 81.9 |       |       |
|                                 |         |                   | 10  |   | 77.1 | 94.8 | 91.1 |       |       |
| Vaquerizo-Villar et al. (2021)  | UofC    | 3196/392          | 1   | CNN architecture  | 90.8 | 36.4 | 80.1 | 0.42  | 0.917 |
|                                 |         |                   | 5   |   | 76.0 | 88.1 | 83.9 |       |       |
|                                 |         |                   | 10  |   | 79.5 | 95.8 | 92.3 |       |       |
| Vaquerizo-Villar et al. (2021)  | CHAT    | 3196/312          | 1   | CNN architecture  | 71.2 | 81.8 | 77.6 | 0.52  | 0.960 |
|                                 |         |                   | 5   |   | 83.7 | 100  | 97.4 |       |       |
|                                 |         |                   | 10  |   | 83.9 | 99.3 | 97.8 |       |       |
| Vaquerizo-Villar et al. (2021)  | BUH     | 3196/231          | 1   | CNN architecture  | 88.8 | 53.2 | 79.2 | 0.42  | 0.583 |
|                                 |         |                   | 5   |   | 61.1 | 93.7 | 83.5 |       |       |
|                                 |         |                   | 10  |   | 65.0 | 96.9 | 91.3 |       |       |

N = Number of subjects, AHI = apnea-hypopnea index, ICC = intra-class correlation coefficient, kappa = Cohen's kappa index, Se = sensitivity (%), Sp = specificity (%), Acc = accuracy (%).

The proposed CNN-based deep-learning model outperformed the DFA-based feature-engineering approach designed in Vaquerizo-Villar et al. (2018a), achieving a slightly higher 4-class kappa and ICC, as well as higher accuracies for the AHI cutoffs of 5 and 10 e/h than Vaquerizo-Villar et al. (2018a). This superior performance is even more noteworthy considering that the optimum CNN model obtained in Vaquerizo-Villar et al. (2021) was trained using only the CHAT dataset, whereas the DFA-based MLP model in Vaquerizo-Villar et al. (2018a) was trained and optimized in the UofC dataset. This highlights the generalization ability of the proposed deep-learning model. Additionally, a similar performance was obtained in Vaquerizo-Villar et al. (2021) and Vaquerizo-Villar et al. (2018c) for the AHI cutoff of 5 e/h. Apart from the aforementioned differences in the training set, in Vaquerizo-Villar et al. (2018c) we only focused on binary classification for the AHI cutoff of 5 e/h, whereas in Vaquerizo-Villar et al. (2021) we

assessed an AHI estimation model, thus emphasizing the reliability of the deep-learning solution.

Despite the fact that a comprehensive comparison with the remaining studies is not possible, in [Vaquerizo-Villar et al. \(2021\)](#) we also achieved a high performance in the CHAT and BUH datasets, specially in the CHAT dataset, where outstanding values of kappa (0.52), ICC (0.960), and accuracies (higher than 95%) for the AHI cutoffs of 5 and 10 e/h were obtained. This reinforces the suitability of deep-learning approaches to identify OSA-related hidden patterns from the oximetry signal in a pediatric OSA context. In addition, our CNN-based model is fed with raw data, thus not requiring human-driven knowledge regarding the SpO<sub>2</sub> information needed. Nonetheless, the interpretation and explanation of the features learned by the CNN is more difficult.

### 6.3.2 Comparison with state-of-the-art studies

Tables 6.3 and 6.4 show the details of state-of-the-art studies aimed at simplifying childhood OSA diagnosis by the use of the oximetry signal. Table 6.3 summarizes the results reported in previous studies based on conventional oximetric indices, whereas Table 6.4 displays results achieved by recent studies using automated signal processing approaches.

As shown in Table 6.3, the diagnostic ability of ODI and clusters of desaturations has been widely assessed ([Brouillette et al., 2000](#); [Chang et al., 2013](#); [Kirk et al., 2003](#); [Ma et al., 2018](#); [Nixon et al., 2004](#); [Tsai et al., 2013](#); [Van Eyck et al., 2015](#); [Velasco et al., 2013](#); [Villa et al., 2015](#)), including in some cases common symptoms ([Chang et al., 2013](#)) and clinical history ([Villa et al., 2015](#)). Regarding their diagnostic performance, these studies achieved accuracies ranging 62%-93% using the AHI threshold of 1 e/h ([Brouillette et al., 2000](#); [Ma et al., 2018](#); [Tsai et al., 2013](#); [Velasco et al., 2013](#); [Villa et al., 2015](#)), 64%-85% for an AHI threshold of 5 e/h ([Chang et al., 2013](#); [Kirk et al., 2003](#); [Ma et al., 2018](#); [Tsai et al., 2013](#); [Villa et al., 2015](#)), and 75%-87% using the AHI threshold of 10 e/h ([Ma et al., 2018](#); [Tsai et al., 2013](#)). In this respect, it is important to highlight that [Ma et al. \(2018\)](#) obtained a substantially lower performance than the reported by [Tsai et al. \(2013\)](#) using the ODI4. In addition, [Van Eyck et al. \(2015\)](#) obtained a different diagnostic performance than [Velasco et al. \(2013\)](#) and [Brouillette et al. \(2000\)](#) prospectively validating the methods proposed in their studies. Apart from the different databases used in these works, this varying diagnostic performance could be accounted for by the fact that these works did not employ any validation strategy to further evaluate

**Table 6.3.** Summary of state-of-the-art studies based on conventional oximetric indices

| Study                     | N<br>(Total/test) | AHI          | Methods<br>(Feature/classification)                                   | Validation          | Se                   | Sp                   | Acc                   |
|---------------------------|-------------------|--------------|---|---------------------|----------------------|----------------------|-----------------------|
| Kirk et al. (2003)        | 58/58             | 5            | ODI3 / Thresholding   | Direct validation** | 66.7                 | 60.0                 | 64.0                  |
| Tsai et al. (2013)        | 148/148           | 1<br>5<br>10 | ODI4 / Thresholding   | No                  | 77.7<br>83.8<br>89.1 | 88.9<br>86.5<br>86.0 | 79.0<br>85.1<br>87.1* |
| Brouillette et al. (2000) | 349/349           | 1            | Clusters of desaturations / Thresholding                              | Direct validation** | 42.9                 | 97.8                 | 64.7                  |
| Velasco et al. (2013)     | 167/167           | 1            | Clusters of desaturations / Thresholding                              | Direct validation** | 86.6                 | 98.9                 | 93.4*                 |
| Van Eyck et al. (2015)    | 130/130           | 2            | Brouillette et al. (2000) criteria and Velasco et al. (2013) criteria | Direct validation** | 58<br>66             | 88<br>69             | 78<br>68              |
| Chang et al. (2013)       | 141/141           | 5            | ODI3 and symptoms / Binary LR   | Direct validation** | 60.0                 | 86.0                 | 71.6                  |
| Villa et al. (2015)       | 268/268           | 1<br>5       | Clusters of desaturations and clinical history / Thresholding         | Direct validation** | 91.6<br>40.6*        | 40.6<br>97.9*        | 85.8<br>69.4*         |
| Ma et al. (2018)          | 32/32             | 1<br>5<br>10 | ODI4 / Thresholding   | No                  | 59.3<br>70.6<br>64.3 | 80.0<br>66.7<br>83.3 | 62.5<br>68.8<br>75.0* |

\* Computed from reported data, \*\* Direct validation of a scoring criteria against AHI from PSG, N = Number of subjects, AHI = apnea-hypopnea index, Se = sensitivity (%), Sp = specificity (%), Acc = accuracy (%).

their methodological approaches.

In this research, we have compared the diagnostic performance of the developed signal processing methodologies with a classical oximetric index, ODI3, obtaining higher diagnostic capability in three independent and large cohorts of pediatric OSA patients. Furthermore, hold-out, bootstrapping, and K-fold cross validation strategies were used in the methodology of all the studies carried out in the Doctoral Thesis (Vaquerizo-Villar et al., 2018a,b,c, 2021), which, together with the large sample size, contribute to a higher generalization ability of our results.

In recent years, automated feature-engineering approaches have been used to enhance the diagnostic ability of the oximetry signal (Álvarez et al., 2017; Álvarez et al., 2018; Crespo et al., 2017, 2018; Garde et al., 2014a; Hornero et al., 2017; Xu et al., 2019). These studies have employed signal processing and pattern recogni-

**Table 6.4.** Summary of state-of-the-art studies based on automated signal processing approaches

| Study                 | N<br>(Total/test) | AHI | Methods<br>(Feature/classification)   | Validation                 | Se   | Sp   | Acc  |
|-----------------------|-------------------|-----|---|----------------------------|------|------|------|
| Garde et al. (2014a)  | 146/146           | 5   | Classical indices, statistical moments, PSD, and nonlinear features / Binary LDA                | Four-fold cross validation | 80.0 | 83.9 | 78.5 |
| Álvarez et al. (2017) | 50/50             | 1   | Classical indices, statistical moments, PSD, and nonlinear features / Binary LR                 | Bootstrapping              | 89.6 | 71.5 | 85.5 |
|                       |                   | 3   |   |                            | 82.9 | 84.4 | 83.4 |
|                       |                   | 5   |   |                            | 82.2 | 83.6 | 82.8 |
| Crespo et al. (2017)  | 146/146           | 3   | Classical indices and nonlinear features / Binary LR  | Bootstrapping              | 84.5 | 83.0 | 83.5 |
| Hornero et al. (2017) | 4191/3602         | 1   | ODI3, statistical moments, PSD, and nonlinear features / Regression MLP                         | Training-Test              | 84.0 | 53.2 | 75.2 |
|                       |                   | 5   |   |                            | 68.2 | 87.2 | 81.7 |
|                       |                   | 10  |   |                            | 68.7 | 94.1 | 90.2 |
| Crespo et al. (2018)  | 176/176           | 1   | Classical indices, statistical moments, PSD, and nonlinear features / LR                        | Bootstrapping              | 93.9 | 37.8 | 84.3 |
|                       |                   | 3   |   |                            | 74.6 | 81.7 | 77.7 |
|                       |                   | 5   |   |                            | 70.0 | 91.4 | 82.7 |
| Xu et al. (2019)      | 432/432           | 1   | ODI3, statistical moments, PSD, and nonlinear features / Regression MLP                         | Training-Test              | 95.3 | 19.1 | 79.6 |
|                       |                   | 5   |   |                            | 77.8 | 80.5 | 79.4 |
|                       |                   | 10  |   |                            | 73.5 | 92.7 | 88.2 |
| Álvarez et al. (2018) | 142/142           | 5   | Classical indices, statistical moments, PSD, nonlinear, and anthropometric features / Binary LR | Bootstrapping              | 73.5 | 89.5 | 83.3 |

N = Number of subjects, AHI = apnea-hypopnea index, Se = sensitivity (%), Sp = specificity (%), Acc = accuracy (%).

tion algorithms, also applying validation strategies to ensure the generalization of their results.

For an AHI cutoff of 1 e/h, the Acc obtained in these works ranged between 75.2% and 85.5% (Álvarez et al., 2017; Crespo et al., 2018; Hornero et al., 2017; Xu et al., 2019). In this research, accuracies were included in this range (77.6% - 82.7%) (Vaquerizo-Villar et al., 2018a, 2021). Additionally, the studies showing higher Acc applied binary classifiers and used small databases (50 subjects in Álvarez et al. (2017) and 176 subjects in Crespo et al. (2018)), whereas in Vaquerizo-Villar et al. (2018a) and Vaquerizo-Villar et al. (2021) we estimated the AHI of each patient using larger cohorts (see Table 6.2).

In the case of an AHI= 5 e/h, the diagnostic accuracy ranged 78.5%-83.3% (Álvarez et al., 2017; Álvarez et al., 2018; Crespo et al., 2018; Garde et al., 2014a; Hornero et al., 2017; Xu et al., 2019). It is remarkable that our deep-learning model in Vaquerizo-Villar et al. (2021) reported higher accuracies in the UofC (83.9%), CHAT (97.4%), and BUH (83.5%) databases. Similarly, our feature-engineering model in Vaquerizo-Villar et al. (2018c) also obtained a higher Acc (84.0%) in the UofC database, whereas the accuracies reported by our feature-engineering models in Vaquerizo-Villar et al. (2018b) and Vaquerizo-Villar et al. (2018c) were within this range (81.3% and 81.9%, respectively).

Finally, in the case of an AHI threshold of 10 e/h, Hornero et al. (2017) and Xu et al. (2019) reported 90.2% and 88.2% Acc, respectively. Vaquerizo-Villar et al. (2018b) achieved a lower accuracy for this cutoff (85.3%). However, our CNN-approach in Vaquerizo-Villar et al. (2021) reported higher accuracies in the UofC (92.3%), CHAT (97.8%), and BUH (91.3%) datasets, whereas our feature-engineering proposal in Vaquerizo-Villar et al. (2018a) obtained 91.1% Acc in the UofC database.

Importantly, our feature-engineering and deep-learning approaches showed a high overall diagnostic ability in comparison with state-of-the-art studies, specially for the AHI cutoffs of 5 e/h and 10 e/h. The high diagnostic performance obtained with the proposed novel feature-engineering methodologies is consistent with the additional OSA-related information that these methods allow to quantify. Furthermore, the overall superior performance of our CNN-based methodology reinforces the ability of deep-learning approaches to learn complex features from oximetry dynamics related to apneic episodes in childhood OSA.

## 6.4 Limitations of the study

The present Doctoral Thesis has shown the utility of novel feature-engineering and deep-learning approaches applied to the oximetry signal for diagnosing pediatric OSA and its severity. However, several limitations need to be considered.

One of the main limitations is that our proposals have not been assessed by population subgroups (i.e., age ranges, sex, and/or BMI groups among others), which hinders to discover in which OSA subgroups our oximetry-based approaches could be more appropriate, as well as to discern new phenotypes within pediatric OSA able to explain differences in the pathophysiology and severity of the disease.

Despite using three datasets involving a large number of pediatric patients,

the sample size is not big enough to optimize and validate each methodology with subgroups of patients according to their clinical and physiological variables, which has resulted in a varying performance obtained with the CHAT, UofC, and BUH databases. Despite this sample size limitation, we used appropriate validation methodologies, so that optimization and validation groups from each database have similar clinical and sociodemographic characteristics.

Likewise, different recording devices and specific protocols were used for oximetry data collection in the three databases, which may influence the performance of the proposed methodologies. This is also a common problem in real-life clinical settings, as there exist multiple and pulse oximetry devices and polysomnography systems, even in the same sleep center. Notwithstanding, our proposals included a pre-processing stage to standardize the oximetry signals obtained from the different acquisition devices.

Another limitation relates to the use of the SpO<sub>2</sub> signal alone to automatically detect pediatric OSA. This limits the diagnostic ability of our proposed methodologies, since some apneic events are not linked to changes in oximetry dynamics (Berry et al., 2012; Marcus et al., 2012). In addition, the total recording time was employed as a substitute for the total sleep time to estimate the AHI, as it is not possible to determine sleep stages from the oximetry signal alone. Nevertheless, our investigation has shown that a thorough analysis of the oximetry signal can reach a remarkable diagnostic performance.

Regarding the deep-learning methodology, we only used CNNs, which were originally designed for image analysis. Nonetheless, Ismail Fawaz et al. (2019) reported that deep-learning architectures based on CNN are the most suitable for time series classification. Additionally, the CNN-based architecture was trained using only the CHAT dataset, as the UofC and BUH datasets do not contain annotations of the time location of apneic events, which may have contributed to a reduced performance in these datasets. However, our results showed that this methodology outperformed conventional methods in the three datasets. In spite of outperforming feature-engineering approaches, our CNN-based model also suffered from a lack of interpretation, which hinders to discover new knowledge regarding childhood OSA. Nonetheless, this 'black box' perception also exists in conventional pattern recognition algorithms.

Another limitation concerns to the use of the AHI as the reference measure for predicting the adverse outcomes of pediatric OSA. In this regard, recent investigations reported that novel measures of hypoxia obtained from the SpO<sub>2</sub> signal are better correlated with mortality, cardiovascular diseases, or cancer incidence

than conventional respiratory indices (i.e., AHI or ODI) in adult OSA patients. Particularly, the nocturnal hypoxemia (Oldenburg et al., 2016), the hypoxic load (Linz et al., 2018), the desaturation severity parameter (Kainulainen et al., 2020, 2019), and the hypoxic burden (Azarbarzin et al., 2019) have been proposed. As these measures have been found to further explain OSA consequences in adults, they could be also useful in the context of pediatric OSA.

The last limitation refers to the place where oximetry signals of the pediatric databases were obtained: supervised hospital facilities. In this respect, it would be interesting to further validate the methodologies proposed during the present Doctoral Thesis in a database of SpO<sub>2</sub> recordings acquired at children's home.



# Chapter 7

## Conclusions

All the scientific articles included in the compendium of publications share a common thread: the application of novel signal processing algorithms to improve the diagnostic capability of the oximetry signal in the simplification of pediatric OSA diagnosis. Feature-engineering and deep-learning methodologies were developed for this purpose. Among the feature-engineering approaches, it has been proposed three novel feature extraction algorithms (bispectrum, wavelet, and DFA) to provide additional OSA-related features from the SpO<sub>2</sub> signal in both the time and frequency domains. Additionally, a CNN-based deep-learning model was used to automatically extract all the relevant information from raw SpO<sub>2</sub> data linked to apneic events. Our results showed that the developed methodologies contribute to increase the diagnostic ability of overnight oximetry in the screening of childhood OSA.

In this chapter, the original contributions of this Doctoral Thesis are stated in section 7.1. Then, the conclusions drawn from this Doctoral Thesis are indicated in section 7.2. Finally, future research lines are listed in section 7.3.

### 7.1 Contributions

The major contributions provided by the compendium of publications of this Thesis are listed below:

- 1) Novel automated feature-engineering and deep-learning models for the analysis of the SpO<sub>2</sub> signal, which have outperformed conventional approaches, thus enhancing the diagnostic capability of nocturnal oximetry in the framework of pediatric OSA ([Vaquerizo-Villar et al., 2018a,b,c, 2021](#)).

- 2) New oximetric indices through the application of bispectral, wavelet, and DFA, which have provided relevant and complimentary information on the changes in the oximetry dynamics associated to pediatric OSA and its severity. Although these methods had shown its usefulness to analyze physiological signals in adult OSA patients, to the extent of our knowledge, this is the first time that these techniques are been applied in the context of childhood OSA (Vaquerizo-Villar et al., 2018a,b,c).
- 3) Optimum subsets obtained with the FCBF method, composed of conventional oximetry variables, as well as the new oximetric features derived from bispectral, wavelet, and DFA (Vaquerizo-Villar et al., 2018a,b,c). This highlights the relevancy and non-redundancy of the novel feature extraction methods (bispectral, wavelet, and DFA).
- 4) High performance pattern recognition models focused on binary classification (Vaquerizo-Villar et al., 2018c), multiclass classification (Vaquerizo-Villar et al., 2018b), and regression (Vaquerizo-Villar et al., 2018a). These models were fed with the optimum subsets of OSA-related features and outperformed conventional approaches, as well as state-of-the-art approaches (Vaquerizo-Villar et al., 2018a,b,c).
- 5) Novel deep-learning model based on CNNs to automatically extract all the relevant information from the SpO<sub>2</sub> signal related to apneic events. This model was validated in a large sample of 3196 SpO<sub>2</sub> recordings from three independent datasets, showing a high diagnostic ability comparing with conventional feature-engineering methodologies and state-of-the-art studies. We believe that this is the first time that deep-learning algorithms are applied in the context of pediatric OSA diagnosis (Vaquerizo-Villar et al., 2021).
- 6) Efficient screening protocols combining abbreviated test (oximetry) and artificial intelligence (pattern recognition and deep learning) able to minimize the number of PSGs. It was shown that these protocols would contribute to reduce the medical costs and waiting lists associated with the diagnosis of childhood OSA, as well as to reduce the children's discomfort during overnight PSG (Vaquerizo-Villar et al., 2018c, 2021).

## 7.2 Main conclusions

The analysis and discussion of the results obtained in the publications composing the compendium (see chapter 5 and chapter 6) lead to draw the principal conclusions of the present Doctoral Thesis, which are listed next:

- 1) The proposed feature-engineering and deep-learning models outperform conventional features from the oximetry signal, as well as state-of-the-art approaches. Thus, the application of novel signal processing techniques allows to increase the diagnostic ability of the SpO<sub>2</sub> signal from nocturnal oximetry in the context of childhood OSA.
- 2) Bispectrum can be used as a complementary tool to classical approaches in the characterization of OSA-related changes in children using SpO<sub>2</sub> recordings. Particularly, the changes in the bispectral amplitude associated to deviations of gaussianity in the oximetry signal ( $M1_{BISP}$ ) and the changes in the bispectral phase associated to a phase coupling between spectral components of the oximetry ( $meanPa_{BISP}$ ) provide additional information to anthropometric parameters, ODI3, and PSD variables in the framework of childhood OSA.
- 3) The DWT is a suitable tool to analyze the non-stationary properties, as well as the low frequency components occurring in the SpO<sub>2</sub> signal owing to pediatric OSA. Specifically, it was found that the concentration of the  $D_9$  coefficients (0.0244-0.0488 Hz) near zero ( $M3_{D_9}$ ), the energy of the  $D_9$  coefficients ( $En_{D_9}$ ), and the changes of the energy distribution ( $WE$ ) in the DWT profile of the oximetry signal provide complimentary information to conventional approaches.
- 4) DFA is an appropriate tool to identify changes in the scaling behavior of the oximetry recordings related to pediatric OSA severity. Our findings suggest that the slope of the short-time scales of the DFA profile ( $slope_1$ ) contains further information that contributes to further characterize OSA-related changes of the oximetry signal in children.
- 5) From all the pattern recognition models, the SVM model fed with ODI3, statistical moments in the time domain, PSD and DWT-derived features has provided solid evidence to detect moderate-to-severe pediatric OSA (AHI  $\geq 5$  e/h), with an Acc of 84.0% and a LR+ of 14.6. This model can be used

as a clinically valuable screening method for moderate-to-severe pediatric OSA patients.

- 6) The CNN-based deep-learning model achieves a higher overall diagnostic performance than feature-engineering approaches in the framework of childhood OSA. Particularly, this model reached accuracies above 80% (97.4%, 83.9%, and 83.5%) for diagnosing moderate-to-severe-OSA and greater than 90% (97.8%, 92.3%, and 91.3%) for the detection of severe patients in the CHAT, UofC and BUH test sets, which also outperformed state-of-the-art studies. This is consistent with the improved predictive performance shown in recent years by deep-learning algorithms in a wide range of domains. Our findings suggest that deep learning could change the paradigm of biomedical data processing in the context of pediatric OSA.
- 7) Deep-learning techniques show a high generalization ability, with a varying diagnostic performance that can be explained by differences in sampling rate, AHI distribution, age range, and patient characteristics among sleep datasets. Hence, clinical and sociodemographic parameters should be considered when validating our proposal in the clinical practice.
- 8) The diagnostic protocol derived from our deep-learning model highlights the clinical applicability of overnight oximetry for the screening of childhood OSA. Particularly, the proposed screening tool would avoid the need for 45%-70% (73.7%, 50.0%, and 45.9%) of complete PSGs in the CHAT, UofC, and BUH datasets. In this way, children would benefit from a more accessible and less intrusive diagnostic test based on the automated analysis of single-channel oximetry.

### 7.3 Future research lines

Several questions that arise from this investigation may be addressed in future work to complement our findings, and investigate other issues beyond the scope of this Doctoral Thesis. Next, the most interesting future research lines are listed:

- 1) The evaluation of our automated signal processing methodologies in subgroups of children showing different clinical characteristics would help to characterize the physiological patterns shared by the OSA pediatric population, as well as to identify those phenotypes within pediatric OSA where oximetry-based approaches achieve a higher performance.

- 2) The field of deep learning is living breakthrough advances thanks to the development of novel deep neural architectures, such as attention or inception networks. In this respect, the application of more advanced deep-learning architectures is another interesting future line of investigation that may contribute to improve the diagnostic performance of oximetry-based approaches.
- 3) The general performance of the oximetry signal may also be increased by using pretrained deep-learning networks designed for time series classification, similar to the pretrained deep-learning architectures existing in the field of image processing.
- 4) Another interesting future research could be the application of eXplainable Artificial Intelligence techniques to detect new patterns/attributes inherent to the oximetry signal linked with the severity of pediatric OSA.
- 5) The evaluation of the proposed methodologies in different types of pulse oximeters and recording systems would help to know if the diagnostic performance is affected by the technical features of the recording equipment, as well as to improve the pre-processing stage
- 6) The acquisition of the PPG signal with the pulse oximetry sensor would also be interesting, since the PPG signal contains information of the changes in the autonomic nervous system and respiratory activity related to sleep stages and apneic events. In this way, the PPG signal may help to improve the diagnostic ability of our proposal.
- 7) One natural way to continue our research would be to validate the proposed methodology in oximetry recordings acquired with portable devices at children's home, as the final goal is to perform at-home screening tests for pediatric OSA based on nocturnal oximetry.
- 8) Another future objective would be to assess the correlation of novel hypoxic measures with cardiovascular, metabolic, behavioral, and neurocognitive variables in pediatric OSA patients, as well as to propose new estimates of the level of hypoxia.



## **Appendix A**

# **Papers included in this Doctoral Thesis**

## Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings

Fernando Vaquerizo-Villar<sup>a</sup>, Daniel Álvarez-González<sup>a,b</sup>, Leila Kheirandish-Gozal<sup>c</sup>, Gonzalo C. Gutiérrez-Tobal<sup>a</sup>, Verónica Barroso-García<sup>a</sup>, Andrea Crespo<sup>a,b</sup>, Félix del Campo<sup>a,b</sup>, David Gozal<sup>c</sup>, Roberto Hornero<sup>a,d</sup>

<sup>a</sup> Biomedical Engineering Group, Universidad de Valladolid, Valladolid, Spain

<sup>b</sup> Servicio de Neumología, Hospital Universitario Río Hortega, Valladolid, Spain

<sup>c</sup> Dept. of Pediatrics, Pritzker School of Medicine, Biological Sciences Division, The University of Chicago, Chicago, United States of America

<sup>d</sup> IMUVA, Instituto de Investigación en Matemáticas, Universidad de Valladolid, Valladolid, Spain

### Computer Methods and Programs in Biomedicine

Volume 156, March 2018, Pages 141-149,

#### Abstract

**Background and objective:** The aim of this study was to assess the utility of bispectrum-based oximetry approaches as a complementary tool to traditional techniques in the screening of pediatric sleep apnea-hypopnea syndrome (SAHS). **Methods:** 298 blood oxygen saturation (SpO<sub>2</sub>) signals from children ranging 0–13 years of age were recorded during overnight polysomnography (PSG). These recordings were divided into three severity groups according to the PSG-derived apnea hypopnea index (AHI): AHI < 5 events per hour (e/h), 5 ≤ AHI < 10 e/h, AHI ≥ 10 e/h. For each pediatric subject, anthropometric variables, 3% oxygen desaturation index (ODI3) and spectral features from power spectral density (PSD) and bispectrum were obtained. Then, the fast correlation-based filter (FCBF) was applied to select a subset of relevant features that may be complementary, excluding those that are redundant. The selected features fed a multi-class multi-layer perceptron (MLP) neural network to build a model to estimate the SAHS severity degrees. **Results:** An optimum subset with features from all the proposed methodological approaches was obtained: variables from bispectrum, as well as PSD, ODI3, Age, and Sex. In the 3-class classification task, the MLP model trained with these features achieved an accuracy of 76.0% and a Cohen's kappa of 0.56 in an independent test set. Additionally, high accuracies were reached using the AHI cutoffs for diagnosis of moderate (AHI = 5 e/h) and severe (AHI = 10 e/h) SAHS: 81.3% and 85.3%, respectively. These results outperformed the diagnostic ability of a MLP model built without using bispectral features. **Conclusions:** Our results suggest that bispectrum provides additional information to anthropometric variables, ODI3 and PSD regarding characterization of changes in the SpO<sub>2</sub> signal caused by respiratory events. Thus, oximetry bispectrum can be a useful tool to provide complementary information for screening of moderate-to-severe pediatric SAHS.

**Keywords:** Sleep apnea-hypopnea syndrome (SAHS), children, oximetry, bispectrum, feature selection, feature classification.

#### 1. Introduction

Childhood sleep apnea-hypopnea syndrome (SAHS) is a breathing disorder characterized by recurrent episodes of complete cessation (apnea) and/or significant reduction (hypopnea) of airflow during sleep due to the presence

of increased upper airway collapsibility (Marcus et al., 2012). According to the American Academy of Pediatrics (AAP), SAHS has a prevalence in the range of 1% to 5% and it may impose many adverse effects on the health and development of infants and young children, such as neurocognitive deficits, cardiometabolic dysfunction, and somatic growth stunting (Marcus et al., 2012).

The gold standard test for pediatric SAHS diagnosis is overnight polysomnography (PSG). PSG requires the patient to spend the night in a specialized sleep laboratory while recording a wide range of biomedical signals (Kaditis et al., 2016; Alonso-Álvarez et al., 2011). Thus, PSG is costly and complex due to the necessary expensive equip-

Email addresses: fernando.vaquerizo@gib.tel.uva.es (Fernando Vaquerizo-Villar<sup>a</sup>), dalvgon@gmail.com (Daniel Álvarez-González<sup>a,b</sup>), lgozal@peds.bsd.uchicago.edu (Leila Kheirandish-Gozal<sup>c</sup>), ggtutob@gmail.com (Gonzalo C. Gutiérrez-Tobal<sup>a</sup>), veronica.barroso@gib.tel.uva.es (Verónica Barroso-García<sup>a</sup>), arespos@saludcastillayleon.es (Andrea Crespo<sup>a,b</sup>), fsas@telefonica.net (Félix del Campo<sup>a,b</sup>), dgozal@peds.bsd.uchicago.edu (David Gozal<sup>c</sup>), robhor@tel.uva.es (Roberto Hornero<sup>a,d</sup>)

ment and intensive labor of medical personnel. It is also especially intrusive for children, due to the use of multiple sensors, and shows limited availability in most places around the world (Nixon et al., 2004; Katz et al., 2012).

These drawbacks, together with the relatively high prevalence of the disease, have led the scientific community to explore the use of simplified screening tests (Nixon et al., 2004). The guidelines of the AAP recommend performing alternative tests when PSG is not available while requiring more conclusive evidences about the efficacy of these tests (Marcus et al., 2012). Thus, a commonly used approach has been the assessment of a reduced set of cardiorespiratory recordings. Cardiorespiratory signals contain essential information about the alterations produced by apneic events in the electrocardiogram (ECG) (Shouldice et al., 2004), pulse rate variability (PRV) (Garde et al., 2014; Dehkordi et al., 2016), airflow (AF) (Gutiérrez-Tobal et al., 2015), photoplethysmography (Gil et al., 2010), oximetry (Garde et al., 2014; Gutiérrez-Tobal et al., 2015; Kirk et al., 2003; Tsai et al., 2013; Brouillette et al., 2000; Velasco et al., 2013; Van Eyck et al., 2015; Chang et al., 2013; Villa et al., 2015; Cohen and De Chazal, 2015; Álvarez et al., 2017), and acoustic pulmonary sounds (Palaniappan et al., 2016, 2017). Among these approaches, nocturnal oximetry is the alternative most frequently advocated. In the nocturnal oximetry, pulse rate and blood oxygen saturation (SpO<sub>2</sub>) signals are recorded with a pulse oximeter probe, typically placed on the earlobe, finger, or toe (Netzer et al., 2001). Moreover, SpO<sub>2</sub> signals can be recorded in an unsupervised way at the patient's home due to the development of commercial portable pulse oximeters (Nixon et al., 2004; Garde et al., 2014). Previous studies have shown the utility of the SpO<sub>2</sub> signal to assist in the SAHS diagnosis in both adults (Marcus et al., 2008; Alvarez et al., 2013) and children (Garde et al., 2014; Gutiérrez-Tobal et al., 2015; Kirk et al., 2003; Tsai et al., 2013; Brouillette et al., 2000; Velasco et al., 2013; Van Eyck et al., 2015; Chang et al., 2013; Villa et al., 2015; Cohen and De Chazal, 2015; Álvarez et al., 2017). In this study, we aim at gaining further insights into the diagnostic ability of the SpO<sub>2</sub> signal in the screening of pediatric SAHS.

Different techniques have been reported to automatically analyze biomedical signals in the context of SAHS. Several studies have assessed the performance of frequency domain features, which reflects the duration and periodicity of respiratory events in children (Shouldice et al., 2004; Garde et al., 2014; Dehkordi et al., 2016; Gutiérrez-Tobal et al., 2015; Gil et al., 2010; Cohen and De Chazal, 2015; Álvarez et al., 2017; Palaniappan et al., 2016, 2017). Power Spectral Density (PSD) is the most common spectral analysis technique in these studies (Shouldice et al., 2004; Garde et al., 2014; Dehkordi et al., 2016; Gutiérrez-Tobal et al., 2015; Cohen and De Chazal, 2015; Álvarez et al., 2017; Palaniappan et al., 2016, 2017). However, the information present in the PSD cannot characterize phase relationships and deviations from gaus-

sianity in a signal (Chua et al., 2010). By contrast, bispectrum is a frequency domain technique defined as the spectral representation of the third order statistic that contains information about the phase of the Fourier transform of a time series. It can detect deviations from linearity, stationarity, and gaussianity in the signal, such as those produced in physiological recordings by respiratory events (Chua et al., 2010).

Based on the aforementioned considerations, we hypothesized that bispectrum analytic could provide additional information about respiratory events, thus being a complementary tool to 3% oxygen desaturation index (ODI3), anthropometric variables, and PSD parameters. Therefore, the aim of this study was to evaluate the complementarity of bispectrum to traditional approaches in the screening of pediatric SAHS using SpO<sub>2</sub> recordings.

We conducted our study in three phases: feature extraction, feature selection, and feature classification. First, anthropometric variables, ODI3 (Taha et al., 1997), and spectral features from PSD and bispectrum were obtained. Then, the fast correlation-based filter (FCBF) method (Yu and Liu, 2004) was applied to select a smaller subset of relevant and non-redundant features. Finally, a multi-layer perceptron (MLP) neural network (Marcus et al., 2008) was applied to this optimum subset for multi-class (3-class) classification in order to estimate the SAHS severity degrees according to the apnea-hypopnea index (AHI) from standard PSG.

## 2. Subjects and signals under study

The dataset was composed of 298 children (166 boys and 132 girls) ranging 0–15 years of age. All children were consecutively and prospectively referred to the Pediatric Sleep Unit at the University of Chicago Medicine Comer Children's Hospital (Chicago, IL, USA) due to clinical symptoms and physical examination findings leading to the clinical suspicion of SAHS. In all participants, an informed consent was obtained as a prerequisite to be included in the study. The Ethical Committee of the University of Chicago Medicine Comer Children's Hospital approved the protocol.

Sleep was monitored using a digital polysomnography system (Polysmith' Nihon Kohden America Inc., CA, USA). SpO<sub>2</sub> recordings were acquired during overnight PSG at a sampling rate of 25 Hz. They were exported and processed offline. Artifacts were discarded from oximetric recordings by removing SpO<sub>2</sub> values below 50% and sudden changes between consecutive SpO<sub>2</sub> sampling intervals  $\geq 4\%/s$  (Magalang et al., 2003).

Sleep and cardiorespiratory events were scored and quantified by specialized technologists who were unaware of the study purpose, and AHI was estimated according to the American Academy of Sleep Medicine guidelines. In this regard, there is no consensus about the AHI cutoff used to determine SAHS in infants (Marcus et al., 2012; Kaditis et al., 2016; Alonso-Álvarez

et al., 2011; Church, 2012; Tan et al., 2014). However, a wide range of studies typically classify children showing  $5 \leq \text{AHI} < 10$  e/h as moderate SAHS and children with  $\text{AHI} \geq 10$  e/h as severe SAHS (Kaditis et al., 2016; Alonso-Álvarez et al., 2011; Church, 2012; Tan et al., 2014). Hence, we have classified the subjects under study into the three groups defined by these commonly used thresholds ( $\text{AHI} < 5$  e/h,  $5 \leq \text{AHI} < 10$  e/h, and  $\text{AHI} \geq 10$  e/h).

The dataset was randomly divided into three sets: feature optimization set (25%), training set (50%), and test set (25%). The first set (feature optimization set) was employed to optimize the feature extraction stage and obtain an optimum subset of features with FCBF. A bootstrap procedure was applied to select the optimum features in order to select a generalizer optimum subset of features (Guyon and Elisseeff, 2003). The second set (training set) was used to select the optimal design parameters of the MLP classifier as well as train the MLP model. Ten-fold cross validation was used to emulate a different dataset when optimizing the MLP design parameters (Witten et al., 2011). Finally, the third set (test set) was employed to assess the diagnostic performance of our proposal in an independent dataset (unknown data). Table 1 shows clinical and demographic data of the population under study. No statistically significant differences ( $p$ -value  $< .05$ ) were found in the Age and Body Mass Index (BMI) between the three groups.

### 3. Methodology

#### 3.1. Feature extraction

Four clinical and signal processing approaches were applied to each subject: anthropometric, *ODI3*, and PSD and bispectrum, which lead to an initial feature set composed of 22 features.

##### 3.1.1. Anthropometric variables

Age, sex and BMI were acquired for each child since the prevalence of childhood SAHS has been associated with these factors in previous studies (Marcus et al., 2012).

##### 3.1.2. Oxygen desaturation index

In order to obtain information about the number of desaturations produced by respiratory events, *ODI3* was computed for each  $\text{SpO}_2$  recording (Kaditis et al., 2016). The definition of a desaturation event employed for computing *ODI3* is based in the study developed by Taha et al. (1997). In this study, a desaturation event occurs when  $\text{SpO}_2$  value decreases at least by 3% with respect to the preceding baseline levels, at a rate between 0.1% and 4%/second, and the  $\text{SpO}_2$  value subsequently returns to the baseline level or increases by at least 3% with respect to the preceding minimum value. The total duration of the event must be between 10 and 60 s.

##### 3.1.3. Power spectral density (PSD)

PSD was estimated for each  $\text{SpO}_2$  recording to explore differences in the spectral information of  $\text{SpO}_2$  signals associated to the duration and recurrence of apneic events. Welch's method was used to estimate PSDs (Welch, 1967), using a Hamming window of 5 min (7500 samples) with 50% overlap and a discrete Fourier transform (DFT) of  $2^{14}$  points.

According to previous research in the context of childhood SAHS diagnosis (Gutiérrez-Tobal et al., 2015), a frequency band of interest was determined as the frequency region of the PSD where there were statistically significant differences ( $p$ -value under .05) between severity groups ( $\text{AHI} < 5$  e/h,  $5 \leq \text{AHI} < 10$  e/h, and  $\text{AHI} \geq 10$  e/h) in the feature optimization set.  $p$ -value was computed between the PSD amplitudes for each pair of severity groups at each frequency using the non-parametric Mann-Whitney  $U$  test. Accordingly, our band of interest was 0.018–0.050 Hz. In this band, higher PSD amplitude is obtained as the severity of SAHS increases.

The following parameters of the PSD were computed in the band of interest:

- First-to-fourth order statistical moments ( $M1f$  –  $M4f$ ) of the PSD amplitudes. The mean ( $M1f$ ), variance ( $M2f$ ), skewness ( $M3f$ ) and kurtosis ( $M4f$ ) quantify the central tendency, dispersion, asymmetry and peakedness of the power spectrum, respectively.
- Relative power ( $P_R$ ).  $P_R$  is defined as the ratio between the power (area enclosed under the PSD) in the band of interest and the total signal power.
- Maximum amplitude ( $MA$ ) and minimum amplitude ( $mA$ ) of the PSD.
- Spectral entropy ( $SE$ ).  $SE$  is an irregularity measure which quantifies the flatness of the PSD (Poza et al., 2007).
- Mobility of the PSD ( $Mb$ ). It is a Hjorth descriptor, which measures the concentration of the signal power. It is defined as the squared root of the ratio between the variance ( $M2f$ ) and the signal power (Blanco-Velasco et al., 2010).

#### 3.2. Bispectrum

High order spectra (HOS) are representations in the frequency domain of high order cumulants of a random process (Chua et al., 2010). PSD is the Fourier transform of the second-order cumulant, while bispectrum and trispectrum are the spectral representations of the third- and fourth-order cumulant, respectively (Chua et al., 2010). Bispectrum can be described as a spectral decomposition of the skewness of a signal over frequency. In

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

| Characteristics                       | All               | AHI < 5           | 5 ≤ AHI < 10      | AHI ≥ 10          |
|---------------------------------------|-------------------|-------------------|-------------------|-------------------|
| <b>All subjects</b>                   |                   |                   |                   |                   |
| Subjects(n)                           | 298               | 164               | 56                | 78                |
| Age(years)                            | 6.0 [4.0, 9.0]    | 7.0 [5.0, 10.0]   | 5.0 [3.0, 8.0]    | 5.5 [3.0, 9.0]    |
| Males (%)                             | 55.7%             | 55.5%             | 57.1%             | 55.1%             |
| BMI(Kg/m <sup>2</sup> )               | 18.4 [16.3, 23.0] | 18.2 [16.2, 22.3] | 18.1 [16.3, 22.6] | 19.1 [16.5, 25.7] |
| Time (h)                              | 7.8 [7.3, 8.4]    | 7.9 [7.3, 8.4]    | 7.9 [7.3, 8.4]    | 7.7 [7.3, 8.3]    |
| AHI (e/h)                             | -                 | 1.9 [1.0, 3.5]    | 7.0 [5.9, 8.5]    | 17.7 [11.7, 27.3] |
| <b>Feature optimization set (25%)</b> |                   |                   |                   |                   |
| Subjects(n)                           | 74                | 41                | 14                | 19                |
| Age(years)                            | 6.0 [3.0, 10.0]   | 7.0 [4.8, 11.0]   | 4.0 [2.0, 8.0]    | 6.0 [3.0, 9.0]    |
| Males (%)                             | 58.1%             | 52.6%             | 71.4%             | 56.1%             |
| BMI(Kg/m <sup>2</sup> )               | 18.2 [16.3, 22.5] | 18.2 [16.8, 22.6] | 19.0 [16.2, 23.0] | 17.7 [15.4, 22.3] |
| Time (h)                              | 8.0 [7.5, 8.3]    | 8.0 [7.5, 8.3]    | 7.9 [7.6, 8.7]    | 7.7 [7.2, 8.2]    |
| AHI (e/h)                             | -                 | 1.7 [1.0, 3.4]    | 6.9 [5.9, 8.1]    | 17.9 [11.5, 26.4] |
| <b>Training set (50%)</b>             |                   |                   |                   |                   |
| Subjects(n)                           | 149               | 82                | 28                | 39                |
| Age(years)                            | 7.0 [4.0, 9.3]    | 7.0 [5.0, 10.0]   | 6.0 [3.0, 8.0]    | 5.0 [3.0, 9.8]    |
| Males (%)                             | 51.7%             | 46.2%             | 46.4%             | 56.1%             |
| BMI(Kg/m <sup>2</sup> )               | 18.5 [16.4, 23.2] | 18.5 [16.3, 21.7] | 17.5 [16.1, 21.1] | 20.1 [17.2, 27.8] |
| Time (h)                              | 7.7 [7.2, 8.4]    | 7.8 [7.2, 8.4]    | 7.9 [7.3, 8.4]    | 7.7 [7.3, 8.4]    |
| AHI (e/h)                             | -                 | 2.0 [1.0, 3.6]    | 7.0 [5.9, 8.5]    | 18.2 [12.0, 27.3] |
| <b>Test set (25%)</b>                 |                   |                   |                   |                   |
| Subjects(n)                           | 75                | 41                | 14                | 20                |
| Age(years)                            | 6.0 [4.0, 8.0]    | 7.0 [5.0, 8.5]    | 5.0 [4.0, 6.0]    | 5.5 [3.0, 8.5]    |
| Males (%)                             | 61.3%             | 75.0%             | 64.3%             | 53.7%             |
| BMI(Kg/m <sup>2</sup> )               | 18.1 [16.0, 23.6] | 18.0 [15.6, 23.7] | 18.5 [16.7, 23.6] | 18.5 [16.2, 24.4] |
| Time (h)                              | 7.8 [7.3, 8.4]    | 7.8 [7.2, 8.3]    | 7.9 [7.3, 8.4]    | 7.9 [7.3, 8.4]    |
| AHI (e/h)                             | -                 | 1.8 [0.9, 3.2]    | 7.0 [5.9, 8.7]    | 17.0 [11.8, 30.2] |

Data are presented as median [interquartile range],  $n$  or  $n(\%)$ , BMI= Body Mass Index, AHI= Apnea Hypopnea Index.

contrast to conventional PSD, bispectrum contains additional information about the phase relationships and deviations from gaussianity, linearity, and stationarity of a signal (Chua et al., 2010).

Let be  $x(n)$  a deterministic and zero-mean signal, the bispectrum is expressed in terms of the Fourier transform of the signal  $X(f)$  (Chua et al., 2010):

$$\begin{aligned}
 B(f_1, f_2) &= \left\{ \sum_{m=-\infty}^{\infty} x(m) \cdot e^{-j(f_1 m)} \right\} \\
 &\quad \left\{ \sum_{k=-\infty}^{\infty} x(k) \cdot e^{-j(f_2 k)} \right\} \\
 &\quad \left\{ \sum_{n=-\infty}^{\infty} x(n) \cdot e^{+j(f_1 + f_2)n} \right\} \\
 &= X(f_1) \cdot X(f_2) \cdot X(f_1 + f_2),
 \end{aligned} \tag{1}$$

where  $f_1$  and  $f_2$  are the frequency indices. Due to the symmetry conditions of the bispectrum, it is sufficient to evaluate the bispectrum in the triangular region  $\Omega$  that satisfies  $f_2 \geq 0$ ,  $f_2 \geq f_1$ ,  $f_1 + f_2 < f_s/2$  where  $f_s$  is the sampling frequency of the signal (Chua et al., 2010). In this study, bispectrum was estimated with a non-parametric approach using a Hamming window of 5 min with 50% overlap and a DFT of  $2^{14}$  points. Figure 1 shows the averaged magnitude of the bispectrum for the

three severity groups. Notice that higher amplitude in the bispectrum is observed at frequencies below 0.03 Hz, as the SAHS severity increases.

The following bispectral features were extracted in the region  $\Omega$  to quantify the differences in the bispectrum between groups (Zhou et al., 2008; Chua et al., 2008):

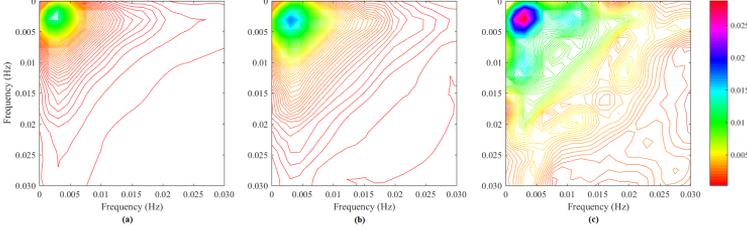
- Mean amplitude of the bispectrum ( $MB1$ ). This parameter is intended to differentiate between signals with similar PSD but different bispectrum (Chua et al., 2008):

$$MB1 = \frac{1}{L} \sum_{f_1, f_2 \in \Omega} |B(f_1, f_2)| \tag{2}$$

where  $L$  is the number of points in the region  $\Omega$ .

- Sum of logarithmic amplitudes of the bispectrum ( $H1$ ), sum of logarithmic amplitudes of elements in the diagonal of the bispectrum ( $H2$ ), and first-order spectral moment of amplitudes of elements in the diagonal of the bispectrum ( $H3$ ) (Zhou et al., 2008). These parameters are related to the moments of the bispectrum (Zhou et al., 2008):

$$H1 = \sum_{f_1, f_2 \in \Omega} \log(|B(f_1, f_2)|), \tag{3}$$



**Figure 1:** Averaged magnitude of the bispectrum for the three SAHS severity groups: (a)  $AHI < 5$  e/h, (b)  $5 \leq AHI < 10$  e/h, and (c)  $AHI \geq 10$  e/h in the feature optimization set.

$$H2 = \sum_{fk \in \Omega_{diagonal}} \log(|B(f_k, f_k)|), \quad (4)$$

$$H3 = \sum_{fk \in \Omega_{diagonal}} k \cdot \log(|B(f_k, f_k)|), \quad (5)$$

where  $\Omega_{diagonal}$  is the diagonal of the bispectrum.

- Normalized bispectral entropy ( $BE1$ ) and normalized bispectral squared entropy ( $BE2$ ), which quantify regularity in the amplitude of the bispectrum (Chua et al., 2008):

$$BE1 = - \sum_{j \in \Omega} p_j \cdot \log(p_j) \quad (6)$$

where

$$p_j = \frac{|B(f_1, f_2)|}{\sum_{f_1, f_2 \in \Omega} |B(f_1, f_2)|} \quad (7)$$

$$BE2 = - \sum_{j \in \Omega} q_j \cdot \log(q_j) \quad (8)$$

where

$$q_j = \frac{|B(f_1, f_2)|^2}{\sum_{f_1, f_2 \in \Omega} |B(f_1, f_2)|^2} \quad (9)$$

- Phase entropy ( $PE$ ), which measures irregularity in the phase of the bispectrum (Chua et al., 2008):

$$PE = - \sum_n p(\Psi_n) \cdot \log(p(\Psi_n)), \quad (10)$$

where  $N$  is the number of bins of the histogram and  $p(\Psi_n)$  is the distribution of the phase angles (Chua et al., 2010; Doane, 1976):

$$p(\Psi_n) = \frac{1}{L} \sum_{f_1, f_2 \in \Omega} \text{Ind}(\phi(B(f_1, f_2)) \in \Psi_n) \quad (11)$$

$$\Psi_n = \left\{ \phi \mid -\pi + \frac{2\pi}{n} \leq \phi < -\pi + \frac{2\pi(n+1)}{N} \right\}, \quad (12)$$

$$n = 0, 1, \dots, N-1$$

where  $\phi$  is the phase angle of the bispectrum,  $\text{Ind}(\cdot)$  is the indicator function, whose value is 1 if  $\phi$  is within the range of histogram bins  $\Psi_n$ , and  $N$  is the number of bins of the histogram, being calculated according to Doane's rule (Doane, 1976):

- Mean ( $meanPa$ ) and variance ( $varPa$ ) of the bispectral invariant ( $P(a)$ ). These features identify a chaotic process with third-order time correlations or phase coupling between spectral components (Chua et al., 2008).  $P(a)$  is the phase of the integrated bispectrum along a radial with slope  $a$  (Chua et al., 2008):

$$P(a) = \arctan\left(\frac{I_i(a)}{I_r(a)}\right) \quad (13)$$

where  $I_r(a)$  and  $I_i(a)$  are the real and imaginary part of  $I(a)$ :

$$I(a) = \int_{f_1=0^+}^{1/1+a} B(f_1, af_1) df_1 = I_r(a) + j \cdot I_i(a), \quad (14)$$

for  $0 \leq a \leq 1$ .

### 3.3. Feature selection: fast correlation-based filter (FCBF)

After the feature extraction stage, FCBF is applied to select a subset of relevant and non-redundant features (Yu and Liu, 2004). FCBF has previously shown its utility in the context of adult SAHS diagnosis (Gutiérrez-Tobal et al., 2013). It is based on symmetrical uncertainty (SU), which is a normalization of the information gain (IG) between two variables (Yu and Liu, 2004). First, features are ranked from the most relevant ones (highest SU with the AHI). Then, the features considered redundant with respect to features that are more relevant were discarded. Thus, an optimum subset with the most relevant and non-redundant features is obtained (Yu and Liu, 2004).

In order to compose an optimum feature subset independent of a particular dataset, 1000 bootstrap replicates were built from our feature optimization set so that the FCBF method was applied to each bootstrapping subset (Guyon and Elisseeff, 2003; Efron and Tibshirani, 1994).

An optimum subset composed of the variables that appear at least in 500 runs was selected.

### 3.4. Feature classification: multi-layer perceptron (MLP) neural network

Artificial neural network are mathematical models inspired in the human brain (Bishop et al., 1995). MLP is probably the most widely used neural network and it has already proven its usefulness in the context of adult SAHS diagnosis with SpO<sub>2</sub> recordings (Marcos et al., 2008). Its architecture consists on several interconnected layers (input, hidden, and output layers) composed of simple units called perceptrons. Each unit is characterized by an activation function and adaptive weights representing connections with units from the subsequent layer.

Since our problem is a 3-class classification task, the output layer has three output neurons, each one representing the posterior probability of belonging to each group. In addition, a configuration with a single hidden layer has been implemented, which may provide universal approximation to any function (Bishop et al., 1995). Weights of the network are randomly initialized. Then, they are optimized using the scaled conjugate gradient with weight decay regularization. It is used to minimize the cross-entropy error function and achieve good generalization, as recommended for classification tasks (Bishop et al., 1995).

The Netlab toolbox was used to implement our MLP classifier (Nabney, 2002). A very complex MLP model leads to overfitting, whereas a very simple model leads to underfitting. Thus, the design parameters of the MLP (the number of units in the hidden layer ( $N_H$ ) and the regularization parameter ( $\alpha$ ) were optimized by means of 10-fold cross-validation using the training set. Then, the MLP model was built using the whole training set with the optimum design parameters.

### 3.5. Statistical Analysis

The Mann Whitney  $U$  test and the Kruskal Wallis test were used to assess statistical differences ( $p$ -value $<.05$ ) between groups. The Bonferroni correction was applied to deal with multiple comparisons. Diagnostic ability of the MLP network was assessed by means of sensitivity (Se, percentage of SAHS positive patients correctly classified), specificity (Sp, percentage of SAHS negative children correctly classified), positive predictive value (PPV, proportion of subjects classified as positive that are true positives), negative predictive value (NPV, proportion of subjects classified as negative that are true negatives), positive likelihood ratio (LR<sup>+</sup>, likelihood ratio for subjects classified as positive), negative likelihood ratio (LR<sup>-</sup>, likelihood ratio for subjects classified as negative), accuracy (Acc, percentage of subjects correctly classified), and Cohen's kappa index (kappa) (Cohen, 1960).

## 4. Results

### 4.1. Feature optimization and selection

A total of 22 features were obtained for each subject: 3 anthropometric variables,  $ODI3$ , 9 parameters from PSD, and 9 bispectral features. Table 2 displays the values of these features for each SAHS severity group in the feature optimization set (median [interquartile range]), along with their corresponding  $p$ -values.  $ODI3$ , 6 out of 9 features from PSD ( $M1f$ ,  $M2f$ ,  $MA$ ,  $mA$ ,  $SE$ , and  $Mb$ ) and 4 out of 9 features from bispectrum ( $MB1$ ,  $H1$ ,  $H2$ , and  $H3$ ) showed statistical significant differences ( $p$ -value $<.05$ ). These features showed higher values as the severity of SAHS increased.

In order to assess the complementarity of bispectrum with respect to the other methodological approaches, two different feature sets were composed. The first one ( $set_{nobis}$ ) consisting of all but bispectrum features and the second one ( $set_{bis}$ ) consisting of all extracted features. FCBF was applied to each bootstrap replication generated with these feature sets ( $set_{nobis}$  and  $set_{bis}$ ) using only the feature optimization group. In both cases, an optimum subset composed of the features selected at least 500 times was obtained. The optimum subset derived when applying FCBF to  $set_{nobis}$  ( $subset_{nobis}$ ) was composed of 3 anthropometric features (Age, Sex, and BMI),  $ODI3$ , and 5 features from the PSD ( $M1f$ ,  $M2f$ ,  $MA$ ,  $P_{Rf}$ , and  $Mb$ ). Regarding the optimum subset obtained when applying FCBF to  $set_{bis}$  ( $subset_{bis}$ ), it was composed of 2 anthropometric features (Age and Sex),  $ODI3$ , 5 features from the PSD ( $M1f$ ,  $M2f$ ,  $MA$ ,  $P_{Rf}$ , and  $Mb$ ), and 2 bispectral features ( $MB1$  and  $meanPa$ ). Notice that two bispectral features were selected: one amplitude bispectral feature ( $MB1$ ) and one phase bispectral feature ( $meanPa$ ).

### 4.2. Model optimization and training

Two MLP networks fed with these optimum subsets of features obtained with FCBF ( $MLP_{nobis}$ :  $subset_{nobis}$ ;  $MLP_{bis}$ :  $subset_{bis}$ ) were designed and trained using the training set.  $N_H$  was varied from 2 up to 50, while  $\alpha$  was varied from 0 up to 5. Kappa was obtained through ten-fold cross validation for each  $N_H$ - $\alpha$  pair, and the optimum values for  $N_H$  and  $\alpha$  were obtained as those for which kappa was higher. Due to the dependence of the network to the initial random values of the weights, kappa was computed and averaged for a total of 10 runs for each  $N_H$ - $\alpha$  pair. Finally, user-dependent network parameters  $N_H = 3$  and  $\alpha = 1$  were chosen using  $subset_{nobis}$  and  $N_H = 4$  and  $\alpha = 2$  were chosen using  $subset_{bis}$ , since those pairs reached the highest kappa. The entire training set was used to train the corresponding MLP models in both cases ( $MLP_{nobis}$  and  $MLP_{bis}$ ).

### 4.3. Diagnostic performance assessment

Table 3 shows the confusion matrices of the MLP models ( $MLP_{nobis}$  and  $MLP_{bis}$ ) in the test group. These matrices show the class estimated by our MLP classifiers for

**Table 2:** Feature values for the SAHS severity groups (median [interquartile range]) in the feature optimization set.

| Features                   | AHI < 5             | 5 ≤ AHI < 10        | AHI ≥ 10            | p-value* |
|----------------------------|---------------------|---------------------|---------------------|----------|
| Age                        | 7.0 [4.8,11.0]      | 4.0 [2.0,8.0]       | 6.0 [3.0,9.0]       | .24      |
| Sex                        | -                   | -                   | -                   | .52      |
| BMI(10 <sup>1</sup> )      | 1.82 [1.68,2.26]    | 1.90 [1.62,2.30]    | 1.77 [1.54,2.23]    | .76      |
| ODI3                       | 2.08 [0.77,3.93]    | 5.82 [3.79,9.28]    | 8.72 [7.23,19.65]   | < .05    |
| M1f                        | 2.95 [1.93,4.23]    | 5.67 [5.10,8.58]    | 14.73 [7.51,28.45]  | < .05    |
| M2f (10 <sup>1</sup> )     | 0.18 [0.05,0.32]    | 1.11 [0.56,1.54]    | 4.07 [1.34,14.39]   | < .05    |
| M3f                        | 0.48 [0.29,0.75]    | 0.68 [0.33,1.04]    | 0.54 [0.32,0.83]    | .54      |
| M4f                        | 2.10 [1.79,2.73]    | 2.33 [1.87,3.06]    | 2.20 [1.60,2.70]    | .65      |
| P <sub>R</sub>             | 0.51 [0.25,0.35]    | 0.24 [0.21,0.34]    | 0.34 [0.30,0.39]    | .05      |
| MA                         | 0.60 [0.33,0.87]    | 1.27 [1.00,1.54]    | 2.92 [1.68,4.59]    | < .05    |
| mA                         | 1.21 [0.84,2.12]    | 2.30 [1.91,2.88]    | 3.46 [2.56,8.59]    | < .05    |
| SE                         | 4.33 [4.26,4.37]    | 4.30 [4.24,4.32]    | 4.20 [4.12,4.30]    | < .05    |
| Mb                         | 0.17 [0.11,0.20]    | 0.28 [0.23,0.32]    | 0.42 [0.25,0.56]    | < .05    |
| MB1 (10 <sup>-1</sup> )    | 0.05 [0.02,0.21]    | 0.23 [0.09,0.63]    | 0.49 [0.17,1.22]    | < .05    |
| H1 (10 <sup>8</sup> )      | -5.78 [-5.90,-5.77] | -5.59 [-5.73,-5.45] | -5.40 [-5.58,-5.28] | < .05    |
| H2 (10 <sup>5</sup> )      | -1.38 [-1.41,-1.33] | -1.32 [-1.36,-1.29] | -1.28 [-1.32,-1.25] | < .05    |
| H3 (10 <sup>8</sup> )      | -2.96 [-3.02,-2.86] | -2.87 [-2.93,-2.79] | -2.77 [-2.86,-2.71] | < .05    |
| BE1                        | 8.51 [7.88,9.45]    | 8.17 [7.71,8.71]    | 8.60 [8.11,8.79]    | .34      |
| BE2                        | 6.08 [4.88,6.62]    | 5.26 [4.43,5.58]    | 6.70 [5.68,7.07]    | .13      |
| PE                         | 2.14 [2.08,2.15]    | 2.12 [2.05,2.14]    | 2.11 [2.08,2.15]    | .64      |
| meanPa (10 <sup>-2</sup> ) | -1.60 [-3.19,0.81]  | 0.89 [-0.47,2.21]   | 0.92 [-1.96,3.20]   | .29      |
| varPa                      | 0.38 [0.26,0.43]    | 0.34 [0.22,0.37]    | 0.33 [0.22,0.44]    | .87      |

\*p-values obtained after Bonferroni correction.

each subject versus the actual SAHS severity group of the subjects in the test set. The overall accuracies (sum of the main diagonal elements) of these models in the test set were 69.3% (MLP<sub>nobis</sub>) and 76.0% (MLP<sub>bis</sub>), whereas the 3-class kappa values were 0.45 (MLP<sub>nobis</sub>) and 0.56 (MLP<sub>bis</sub>). Table 4 shows the diagnostic ability of these models for AHI cutoffs = 5 and 10 e/h.

Notice that the results obtained with the model MLP<sub>bis</sub> outperformed MLP<sub>nobis</sub> in terms of Se, Sp, PPV, NPV, LR+, LR-, Acc, and kappa for both cutoffs.

## 5. Discussion

This study assessed the usefulness of bispectrum to provide additional information from SpO<sub>2</sub> recordings in the screening of pediatric SAHS. The results obtained with our proposed approach suggest that the information provided by bispectrum is relevant and complementary.

Our findings showed that significantly higher values in 4 out of 9 features from bispectrum (*MB1*, *H1*, *H2*, and *H3*) are present in the subjects with the most severe degrees of SAHS. The statistical differences between groups of these bispectral features are consistent with the higher values of the bispectrum observed in Figure 1. MLP<sub>bis</sub>, which was fed with optimum features from all signal processing approaches, outperformed the neural network without information from bispectrum (MLP<sub>nobis</sub>). Regarding the optimum feature subset, FCBF automatically selected Age, Sex (anthropometric); *ODI3* (oximetric index); *M1f*, *M2f*, *MA*, *P<sub>R</sub>*, *Mb* (PSD); *MB1* and *meanPa* (bispectrum). Moreover, the results obtained in this stage suggest that information from bispectrum-based variables is complementary to that obtained from

conventional approaches.

In the test set, the proposed 3-class neural network (MLP<sub>bis</sub>) achieved an overall Acc of 76%, as well as kappa = 0.56, with 81.3% Acc and 85.3% Acc for the common cutoffs AHI = 5 e/h and AHI = 10 e/h, respectively. It is remarkable to say that, with our MLP model (MLP<sub>bis</sub>), a high positive predictive value (95.5%) is obtained for an AHI cutoff of 5 e/h, whereas a high negative predictive value (86.7%) is obtained for an AHI cutoff of 10 e/h. These cutoffs (5 and 10 e/h) were not arbitrary selected. They are commonly employed in clinical settings to define the boundary for moderate (5 ≤ AHI < 10 e/h) and severe (AHI ≥ 10 e/h) SAHS (Kaditis et al., 2016; Alonso-Álvarez et al., 2011; Church, 2012; Tan et al., 2014). For patients with an AHI ≥ 5 e/h, treatment with adenotonsillectomy is recommended (Kaditis et al., 2016). Furthermore, children with an AHI ≥ 10 e/h have an increased risk for cardiac strain and overnight observation is recommended after treatment. In this sense, continuous positive airway pressure (CPAP) is recommended in these cases when other treatment strategies such as surgery have failed (Kaditis et al., 2016).

To the best of our knowledge, this is the first study in the context of pediatric SAHS using bispectrum. Two parameters from bispectrum, *MB1* and *meanPa*, were involved in the optimum feature subset obtained with FCBF. These features contain information about the amplitude (*MB1*) and the phase (*meanPa*) of the bispectrum. Thus, the changes in the amplitude and phase of the bispectrum of oximetric recordings detected by *MB1* and *meanPa* can provide additional information about oximetry recordings to assist in pediatric SAHS screening.

**Table 3:** Confusion matrices of the MLP models in the test set. Regarding the model MLP<sub>nobis</sub> average Acc = 69.3% and kappa = 0.45, whereas for model MLP<sub>bis</sub> average Acc = 76.0% and kappa = 0.56.

| Estimated |              | MLP <sub>nobis</sub> |              |          | MLP <sub>bis</sub> |              |          |
|-----------|--------------|----------------------|--------------|----------|--------------------|--------------|----------|
|           |              | AHI < 5              | 5 ≤ AHI < 10 | AHI ≥ 10 | AHI < 5            | 5 ≤ AHI < 10 | AHI ≥ 10 |
| Actual    | AHI < 5      | 37                   | 3            | 1        | 40                 | 0            | 1        |
|           | 5 ≤ AHI < 10 | 6                    | 6            | 2        | 7                  | 5            | 2        |
|           | AHI ≥ 10     | 7                    | 4            | 9        | 6                  | 2            | 12       |

**Table 4:** Diagnostic ability of the MLP models in the test set for AHI cutoffs = 5 e/h and 10 e/h.

| AHI cutoff = 5 e/h   |      |      |      |      |       |      |      |
|----------------------|------|------|------|------|-------|------|------|
| Features             | Se   | Sp   | PPV  | NPV  | LR+   | LR-  | Acc  |
| MLP <sub>nobis</sub> | 61.8 | 90.2 | 84.0 | 74.0 | 6.33  | 0.42 | 77.3 |
| MLP <sub>bis</sub>   | 61.8 | 97.6 | 95.5 | 75.5 | 25.32 | 0.39 | 81.3 |
| AHI cutoff = 10 e/h  |      |      |      |      |       |      |      |
| Features             | Se   | Sp   | PPV  | NPV  | LR+   | LR-  | Acc  |
| MLP <sub>nobis</sub> | 45.0 | 94.5 | 75.0 | 82.5 | 8.25  | 0.58 | 81.3 |
| MLP <sub>bis</sub>   | 60.0 | 94.5 | 80.0 | 86.7 | 11.00 | 0.42 | 85.3 |

Table 5 shows the performance of previous research focused on the use of SpO<sub>2</sub> recordings in the screening of pediatric SAHS (Garde et al., 2014; Gutiérrez-Tobal et al., 2015; Kirk et al., 2003; Tsai et al., 2013; Brouillette et al., 2000; Velasco et al., 2013; Van Eyck et al., 2015; Chang et al., 2013; Villa et al., 2015; Cohen and De Chazal, 2015; Álvarez et al., 2017). ODI has been used for this task (Kirk et al., 2003; Tsai et al., 2013). Kirk et al. (2003) reached 67% Se, 60% Sp, and 64% Acc (AHI ≥ 5) using ODI3. Tsai et al. (2013) reported 79.0% Acc for AHI ≥ 1 (77.7% Se and 88.9% Sp), 85.1% Acc for AHI ≥ 5 (83.8% Se and 86.5% Sp), and 87.1% Acc for AHI ≥ 10 (89.1% Se and 86.0% Sp) using 4% ODI (ODI4) in a multiclass task. Nevertheless, in this study, ODI4 cutoff values for each severity group were optimized and validated using the same population (Tsai et al., 2013).

Clusters of desaturations have been also assessed (Brouillette et al., 2000; Velasco et al., 2013; Van Eyck et al., 2015). Brouillette et al. (2000) achieved 42.9% Se, 97.8% Sp, and 64.8% Acc (AHI ≥ 1), whereas Velasco et al. (2013) reached 86.6% Se, 98.9% Sp, and 93.4% Acc (AHI ≥ 1). However, the latter study only included patients with adenotonsillar hypertrophy, which limits its generalization (Velasco et al., 2013). Moreover, Van Eyck et al. (2015) achieved moderate Acc results when validating the methodologies proposed by Brouillette et al. (2000) (58% Se, 88% Sp, and 78% Acc) and Velasco et al. (2013) (57% Se, 73.0% Sp, and 68% Acc) in a sample of obese patients using AHI ≥ 2 e/h as cutoff. Van Eyck et al. (2015) also assessed ODI3 reaching low diagnostic ability (66% Se, 69% Sp, and 68% Acc). However, these studies only assess the presence of SAHS in children without taking into account of its severity.

Common symptoms and clinical history have been also involved in pediatric SAHS screening tools (Chang et al., 2013; Villa et al., 2015). Chang et al. (2013) used ODI3 and common symptoms to assess both a discriminative score and a logistic regression (LR) classifier (Chang et al., 2013). The LR model achieved 76.6% Acc, whereas the discriminative score reached 60% Se, 86% Sp, and 72% Acc (AHI ≥ 5). Recently, Villa et al. (2015) developed a multiclass algorithm using both clinical history and the McGill oximetry score, which was defined by Nixon et al. (2004). This paper reported 57.4% Acc in the multiclass classification task (AHI < 1 e/h, 1 ≤ AHI < 5 e/h, and AHI ≥ 5 e/h). From their confusion matrix, diagnostic performance metrics were computed: 85.8% Acc for AHI ≥ 1 (91.6% Se and 40.6% Sp), 69.4% Acc for AHI ≥ 5 (40.6% Se and 97.9% Sp), and overall kappa = 0.30.

Previous studies assessed the joint use of parameters from SpO<sub>2</sub> and other cardiorespiratory signals (Garde et al., 2014; Gutiérrez-Tobal et al., 2015; Cohen and De Chazal, 2015). Cohen and De Chazal (2015) applied linear discriminant analysis (LDA) to automatic features computed from SpO<sub>2</sub> and ECG recordings. This model achieved 58.1% Se, 67.0% Sp, and 66.7% Acc (AHI ≥ 5). Gutiérrez-Tobal et al. (2015) assessed a LR model built with ODI3 from SpO<sub>2</sub> and PSD features from AF, achieving average 85.9% Se, 87.4% Sp, and 86.3% Acc (AHI ≥ 3) using a bootstrap validation approach. Garde et al. (2014) built a LDA model using features from PRV and SpO<sub>2</sub> recordings. This model was validated using 4-fold cross validation and achieved 88.4% Se, 83.6% Sp, and 85.0% Acc (AHI ≥ 5). In contrast to these studies, our methods reached high diagnostic ability by the exclusive use of single-channel SpO<sub>2</sub> as the only signal involved.

Finally, Álvarez et al. (2017) assessed oximetry-based LR models for different AHI cutoffs (1, 3, and 5 e/h), reaching 85.5% Acc (89.6% Se and 71.5% Sp), 83.4% Acc (82.9% Se and 84.4% Sp), and 82.8% Acc (82.2% Se and 83.6% Sp), respectively. They used bootstrapping to validate results from a small sample size (50 children). While they focused on low severity AHI cutoffs, our current proposed methodology reached high diagnostic performance in the detection of moderate-to-severe pediatric SAHS (AHI ≥ 5, 10 e/h) in an independent test set from a large database (298 children).

Although we present compelling evidence on the usefulness of our method, some limitations have to be taken

**Table 5:** Summary of previous relevant published studies in the context of automated analysis of SpO<sub>2</sub> recordings to assist in the diagnosis of pediatric SAHS.

| Study                         | Subjects(n) | Signal                   | AHI (e/h)       | Methods  | Validation                         | Se                   | Sp                   | Acc                   |
|-------------------------------|-------------|--------------------------|-----------------|--|------------------------------------|----------------------|----------------------|-----------------------|
| Kirk et al. (2005)            | 58          | SpO <sub>2</sub>         | 5               | ODI3   | Direct validation**                | 67                   | 60                   | 64*                   |
| Tsai et al. (2013)            | 148         | SpO <sub>2</sub>         | 1<br>5<br>10    | ODI4   | No                                 | 77.7<br>83.8<br>89.1 | 88.9<br>86.5<br>86.0 | 79.0<br>85.1<br>87.1* |
| Brouillette et al. (2000)     | 349         | SpO <sub>2</sub>         | 1               | Clusters of desaturations  | Direct validation**                | 42.9                 | 97.8                 | 64.7*                 |
| Velasco et al. (2013)         | 167         | SpO <sub>2</sub>         | 1               | Clusters of desaturations  | Direct validation**                | 86.6                 | 98.9                 | 93.4*                 |
| Van Eyck et al. (2015)        | 130         | SpO <sub>2</sub>         | 2               | ODI3, Brouillette et al. (2000), and Velasco et al. (2013)             | Train-test for ODI3                | 57<br>58<br>66       | 73<br>88<br>69       | 68<br>78<br>68*       |
| Chang et al. (2013)           | 141         | SpO <sub>2</sub>         | 5               | ODI3 and symptoms  | Direct validation**                | 60                   | 86                   | 72*                   |
| Villa et al. (2015)           | 268         | SpO <sub>2</sub>         | 1<br>5          | McGill oximetric score and clinical history                            | Direct validation**                | 91.6<br>40.6*        | 40.6<br>97.9*        | 85.8<br>69.4*         |
| Cohen and De Chazal (2015)    | 288         | ECG and SpO <sub>2</sub> | Event detection | Statistical parameters, classical indices, and PSD                     | Loocv                              | 58.1                 | 67.0                 | 66.7                  |
| Gutiérrez-Tobal et al. (2015) | 50          | AF and SpO <sub>2</sub>  | 3               | PSD (AF) and ODI3 (SpO <sub>2</sub> )                                  | Bootstrap 0.632                    | 85.9                 | 87.4                 | 86.3                  |
| Garde et al. (2014)           | 146         | PRV and SpO <sub>2</sub> | 5               | Statistical parameters, nonlinear features, classical indices, and PSD | Four-fold cross validation         | 88.4                 | 83.6                 | 85.0                  |
| Álvarez et al. (2017)         | 50          | SpO <sub>2</sub>         | 1<br>3<br>5     | Statistical parameters, nonlinear features, classical indices, and PSD | Bootstrap 0.632                    | 89.6<br>82.9<br>82.2 | 71.5<br>84.4<br>83.6 | 85.5<br>83.4<br>82.8  |
| Our proposal                  | 298         | SpO <sub>2</sub>         | 5<br>10         | Bispectrum, PSD, ODI3, anthropometric variables                        | Feature optimization-training-test | 61.8<br>60.0         | 97.6<br>94.5         | 81.3<br>85.3          |

\*Computed from reported data, \*\* Direct validation of a scoring criteria against AHI from PSG, loocv= leave-one-out cross validation.

into account. First, there were less subjects showing an AHI in the ranges  $5 \leq \text{AHI} < 10$  and  $\text{AHI} \geq 10$  e/h in the cohort. This is one possible reason for the slight tendency of the MLP classifier to underestimate for lower SAHS severity groups. A larger sample size, balancing the proportion of subjects among classes, would likely minimize this effect. Another limitation concerns the only detection of moderate ( $5 \leq \text{AHI} < 10$  e/h) to severe ( $\text{AHI} \geq 10$  e/h) patients, while avoiding the evaluation of the presence of SAHS in subjects with  $\text{AHI} < 5$  e/h. However, while moderate to severe subjects are treated regardless the presence of co-morbidities, this group only requires treatment if neurocognitive or developmental deficits are concurrently present, and this latter feature cannot be evaluated by the AHI or any other PSG-derived variable (Tan et al., 2017). Furthermore, since our methodology aims at simplifying the detection of pediatric SAHS, it would be also useful to validate this proposal using oximetry recordings obtained in unsupervised children at home. Finally, the only use of MLP for classification is another limitation of our study. In this sense, the application of

more advanced machine learning algorithms for classification, such as ensemble learning classifiers, could be potentially useful to enhance our methodology.

In summary, a high diagnostic performance was achieved with a multiclass MLP model built with bispectral features, together with anthropometric variables, ODI3, and PSD parameters, in an independent set using a large database of oximetry recordings. Thus, bispectrum contains additional and complementary information to the other methodological approaches when aiming to further characterize desaturation events in the context of SAHS screening in children. Furthermore, this model outperformed previous results obtained by state-of-the-art studies. Therefore, bispectrum could be potentially used as a complementary tool in the analysis of oximetry recordings to help in the screening of moderate-to-severe childhood SAHS.

#### Conflict of interest

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

### Ethical approval

In all participants, the informed consent to be included in the research was obtained and the Ethical Committee of the University of Chicago Medicine approved the protocol.

### Authorship responsibility

- The material in this manuscript is original and contains no matter libelous or otherwise unlawful.
- The manuscript represents valid work and that neither this manuscript nor any other with substantially similar content under my authorship has been published or is being considered for publication elsewhere.
- I have participated sufficiently in the work to take public responsibility for all its content.

### Acknowledgements

This work was supported by 'Consejería de Educación de la Junta de Castilla y León and FEDER' under project VA037U16, by 'Ministerio de Economía y Competitividad' under projects TEC2014-53196-R and RTC-2015-3446-1, by 'European Commission' and 'European Regional Development Fund' under project 'Análisis y correlación entre el genoma completo y la actividad cerebral para la ayuda en el diagnóstico de la enfermedad de Alzheimer' ('Cooperation Programme Interreg V-A Spain-Portugal POCTEP 2014–2020'), and the project 153/2015 de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR).

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

### References

- Alonso-Álvarez, M. L., Canet, T., Cubell-Alarco, M., Estivill, E., Fernández, E., Gozal, D., Jurado-Luque, M. J., Lluich-Roselló, M. A., Martínez-Pérez, F., Merino-Andren, M., Pin-Arboledas, G., Roure, N., Sanmartí, F. X., Sans-Capdevila, O., Segarra-Isern, J., 2011. Documento de consenso del síndrome de apneas-hipopneas durante el sueño en niños. *Arch Bronconeumol* 47 (Supl 5), 2–18.
- Álvarez, D., Alonso-Álvarez, M. L., Gutiérrez-Tobal, G. C., Crespo, A., Kheirandish-Gozal, L., Gozal, D., Terán-Santos, J., Campo, F. D., 2017. Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal Oximetry: An Alternative to Respiratory Polysomnography in Unattended Settings. *J Clin Sleep Med* 13 (5), 7–11.
- Alvarez, D., Hornero, R., Marcos, J. V., Wessel, N., Penzel, T., Glos, M., Del Campo, F., 2013. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis. *International journal of neural systems* 23 (05), 1350020.
- Bishop, C. M., et al., 1995. *Neural networks for pattern recognition*. Oxford university press.
- Blanco-Velasco, M., Cruz-Roldan, F., Godino-Llorente, J. L., Barner, K. E., 2010. Nonlinear Trend Estimation of the Ventricular Repolarization Segment for T-Wave Alternans Detection. *IEEE transactions on bio-medical engineering* 57 (10), 2402–2412.
- Brouillette, R. T., Morielli, A., Leimanis, A., Waters, K. A., Luciano, R., Ducharme, F. M., 2000. Nocturnal pulse oximetry as an abbreviated testing modality for pediatric obstructive sleep apnea. *Pediatrics* 105 (2), 405–412.
- Chang, L., Wu, J., Cao, L., 2013. Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea. *International Journal of Pediatric Otorhinolaryngology* 77 (5), 365–371.
- Chua, K., Chandran, V., Acharya, U. R., Min, C., 2010. Application of higher order statistics/spectra in biomedical signals—A review. *Medical Engineering and Physics* 32 (7), 679–689.
- Chua, K. C., Chandran, V., Acharya, U., Lim, C., 2008. Cardiac state diagnosis using higher order spectra of heart rate variability. *Journal of medical engineering & technology* 32 (2), 145–155.
- Church, G. D., 2012. The Role of Polysomnography in Diagnosing and Treating Obstructive Sleep Apnea in Pediatric Patients. *Current problems in pediatric and adolescent health care* 42 (1), 22–25.
- Cohen, G., De Chazal, P., 2015. Automated detection of sleep apnea in infants: A multi-modal approach. *Computers in biology and medicine* 63, 118–123.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20 (1), 37–46.
- Dehkordi, P., Garde, A., Karlen, W., Petersen, C. L., Wensley, D., Dumont, G. A., Ansermino, J. M., 2016. Evaluation of cardiac modulation in children in response to apnea/hypopnea using the phone oximeter™. *Physiological measurement* 37 (2), 187.
- Doane, D. P., 1976. Aesthetic frequency classifications. *The American Statistician* 30 (4), 181–183.
- Efron, B., Tibshirani, R. J., 1994. *An introduction to the bootstrap*. CRC press.
- Garde, A., Dehkordi, P., Karlen, W., Wensley, D., Ansermino, J. M., Dumont, G. A., 2014. Development of a screening tool for sleep disordered breathing in children using the phone oximeter™. *PloS one* 9 (11), e112959.
- Gil, E., Bailón, R., Vergara, J. M., Laguna, P., 2010. Ptt variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of ppg signal in children. *IEEE Transactions on Biomedical Engineering* 57 (5), 1079–1088.
- Gutiérrez-Tobal, G. C., Alonso-Álvarez, M. L., Álvarez, D., del Campo, F., Terán-Santos, J., Hornero, R., 2015. Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home. *Biomedical Signal Processing and Control* 18, 401–407.
- Gutiérrez-Tobal, G. C., Álvarez, D., Marcos, J. V., Del Campo, F., Hornero, R., 2015. Pattern recognition in airflow recordings to assist in the sleep apnoea-hypopnoea syndrome diagnosis. *Medical & biological engineering & computing* 51 (12), 1367–1380.
- Guyon, I., Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research (JMLR)* 3 (3), 1157–1182.
- Kaditis, A., Kheirandish-Gozal, L., Gozal, D., 2016. Pediatric OSAS: Oximetry can provide answers when polysomnography is not available. *Sleep Medicine Reviews* 27, 96–105.
- Katz, E. S., Mitchell, R. B., Ambrosio, C. M. D., 2012. Obstructive Sleep Apnea in Infants. *American journal of respiratory and critical care medicine* 185 (8), 805–816.
- Kirk, V. G., Bohn, S. G., Flemons, W. W., Remmers, J. E., 2003. Comparison of Home Oximetry Monitoring With Laboratory Polysomnography in Children. *Chest* 124 (5), 1702–1708.
- Magalang, U. J., Dmochowski, J., Veeramachaneni, S., Draw, A., Mador,

- J., El-Solh, A., Grant, B. J. B., 2005. Prediction of the Apnea-Hypopnea Index From Overnight Pulse Oximetry. *CHEST Journal* 124 (5), 1694–1701.
- Marcos, J. V., Hornero, R., Álvarez, D., Del Campo, F., Zamarrón, C., López, M., 2008. Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. *computer methods and programs in biomedicine* 92 (1), 79–89.
- Marcus, C. L., Brooks, L. J., Ward, S. D., Draper, K. A., Gozal, D., Halbower, A. C., Jones, J., Lehmann, C., Schechter, M. S., Sheldon, S., Shiffman, R. N., Spruyt, K., 2012. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics* 130 (3), e714–e755.
- Nabney, I., 2002. *NETLAB: algorithms for pattern recognition*. Springer Science & Business Media.
- Netzer, N., Eliasson, A. H., Netzer, C., Kristo, D. A., 2001. Overnight pulse oximetry for sleep-disordered breathing in adults: a review. *Chest* 120 (2), 625–633.
- Nixon, G. M., Kermack, A. S., Davis, G. M., Manoukian, J. J., Brown, A., Brouillette, R. T., 2004. Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry. *Pediatrics* 115 (1), e19–e25.
- Palaniappan, R., Sundaraj, K., Sundaraj, S., 2017. Adaptive neuro-fuzzy inference system for breath phase detection and breath cycle segmentation. *Computer methods and programs in biomedicine* 145, 67–72.
- Palaniappan, R., Sundaraj, K., Sundaraj, S., Huliraj, N., Revadi, S., 2016. A novel approach to detect respiratory phases from pulmonary acoustic signals using normalised power spectral density and fuzzy inference system. *The clinical respiratory journal* 10 (4), 486–494.
- Poza, J., Hornero, R., Abásolo, D., Fernández, A., 2007. Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease. *Medical Engineering and Physics* 29 (10), 1075–1083.
- Shouldice, R. B., O'Brien, L. M., O'Brien, C., de Chazal, P., Gozal, D., Heneghan, C., 2004. Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features. *Sleep* 27 (4), 784–792.
- Taha, B. H., Dempsey, J. A., Weber, S. M., Badr, M. S., Skatrud, J. B., Young, T. B., Jacques, A. J., Seow, K., 1997. Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep* 20 (11), 991–1001.
- Tan, H.-L., Alonso Alvarez, M. L., Tsaoussoglou, M., Weber, S., Kaditis, A. G., 2017. When and why to treat the child who snores? *Pediatric pulmonology* 52 (5), 399–412.
- Tan, H.-L., Gozal, D., Ramirez, H. M., Bandla, H. P. R., Kheirandish-Gozal, L., 2014. Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea. *Sleep* 37 (2), 255–260.
- Tsai, C.-M., Kang, C.-H., Su, M.-C., Lin, H.-C., Huang, E.-Y., Chen, C.-C., Hung, J.-C., Niu, C.-K., Liao, D.-L., Yu, H.-R., 2013. Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. *International journal of pediatric otorhinolaryngology* 77 (8), 1286–1290.
- Van Eyck, A., Lambrechts, C., Vanheeswijck, L., 2015. The role of nocturnal pulse oximetry in the screening for obstructive sleep apnea in obese children and adolescents. *Sleep Medicine* 16 (11), 1409–1412.
- Velasco, C. T., Suárez, M. D., Figueroa, J. M., Turienzo, M. D., Len, F. L., Mansilla, E., 2013. Pulse oximetry recording in children with adenotonsillar hypertrophy : usefulness in the diagnosis of obstructive sleep apnea syndrome. *Archivos argentinos de pediatría* 111 (3), 196–201.
- Villa, M. P., Pietropaoli, N., Supino, M. C., Vitelli, O., Rabasco, J., Evangelisti, M., Del Pozzo, M., Kaditis, A. G., 2015. Diagnosis of pediatric obstructive sleep apnea syndrome in settings with limited resources. *JAMA Otolaryngology–Head & Neck Surgery* 141 (11), 990–996.
- Welch, P., 1967. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* 15 (2), 70–73.
- Witten, I., Frank, E., Hall, M., 2011. *Data Mining*. Morgan Kaufmann.
- Yu, L., Liu, H., 2004. Efficient Feature Selection via Analysis of Relevance and Redundancy. *Journal of Machine Learning Research* 5, 1205–1224.
- Zhou, S.-M., Gan, J. Q., Sepulveda, F., 2008. Classifying mental tasks based on features of higher-order statistics from eeg signals in brain-computer interface. *Information Sciences* 178 (6), 1629–1640.

## Wavelet analysis of oximetry recordings to assist in the automated detection of moderate-to-severe pediatric sleep apnea-hypopnea syndrome

Fernando Vaquerizo-Villar<sup>a</sup>, Daniel Álvarez-González<sup>a,b</sup>, Leila Kheirandish-Gozal<sup>c</sup>, Gonzalo C. Gutiérrez-Tobal<sup>a</sup>, Verónica Barroso-García<sup>a</sup>, Andrea Crespo<sup>a,b</sup>, Félix del Campo<sup>a,b</sup>, David Gozal<sup>c</sup>, Roberto Hornero<sup>a,d</sup>

<sup>a</sup> Biomedical Engineering Group, Universidad de Valladolid, Valladolid, Spain

<sup>b</sup> Pneumology Service, Hospital Universitario Río Hortega, Valladolid, Spain

<sup>c</sup> Department of Child Health, The University of Missouri School of Medicine, Columbia, Missouri, United States of America

<sup>d</sup> IMUVA, Instituto de Investigación en Matemáticas, Universidad de Valladolid, Valladolid, Spain

### PLoS ONE

Volume 13 (12), December 2018, Pages e0208502

### Abstract

**Background:** The gold standard for pediatric sleep apnea hypopnea syndrome (SAHS) is overnight polysomnography, which has several limitations. Thus, simplified diagnosis techniques become necessary. **Objective:** The aim of this study is twofold: (i) to analyze the blood oxygen saturation (SpO<sub>2</sub>) signal from nocturnal oximetry by means of features from the wavelet transform in order to characterize pediatric SAHS; (ii) to evaluate the usefulness of the extracted features to assist in the detection of pediatric SAHS. **Methods:** 981 SpO<sub>2</sub> signals from children ranging 2–13 years of age were used. Discrete wavelet transform (DWT) was employed due to its suitability to deal with non-stationary signals as well as the ability to analyze the SAHS-related low frequency components of the SpO<sub>2</sub> signal with high resolution. In addition, 3% oxygen desaturation index (ODI3), statistical moments and power spectral density (PSD) features were computed. Fast correlation-based filter was applied to select a feature subset. This subset fed three classifiers (logistic regression, support vector machines (SVM), and multilayer perceptron) trained to determine the presence of moderate-to-severe pediatric SAHS (apnea-hypopnea index cutoff  $\geq 5$  events per hour). **Results:** The wavelet entropy and features computed in the  $D_9$  detail level of the DWT reached significant differences associated with the presence of SAHS. All the proposed classifiers fed with a selected feature subset composed of ODI3, statistical moments, PSD, and DWT features outperformed every single feature. SVM reached the highest performance. It achieved 84.0% accuracy (71.9% sensitivity, 91.1% specificity), outperforming state-of-the-art studies in the detection of moderate-to-severe SAHS using the SpO<sub>2</sub> signal alone. **Conclusion:** Wavelet analysis could be a reliable tool to analyze the oximetry signal in order to assist in the automated detection of moderate-to-severe pediatric SAHS. Hence, pediatric subjects suffering from moderate-to-severe SAHS could benefit from an accurate simplified screening test only using the SpO<sub>2</sub> signal.

### 1. Introduction

The American Academy of Pediatrics (AAP) defines pediatric sleep apnea-hypopnea syndrome (SAHS) as a breathing disorder characterized by recurrent episodes of complete cessation (apnea) and/or significant reduction (hypopnea) of airflow during sleep (Marcus et al., 2012). SAHS is a highly prevalent condition among children (in the range of 1% to 5%) that may lead to many adverse consequences on the overall health and quality of life, such as

cognitive deficits, behavioral abnormalities, sleepiness, systemic inflammation, and cardiac and metabolic derangements (Marcus et al., 2012).

The gold standard technique for pediatric SAHS diagnosis is overnight polysomnography (PSG). It involves recording a wide range of biomedical signals in a specialized sleep laboratory (Kaditis et al., 2016a; Alonso-Álvarez et al., 2011). These recordings are used to score apneas and hypopneas in order to compute the apnea-hypopnea index (AHI), defined as the number of apneas and hypopneas per hour (e/h) of sleep. AHI is the clinical variable used to establish SAHS. The diagnosis of moderate-to-severe pediatric SAHS is confirmed when they present an  $AHI \geq 5$  e/h, irrespective of other comorbidities (Marcus et al., 2012). These children are at increased risk of suffering from the major negative consequences of the disease (Kaditis et al., 2016a; Hunter et al.,

*Email addresses:* fernando.vaquerizo@gib.tel.uva.es (Fernando Vaquerizo-Villar<sup>a</sup>), dalvgon@gmail.com (Daniel Álvarez-González<sup>a,b</sup>), lgozal@peds.bsd.uchicago.edu (Leila Kheirandish-Gozal<sup>c</sup>), gguttob@gmail.com (Gonzalo C. Gutiérrez-Tobal<sup>a</sup>), veronica.barroso@gib.tel.uva.es (Verónica Barroso-García<sup>a</sup>), arespos@saludcastillayleon.es (Andrea Crespo<sup>a,b</sup>), fsas@telefonica.net (Félix del Campo<sup>a,b</sup>), dgozal@peds.bsd.uchicago.edu (David Gozal<sup>c</sup>), robhor@tel.uva.es (Roberto Hornero<sup>a,d</sup>)

2016; Church, 2012). Thus, to expedite the diagnosis and treatment is essential in these patients. In this sense, surgical treatment with adenotonsillectomy is consistently recommended for children suffering from SAHS with an  $AHI \geq 5$  e/h (Tan et al., 2014). This treatment leads to an improvement in the condition in the majority of pediatric patients who suffer from moderate-to-severe childhood SAHS (Marcus et al., 2012). However, in spite of the PSG serving as the current recommended diagnostic gold standard, it is costly and complex due to the necessary equipment and trained staff, as well as highly intrusive due to the use of multiple sensors. In addition, it is a time-demanding method that shows limited availability and absent scalability, thereby delaying the diagnosis and treatment of SAHS patients (Nixon et al., 2004; Katz et al., 2012).

These drawbacks have led to extensive exploration of the use of simplified diagnostic techniques (Brockmann et al., 2013; Kaditis et al., 2016b). One common approach is the analysis of a reduced set of cardiorespiratory signals involved in PSG. In this regard, overnight oximetry is a common alternative due to its reliability, simplicity, and suitability for children (Nixon et al., 2004; Garde et al., 2014a). Nocturnal oximetry records the blood oxygen saturation ( $SpO_2$ ) signal, which provides a numerical measure of the oxygen content in hemoglobin (Berry et al., 2012). Apneic events result in decreases in blood oxygen levels and such events are termed oxyhemoglobin desaturations (Berry et al., 2012). Hence, the  $SpO_2$  signal contains useful information to detect pediatric SAHS. Previous studies have shown the usefulness of automated analysis of the  $SpO_2$  signal from nocturnal oximetry to assist in the screening of moderate-to-severe pediatric SAHS (Kirk et al., 2003; Tsai et al., 2013; Chang et al., 2013; Villa et al., 2015; Álvarez et al., 2017; Vaquerizo-Villar et al., 2018; Hornero et al., 2017). However, the results obtained in these studies indicate that an accurate diagnosis of pediatric SAHS is difficult, and in fact, substantially more difficult than in adults, particularly because the frequency of apneic events and reductions in  $SpO_2$  is markedly lower in children. Thus, further scientific evidence is still necessary before the diagnostic ability of the  $SpO_2$  signal can be widely implemented as a pragmatic tool to assist in an automated detection of childhood SAHS.

Different signal processing techniques have already been applied to characterize the changes produced in the  $SpO_2$  signal as elicited by apneic events. Conventional oximetry indices, statistical measures, nonlinear parameters, and spectral analysis from the  $SpO_2$  recordings have all been evaluated (Kirk et al., 2003; Tsai et al., 2013; Chang et al., 2013; Villa et al., 2015; Álvarez et al., 2017; Vaquerizo-Villar et al., 2018; Hornero et al., 2017). Among these approaches, the use of spectral analysis is a common choice due to the recurrence of apneic events. In this sense, previous studies have assessed features extracted from the power spectral density (PSD) and bispec-

trum (Álvarez et al., 2017; Vaquerizo-Villar et al., 2018; Hornero et al., 2017). However, these methods are based on the Short-Time Fourier Transform (STFT), thus having a fixed time-frequency resolution (Rioul and Vetterli, 1991). In contrast, wavelet transform (WT) offers high frequency resolution at low frequencies as well as high time resolution at high frequencies (Rioul and Vetterli, 1991; Daubechies, 1990). This property makes WT a potentially more suitable technique to accurately detect low frequency components, such as those associated with the duration of  $SpO_2$  desaturations. Additionally, WT is also suitable to analyze non-stationarities like those occurring in the  $SpO_2$  signal by apnea-hypopnea events. In this sense, wavelet analysis has proven its usefulness to detect changes produced in biomedical signals by apneic events among adult SAHS patients (Fontenla-Romero et al., 2005; Tagluk and Sezgin, 2011; Khandoker et al., 2008; Lin et al., 2006; Mendez et al., 2010; Lee et al., 2004). Nevertheless, only two single preliminary studies by our group evaluated the usefulness of the wavelet analysis in the detection of pediatric SAHS using the  $SpO_2$  signal (Sedano et al., 2017; Vaquerizo-Villar et al., 2017). Therefore, additional research is clearly needed to further corroborate previous findings in a small cohort and to assess the usefulness of wavelet analysis of  $SpO_2$  in the diagnosis of pediatric SAHS. Thus, we propose to develop a more exhaustive wavelet analysis with a larger database of 981 overnight  $SpO_2$  recordings.

We hypothesized that the multiresolution analysis afforded by the WT could provide a set of useful features to precisely characterize changes occurring in the  $SpO_2$  signal associated with pediatric SAHS. Consequently, the aim of this study was twofold: (i) to analyze oximetry dynamics by means of WT-derived features in order to characterize differences associated with the presence of SAHS; and (ii) to assess the usefulness of these features to assist in an automated detection of moderate-to-severe pediatric SAHS.

## 2. Subjects and signals under study

The database is composed of 981 pediatric subjects (602 males and 379 females) ranging from 2 to 10 years of age. All children were referred to the Pediatric Sleep Unit at the University of Chicago Medicine-Comer Children's Hospital (Chicago, IL, USA) in the context of clinical suspicion of SAHS. All legal caretakers of the children gave their informed consent as a prerequisite to be part of the study and the Ethics Committee of the hospital approved the protocols (#11-0268-AM017, #09-115-B-AM031, and #IRB14-1241).

Children's sleep was monitored using a digital polysomnography system (Nihon Kohden America Inc., CA, USA).  $SpO_2$  recordings were acquired during overnight polysomnography at sampling rates of 25, 200, or 500 Hz. In a preprocessing stage, artifacts were removed by discarding those  $SpO_2$  values below 50% and

**Table 1:** Demographic and clinical characteristics of the patient groups under study.

|                                 | All                  | Optimization set     | Cross-validation set |
|---------------------------------|----------------------|----------------------|----------------------|
| <b>Subjects (n)</b>             | 981                  | 589                  | 392                  |
| <b>Age (years)</b>              | 6 [3, 9]             | 6 [3, 8]             | 6 [3, 9]             |
| <b>Males (n)</b>                | 602 (61.4%)          | 347 (58.9%)          | 255 (65.1%)          |
| <b>BMI (Kg/m<sup>2</sup>)</b>   | 17.9<br>[15.8, 21.9] | 17.6<br>[15.9, 22.0] | 18.1<br>[15.8, 21.7] |
| <b>AHI (e/h)</b>                | 3.8 [1.5, 9.3]       | 4.1 [1.7, 9.9]       | 3.3 [1.4, 7.8]       |
| <b>Group AHI &lt; 5 e/h (n)</b> | 576 (58.7%)          | 330 (56.0%)          | 246 (62.8%)          |
| <b>Group AHI &lt; 5 e/h (n)</b> | 405 (41.3%)          | 259 (44.0%)          | 146 (37.2%)          |

Data are presented as median [interquartile range],  $n$  or  $n(\%)$ , BMI= Body Mass Index, AHI= Apnea Hypopnea Index.

those intervals with a slope higher than 4%/s (Magalang et al., 2003). Then, SpO<sub>2</sub> recordings were resampled to a common rate of 25 Hz, as recommended by the American Academy of Sleep Medicine (AASM) (Berry et al., 2012), and were rounded to the second decimal place in order to have the same resolution (Garde et al., 2014b). The guidelines of the AASM were used by a certified pediatric sleep specialist to quantify sleep and cardiorespiratory events. The AHI was subsequently derived in order to diagnose pediatric SAHS. An AHI of 5 e/h was the threshold used to establish moderate-to-severe SAHS because of the enhanced risk of morbidity and thus the importance of an early detection and treatment in these cases. According to this AHI-based cutoff, 405 children were in the group AHI ≥ 5 e/h, whereas 576 children were in the group AHI < 5 e/h.

The dataset was randomly divided into an optimization set (60%) and a cross-validation set (40%) (Hornero et al., 2017). Table 1 shows demographic and clinical data of the population under study (median [interquartile range] or  $n$  (%)). No statistically significant differences ( $p$ -value < .01) emerged in either age or body mass index (BMI) between optimization and cross-validation groups.

### 3. Methods

Our methodology is divided into three steps: feature extraction, selection, and classification. In the first step, the wavelet transform was applied to analyze each SpO<sub>2</sub> signal. A set of features was computed using the discrete wavelet transform (DWT) to characterize the changes produced in SpO<sub>2</sub> recordings due to SAHS. In addition, 3% oxygen desaturation index (ODI3), statistical moments in the time domain and PSD features, which are common features from the SpO<sub>2</sub> signal (Álvarez et al., 2017; Hornero et al., 2017), were obtained to compose a wide initial feature set with relevant as well as complementary information. In the second step, a feature subset was selected using the fast correlation-based filter (FCBF) method (Yu and Liu, 2004). Finally, binary logistic regression (LR) (Hosmer and Lemeshow, 2004), support vector

machines (SVM) (Bishop, 2006) and multi-layer perceptron (MLP) neural network (Bishop et al., 1995) classifiers were trained using this selected feature subset in order to detect moderate-to-severe pediatric SAHS.

Figure 1 shows the validation approach employed in each methodological step. The first set (optimization set) was employed to perform descriptive analysis of the extracted features, select a subset of features with FCBF, and select the optimal design parameters of the SVM and MLP classifiers. Bootstrapping has been employed in the feature selection stage, in order to avoid overfitting (Witten et al., 2011) In the same way, 10-fold stratified cross validation has been applied to optimize the design parameters of SVM and MLP. The second set (cross-validation set) was used to evaluate the diagnostic performance of the single features and classifiers. Stratified  $K$ -fold cross validation ( $K = 5$ ) was applied for this purpose (Steierberg and Vergouwe, 2014).

#### 3.1. Feature extraction

##### 3.1.1. Discrete Wavelet Transform

WT can be seen as the decomposition of a signal  $x(t)$  onto a set of basis functions, called wavelets (Rioul and Vetterli, 1991). Wavelets are obtained by time translations and scaling of a unique function called the mother wavelet. WT can be seen as an extension of the Fourier transform where, instead of analyzing a single scale, a multiscale analysis is performed. This multiscale property of the WT allows decomposing a signal into a set of scales, where each scale analyzes a different frequency range of the signal. WT can be continuous (Continuous Wavelet Transform, CWT) or discrete (DWT), depending on the scale and translation values (Rioul and Vetterli, 1991). CWT computes WT for each scale, whereas DWT only computes WT for dyadic (power of 2) scales, thus presenting lower complexity and higher computational efficiency than CWT (Cvetkovic et al., 2008). Consequently, DWT was chosen in this study. In addition, it has previously shown its usefulness to detect different frequency components in physiological signals associated to SAHS events in adult patients (Fontenla-Romero et al., 2005; Tagluk and Sezgin, 2011; Khandoker et al., 2008; Lin et al., 2006; Mendez et al., 2010; Lee et al., 2004).

Figure 2 shows how DWT is computed. In Figure 2A, the decomposition process of a SpO<sub>2</sub> signal  $x[n]$  using DWT, the so-called subband coding scheme, is illustrated. It is a filter-bank tree where each stage consists of a high pass-filter  $g[n]$  (the mother wavelet) and a low pass filter  $h[n]$  (the mirror version of the mother wavelet), followed by a subsampling process of factor two (Rioul and Vetterli, 1991). The relationship between these two filters is as follows (Rioul and Vetterli, 1991):

$$g[L - 1 - n] = (-1)^n \cdot h[n], \quad (1)$$

where  $L$ , an even number, is the length of the filter. First,  $x[n]$  is decomposed in an approximation signal (lowpass

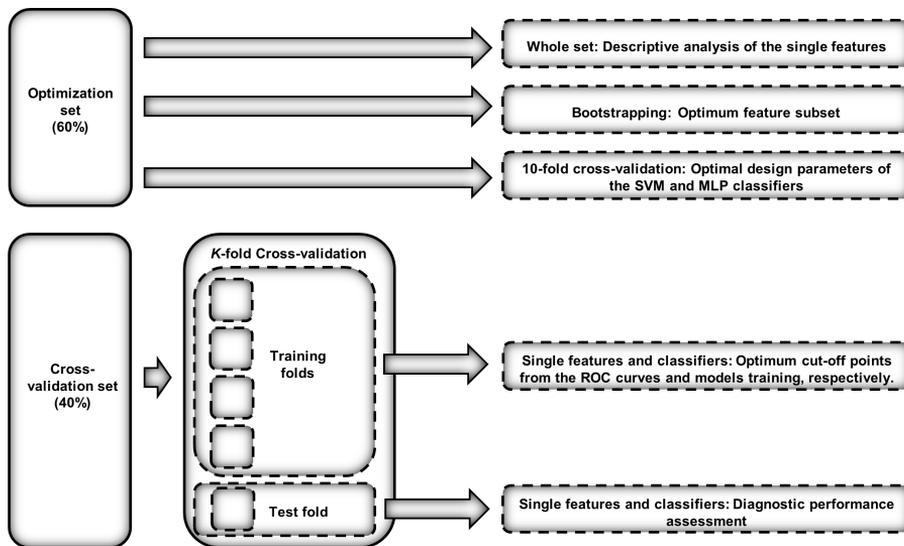


Figure 1: Validation approach employed in each methodological step of the study.

version),  $A_1$ , and a detail signal (highpass version),  $D_1$ . Then,  $A_1$  is further decomposed into another approximation signal,  $A_2$ , and another detail signal,  $D_2$ . Each iteration increases the frequency resolution of the approximation and the detail version by two, as well as decreases the number of samples of both approximation and detail signals. This process continues until the maximum detail level of the signal,  $N = \log_2(M)$  is reached, being  $M$  the length of  $x[n]$  (Rosso et al., 2001). At each level ( $i = 1, 2, \dots, N$ ), the approximation signal,  $A_i$ , and the detail signal,  $D_i$ , can be computed as follows:

$$D_i[k] = \sum_n A_{i-1}[n] \cdot g[2k - n], \quad (2)$$

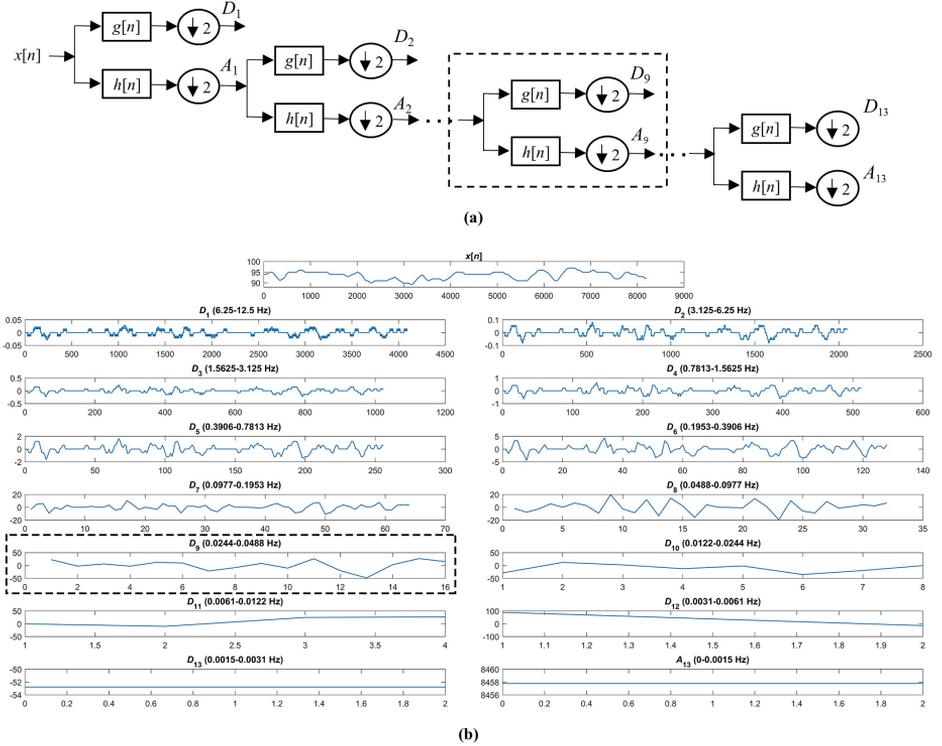
$$A_i[k] = \sum_n A_{i-1}[n] \cdot h[2k - n], \quad (3)$$

where  $A_{i-1}$  is the approximation signal in the level  $i - 1$ . In the level 1,  $A_0$  is the original signal  $x[n]$ . Figure 2B shows an example of  $\text{SpO}_2$  signal,  $x[n]$ , the detail signal  $D_i$  obtained at each level  $i$  of the DWT decomposition, and the approximation signal  $A_N$  obtained at the level  $N$  of the DWT decomposition.

DWT was applied to the upper power of 2 for 5 minute segments ( $M = 2^{13}$  samples (5.5 minutes)) and, consequently,  $N = 13$  (Tagluk and Sezgin, 2011). In this

study, the Haar wavelet was chosen as mother wavelet. The reason is twofold (Lee et al., 2004): (i) its suitability for picking up abrupt changes, which is appropriate to detect the changes produced in the  $\text{SpO}_2$  values due to apneic events; and (ii) its smoothing feature, which does not distort the original form of the  $\text{SpO}_2$  signal. At each level of the decomposition, detail coefficients contain information about a different frequency band, as stated in Figure 2B. We focused on the detail coefficients of the level 9 ( $D_9$ , i.e., 0.0244–0.0488 Hz), since it is the level which is contained in the band of interest previously related to the recurrence of apneic events (Vaquerizo-Villar et al., 2018).  $\text{SpO}_2$  signal presents both drops and rises associated to apneic events, which result in decreased and increased values in  $D_9$  coefficients, respectively. The information contained in the  $D_9$  coefficients may be canceled due to the presence of both positive and negative values, such as mean or skewness. To avoid this, the absolute values of the DWT coefficients were used. The following seven features were extracted from the DWT coefficients:

- Statistical moments of the  $D_9$  coefficients (Mean ( $M1_{D_9}$ ), variance ( $M2_{D_9}$ ), skewness ( $M3_{D_9}$ ) and kurtosis ( $M4_{D_9}$ )).  $M1_{D_9}$ – $M4_{D_9}$  measure the central tendency, dispersion, asymmetry and peakedness of the data, respectively.



**Figure 2: DWT computation.**(A) Decomposition process of a signal using DWT. (B) Original SpO<sub>2</sub> signal, detail signals at each decomposition level and approximation signal at the maximum level of the decomposition.

- Maximum amplitude of the  $D_9$  coefficients ( $Max_{D_9}$ ). It quantifies the highest amplitude in this frequency band.
- Energy of the  $D_9$  coefficients ( $En_{D_9}$ ). It measures the averaged quadratic amplitude of the signal in  $D_9$ . It is computed as follows:

$$En_{D_9} = \sum_k |D_9[k]|^2 \quad (4)$$

- Wavelet Entropy ( $WE$ ), which measures the irregularity introduced in the DWT. It was extracted in order to obtain information about the changes produced in the energy distribution of the different detail levels of the DWT of the SpO<sub>2</sub> signal by apneic

events (Rosso et al., 2001). It is computed as follows:

$$WE = - \sum_{i=1}^N p_i \log(p_i), \quad (5)$$

where  $p_i$  is the relative wavelet energy at the detail level  $D_i$ :

$$p_i = \frac{En_{D_i}}{\sum_{i=1}^N En_{D_i}} \quad (6)$$

where  $En_{D_i}$  is the wavelet energy at the detail level  $D_i$ :

$$En_{D_i} = \sum_k |D_i[k]|^2 \quad (7)$$

### 3.1.2. Conventional features from the SpO<sub>2</sub> signal

In order to enhance the diagnostic ability of our proposal, the following features, that are common parameters of the oximetry signal (Álvarez et al., 2017; Hornero

et al., 2017), were computed:

- *ODI3*. It was estimated as the number of desaturations of at least 3% from preceding baseline per hour of recording (Taha et al., 1997). This parameter has shown its usefulness in clinical studies, even though it underestimates AHI (Kirk et al., 2003; Tsai et al., 2013; Chang et al., 2013).
- Statistical moments. First-to-fourth order statistical moments were computed from the SpO<sub>2</sub> signal in the time domain ( $M1_T$  -  $M4_T$ ): mean ( $M1_T$ ), variance ( $M2_T$ ), skewness ( $M3_T$ ), and kurtosis ( $M4_T$ ) (Álvarez et al., 2017; Hornero et al., 2017). These features measure the central tendency, dispersion, asymmetry, and peakedness of the data, respectively.
- PSD features. PSD was estimated using the Welch's method ( $2^{13}$ -sample Hamming window, 50% overlap and  $2^{14}$ -points DFT) (Welch, 1967). The following features were obtained: first-to-fourth order statistical moments ( $M1_{PSD}$  -  $M4_{PSD}$ ) and maximum amplitude ( $Max_{PSD}$ ) from the band of interest determined in (Vaquerizo-Villar et al., 2018) (0.018–0.050 Hz) and spectral entropy ( $SE_{PSD}$ ) in the full spectrum. These features provide information about the recurrence and duration of apneic events.

### 3.2. Feature selection: fast correlation-based filter (FCBF)

The FCBF method was applied to select a subset of relevant and non-redundant features (Yu and Liu, 2004). FCBF is a feature selection algorithm that has previously shown its usefulness in the context of pediatric SAHS (Vaquerizo-Villar et al., 2018; Hornero et al., 2017). First, FCBF computes the symmetrical uncertainty (*SU*) between each feature ( $x_i$ ) and the AHI ( $y$ ). *SU* is a normalization of the information gain between two variables. *SU* is computed as follows (Yu and Liu, 2004):

$$SU(x_i, y) = 2 \cdot \left( \frac{IG(x_i|y)}{H(x_i) + H(y)} \right), i = 1, 2, \dots, N, \quad (8)$$

where  $IG(x_i|y) = H(x_i) - H(x_i|y)$ ,  $N$  is the total number of features extracted and  $H$  refers to Shannon's entropy [32]. According to their *SU* value (between 0 and 1), features are ranked from the most relevant (highest *SU* with the AHI) to the least relevant one (lowest *SU* with the AHI). Then, a redundancy analysis is performed. *SU* between each pair of features ( $x_j$ ,  $x_i$ ) is computed. Features  $x_j$  sharing more information with a more relevant one than with the AHI ( $SU(x_j|x_i) \geq SU(x_j|y)$ ) were discarded. Finally, an optimum subset composed of the features not discarded in this process is obtained.

A bootstrap approach was employed in order to obtain a subset of features independent of a particular dataset. In this regard, FCBF was applied to 1000 bootstrap replicates built with a sample with replacement procedure from the optimization set (Efron and Tibshirani, 1994; Guyon and

Elisseeff, 2003). Those variables that were selected with FCBF more than 500 times (50% of runs) formed the feature subset (Vaquerizo-Villar et al., 2018; Hornero et al., 2017).

### 3.3. Feature classification

In this study, we employed LR, SVM, and MLP, which are well-known algorithms in the context of binary classification. Particularly, these algorithms were applied to assign each subject to the groups AHI < 5 e/h and AHI ≥ 5 e/h (Hosmer and Lemeshow, 2004; Bishop, 2006; Bishop et al., 1995).

#### 3.3.1. Logistic regression

LR is a standard machine learning approach for binary classification. Given a set of input features, LR estimates the posterior probability of a given instance (subject) belonging to one of two mutually exclusive groups (AHI < 5 e/h and AHI ≥ 5 e/h) by the use of the logistic function (Hosmer and Lemeshow, 2004):

$$p(C_l|x_k) = \frac{1}{1 + \exp -(\beta_0 + \beta_1 \cdot x_{1,k} + \dots + \beta_N \cdot x_{N,k})}, \quad (9)$$

where  $C_l$  represents the two groups (AHI < 5 e/h and AHI ≥ 5 e/h),  $\beta = \beta_0, \beta_1, \dots, \beta_N$  are the coefficients of the model for each input feature,  $x_k = x_{1,k}, \dots, x_{N,k}$ , is the input pattern for the instance  $k$ , and  $N$  is the number of features. A Bernoulli distribution is used to model the probability density function and  $\beta$  coefficients are optimized using the maximum likelihood ratio (Hosmer and Lemeshow, 2004).

#### 3.3.2. Support vector machines

A SVM is a binary classifier that searches for the best hyperplane that separates instances from the classes under study (Bishop, 2006). The hyperplane has the following expression (Bishop, 2006):

$$y(x, w) = w^T \cdot \phi(x) + w_0 \quad (10)$$

where  $x \in R_N$  is the input pattern of dimension  $N$  (number of features),  $\phi(x) \in R_P$  transforms the data into a high-dimensional space  $P > N$ , and  $w$  is the weight vector. The weight vector  $w$  is optimized in order to maximize the margin of separation between the two groups (Bishop, 2006). A regularization parameter  $C$  was applied to control the trade-off between maximizing the margin of separation between groups and obtaining a good generalization ability in an independent set (Bishop, 2006). The optimization problem of SVM is formulated using Lagrange multipliers:

$$y(x, w) = - \sum_{i \in S} \eta^i K(x^i, x) + w_0, \quad (11)$$

where  $S$  is a subset of indices  $\{1, \dots, L\}$  corresponding to the non-zero Lagrange multipliers (support vectors)  $\eta^i$ ,

$L$  is the number of observations in the training set,  $t^i$  are the output labels ( $\pm 1$  for the  $\text{AHI} < 5$  and  $\text{AHI} \geq 5$  e/h groups), and  $K(\cdot, \cdot)$  is the Kernel function in the transformed space. In this study, a linear kernel was used, which has previously shown its usefulness in the context of adult SAHS (Alvarez et al., 2013). The value of  $C$  was optimized by means of 10-fold cross-validation using the optimization set.

### 3.3.3. Multi-layer perceptron neural network

A MLP is an artificial neural network arranged in several fully connected layers: input, hidden, and output layers (Bishop et al., 1995). These layers are composed of computing units called perceptrons or neurons. Each neuron consists of an activation function  $g^k(\cdot)$  and adaptive weights  $w_{kj}$  that interconnect the neuron with neurons from the subsequent layer (Bishop et al., 1995). The input layer was composed of one neuron for each input feature. Additionally, a configuration with one single hidden layer with a hyperbolic tangent activation function was applied since it provides a fast convergence for the training algorithm (Bishop et al., 1995). This configuration can provide universal approximation to any continuous function with the only condition that there are enough hidden units (Bishop et al., 1995; Hornik, 1991). Finally, two neurons composed the output layer, since our problem is a binary classification task. A logistic sigmoid activation function has been used in the output layer, because it allows the output neurons to be interpreted probabilistically (Bishop et al., 1995):

$$y_k = g^k \left\{ \sum_{j=1}^{N_H} w_{kj} g^j \left\{ \sum_{i=1}^N w_{ji} x_i + b_j \right\} + b_k \right\}, \quad (12)$$

where  $y_k$  are the outputs neurons,  $w_{kj}$  are the weights connecting the hidden layer to the output layer,  $w_{ji}$  are the weights connecting the input layer to the hidden layer,  $b_j$  and  $b_k$  are the bias associated to the hidden and the output units, respectively,  $x_i$  is the feature  $i$ ,  $g^k(\cdot)$  and  $g^j(\cdot)$  are the activation functions of the output and hidden layer, respectively,  $N_H$  is the number of neurons in the hidden layer, and  $N$  is the number of input features (Bishop et al., 1995). Random initialization was performed for the weights of the network. Then, the scaled conjugate gradient algorithm with weight-decay regularization was used to optimize the weights (Bishop et al., 1995).  $N_H$  and the regularization parameter ( $\alpha$ ) were optimized by means of 10-fold cross-validation using the optimization set.

### 3.4. Statistical Analysis

The software tools Matlab version R2017a was used for performing signal processing and statistical analyses. Normality and homoscedasticity tests showed that extracted parameters were not normality distributed

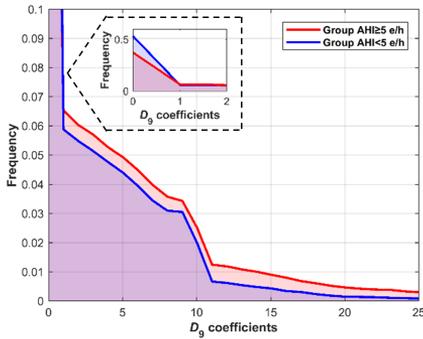
and had different variances. Consequently, the Mann-Whitney  $U$  test was applied to search for statistical significant differences in the extracted features ( $p$ -value  $< 0.01$ ) between groups. Diagnostic performance was assessed by means of sensitivity (Se, percentage of patients with an  $\text{AHI} \geq 5$  e/h correctly classified), specificity (Sp, percentage of children with an  $\text{AHI} < 5$  e/h correctly classified), positive predictive value (PPV, proportion of subjects classified as positive that are true positives), negative predictive value (NPV, proportion of subjects classified as negative that are true negatives), positive likelihood ratio (LR+, likelihood ratio for subjects classified as positive), negative likelihood ratio (LR-, likelihood ratio for subjects classified as negative), and accuracy (Acc, percentage of subjects correctly classified).

$K$ -fold stratified cross validation ( $K = 5$ ) was applied to assess the performance of the extracted features and the binary classifiers (Steyerberg and Vergouwe, 2014). The cross-validation set was randomly divided into  $K$  subsets, preserving the proportion of subjects belonging to the groups  $\text{AHI} < 5$  e/h and  $\text{AHI} \geq 5$  e/h.  $K-1$  folds formed the training folds (80% of the cross-validation set), whereas the remaining one formed the test fold (20% of the cross-validation set). Accordingly, Receiver Operating Characteristics (ROC) curves were used to obtain optimum classification cut-off points for the single features using the  $K-1$  training folds. Similarly, the classification algorithms were trained using the training folds. Then, the diagnostic performance of the single features and the LR, SVM, and MLP classifiers was measured using the test fold. This process was repeated  $K$  times, so each fold was considered once as the test fold. Finally, all the metrics are averaged across the  $K = 5$  iterations.

## 4. Results

### 4.1. Feature separability

A total of seven DWT-derived features were obtained for each  $\text{SpO}_2$  recording (S1 Table). Figure 3 shows the histogram of the  $D_9$  coefficients in the optimization set for the groups  $\text{AHI} < 5$  e/h and  $\text{AHI} \geq 5$  e/h. It can be observed that  $D_9$  coefficients are more concentrated near zero in the  $\text{AHI} < 5$  e/h group, whereas in the group  $\text{AHI} \geq 5$  e/h these coefficients are more disperse. Table 2 shows the median and interquartile range of all these extracted features in the optimization set for both groups. All features showed significant statistical differences ( $p$ -value  $< 0.01$ ) between groups.  $M1_{D_9}$ ,  $M2_{D_9}$ ,  $Max_{D_9}$ ,  $En_{D_9}$ , and  $WE$  showed higher values in the  $\text{AHI} \geq 5$  e/h group, whereas  $M3_{D_9}$  and  $M4_{D_9}$  showed higher values in the  $\text{AHI} < 5$  e/h group.  $ODI3$ , statistical moments and PSD features were also computed for each  $\text{SpO}_2$  recording (S1 Table).  $ODI3$ , 3 out of 4 statistical moments ( $M1_T$ ,  $M2_T$ , and  $M3_T$ ) and 3 out of 6 spectral features ( $M1_{PSD}$ ,  $M2_{PSD}$ , and  $Max_{PSD}$ ) also showed significant statistical differences ( $p$ -value  $< 0.01$ ), which agrees with previ-



**Figure 3:** Histogram of the  $D_9$  coefficients for each group in the optimization set.

**Table 2:** DWT-derived features for each group in the optimization set.

| Features           | AHI < 5           | AHI ≥ 5           | $p$ -value |
|--------------------|-------------------|-------------------|------------|
| $M1_{D_9}$         | 3.04 [2.26, 3.92] | 5.36 [3.77, 7.70] | $p < .01$  |
| $M2_{D_9}$         | 3.78 [3.23, 4.63] | 5.73 [4.30, 7.57] | $p < .01$  |
| $M3_{D_9}$         | 1.31 [1.20, 1.44] | 1.19 [1.06, 1.32] | $p < .01$  |
| $M4_{D_9} (10^2)$  | 3.58 [1.03, 7.69] | 0.06 [0.04, 2.69] | $p < .01$  |
| $Max_{D_9} (10^1)$ | 1.23 [1.04, 1.55] | 1.96 [1.42, 2.62] | $p < .01$  |
| $En_{D_9} (10^5)$  | 0.54 [0.37, 0.89] | 1.54 [0.78, 2.96] | $p < .01$  |
| $WE (10^{-4})$     | 1.83 [1.18, 2.86] | 4.27 [2.52, 9.41] | $p < .01$  |

ous studies (Álvarez et al., 2017; Vaquerizo-Villar et al., 2018).

#### 4.2. Optimum feature subset

FCBF was applied to each bootstrap replicate from the optimization set, each one composed of all the extracted features ( $ODI3$ , statistical moments, PSD, and DWT features).  $ODI3$ , 1 statistical moment ( $M2_T$ ), 3 features from PSD ( $M2_{PSD}$ ,  $M3_{PSD}$ , and  $Max_{PSD}$ ), and 3 DWT-derived features ( $M3_{D_9}$ ,  $En_{D_9}$ , and  $WE$ ) were selected more than 50% of times (500) (S2 Table). Thus, these features formed the selected feature subset (Vaquerizo-Villar et al., 2018; Hornero et al., 2017). Notice that features from all the different methodological approaches were selected.

#### 4.3. Classification models optimization

LR, SVM, and MLP classifiers were designed using the selected feature subset obtained with FCBF ( $ODI3$ ,  $M2_T$ ,  $M2_{PSD}$ ,  $M3_{PSD}$ ,  $Max_{PSD}$ ,  $M3_{D_9}$ ,  $En_{D_9}$ , and  $WE$ ). Optimum values for the design parameters of the SVM (regularization parameter:  $C$ ) and MLP classifiers (number of neurons in the hidden layer:  $N_H$ ; regularization parameter:  $\alpha$ ) were obtained as those for which the Acc of the

classifiers was the highest in the optimization set. Concerning SVM, the following values of  $C$  were assessed:  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ , ...,  $10^4$ ,  $10^5$ . The optimum value of the input parameter  $C$  was  $10^3$ , which maximizes Acc. Regarding MLP,  $N_H$  was varied from 2 up to 50 and  $\alpha$  was varied from 0 up to 10. Since the network depends on the initial random values of the weights, the accuracy was computed and averaged for a total of 10 runs for each pair  $N_H$ - $\alpha$ . Finally, user-dependent network parameters  $N_H=5$  and  $\alpha=1$  were chosen since this pair reached the highest accuracy.

#### 4.4. Diagnostic performance

The value of all the extracted features ( $ODI3$ , statistical moments, PSD, and DWT features) and the classification score of the LR, SVM, and MLP classifiers were obtained for each subject in the cross-validation set (S3 Table). Table 3 shows the diagnostic ability of each single feature in the cross-validation set obtained using optimum cut-off point obtained from the ROC curve. Most of the DWT-derived features (5 out of 7) showed accuracies near 80%. In this regard,  $Max_{D_9}$  achieved the highest performance ( $81.7 \pm 5.6\%$  Acc, with  $75.4 \pm 7.1\%$  Se and  $85.4 \pm 6.8\%$  Sp), outperforming statistical moments and PSD features. Only  $ODI3$  achieved slightly higher Acc than  $Max_{D_9}$ , reaching  $81.9 \pm 7.2\%$  Acc ( $78.1 \pm 7.3\%$  Se and  $84.2 \pm 8.1\%$  Sp). Table 4 shows the diagnostic performance of LR, SVM, and MLP classifiers, which were trained using the selected feature subset ( $ODI3$ ,  $M2_T$ ,  $M2_{PSD}$ ,  $M3_{PSD}$ ,  $Max_{PSD}$ ,  $M3_{D_9}$ ,  $En_{D_9}$ , and  $WE$ ) obtained with FCBF, in the cross validation set. These classifiers showed high diagnostic performance, outperforming all the extracted features in terms of Sp, PPV, LR+, and Acc. SVM achieved the highest accuracy ( $84.0 \pm 5.2\%$  Acc, with  $71.9 \pm 4.4\%$  Se and  $91.1 \pm 7.2\%$  Sp) for the cutoff of 5 e/h.

## 5. Discussion

In the present study, we examined the usefulness of wavelet analysis to identify features that characterize oximetry dynamics in order to expedite detection of moderate-to-severe pediatric SAHS.  $WE$  and features from the coefficients in  $D_9$  ( $M1_{D_9}$ - $M4_{D_9}$ ,  $Max_{D_9}$ , and  $En_{D_9}$ ) were obtained from the DWT of each  $SpO_2$  recording.  $D_9$  (0.0244–0.0488 Hz) was chosen according to a previous study in the context of pediatric SAHS (Vaquerizo-Villar et al., 2018), and is related to the duration and frequency of the  $SpO_2$  desaturations associated with apneic events (Taha et al., 1997). Statistically significant differences ( $p$ -value  $< 0.01$ ) emerged in all DWT-derived features between the groups AHI  $< 5$  e/h and AHI  $\geq 5$  e/h in the optimization set (Table 2). The higher values showed by  $M1_{D_9}$ ,  $Max_{D_9}$ , and  $En_{D_9}$  in the AHI  $\geq 5$  e/h group agree with a higher amplitude of the histogram for high values of the  $D_9$  coefficients in this group. In addition, the  $SpO_2$  drops and rises caused by apneic events

**Table 3:** Diagnostic ability of the proposed features (*ODI3*, statistical moments, PSD, and DWT) in the cross-validation set.

| Feature                  | Se        | Sp        | PPV       | NPV      | LR+     | LR-       | Acc      |
|--------------------------|-----------|-----------|-----------|----------|---------|-----------|----------|
| <i>ODI3</i>              | 78.1±7.3  | 84.2±8.1  | 75.2±10.2 | 86.5±5.0 | 6.1±2.9 | 0.27±0.11 | 81.9±7.2 |
| <i>M1<sub>T</sub></i>    | 62.3±6.8  | 65.0±2.6  | 51.4±2.1  | 74.6±3.6 | 1.8±0.2 | 0.58±0.10 | 64.0±2.3 |
| <i>M2<sub>T</sub></i>    | 72.6±13.6 | 67.1±6.6  | 56.7±2.8  | 81.2±6.6 | 2.2±0.3 | 0.40±0.17 | 69.2±3.1 |
| <i>M3<sub>T</sub></i>    | 65.0±8.5  | 61.4±6.8  | 50.1±2.8  | 74.9±2.8 | 1.7±0.2 | 0.57±0.09 | 62.7±2.7 |
| <i>M4<sub>T</sub></i>    | 60.9±15.6 | 49.9±8.4  | 41.6±5.0  | 69.0±7.5 | 1.2±0.3 | 0.78±0.26 | 54.0±5.2 |
| <i>M1<sub>PSD</sub></i>  | 75.5±7.9  | 82.5±7.4  | 73.0±8.5  | 85.1±3.5 | 5.3±3.1 | 0.30±0.08 | 79.9±3.8 |
| <i>M2<sub>PSD</sub></i>  | 69.8±7.3  | 83.4±5.2  | 71.8±6.2  | 82.5±3.0 | 4.5±1.4 | 0.36±0.08 | 78.3±3.2 |
| <i>M3<sub>PSD</sub></i>  | 47.2±11.7 | 58.1±11.9 | 40.4±4.1  | 65.0±2.8 | 1.2±0.2 | 0.91±0.12 | 54.1±4.5 |
| <i>M4<sub>PSD</sub></i>  | 63.6±8.3  | 47.1±6.2  | 41.7±4.2  | 68.7±6.1 | 1.2±0.2 | 0.79±0.23 | 53.3±5.0 |
| <i>Max<sub>PSD</sub></i> | 78.1±8.8  | 75.2±9.9  | 66.2±6.9  | 85.6±3.6 | 3.5±1.1 | 0.29±0.09 | 76.3±4.3 |
| <i>SE<sub>PSD</sub></i>  | 48.6±14.4 | 61.8±11.8 | 43.0±4.8  | 67.3±3.3 | 1.3±0.3 | 0.82±0.12 | 56.9±4.2 |
| <i>M1<sub>D9</sub></i>   | 75.4±9.1  | 82.6±7.8  | 72.2±10.2 | 84.0±5.1 | 5.2±2.7 | 0.32±0.12 | 79.1±6.2 |
| <i>M2<sub>D9</sub></i>   | 74.7±6.1  | 81.7±6.5  | 71.5±6.9  | 84.6±3.0 | 4.6±1.7 | 0.31±0.07 | 79.1±3.2 |
| <i>M3<sub>D9</sub></i>   | 58.3±9.2  | 63.4±6.5  | 48.7±3.1  | 72.1±3.3 | 1.6±0.2 | 0.66±0.10 | 61.5±3.2 |
| <i>M4<sub>D9</sub></i>   | 71.2±6.7  | 64.6±5.7  | 54.6±3.3  | 79.2±4.0 | 2.0±0.3 | 0.45±0.10 | 67.1±3.5 |
| <i>Max<sub>D9</sub></i>  | 75.4±7.1  | 85.4±6.8  | 76.0±9.0  | 85.4±4.3 | 6.2±2.8 | 0.29±0.10 | 81.7±5.6 |
| <i>En<sub>D9</sub></i>   | 78.8±4.4  | 81.7±5.2  | 72.2±5.5  | 86.7±2.4 | 4.6±1.4 | 0.26±0.05 | 80.6±3.4 |
| <i>WE</i>                | 76.0±8.2  | 78.4±5.6  | 68.0±3.8  | 84.9±3.5 | 3.6±0.7 | 0.30±0.09 | 77.6±2.5 |

**Table 4:** Diagnostic ability of the LR, SVM, and MLP models in the cross-validation set.

| Feature    | Se       | Sp       | PPV       | NPV      | LR+       | LR-       | Acc      |
|------------|----------|----------|-----------|----------|-----------|-----------|----------|
| <b>LR</b>  | 72.6±4.7 | 90.2±6.2 | 82.3±8.8  | 84.7±2.8 | 9.8±5.5   | 0.31±0.06 | 83.7±4.9 |
| <b>SVM</b> | 71.9±4.4 | 91.1±7.2 | 83.8±10.8 | 84.5±2.6 | 14.6±12.9 | 0.31±0.06 | 84.0±5.2 |
| <b>MLP</b> | 73.3±6.6 | 89.0±6.9 | 80.7±9.2  | 84.9±3.3 | 9.0±5.8   | 0.30±0.08 | 83.2±5.2 |

are reflected in a higher dispersion in the histogram of  $D_9$  coefficients, as reported by the higher values of  $M2_{D_9}$  in the  $AHI \geq 5$  e/h group. In contrast, the lower values that  $M3_{D_9}$  and  $M4_{D_9}$  as reflected in the  $AHI \geq 5$  e/h group indicate that the variations produced in the  $SpO_2$  signal due to apneic events result in values less proximal to zero in the histogram of the  $D_9$  coefficients. Finally, the higher irregularity reported by *WE* in the SAHS positive group suggests that apnea-hypopnea events alter the energy distribution of the whole DWT profile of the  $SpO_2$  signal.

Regarding the diagnostic performance of the proposed features, *ODI3* and *Max<sub>D9</sub>* reached similar Acc in the cross-validation set, higher than the remaining features. In addition, higher accuracies were generally obtained with the DWT-derived features with respect to statistical moments and features from PSD. This suggests that DWT is a useful approach to analyze the changes produced in the  $SpO_2$  signal associated to SAHS. In the feature selection stage, a feature subset composed of *ODI3* (conventional oximetric index); *M2<sub>T</sub>* (time); *M2<sub>PSD</sub>*, *M3<sub>PSD</sub>*, and *Max<sub>PSD</sub>* (PSD), and *M3<sub>D9</sub>*, *En<sub>D9</sub>* y *WE* (DWT) was obtained with FCBF. LR, SVM, and MLP models built with this subset obtained high diagnostic performance for the detection of moderate-to-severe SAHS ( $AHI \geq 5$  e/h), improving the diagnostic ability of the single features (Table 3) in terms of Sp, PPV, LR+, and Acc. It is worthy to note that the SVM model achieved the highest average Acc (84.0%), Sp (91.1%), PPV (83.8%), and LR+ (14.6) among the single features and binary classifiers. In addition, SVM reached similar NPV and LR- to LR, MLP, *ODI3*

and the remaining features. A high LR+ is especially important for screening tests (Álvarez et al., 2017; Deeks and Altman, 2004). In this sense, a LR+ greater than 10 is considered to provide strong evidence to confirm diagnoses (Deeks and Altman, 2004). Thus, our method is especially useful to confirm the presence of pediatric SAHS.

Three DWT features were involved in the feature subset obtained with FCBF: *M3<sub>D9</sub>*, *En<sub>D9</sub>* and *WE*. As aforementioned, these features provide information about the concentration of the  $D_9$  coefficients near zero (*M3<sub>D9</sub>*), the amplitude of the  $D_9$  coefficients (*En<sub>D9</sub>*), and the irregularity of the distribution of the whole DWT profile of the  $SpO_2$  signal (*WE*). According to our results, *M3<sub>D9</sub>*, *En<sub>D9</sub>* and *WE* provide both relevant and complementary (non-redundant) information on the changes occurring in the  $SpO_2$  signal due to SAHS. This is consistent with the different properties of the  $SpO_2$  signal these DWT-derived features quantify. The fact that a high performance was reached with the three classification algorithms reinforces the notion that DWT is a useful method to analyze the  $SpO_2$  signal in the context of pediatric SAHS.

To the best of our knowledge, this is the first study assessing wavelet analysis of  $SpO_2$  recordings in the context of pediatric SAHS. Our results suggest that DWT is an appropriate tool to analyze the low frequency components of the  $SpO_2$  signals related to the duration of the desaturations caused by apnea-hypopnea events since it provides high resolution at low frequencies of the power spectrum (Rioul and Vetterli, 1991; Daubechies, 1990).

This assumption is further supported by previous studies, whereby DWT was also applied to quantify the frequency components of different biomedical signals associated to respiratory events in the context of adult SAHS (Khandoker et al., 2008; Tagluk and Sezgin, 2011). Additionally, the favorable performance of our approach may be due to the suitability of the WT to analyze non-stationary properties of a signal (Rioul and Vetterli, 1991; Daubechies, 1990), which is appropriate to events such as the non-stationary changes of the SpO<sub>2</sub> signal associated with apneic events. The high resolution afforded by WT at low frequencies, as well as its suitability to analyze non-stationary signals clearly support the contention that DWT is more appropriate than conventional spectral analysis techniques to analyze the SpO<sub>2</sub> signal (Rioul and Vetterli, 1991; Daubechies, 1990).

Table 5 shows the performance of previous studies focused on the automated analysis of SpO<sub>2</sub> as an alternative to PSG in the screening of moderate-to-severe pediatric SAHS (Kirk et al., 2003; Tsai et al., 2013; Chang et al., 2013; Villa et al., 2015; Álvarez et al., 2017; Vaquerizo-Villar et al., 2018; Hornero et al., 2017). Oxygen desaturation index and clusters of desaturations have been employed for this task (Kirk et al., 2003; Tsai et al., 2013; Chang et al., 2013; Villa et al., 2015). Kirk et al. (2003) applied *ODI3*, reaching 67.0% Se, 60.0% Sp, and 64.0% Acc. Tsai et al. (2013) obtained 83.8% Se, 86.5% Sp, and 85.1% Acc using 4% *ODI* (*ODI4*). However, *ODI4* cut-off values were optimized and validated using the same population, such that no true post-hoc verification was achieved. Chang et al. (2013) combined *ODI3* with common symptoms to assess a discriminative score, reaching 60% Se, 86% Sp, and 72% Acc. Villa et al. (2015) reported 69.4% Acc (40.6% Se and 97.9% Sp) combining clusters of desaturations and clinical history in a discriminative score. Our approach achieved a high the diagnostic performance while also strengthening its validity since the methods were derived using not only a much larger sample size, but also applying a cross validation approach to validate the results.

In order to increase the diagnostic ability of the SpO<sub>2</sub> signal, conventional oximetric indices have been combined with features from other signal processing approaches in studies developed by our group (Álvarez et al., 2017; Vaquerizo-Villar et al., 2018; Hornero et al., 2017). Álvarez et al. (2017) assessed LR models fed with conventional oximetric indices, statistical parameters, PSD, and nonlinear features. These models were validated using a bootstrap procedure, reaching 82.8% Acc (82.2% Se and 83.6% Sp). Vaquerizo-Villar et al. (2018) assessed the usefulness of oximetry bispectrum. A multiclass multi-layer perceptron (MLP) model fed with *ODI3*, anthropometrical variables, PSD, and bispectral features reached 61.8% Se, 97.6% Sp, and 81.3% Acc in an independent test set, outperforming a MLP classifier built without bispectral features. Finally, Hornero et al. (2017) analyzed 4,191 SpO<sub>2</sub> recordings obtained from 13 sleep laboratories in a

multicenter international study. A MLP regression model with *ODI3* and the skewness of the PSD reached 68.2% Se, 87.2% Sp, and 81.7% Acc. In contrast with the findings of these studies, our current results achieved improved diagnostic ability for the screening of moderate-to-severe SAHS with the use of DWT-derived features. This suggests that wavelet analysis could enhance the detection of this clinically important and vulnerable group of SAHS severity from single-channel oximetry recordings. In these patients, it is essential to early detect this condition, since they are more likely to suffer from morbidities such as decreases in cognitive performance (Kaditis et al., 2016a; Hunter et al., 2016), as well as an increased C-reactive protein level due to systemic inflammation (Church, 2012). Moreover, an AHI > 5 e/h is also associated with increased systemic blood pressure measurements and an increased risk for cardiac strain (Kaditis et al., 2016a). All these important negative consequences highlight the necessity of an early detection of moderate-to-severe pediatric SAHS (Kaditis et al., 2016a).

Notwithstanding the highly promising results of our current approach, several limitations must be considered. First, the exclusive use of the SpO<sub>2</sub> signal to detect SAHS may restrict the spectrum of physiological perturbations being detected by the oximetry signal, such as electroencephalographic arousals or reductions in airflow and increased intrathoracic pressure swings (Marcus et al., 2012). In this regard, the combination of SpO<sub>2</sub> with other physiological signals from PSG could potentially enhance the performance of our proposed method but at the cost of adding significant complexity to the test. In addition, future research efforts may prospectively focus on identifying a specific mother wavelet for this task. However, our proposed approach achieved high performance with the Haar's mother wavelet. Of note, the lack of universally accepted AHI severity cutoffs is another limitation that affects our study. Nevertheless, we have assessed the diagnostic ability of our proposal using an AHI cutoff of 5 e/h, a widely used criterion in the clinical decision making leading to the recommendation of surgical treatment (Kaditis et al., 2016a,b). Finally, it would be an interesting future goal to further validate our methodology in a larger sample of unattended oximetry recordings obtained at patients' homes.

## 6. Conclusions

The application of WT has enabled the identification of features with the ability to characterize the effects of SAHS in the overnight oximetry profile of children. Features computed in the  $D_9$  detail level of the DWT as well as  $WE$  reached significant differences associated with the presence of SAHS. DWT has been found to provide complementary information to conventional approaches. Additionally, high diagnostic performance was reached using different reference binary classifiers, which emphasizes the usefulness of the DWT to provide discriminant

**Table 5:** Summary of the state-of-the-art studies in the context of detection of moderate-to-severe pediatric SAHS using SpO<sub>2</sub> recordings.

| Study                          | Subjects(n) | Methods  | Validation                         | Se    | Sp    | Acc   |
|--------------------------------|-------------|--|------------------------------------|-------|-------|-------|
| Kirk et al. (2003)             | 58          | ODI3   | Direct validation**                | 67    | 60    | 64*   |
| Tsai et al. (2013)             | 148         | ODI4   | No                                 | 83.8  | 86.5  | 85.1* |
| Chang et al. (2013)            | 141         | ODI3 and symptoms  | Direct validation**                | 60    | 86    | 72*   |
| Villa et al. (2015)            | 268         | Clusters of desaturations and clinical history                           | Direct validation**                | 40.6* | 97.9* | 69.4* |
| Álvarez et al. (2017)          | 50          | Statistical moments, spectral, nonlinear features, and classical indices | Bootstrap 0.652                    | 82.2  | 83.6  | 82.8  |
| Vaquerizo-Villar et al. (2018) | 298         | Bispectrum, PSD, ODI3, anthropometric variables                          | Feature optimization-training-test | 61.8  | 97.6  | 81.3  |
| Hornero et al. (2017)          | 4191        | Statistical moments, PSD, nonlinear features, and ODI3                   | Training-test                      | 68.2  | 87.2  | 81.7  |
| Our proposal                   | 981         | ODI3, Statistical moments, PSD, and DWT features                         | Optimization- cross validation     | 71.9  | 91.1  | 84.0  |

\*Computed from reported data, \*\* Direct validation of a scoring criteria against AHI from PSG.

information from oximetry signals. These results suggest that wavelet analysis could be useful to further characterize the oximetry signal and improve the diagnostic performance and implementation of abbreviated screening test for pediatric SAHS.

### Supporting information

**S1 Table** Actual AHI from PSG and values of all the extracted features from the SpO<sub>2</sub> signal (ODI3, statistical moments, PSD, and DWT features) of each subject in the optimization set. (XLSX)

**S2 Table** The number of times each feature was selected with FCBF in the optimization set. (XLSX)

**S3 Table** Actual AHI from PSG, values of all the extracted features from the SpO<sub>2</sub> signal (ODI3, statistical moments, PSD, and DWT features), and the classification scores of LR, SVM, and MLP of each subject in the cross-validation set. (XLSX)

### Competing interests

The authors have declared that no competing interests exist.

### Copyright

©2018 Vaquerizo-Villar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Data availability

All relevant data are within the paper and its Supporting Information files.

### Authorship contributions

**Funding acquisition:** Félix del Campo, Roberto Hornero.

**Investigation:** Fernando Vaquerizo-Villar, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero.

**Methodology:** Fernando Vaquerizo-Villar, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero.

**Project administration:** David Gozal, Roberto Hornero.

**Software:** Fernando Vaquerizo-Villar, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Félix del Campo.

**Supervision:** Daniel Álvarez, Roberto Hornero.

**Validation:** Fernando Vaquerizo-Villar, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal.

**Writing – original draft:** Fernando Vaquerizo-Villar, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero.

### Funding

This work was supported by 'Agencia Estatal de Investigación del Ministerio de Ciencia, Innovación y Universidades' and 'European Regional Development Fund (FEDER)' under projects DPI2017-84280-R, RTC-2015-3446-1, and 0378\_AD\_EEGWA\_2\_P, by 'Consejería de Educación de la Junta de Castilla y León and FEDER' under project VA037U16, and by 'European Commission' and 'FEDER' under project 'Análisis y correlación entre

el genoma completo y la actividad cerebral para la ayuda en el diagnóstico de la enfermedad de Alzheimer' ('Cooperation Programme Interreg V-A Spain-Portugal POCTEP 2014–2020'). F. Vaquerizo-Villar was in receipt of a 'Ayuda para contratos predoctorales para la Formación de Profesorado Universitario (FPU)' grant from the 'Ministerio de Educación, Cultura y Deporte' (FPU16/02938). V. Barroso-García was in a receipt of a 'Ayuda para financiar la contratación predoctoral de personal investigador' grant from the 'Consejería de Educación de la Junta de Castilla y León' and the European Social Fund. D. Álvarez was in receipt of a 'Juan de la Cierva' grant from MINECO (IJCI-2014-22664). L. Kheirandish-Gozal was supported by National Institutes of Health (NIH) grant HL130984 and D. Gozal by NIH grant HL-65270.

## References

- Alonso-Álvarez, M. L., Canet, T., Cubell-Alarco, M., Estivill, E., Fernández, E., Gozal, D., Jurado-Luque, M. J., Lluich-Roselló, M. A., Martínez-Pérez, F., Merino-Andren, M., Pin-Arboledas, G., Roure, N., Sanmartí, F. X., Sans-Capdevila, O., Segarra-Isern, J., 2011. Documento de consenso del síndrome de apneas-hipopneas durante el sueño en niños. *Arch Bronconeumol* 47 (Supl 5), 2–18.
- Álvarez, D., Alonso-Álvarez, M. L., Gutiérrez-Tobal, G. C., Crespo, A., Kheirandish-Gozal, L., Gozal, D., Terán-Santos, J., Campo, F. D., 2017. Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal Oximetry: An Alternative to Respiratory Polysomnography in Unattended Settings. *J Clin Sleep Med* 13 (5), 7–11.
- Álvarez, D., Hornero, R., Marcos, J. V., Wessel, N., Penzel, T., Glos, M., Del Campo, F., 2015. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis. *International journal of neural systems* 23 (05), 1350020.
- Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., Marcus, C. L., Mehra, R., Parthasarathy, S., Quan, S. F., Redline, S., Strohl, K. P., Davidson Ward, S. L., Tangredi, M. M., 2012. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. *Journal of clinical sleep medicine* 8 (5), 597.
- Bishop, C. M., 2006. Pattern recognition and machine learning. Springer.
- Bishop, C. M., et al., 1995. Neural networks for pattern recognition. Oxford university press.
- Brockmann, P. E., Schaefer, C., Poets, A., Poets, C. F., Urschitz, M. S., 2013. Diagnosis of obstructive sleep apnea in children: a systematic review. *Sleep medicine reviews* 17 (5), 331–340.
- Chang, L., Wu, J., Cao, L., 2013. Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea. *International Journal of Pediatric Otorhinolaryngology* 77 (3), 365–371.
- Church, G. D., 2012. The Role of Polysomnography in Diagnosing and Treating Obstructive Sleep Apnea in Pediatric Patients. *Current problems in pediatric and adolescent health care* 42 (1), 22–25.
- Cvetkovic, D., Übeyli, E. D., Cosic, I., 2008. Wavelet transform feature extraction from human ppg, ecg, and eeg signal responses to elf pemf exposures: A pilot study. *Digital signal processing* 18 (5), 861–874.
- Daubechies, I., 1990. The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE transactions on information theory* 36 (5), 961–1005.
- Deeks, J. J., Altman, D. G., 2004. Diagnostic tests 4: likelihood ratios. *Bmj* 329 (7458), 168–169.
- Efron, B., Tibshirani, R. J., 1994. An introduction to the bootstrap. CRC press.
- Fontela-Romero, O., Guijarro-Berdinas, B., Alonso-Betzanos, A., Moret-Bonillo, V., 2005. A new method for sleep apnea classification using wavelets and feedforward neural networks. *Artificial Intelligence in Medicine* 34 (1), 65–76.
- Garde, A., Dehkordi, P., Karlen, W., Wensley, D., Ansermino, J. M., Dumont, G. A., 2014a. Development of a screening tool for sleep disordered breathing in children using the phone oximeter™. *PLoS one* 9 (11), e112959.
- Garde, A., Karlen, W., Dehkordi, P., Ansermino, J. M., Dumont, G. A., 2014b. Oxygen saturation resolution influences regularity measurements. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 2257–2260.
- Guyon, I., Elisseeff, A., 2005. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research (JMLR)* 3 (5), 1157–1182.
- Hornero, R., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Philby, M. F., Alonso-Álvarez, M. L., Álvarez, D., Dayyat, E. A., Xu, Z., Huang, Y.-S., Tamae Kakazu, M., Li, A. M., Van Eyck, A., Brockmann, P. E., Ehsan, Z., Simakajornboon, N., Kaditis, A. G., Vaquerizo-Villar, F., Crespo Sedano, A., Sans Capdevila, O., von Lukowicz, M., Terán-Santos, J., Del Campo, F., Poets, C. F., Ferreira, R., Bertran, K., Zhang, Y., Schuen, J., Verhulst, S., Gozal, D., 2017. Nocturnal Oximetry-based Evaluation of Habitually Snoring Children. *American Journal of Respiratory and Critical Care Medicine* 196 (12), 1591–1598.
- Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. *Neural networks* 4 (2), 251–257.
- Hosmer, D., Lemeshow, S., 2004. Applied logistic regression. John Wiley & Sons.
- Hunter, S. J., Gozal, D., Smith, D. L., Philby, M. F., Kaylegian, J., Kheirandish-Gozal, L., 2016. Effect of sleep-disordered breathing severity on cognitive performance measures in a large community cohort of young school-aged children. *American journal of respiratory and critical care medicine* 194 (6), 739–747.
- Kaditis, A., Kheirandish-Gozal, L., Gozal, D., 2016a. Pediatric OSAS: Oximetry can provide answers when polysomnography is not available. *Sleep Medicine Reviews* 27, 96–105.
- Kaditis, A. G., Alvarez, M. L. A., Boudewyns, A., Alexopoulos, E. I., Ersu, R., Joosten, K., Larramona, H., Miano, S., Narang, I., Trang, H., Tsaoussoglou, M., Vandenbussche, N., Villa, M. P., Waardenburg, D. V., Weber, S., Verhulst, S., 2016b. Obstructive sleep disordered breathing in 2- to 18-year-old children: Diagnosis and management. *European Respiratory Journal* 47 (1), 69–94.
- Katz, E. S., Mitchell, R. B., Ambrosio, C. M. D., 2012. Obstructive Sleep Apnea in Infants. *American journal of respiratory and critical care medicine* 185 (8), 805–816.
- Khandoker, A. H., Palaniswami, M., Karmakar, C. K., 2008. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings. *IEEE transactions on information technology in biomedicine* 13 (1), 37–48.
- Kirk, V. G., Bohn, S. G., Flemons, W. W., Remmers, J. E., 2005. Comparison of Home Oximetry Monitoring With Laboratory Polysomnography in Children. *Chest* 124 (5), 1702–1708.
- Lee, Y., Bister, M., Blanchfield, P., Salleh, Y., 2004. Automated detection of obstructive apnea and hypopnea events from oxygen saturation signal. In: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 1. IEEE, pp. 321–324.
- Lin, R., Lee, R.-G., Tseng, C.-L., Zhou, H.-K., Chao, C.-F., Jiang, J.-A., 2006. A new approach for identifying sleep apnea syndrome using wavelet transform and neural networks. *Biomedical Engineering: Applications, Basis and Communications* 18 (05), 138–143.
- Magalang, U. J., Dmochowski, J., Veeramachaneni, S., Draw, A., Mador, J., El-Solh, A., Grant, B. J. B., 2003. Prediction of the Apnea-Hypopnea Index From Overnight Pulse Oximetry. *CHEST Journal* 124 (5), 1694–1701.
- Marcus, C. L., Brooks, L. J., Ward, S. D., Draper, K. A., Gozal, D., Halbower, A. C., Jones, J., Lehmann, C., Schechter, M. S., Sheldon, S., Shiffman, R. N., Spruyt, K., 2010. Automatic screening of obstructive sleep apnea from the ecg based on empirical mode decomposition

- and wavelet analysis. *Physiological measurement* 31 (5), 275.
- Nixon, G. M., Kermack, A. S., Davis, G. M., Manoukian, J. J., Brown, A., Brouillette, R. T., 2004. Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry. *Pediatrics* 113 (1), e19–e25.
- Rioul, O., Vetterli, M., 1991. Wavelets and signal processing. *IEEE signal processing magazine* 8 (4), 14–38.
- Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schürmann, M., Başar, E., 2001. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of neuroscience methods* 105 (1), 65–75.
- Sedano, A. C., Vaquerizo-Villar, F., Álvarez, D., Gutiérrez-Tobal, G. C., Barroso-García, V., Cerezo, A., López, G., Kheirandish-Gozal, L., Gozal, D., Hornero, R., et al., 2017. Automated detection of childhood sleep apnea using discrete wavelet transform of nocturnal oximetry and anthropometric variables.
- Steyerberg, E. W., Vergouwe, Y., 2014. Towards better clinical prediction models: seven steps for development and an abcd for validation. *European heart journal* 35 (29), 1925–1931.
- Tagluk, M. E., Sezgin, N., 2011. A new approach for estimation of obstructive sleep apnea syndrome. *Expert Systems with Applications* 38 (5), 5546–5551.
- Taha, B. H., Dempsey, J. A., Weber, S. M., Badr, M. S., Skatrud, J. B., Young, T. B., Jacques, A. J., Seow, K., 1997. Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep* 20 (11), 991–1001.
- Tan, H.-L., Gozal, D., Ramirez, H. M., Bandla, H. P. R., Kheirandish-Gozal, L., 2014. Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea. *Sleep* 37 (2), 255–260.
- Tsai, C.-M., Kang, C.-H., Su, M.-C., Lin, H.-C., Huang, E.-Y., Chen, C.-C., Hung, J.-C., Niu, C.-K., Liao, D.-L., Yu, H.-R., 2015. Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. *International journal of pediatric otorhinolaryngology* 77 (8), 1286–1290.
- Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., del Campo, F., Gozal, D., Hornero, R., 2018. Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings. *Computer Methods and Programs in Biomedicine* 156, 141–149.
- Vaquerizo-Villar, F., Gutiérrez-Tobal, G. C., Barroso-García, V., Kheirandish-Gozal, L., Crespo, A., del Campo, F., Gozal, D., Hornero, R., et al., 2017. Usefulness of discrete wavelet transform in the analysis of oximetry signals to assist in childhood sleep apnea-hypopnea syndrome diagnosis. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 3753–3756.
- Villa, M. P., Pietropaoli, N., Supino, M. C., Vitelli, O., Rabasco, J., Evangelisti, M., Del Pozzo, M., Kaditis, A. G., 2015. Diagnosis of pediatric obstructive sleep apnea syndrome in settings with limited resources. *JAMA Otolaryngology-Head & Neck Surgery* 141 (11), 990–996.
- Welch, P., 1967. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* 15 (2), 70–75.
- Witten, I., Frank, E., Hall, M., 2011. *Data Mining*. Morgan Kaufmann.
- Yu, L., Liu, H., 2004. Efficient Feature Selection via Analysis of Relevance and Redundancy. *Journal of Machine Learning Research* 5, 1205–1224.

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

Fernando Vaquerizo-Villar<sup>a</sup>, Daniel Álvarez-González<sup>a,b</sup>, Leila Kheirandish-Gozal<sup>c</sup>, Gonzalo C. Gutiérrez-Tobal<sup>a</sup>, Verónica Barroso-García<sup>a</sup>, Andrea Crespo<sup>a,b</sup>, Félix del Campo<sup>a,b</sup>, David Gozal<sup>c</sup>, Roberto Hornero<sup>a,d</sup>

<sup>a</sup> Biomedical Engineering Group, Universidad de Valladolid, Valladolid, Spain

<sup>b</sup> Servicio de Neumología, Hospital Universitario Río Hortega, Valladolid, Spain

<sup>c</sup> Department of Child Health, The University of Missouri School of Medicine, Columbia, MO, United States of America

<sup>d</sup> IMUVA, Instituto de Investigación en Matemáticas, Universidad de Valladolid, Valladolid, Spain

### Physiological Measurement

Volume 39, November 2018, Pages 11406,

### Abstract

**Objective:** To evaluate whether detrended fluctuation analysis (DFA) provides information that improves the diagnostic ability of the oximetry signal in the diagnosis of paediatric sleep apnoea–hypopnoea syndrome (SAHS). **Approach:** A database composed of 981 blood oxygen saturation (SpO<sub>2</sub>) recordings in children was used to extract DFA-derived features in order to quantify the scaling behaviour and the fluctuations of the SpO<sub>2</sub> signal. The 3% oxygen desaturation index (ODI3) was also computed for each subject. Fast correlation-based filter (FCBF) was then applied to select an optimum subset of relevant and non-redundant features. This subset fed a multi-layer perceptron (MLP) neural network to estimate the apnoea–hypopnoea index (AHI). **Main results:** ODI3 and four features from the DFA reached significant differences associated with the severity of SAHS. An optimum subset composed of the slope in the first scaling region of the DFA profile and the ODI3 was selected using FCBF applied to the training set (60% of samples). The MLP model trained with this feature subset showed good agreement with the actual AHI, reaching an intra-class correlation coefficient of 0.891 in the test set (40% of samples). Furthermore, the estimated AHI showed high diagnostic ability, reaching an accuracy of 82.7%, 81.9%, and 91.1% using three common AHI cut-offs of 1, 5, and 10 events per hour (e/h), respectively. These results outperformed the overall performance of ODI3. **Significance:** DFA may serve as a reliable tool to improve the diagnostic performance of oximetry recordings in the evaluation of paediatric patients with symptoms suggestive of SAHS.

**Keywords:** blood oxygen saturation (SpO<sub>2</sub>), detrended fluctuation analysis (DFA), feature selection, apnoea–hypopnoea index (AHI) estimation, paediatric sleep apnoea–hypopnoea syndrome (SAHS)

### 1. Introduction

Childhood sleep apnoea–hypopnoea syndrome (SAHS) is a breathing disorder whereby paediatric subjects manifest recurrent episodes of either complete cessation (apnoea) or significant reductions (hypopnoea) of airflow while sleeping (Marcus et al., 2012). Paediatric SAHS has become a major health problem due to its high prevalence and negative effects. SAHS has an estimated prevalence

in the range of 1%–5% in the general paediatric population (Marcus et al., 2012). In addition, cognitive deficits, behavioural abnormalities, daytime sleepiness, cardiac and metabolic derangements, and systemic inflammation are all morbid consequences that adversely affect the optimal development of children affected by SAHS (Marcus et al., 2012).

Based on the aforementioned considerations, an early diagnosis of paediatric SAHS is vital. The gold standard diagnostic approach to childhood SAHS is overnight polysomnography (PSG) (Marcus et al., 2012). It requires patients to spend the night in a specialised sleep laboratory while being recorded for a wide range of biomedical signals, including electrocardiogram, electroencephalogram, electrooculogram, submental and leg electromyogram, oronasal airflow, and blood oxygen saturation (SpO<sub>2</sub>) by pulse oximetry (Kaditis et al., 2016; Alonso-

Email addresses: fernando.vaquerizo@gib.tel.uva.es (Fernando Vaquerizo-Villar<sup>a</sup>), dalvgon@gmail.com (Daniel Álvarez-González<sup>a,b</sup>), lgozal@peds.bsdu.uchicago.edu (Leila Kheirandish-Gozal<sup>c</sup>), gguttob@gmail.com (Gonzalo C. Gutiérrez-Tobal<sup>a</sup>), veronica.barroso@gib.tel.uva.es (Verónica Barroso-García<sup>a</sup>), acre-spos@saludcastillayleon.es (Andrea Crespo<sup>a,b</sup>), fsas@telefonica.net (Félix del Campo<sup>a,b</sup>), dgozal@peds.bsdu.uchicago.edu (David Gozal<sup>c</sup>), robhor@tel.uva.es (Roberto Hornero<sup>a,d</sup>)

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

Considering the inherent disadvantages and limitations of PSG, along with the need for an early and timely diagnosis of SAHS, the search for simplified alternative techniques has emerged in recent years. In this regard, one common approach consists of the automated analysis of a reduced subset of cardiorespiratory signals that is normally included in the overnight PSG. One of these alternatives is nocturnal pulse oximetry (NPO), which records the blood oxygen saturation signal ( $\text{SpO}_2$ ) with a pulse oximeter probe, usually placed on a finger (Netzer et al., 2001). NPO can be readily performed without the need for professional supervision in the patient's home and is widely available, as reflected by the large number of commercially available portable pulse oximeters (Nixon et al., 2004; Garde et al., 2014). Thus, NPO is a technically simple test for children, and the  $\text{SpO}_2$  signal from NPO provides moment-to-moment oxygen content in haemoglobin (McClatchey, 2002), a signal that contains essential information about the apnoeic events from SAHS, since these events induce recurrent decreases in blood oxygen levels, otherwise termed oxygen desaturations (Berry et al., 2012).

Previous studies have examined the  $\text{SpO}_2$  signal as a potential alternative to PSG in the screening of paediatric SAHS. These studies employed different signal processing techniques (Kirk et al., 2003; Tsai et al., 2013; Garde et al., 2014; Van Eyck et al., 2015; Álvarez et al., 2017; Crespo et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018), and more specifically, conventional oximetry indices, common statistics, frequency domain analysis techniques, and nonlinear methods. Among these approaches, nonlinear parameters proved useful to characterise the oxygen desaturations caused by apnoeic events in adults and children. However, a recent study using a very large database of 4191 paediatric subject recordings showed that traditional nonlinear metrics (central tendency measure, Lempel–Ziv complexity and sample entropy) were redundant with respect to the 3% oxygen desaturation index (*ODI3*) (Hornero et al., 2017), an oximetry index commonly used in the clinical practice for simplified screening purposes. Therefore, additional research is needed to find alternative and better performant nonlinear methods that may provide further insights into the properties of the oximetry signal and allow for the extraction of additional information to that provided by *ODI3*. In this regard, detrended fluctuation analysis (DFA) is a nonlinear analysis technique

widely used to detect the correlation properties of a non-stationary signal (Peng et al., 1994, 1995). DFA computes the logarithm of the fluctuation function of a time series versus the logarithm of a window time length (scale). DFA provides a quantitative parameter, the scaling exponent ( $\alpha$ ), which measures the linear relationship between the fluctuation function and the scale (Peng et al., 1995). The variation of  $\alpha$  value for different ranges of scales (different window time lengths) identifies regions with different correlations (Peng et al., 1995). In this sense, the scaling behaviour of a signal is given by the different regions observed in the DFA profile and the value of  $\alpha$  in these regions (Peng et al., 1995). Thus, DFA is a useful tool to analyse signals with segments that modify its scaling behaviour, such as random spikes or segments which have a different local behaviour (Chen et al., 2002; Hua and Yu, 2017). Apnoeic events produce random spikes and/or irregular fluctuations in the  $\text{SpO}_2$  signal. Hence, DFA could be useful to analyse the oximetry signal in the context of SAHS.

Previous work has suggested the ability of DFA to analyse the correlation properties of physiological signals in the context of both adult and paediatric SAHS (Lee et al., 2002; Penzel et al., 2003; Dehkordi et al., 2016; Kaimakamis et al., 2016; Hua and Yu, 2017). Hua and Yu (2017) applied DFA to  $\text{SpO}_2$  signals in the context of diagnosing adult SAHS. However, no studies have focused on applying DFA to  $\text{SpO}_2$  recordings in the context of paediatric SAHS.  $\text{SpO}_2$  signal properties in children differ from those of adults. Furthermore, the frequency of events that are required to define abnormality or severity markedly differ between adults and children. In addition, scoring rules for apnoeas and hypopnoeas are also more restrictive in the case of paediatric SAHS (Berry et al., 2012). Thus, the diagnosis of SAHS in children is vastly more challenging than in adults.

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

## 2. Methods

### 2.1. Subjects and signals under study

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

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

A digital polysomnography system (Polysmith; Nihon Kohden America Inc., CA, USA) was used to monitor the childrens' sleep. SpO<sub>2</sub> recordings were obtained during overnight PSG at sampling rates of 25 Hz, 200 Hz, or 500 Hz. They were exported and processed offline. Artefacts were rejected from oximetric recordings by removing those SpO<sub>2</sub> values below 50% and sudden changes between consecutive SpO<sub>2</sub> samples faster than 4%/second (Magalang et al., 2005). Then, a non-overlapping averaging-window of 1s was applied (effective sampling rate = 1 Hz) to speed up the signal processing stage, which has been found to be appropriate to perform a multiscale analysis of the oximetry signal (Crespo et al., 2017; Hua and Yu, 2017). This window size is lower than 3s, which is the maximum averaging-time recommended by the American Academy of Sleep Medicine (AASM) (Berry et al., 2012). The resolution of the SpO<sub>2</sub> signals was set to two decimal points to ensure the resolution was the same (Hornero et al., 2017).

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

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

## 2.2. Automated signal processing

Our approach consisted of three sequential stages. First, features derived from DFA and *ODI3* were obtained from the SpO<sub>2</sub> recording of each subject. Then, a smaller subset of relevant and non-redundant features was selected using the fast correlation-based filter (FCBF) method (Yu and Liu, 2004). Finally, a multi-layer perceptron (MLP) neural network (Bishop et al., 1995) was applied to this optimum subset in order to estimate the AHI of each patient.

### 2.2.1. Detrended fluctuation analysis

DFA performs a multiscale analysis of a time series to study its correlation properties (Peng et al., 1994). The DFA profile shows changes in the correlation properties

for different ranges of scales, termed 'crossovers', which may be caused by different non-stationarities in the signal such as (Chen et al., 2002): (i) segments removed from the signal; (ii) random spikes with variable amplitude; (iii) segments with different local behaviour. Segments of the SpO<sub>2</sub> signal associated with apnoeic events typically have different statistical properties, presenting fluctuations and spikes (Crespo et al., 2017; Hua and Yu, 2017). Thus, these properties of the SpO<sub>2</sub> signal may be reflected in the DFA profile.

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

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

$$y(i) = \sum_{j=1}^i x(j) - x_{avg}, i = 1, \dots, N, \quad (1)$$

where  $x_{avg}$  is the average of the whole signal  $x(t)$ , and  $N$  is the length of the SpO<sub>2</sub> signal.

2. The integrated signal  $y(i)$  is divided into  $B$  non-overlapping windows of equal size. In the case of SpO<sub>2</sub> recordings, the minimum length of the signal is 3 h (10800 samples) to ensure there were enough sleep cycles (Berry et al., 2012). Thus, the length of each window (i.e. the scale),  $k$ , is between 3 and 1080, since the maximum box size in DFA must be one-tenth of the signal length (Chen et al., 2002).

3. For each window  $b$  ( $b = 1, \dots, B$ ), the local trend was obtained as a straight line,  $y^b$ , estimated by applying a least squares fitting to  $y(i)$ .

4. The variance of the fluctuation in each window,  $F_b^2(k)$ , is defined as follows:

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

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

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

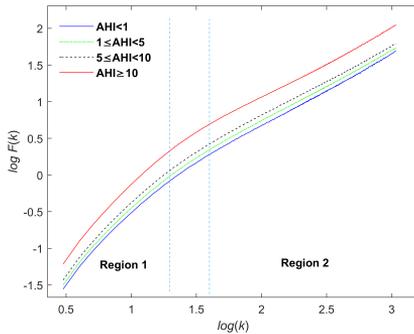
Steps 2-5 are iterated until the highest scale is used.

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

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

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

Data are presented as median [interquartile range],  $n$  or  $n(\%)$ , BMI= Body Mass Index, AHI= Apnea Hypopnea Index.

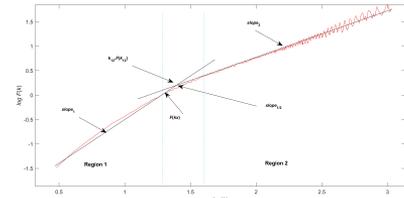


**Figure 1:** Averaged DFA profile for the four SAHS severity groups: (a) AHI < 1 e/h, (b)  $1 \leq$  AHI < 5 e/h, (c)  $5 \leq$  AHI < 10 e/h, and (d) AHI  $\leq$  10 e/h in the training set.

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

A crossover is produced in the space between these two regions of the DFA profile. Robust linear regression (Hua and Yu, 2017) was applied to estimate the line that fits both regions for each SpO<sub>2</sub> recording.

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



**Figure 2:** Illustrative example of the DFA plot of the SpO<sub>2</sub> signal of a patient from the training set.

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

It is expected that these parameters allow for the quantification of the differences in the scaling behaviour and the fluctuations of the SpO<sub>2</sub> signal are associated with the severity of SAHS, as shown in Figure 1.

### 2.2.2. Oxygen desaturation index

ODI3 was computed as the number of oxygen desaturations from the preceding baseline greater than or equal to 3% per hour of recording (Taha et al., 1997). This clinical parameter has commonly been used in the SAHS context (Kirk et al., 2003; Chang et al., 2013; Tsai et al., 2013). Higher values of ODI3 are expected in patients with a higher severity of SAHS, since oxygen desaturations are associated with apnoea or hypopnoea events (Berry et al., 2012).

### 2.3. Feature selection: FCBF

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

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

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

According to their  $SU$  value, features are ranked from the most relevant (highest  $SU$  with the AHI) to the least relevant one (lowest  $SU$  with the AHI). Different  $SU$ -based thresholds can be used to discard non-relevant features. Nevertheless, the number of features comprising our original feature set is not high. Therefore, as proposed by Yu and Liu (2004), no relevance threshold was applied to discard non-relevant features in order to maximize the relevancy of information derived from oximetry (Gutiérrez-Tobal et al., 2018; Hornero et al., 2017). In this regard, a feature that is useless by itself still may provide useful information when being selected with others (Guyon and Elisseeff, 2003). A redundancy analysis of each feature is then performed. The  $SU$  value between each pair of features ( $x_i, x_j$ ) is computed, beginning with the most relevant one (Yu and Liu 2004). When  $SU(x_i|x_j) \geq SU(x_j|y)$ , the feature  $x_j$  is considered redundant with respect to the feature  $x_i$  and discarded. In this way, an optimum subset composed of the most rele-

vant and non-redundant features is obtained (Yu and Liu, 2004).

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

### 2.4. AHI estimation: MLP neural network

MLP was applied to estimate the AHI of the subjects under study using the optimum feature subset obtained with FCBF. MLP is one of the most widely used artificial neural networks (ANNs). This ANN has already demonstrated its usefulness in the screening of paediatric SAHS diagnosis using SpO<sub>2</sub> recordings (Hornero et al., 2017). MLP is arranged in several interconnected layers (input, hidden layers, and output) composed of simple units called perceptrons or neurons (Bishop et al., 1995). Each neuron consists of an activation function  $g_i$  and adaptive weights  $w_{jk}$  representing connections with neurons from the following layer. In our case, the output layer has one neuron  $y$ , which represents the estimated AHI. Additionally, a single hidden layer configuration was implemented, since it is able to provide universal approximation to any function (Bishop et al., 1995). Thus, the output unit in our MLP architecture is calculated as follows:

$$y_k = g_t \left\{ \sum_{j=1}^{N_H} w_{jk} g_t \left\{ \sum_{i=1}^d w_{ij} x_i + b_j \right\} + b_k \right\}, \quad (5)$$

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

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

## 2.5. Statistical Analysis

Matlab R2016a (The MathWorks Inc., Natick, MA, USA) was used to implement automated signal processing algorithms, as well as to perform statistical analyses. The Kruskal–Wallis test was used to assess the statistical differences ( $p$ -value $<0.01$ ) between groups, since the extracted features did not pass the Lilliefors normality test. The Bonferroni correction was applied to deal with multiple comparisons. Both agreement between estimated AHI ( $AHI_{MLP}$ ) and actual AHI ( $AHI_{PSG}$ ), as well as agreement between  $ODI3$  and  $AHI_{PSG}$  were assessed by means of Bland–Altman plots and the intra-class correlation coefficient (ICC). Cohen’s kappa index (kappa) was used to measure the agreement between  $AHI_{MLP}$  and  $AHI_{PSG}$ , the agreement between  $ODI3$  and  $AHI_{PSG}$  to estimate the severity of SAHS (Cohen, 1960). The diagnostic ability of  $ODI3$  and  $AHI_{MLP}$  was assessed in terms of sensitivity (Se, percentage of SAHS positive patients correctly classified), specificity (Sp, percentage of SAHS negative children correctly classified), positive predictive value (PPV, proportion of subjects classified as positive that are true positives), negative predictive value (NPV, proportion of subjects classified as negative that are true negatives), positive likelihood ratio (LR+, likelihood ratio for subjects classified as positive), negative likelihood ratio (LR-, likelihood ratio for subjects classified as negative), accuracy (Acc, percentage of subjects correctly classified), and area under the ROC curve (AUC). A bootstrapping approach was employed in order to compare the ICC, kappa, overall Acc (four classes), and AUC values between  $ODI3$  and  $AHI_{MLP}$ . The number of bootstrap replicates built from the test data was set to 1000, since it ensures a proper estimation of the 95% confidence interval (Efron and Tibshirani, 1994). ICC, kappa, overall Acc, and AUC values were obtained for  $ODI3$  and  $AHI_{MLP}$  from each of these replicates. Then, the  $p$ -value between  $ODI3$  and  $AHI_{MLP}$  was computed for each of these metrics according to the Mann–Whitney  $U$  test.

## 3. Results

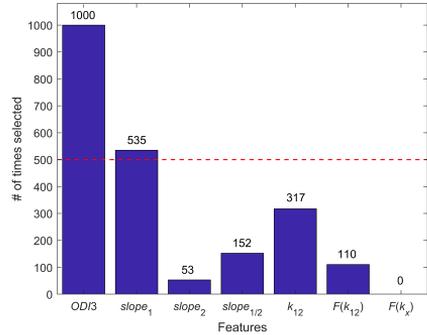
### 3.1. Training set

#### 3.1.1. Features separability

A total of seven features were obtained for each subject:  $ODI3$ , and six DFA-derived features. Table 2 shows the median and interquartile range of these features in the training set for each SAHS severity group, along with their corresponding  $p$ -values.  $ODI3$  and four out of six DFA-derived features ( $slope_1$ ,  $slope_{1/2}$ ,  $F(k_{12})$ , and  $F(k_e)$ ) showed statistically significant differences ( $p$ -value $<0.01$  after Bonferroni correction).

#### 3.1.2. Optimum feature subset

Figure 3 displays the histogram with the number of times that each feature was selected over the 1000 bootstrap replicates.  $ODI3$  was selected all the time,



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

which agrees with previous studies (Hornero et al 2017, Vaquerizo-Villar et al 2018). Additionally,  $slope_1$  was selected more than half the time (535). Thus,  $ODI3$  and  $slope_1$  were chosen as the optimum subset.

#### 3.1.3. MLP model optimisation and training

The MLP network was designed and trained using this optimum feature subset ( $ODI3$  and  $slope_1$ ). In order to find the optimum values,  $N_H$  and  $\alpha$  were varied from  $N_H=2$  up to  $N_H=30$  and  $\alpha=0$  up to  $\alpha=10$ , respectively. For each  $N_H$ - $\alpha$  pair, kappa was obtained through ten-fold cross validation. Since the network is sensitive to the initial random values of the weights, kappa was computed on the cross validation set and averaged for a total of ten runs for each  $N_H$ - $\alpha$  pair. Figure 4 shows the kappa value obtained for each  $N_H$ - $\alpha$  pair. According to this figure, the optimum values  $N_H=5$  and  $\alpha=6$  were obtained as those for which kappa was higher. Finally, the optimum feature subset ( $ODI3$  and  $slope_1$ ) from the entire training set was used to train the MLP model ( $AHI_{MLP}$ ) with these optimum user-dependent network parameters.

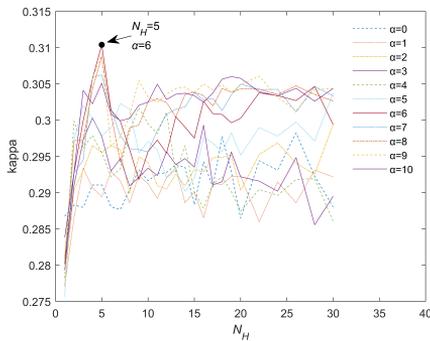
#### 3.2. Test set

Figures 5(a) and (b) show the Bland–Altman plots of  $ODI3$  and  $AHI_{MLP}$  compared with  $AHI_{PSG}$ , respectively, in the test set. ICC between  $ODI3$  and  $AHI_{MLP}$  with  $AHI_{PSG}$  is also shown.  $AHI_{MLP}$  reached a lower mean difference (bias) with  $AHI_{PSG}$  than  $ODI3$  (0.75 versus -1.65), whereas  $ODI3$  achieved a slightly lower confidence interval than  $AHI_{MLP}$  (23.2 versus 24.3). Notice that  $ODI3$  underestimates AHI, whereas  $AHI_{MLP}$  corrects this behaviour by showing a slight overestimation. In addition,  $AHI_{MLP}$  achieved better agreement with  $AHI_{PSG}$  (ICC=0.891) than  $ODI3$  (ICC=0.866). Regarding the diagnostic performance, table 3 shows the confusion matrices of  $ODI3$  and  $AHI_{MLP}$  in the test group. These matri-

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

| Features                            | AHI < 1            | 1 ≤ AHI < 5       | 5 ≤ AHI < 10     | AHI ≥ 10           | p-value* |
|-------------------------------------|--------------------|-------------------|------------------|--------------------|----------|
| <i>ODI3</i>                         | 1.04 [0.52,2.47]   | 2.03 [0.93,3.89]  | 3.69 [1.94,7.23] | 12.35 [6.65,24.49] | < .01    |
| <i>slope</i> <sub>1</sub>           | 1.63 [1.58,1.68]   | 1.64 [1.58,1.70]  | 1.67 [1.60,1.71] | 1.74 [1.66,1.79]   | < .01    |
| <i>slope</i> <sub>1/2</sub>         | 0.96 [0.90,1.05]   | 0.95 [0.87,1.03]  | 0.92 [0.85,1.02] | 0.94 [0.88,1.01]   | .18      |
| <i>k</i> <sub>12</sub>              | 1.66 [1.53,1.82]   | 1.69 [1.55,1.87]  | 1.77 [1.60,1.94] | 1.82 [1.68,1.95]   | < .01    |
| <i>F</i> ( <i>k</i> <sub>12</sub> ) | 1.33 [1.23,1.42]   | 1.36 [1.26,1.44]  | 1.38 [1.29,1.45] | 1.34 [1.23,1.42]   | .04      |
| <i>F</i> ( <i>k</i> <sub>2</sub> )  | 0.01 [-0.18,0.18]  | 0.12 [-0.12,0.26] | 0.22 [0.04,0.38] | 0.42 [0.16,0.61]   | < .01    |
| <i>F</i> ( <i>k</i> <sub>α</sub> )  | -0.05 [-0.13,0.04] | 0.02 [-0.07,0.11] | 0.10 [0.00,0.20] | 0.31 [0.18,0.52]   | < .01    |

\*p-values obtained after Bonferroni correction.

**Figure 4:** Averaged kappa for each  $N_H$ - $\alpha$  pair.

ces show the class predicted by both original *ODI3* and  $AHI_{MLP}$  for each subject versus the actual SAHS severity group, according to  $AHI_{PSG}$ . Using *ODI3*, 55.4% of the subjects (217/392) were correctly assigned to their actual group of SAHS severity (sum of the main diagonal elements of the matrix). Conversely,  $AHI_{MLP}$  rightly assigned 60.0% (235/392) of the subjects to their SAHS severity group. Kappa values were 0.355 (*ODI3*) and 0.412 ( $AHI_{MLP}$ ). Table 4 shows the diagnostic ability of both *ODI3* and  $AHI_{MLP}$  for the  $AHI_{PSG}$ -based cut-offs of 1, 5, and 10 e/h.  $AHI_{MLP}$  outperformed single *ODI3* in terms of ICC, overall Acc and kappa. Additionally, our  $AHI_{MLP}$  reached higher Acc for the single AHI cut-offs of 1 and 10 e/h. With respect to the comparison of the results of *ODI3* and  $AHI_{MLP}$ , statistically significant higher values ( $p$ -value < 0.01) were obtained using  $AHI_{MLP}$  in the case of ICC, kappa, and overall Acc. In addition, statistically significant differences were found for the AHI cut-offs of 5 and 10 e/h between AUC of *ODI3* and  $AHI_{MLP}$ .

#### 4. Discussion

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

to  $SpO_2$  recordings is novel in the context of paediatric SAHS. Our proposed approach shows a high diagnostic ability which outperforms the conventional oximetric index *ODI3*.

*ODI3* and four out of six features from DFA (*slope*<sub>1</sub>, *slope*<sub>1/2</sub>, *F*(*k*<sub>12</sub>), and *F*(*k*<sub>α</sub>)) reached significantly higher values that were associated with increased severity of SAHS. The statistical differences shown by these DFA-derived parameters indicate that the scaling behaviour of the DFA profile of the  $SpO_2$  signal is affected in the presence of SAHS, as illustrated by Figure 1. This change in the correlation properties of the  $SpO_2$  signal along time scales may be caused by the presence of spikes or segments with different statistical properties (Chen et al., 2002; Hua and Yu, 2017). Figure 1 shows two regions with different scaling exponents (correlation)—one region for short-time scales (region 1) and another region for long-time scales (region 2). Two scaling regions were also obtained in the studies developed by Dehkordi et al. (2016) and Penzel et al. (2003). Dehkordi et al. (2016) and Penzel et al. (2003) applied DFA to analyse the scaling behaviour of the pulse rate variability (PRV) and heart rate variability (HRV) signals in the context of SAHS, respectively. According to these studies (Dehkordi et al., 2016; Penzel et al., 2003), the time scales of these regions may be related to the duration of apnoeic events. In these studies, short-time scales relate to the effects of respiration on the heart rate, whereas long time scales relate to the effects of sleep stages and circadian rhythm (Penzel et al., 2003).

According to the physiological interpretation of both regions in the DFA profile, the higher values shown by *slope*<sub>1</sub> and *slope*<sub>1/2</sub> that are associated with the severity of SAHS may be related to the variations in the  $SpO_2$  signals caused by respiratory events (Hua and Yu, 2017; Peng et al., 1995; Penzel et al., 2003) which directly affect the oximetry dynamics. On the contrary, *slope*<sub>2</sub> did not show statistically significant differences. According to Penzel et al. (2003), *slope*<sub>2</sub> is related to the effects of slower brain functions on the HRV signal. Nevertheless, slower brain functions may not be related to the effects of SAHS in the oximetry signal. That is one possible reason why the value of *slope*<sub>2</sub> does not increase with the severity of SAHS. Figure 1 also shows higher values of *F*(*k*) in the  $SpO_2$  signal as the severity of SAHS increases. These differences may be due to the fluctuations

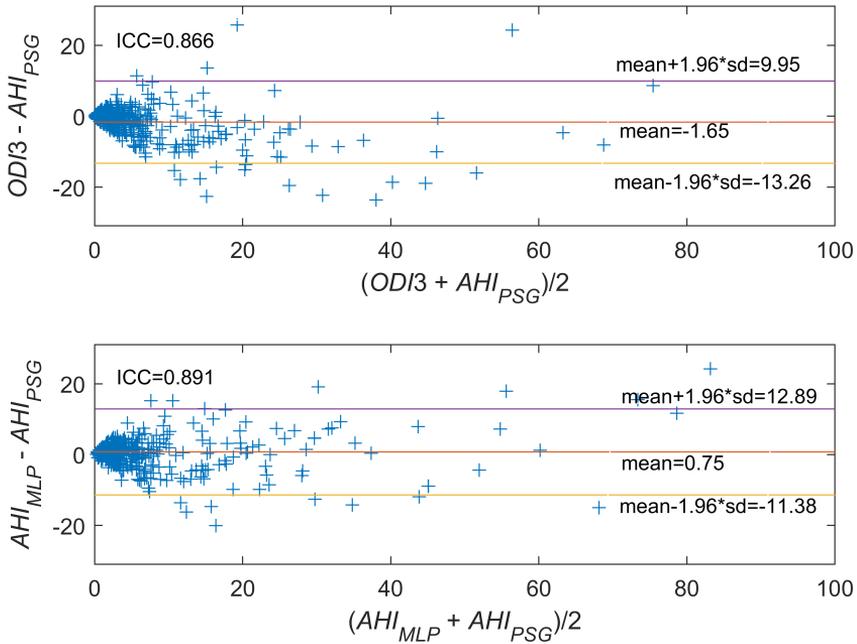


Figure 5: Bland–Altman plots comparing (a)  $ODI3$  with  $AHI_{PSG}$  and (b)  $AHI_{MLP}$  with  $AHI_{PSG}$ .

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

|             |                   | $ODI3$    |                  |                   |               | $AHI_{MLP}$ |                  |                   |               |
|-------------|-------------------|-----------|------------------|-------------------|---------------|-------------|------------------|-------------------|---------------|
|             |                   | $AHI < 1$ | $1 \leq AHI < 5$ | $5 \leq AHI < 10$ | $AHI \geq 10$ | $AHI < 1$   | $1 \leq AHI < 5$ | $5 \leq AHI < 10$ | $AHI \geq 10$ |
| $AHI_{PSG}$ | $AHI < 1$         | 39        | 36               | 1                 | 1             | 18          | 55               | 3                 | 1             |
|             | $1 \leq AHI < 5$  | 47        | 107              | 12                | 3             | 8           | 125              | 33                | 3             |
|             | $5 \leq AHI < 10$ | 5         | 33               | 17                | 8             | 1           | 22               | 28                | 12            |
|             | $AHI \geq 10$     | 0         | 13               | 16                | 54            | 0           | 8                | 11                | 64            |

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

Regarding the results of the feature selection stage, figure 3 shows that only  $ODI3$  and  $slope_1$  were selected more than 500 times after the bootstrapping approach. The remaining features showed high redundancy.  $ODI3$  and  $slope_1$  come from different methodological approaches. Therefore, this suggests that information from DFA is complementary to that obtained from the conventional  $ODI3$ . As aforementioned, an MLP neural network fed with this optimum subset outperformed the  $ODI3$  (Tables 3 and 4). A better agreement with the  $AHI_{PSG}$  was achieved with our  $AHI_{MLP}$ , as well as a higher diagnostic

**Table 4:** Diagnostic ability of *ODI3* and *AHI<sub>MLP</sub>* in the test set for AHI cut-offs = 1, 5, and 10 e/h.

| AHI cutoff = 1 e/h       |      |      |      |      |      |      |      |       |
|--------------------------|------|------|------|------|------|------|------|-------|
| Method                   | Se   | Sp   | PPV  | NPV  | LR+  | LR-  | Acc  | AUC   |
| <i>ODI3</i>              | 85.5 | 50.6 | 87.4 | 42.9 | 1.7  | 0.35 | 77.0 | 0.811 |
| <i>AHI<sub>MLP</sub></i> | 97.1 | 23.5 | 83.9 | 66.7 | 1.3  | 0.12 | 82.7 | 0.813 |
| AHI cutoff = 5 e/h       |      |      |      |      |      |      |      |       |
| Method                   | Se   | Sp   | PPV  | NPV  | LR+  | LR-  | Acc  | AUC   |
| <i>ODI3</i>              | 65.1 | 93.1 | 84.8 | 81.8 | 9.4  | 0.37 | 82.7 | 0.883 |
| <i>AHI<sub>MLP</sub></i> | 78.8 | 83.7 | 74.2 | 86.9 | 4.8  | 0.25 | 81.9 | 0.888 |
| AHI cutoff = 10 e/h      |      |      |      |      |      |      |      |       |
| Method                   | Se   | Sp   | PPV  | NPV  | LR+  | LR-  | Acc  | AUC   |
| <i>ODI3</i>              | 65.1 | 96.1 | 81.8 | 91.1 | 16.7 | 0.36 | 89.5 | 0.921 |
| <i>AHI<sub>MLP</sub></i> | 77.1 | 94.8 | 80.0 | 93.9 | 14.9 | 0.24 | 91.1 | 0.930 |

ability to predict SAHS severity. This highlights the usefulness of FCBF, the feature selection method employed in our proposal. According to our results,  $slope_1$ , which was involved in the optimum subset, quantifies changes in the scaling behaviour of the DFA profile that provides additional information regarding oximetry dynamics able to enhance its diagnostic ability.

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

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

Recent studies have focused on the application of automated signal processing approaches to enhance the diagnostic ability of the  $SpO_2$  signal (Garde et al., 2014; Álvarez et al., 2017; Crespo et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018). From these studies, only Hornero et al. (2017) assessed an AHI estimation model. Hornero et al. (2017) built an MLP regression model with *ODI3* and the skewness of the PSD extracted from 4191  $SpO_2$  recordings from 13 sleep laboratories worldwide. Our study outperformed the state-of-the-art approaches except the performance reported by the study of Álvarez et al. (2017), which achieved higher accuracies for the AHI cut-offs of 1 and 5 e/h. However, the database used by Álvarez et al. (2017) had only 50 subjects. As a consequence, their results are less generalizable, and they performed binary classification instead of estimating the AHI of each patient.

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

**Table 5:** Summary of state-of-the-art in the context of the analysis of SpO<sub>2</sub> to assist in the diagnosis of paediatric SAHS.

| Study                          | Subjects (n) | AHI cut-off  | Methods  | Validation                         | Se (%)               | Sp (%)               | Acc (%)               |
|--------------------------------|--------------|--------------|--|------------------------------------|----------------------|----------------------|-----------------------|
| Kirk et al. (2003)             | 58           | 5            | <i>ODI3</i>  | Direct validation**                | 67                   | 60                   | 64*                   |
| Tsai et al. (2013)             | 148          | 1<br>5<br>10 | <i>ODI4</i>  | No                                 | 77.7<br>85.8<br>89.1 | 88.9<br>86.5<br>86.0 | 79.0<br>85.1<br>87.1* |
| Garde et al. (2014)            | 146          | 5            | Statistical, nonlinear features, classical indices, and PSD            | Four-fold cross validation         | 80.0                 | 83.9                 | 78.5                  |
| Van Eyck et al. (2015)         | 130          | 2            | <i>ODI3</i> and clusters of desaturations                              | Train-test for <i>ODI3</i>         | 57<br>58<br>66       | 73<br>88<br>69       | 68<br>78<br>68*       |
| Álvarez et al. (2017)          | 50           | 1<br>3<br>5  | Statistical parameters, nonlinear features, classical indices, and PSD | Bootstrap 0.632                    | 89.6<br>82.9<br>82.2 | 71.5<br>84.4<br>83.6 | 85.5<br>83.4<br>82.8  |
| Crespo et al. (2017)           | 50           | 3            | Multiscale entropy and classical indices                               | Bootstrap 0.632                    | 84.5                 | 83.0                 | 83.5                  |
| Hornero et al. (2017)          | 4191         | 1<br>5<br>10 | Statistical, nonlinear features, PSD, and <i>ODI3</i>                  | Training-test                      | 84.0<br>68.2<br>68.7 | 53.2<br>87.2<br>94.1 | 75.2<br>81.7<br>90.2  |
| Vaquerizo-Villar et al. (2018) | 298          | 5<br>10      | Bispectrum, PSD, <i>ODI3</i> , anthropometric variables                | Feature optimization-training-test | 61.8<br>60.0         | 97.6<br>94.5         | 81.3<br>85.5          |
| Our proposal                   | 981          | 1<br>5<br>10 | DFA and <i>ODI3</i>  | Training-test                      | 97.1<br>78.8<br>77.1 | 23.3<br>83.7<br>94.8 | 82.7<br>81.9<br>91.1  |

\*Computed from reported data, \*\* Direct validation of a scoring criteria against AHI from PSG, loocv= leave-one-out cross validation.

ter could facilitate its use in ambulatory settings.

## 5. Conclusion

In summary, we investigated the usefulness of DFA to obtain additional information from SpO<sub>2</sub> recordings in order to simplify the detection of paediatric SAHS. Four features extracted from DFA showed significant differences between the SAHS severity groups. An optimum subset composed of *ODI3* and  $slope_1$  was obtained with FCBF, which suggests that these features are complementary and non-overlapping. An MLP model fed with this optimum subset achieved a good agreement with the AHI from PSG, obtaining 0.891 ICC and 0.412 kappa, as well as high diagnostic ability. This MLP model achieved better agreement (ICC and kappa) than *ODI3*, as well as higher accuracies for the cut-offs of 1 and 10 e/h. Our methodology achieved a high diagnostic performance in comparison with state-of-the-art techniques. This suggests that the changes in the scaling behaviour of the DFA profile quantified by  $slope_1$  can provide additional information to enhance the diagnostic ability of the oximetry signal in the context of paediatric SAHS.

## Conflict of interest

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

## Ethical approval

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

## Authorship responsibility

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

## Acknowledgements

This work was supported by 'Agencia Estatal de Investigación del Ministerio de Ciencia, Innovación y Universidades' and 'European Regional Development Fund (FEDER)' under projects DPI2017-84280-R, RTC-2015-3446-1, and by 'European Commission' and 'FEDER' under project 'Análisis y correlación entre el genoma completo y la actividad cerebral para la ayuda en el di-

agnóstico de la enfermedad de Alzheimer' ('Cooperation Programme Interreg V-A Spain-Portugal POCTEP 2014–2020').

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

## References

- Alonso-Álvarez, M. L., Canet, T., Cubell-Alarco, M., Estivill, E., Fernández, E., Gozal, D., Jurado-Luque, M. J., Lluich-Roselló, M. A., Martínez-Pérez, F., Merino-Andren, M., Pin-Arboledas, G., Roure, N., Sanmartí, F. X., Sans-Capdevila, O., Segarra-Isern, J., 2011. Documento de consenso del síndrome de apneas-hipopneas durante el sueño en niños. *Arch Bronconeumol* 47 (Supl 5), 2–18.
- Álvarez, D., Alonso-Álvarez, M. L., Gutiérrez-Tobal, G. C., Crespo, A., Kheirandish-Gozal, L., Gozal, D., Terán-Santos, J., Campo, F. D., 2017. Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal Oximetry : An Alternative to Respiratory Polysomnography in Unattended Settings. *J Clin Sleep Med* 13 (5), 7–11.
- Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., Marcus, C. L., Mehra, R., Parthasarathy, S., Quan, S. F., Redline, S., Strohl, K. P., Davidson Ward, S. L., Tangredi, M. M., 2012. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. *Journal of clinical sleep medicine* 8 (5), 597.
- Bishop, C. M., et al., 1995. Neural networks for pattern recognition. Oxford university press.
- Chang, L., Wu, J., Cao, L., 2013. Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea. *International Journal of Pediatric Otorhinolaryngology* 77 (5), 365–371.
- Chen, Z., Ivanov, P. C., Hu, K., Stanley, H. E., 2002. Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 65 (4), 15.
- Church, G. D., 2012. The Role of Polysomnography in Diagnosing and Treating Obstructive Sleep Apnea in Pediatric Patients. *Current problems in pediatric and adolescent health care* 42 (1), 22–25.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20 (1), 37–46.
- Crespo, A., Álvarez, D., Gutiérrez-Tobal, G. C., Vaquerizo-Villar, F., Barroso-García, V., Alonso-Álvarez, M. L., Terán-Santos, J., Hornero, R., del Campo, F., 2017. Multiscale Entropy Analysis of Unattended Oximetric Recordings to Assist in the Screening of Paediatric Sleep Apnoea at Home. *Entropy* 19 (6), 284.
- Dehkordi, P., Garde, A., Karlen, W., Petersen, C. L., Wensley, D., Dumont, G. A., Ansermino, J. M., 2016. Evaluation of cardiac modulation in children in response to apnea/hypopnea using the phone oximeter™. *Physiological measurement* 37 (2), 187.
- Efron, B., Tibshirani, R. J., 1994. An introduction to the bootstrap. CRC press.
- Garde, A., Dehkordi, P., Karlen, W., Wensley, D., Ansermino, J. M., Dumont, G. A., 2014. Development of a screening tool for sleep disordered breathing in children using the phone oximeter™. *PLoS one* 9 (11), e112959.
- Gutiérrez-Tobal, G. C., Álvarez, D., Crespo, A., Del Campo, F., Hornero, R., 2018. Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings. *IEEE journal of biomedical and health informatics* 23 (2), 882–892.
- Guyon, I., Elisseeff, A., 2005. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research (JMLR)* 5 (3), 1157–1182.
- Hornero, R., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Philby, M. F., Alonso-Álvarez, M. L., Álvarez, D., Dayyat, E. A., Xu, Z., Huang, Y.-S., Tamae Kakazu, M., Li, A. M., Van Eyck, A., Brockmann, P. E., Ehsan, Z., Simakajornboon, N., Kaditis, A. G., Vaquerizo-Villar, F., Crespo Sedano, A., Sans Capdevila, O., von Lukovits, M., Terán-Santos, J., Del Campo, F., Poets, C. F., Ferreira, R., Bertran, K., Zhang, Y., Schuen, J., Verhulst, S., Gozal, D., 2017. Nocturnal Oximetry-based Evaluation of Habitually Snoring Children. *American Journal of Respiratory and Critical Care Medicine* 196 (12), 1591–1598.
- Hua, C.-C., Yu, C.-C., 2017. Detrended fluctuation analysis of oxyhemoglobin saturation by pulse oximetry in sleep apnea syndrome. *Journal of Medical and Biological Engineering* 37 (6), 791–799.
- Kaditis, A., Kheirandish-Gozal, L., Gozal, D., 2016. Pediatric OSAS : Oximetry can provide answers when polysomnography is not available. *Sleep Medicine Reviews* 27, 96–105.
- Kaimakamis, E., Tsara, V., Bratsas, C., Sichelitidis, L., Karvounis, C., Maglaveras, N., 2016. Evaluation of a decision support system for obstructive sleep apnea with nonlinear analysis of respiratory signals. *PLoS one* 11 (3), e0150163.
- Katz, E. S., Mitchell, R. B., Ambrosio, C. M. D., 2012. Obstructive Sleep Apnea in Infants. *American journal of respiratory and critical care medicine* 185 (8), 805–816.
- Kirk, V. G., Bohn, S. G., Flemons, W. W., Remmers, J. E., 2003. Comparison of Home Oximetry Monitoring With Laboratory Polysomnography in Children. *Chest* 124 (5), 1702–1708.
- Lee, J.-M., Kim, D.-J., Kim, I.-Y., Park, K.-S., Kim, S. I., 2002. Detrended fluctuation analysis of eeg in sleep apnea using mit/bih polysomnography data. *Computers in biology and medicine* 32 (1), 37–47.
- Magalang, U. J., Dmochowski, J., Veeramachaneni, S., Draw, A., Mador, J., El-Solh, A., Grant, B. J. B., 2003. Prediction of the Apnea-Hypopnea Index From Overnight Pulse Oximetry. *CHEST Journal* 124 (5), 1694–1701.
- Marcus, C. L., Brooks, L. J., Ward, S. D., Draper, K. A., Gozal, D., Halbower, A. C., Jones, J., Lehmann, C., Schechter, M. S., Sheldon, S., Shiffman, R. N., Spruyt, K., 2012. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics* 130 (3), e714–e755.
- McClatchey, K. D., 2002. Clinical laboratory medicine. Lippincott Williams & Wilkins.
- Nabney, I., 2002. NETLAB: algorithms for pattern recognition. Springer Science & Business Media.
- Netzer, N., Eliasson, A. H., Netzer, C., Kristo, D. A., 2001. Overnight pulse oximetry for sleep-disordered breathing in adults: a review. *Chest* 120 (2), 625–633.
- Nixon, G. M., Kermack, A. S., Davis, G. M., Manoukian, J. J., Brown, A., Brouillette, R. T., 2004. Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry. *Pediatrics* 115 (1), e19–e25.
- Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., Goldberger, A. L., 1994. Mosaic organization of DNA nucleotides. *Physical review e* 49 (2), 1685.
- Peng, C.-K., Havlin, S., Stanley, H. E., Goldberger, A. L., 1995. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: an interdisciplinary journal of nonlinear science* 5 (1), 82–87.
- Penzel, T., Kantelhardt, J., Grote, L., Peter, J., Bunde, A., 2003. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE Transactions on Biomedical Engineering* 50 (10), 1145–1151.
- Taha, B. H., Dempsey, J. A., Weber, S. M., Badr, M. S., Skatrud, J. B., Young, T. B., Jacques, A. J., Seow, K., 1997. Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep* 20 (11), 991–1001.
- Tan, H.-L., Gozal, D., Ramirez, H. M., Bandla, H. P. R., Kheirandish-

- Gozal, L., 2014. Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea. *Sleep* 37 (2), 255–260.
- Tsai, C.-M., Kang, C.-H., Su, M.-C., Lin, H.-C., Huang, E.-Y., Chen, C.-C., Hung, J.-C., Niu, C.-K., Liao, D.-L., Yu, H.-R., 2013. Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. *International journal of pediatric otorhinolaryngology* 77 (8), 1286–1290.
- Van Eyck, A., Lambrechts, C., Vanheeswijck, L., 2015. The role of nocturnal pulse oximetry in the screening for obstructive sleep apnea in obese children and adolescents. *Sleep Medicine* 16 (11), 1409–1412.
- Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., del Campo, F., Gozal, D., Hornero, R., 2018. Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings. *Computer Methods and Programs in Biomedicine* 156, 141–149.
- Yu, L., Liu, H., 2004. Efficient Feature Selection via Analysis of Relevance and Redundancy. *Journal of Machine Learning Research* 5, 1205–1224.

## A convolutional neural network architecture to enhance oximetry ability to diagnose pediatric obstructive sleep apnea

Fernando Vaquerizo-Villar<sup>1,2</sup>, Daniel Álvarez-González<sup>1,2,3</sup>, Leila Kheirandish-Gozal<sup>4</sup>, Gonzalo C. Gutiérrez-Tobal<sup>1,2</sup>, Verónica Barroso-García<sup>1,2</sup>, Eduardo Santamaría-Vázquez<sup>1,2</sup>, Félix del Campo<sup>1,2,3</sup>, David Gozal<sup>4</sup>, Roberto Hornero<sup>1,2</sup>

<sup>1</sup> Biomedical Engineering Group, E.T.S.I. Telecomunicación, University of Valladolid, Paseo de Belén 15, 47011, Valladolid, Spain

<sup>2</sup> Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN), Spain

<sup>3</sup> Hospital Universitario Río Hortega of Valladolid, Spain

<sup>4</sup> The University of Missouri School of Medicine, Columbia, Missouri, United States of America

IEEE Journal of Biomedical and Health Informatics

Volume 25 (8), August 2021, Pages 2906–2916

### Abstract

This study aims at assessing the usefulness of deep learning to enhance the diagnostic ability of oximetry in the context of automated detection of pediatric obstructive sleep apnea (OSA). A total of 3196 blood oxygen saturation (SpO<sub>2</sub>) signals from children were used for this purpose. A convolutional neural network (CNN) architecture was trained using 20-min SpO<sub>2</sub> segments from the training set (859 subjects) to estimate the number of apneic events. CNN hyperparameters were tuned using Bayesian optimization in the validation set (1402 subjects). This model was applied to three test sets composed of 312, 392, and 231 subjects from three independent databases, in which the apnea-hypopnea index (AHI) estimated for each subject (AHI<sub>CNN</sub>) was obtained by aggregating the output of the CNN for each 20-min SpO<sub>2</sub> segment. AHI<sub>CNN</sub> outperformed the 3% oxygen desaturation index (ODI3), a clinical approach, as well as the AHI estimated by a conventional feature-engineering approach based on multi-layer perceptron (AHI<sub>MLP</sub>). Specifically, AHI<sub>CNN</sub> reached higher four-class Cohen's kappa in the three test databases than ODI3 (0.515 vs 0.417, 0.422 vs 0.372, and 0.423 vs 0.369) and AHI<sub>MLP</sub> (0.515 vs 0.377, 0.422 vs 0.381, and 0.423 vs 0.306). In addition, our proposal outperformed state-of-the-art studies, particularly for the AHI severity cutoffs of 5 e/h and 10 e/h. This suggests that the information automatically learned from the SpO<sub>2</sub> signal by deep-learning techniques helps to enhance the diagnostic ability of oximetry in the context of pediatric OSA.

**Keywords:** Oximetry, deep learning, convolutional neural networks (CNN), apnea–hypopnea index (AHI), pediatric obstructive sleep apnea (OSA).

### 1. Introduction

Obstructive sleep apnea (OSA) is a highly prevalent condition among the pediatric population (1%–5%) (Marcus et al., 2012). Pediatric OSA is characterized by recurrent respiratory pauses (apneas) and airflow reductions (hypopneas), which leads to oxygen desaturations and arousals that cause restless sleep (Marcus et al., 2012; Berry et al., 2012). Untreated OSA is associated to metabolic and cardiovascular malfunctioning, as well as

neurobehavioral abnormalities that diminish children's health and quality of life (Marcus et al., 2012; Alonso-Álvarez et al., 2011).

The gold standard diagnosis test is polysomnography (PSG) (Marcus et al., 2012). PSG requires children to spend the night in a specialized sleep unit while being recorded up to 32 biomedical signals (Alonso-Álvarez et al., 2011; Kaditis et al., 2016). These recordings are used to score apneas and hypopneas in order to obtain the apnea-hypopnea index (AHI), which is the clinical variable used to diagnose OSA (Berry et al., 2012). Despite its effectiveness, several limitations of PSG have been pointed out (Katz et al., 2012; Kheirandish-Gozal, 2010), including its complexity, cost, high intrusiveness, and limited availability. This results in a delay in the diagnosis and treatment of OSA of the affected children (Nixon et al., 2004).

In order to overcome these limitations, the scientific

Email addresses: fernando.vaquerizo@gib.tel.uva.es (Fernando Vaquerizo-Villar<sup>1,2</sup>), dalvgon@gmail.com (Daniel Álvarez-González<sup>1,2,3</sup>), gozall@health.missouri.edu (Leila Kheirandish-Gozal<sup>4</sup>), gguttob@gmail.com (Gonzalo C. Gutiérrez-Tobal<sup>1,2</sup>), veronica.barroso@gib.tel.uva.es (Verónica Barroso-García<sup>1,2</sup>), eduardo.santamaria@gib.tel.uva.es (Eduardo Santamaría-Vázquez<sup>1,2</sup>), fsas@telefonica.net (Félix del Campo<sup>1,2,3</sup>), gozald@health.missouri.edu (David Gozal<sup>4</sup>), robhor@tel.uva.es (Roberto Hornero<sup>1,2</sup>)

community has explored the use of simplified tests that increase the accessibility and effectiveness of pediatric OSA diagnosis. In this respect, the blood oxygen saturation signal ( $\text{SpO}_2$ ) from nocturnal oximetry has been frequently proposed as a clinically valuable tool for the screening of OSA in children due to its simplicity, reliability and suitability (del Campo et al., 2018; Netzer et al., 2001). The  $\text{SpO}_2$  signal measures the oxygen content in the hemoglobin of the blood (McClatchey, 2002), thus containing information of the oxygen desaturations associated to apneic events from OSA (Berry et al., 2012).

Promising results have been obtained in previous studies from the automated analysis of the  $\text{SpO}_2$  signal following a feature-engineering methodology (Tsai et al., 2013; Villa et al., 2015; Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019). First, a set of hand-crafted features from oximetry was obtained using different signal processing algorithms: conventional oximetric indices, statistical parameters, non-linear methods, and frequency domain techniques (Tsai et al., 2013; Villa et al., 2015; Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019). Then, thresholding rules (Tsai et al., 2013; Villa et al., 2015) and machine-learning algorithms (Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019) were used with these features to determine the presence and severity of pediatric OSA. Nonetheless, these conventional feature-engineering approaches require considerable knowledge in order to identify, *a priori*, a set of relevant features to extract from the data (Goodfellow et al., 2016). In addition, the level of abstraction that classical methods provide is low, which limits their ability to identify complex patterns in the data (Goodfellow et al., 2016). This may result in missing relevant information from the  $\text{SpO}_2$  signals linked to apneic events.

These issues can be solved by using deep-learning algorithms, which automatically learn complex patterns for detection or classification tasks from raw data using architectures with multiple levels of representation (Goodfellow et al., 2016). These algorithms have beaten conventional approaches in many fields, such as image recognition, language processing, and time series analysis (Goodfellow et al., 2016). In the OSA context, recent studies have focused on the application of deep-learning techniques to detect sleep stages (Faust et al., 2019), apneic events (Mostafa et al., 2019), and/or estimate AHI in adult OSA patients (Mostafa et al., 2019). Their findings suggest that deep-learning algorithms are appropriate to analyze different physiological signals from PSG, such as electrocardiogram, electroencephalogram, airflow, or oximetry (Faust et al., 2019; Mostafa et al., 2019).

Specifically, the majority of these studies employed deep-learning architectures based on convolutional neural networks (CNN) (Mostafa et al., 2019), which are the

most widely-used deep-learning algorithm (Goodfellow et al., 2016). Despite being originally inspired for image analysis, CNNs have proven its suitability for time series classification in a big variety of domains (Ismail Fawaz et al., 2019), including biomedical signal analysis (Faust et al., 2018; Roy et al., 2019; Ebrahimi et al., 2020; Murat et al., 2020). CNN have a multi-layer architecture, with shared weights, sparse connections, and pooling operations, which allows them to identify both short- and long-term patterns occurring in different parts of the time series (Ebrahimi et al., 2020), while reducing the computational cost of other deep-learning algorithms (Goodfellow et al., 2016). This CNN property may be useful to identify desaturations in the  $\text{SpO}_2$  signal associated to apneic events that may occur at different times. In addition, CNNs provide higher levels of representation that allow to learn more complex features (Goodfellow et al., 2016), which may be useful to detect complex patterns in long segments of the  $\text{SpO}_2$  signal, such as clusters of desaturations (Brouillette et al., 2000).

The novelty of this research is the use of a new deep-learning model based on CNN that allows to accurately diagnose pediatric OSA with a high generalization ability from the raw oximetry signal. We hypothesize that deep-learning approaches could help to automatically extract the relevant information of the oximetry signal in the context of pediatric OSA diagnosis. Consequently, the main goal of this study is to evaluate the usefulness of deep-learning to estimate the AHI from overnight oximetry in children with suspected OSA. To achieve this goal, a CNN architecture is trained to estimate the number of apneic events from 20-min  $\text{SpO}_2$  segments, which is a novel approach in the context of pediatric OSA. The output of the CNN for each segment is then aggregated to estimate the AHI in pediatric OSA patients using a large cohort of 3196  $\text{SpO}_2$  recordings from three different datasets.

One related conference paper developed by our own group has already been published showing preliminary results (Vaquerizo-Villar et al., 2019). Despite the fact that our previous work also applied CNNs to analyze  $\text{SpO}_2$  recordings, there are some essential differences with this research. Our main contribution is that our deep-learning based methodology allows to diagnose pediatric OSA using the oximetry signal. In this sense, our previous work showed promising results in detecting apneic events (event-based approach) from the oximetry signal using a CNN (Vaquerizo-Villar et al., 2019). In the current study, we have investigated whether those indications may be extended to obtain a new deep-learning model based on CNN that allows to directly estimate the AHI, thus being able to conduct a complete automatic diagnosis (subject-based), including the assessment of the pediatric OSA severity degrees. Instead of training a CNN to detect individual apneic events (binary output), in this research we have trained a CNN to regress the number of apneic events in  $\text{SpO}_2$  segments (continuous output), which allows to accurately analyze  $\text{SpO}_2$  segments with several

apneic events. A two-step aggregation procedure (averaging plus linear regression) has been also included to accurately estimate the AHI of each subject from the outputs of the CNN. We have also incorporated novel elements to improve the training and optimization process of the deep-learning model (Huber loss, batch shuffling, learning rate scheduler, early stopping, and Bayesian optimization). Additionally, in the present study, we have designed and prospectively assessed a new model using three independent datasets, leading to a sample size seven times larger than in our preliminary work (3196 vs. 453). This contributes to increase the generalization ability of our current proposal. Finally, another contribution of our work is that we have performed a thorough comparison with two conventional approaches to properly assess the validity of our proposal. Particularly, we have compared the results of the proposed approach with the 3% oxygen desaturation index (ODI3), a conventional clinical approach commonly used for comparison purposes (Tsai et al., 2013; Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019), as well as with the AHI estimated by a classical feature-engineering approach.

## 2. Subjects and signals under study

A total of 3196 sleep studies of children ranging from 0 to 18 years of age composed the population under study. Three large datasets were used: (i) the Childhood Adenotonsillectomy Trial (CHAT) dataset, a public multicenter database composed of 1638 sleep studies (clinical trial identifier: NCT00560859) (Marcus et al., 2013; Redline et al., 2011); (ii) the University of Chicago (UofC) dataset, a private database composed of 980 pediatric subjects; and (iii) the Burgos University Hospital (BUH) dataset, a private database composed of 578 subjects. All subjects from the three datasets were referred to overnight PSG due to clinical suspicion of OSA. An informed consent was obtained from all legal caretakers of the children and the Ethics Committee of the different sleep centers involved in the study approved the protocols.

SpO<sub>2</sub> recordings were acquired during PSG using sampling rates ranging from 1 to 512 Hz. The guidelines of the AASM were used to quantify sleep and score apneas and hypopneas by pediatric sleep specialists from the different centers (Berry et al., 2012; Iber et al., 2007). The AHI, obtained as the number of apneas and hypopneas per hour of sleep, was used to diagnose pediatric OSA (Berry et al., 2012). Common clinically used AHI cutoffs of 1, 5, and 10 events per hour (e/h) were used in this study to classify children into four OSA severity degrees: no-OSA (AHI < 1 e/h), mild OSA (1 ≤ AHI < 5 e/h), moderate OSA (5 ≤ AHI < 10 e/h), and severe OSA (AHI ≥ 10 e/h) (Alonso-Álvarez et al., 2011; Church, 2012; Tan et al., 2014).

Data was divided into three sets: training set, employed to train the deep-learning algorithms; validation

set, used for hyperparameters optimization; and test set, employed to evaluate the diagnostic performance of the deep-learning methods. Only the CHAT database contains annotations of time location of apnea and hypopnea events, which are needed in the deep-learning models as the output labels for training. Accordingly, the training set was composed of 859 SpO<sub>2</sub> recordings from the baseline (453 subjects) and follow-up groups (406 subjects) of the CHAT database (Redline et al., 2011). The subjects from the remaining group of the CHAT dataset, non-randomized (779 subjects), as well as the subjects of the UofC and BUH sets, were randomly divided into a validation set (60%) and a test set (40%), being 60%-40% a common proportion used in previous studies for validation and test purposes (Hornero et al., 2017; Vaquerizo-Villar et al., 2018a).

In this way, the validation set was composed of 1402 SpO<sub>2</sub> recordings from the CHAT (467, 60% of the 779 subjects from the nonrandomized group), UofC (588, 60% of the 980 subjects), and BUH (347, 60% of the 578 subjects) databases, whereas the test set was composed of 312 subjects from the CHAT dataset (40% of the nonrandomized group), 392 subjects from the UofC dataset (40%), and 231 subjects from the BUH dataset (40%). Table 1 shows clinical and demographic data from the subjects under study.

## 3. Methodology

### 3.1. Proposed CNN model

The proposed solution, depicted in Figure 1, consists of three steps: (1) signals segmentation; (2) CNN architecture; and (3) AHI estimation.

#### 3.1.1. Signals segmentation

First, SpO<sub>2</sub> recordings were down-sampled to a sample rate of 1 Hz in order to homogenize the frequency. SpO<sub>2</sub> signals from each subject were then divided into 20-min segments (1200 samples), as shown in Figure 1 (a). This segment size (20-min) allows to detect clusters of desaturations, which have a minimum duration of 10-min (Brouillette et al., 2000). Finally, each 20-min SpO<sub>2</sub> segment in the training set is labelled with the annotations provided by sleep technicians (Redline et al., 2011). The output label for each segment was obtained as the number of apnea and hypopnea events associated to 3% oxygen desaturations occurring in these 20 minutes, according to the annotation event files of the CHAT dataset (Redline et al., 2011).

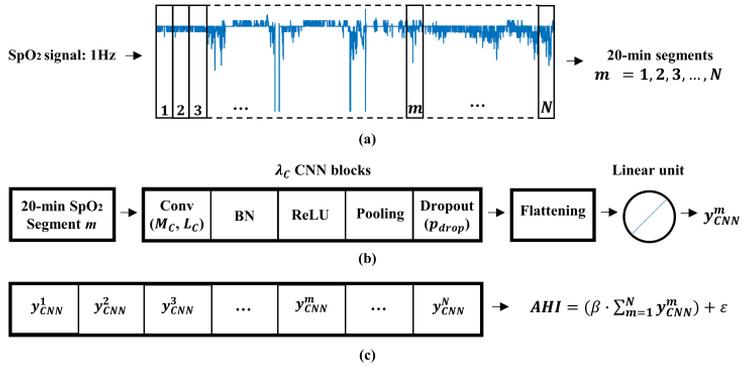
#### 3.1.2. CNN architecture

CNN are the most popular deep-learning technique to process multidimensional arrays, such as 1D signals or 2D images. In this study, CNN were used to process raw oximetry data. Figure 1 (b) shows the architecture of the proposed CNN. The input of the CNN architecture is the 20-min SpO<sub>2</sub> segment. The CNN architecture processes

**Table 1:** Demographic and clinical data from children under study.

|                                 | All                  | Training set         | Validation set       | CHAT Test set        | UofC Test set        | BUH Test set         |
|---------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| SpO <sub>2</sub> recordings (n) | 3196                 | 859                  | 1402                 | 312                  | 392                  | 231                  |
| Age (years)                     | 6 [5, 8]             | 7 [6, 8]             | 6 [4, 8]             | 7 [6, 8]             | 6 [3, 9]             | 5 [4, 7]             |
| Males (%)                       | 1735 (54.6%)         | 417 (48.5%)          | 740 (52.8%)          | 143 (45.8%)          | 254 (64.8%)          | 160 (69.3%)          |
| BMI (Kg/m <sup>2</sup> )        | 17.2<br>[15.4, 21.1] | 17.3<br>[15.5, 22.3] | 16.9<br>[15.2, 20.7] | 17.1<br>[15.4, 19.9] | 18.1<br>[15.8, 21.7] | 16.0<br>[14.7, 18.0] |
| AHI (e/h)                       | 2.1<br>[0.7, 6.3]    | 3.1<br>[1.4, 6.9]    | 1.7<br>[0.6, 5.9]    | 0.8<br>[0.4, 1.7]    | 3.3<br>[1.4, 7.8]    | 2.3<br>[0.9, 6.4]    |
| AHI < 1 (%)                     | 1015 (31.8%)         | 173 (20.1%)          | 516 (36.8%)          | 187 (59.9%)          | 77 (19.6%)           | 62 (26.8%)           |
| 1 ≤ AHI < 5 (%)                 | 1230 (38.5%)         | 395 (46.0%)          | 495 (35.2%)          | 76 (24.4%)           | 169 (43.1%)          | 97 (42.0%)           |
| 5 ≤ AHI < 10 (%)                | 447 (14.0%)          | 170 (19.8%)          | 164 (11.7%)          | 18 (7.8%)            | 63 (16.1%)           | 32 (13.9%)           |
| AHI ≥ 10 (%)                    | 504 (15.8%)          | 121 (14.1%)          | 229 (16.3%)          | 31 (9.9%)            | 83 (21.2%)           | 40 (17.3%)           |

Data are presented as median [interquartile range],  $n$  or  $n(\%)$ , BMI= Body Mass Index, AHI= Apnea Hypopnea Index.

**Figure 1:** Overview of the proposed methodology. (a) Signals segmentation, (b) CNN architecture, and (c) AHI estimation.

the input by the use of  $\lambda_C$  stacked convolutional blocks, each one composed of: convolutional layer, batch normalization, activation, pooling, and dropout (Goodfellow et al., 2016).

The convolutional layer extracts feature maps from the input data  $a[n]$  using convolutional filters (kernels) (Goodfellow et al., 2016):

$$x_i^l[n] = \sum_{k=1}^{L_C} w_k^l * a_i[n - k + 1] + b_k^l, \quad (1)$$

where  $x_i^l$  is the  $l$ 'th feature map ( $l = 1, \dots, M_C$ , being  $M_C$  the number of filters) in the convolutional block  $i = 1, \dots, \lambda_C$ ,  $w_k^l$  and  $b_k^l$  are the weights and biases of each convolutional kernel in the convolutional block  $i$ , and  $L_C$  is the kernel size.

After the convolution, batch normalization is applied to normalize the feature maps (Goodfellow et al., 2016). Then, a non-linear function is used to decide which feature maps are activated, depending on a rule or a threshold (Goodfellow et al., 2016). In this study, a rectified

linear unit (ReLU) activation function, which is the standard choice for deep-learning (Goodfellow et al., 2016), was used:

$$f(x) = \max(0, x) \quad (2)$$

The output of the ReLU is fed into a max-pooling layer, which applies a maximum operation with a pool factor  $K=2$ , which is a widely used value, in order to reduce dimensionality as well as computational cost (Goodfellow et al., 2016). Finally, dropout operation was included in the training phase in order to avoid overfitting (Goodfellow et al., 2016). Dropout randomly removes some units with a probability  $p_{drop}$  at each batch of a training epoch (Goodfellow et al., 2016).

After  $\lambda_C$  convolutional blocks, a flattening layer is used to transform the 2-D feature maps into a 1-D series (Goodfellow et al., 2016). Then, a linear activation unit is used to obtain the output of the network,  $y_{CNN}^m$  which accounts for the apneic events associated to desaturations for the corresponding input 20-min SpO<sub>2</sub> segment.

### 3.1.3. AHI estimation

Based on the output  $y_{CNN}^m$  of the CNN for each 20-min SpO<sub>2</sub> segment  $m = 1, 2, 3, \dots, N$ , the AHI of each patient can be estimated. First, the average of the output of the CNN obtained for each SpO<sub>2</sub> segment is computed:

$$y_{CNN}^{avg} = \frac{\sum_{m=1}^N y_{CNN}^m}{N} \quad (3)$$

where  $N$  is the number of 20-min SpO<sub>2</sub> segments of the oximetry signal. This step is necessary as the number of 20-min SpO<sub>2</sub> segments is different for each patient. Then, the AHI is obtained using the following expression, as shown in Figure 1 (c):

$$AHI = (\beta \cdot y_{CNN}^{avg}) + \epsilon \quad (4)$$

where  $\beta$  and  $\epsilon$  are the intercept and disturbance term of a linear regression model, which was fitted using the validation set. This linear regression corrects the trend of the CNN to underestimate the AHI, which is caused by (Devi-aene et al., 2018): (i) the AHI estimated by the CNN is obtained using the total recording time, while the AHI from PSG uses total sleep time; (ii) there are apneic events that are not associated to oxygen desaturations, so they cannot be detected by the CNN.

### 3.2. CNN training and optimization process

The training data were fed into the CNN in batches of 100 during 500 epochs. He-normal method was used to initialize the weights and biases of each layer (He et al., 2015). Then, the adaptive moment estimation (Adam) algorithm was used with an initial learning rate of 0.001 to update the weights and biases in each training batch (Kingma and Ba, 2014). Huber loss (Huber, 1964) was the function used to minimize Adam algorithm in the validation set. This loss function has a tunable hyperparameter, delta ( $\delta$ ), that allows to control the importance of outliers (Huber, 1964):

$$L(y^m, y_{CNN}^m) = \begin{cases} \frac{1}{2} \cdot (y^m - y_{CNN}^m)^2, & |y^m - y_{CNN}^m| \geq \delta \\ y^m \cdot (|y^m - y_{CNN}^m| - \frac{1}{2} \cdot \delta), & \text{otherwise} \end{cases} \quad (5)$$

where  $y^m$  is the target variable and  $y_{CNN}^m$  is the output of the CNN for a segment  $m$ . At each training epoch, training data were shuffled in order to improve the convergence of the optimization algorithm (Goodfellow et al., 2016), so the batches were different. In addition, the learning rate was decreased by a factor of 2 after 10 epochs of non-improvement in the loss function value of the validation set, which helps to obtain a converged stable set of final weights (Goodfellow et al., 2016). Finally, early stopping (Goodfellow et al., 2016) was applied to stop training after 30 epochs of non-improvement in order to reduce the training time, restoring weights to those that achieved the best performance in the validation set.

The hyperparameters of the CNN architecture to optimize were the number of filters in each convolutional layer ( $M_C$ ), the kernel size of each convolutional layer ( $L_C$ ), the number of CNN blocks ( $\lambda_C$ ), the dropout probability ( $p_{drop}$ ), and the delta parameter of the Huber loss ( $\delta$ ). Bayesian optimization with tree-structured Parzen estimator (BO-TPE) (Bergstra et al., 2011) implemented in Hyperopt library (Bergstra et al., 2015) was used to obtain the optimum values of these hyperparameters. BO-TPE is considered more efficient than grid search or random search for hyperparameters optimization, since it uses past evaluation results to form a probabilistic model that attempts to optimize the objective function in an iterative way (Snoek et al., 2012).

Keras framework with Tensorflow backend was used to implement the CNN-based architecture (Chollet, 2015). CNNs were trained on a NVIDIA GeForce RTX 2080 GPU in a Windows 10 environment.

### 3.3. Comparison with conventional approaches

The following conventional methods have been applied in order to compare the diagnostic performance of the proposed deep-learning model:

- 1) *Clinical approach*: ODI3. ODI3 was estimated as the number of desaturations of at least 3% per hour of recording (Taha et al., 1997). This parameter has shown its usefulness in the clinical OSA context, and is usually employed for comparison purposes (Tsai et al., 2013; Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019).
- 2) *Classical feature-engineering approach*: multilayer perceptron (MLP) neural network trained using features extracted from the 20-min SpO<sub>2</sub> segments. This approach is divided into the following common four steps: (i) signal preprocessing, where artifacts were removed from SpO<sub>2</sub> recordings following the methodology employed in previous studies (Hornero et al., 2017; Vaquerizo-Villar et al., 2018b,a); (ii) feature extraction, where up to 23 features were extracted from each 20-min SpO<sub>2</sub> segment, the same features as in the previous study by Hornero et al. (2017); (iii) segment-based AHI estimation; where a MLP model was trained with the set of 23 SpO<sub>2</sub> features to estimate the number of apneic events associated to desaturations in each segment; (iv) subject-based AHI estimation, where the AHI of each subject is obtained from the output of the MLP for each 20-min SpO<sub>2</sub> segment using the procedure described in Section 3.1.3.

### 3.4. Statistical Analysis

The agreement between the estimated AHI by the CNN architecture (AHI<sub>CNN</sub>) and the actual AHI from PSG

( $AHI_{PSG}$ ) was assessed by means of scatter and error distribution plots, as well as the intra-class correlation coefficient (ICC) and root mean square error (RMSE). The overall agreement of  $AHI_{CNN}$  to estimate the severity of OSA was assessed by means of the confusion matrices, as well as Cohen's kappa index (kappa) and 4-class accuracy. ICC, RMSE, kappa, and 4-class accuracy were also obtained for ODI3 and the AHI estimated by the MLP ( $AHI_{MLP}$ ). Additionally, the diagnostic ability of  $AHI_{CNN}$  was assessed for each of the AHI cutoffs that define the OSA severity degrees (1, 5, and 10 e/h) by means of sensitivity (Se, percentage of OSA positive patients rightly classified), specificity (Sp, percentage of OSA negative children rightly classified), positive predictive value (PPV, proportion of positive test results that are true positives), negative predictive value (NPV, proportion of negative test results that are true negatives), positive likelihood ratio (LR+,  $Se/(1-Sp)$ ), negative likelihood ratio (LR-,  $(1-Se)/Sp$ ), and accuracy (Acc, percentage of subjects correctly classified).

## 4. Results

### 4.1. Training and validation sets

Training and validation sets were used to optimize the CNN architecture. BO-TPE was used to find the optimum values of the hyperparameters of the CNN architecture:  $M_C$ ,  $L_C$ ,  $\lambda_C$ ,  $p_{drop}$ , and  $\delta$ . The search space of the BO-TPE is shown in Table 2. The training set was used to train the CNN models at each iteration of the BO-TPE procedure, whereas kappa was obtained in the validation set as the objective function to optimize. The training of most of the CNNs was finished by early stopping criterion after 80-120 epochs, thus contributing to reduce the training time. The results of the BO-TPE algorithm are shown in Figure 2. For each hyperparameter, the values of kappa in the validation set are given. These values are represented in a boxplot. It can be seen that there is not a high dependence of kappa on the hyperparameter values. Slightly higher overall kappa values are obtained when  $\lambda_C = 6$  and  $L_C = 5$ , as well as with increasing values of  $M_C$  and decreasing values of  $\delta$ , whereas  $p_{drop}$  had little effect on the value of kappa. Finally,  $M_C = 64$ ,  $L_C = 5$ ,  $\lambda_C = 6$ ,  $p_{drop} = 0.1$ , and  $\delta = 1.5$  were obtained as the optimum values for the hyperparameters, since this combination reached the highest kappa, as shown in Table 2.

### 4.2. Test set

#### 4.2.1. Diagnostic performance of the CNN model

Figure 3 shows the scatter plots of  $AHI_{CNN}$  compared to  $AHI_{PSG}$  in the CHAT, UofC and BUH test sets, respectively. ICC and RMSE are also shown. Points of the scatter plot of  $AHI_{CNN}$  in the CHAT test set are more concentrated near the diagonal line, which is reflected in a higher agreement (ICC=0.960 and RMSE=2.89) than in the UofC (ICC=0.917 and RMSE=5.45) and BUH test sets (ICC=0.583 and RMSE=10.44). Figure 4 shows the error distribution

**Table 2:** Search space of BO-TPE for the CNN hyperparameters.

| Hyperparameter | Search space  | Optimum value |
|----------------|---------------|---------------|
| $M_C$          | 8, 16, 32, 64 | 64            |
| $L_C$          | 3, 5, 7       | 5             |
| $\lambda_C$    | 4, 5, 6, 7, 8 | 6             |
| $p_{drop}$     | 0:0.25:0.5    | 0.1           |
| $\delta$       | 0.5:0.5:6     | 1.5           |

BO-TPE= Bayesian optimization with tree-structured Parzen estimator; CNN = Convolutional neural network;  $M_C$  = number of filters;  $L_C$  = kernel size;  $\lambda_C$  = number of convolutional blocks;  $p_{drop}$  = dropout probability;  $\delta$  = delta value of the Huber loss

plots of  $AHI_{CNN}$  in the three test sets. Mean error was low in the three test sets. Nonetheless, 95% confidence intervals of  $AHI_{CNN}$  were higher in the UofC (21.69 e/h) and BUH (28.84 e/h) test sets than in the CHAT test set (12.80 e/h). In addition, there are some outliers in  $AHI_{CNN}$  that can be observed in the UofC and BUH sets, as reported by the maximum error.

Figure 5 shows the confusion matrices of  $AHI_{CNN}$ , evaluated in the three test sets.  $AHI_{CNN}$  rightly assigned 72.8% (227/312), 60.2% (236/392), and 61.0% (141/231) of subjects in the three test sets to their actual OSA severity group. Table 3 shows diagnostic ability statistics of  $AHI_{CNN}$  for the AHI severity cutoffs of 1, 5, and 10 e/h, which are derived from the confusion matrix. Notice that  $AHI_{CNN}$  reached a higher kappa in the CHAT test set (0.515) than in the UofC (0.422) and BUH test sets (0.423). Higher performance metrics were obtained in the CHAT test set for the three AHI cutoffs, especially for the AHI cutoffs of 5 and 10 e/h.

#### 4.2.2. Comparison with conventional approaches

Table 4 shows the comparison of  $AHI_{CNN}$  with ODI3 and  $AHI_{MLP}$  in the three test sets. It can be seen that  $AHI_{CNN}$  showed a higher diagnostic capability than ODI3 and  $AHI_{MLP}$  in the CHAT, UofC, and BUH test sets, as derived from the values of ICC, RMSE, kappa, and 4-class accuracy.

Table 5 summarizes the comparison of the performance of our proposal with state-of-the-art studies aimed at simplifying the detection of pediatric OSA and its severity using the SpO<sub>2</sub> signal (Tsai et al., 2013; Villa et al., 2015; Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019). Notice that none of the studies that employed a validation approach reported a higher accuracy for the AHI cutoffs of 5 e/h and 10 e/h than the proposed CNN-based architecture in the CHAT, UofC, and BUH datasets.

## 5. Discussion

In the present study, we assessed the potential usefulness of a new CNN architecture to enhance the diagnostic ability of the oximetry signal in the context of pediatric OSA. A CNN-based deep-learning model was built

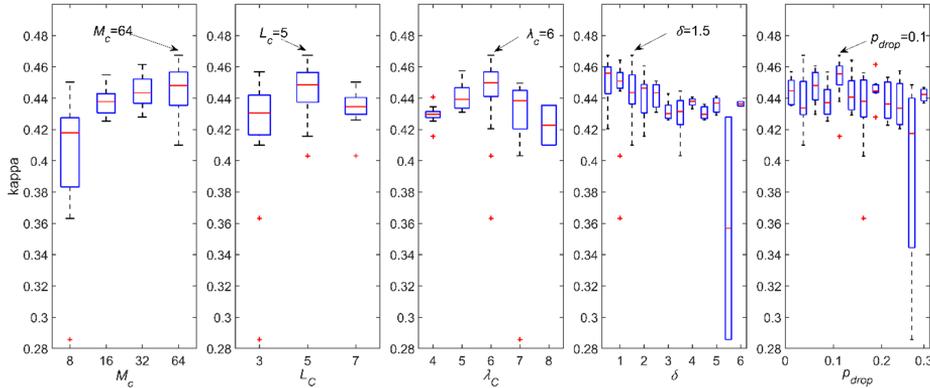


Figure 2: Results of the BO-TPE for every hyperparameter in the validation set.

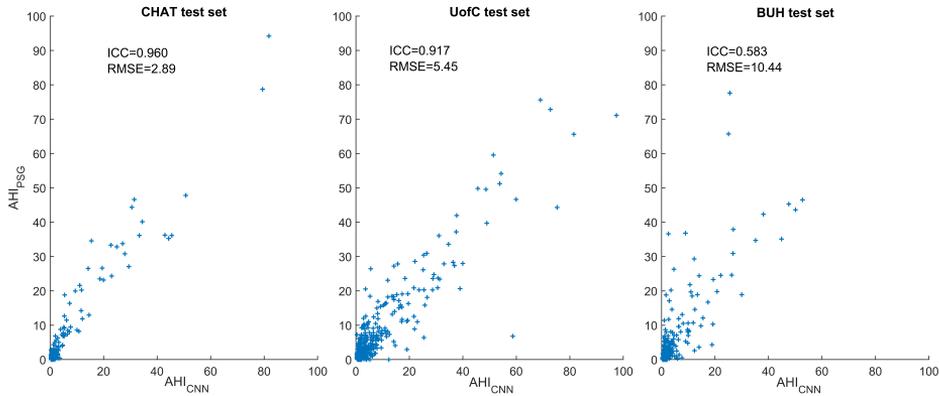


Figure 3: Scatter plots comparing  $AHI_{CNN}$  with  $AHI_{PSG}$  in the CHAT, UofC, and BUH test databases.

Table 3: Diagnostic ability of  $AHI_{CNN}$  for the AHI cutoffs=1 e/h, 5 e/h, and 10 e/h in the CHAT, UofC, and BUH test databases.

| Estimated AHI | CHAT test set |           |            | UofC test set |           |            | BUH test set |           |            |
|---------------|---------------|-----------|------------|---------------|-----------|------------|--------------|-----------|------------|
|               | AHI=1 e/h     | AHI=5 e/h | AHI=10 e/h | AHI=1 e/h     | AHI=5 e/h | AHI=10 e/h | AHI=1 e/h    | AHI=5 e/h | AHI=10 e/h |
| Se (%)        | 71.2          | 83.7      | 83.9       | 90.8          | 76.0      | 79.5       | 88.8         | 61.1      | 65.0       |
| Sp (%)        | 81.8          | 100       | 99.5       | 36.4          | 88.6      | 95.8       | 53.2         | 93.7      | 96.9       |
| PPV (%)       | 72.4          | 100       | 92.9       | 85.4          | 79.8      | 83.5       | 83.8         | 81.5      | 81.3       |
| NPV (%)       | 81.0          | 97.0      | 98.2       | 49.1          | 86.2      | 94.6       | 63.5         | 84.2      | 93.0       |
| LR+           | 3.92          | N.D       | 117.84     | 1.43          | 6.68      | 18.90      | 1.90         | 9.72      | 20.69      |
| LR-           | 0.35          | 0.16      | 0.16       | 0.25          | 0.27      | 0.21       | 0.21         | 0.42      | 0.36       |
| Acc (%)       | 77.6          | 97.4      | 97.8       | 80.1          | 83.9      | 92.3       | 79.2         | 85.5      | 91.3       |
| kappa         |               | 0.515     |            |               | 0.422     |            |              | 0.423     |            |

CNN = Convolutional neural network, AHI = apnea-hypopnea index, Se = sensitivity (%), Sp = specificity (%), PPV = positive predictive value (%), NPV = negative predictive value (%), LR+ = positive likelihood ratio, LR- = negative likelihood ratio, Acc = accuracy (%), kappa = Cohen's kappa index, N.D = not defined, CHAT = Childhood Adenotonsillectomy Trial, UofC = University of Chicago, BUH = Burgos University Hospital.

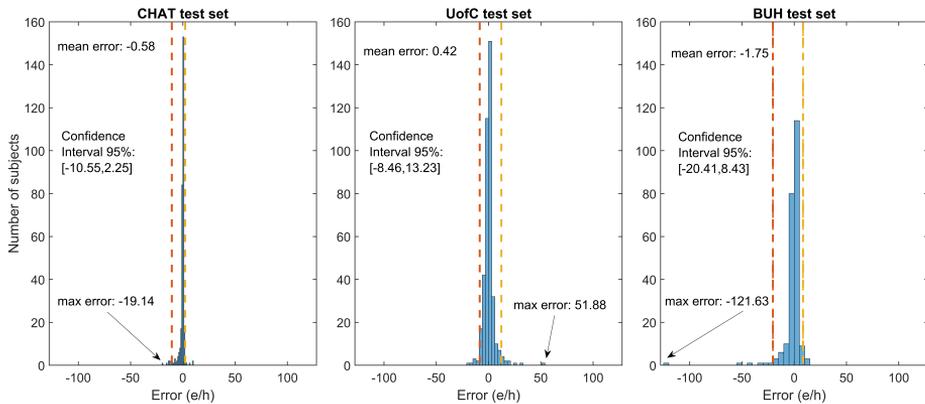


Figure 4: Error distribution of  $AHI_{CNN}$  in the CHAT, UofC, and BUH test databases.

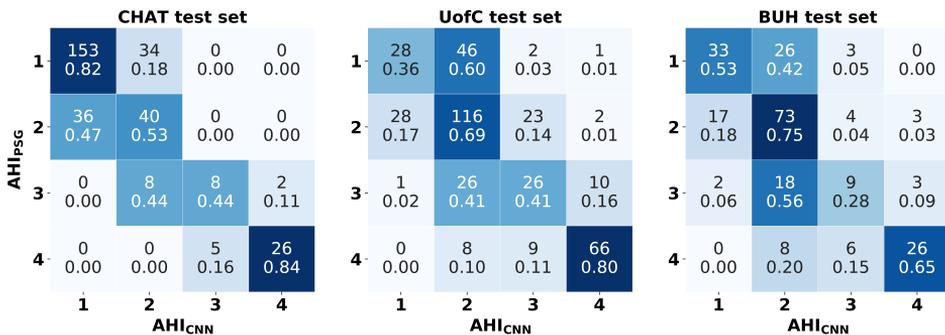


Figure 5: Confusion matrices of  $AHI_{CNN}$  in the CHAT, UofC, and BUH test databases. 1: No-OSA ( $AHI < 1$  e/h); 2: Mild OSA ( $1 \leq AHI < 5$  e/h); 3: Moderate OSA ( $5 \leq AHI < 10$  e/h); 4: Severe OSA ( $AHI \geq 10$  e/h).

and trained to estimate pediatric OSA severity using raw  $SpO_2$  data. This model was validated in a database of 3196  $SpO_2$  recordings from three different datasets. The proposed CNN model showed a high diagnostic ability, improving the diagnostic performance of  $ODI_3$  and  $AHI_{MLP}$ .

### 5.1. CNN architecture

To the best of our knowledge, this is the first study that provides a deep-learning model able to automatically detect pediatric OSA and its severity from the oximetry signal. Our results showed that the proposed CNN-based architecture is able to discern patterns linked with apneic events present in the oximetry signal of children with OSA. Recent studies have also shown the usefulness of deep-learning to analyze different physiological

signals from PSG in adult OSA patients [21]. In this regard, the studies developed by Biswal et al. (2018), Choi et al. (2018), Van Steenkiste et al. (2018), and Nikkonen et al. (2019) reached accuracies in the range 57%-91% to classify subjects into the four adult OSA severity degrees ( $AHI < 5$ ,  $5 \leq AHI < 15$ ,  $15 \leq AHI < 30$ , and  $AHI \geq 30$  e/h). Despite some of these studies reported higher accuracies, they focus on adult patients, whereas our study applies a CNN to the context of pediatric OSA. In this respect, scoring rules for apnea and hypopnea events are more restrictive in children than in adults (Iber et al., 2007). In addition,  $AHI$  cutoffs for mild, moderate and severe OSA are lower in children (1, 5, and 10 e/h in contrast to 5, 15, and 30 e/h), which changes the diagnosis and treatment of these patients (Alonso-Álvarez et al., 2011; Church, 2012; Epstein et al., 2009). Due to these remarkable dif-

**Table 4:** Diagnostic performance of  $AHI_{CNN}$  vs.  $ODI3$  and  $AHI_{MLP}$  in the CHAT, UofC, and BUH test databases.

| Test set | Method      | ICC   | RMSE  | 4-class kappa | 4-class Acc |
|----------|-------------|-------|-------|---------------|-------------|
| CHAT     | $AHI_{CNN}$ | 0.960 | 2.89  | 0.515         | 72.8        |
|          | $ODI3$      | 0.871 | 4.63  | 0.417         | 65.1        |
|          | $AHI_{MLP}$ | 0.832 | 5.51  | 0.377         | 63.3        |
| UofC     | $AHI_{CNN}$ | 0.917 | 5.45  | 0.422         | 60.2        |
|          | $ODI3$      | 0.861 | 6.21  | 0.372         | 56.6        |
|          | $AHI_{MLP}$ | 0.890 | 6.02  | 0.381         | 56.9        |
| BUH      | $AHI_{CNN}$ | 0.583 | 10.44 | 0.423         | 61.0        |
|          | $ODI3$      | 0.520 | 10.64 | 0.369         | 57.6        |
|          | $AHI_{MLP}$ | 0.500 | 11.05 | 0.306         | 52.4        |

$AHI_{CNN}$  = apnea-hypopnea index (AHI) estimated by our convolutional neural network architecture,  $ODI3$  = 3% oxygen desaturation index,  $AHI_{MLP}$  = AHI estimated by the multi-layer perceptron neural network trained with features from the blood oxygen saturation ( $SpO_2$ ) signal, ICC = intra-class correlation coefficient, RMSE = root mean squared error, kappa = Cohen's kappa index, CHAT = Childhood Adenotonsillectomy Trial, UofC = University of Chicago, BUH = Burgos University Hospital.

ferences, automated diagnosis of OSA is more challenging in children and thus higher performances are commonly reached in adult patients.

These aforementioned studies in the context of adult OSA applied different deep-learning architectures to raw PSG signals: recurrent neural networks (RNN) (Biswal et al., 2018; Van Steenkiste et al., 2018), multi-layer perceptron (MLP) (Nikkonen et al., 2019), and CNN (Choi et al., 2018). From these architectures, CNN hold advantage over RNN and MLP in terms of computational cost, since they do not include recurrent and/or fully-connected layers. This facilitates the integration of the proposed architecture in wearable and portable devices. In order to corroborate the suitability of CNNs for our problem, we also applied the RNN architecture proposed by Van Steenkiste et al. (2018). This RNN architecture did not obtain a better performance than our CNN model ( $AHI_{CNN}$  outperformed the RNN architecture in terms of ICC: 0.960 vs 0.921, 0.917 vs 0.812, and 0.583 vs 0.480 in the CHAT, UofC, and BUH test sets), while having a higher computational cost. This agrees with a recent review of deep learning for time series classification (TSC), where CNN-based architectures achieved the highest performance for TSC in an experiment where more than 8000 deep-learning models were trained and assessed on 97 different time series datasets (Ismail Fawaz et al., 2019).

With respect to the hyperparameters of the CNN architecture, Figure 2 shows the low dependence of kappa from the validation set on the optimum hyperparameter values, which highlights the reliability of the proposed solution to automatically learn OSA-related features from the oximetry signal. We also assessed the effect of varying the segment size and the overlap between segments. Different values of the segment size (5 min, 10 min, 30 min, and 60 min), and overlap (50%, and 90%)

were tested. Regarding segment size, none of the tested values achieved higher kappa in the validation set than the segment size of our optimum CNN model (20 min), which is appropriate to detect clusters of desaturations (Brouillette et al., 2000). Changing the overlap between segments did not result in a better performance while significantly increased training and validation process.

## 5.2. Diagnostic performance

As aforementioned, the AHI estimated by our proposed optimum CNN architecture ( $AHI_{CNN}$ ) outperformed a conventional clinical index  $ODI3$  as well as a classical feature-engineering approach ( $AHI_{MLP}$ ) in the three test sets. Our  $AHI_{CNN}$  achieved a higher overall agreement with  $AHI_{PSG}$ , as well as a higher diagnostic capability to predict pediatric OSA severity. In contrast to traditional clinical ( $ODI3$ ) and feature-engineering ( $AHI_{MLP}$ ) approaches,  $AHI_{CNN}$  automatically learns features from the  $SpO_2$  recordings associated to apneic events through a multi-layer architecture that provides a high level of abstraction. According to our results, CNNs can detect additional information on the OSA-related changes occurring in the  $SpO_2$  signal that helps to enhance its diagnostic ability.

Looking at the confusion matrices of Fig. 5, it can be seen that 95.2% (BUH), 96.1% (UofC), and 100% (CHAT) of class 1 (no-OSA) patients have an estimated  $AHI_{CNN} < 5$  e/h (class 1 or class 2). In addition, 94.4% (BUH), 97.8% (UofC), and 100% (CHAT) of subjects with an  $AHI_{CNN} \geq 5$  e/h actually show an  $AHI_{PSG} \geq 1$  e/h, whereas 90.6% (BUH), 96.2% (UofC), and 100% (CHAT) predicted as severe OSA ( $AHI_{CNN} \geq 10$  e/h) are at least moderate OSA patients. Hence, a possible screening protocol can be derived to show the clinical usefulness of our proposal as follows: *i*) if  $AHI_{CNN} < 1$  e/h, discard the presence of OSA because most of these patients (96.2% in BUH, 98.2% in UofC, and 100% in CHAT) will have an  $AHI_{PSG} < 5$  e/h. If symptoms persist, these children may be eventually referred to PSG, as recommended by Alonso-Álvarez et al. (2011); *ii*) if  $1 \leq AHI_{CNN} < 5$  e/h, suggest PSG, since doubts arise about the actual diagnosis of the patients; *iii*) if  $5 \leq AHI_{CNN} < 10$  e/h, consider treatment, since most probably (86.4% in BUH, 96.7% in UofC, and 100% in CHAT) these subjects have at least a mild degree of OSA; *iv*) if  $AHI_{CNN} \geq 10$  e/h, suggest treatment, since most of these children (90.6% in BUH, 96.2% in UofC, and 100% in CHAT) have an  $AHI_{PSG} \geq 5$  e/h, and also consider a further observation of these patients, since they are prone to have residual OSA after PSG (Marcus et al., 2012). This screening protocol would avoid the need for 45.9% (BUH), 50.0% (UofC), and 73.7% (CHAT) of complete PSGs, thus contributing to a reduction in the waiting lists and medical costs associated with the diagnosis of OSA, as well as to provide a more suitable diagnostic procedure for children.

Comparing the results of the proposed approach in the three test sets, it is important to highlight the high diagnostic performance obtained by  $AHI_{CNN}$  in CHAT, where

**Table 5:** Summary of state-of-the-art studies in the context of pediatric OSA detection using SpO<sub>2</sub> recordings

| Study                           | N (Total/test) | AHI (e/h) | Methods (Feature/classification)   | Validation                         | Se   | Sp   | Acc  |
|---------------------------------|----------------|-----------|--|------------------------------------|------|------|------|
| Tsai et al. (2015)              | 148/148        | 1         | ODI4 / Thresholding  | No                                 | 77.7 | 88.9 | 79.0 |
|                                 |                | 5         |  |                                    | 83.8 | 86.5 | 85.1 |
|                                 |                | 10        |  |                                    | 89.1 | 86.0 | 87.1 |
| Villa et al. (2015)             | 268/268        | 1         | Clusters of desaturations and clinical history / Thresholding                          | Direct validation*                 | 91.6 | 40.6 | 85.8 |
|                                 |                | 5         |  |                                    | 40.6 | 97.9 | 69.4 |
| Álvarez et al. (2017)           | 50/50          | 1         | Classical indices, statistical moments, PSD, and nonlinear features / Binary LR        | Bootstrapping                      | 89.6 | 71.5 | 85.5 |
|                                 |                | 3         |  |                                    | 82.9 | 84.4 | 83.4 |
|                                 |                | 5         |  |                                    | 82.2 | 83.6 | 82.8 |
| Hornero et al. (2017)           | 4191/3602      | 1         | ODI3, statistical moments, PSD, and nonlinear features / Regression MLP                | Training-test                      | 84.0 | 53.2 | 75.2 |
|                                 |                | 5         |  |                                    | 68.2 | 87.2 | 81.7 |
|                                 |                | 10        |  |                                    | 68.7 | 94.1 | 90.2 |
| Vaquerizo-Villar et al. (2018b) | 298/75         | 5         | Bispectrum, spectral features, ODI3, and anthropometric variables / Multi-class MLP    | Feature optimization-training-test | 61.8 | 97.6 | 81.3 |
|                                 |                | 10        |  |                                    | 60.0 | 94.5 | 85.3 |
| Crespo et al. (2018)            | 176/176        | 1         | Classical indices, statistical moments, PSD, and nonlinear features / LDA, QDA, and LR | Bootstrapping                      | 93.9 | 37.8 | 84.3 |
|                                 |                | 3         |  |                                    | 74.6 | 81.7 | 77.7 |
|                                 |                | 5         |  |                                    | 70.0 | 91.4 | 82.7 |
| Vaquerizo-Villar et al. (2018a) | 981/392        | 1         | DFA and ODI3 / Regression MLP  | Training-test                      | 97.1 | 23.3 | 82.7 |
|                                 |                | 5         |  |                                    | 78.8 | 83.7 | 81.9 |
|                                 |                | 10        |  |                                    | 77.1 | 94.8 | 91.1 |
| Xu et al. (2019)                | 432/432        | 1         | ODI3, statistical moments, PSD, and nonlinear features / Regression MLP                | Training-test                      | 95.3 | 19.1 | 79.6 |
|                                 |                | 5         |  |                                    | 77.8 | 80.5 | 79.4 |
|                                 |                | 10        |  |                                    | 73.5 | 92.7 | 88.2 |
| Our proposal: CHAT set          | 3196/512       | 1         | CNN architecture   | Training-validation-test           | 71.2 | 81.8 | 77.6 |
|                                 |                | 5         |  |                                    | 83.7 | 100  | 97.4 |
|                                 |                | 10        |  |                                    | 83.9 | 99.3 | 97.8 |
| Our proposal: UofC set          | 3196/392       | 1         | CNN architecture   | Training-validation-test           | 90.8 | 36.4 | 80.1 |
|                                 |                | 5         |  |                                    | 76.0 | 88.1 | 83.9 |
|                                 |                | 10        |  |                                    |      | 95.8 | 92.3 |
| Our proposal: BUH set           | 3196/231       | 1         | CNN architecture   | Training-validation-test           | 88.8 | 53.2 | 79.2 |
|                                 |                | 5         |  |                                    | 61.1 | 93.7 | 83.5 |
|                                 |                | 10        |  |                                    | 65.0 | 96.9 | 91.3 |

\* Direct validation of a scoring criteria against AHI from PSG, N= Number of subjects, CNN = Convolutional Neural Networks, AHI= apnea-hypopnea index, Se = sensitivity (%), Sp = specificity (%), Acc = accuracy (%) PSD= power spectral density, ODI3= 3% oxygen desaturation index, ODI4= 4% oxygen desaturation index, LR = Logistic Regression, MLP = Multi-layer perceptron, LDA = Linear Discriminant analysis, QDA = Quadratic discriminant analysis, CHAT = Childhood Adenotonsillectomy Trial, UofC = University of Chicago, BUH = Burgos University Hospital.

there is a higher increase in the performance of AHI<sub>CNN</sub> with respect to ODI3 and AHI<sub>MLP</sub> in terms of overall accuracy, kappa, RMSE, and ICC. The proposed CNN model also performed well in the UofC and BUH datasets. Despite not being as remarkable as in the CHAT dataset, AHI<sub>CNN</sub> also outperformed ODI3 and AHI<sub>MLP</sub> in most of the performance metrics. As it can be seen in the scatter plots (Figure 3), error distribution plots (Figure 4), and confusion matrices (Figure 5), AHI<sub>CNN</sub> performed better in the CHAT dataset than in the UofC and BUH datasets. However, the results are still remarkable considering that the optimum CNN model was trained in the CHAT dataset. In this sense, Collop (2002) state that there is a high variability in the scoring of polysomnographies among different sleep technologists, which may affect the external assessment of our proposed deep-learning methodology in two independent databases. In the current work, we tried to minimize this variability by using a validation set composed of subjects from the three

datasets to optimize the hyperparameters of the CNN architecture.

The varying diagnostic performance could also be due to some differences in the clinical characteristics among datasets. As observed in the scatter plots (Figure 3), AHI from PSG has a different distribution in each dataset. The mean values of AHI are 4.2 e/h, 9.3 e/h, and 5.9 e/h in the CHAT, UofC and BUH test sets. In addition, interquartile range are also different: 0.4-1.7 in the CHAT dataset, 1.5-9.3 in the UofC dataset, and 0.6-5.3 in the BUH dataset. The age of children is also different in each dataset. CHAT is composed of children ranging 5 to 10 years of age, whereas UofC dataset is composed of children from 0 to 13 years of age and children in the BUH dataset range from 0 up to 18 years of age. Sampling rate values of SpO<sub>2</sub> recordings also vary among datasets: (i) 1, 2, 10, 12, 16, 200, 256, and 512 Hz in the CHAT dataset; (ii) 25, 200, and 500 Hz in the UofC dataset; (iii) 200 Hz in the BUH dataset. Finally, the population

groups of CHAT and UofC datasets are children from the United States of America (USA), whereas BUH dataset is composed of Spanish patients. In this respect, there are differences in race and obesity prevalence between these countries. Health system is also different: mostly public in Spain vs. private in USA. This influences the socio-economic level of the patients, thus having a considerable impact on the health condition. Consequently, these differences in sampling rate values, age range, AHI distribution, and patient characteristics among countries may have resulted in a lower diagnostic performance in the UofC and BUH datasets. This agrees with previous studies that also reported differences in the diagnostic performance among sleep datasets with different clinical characteristics (Biswal et al., 2018; Nikkonen et al., 2019; Alvarez et al., 2013).

### 5.3. Comparison with state-of-the-art studies

Table 5 shows the details of previous studies focused on the analysis of the SpO<sub>2</sub> signal in the automated detection of pediatric OSA and its severity (Tsai et al., 2013; Villa et al., 2015; Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019). The first studies focused on the use of conventional oximetric indices (Tsai et al., 2013; Villa et al., 2015). Nonetheless, these studies did not employ a hold-out approach to further assess their methodological approaches.

Recent studies focused on the use of automated signal processing and machine learning methods to enhance the diagnostic ability of the oximetry signal (Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019). These studies followed a three-stage feature-engineering methodology to detect pediatric OSA and its severity (Álvarez et al., 2017; Hornero et al., 2017; Vaquerizo-Villar et al., 2018b; Crespo et al., 2018; Vaquerizo-Villar et al., 2018a; Xu et al., 2019). The diagnostic accuracies reported in these studies ranged between 75.2% and 85.5% Acc for an AHI cutoff of 1 e/h, 79.4%-82.8% Acc using an AHI cutoff of 5 e/h, and 85.3%-91.1% using an AHI cutoff of 10 e/h. From these studies, only Hornero et al. (2017), Xu et al. (2019), and Vaquerizo-Villar et al. (2018a) evaluated the diagnostic performance of an AHI estimation model for the common AHI cutoffs of 1, 5, and 10 e/h. As aforementioned, our optimum CNN model showed a higher diagnostic performance in the CHAT, UofC, and BUH datasets than state-of-the-art studies for the AHI cutoffs of 5 and 10 e/h. In addition, a higher Sp for the AHI cutoff of 1 e/h was obtained in the CHAT, UofC, and BUH datasets than the reported by Xu et al. (2019) and Vaquerizo-Villar et al. (2018a), which is useful to discard the presence of OSA in pediatric patients. Beyond the superior performance of our CNN model, it uses raw data, i.e., does not require neither prior pre-processing nor human-driven assumptions regarding the SpO<sub>2</sub> information needed.

### 5.4. Limitations

In spite of the promising results of our proposal, some limitations should be considered. First, the CNN model training procedures were conducted using only the CHAT database, since the other two datasets do not contain the annotation files with the time locations of apneic events. This, together with the differences in sampling rate values, age range, AHI distribution, and patient characteristics among countries, may have resulted in a lower diagnostic performance in the UofC and BUH datasets. Nonetheless, our proposed approach showed a higher diagnostic ability than a conventional clinical index, ODI3, as well as a classical feature-engineering approach, AHI<sub>MLP</sub>, in all the datasets. Another limitation is that different optimization runs could result in different values of the hyperparameters, as shown in Figure 2. However, preliminary analysis on our data showed that kappa values in the validation set were similar among different runs, which highlights the reliability of our CNN architecture. Regarding the explanation of the features extracted by the CNN, a new limitation arises. In this regard, the application of methods for explainable deep-learning models would help to further understand the perturbations in oximetry dynamics caused by apneic events, as well as the influence of the different elements of the CNN architecture. Future research may also focus on the use of pretrained deep-learning architectures especially suited for the time series classification field, which might increase the diagnostic performance of traditional architectures based on CNN and RNN, analogous to the pretrained deep-learning networks existing in the image processing field (Canziani et al., 2016). Another limitation is that we used the AHI without including central sleep apnea (CSA) events, as originally conducted in the study that designed the CHAT database (Marcus et al., 2013). In this respect, our proposal could also be used to estimate other physiological parameters, such as the apnea index, obstructive apnea index, central apnea index, and/or ODI. Additionally, the use of SpO<sub>2</sub> together with other physiological signals from PSG may help to improve the diagnostic ability of our proposal at the cost of higher complexity in the test, since some physiological perturbation of apneic events may not be detected by the oximetry signal alone (Marcus et al., 2012). Finally, another future goal would be further validation of our proposed methodology in a database of oximetry signals recorded at home.

## 6. Conclusions

In summary, we investigated the ability of a novel deep-learning model based on CNN to automatically detect pediatric OSA and its severity from the raw oximetry signal. Our results suggest that deep learning is an appropriate tool to automatically learn discriminative features from oximetry dynamics associated to apneic events. The proposed CNN architecture reached a high

diagnostic performance, outperforming the ODI3, a clinical approach, as well as the  $AH_{MLP}$  from a conventional feature-engineering approach. In addition, we achieved higher performance than the reported by previous studies, particularly for moderate-to-severely affected children. The extensive validation of our proposal in three independent datasets as well as the design of a screening protocol highlight the applicability of our results. Therefore, we conclude that deep-learning techniques could be potentially used to enhance the diagnostic ability of the oximetry signal in the context of pediatric OSA.

### Funding

This work was supported by ‘Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación’ and ‘European Regional Development Fund (FEDER)’ under projects DPI2017-84280-R and RTC-2017-6516-1, by European Commission and ‘FEDER’ under project ‘Análisis y correlación entre la epigenética y la actividad cerebral para evaluar el riesgo de migraña crónica y episódica en mujeres’ (‘Cooperation Programme Interreg V-A Spain-Portugal POCTEP 2014–2020’), by ‘Sociedad Española de Neumología y Cirugía Torácica (SEPAR)’ under project 649/2018, by ‘Sociedad Española de Sueño (SES)’ under project ‘Beca de Investigación SES 2019’, and by ‘Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN), Spain’ through ‘Instituto de Salud Carlos III’ co-funded with FEDER funds. The Childhood Adenotonsillectomy Trial (CHAT) was supported by ‘National Institutes of Health (NIH)’ grants HL083075, HL083129, UL1-RR-024134, and UL1 RR024989. The National Sleep Research Resource was supported by the ‘National Heart, Lung, and Blood Institute’ (R24 HL114473, 75N92019R002).

F. Vaquerizo-Villar was in receipt of a ‘Ayuda para contratos predoctorales para la Formación de Profesorado Universitario (FPU)’ grant from the ‘Ministerio de Educación, Cultura y Deporte’ (FPU16/02938). D. Álvarez is supported by a ‘Ramón y Cajal’ grant (RYC2019-028566-I) from the ‘Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación’ co-funded by the European Social Fund (ESF). V. Barroso-García and E. Santamaría-Vázquez were in a receipt of a ‘Ayuda para financiar la contratación predoctoral de personal investigador’ grant from the ‘Consejería de Educación de la Junta de Castilla y León’ and the ‘European Social Fund’. L. Kheirandish-Gozal and D. Gozal were supported by NIH grants HL130984, HL140548, and AG061824.

### Acknowledgments

In the loving memory of Dr. Joaquín Terán-Santos and Dra. María. L. Alonso-Álvarez, for all their help and encouragement, and for providing us with the BUH database.

### References

- Alonso-Álvarez, M. L., Canet, T., Cubell-Alarco, M., Estivill, E., Fernández, E., Gozal, D., Jurado-Luque, M. J., Lluch-Roselló, M. A., Martínez-Pérez, F., Merino-Andren, M., Pin-Arboledas, G., Roure, N., Sanmartí, F. X., Sans-Capdevila, O., Segarra-Isern, J., 2011. Documento de consenso del síndrome de apneas-hipopneas durante el sueño en niños. *Arch Bronconeumol* 47 (Supl 5), 2–18.
- Álvarez, D., Alonso-Álvarez, M. L., Gutiérrez-Tobal, G. C., Crespo, A., Kheirandish-Gozal, L., Gozal, D., Terán-Santos, J., Campo, F. D., 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 (5), 7–11.
- Álvarez, D., Hornero, R., Marcos, J. V., Wessel, N., Penzel, T., Glos, M., Del Campo, F., 2013. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis. *International journal of neural systems* 23 (05), 1350020.
- Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. In: 25th annual conference on neural information processing systems (NIPS 2011). pp. 1–9.
- Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., Cox, D. D., 2015. Hyperopt: A Python library for model selection and hyperparameter optimization. *Computational Science and Discovery* 8 (1), 014008.
- Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., Marcus, C. L., Mehra, R., Parthasarathy, S., Quan, S. F., Redline, S., Strohl, K. P., Davidson Ward, S. L., Tangredi, M. M., 2012. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. *Journal of clinical sleep medicine* 8 (5), 597.
- Biswal, S., Sun, H., Goparaju, B., Brandon Westover, M., Sun, J., Bianchi, M. T., 2018. Expert-level sleep scoring with deep neural networks. *Journal of the American Medical Informatics Association* 25 (12), 1643–1650.
- Brouillette, R. T., Morielli, A., Leimanis, A., Waters, K. A., Luciano, R., Ducharme, F. M., 2000. Nocturnal pulse oximetry as an abbreviated testing modality for pediatric obstructive sleep apnea. *Pediatrics* 105 (2), 405–412.
- Canziani, A., Paszke, A., Curlicio, E., 2016. An analysis of deep neural network models for practical applications. arXiv preprint arXiv:1605.07678.
- Choi, S. H., Yoon, H., Kim, H. S., Kim, H. B., Kwon, H. B., Oh, S. M., Lee, Y. J., Park, K. S., 2018. Real-time apnea-hypopnea event detection during sleep by convolutional neural networks. *Computers in Biology and Medicine* 100 (February), 123–131.
- Chollet, F., 2015. Keras. URL <https://github.com/fchollet/keras>
- Church, G. D., 2012. The Role of Polysomnography in Diagnosing and Treating Obstructive Sleep Apnea in Pediatric Patients. *Current problems in pediatric and adolescent health care* 42 (1), 22–25.
- Collop, N. A., 2002. Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep medicine* 5 (1), 43–47.
- Crespo, A., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Cerezo-Hernández, A., Gozal, D., Hornero, R., Del Campo, F., 2018. Assessment of oximetry-based statistical classifiers as simplified screening tools in the management of childhood obstructive sleep apnea. *Sleep and Breathing* 22 (4), 1063–1073.
- del Campo, F., Crespo, A., Cerezo-Hernández, A., Gutiérrez-Tobal, G. C., Hornero, R., Álvarez, D., 2018. Oximetry use in obstructive sleep apnea. *Expert Review of Respiratory Medicine* 12 (8), 665–681.
- Deviaene, M., Testelmans, D., Buysse, B., Borzé, P., Van Huffel, S., Varon, C., 2018. Automatic screening of sleep apnea patients based on the spo 2 signal. *IEEE journal of biomedical and health informatics* 23 (2), 607–617.
- Ebrahimi, Z., Loni, M., Daneshlab, M., Gharebaghi, A., 2020. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X* 7, 100033.
- Epstein, L. J., Kristo, D., Strollo, P. J., Friedman, N., Malhotra, A., Patil, S. P., Ramar, K., Rogers, R., Schwab, R. J., Weaver, E. M., Weinstein, M. D., 2009. Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine, 2009. Clinical guideline for the evalu-

- ation, management and long-term care of obstructive sleep apnea in adults. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine* 5 (3), 263–76.
- Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., Acharya, U. R., 2018. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine* 161, 1–15.
- Faust, O., Razaghi, H., Barika, R., Ciaccio, E. J., Acharya, U. R., 2019. A review of automated sleep stage scoring based on physiological signals for the new millennia. *Computer methods and programs in biomedicine* 176, 81–91.
- Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. *Deep learning*. Vol. 1. MIT press Cambridge.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*. pp. 1026–1034.
- Hornero, R., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Philby, M. F., Alonso-Álvarez, M. L., Álvarez, D., Dayyat, E. A., Xu, Z., Huang, Y.-S., Tamae Kakazu, M., Li, A. M., Van Eyck, A., Brockmann, P. E., Ehsan, Z., Simakajornboon, N., Kaditis, A. G., Vaquerizo-Villar, F., Crespo Sedano, A., Sans Capdevila, O., von Lukowicz, M., Terán-Santos, J., Del Campo, F., Poets, C. F., Ferreira, R., Bertran, K., Zhang, Y., Schuen, J., Verhulst, S., Gozal, D., 2017. Nocturnal Oximetry-based Evaluation of Habitually Snoring Children. *American Journal of Respiratory and Critical Care Medicine* 196 (12), 1591–1598.
- Huber, P. J., 1964. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* 35 (1), 73–101.
- Iber, C., Ancoli-Israel, S., Chesson, A., Quan, S. F., 2007. *The AASM Manual for the Scoring of Sleep and Associated Events: Clinical Terminology and Technical Specification*. *Journal of Clinical Sleep Medicine* 5 (7), 752.
- Ismael Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P. A., 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33 (4), 917–963.
- Kaditis, A., Kheirandish-Gozal, L., Gozal, D., 2016. Pediatric OSAS : Oximetry can provide answers when polysomnography is not available. *Sleep Medicine Reviews* 27, 96–105.
- Katz, E. S., Mitchell, R. B., Ambrosio, C. M. D., 2012. Obstructive Sleep Apnea in Infants. *American journal of respiratory and critical care medicine* 185 (8), 805–816.
- Kheirandish-Gozal, L., 2010. What is “abnormal” in pediatric sleep? *Respiratory Care* 55 (10), 1366–1376.
- Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Marcus, C. L., Brooks, L. J., Ward, S. D., Draper, K. A., Gozal, D., Halbower, A. C., Jones, J., Lehmann, C., Schechter, M. S., Sheldon, S., Shiffman, R. N., Spruyt, K., 2012. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics* 130 (3), e714–e755.
- Marcus, C. L., Moore, R. H., Rosen, C. L., Giordani, B., Garetz, S. L., Taylor, H. G., Mitchell, R. B., Amin, R., Katz, E. S., Arens, R., Paruthi, S., Muzumdar, H., Gozal, D., Thomas, N. H., Ware, J., Beebe, D., Snyder, K., Elden, L., Sprecher, R. C., Willging, P., Jones, D., Bent, J. P., Hoban, T., Chervin, R. D., Ellenberg, S. S., Redline, S., 2013. A Randomized Trial of Adenotonsillectomy for Childhood Sleep Apnea. *New England Journal of Medicine* 368 (25), 2366–2376.
- McClatchey, K. D., 2002. *Clinical laboratory medicine*. Lippincott Williams & Wilkins.
- Mostafa, S. S., Mendonça, F., G Ravelo-García, A., Morgado-Dias, F., 2019. A systematic review of detecting sleep apnea using deep learning. *Sensors* 19 (22), 4954.
- Murat, F., Yildirim, O., Talo, M., Baloglu, U. B., Demir, Y., Acharya, U. R., 2020. Application of deep learning techniques for heartbeats detection using ecg signals-analysis and review. *Computers in biology and medicine*, 105726.
- Netzer, N., Eliasson, A. H., Netzer, C., Kristo, D. A., 2001. Overnight pulse oximetry for sleep-disordered breathing in adults: a review. *Chest* 120 (2), 625–635.
- Nikkonen, S., Afara, I. O., Leppänen, T., Töyräs, J., 2019. Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea. *Scientific reports* 9 (1), 13200.
- Nixon, G. M., Kermack, A. S., Davis, G. M., Manoukian, J. J., Brown, A., Brouillette, R. T., 2004. Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry. *Pediatrics* 113 (1), e19–e25.
- Redline, S., Amin, R., Beebe, D., Chervin, R. D., Garetz, S. L., Giordani, B., Marcus, C. L., Moore, R. H., Rosen, C. L., Arens, R., Gozal, D., Katz, E. S., Mitchell, R. B., Muzumdar, H., Taylor, H., Thomas, N., Ellenberg, S., 2011. The Childhood Adenotonsillectomy Trial (CHAT): Rationale, Design, and Challenges of a Randomized Controlled Trial Evaluating a Standard Surgical Procedure in a Pediatric Population. *Sleep* 34 (11), 1509–1517.
- Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., Faubert, J., 2019. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering* 16 (5), 051001.
- Snoek, J., Larochelle, H., Adams, R. P., 2012. Practical bayesian optimization of machine learning algorithms. *arXiv preprint arXiv:1206.2944*.
- Taha, B. H., Dempsey, J. A., Weber, S. M., Badr, M. S., Skatrud, J. B., Young, T. B., Jacques, A. J., Seow, K., 1997. Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep* 20 (11), 991–1001.
- Tan, H.-L., Gozal, D., Ramirez, H. M., Bandla, H. P. R., Kheirandish-Gozal, L., 2014. Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea. *Sleep* 37 (2), 255–260.
- Tsai, C.-M., Kang, C.-H., Su, M.-C., Lin, H.-C., Huang, E.-Y., Chen, C.-C., Hung, J.-C., Niu, C.-K., Liao, D.-L., Yu, H.-R., 2013. Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. *International journal of pediatric otorhinolaryngology* 77 (8), 1286–1290.
- Van Steenkiste, T., Groenendaal, W., Deschrijver, D., Dhaene, T., 2018. Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks. *IEEE journal of biomedical and health informatics* 23 (6), 2554–2564.
- Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., Del Campo, F., Gozal, D., Hornero, R., 2018a. Detrended fluctuation analysis of the oximetry signal to assist in paediatric sleep apnoea-hypopnoea syndrome diagnosis. *Physiological measurement* 39 (11), 114006.
- Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., Del Campo, F., Gozal, D., Hornero, R., 2018b. Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings. *Computer Methods and Programs in Biomedicine* 156, 141–149.
- Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Del Campo, F., Gozal, D., Hornero, R., 2019. Convolutional neural networks to detect pediatric apnea-hypopnea events from oximetry. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 3555–3558.
- Villa, M. P., Pietropaoli, N., Supino, M. C., Vitelli, O., Rabasco, J., Evangelisti, M., Del Pozzo, M., Kaditis, A. G., 2015. Diagnosis of pediatric obstructive sleep apnea syndrome in settings with limited resources. *JAMA Otolaryngology-Head & Neck Surgery* 141 (11), 990–996.
- Xu, Z., Gutiérrez-Tobal, G. C., Wu, Y., Kheirandish-Gozal, L., Ni, X., Hornero, R., Gozal, D., 2019. Cloud algorithm-driven oximetry-based diagnosis of obstructive sleep apnoea in symptomatic habitually snoring children. *European Respiratory Journal* 53 (2).

# Appendix B

## Scientific achievements

### B.1 Publications

#### B.1.1 Papers indexed in the Journal Citation Reports (JCR)

1. Adrián Martín-Montero, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, **Fernando Vaquerizo-Villar**, Daniel Álvarez, Félix del Campo, David Gozal, Roberto Hornero, “Heart Rate Variability as a Potential Biomarker of Pediatric Obstructive Sleep Apnea Resolution”, *Sleep*, Accepted, 2021. DOI: 10.1093/sleep/zsab214. Impact factor in 2020 (last year available): 5.849, Q1 in “NEUROSCIENCES” (JCR-WOS).
2. Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Andrea Crespo, Leila Kheirandish-Gozal, David Gozal, Félix del Campo, Roberto Hornero, “Ensemble-learning regression to estimate sleep apnea severity using at-home oximetry in adults”, *Applied Soft Computing*, November, 2021. DOI: 10.1016/j.asoc.2021.107827. Impact factor in 2020 (last year available): 6.725, D1 in “COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS” (JCR-WOS).
3. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Eduardo Santamaría-Vázquez, Félix del Campo, David Gozal, Roberto Hornero, “A convolutional neural network architecture to enhance oximetry ability to diagnose pediatric obstructive sleep apnea”, *IEEE Journal of Biomedical and Health Informatics*, vol. 25 (8), August, 2021, DOI: 10.1109/JBHI.2020.3048901. Im-

- pact factor in 2020 (last year available): 5.772, D1 in "MATHEMATICAL & COMPUTATIONAL BIOLOGY" (JCR-WOS).
4. Hua Qin, Nicolas Steenbergen, Martin Glos, Niels Wessel, Jan Kraemer, **Fernando Vaquerizo-Villar**, Thomas Penzel, "The Different Facets of Heart Rate Variability in Obstructive Sleep Apnea", *Frontiers in Psychiatry*, July, 2021, DOI: 10.3389/fpsyt.2021.642333. Impact factor in 2020 (last year available): 4.157, Q2 in "PSYCHIATRY" (JCR-WOS).
  5. Hua Qin, Brendan T. Keenan, Diego R. Mazotti, **Fernando Vaquerizo-Villar**, Jan F. Kraemer, Niels Wessel, Sergio Tufik, Lia Bittencourt, Peter A. Cistulli, Philip de Chazal, Kate Sutherland, Bhajan Singh, Allan I. Pack, Ning-Hung Chen, Ingo Fietze, Thorarinn Gislason, Steven Holfinger, Ulysses J. Magalang, Thomas Penzel, "Heart Rate Variability during Wakefulness as a Marker of Obstructive Sleep Apnea Severity", *Sleep*, vol. 44 (5), May, 2021, DOI: 10.1093/sleep/zsab018. Impact factor in 2020 (last year available): 5.849, Q1 in "NEUROSCIENCES" (JCR-WOS).
  6. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Félix del Campo, David Gozal, and Roberto Hornero, "Wavelet Analysis of Overnight Airflow to Detect Obstructive Sleep Apnea in Children", *Sensors*, vol. 21 (4), pp. 1491, February, 2021, DOI: 10.3390/s21041491. Impact factor in 2020 (last year available): 3.576, Q1 in "INSTRUMENTS & INSTRUMENTATION" (JCR-WOS).
  7. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Félix del Campo, David Gozal, and Roberto Hornero, "Bispectral Analysis of Overnight Airflow to Improve the Pediatric Sleep Apnea Diagnosis", *Computers in Biology and Medicine*, vol. 129, pp. 104167, February, 2021, DOI: 10.1016/j.combiomed.2020.104167. Impact factor in 2020 (last year available): 4.589, Q1 in "MATHEMATICAL & COMPUTATIONAL BIOLOGY" (JCR-WOS).
  8. Daniel Álvarez, Carmen A. Arroyo, Julio de Frutos, Andrea Crespo, Ana Cerezo-Hernández, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Fernando Moreno, Tomás Ruíz-Albi, Roberto Hornero, Félix del Campo, "Assessment of Nocturnal Autonomic Cardiac Imbalance in Positional Obstructive Sleep Apnea. A Multiscale

- Nonlinear Approach", *Entropy*, vol. 22 (12), pp. 1404, December, 2020, DOI: 10.3390/e22121404. Impact factor in 2020: 2.524, Q2 in "PHYSICS, MULTIDISCIPLINARY" (JCR-WOS).
9. Eduardo Santamaría-Vázquez, Víctor Martínez-Cagigal, **Fernando Vaquerizo-Villar**, Roberto Hornero, "EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-based Brain-Computer Interfaces", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28 (12), pp. 2773 - 2782, December, 2020, DOI: 10.1109/TNSRE.2020.3048106. Impact factor in 2020: 3.802, D1 in "REHABILITATION" (JCR-WOS).
  10. Daniel Álvarez, Ana Cerezo-Hernández, Andrea Crespo, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Fernando Moreno, Ainhoa Arroyo, Tomás Ruíz-Albi, Félix del Campo, Roberto Hornero, "A machine learning-based test for adult sleep apnoea screening at home using oximetry and airflow", *Scientific Reports*, vol. 10, pp. 5332, March, 2020, DOI: 10.1038/s41598-020-62223-4. Impact factor in 2020: 4.379, Q1 in "MULTIDISCIPLINARY SCIENCES" (JCR-WOS).
  11. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Pablo Núñez, Félix del Campo, David Gozal, and Roberto Hornero, "Usefulness of recurrence plots from airflow recordings to aid in paediatric sleep apnoea diagnosis", *Computer Methods and Programs in Biomedicine*, vol. 183, pp. 105083, January, 2020, DOI: 10.1016/j.cmpb.2019.105083. Impact factor in 2020: 5.428, Q1 in "COMPUTER SCIENCE, THEORY & METHODS" (JCR-WOS).
  12. Daniel Álvarez, Ana Sánchez, Ana Andrés, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Roberto Hornero, Félix del Campo, "Influence of chronic obstructive pulmonary disease and moderate-to-severe sleep apnoea in overnight cardiac autonomic modulation: time, frequency and non-linear analyses", *Entropy*, vol. 21 (4), pp. 381, April, 2019, DOI: 10.3390/e21040381. Impact factor in 2019: 2.494, Q2 in "PHYSICS, MULTIDISCIPLINARY" (JCR-WOS).
  13. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Wavelet analysis of oximetry

- recordings to assist in the automated detection of moderate-to-severe pediatric sleep apnea-hypopnea syndrome”, *PLOS One*, vol. 120, pp. 155–166, December, 2018, DOI: 10.1371/journal.pone.0208502. Impact factor in 2018: 2.776, Q2 in “MULTIDISCIPLINARY SCIENCES” (JCR-WOS).
14. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, “Detrended fluctuation analysis of the oximetry signal to assist in paediatric sleep apnoea-hypopnoea syndrome diagnosis”, *Physiological Measurement*, vol. 39(11), p. 114006, November, 2018, DOI: 10.1088/1361-6579/aae66a. Impact factor in 2018: 2.246, Q3 in “ENGINEERING, BIOMEDICAL” (JCR-WOS).
  15. Daniel Álvarez, Andrea Crespo, **Fernando Vaquerizo-Villar**, Gonzalo C. Gutiérrez-Tobal, Ana Cerezo-Hernández, Verónica Barroso-García, J. Mark Ansermino, Guy A. Dumont, Roberto Hornero, Félix del Campo, Ainara Garde, “Symbolic dynamics to enhance diagnostic ability of portable oximetry from the Phone Oximeter in the detection of paediatric sleep apnoea”, *Physiological Measurement*, vol. 39 (10), pp. 104002, October, 2018, DOI: 10.1088/1361-6579/aae2a8. Impact factor in 2018: 2.246, Q3 in “ENGINEERING, BIOMEDICAL” (JCR-WOS).
  16. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, “Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings”, *Computer Methods and Programs in Biomedicine*, vol. 156, p. 141-149, March, 2018, DOI: 10.1016/j.cmpb.2017.12.020. Impact factor in 2018: 3.424, Q1 in “COMPUTER SCIENCE, THEORY & METHODS” (JCR-WOS).
  17. Roberto Hornero, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Mona F. Philby, María Luz Alonso-Álvarez, Daniel Álvarez, Ehab A. Dayyat, Zhifei Xu, Yu-Shu Huang, Maximiliano Tamae Kakazu, Albert M. Li, Annelies Van Eyck, Pablo E. Brockmann, Zarmina Ehsan, Narong Simakajornboon, Athanasios G. Kaditis, **Fernando Vaquerizo-Villar**, Andrea Crespo, Óscar Sans-Capdevila, Magnus von Lukowicz, Joaquín Terán-Santos, Félix del Campo, Christian F. Poets, Rosario Ferreira, Katalina Bertran, Yamei Zhang, John Schuen, Stijn Verhulst, David Gozal, “Nocturnal Oximetry-based Evaluation of Habitually Snoring Children”, *American*

*Journal of Respiratory and Critical Care Medicine*, vol. vol. 196 (12), pp. 1591-1598, December, 2017, DOI: 10.1164/rccm.201705-0930OC. Impact factor in 2017: 15.239, D1 in "RESPIRATORY SYSTEM" (JCR-WOS).

18. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Andrea Crespo, Félix del Campo, David Gozal, and Roberto Hornero, "Irregularity and Variability Analysis of Airflow Recordings to Facilitate the Diagnosis of Paediatric Sleep Apnoea-Hypopnoea Syndrome", *Entropy*, vol. 19, pp. 447, September, 2017, DOI: 10.3390/e19090447. Impact factor in 2017: 2.305, Q2 in "PHYSICS, MULTIDISCIPLINARY" (JCR-WOS).
19. Andrea Crespo, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, María Luz Alonso-Álvarez, Joaquín Terán-Santos, Roberto Hornero, Félix del Campo, "Multiscale Entropy Analysis of Unattended Oximetric Recordings to Assist in the Screening of Paediatric Sleep Apnoea at Home", *Entropy*, vol. 19, pp. 284, June, 2017, DOI: 10.3390/e1906028. Impact factor in 2017: 2.305, Q2 in "PHYSICS, MULTIDISCIPLINARY" (JCR-WOS).

### B.1.2 International conferences

1. Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, Andrea Crespo, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Fernando Moreno, Félix del Campo, Roberto Hornero, "Assessment of least square boosting to estimate apnea-hypopnea index from at-home oximetry recordings", *8<sup>th</sup> European Medical and Biological Engineering Conference (EMBE 2020)*, ISBN: 978-961-243-411-3, pp. 67, Portoroz (Slovenia), November 29 - December 3, 2020.
2. Javier Gómez-Pilar, Georg Northoff, **Fernando Vaquerizo-Villar**, Jesús Poza, Gonzalo C. Gutiérrez-Tobal, Roberto Hornero, "Intraindividual Characterization of the Sleep Spindle Variability in Healthy Subjects", *42<sup>th</sup> Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2020)*, ISBN: 978-1-7281-1990-8, pp. 3473-3476, Montreal (Canada), July 20 - July 24, 2020, DOI: 10.1109/EMBC44109.2020.9176315.
3. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Javier Gómez-Pilar, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Automatic Assessment of Pediatric Sleep Apnea Severity Using Overnight Oximetry and Convolutional

- Neural Networks,” *42th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2020)*, ISBN: 978-1-7281-1990-8, pp. 633-636, Montreal (Canada), July 20 - July 24, 2020, DOI: 10.1109/EMBC44109.2020.9176342.
4. María García, Carlos Gómez, Jesús Poza, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, Javier Gómez-Pilar, **Fernando Vaquerizo-Villar**, Pablo Núñez, Roberto Hornero, “Tutorización de Trabajos Fin de Grado y Trabajos Fin de Máster empleando Trello y actividades formativas específicas”, *Congreso Internacional Virtual USATIC 2020. Ubicuo y Social: Aprendizaje con TIC*, ISBN: 978-84-18321-01-6, pp. 36, Online (Spain), July 8 - July 10, 2020.
  5. Jesús Poza, Roberto Hornero, **Fernando Vaquerizo-Villar**, Daniel Álvarez, “Desarrollo y evaluación del modelo de aprendizaje inverso en la docencia de “Tratamiento de Señales Biomédicas””, *I Congreso Internacional de Innovación Docente e Investigación en Educación Superior*, ISBN: 978-84-09-16343-4, pp. 621, Madrid (Spain), November 20 - November 22, 2019.
  6. Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, Andrea Crespo, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Fernando Moreno, Félix del Campo, Roberto Hornero, “A bagging-based automatic method to estimate apnea-hypopnea index from home-oximetry recordings”, *European Respiratory Society (ERS) International Congress 2019*, Madrid (Spain), September 28 - October 2, 2019, DOI: 10.1183/13993003.congress-2019.PA822.
  7. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Félix del Campo, David Gozal, Roberto Hornero, “Usefulness of Spectral Analysis of Respiratory Rate Variability to Help in Pediatric Sleep Apnea-Hypopnea Syndrome Diagnosis”, *41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2019)*, ISBN: 978-1-5386-1311-5, pp. 4580-4583, Berlin (Germany), July 23 - July 27, 2019, DOI: 10.1109/EMBC.2019.8857719.
  8. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Félix del Campo, David Gozal, Roberto Hornero, “Convolutional Neural Networks to Detect Pediatric Apnea- Hypopnea Events from Oximetry,”, *41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*

- (EMBC 2019), ISBN: 978-1-5386-1311-5, pp. 3555-3558, Berlin (Germany), July 23 - July 27, 2019, DOI: 10.1109/EMBC.2019.8857934.
9. Daniel Álvarez, Ainara Garde, Andrea Crespo, **Fernando Vaquerizo-Villar**, Gonzalo C. Gutiérrez-Tobal, Ana Cerezo-Hernández, J. Mark Ansermino, Guy A. Dumont, Roberto Hornero, Félix del Campo, "Usefulness of symbolic dynamics to characterize oximetric recordings from a smartphone in the detection of pediatric sleep apnea", *European Respiratory Society (ERS) International Congress 2018*, pp. 52:PA554, París (France), September 15 - September 19, 2018, DOI: 10.1183/13993003.congress-2018.PA554.
  10. Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, **Fernando Vaquerizo-Villar**, Daniel Álvarez, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Bispectral Analysis to Enhance Oximetry as a Simplified Alternative for Pediatric Sleep Apnea Diagnosis", *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2018)*, ISBN: 978-1-5386-3646-6, pp. 175-178, Honolulu (United States), July 17 - July 21, 2018, DOI: 10.1109/EMBC.2018.8512248.
  11. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Leila Kheirandish-Gozal, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Improving the Diagnostic Ability of Oximetry Recordings in Pediatric Sleep Apnea-Hypopnea Syndrome by Means of Multi-Class AdaBoost", *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2018)*, ISBN: 978-1-5386-3646-6, pp. 167-170, Honolulu (United States), July 17 - July 21, 2018, DOI: 10.1109/EMBC.2018.8512264.
  12. Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Adrián Martín-Montero, Andrea Crespo, Félix del Campo, Roberto Hornero, "Pulse rate variability analysis to enhance oximetry as at-home alternative for sleep apnea diagnosing", *World Congress on Medical Physics & Biomedical Engineering (IUPESM 2018)*, ISBN: 978-981-10-9038-7, pp. 213-218, Praga (Czech Republic), June 3 - June 8, 2018, DOI: 10.1007/978-981-10-9038-7\_39.
  13. Andrea Crespo, **Fernando Vaquerizo-Villar**, Daniel Álvarez González, Gonzalo C. Gutiérrez Tobal, Verónica Barroso-García, Ana Cerezo-Hernández, Graciela López-Muñiz, Leila Kheirandish-Gozal, David Gozal,

- Roberto Hornero, Félix del Campo, "Automated detection of childhood sleep apnea using discrete wavelet transform of nocturnal oximetry and anthropometric variables", *European Respiratory Society (ERS) International Congress 2017*, Milán (Italy), September 9 - September 13, 2017, DOI: 10.1183/1393003.congress-2017.PA1308.
14. Gonzalo C. Gutiérrez-Tobal, Julio de Frutos, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Andrea Crespo, Félix del Campo, Roberto Hornero, "A Bayesian Neural Network Approach to Compare the Spectral Information from Nasal Pressure and Thermistor Air-flow in the Automatic Sleep Apnea Severity Estimation", *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2017)*, ISBN: 978-1-5090-2809-2, pp. 3741-3744, Jeju (South Korea), July 11 - July 15, 2017, DOI: 10.1109/EMBC.2017.8037670.
  15. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Leila Kheirandish-Gozal, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Usefulness of discrete wavelet transform in the analysis of oximetry signals to assist in childhood sleep apnea-hypopnea syndrome diagnosis", *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2017)*, ISBN: 978-1-5090-2809-2, pp. 3753-3756, Jeju (South Korea), July 11 - July 15, 2017, DOI: 10.1109/EMBC.2017.8037673.
  16. Carlos Gómez, **Fernando Vaquerizo-Villar**, Jesús Poza, Saúl J. Ruíz-Gómez, Miguel A. Tola-Arribas, Mónica Cano, Roberto Hornero, "Bispectral Analysis of Spontaneous EEG Activity from Patients with Moderate Dementia Due to Alzheimer's Disease", *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2017)*, ISBN: 978-1-5090-2809-2, pp. 422-425, Jeju (South Korea), July 11 - July 15, 2017, DOI: 10.1109/EMBC.2017.8036852.
  17. Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, Andrea Crespo, Ainhoa Arroyo, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Félix del Campo, Roberto Hornero, "Multi-Class AdaBoost to Detect Sleep Apnea-Hypopnea Syndrome Severity from Oximetry Recordings Obtained at Home", *Global Medical Engineering Physics Exchanges (GMEPE/PAHCE) 2016*, ISBN: 978-1-5090-2485-8, pp. 95-99, Madrid (España), April 4 - April 11, 2016, DOI: 10.1109/GMEPE-PAHCE.2016.7504632.

18. Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Andrea Crespo, Ainhoa Arroyo, Félix del Campo, Roberto Hornero, "Automated Analysis of Unattended Portable Oximetry by means of Bayesian Neural Networks to Assist in the Diagnosis of Sleep Apnea", *Global Medical Engineering Physics Exchanges (GMEPE/PAHCE) 2016*, ISBN: 978-1-5090-2485-8, pp. 79-82, Madrid (España), April 4 - April 11, 2016, DOI: 10.1109/GMEPE-PAHCE.2016.7504628.

### B.1.3 National conferences

1. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Félix del Campo, David Gozal, Roberto Hornero, "Análisis del flujo aéreo nocturno mediante wavelets para la ayuda en el diagnóstico de la apnea del sueño infantil", *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2020)*, ISBN: 978-84-09-25491-0, pp. 252-255, Valladolid (Spain), November 25 - November 27, 2020
2. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Eduardo Santamaría-Vázquez, Félix del Campo, David Gozal, Roberto Hornero, "Modelo de deep learning basado en la arquitectura Inception para el diagnóstico de la apnea del sueño infantil mediante la señal de oximetría", *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2020)*, ISBN: 978-84-09-25491-0, pp. 340-343, Valladolid (Spain), November 25 - November 27, 2020.
3. Daniel Álvarez, Andrea Crespo, Ana Cerezo-Hernández, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Fernando Moreno, Carmen A. Arroyo, Roberto Hornero, Félix del Campo, "Estudio de la asociación entre el índice de calidad del sueño de Pittsburgh y los índices polisomnográficos de severidad del SAHS", *53º Congreso Nacional de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR 2020)*, pp. 432, Online (Spain), November 12 - November 14, 2020.
4. Adrián Martín-Montero, Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Jorge Jiménez-García, Leila Kheirandish-Gozal, Félix del Campo, David Gozal, Roberto Hornero, "Utilidad de nuevas bandas espectrales en la señal de HRV para ayudar

- en el diagnóstico de la apnea del sueño infantil”, *XXXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2019)*, ISBN: 978-84-09-16707-4, pp. 295-298, Santander (Spain), November 27 - November 29, 2019.
5. Jorge Jiménez-García, Gonzalo C. Gutiérrez-Tobal, María García, Daniel Álvarez, Verónica Barroso-García, **Fernando Vaquerizo-Villar**, Adrián Martín-Montero, Félix del Campo, Leila Kheirandish-Goza, David Goza, Roberto Hornero, “Evaluación de la información espectral de las señales de flujo aéreo y saturación de oxígeno en sangre para la ayuda al diagnóstico de la apnea del sueño infantil”, *XXXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2019)*, ISBN: 978-84-09-16707-4, pp. 25-28, Santander (Spain), November 27 - November 29, 2019.
  6. Daniel Álvarez, Ana Cerezo-Hernández, Andrea Crespo, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Fernando Moreno, Ainhoa Arroyo, Tomás Ruíz-Albi, Roberto Hornero, Félix del Campo, “Estimación automática del índice de apnea-hipopnea mediante los registros de oximetría y flujo aéreo realizados en el domicilio”, *52º Congreso Nacional de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR 2019)*, pp. 432, Santiago de Compostela (Spain), June 13 - June 16, 2019.
  7. Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, Andrea Crespo, Fernando Moreno, **Fernando Vaquerizo-Villar**, Ana Cerezo-Hernández, Verónica Barroso-García, Félix del Campo, Roberto Hornero, “Screening automático de la apnea del sueño a partir de la pulsioximetría domiciliaria”, *52º Congreso Nacional de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR 2019)*, pp. 456-457, Santiago de Compostela (Spain), June 13 - June 16, 2019.
  8. Milko D. Terranova-Ríos, Ainhoa Arroyo, Daniel Álvarez, Ana Cerezo-Hernández, Andrea Crespo, Fernando Moreno, Julio de Frutos, Tomás Ruíz-Albi, Ana Andrés, Ana Gómez, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Roberto Hornero, Félix del Campo, “Estudio comparativo entre la polisomnografía hospitalaria y domiciliaria en pacientes con sospecha moderada-alta de síndrome de apnea-hipopnea del sueño. Resultados finales”, *52º Congreso Nacional de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR 2019)*, pp. 433, Santiago de Compostela (Spain), June 13 - June 16, 2019.

9. Ana Cerezo-Hernández, Daniel Álvarez, Ana Sánchez, Ana Andrés, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Roberto Hornero, Tomás Ruíz-Albi, Félix del Campo, "Estudio del efecto combinado de la enfermedad pulmonar obstructiva crónica y del síndrome de apnea obstructiva del sueño moderado-a-severo en la modulación nocturna de la función autonómica cardiaca", *XXXVIII Congreso de la Sociedad Castellano-Leonesa y Cantabria de Patología Respiratoria (SOCALPAR 2019)*, Zamora (Spain), May 10 - May 11, 2019.
10. Carolina A. Urbina, Daniel Álvarez, Ana Cerezo-Hernández, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Fernando Moreno, Ainhoa Arroyo, Julio de Frutos, Roberto Hornero, Félix del Campo, "Diseño y evaluación de un test simplificado de detección de SAHS basado en el análisis automático de las señales de saturación de oxígeno y flujo aéreo adquiridas en el domicilio", *XXXVIII Congreso de la Sociedad Castellano-Leonesa y Cantabria de Patología Respiratoria (SOCALPAR 2019)*, Zamora (Spain), May 10 - May 11, 2019.
11. Daniel Álvarez, Ana Cerezo-Hernández, Andrea Crespo, Gonzalo C. Gutierrez-Tobal, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Carmen A. Arroyo, Tomás Ruíz-Albi, Roberto Hornero, Félix del Campo, "Comparación de tests de screening automático de estimación del índice de apnea-hipopnea basados en oximetría y flujo aéreo no supervisados frente a la polisomnografía domiciliaria", *XXVII Reunión Anual de la Sociedad Española del Sueño (SES 2019)*, Vitoria (Spain), April 11 - April 13, 2019.
12. Adrián Martín-Montero, Gonzalo C. Gutiérrez-Tobal, Jesús Poza, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Saúl J. Ruíz Gómez, Leila Kheirandish-Gozal, Félix del Campo, David Gozal, Roberto Hornero, "Caracterización de la apnea del sueño infantil mediante nuevas bandas espectrales del EEG", *XXXVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2018)*, ISBN: 978-84-09-06253-9, pp. 249-252, Ciudad Real (Spain), November 21 - November 23, 2018.
13. Daniel Álvarez, Ainara Garde, Andrea Crespo, **Fernando Vaquerizo-Villar**, Gonzalo C. Gutiérrez-Tobal, Ana Cerezo-Hernández, J. Mark Ansermino, Guy A. Dumont, Roberto Hornero, Félix del Campo, "Caracterización del registro portátil de oximetría nocturna mediante técnicas de dinámica simbólica en niños con sospecha de apnea del sueño", *51º Congreso Nacional de*

- la Sociedad Española de Neumología y Cirugía Torácica (SEPAR 2018)*, pp. 386, Palma de Mallorca (Spain), May 31 - June 3, 2018.
14. Daniel Álvarez, Ainara Garde, Andrea Crespo, **Fernando Vaquerizo-Villar**, Gonzalo C. Gutiérrez-Tobal, Ana Cerezo-Hernández, J. Mark Ansermino, Guy A. Dumont, Roberto Hornero, Félix del Campo, "Utilidad de la dinámica simbólica para el análisis del registro portátil de oximetría mediante un smartphone en la ayuda al diagnóstico de la apnea del sueño infantil", *XXVI Reunión Anual de la Sociedad Española del Sueño (SES 2018)*, Barcelona (Spain), April 26 - April 28, 2018.
  15. Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, Verónica Barroso-García, **Fernando Vaquerizo-Villar**, Adrián Martín-Montero, Andrea Crespo, Félix del Campo, Roberto Hornero, "Aplicación de la entropía espectral a la señal de variabilidad de pulso para incrementar el potencial de la oximetría en el diagnóstico de la apnea del sueño a domicilio", *XXVI Reunión Anual de la Sociedad Española del Sueño (SES 2018)*, Barcelona (Spain), April 26 - April 28, 2018.
  16. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Utilidad de los patrones binarios locales aplicados a la señal de oximetría en la ayuda al diagnóstico del síndrome de la apnea-hipopnea del sueño en niños", *XXVI Reunión Anual de la Sociedad Española del Sueño (SES 2018)*, Barcelona (Spain), April 26 - April 28, 2018.
  17. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Roberto Romero-Oraá, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Análisis de diferencias de segundo orden aplicado a la señal de flujo aéreo mono-canal para la ayuda al diagnóstico del síndrome de la apnea-hipopnea del sueño en niños", *XXXV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2017)*, ISBN: 978-84-9082-797-0, pp. 481-484, Bilbao (Spain), November 29 - December 1, 2017.
  18. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Verónica Barroso-García, Roberto Romero-Oraá, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Análisis de fluctuaciones sin tendencias (DFA) en los registros de oximetría

- para la ayuda en el diagnóstico del síndrome de la apnea-hipopnea del sueño infantil”, *XXXV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2017)*, ISBN: 978-84-9082-797-0, pp. 209-212, Bilbao (Spain), November 29 - December 1, 2017.
19. Daniel Álvarez, **Fernando Vaquerizo-Villar**, Andrea Crespo, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Ana Cerezo-Hernández, Graciela López-Muñiz, Leila Kheirandish-Goza, David Goza, Roberto Hornero, Félix del Campo, “Transformada wavelet de la señal de oximetría nocturna y variables antropométricas en la ayuda al diagnóstico automático de la apnea del sueño infantil”, *XXV Reunión Anual de la Sociedad Española del Sueño (SES 2017)*, Santander (Spain), April 20 - April 22, 2017.
  20. Gonzalo C. Gutiérrez-Tobal, Julio de Frutos, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Andrea Crespo, Félix del Campo, Roberto Hornero, “Estimación de la severidad de la apnea del sueño mediante redes neuronales bayesianas entrenadas con información espectral del flujo aéreo de sonda de presión y termistor”, *XXV Reunión Anual de la Sociedad Española del Sueño (SES 2017)*, Santander (Spain), April 20 - April 22, 2017.
  21. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Goza, Daniel Álvarez, **Fernando Vaquerizo-Villar**, Andrea Crespo, Félix del Campo, David Goza, Roberto Hornero, “Análisis espectral de la señal de flujo aéreo como ayuda al diagnóstico del síndrome de apnea-hipopnea del sueño en niños”, *XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2016)*, ISBN: 978-84-9048-531-6, pp. 228-231, Valencia (Spain), November 23 - November 25, 2016.
  22. **Fernando Vaquerizo-Villar**, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Leila Kheirandish-Goza, Andrea Crespo, Félix del Campo, David Goza, Roberto Hornero, “Análisis de la señal de oximetría mediante la densidad espectral de potencia y bispectrum en la ayuda al diagnóstico de la apnea infantil”, *XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2016)*, ISBN: 978-84-9048-531-6, pp. 202-205, Valencia (Spain), November 23 - November 25, 2016.
  23. Andrea Crespo, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, **Fernando Vaquerizo-Villar**, Leila Kheirandish-Goza, Roberto Hornero, David Goza,

Félix del Campo, “Análisis automático mediante regresión logística de la señal de oximetría nocturna en niños con sospecha del síndrome de apnea-hipopnea del sueño”, *XXIV Reunión Anual de la Sociedad Española del Sueño (SES 2016)*, Valladolid (Spain), March 31 - April 2, 2016.

24. Gonzalo C. Gutiérrez-Tobal, Daniel Álvarez, Andrea Crespo, **Fernando Vaquerizo-Villar**, Verónica Barroso-García, Leila Kheirandish-Goza, David Gozal, Félix del Campo, Roberto Hornero, “Modelos de máquinas de vector soporte aplicados sobre la oximetría nocturna para la detección automática de niños con síndrome de la apnea-hipopnea del sueño severo”, *XXIV Reunión Anual de la Sociedad Española del Sueño (SES 2016)*, Valladolid (Spain), March 31 - April 2, 2016.

## B.2 International internship

Three-month research internship at the *Charité UniversitätsMedizin Berlin*, Germany.

### i. Purpose of the internship

The main purpose of the research stay was to deepen into the application of deep-learning techniques to physiological signals in the pediatric OSA context. To achieve that objective, the developed specific objectives were proposed: (1) obtaining of a database of pediatric OSA patients with annotations of the time location of sleep stages and apneic events; (2) a state-of-the-art revision of deep-learning algorithms; (3) application of a CNN architecture to detect apneic events from oximetry; (4) application of a CNN architecture to the oximetry signal to estimate the AHI in pediatric OSA patients; (5) application of a CNN architecture to PPG signals to detect sleep stages; (6) statistical analysis of the results; and (7) presentation and publication of the main findings in conferences and JCR journals, respectively. Apart from the main internship activity, the collaboration between members from the *Charité UniversitätsMedizin Berlin*, the University of Valladolid, and the Cardiovascular Physics research group from the Humboldt University has been promoted with the joint involvement in different investigations. This ongoing collaboration has hitherto led to the publication of 2 JCR article and 1 international conference paper by members from these institutions.

### ii. Quality indicators of the institutions

With more than 300 years of history, *Charité UniversitätsMedizin Berlin* is

a multicenter hospital complex, international reference, where 3700 scientists perform their research activity. Among its former members, there have been 11 researchers who have obtained a Nobel prize. This stay was held at the International Center of Sleep Medicine (ICSM) of the *Charité UniversitätsMedizin Berlin*, an expert research center on the area of signal processing in the context of sleep disorders. During the stay, there has also been collaboration with the Cardiovascular Physics research group in the Humboldt University, Berlin. In this respect, *Charité UniversitätsMedizin Berlin* and Humboldt University of Berlin were placed between positions 51 and 75 in the Ranking of Shanghai in 2020. The scientific chair of the ICSM, Professor Thomas Penzel, who was also the supervisor of the internship, is currently the principal investigator of the international research project entitled “Discovery of fundamental mechanisms of sleep for breakthrough technologies of neurorehabilitation medicine”. He has also participated in important European Union (EU) funded projects ENN, ANNDEE, SIESTA, in the FP6 projects BIOSIM, SENSATION, ENN-ICS and DAPHNET and the FP7 project HIVE. His extensive curriculum also includes more than 250 JCR articles, book chapters, and books in the field of sleep medicine, reaching a Hirsch index of 51. He has also received important recognition: the “Bial Award” for clinical medicine in Portugal, 2001; the “Bill Gruen Award” for Innovations in Sleep Research by the Sleep Research Society in 2008; the “Innovations in Biomedical technology Award” by the German Ministry for Education and Research, 2008; the “Somnus Award” for excellent service in sleep medicine by *Schlafmagazin* in Germany, 2012; and the “Distinguished development Award” by the Chinese sleep research society in 2014. To date, he has also supervised more than 19 Doctoral Thesis.

### B.3 Grants

- 07/2019: **‘Ayudas por la Asistencia a Cursos, Congresos y Jornadas Relevantes para el Desarrollo de Tesis Doctorales’ grant** (Berlin, Germany, 25/07/2019-27/07/2019), funded by the Universidad de Valladolid.
- 06/2019: **‘Ayudas a la movilidad para estancias breves y traslados temporales’ grant** (Berlin, Germany, 01/05/2019-31/05/2019), funded by the Ministerio de Educación y Formación Profesional from the Spanish Government.

- 02/2019: **Erasmus + scholarship to perform a three month internship** (Berlin, Germany, 01/05/2019-31/05/2019), funded by the Universidad de Valladolid with funds from the EU.
- 04/2018: **Erasmus + scholarship to assist to a two weeks English course** (Portsmouth, United Kingdom, 01/04/2018-15/04/2018), funded by the Universidad de Valladolid with funds from the EU.
- 09/2017: **‘Ayuda para contratos predoctorales para la Formación de Profesorado Universitario (FPU)’ grant** (FPU16/02938), funded by the Ministerio de Educación, Cultura y Deporte from the Spanish Government.
- 07/2017: **‘Ayuda para financiar la contratación predoctoral de personal investigador’ grant**, funded by the Consejería de Educación de la Junta de Castilla y León from Spain and the European Social Fund.
- 04/2017: **Erasmus + scholarship to assist to a two weeks English course** (Bristol, United Kingdom, 08/04/2017-22/04/2017), funded by the Universidad de Valladolid with funds from the EU.
- 11/2014: **Grant for the collaboration in research tasks in the Department of Signal Theory and Communications from the University of Valladolid** (Bristol, United Kingdom, 19/11/2014-31/07/2015), funded by the Consejo Social de la Universidad de Valladolid.

## B.4 Awards and honors

- 09/2019: **SOCALPAR DUE 2019 Award**, for the project entitled “Diseño y validación de nuevos modelos de oximetría basados en arquitecturas de deep learning para la clasificación automática de eventos respiratorios en pacientes con sospecha de síndrome de apnea-hipopnea del sueño”, conducted by Félix del Campo-Matías, Julio F. De Frutos-Arribas, Tomás Ruíz-Albi, C. Ainhoa Arroyo-Domingo, Ana Cerezo-Hernández, Daniel Álvarez González, Fernando Moreno-Torrero, Roberto Hornero-Sánchez, Gonzalo C. Gutiérrez-Tobal, and **Fernando Vaquerizo-Villar**.
- 06/2018: **“Selección Plata” Award in the Respiratory Sleep Disorders field at the SEPAR 2018**, for the conference paper entitled “Caracterización del registro portátil de oximetría nocturna mediante técnicas de dinámica simbólica en niños con sospecha de apnea del sueño”, conducted

by Daniel Álvarez-González, Ainara Garde-Martínez, Andrea Crespo-Sedano, **Fernando Vaquerizo-Villar**, Ana Cerezo-Hernández, J. Mark Ansermino, Guy A Dumont, Roberto Hornero Sánchez, and Félix del Campo-Matías.

- 04/2017: **Prize for the “Taller de preparación de propuestas INNOvadoras para participar en el proyecto europeo INNOLABS”**, for the project entitled “Diagnóstico y estimación de la severidad del Síndrome de Apnea-Hipopnea del Sueño mediante procesado automático de señales oximétricas”, conducted by Verónica Barroso-García, **Fernando Vaquerizo-Villar**, Gonzalo C. Gutiérrez-Tobal, and Roberto Hornero.
- 04/2017: **Award for the second best conference paper in the “XXV Reunión anual de la Sociedad Española del Sueño ”**, for the conference paper entitled “Transformada wavelet de la señal de oximetría nocturna y variables antropométricas en la ayuda al diagnóstico de la apnea del sueño infantil”, conducted by Daniel Álvarez-González, **Fernando Vaquerizo-Villar**, Andrea Crespo, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, Ana Cerezo, Graciela López, Leila Kheirandish-Gozal, David Gozal, Roberto Hornero, and Félix del Campo.
- 05/2016: **SOCALPAR 2016 Award**, for the project entitled “Utilidad de una red neuronal basada en características demográficas y de oximetría nocturna como método ayuda al diagnóstico del síndrome de apnea-hipopnea obstructiva del sueño en niños”, conducted by Félix del Campo, Daniel Álvarez, Andrea Crespo, Tania Álvaro, Gonzalo C. Gutiérrez-Tobal, Ainhoa Arroyo, Julio De Frutos, Tomás Ruíz, Verónica Barroso, **Fernando Vaquerizo-Villar**, David Gozal, and Roberto Hornero.
- 12/2014: **Prize for Special Achievement in the Bachelor Degree in Specific Telecommunications Engineering**, due to the obtaining of the highest marks of his class.



# Apéndice C

## Resumen en castellano

### C.1 Introducción

La apnea obstructiva del sueño (AOS) es un trastorno respiratorio de elevada prevalencia (1 % - 5 %) en la población infantil (Marcus et al., 2012). La AOS infantil se caracteriza por la recurrencia de episodios de ausencia total (apneas) o parcial (hipopneas) de respiración durante el sueño de los niños, que derivan en una arquitectura del sueño fragmentada y no reparadora (Marcus et al., 2012). En caso de no ser tratada, la AOS infantil puede tener importantes consecuencias adversas que afecten a los sistemas cardiovascular, neurocognitivo, metabólico y al comportamiento, resultando por tanto en una disminución de la salud y calidad de vida de los niños afectados (Capdevila et al., 2008).

La prueba de referencia para el diagnóstico de la AOS en niños es la polisomnografía (PSG) nocturna (Marcus et al., 2012), que requiere la presencia de los niños durante una noche completa en una unidad del sueño pediátrica, donde se monitorizan hasta 32 señales biomédicas. Estas señales son posteriormente analizadas por especialistas del sueño para obtener el índice de apnea-hipopnea (IAH) por hora de sueño, que es la variable polisomnográfica utilizada para el diagnóstico de la AOS (Tan et al., 2014). A pesar de su efectividad, la PSG es una prueba compleja, debido a la necesidad de registrar múltiples señales biomédicas (Tan et al., 2015). Además, los niños tienen que dormir fuera del entorno habitual con múltiples sensores colocados por el cuerpo, lo cual es especialmente incómodo e intrusivo (Katz et al., 2012). Finalmente, esta prueba presenta disponibilidad limitada, lo cual retrasa el acceso al diagnóstico y tratamiento de los pacientes (Tan et al., 2015).

Como consecuencia de estas limitaciones, en los últimos años se han desarrollado pruebas diagnósticas más sencillas como alternativa a la PSG (Kaditis et al., 2016b; Marcus et al., 2012). En este contexto, un enfoque habitual consiste en el análisis automático de la señal de saturación de oxígeno en sangre ( $SpO_2$ ) procedente de la oximetría nocturna, debido a su fiabilidad, sencillez e idoneidad para los niños (del Campo et al., 2018). La señal de  $SpO_2$  proporciona una medida indirecta de la cantidad de oxígeno en la sangre, conteniendo por tanto información sobre las desaturaciones de oxígeno asociadas a los eventos de apnea e hipopnea propios de la AOS infantil (Berry et al., 2012).

Múltiples estudios han demostrado la utilidad del análisis automático de la señal de  $SpO_2$  en la ayuda al diagnóstico de la AOS en adultos (del Campo et al., 2018). En el contexto de la AOS infantil, los estudios precedentes han analizado la señal de  $SpO_2$  mediante una metodología de *feature engineering* similar a la empleada en adultos, basada en la extracción de características, selección de variables y reconocimiento de patrones (del Campo et al., 2018). Sin embargo, estos estudios obtuvieron un rendimiento diagnóstico menor al obtenido en la población adulta (del Campo et al., 2018). Además, se observó una alta redundancia en las características extraídas de la señal de oximetría (Hornero et al., 2017), destacando la necesidad de buscar nuevos algoritmos de procesamiento de señal que proporcionen información adicional sobre la señal de  $SpO_2$  para las particularidades de la AOS infantil.

En esta Tesis Doctoral, se presenta un compendio formado por cuatro artículos publicados en revistas científicas indexadas en el *Journal Citation Reports* (JCR) entre los años 2018 y 2021. Las tres primeras publicaciones se centran en la aplicación de tres métodos de extracción de características novedosos en el contexto de la AOS infantil, que permitan obtener información adicional de la señal de oximetría. En la primera publicación, se aplicó el *bispectrum* (Vaquerizo-Villar et al., 2018b), un método de análisis en el dominio de la frecuencia que, a diferencia del análisis espectral convencional, permite identificar relaciones de fase y desviaciones de la linealidad y gaussianidad de la señal de oximetría (Chua et al., 2010). Siguiendo la caracterización de la señal de oximetría en el dominio de la frecuencia, en la segunda publicación se ha utilizado la transformada *wavelet* (Vaquerizo-Villar et al., 2018c), que proporciona una resolución óptima tiempo-frecuencia y además es adecuada para analizar las propiedades no estacionarias de la señal de  $SpO_2$  asociadas a los eventos apneicos (Rioul and Vetterli, 1991). El tercer artículo (Vaquerizo-Villar et al., 2018a) se ha centrado en mejorar la caracterización de las propiedades no lineales y no estacionarias de la señal de  $SpO_2$  median-

te el análisis de fluctuaciones sin tendencias (*detrended fluctuation analysis*, DFA) (Vaquerizo-Villar et al., 2018a), un método de análisis no lineal que permite detectar cambios en las propiedades de correlación de la señal de SpO<sub>2</sub> a lo largo de las escalas temporales (Chen et al., 2002; Hua and Yu, 2017). Finalmente, en la cuarta publicación de la Tesis Doctoral (Vaquerizo-Villar et al., 2021) se emplea una metodología de *deep learning* que, a diferencia de la metodología de *feature engineering* utilizada en las tres primeras publicaciones, permite identificar de manera automática nuevas características que no son conocidas a priori por los expertos diseñadores del estudio (LeCun et al., 2015). Específicamente, en Vaquerizo-Villar et al. (2021) se evalúa la utilidad de una red neuronal convolucional (*convolutional neural networks*, CNN), una técnica de *deep learning*, para extraer de manera automática toda la información relevante de la señal de SpO<sub>2</sub> asociada a eventos apneicos.

## C.2 Hipótesis y objetivos

La señal de SpO<sub>2</sub> permite detectar las desaturaciones de oxígeno asociadas a eventos apneicos, lo que, junto con la facilidad de adquisición de esta señal, ha llevado a su uso como herramienta de simplificación de diagnóstico de la AOS infantil (del Campo et al., 2018). En este sentido, la primera de las hipótesis en las que se sustenta la investigación desarrollada es que *la señal de oximetría por sí misma contiene información suficiente para la simplificación del diagnóstico de la AOS infantil*. Sin embargo, estudios previos han encontrado una gran redundancia en las características convencionales que se suelen extraer de la señal de SpO<sub>2</sub> (del Campo et al., 2018; Hornero et al., 2017). Por lo tanto, se hipotetiza que *métodos novedosos de extracción de características podrían mejorar la caracterización de los cambios producidos en la señal de SpO<sub>2</sub> relacionados con los recurrentes eventos de apnea e hipopnea típicos de la AOS infantil*. A pesar de su utilidad, los métodos de *feature engineering* tienen una capacidad limitada a la hora de obtener toda la información de los datos debido a dos grandes limitaciones (LeCun et al., 2015): (i) requieren un gran conocimiento del campo bajo estudio para diseñar el proceso de extracción de características; (ii) tienen un nivel de abstracción bajo que no les permite detectar patrones de alta complejidad en los datos. En este sentido, los algoritmos de *deep learning* son capaces de aprender patrones complejos de manera automática a partir de la señal en crudo, por lo que se hipotetiza que *los algoritmos de deep learning podrían extraer de la señal de SpO<sub>2</sub> toda la información relacionada con la AOS infantil*. En base a estas suposiciones, se fundamenta la hipótesis global de esta Te-

sis Doctoral: "La aplicación de técnicas novedosas de extracción de características y de *deep learning* permite encontrar patrones ocultos en las desaturaciones asociados a eventos apneicos, mejorando así la capacidad diagnóstica de la señal de  $SpO_2$  en el contexto de la AOS infantil."

Definidas las hipótesis, el objetivo general de la Tesis es *diseñar, desarrollar y evaluar nuevos modelos automáticos de ayuda a la toma de decisiones clínicas en el contexto de la AOS infantil basados en el procesado avanzado de la señal de oximetría nocturna*. Para llevar a cabo este objetivo, se proponen los siguientes objetivos específicos:

- I. Mejorar la caracterización de los cambios en la señal de  $SpO_2$ , tanto en el dominio del tiempo como en el dominio de la frecuencia, relacionados con los eventos apneicos típicos de la AOS infantil.
- II. Encontrar nuevas características de la señal de oximetría capaces de proporcionar información relevante y complementaria a los parámetros convencionales.
- III. Diseñar y optimizar modelos de reconocimiento de patrones de alto rendimiento que abordan la detección automática de la AOS infantil y su severidad a partir de subconjuntos óptimos de características de la señal de  $SpO_2$ .
- IV. Investigar y desarrollar arquitecturas novedosas de *deep learning* capaces de aprender de manera automática toda la información relevante de la señal de oximetría en el contexto de la AOS infantil.

### C.3 Sujetos y señales

Durante la Tesis Doctoral se han analizado tres bases de datos diferentes: (i) la base de datos pública *Childhood Adenotonsillectomy Trial* (CHAT), compuesta por 1638 estudios del sueño; (ii) la base de datos procedente de la Universidad de Chicago (*University of Chicago, UofC*), compuesta por 981 sujetos pediátricos; y (iii) la base de datos del Hospital Universitario de Burgos (*Burgos University Hospital, BUH*), compuesta por 578 sujetos. Todas ellas contienen registros de  $SpO_2$  de sujetos pediátricos de 0 a 18 años de edad que fueron derivados a la PSG nocturna por sospecha de AOS.

Los registros de  $SpO_2$  fueron adquiridos durante la PSG utilizando frecuencias de muestreo de 1 a 512 Hz. Se siguieron las normas de la Academia Ame-

ricana de Medicina del Sueño (AASM) (Berry et al., 2012; Iber et al., 2007) para obtener el IAH de cada PSG nocturna, que determina el número de eventos de apnea e hipopnea por hora de sueño (e/h). En base al IAH, se han utilizado los puntos de corte habituales (1 e/h, 5 e/h y 10 e/h) para determinar los grupos de severidad manejados en la práctica clínica: no AOS (IAH < 1 e/h), AOS leve ( $1 \leq \text{IAH} < 5$  e/h), AOS moderado ( $5 \leq \text{IAH} < 10$  e/h) y AOS severo ( $\text{IAH} \geq 10$  e/h) (Alonso-Álvarez et al., 2011; Church, 2012; Tan et al., 2014). Además, el umbral de 5 e/h, utilizado habitualmente para recomendar el tratamiento quirúrgico en aquellos niños con un mayor riesgo de desarrollar comorbilidades (Tan et al., 2014), ha sido empleado para establecer la presencia o ausencia de la AOS infantil desde un enfoque binario: AOS negativo ( $\text{AHI} < 5$  e/h) y AOS positivo ( $\text{AHI} \geq 5$  e/h). Las siguientes tablas incluyen las principales variables clínicas y sociodemográficas, incluyendo edad, porcentaje de sujetos masculinos, e índice de masa corporal (IMC) de las tres bases de datos: CHAT (Tabla C.1), UofC (Tabla C.2) y BUH (Tabla C.3).

## C.4 Métodos

Las señales de  $\text{SpO}_2$  se analizaron utilizando metodologías de *feature engineering* (Vaquerizo-Villar et al., 2018a,b,c) y *deep learning* (Vaquerizo-Villar et al., 2021). En ambas metodologías, se comenzó con una etapa de preprocesado de señal, que se encargó de adaptar la señal de oximetría a los requisitos de los diferentes algoritmos de procesado.

Tras el preprocesado, la metodología de *feature engineering* se desarrolló en

**Tabla C.1.** Datos clínicos y sociodemográficos de la base de datos CHAT.

|                                 | Todos                | no AOS               | AOS leve             | AOS moderado         | AOS severo           |
|---------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Registros de $\text{SpO}_2$ (%) | 1638<br>(100 %)      | 637<br>(38.9 %)      | 609<br>(37.2 %)      | 205<br>(12.5 %)      | 187<br>(11.4 %)      |
| Edad (años)                     | 7<br>[6, 8]          |
| Hombres (%)                     | 602<br>(47.4 %)      | 297<br>(46.6 %)      | 287<br>(47.1 %)      | 101<br>(49.3 %)      | 92<br>(49.2 %)       |
| IMC ( $\text{kg}/\text{m}^2$ )  | 17.3<br>[15.8, 21.7] | 17.0<br>[15.5, 19.6] | 17.4<br>[15.6, 21.7] | 18.6<br>[15.4, 23.3] | 18.9<br>[16.0, 24.3] |
| IAH (e/h)                       | 1.6<br>[0.6, 4.7]    | 0.4<br>[0.2, 0.7]    | 2.2<br>[1.5, 3.2]    | 7.1<br>[5.9, 8.4]    | 17.9<br>[12.8, 26.9] |

Los datos se presentan como mediana [rango intercuartil],  $n$  o  $n$  (%). AOS = apnea obstructiva del sueño.

**Tabla C.2.** Datos clínicos y sociodemográficos de la base de datos UofC.

|                                   | Todos                | no AOS               | AOS leve             | AOS moderado         | AOS severo           |
|-----------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Registros de SpO <sub>2</sub> (%) | 981<br>(100%)        | 175<br>(17.8%)       | 401<br>(40.9%)       | 176<br>(17.9%)       | 229<br>(23.4%)       |
| Edad (años)                       | 6<br>[3, 9]          | 7<br>[4, 10]         | 6<br>[4, 9]          | 5<br>[2, 8]          | 4<br>[2, 8]          |
| Hombres (%)                       | 602<br>(61.4%)       | 109<br>(62.3%)       | 247<br>(61.6%)       | 107<br>(60.8%)       | 139<br>(60.7%)       |
| IMC (kg/m <sup>2</sup> )          | 17.9<br>[15.8, 21.9] | 17.7<br>[15.5, 20.9] | 17.7<br>[15.9, 21.2] | 18.6<br>[16.2, 24.0] | 18.3<br>[16.0, 23.2] |
| IAH (e/h)                         | 3.8<br>[1.5, 9.3]    | 0.5<br>[0.1, 0.8]    | 2.5<br>[1.7, 3.5]    | 6.8<br>[5.8, 8.3]    | 19.1<br>[13.9, 31.1] |

Los datos se presentan como mediana [rango intercuartil],  $n$  o  $n$  (%). AOS = apnea obstructiva del sueño.

**Tabla C.3.** Datos clínicos y sociodemográficos de la base de datos BUH.

|                                   | Todos                | no AOS               | AOS leve             | AOS moderado         | AOS severo           |
|-----------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Registros de SpO <sub>2</sub> (%) | 578<br>(100%)        | 205<br>(35.5%)       | 220<br>(38.1%)       | 65<br>(11.3%)        | 88<br>(15.2%)        |
| Edad (años)                       | 5<br>[4, 7]          | 6<br>[4, 8]          | 5<br>[3, 6]          | 5<br>[3, 6]          | 4<br>[3, 5]          |
| Hombres (%)                       | 356<br>(61.6%)       | 127<br>(62.0%)       | 129<br>(58.7%)       | 38<br>(58.5%)        | 62<br>(70.5%)        |
| IMC (kg/m <sup>2</sup> )          | 16.0<br>[14.6, 18.2] | 16.1<br>[14.5, 18.8] | 16.0<br>[14.7, 17.7] | 15.4<br>[14.6, 18.1] | 16.1<br>[14.7, 17.3] |
| IAH (e/h)                         | 1.8<br>[0.6, 5.3]    | 0.4<br>[0.0, 0.6]    | 2.1<br>[1.5, 3.4]    | 6.9<br>[5.8, 8.1]    | 24.3<br>[14.8, 34.9] |

Los datos se presentan como mediana [rango intercuartil],  $n$  o  $n$  (%). AOS = apnea obstructiva del sueño.

tres etapas fundamentales (Vaquerizo-Villar et al., 2018a,b,c): (1) extracción de características, (2) selección de características y (3) reconocimiento de patrones. En primer lugar, se llevó a cabo una fase de extracción de características para obtener un conjunto de parámetros que reflejen la información contenida en las señales de SpO<sub>2</sub> sobre los eventos apneicos de la AOS. Para ello, se emplearon métodos de distinta naturaleza: parámetros estadísticos, índices de oximetría clásicos, análisis en el dominio de la frecuencia y análisis no lineal. Como se ha comentado anteriormente, en esta Tesis Doctoral se ha evaluado de manera particular la utilidad de tres métodos novedosos para proporcionar información adicional y complementaria a los métodos tradicionales relacionada con la AOS infantil y su severidad: (i) el *bispectrum* (Vaquerizo-Villar et al., 2018b), técnica de análisis en el dominio de la frecuencia que permite detectar relaciones de fase y desviaciones de la linealidad y la gaussianidad en la señal de SpO<sub>2</sub> (Chua et al., 2010); (ii) la transformada *wavelet* (Vaquerizo-Villar et al., 2018c), que permite analizar

la señal de SpO<sub>2</sub> con una gran resolución espectral a frecuencias bajas y una alta resolución temporal en el rango de altas frecuencias (Rioul and Vetterli, 1991); (iii) el DFA (Vaquerizo-Villar et al., 2018a), método no lineal que permite detectar cambios en las propiedades de correlación de la señal de SpO<sub>2</sub> relacionados con la escala temporal causados por *spikes* aleatorios y/o segmentos con distinto comportamiento local (Chen et al., 2002; Hua and Yu, 2017).

Tras la extracción de características, se llevó a cabo una etapa de selección de variables para obtener subconjuntos óptimos de características relacionadas con la AOS infantil (Vaquerizo-Villar et al., 2018a,b,c). Para ello, se ha empleado el algoritmo *fast correlation-based filter* (FCBF), que permite obtener un subconjunto de características relevantes y no redundantes. A partir de estos subconjuntos óptimos, en la etapa de reconocimiento de patrones se aplicaron algoritmos de *machine learning* para obtener un diagnóstico automático acerca de la presencia y severidad de la AOS infantil. Concretamente, se han aplicado tres métodos diferentes (Vaquerizo-Villar et al., 2018a,b,c): regresión logística (*logistic regression*, LR), máquinas vector soporte (*support vector machines*, SVM), y red neuronal perceptrón multicapa (*multiLayer perceptron*, MLP).

Al contrario que los enfoques de *feature engineering*, que requieren determinar las características a extraer de los datos, los métodos de *deep learning* son capaces de aprender de manera automática las particularidades inherentes de los mismos mediante el uso de arquitecturas multicapa con múltiples niveles de representación (LeCun et al., 2015). En esta Tesis Doctoral se ha aplicado una arquitectura de *deep learning* basada en CNNs (Vaquerizo-Villar et al., 2021). Las CNNs tienen una arquitectura multicapa, con pesos compartidos, conexiones dispersas y operaciones de reducción de dimensionalidad (*pooling*), que permite identificar, a partir de la señal en crudo, características complejas presentes en distintas partes de la señal (LeCun et al., 2015). Concretamente, se ha utilizado una arquitectura CNN formada por una serie de bloques convolucionales, cada uno de ellos compuesto por: (i) una capa convolucional; (ii) normalización (*batch normalization*); (iii) función de activación; (iv) *pooling*; y (v) regularización (*dropout*). Esta arquitectura se ha entrenado para estimar el número de eventos apneicos en cada segmento de 20 minutos de la señal de SpO<sub>2</sub>. A partir de la salida de la CNN para segmento, el IAH de cada sujeto se ha calculado mediante un procedimiento de agregación basado en regresión lineal, obteniendo así el diagnóstico automático de la AOS infantil y su severidad (Vaquerizo-Villar et al., 2021).

Finalmente, se han empleado las siguientes técnicas de análisis estadístico para interpretar y evaluar los resultados obtenidos con las metodologías de proce-

sado de señal desarrolladas en esta Tesis Doctoral: (i) test de hipótesis estadística; (ii) métricas de rendimiento diagnóstico; (iii) medidas de concordancia; y (iv) estrategias de validación.

## C.5 Resultados y discusión

La tabla C.4 muestra el rendimiento diagnóstico global obtenido con las metodologías de *feature engineering* (Vaquerizo-Villar et al., 2018a,b,c) y *deep learning* (Vaquerizo-Villar et al., 2021) desarrolladas en el compendio de publicaciones. Se puede observar cómo se ha alcanzado un alto rendimiento diagnóstico con los distintos enfoques metodológicos, con precisiones en un rango de 77.6 %-82.7 %, 81.3 %-97.4 % y 85.3 %-97.8 % para los umbrales del IAH de 1, 5 y 10 e/h, respectivamente.

Con respecto a las metodologías de *feature engineering*, se ha mejorado la caracterización de los cambios inducidos por la AOS infantil en la señal de SpO<sub>2</sub> en el dominio del tiempo y en el dominio de la frecuencia utilizando el *bispectrum*, la transformada *wavelet* y el DFA (Vaquerizo-Villar et al., 2018a,b,c). El *bispectrum* ha permitido detectar relaciones de fase y desviaciones de la linealidad y la gaussianidad de la señal de SpO<sub>2</sub> que proporcionan información adicional y complementaria a los enfoques convencionales (Vaquerizo-Villar et al., 2018b). Concretamente, el modelo MLP entrenado en Vaquerizo-Villar et al. (2018b) diseñado con un subconjunto óptimo que incluía variables extraídas del *bispectrum* ha superado la capacidad diagnóstica de un modelo MLP entrenado sin características del *bispectrum*. Además, la gran resolución proporcionada por la transformada *wavelet* discreta (*discrete wavelet transform*, DWT) a bajas frecuencias, así como su adecuación para señales no estacionarias, ha demostrado ser útil para analizar los cambios producidos en la señal de SpO<sub>2</sub> por los eventos apneicos. En concreto, el modelo SVM obtenido en Vaquerizo-Villar et al. (2018c) ha demostrado ser útil como herramienta de *screening* para confirmar la presencia de AOS moderado-a-severo en niños. Finalmente, la aplicación del DFA ha permitido detectar cambios en las propiedades de la señal de SpO<sub>2</sub> a lo largo de las escalas temporales relacionadas con el grado de severidad de la AOS infantil. El modelo de regresión MLP obtenido en Vaquerizo-Villar et al. (2018a) ha permitido superar la precisión de los índices de oximetría clásicos, mejorando la capacidad diagnóstica de la oximetría en el contexto de la AOS infantil.

Por otro lado, la metodología de *deep learning* propuesta ha demostrado ser una herramienta útil para aprender de manera automática características de la

**Tabla C.4.** Resumen del rendimiento diagnóstico de los métodos desarrollados en el compendio de publicaciones.

| Estudio                          | Base de datos | N (Total/test) | IAH          | Métodos (Características/ Clasificación)   | S                    | E                    | P                    | kappa | CCI   |
|----------------------------------|---------------|----------------|--------------|--|----------------------|----------------------|----------------------|-------|-------|
| (Vaquerizo-Villar et al., 2018b) | UofC          | 298/75         | 5<br>10      | <i>Bispectrum</i> , características de la PSD, ODI3 y variables antropométricas / MLP multiclase | 61.8<br>60.0         | 97.6<br>94.5         | 81.3<br>85.3         | 0.56  | -     |
| (Vaquerizo-Villar et al., 2018c) | UofC          | 981/392        | 5            | Transformada wavelet, ODI3, momentos estadísticos y características de la PSD/ SVM binario       | 71.9                 | 91.1                 | 84.0                 | -     | -     |
| (Vaquerizo-Villar et al., 2018b) | UofC          | 981/392        | 1<br>5<br>10 | DFA y ODI3 / MLP de regresión  | 97.1<br>78.8<br>77.1 | 23.3<br>83.7<br>94.8 | 82.7<br>81.9<br>91.1 | 0.41  | 0.891 |
| (Vaquerizo-Villar et al., 2021)  | UofC          | 3196/392       | 1<br>5<br>10 | Arquitectura CNN   | 90.8<br>76.0         | 36.4<br>88.1<br>95.8 | 80.1<br>83.9<br>92.3 | 0.42  | 0.917 |
| (Vaquerizo-Villar et al., 2021)  | CHAT          | 3196/312       | 1<br>5<br>10 | Arquitectura CNN   | 71.2<br>83.7<br>83.9 | 81.8<br>100<br>99.3  | 77.6<br>97.4<br>97.8 | 0.52  | 0.960 |
| (Vaquerizo-Villar et al., 2021)  | BUH           | 3196/231       | 1<br>5<br>10 | Arquitectura CNN   | 88.8<br>61.1<br>65.0 | 53.2<br>93.7<br>96.9 | 79.2<br>83.5<br>91.3 | 0.42  | 0.583 |

N = número de sujetos, IAH= índice de apnea-hipopnea, S = sensibilidad (%), Sp = especificidad (%), P = precisión (%), kappa=kappa de Cohen, CCI = coeficiente de correlación intra-clase, ODI3= índice de desaturación de oxígeno del 3%, PSD = densidad espectral de potencia, DFA = análisis de fluctuaciones sin tendencias, MLP = perceptrón multicapa, SVM = máquina vector soporte, CNN = red neuronal convolucional, UofC = Universidad de Chicago, CHAT = *Childhood Adenotonsillectomy Trial*, BUH = Hospital universitario de Burgos.

señal de oximetría asociadas a los eventos apneicos (Vaquerizo-Villar et al., 2021). La arquitectura de *deep learning* basada en CNN ha obtenido un alto rendimiento diagnóstico en tres bases independientes (UofC, CHAT, y BUH), superando al índice de desaturación de oxígeno del 3% (ODI3) y a un modelo MLP de *feature engineering*. Además, la aplicación de este modelo de *deep learning* en un protocolo clínico de *screening* permite reducir la necesidad de realizar la PSG completa en el 45.9% (BUH), 50.0% (UofC) y 73.7% (CHAT) de los sujetos pediátricos, contribuyendo así a la reducción de listas de espera y costes médicos asociados al

diagnóstico de la AOS en niños. No obstante, se observó un rendimiento diagnóstico inferior en las bases de datos UofC y BUH que en la base de datos de CHAT, lo cual puede deberse a diferencias en la distribución del IAH, la edad de los sujetos, la frecuencia de muestreo de los registros de SpO<sub>2</sub> y los grupos de población de las distintas bases de datos (Vaquerizo-Villar et al., 2021). Esto concuerda con estudios previos que también obtuvieron diferencias en el rendimiento diagnóstico entre bases de datos de sueño con distintas características clínicas (Alvarez et al., 2013; Biswal et al., 2018; Nikkonen et al., 2019).

Al comparar los resultados obtenidos en los distintos artículos de esta Tesis, se observa como el conjunto de test de la base de datos UofC empleado en Vaquerizo-Villar et al. (2018c), Vaquerizo-Villar et al. (2018a) y Vaquerizo-Villar et al. (2021) está compuesto por los mismos sujetos. En este sentido, destaca el mayor rendimiento diagnóstico global obtenido con la metodología de *deep learning* en comparación con las metodologías de *feature engineering* utilizadas en Vaquerizo-Villar et al. (2018c) y Vaquerizo-Villar et al. (2018a). Esto resalta la capacidad de generalización de los métodos de *deep learning* y su idoneidad para identificar de manera automática los cambios relacionados con la AOS infantil en la señal de SpO<sub>2</sub>. Sin embargo, como limitación cabe destacar que es más difícil explicar e interpretar las características extraídas por la arquitectura de *deep learning*.

En la tabla C.5 se muestran de forma resumida los resultados obtenidos en estudios previos centrados en la simplificación del diagnóstico de la AOS infantil mediante el análisis automático de la señal de oximetría basado en técnicas de procesado de señal (Álvarez et al., 2017; Álvarez et al., 2018; Crespo et al., 2017, 2018; Garde et al., 2014a; Hornero et al., 2017; Xu et al., 2019). Como puede observarse, las metodologías propuestas de *feature engineering* y *deep learning* alcanzaron un elevado rendimiento diagnóstico en comparación con los estudios del estado del arte, especialmente para los puntos de corte del IAH de 5 y 10 e/h. Esto sugiere que los métodos novedosos de procesado de señal desarrollados en esta Tesis Doctoral permiten extraer información adicional acerca de los cambios en la señal de SpO<sub>2</sub> asociados a la AOS infantil.

**Tabla C.5.** Resumen del estado del arte de los estudios basados en enfoques de procesamiento automático de la señal de oximetría.

| Estudio                | N<br>(Total/test) | IAH | Métodos<br>(Características/<br>Clasificación)   | S    | E    | P    |
|------------------------|-------------------|-----|--|------|------|------|
| (Garde et al., 2014a)  | 146/146           | 5   | Índices de oximetría, momentos estadísticos y características de la PSD y no lineales/ LDA binario               | 80.0 | 83.9 | 78.5 |
| (Álvarez et al., 2017) | 50/50             | 1   | Índices de oximetría, momentos estadísticos y características de la PSD y no lineales /LR binario                | 89.6 | 71.5 | 85.5 |
|                        |                   | 3   |  | 82.9 | 84.4 | 83.4 |
|                        |                   | 5   |  | 82.2 | 83.6 | 82.8 |
| (Crespo et al., 2017)  | 146/146           | 3   | Índices de oximetría y características no lineales / LR binario  | 84.5 | 83.0 | 83.5 |
| (Hornero et al., 2017) | 4191/3602         | 1   | ODI3, momentos estadísticos y características de la PSD y no lineales / MLP de regresión                         | 84.0 | 53.2 | 75.2 |
|                        |                   | 5   |  | 68.2 | 87.2 | 81.7 |
|                        |                   | 10  |  | 68.7 | 94.1 | 90.2 |
| (Crespo et al., 2018)  | 176/176           | 1   | Índices de oximetría, momentos estadísticos y características de la PSD y no lineales /LR                        | 93.9 | 37.8 | 84.3 |
|                        |                   | 3   |  | 74.6 | 81.7 | 77.7 |
|                        |                   | 5   |  | 70.0 | 91.4 | 82.7 |
| (Xu et al., 2019)      | 432/432           | 1   | ODI3, momentos estadísticos y características de la PSD y no lineales / MLP de regresión                         | 95.3 | 19.1 | 79.6 |
|                        |                   | 5   |  | 77.8 | 80.5 | 79.4 |
|                        |                   | 10  |  | 73.5 | 92.7 | 88.2 |
| (Álvarez et al., 2018) | 142/142           | 5   | Índices de oximetría, características de la PSD, parámetros no lineales y variables antropométricas / LR binario | 73.5 | 89.5 | 83.3 |

LDA = análisis discriminante lineal, LR = regresión logística, MLP= perceptrón multicapa, IAH= índice de apnea-hipopnea, S = sensibilidad (%), Sp = especificidad (%), P = precisión (%), ODI3= índice de desaturación de oxígeno del 3%, PSD = densidad espectral de potencia.

## C.6 Conclusiones

A raíz del análisis de los resultados obtenidos en esta Tesis Doctoral se han obtenido las siguientes conclusiones:

- 1) Los modelos propuestos de *feature engineering* y *deep learning* han superado la capacidad diagnóstica de las características comúnmente extraídas de la señal de oximetría, así como a los resultados obtenidos en estudios previos. Por tanto, el uso de nuevas técnicas de procesamiento automático es útil para incrementar la capacidad diagnóstica de la señal de SpO<sub>2</sub>.

- 2) El *bispectrum* puede ser empleado como herramienta complementaria a los métodos tradicionales a la hora de caracterizar los cambios producidos en la señal de SpO<sub>2</sub> relacionados con la AOS infantil. Concretamente, los cambios en la amplitud del *bispectrum* relacionados con desviaciones de la gaussianidad y los cambios en la fase del *bispectrum* relacionados con el acoplamiento en fase entre componentes espectrales de la señal de oximetría, proporcionan información adicional a las variables antropométricas, al ODI3 y a las características extraídas de la densidad espectral de potencia.
- 3) La transformada *wavelet* es una herramienta adecuada para analizar las propiedades no estacionarias, así como las componentes de baja frecuencia, de la señal de SpO<sub>2</sub> relacionadas con la AOS infantil. Específicamente, la concentración en torno a cero de los coeficientes de la DWT en la banda 0.0244-0.0488 Hz, la energía de los coeficientes de la DWT en la banda 0.0244-0.0488 Hz y los cambios en la distribución de energía de la DWT de la señal de oximetría, proporcionan información complementaria a los métodos convencionales.
- 4) El DFA es una herramienta apropiada para caracterizar los cambios producidos en las propiedades de la señal de oximetría lo largo de las escalas temporales relacionadas con la AOS infantil y su severidad. Los resultados obtenidos sugieren que la pendiente en las primeras escalas temporales del perfil del DFA contiene información adicional que permite mejorar la caracterización de los cambios inducidos en la señal de SpO<sub>2</sub> por los episodios apnéicos típicos de la AOS infantil.
- 5) De todos los métodos de reconocimiento de patrones, el modelo SVM diseñado con el ODI3, momentos estadísticos y características extraídas de la densidad espectral de potencia y la transformada *wavelet*, ha proporcionado evidencias sólidas para establecer la presencia de AOS moderado-a-severo (AHI  $\geq 5$  e/h) en sujetos pediátricos, logrando una precisión del 84.0 % y una razón de verosimilitud positiva de 14.6. Este modelo podría ser utilizado como herramienta de screening para niños con AOS moderado-a-severo.
- 6) El modelo de *deep learning* basado en CNN alcanzó un rendimiento diagnóstico superior al obtenido con los enfoques de *feature engineering* en el contexto de la AOS infantil. Concretamente, este modelo alcanzó precisiones por encima del 80 % para el diagnóstico de AOS moderado-a-severo (97.4 %, 83.9 % y 83.5 %) y superiores al 90 % (97.8 %, 92.3 % y 91.3 %) para

la detección de AOS severo en los conjuntos de test de CHAT, UofC y BUH, respectivamente. Esto concuerda con la mayor capacidad predictiva mostrada en los últimos años por los algoritmos de *deep learning* en numerosos ámbitos. Nuestros hallazgos sugieren que la metodología de *deep learning* podría cambiar el paradigma de procesamiento de datos biomédicos en relación con la AOS infantil.

- 7) El modelo de *deep learning* mostró una gran capacidad de generalización, aunque el rendimiento diagnóstico difirió entre las bases de datos, lo cual puede estar influenciado por diferencias en la frecuencia de muestreo, rangos de edad, distribución del IAH y grupos de población de las tres bases de datos. Por tanto, los parámetros clínicos y sociodemográficos deberían ser considerados a la hora de validar nuestra propuesta en la práctica clínica.
- 8) Los protocolos diagnósticos derivados de nuestra propuesta ponen de manifiesto la aplicabilidad clínica de la señal de oximetría en el *screening* de la AOS infantil. En concreto, el protocolo de *screening* propuesto contribuiría a reducir el número de PSGs en un 45 %-70 % (73.7 %, 50.0 %, y 45.9 %) de los sujetos pediátricos en las bases de datos de CHAT, UofC y BUH. De este modo, los niños afectados se beneficiarían de un test diagnóstico más accesible y de menor intrusividad basado en el análisis automático de la señal de SpO<sub>2</sub>.



# Bibliography

- Alonso-Álvarez, M. L., Canet, T., Cubell-Alarco, M., Estivill, E., Fernández, E., Gozal, D., Jurado-Luque, M. J., Lluch-Roselló, M. A., Martínez-Pérez, F., Merino-Andren, M., Pin-Arboledas, G., Roure, N., Sanmartí, F. X., Sans-Capdevila, O., Segarra-Isern, J., 2011. Documento de consenso del síndrome de apneas-hipopneas durante el sueño en niños. *Arch Bronconeumol* 47 (Supl 5), 2–18.
- Álvarez, D., Alonso-Álvarez, M. L., Gutiérrez-Tobal, G. C., Crespo, A., Kheirandish-Gozal, L., Gozal, D., Terán-Santos, J., Campo, F. D., 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 (5), 7–11.
- Álvarez, D., Crespo, A., Vaquerizo-Villar, F., Gutiérrez-Tobal, G. C., Cerezo-Hernández, A., Barroso-García, V., Ansermino, J. M., Dumont, G. A., Hornero, R., Del Campo, F., Garde, A., 2018. Symbolic dynamics to enhance diagnostic ability of portable oximetry from the phone oximeter in the detection of paediatric sleep apnoea. *Physiological measurement* 39 (10), 104002.
- Alvarez, D., Hornero, R., Marcos, J. V., Wessel, N., Penzel, T., Glos, M., Del Campo, F., 2013. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis. *International journal of neural systems* 23 (05), 1350020.
- Atri, R., Mohebbi, M., 2015. Obstructive sleep apnea detection using spectrum and bispectrum analysis of single-lead ecg signal. *Physiological measurement* 36 (9), 1963.
- Azarbarzin, A., Sands, S. A., Stone, K. L., Taranto-Montemurro, L., Messineo, L., Terrill, P. I., Ancoli-Israel, S., Ensrud, K., Purcell, S., White, D. P., et al., 2019. The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the osteoporotic fractures in men study and the sleep heart health study. *European heart journal* 40 (14), 1149–1157.
- Barroso-García, V., Gutiérrez-Tobal, G. C., Kheirandish-Gozal, L., Álvarez, D., Vaquerizo-Villar, F., del Campo, F., Gozal, D., Hornero, R., 2020. Usefulness of recurrence plots from airflow recordings to aid in paediatric sleep apnoea diagnosis. *Computer Methods and Programs in Biomedicine* 183, 105083.
- Barroso-García, V., Gutiérrez-Tobal, G. C., Kheirandish-Gozal, L., Vaquerizo-Villar, F., Álvarez, D., del Campo, F., Gozal, D., Hornero, R., 2021. Bispectral analysis of overnight airflow to improve the pediatric sleep apnea diagnosis. *Computers in Biology and Medicine* 129, 104167.
- Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. In: 25th annual conference on neural information processing systems (NIPS 2011). pp. 1–9.

- Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., Marcus, C. L., Mehra, R., Parthasarathy, S., Quan, S. F., Redline, S., Strohl, K. P., Davidson Ward, S. L., Tangredi, M. M., 2012. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. *Journal of clinical sleep medicine* 8 (5), 597.
- Bertoni, D., Isaiah, A., 2019. Towards patient-centered diagnosis of pediatric obstructive sleep apnea—a review of biomedical engineering strategies. *Expert review of medical devices* 16 (7), 617–629.
- Bishop, C. M., 2006. *Pattern Recognition and Machine Learning*. Springer.
- Biswal, S., Sun, H., Goparaju, B., Brandon Westover, M., Sun, J., Bianchi, M. T., 2018. Expert-level sleep scoring with deep neural networks. *Journal of the American Medical Informatics Association* 25 (12), 1643–1650.
- Blanco-Velasco, M., Cruz-Roldan, F., Godino-Llorente, J. I., Barner, K. E., 2010. Nonlinear Trend Estimation of the Ventricular Repolarization Segment for T-Wave Alternans Detection. *IEEE transactions on bio-medical engineering* 57 (10), 2402–2412.
- Bronzino, J. D., 2000. *Biomedical Engineering Handbook 2*. Vol. 2. Springer Science & Business Media.
- Brouillette, R. T., Morielli, A., Leimanis, A., Waters, K. A., Luciano, R., Ducharme, F. M., 2000. Nocturnal pulse oximetry as an abbreviated testing modality for pediatric obstructive sleep apnea. *Pediatrics* 105 (2), 405–412.
- Capdevila, O. S., Kheirandish-Gozal, L., Dayyat, E., Gozal, D., 2008. Pediatric obstructive sleep apnea: Complications, management, and long-term outcomes. *Proceedings of the American Thoracic Society* 5 (2), 274–282.
- Chan, E. D., Chan, M. M., Chan, M. M., 2013. Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations. *Respiratory medicine* 107 (6), 789–799.
- Chang, L., Wu, J., Cao, L., 2013. Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea. *International Journal of Pediatric Otorhinolaryngology* 77 (3), 365–371.
- Chen, C.-C., Barnhart, H. X., 2008. Comparison of icc and ccc for assessing agreement for data without and with replications. *Computational statistics & data analysis* 53 (2), 554–564.
- Chen, Z., Ivanov, P. C., Hu, K., Stanley, H. E., 2002. Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 65 (4), 15.
- Choi, S. H., Yoon, H., Kim, H. S., Kim, H. B., Kwon, H. B., Oh, S. M., Lee, Y. J., Park, K. S., 2018. Real-time apnea-hypopnea event detection during sleep by convolutional neural networks. *Computers in Biology and Medicine* 100 (February), 123–131.
- Chua, K., Chandran, V., Acharya, U. R., Min, C., 2010. Application of higher order statistics/spectra in biomedical signals—A review. *Medical Engineering and Physics* 32 (7), 679–689.

- Chua, K. C., Chandran, V., Acharya, U., Lim, C., 2008. Cardiac state diagnosis using higher order spectra of heart rate variability. *Journal of medical engineering & technology* 32 (2), 145–155.
- Church, G. D., 2012. The Role of Polysomnography in Diagnosing and Treating Obstructive Sleep Apnea in Pediatric Patients. *Current problems in pediatric and adolescent health care* 42 (1), 22–25.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20 (1), 37–46.
- Cohen, M. E., Hudson, D. L., 2000. New chaotic methods for biomedical signal analysis. In: *IEEE EMBS International Conference on Information Technology Applications in Biomedicine*. pp. 123–128.
- Cohen, M. E., Hudson, D. L., Deedwania, P. C., 1996. Applying continuous chaotic modeling to cardiac signal analysis. *IEEE Engineering in Medicine and Biology Magazine* 15 (5), 97–102.
- Collop, N. A., 2002. Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep medicine* 3 (1), 43–47.
- Costa, M., Goldberger, A. L., Peng, C., 2005. Multiscale entropy analysis of biological signals. *Physical review E* 71 (2), 1–18.
- Crespo, A., Álvarez, D., Gutiérrez-Tobal, G. C., Vaquerizo-Villar, F., Barroso-García, V., Alonso-Álvarez, M. L., Terán-Santos, J., Hornero, R., del Campo, F., 2017. Multiscale Entropy Analysis of Unattended Oximetric Recordings to Assist in the Screening of Paediatric Sleep Apnoea at Home. *Entropy* 19 (6), 284.
- Crespo, A., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Cerezo-Hernández, A., Gozal, D., Hornero, R., Del Campo, F., 2018. Assessment of oximetry-based statistical classifiers as simplified screening tools in the management of childhood obstructive sleep apnea. *Sleep and Breathing* 22 (4), 1063–1073.
- Davis, K. F., Parker, K. P., Montgomery, G. L., 2004. Sleep in infants and young children: Part one: normal sleep. *Journal of Pediatric Health Care* 18 (2), 65–71.
- Daw, C. S., Finney, C. E. A., Tracy, E. R., 2003. A review of symbolic analysis of experimental data. *Review of Scientific instruments* 74 (2), 915–930.
- Dehkordi, P., Garde, A., Karlen, W., Petersen, C. L., Wensley, D., Dumont, G. A., Ansermino, J. M., 2016. Evaluation of cardiac modulation in children in response to apnea/hypopnea using the phone oximeter™. *Physiological measurement* 37 (2), 187.
- Dehlink, E., Tan, H.-L., 2016. Update on paediatric obstructive sleep apnoea. *Journal of thoracic disease* 8 (2), 224.
- del Campo, F., Crespo, A., Cerezo-Hernández, A., Gutiérrez-Tobal, G. C., Hornero, R., Álvarez, D., 2018. Oximetry use in obstructive sleep apnea. *Expert Review of Respiratory Medicine* 12 (8), 665–681.
- Deviaene, M., Testelmans, D., Buyse, B., Borzé, P., Van Huffel, S., Varon, C., 2018. Automatic screening of sleep apnea patients based on the spo 2 signal. *IEEE journal of biomedical and health informatics* 23 (2), 607–617.

- Doane, D. P., 1976. Aesthetic frequency classifications. *The American Statistician* 30 (4), 181–183.
- Ebrahimi, Z., Loni, M., Daneshtalab, M., Gharehbaghi, A., 2020. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X 7*, 100033.
- Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., Acharya, U. R., 2018. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine* 161, 1–13.
- Faust, O., Razaghi, H., Barika, R., Ciaccio, E. J., Acharya, U. R., 2019. A review of automated sleep stage scoring based on physiological signals for the new millennia. *Computer methods and programs in biomedicine* 176, 81–91.
- Flemons, W. W., Littner, M. R., 2003. Measuring agreement between diagnostic devices. *Chest* 124 (4), 1535–1542.
- Fontenla-Romero, O., Guijarro-Berdinas, B., Alonso-Betanzos, A., Moret-Bonillo, V., 2005. A new method for sleep apnea classification using wavelets and feedforward neural networks. *Artificial Intelligence in Medicine* 34 (1), 65–76.
- Garde, A., Dehkordi, P., Karlen, W., Wensley, D., Ansermino, J. M., Dumont, G. A., 2014a. Development of a screening tool for sleep disordered breathing in children using the phone oximeter™. *PloS one* 9 (11), e112959.
- Garde, A., Karlen, W., Dehkordi, P., Ansermino, J. M., Dumont, G. A., 2014b. Oxygen Saturation Resolution Influences Regularity Measurements. In: *36th Annual International Conference of the IEEE In Engineering in Medicine and Biology Society (EMBC)*. pp. 2257–2260.
- Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. *Deep learning*. Vol. 1. MIT press Cambridge.
- Gutiérrez-Tobal, G. C., Álvarez, D., Kheirandish-Gozal, L., Del Campo, F., Gozal, D., Hornero, R., 2021. Reliability of machine learning to diagnose pediatric obstructive sleep apnea: Systematic review and meta-analysis. *Pediatric Pulmonology*.
- Guyon, I., 2003. An Introduction to Variable and Feature Selection. *The Journal of Machine Learning Research* 3, 1157–1182.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*. pp. 1026–1034.
- Hornero, R., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Philby, M. F., Alonso-Álvarez, M. L., Álvarez, D., Dayyat, E. A., Xu, Z., Huang, Y.-S., Tamae Kakazu, M., Li, A. M., Van Eyck, A., Brockmann, P. E., Ehsan, Z., Simakajornboon, N., Kaditis, A. G., Vaquerizo-Villar, F., Crespo Sedano, A., Sans Capdevila, O., von Lukowicz, M., Terán-Santos, J., Del Campo, F., Poets, C. F., Ferreira, R., Bertran, K., Zhang, Y., Schuen, J., Verhulst, S., Gozal, D., 2017. Nocturnal Oximetry-based Evaluation of Habitually Snoring Children. *American Journal of Respiratory and Critical Care Medicine* 196 (12), 1591–1598.
- Hosmer, D., Lemeshow, S., 2004. *Applied logistic regression*. John Wiley & Sons.

- Hua, C.-C., Yu, C.-C., 2017. Detrended fluctuation analysis of oxyhemoglobin saturation by pulse oximetry in sleep apnea syndrome. *Journal of Medical and Biological Engineering* 37 (6), 791–799.
- Huber, P. J., 1964. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* 35 (1), 73–101.
- Hunter, S. J., Gozal, D., Smith, D. L., Philby, M. F., Kaylegian, J., Kheirandish-Gozal, L., 2016. Effect of sleep-disordered breathing severity on cognitive performance measures in a large community cohort of young school-aged children. *American Journal of Respiratory and Critical Care Medicine* 194 (6), 739–747.
- Iber, C., Ancoli-Israel, S., Chesson, A., Quan, S. F., 2007. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification. *Journal of Clinical Sleep Medicine* 3 (7), 752.
- Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P. A., 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33 (4), 917–963.
- Jahmunah, V., Oh, S. L., Wei, J. K. E., Ciaccio, E. J., Chua, K., San, T. R., Acharya, U. R., 2019. Computer-aided diagnosis of congestive heart failure using ecg signals—a review. *Physica Medica* 62, 95–104.
- Jiménez-García, J., Gutiérrez-Tobal, G. C., García, M., Kheirandish-Gozal, L., Martín-Montero, A., Álvarez, D., del Campo, F., Gozal, D., Hornero, R., 2020. Assessment of Airflow and Oximetry Signals to Detect Pediatric Sleep Apnea-Hypopnea Syndrome Using AdaBoost. *Entropy* 22 (6), 670.
- Jobson, J. D., 2012. *Applied multivariate data analysis: regression and experimental design*. Springer Science & Business Media.
- Kaditis, A., Kheirandish-Gozal, L., Gozal, D., 2016a. Pediatric OSAS : Oximetry can provide answers when polysomnography is not available. *Sleep Medicine Reviews* 27, 96–105.
- Kaditis, A. G., Alvarez, M. L. A., Boudewyns, A., Alexopoulos, E. I., Ersu, R., Joosten, K., Larra-mona, H., Miano, S., Narang, I., Trang, H., Tsaoussoglou, M., Vandebussche, N., Villa, M. P., Waardenburg, D. V., Weber, S., Verhulst, S., 2016b. Obstructive sleep disordered breathing in 2- to 18-year-old children: Diagnosis and management. *European Respiratory Journal* 47 (1), 69–94.
- Kaimakamis, E., Tsara, V., Bratsas, C., Sichletidis, L., Karvounis, C., Maglaveras, N., 2016. Evaluation of a decision support system for obstructive sleep apnea with nonlinear analysis of respiratory signals. *PloS one* 11 (3), e0150163.
- Kainulainen, S., Duce, B., Korkalainen, H., Oksenberg, A., Leino, A., Arnardottir, E. S., Kulkas, A., Myllymaa, S., Töyräs, J., Leppänen, T., 2020. Severe desaturations increase psychomotor vigilance task-based median reaction time and number of lapses in obstructive sleep apnoea patients. *European Respiratory Journal* 55 (4).
- Kainulainen, S., Töyräs, J., Oksenberg, A., Korkalainen, H., Sefa, S., Kulkas, A., Leppänen, T., 2019. Severity of desaturations reflects osa-related daytime sleepiness better than ahi. *Journal of Clinical Sleep Medicine* 15 (8), 1135–1142.

- Kapur, V. K., Auckley, D. H., Chowdhuri, S., Kuhlmann, D. C., Mehra, R., Ramar, K., Harrod, C. G., 2017. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an american academy of sleep medicine clinical practice guideline. *Journal of Clinical Sleep Medicine* 13 (3), 479–504.
- Katz, E. S., Mitchell, R. B., Ambrosio, C. M. D., 2012. Obstructive Sleep Apnea in Infants. *American journal of respiratory and critical care medicine* 185 (8), 805–816.
- Khandoker, A. H., Palaniswami, M., Karmakar, C. K., 2008. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings. *IEEE transactions on information technology in biomedicine* 13 (1), 37–48.
- Kheirandish-Gozal, L., 2010. What is “abnormal” in pediatric sleep? *Respiratory Care* 55 (10), 1366–1376.
- Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kirk, V. G., Bohn, S. G., Flemons, W. W., Remmers, J. E., 2003. Comparison of Home Oximetry Monitoring With Laboratory Polysomnography in Children. *Chest* 124 (5), 1702–1708.
- Krishnan, S., Athavale, Y., 2018. Trends in biomedical signal feature extraction. *Biomedical Signal Processing and Control* 43, 41–63.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *nature* 521 (7553), 436–444.
- Lee, J.-M., Kim, D.-J., Kim, I.-Y., Park, K.-S., Kim, S. I., 2002. Detrended fluctuation analysis of eeg in sleep apnea using mit/bih polysomnography data. *Computers in biology and medicine* 32 (1), 37–47.
- Lee, Y., Bister, M., Blanchfield, P., Salleh, Y., 2004. Automated detection of obstructive apnea and hypopnea events from oxygen saturation signal. In: *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 1. IEEE, pp. 321–324.
- Lempel, A., Ziv, J., 1976. On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory* 22 (1), 75–81.
- Lin, R., Lee, R.-G., Tseng, C.-L., Zhou, H.-K., Chao, C.-F., Jiang, J.-A., 2006. A new approach for identifying sleep apnea syndrome using wavelet transform and neural networks. *Biomedical Engineering: Applications, Basis and Communications* 18 (03), 138–143.
- Linz, D., Colling, S., Nußstein, W., Debl, K., Hohl, M., Fellner, C., Böhm, M., Maier, L. S., Hamer, O. W., Buchner, S., et al., 2018. Nocturnal hypoxemic burden is associated with epicardial fat volume in patients with acute myocardial infarction. *Sleep and Breathing* 22 (3), 703–711.
- Loughlin, G., Brouillette, R., Brooke, L., Carroll, J., Chipps, B., England, S., Ferber, P., Ferraro, N., Gaultier, C., Givan, D., Haddad, G. G., R., M. B., mallory G. B., T., N. I., Rosen, C. L., T., T. B., L., D.-W. S., E., W.-M. D., Wohl, M. E., 1996. Standards and indications for cardiopulmonary sleep studies in children. *American Journal of Respiratory and Critical Care Medicine* 153 (2), 866–878.

- Ma, J.-R., Huang, J.-J., Chen, Q., Wu, H.-T., Xiao, K.-L., Zhang, Y.-T., 2018. Value of pulse oximetry watch for diagnosing pediatric obstructive sleep apnea/hypopnea syndrome. *Acta otolaryngologica* 138 (2), 175–179.
- Magalang, U. J., Dmochowski, J., Veeramachaneni, S., Draw, A., Mador, J., El-Solh, A., Grant, B. J. B., 2003. Prediction of the Apnea-Hypopnea Index From Overnight Pulse Oximetry. *CHEST Journal* 124 (5), 1694–1701.
- Marcus, C. L., Brooks, L. J., Ward, S. D., Draper, K. A., Gozal, D., Halbower, A. C., Jones, J., Lehmann, C., Schechter, M. S., Sheldon, S., Shiffman, R. N., Spruyt, K., 2012. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics* 130 (3), e714–e755.
- Marcus, C. L., Moore, R. H., Rosen, C. L., Giordani, B., Garetz, S. L., Taylor, H. G., Mitchell, R. B., Amin, R., Katz, E. S., Arens, R., Paruthi, S., Muzumdar, H., Gozal, D., Thomas, N. H., Ware, J., Beebe, D., Snyder, K., Elden, L., Sprecher, R. C., Willging, P., Jones, D., Bent, J. P., Hoban, T., Chervin, R. D., Ellenberg, S. S., Redline, S., 2013. A Randomized Trial of Adenotonsillectomy for Childhood Sleep Apnea. *New England Journal of Medicine* 368 (25), 2366–2376.
- Martin, M., Plastino, A., Rosso, O., 2003. Statistical complexity and disequilibrium. *Physics Letters A* 311 (2-3), 126–132.
- Mendez, M., Corthout, J., Van Huffel, S., Matteucci, M., Penzel, T., Cerutti, S., Bianchi, A. M., 2010. Automatic screening of obstructive sleep apnea from the ecg based on empirical mode decomposition and wavelet analysis. *Physiological measurement* 31 (3), 273.
- Mostafa, S. S., Mendonça, F., G Ravelo-García, A., Morgado-Dias, F., 2019. A systematic review of detecting sleep apnea using deep learning. *Sensors* 19 (22), 4934.
- Mostafa, S. S., Mendonça, F., Morgado-Dias, F., Ravelo-García, A., 2017. Spo2 based sleep apnea detection using deep learning. In: 2017 IEEE 21st international conference on intelligent engineering systems (INES). IEEE, pp. 91–96.
- Murat, F., Yildirim, O., Talo, M., Baloglu, U. B., Demir, Y., Acharya, U. R., 2020. Application of deep learning techniques for heartbeats detection using ecg signals-analysis and review. *Computers in biology and medicine*, 103726.
- Najarian, K., Splinter, R., 2012. *Biomedical signal and image processing*. Taylor & Francis.
- Netzer, N., Eliasson, A. H., Netzer, C., Kristo, D. A., 2001. Overnight pulse oximetry for sleep-disordered breathing in adults: a review. *Chest* 120 (2), 625–633.
- Nikkonen, S., Afara, I. O., Leppänen, T., Töyräs, J., 2019. Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea. *Scientific reports* 9 (1), 13200.
- Ning, T., Bronzino, J., 1990. Autoregressive and bispectral analysis techniques: Eeg applications. *IEEE Engineering in Medicine and Biology Magazine* 9 (1), 47–50.
- Nixon, G. M., Kermack, A. S., Davis, G. M., Manoukian, J. J., Brown, A., Brouillette, R. T., 2004. Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry. *Pediatrics* 113 (1), e19–e25.

- Oldenburg, O., Wellmann, B., Buchholz, A., Bitter, T., Fox, H., Thiem, U., Horstkotte, D., Wegscheider, K., 2016. Nocturnal hypoxaemia is associated with increased mortality in stable heart failure patients. *European heart journal* 37 (21), 1695–1703.
- Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., Goldberger, A. L., 1994. Mosaic organization of DNA nucleotides. *Physical review e* 49 (2), 1685.
- Penzel, T., Kantelhardt, J., Grote, L., Peter, J., Bunde, A., 2003. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE Transactions on Biomedical Engineering* 50 (10), 1143–1151.
- Pincus, S. M., 1991. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* 88 (6), 2297–2301.
- Poza, J., Hornero, R., Abásolo, D., Fernández, A., 2007. Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease. *Medical Engineering and Physics* 29 (10), 1073–1083.
- Rangayyan, R. M., 2015. *Biomedical signal analysis*. John Wiley & Sons.
- Redline, S., Amin, R., Beebe, D., Chervin, R. D., Garetz, S. L., Giordani, B., Marcus, C. L., Moore, R. H., Rosen, C. L., Arens, R., Gozal, D., Katz, E. S., Mitchell, R. B., Muzumdar, H., Taylor, H., Thomas, N., Ellenberg, S., 2011. The Childhood Adenotonsillectomy Trial (CHAT): Rationale, Design, and Challenges of a Randomized Controlled Trial Evaluating a Standard Surgical Procedure in a Pediatric Population. *Sleep* 34 (11), 1509–1517.
- Richman, J. S., Moorman, J. R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* 278 (6), H2039–H2049.
- Rioul, O., Vetterli, M., 1991. Wavelets and signal processing. *IEEE signal processing magazine* 8 (4), 14–38.
- Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schürmann, M., Başar, E., 2001. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of neuroscience methods* 105 (1), 65–75.
- Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., Faubert, J., 2019. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering* 16 (5), 051001.
- Snoek, J., Larochelle, H., Adams, R. P., 2012. Practical bayesian optimization of machine learning algorithms. *arXiv preprint arXiv:1206.2944*.
- Sörnmo, L., Laguna, P., 2005. *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press.
- Standards of Practice Committee of the American Sleep Disorders Association, 1994. Practice parameters for the use of portable recording in the assessment of obstructive sleep apnea. *Sleep* 17 (4), 372–377.
- Steyerberg, E. W., Vergouwe, Y., 2014. Towards better clinical prediction models: seven steps for development and an abcd for validation. *European heart journal* 35 (29), 1925–1931.

- Tagluk, M. E., Sezgin, N., 2011. A new approach for estimation of obstructive sleep apnea syndrome. *Expert Systems with Applications* 38 (5), 5346–5351.
- Taha, B. H., Dempsey, J. A., Weber, S. M., Badr, M. S., Skatrud, J. B., Young, T. B., Jacques, A. J., Seow, K., 1997. Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep* 20 (11), 991–1001.
- Tan, H.-L., Alonso Alvarez, M. L., Tsaoussoglou, M., Weber, S., Kaditis, A. G., 2017. When and why to treat the child who snores? *Pediatric pulmonology* 52 (3), 399–412.
- Tan, H.-L., Gozal, D., Ramirez, H. M., Bandla, H. P. R., Kheirandish-Gozal, L., 2014. Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea. *Sleep* 37 (2), 255–260.
- Tan, H. L., Kheirandish-Gozal, L., Gozal, D., 2015. Pediatric home sleep apnea testing slowly getting there! *Chest* 148 (6), 1382–1395.
- Tsai, C.-M., Kang, C.-H., Su, M.-C., Lin, H.-C., Huang, E.-Y., Chen, C.-C., Hung, J.-C., Niu, C.-K., Liao, D.-L., Yu, H.-R., 2013. Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. *International journal of pediatric otorhinolaryngology* 77 (8), 1286–1290.
- Van Eyck, A., Lambrechts, C., Vanheeswijck, L., 2015. The role of nocturnal pulse oximetry in the screening for obstructive sleep apnea in obese children and adolescents. *Sleep Medicine* 16 (11), 1409–1412.
- Van Steenkiste, T., Groenendaal, W., Deschrijver, D., Dhaene, T., 2018. Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks. *IEEE journal of biomedical and health informatics* 23 (6), 2354–2364.
- Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., Del Campo, F., Gozal, D., Hornero, R., 2018a. Detrended fluctuation analysis of the oximetry signal to assist in paediatric sleep apnoea-hypopnoea syndrome diagnosis. *Physiological measurement* 39 (11), 114006.
- Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., del Campo, F., Gozal, D., Hornero, R., 2018b. Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings. *Computer Methods and Programs in Biomedicine* 156, 141–149.
- Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., del Campo, F., Gozal, D., Hornero, R., 2018c. Wavelet analysis of oximetry recordings to assist in the automated detection of moderate-to-severe pediatric sleep apnea-hypopnea syndrome. *PLoS one* 13 (12), e0208502.
- Vaquerizo-Villar, F., Alvarez, D., Kheirandish-Gozal, L., Gutierrez-Tobal, G. C., Barroso-Garcia, V., Santamaria-Vazquez, E., Del Campo, F., Gozal, D., Hornero, R., 2021. A convolutional neural network architecture to enhance oximetry ability to diagnose pediatric obstructive sleep apnea. *IEEE Journal of Biomedical and Health Informatics*.

- Velasco, C. T., Suárez, M. D., Figueroa, J. M., Turienzo, M. D., Len, F. L., Mansilla, E., 2013. Pulse oximetry recording in children with adenotonsillar hypertrophy : usefulness in the diagnosis of obstructive sleep apnea syndrome. *Archivos argentinos de pediatría* 111 (3), 196–201.
- Villa, M. P., Pietropaoli, N., Supino, M. C., Vitelli, O., Rabasco, J., Evangelisti, M., Del Pozzo, M., Kaditis, A. G., 2015. Diagnosis of pediatric obstructive sleep apnea syndrome in settings with limited resources. *JAMA Otolaryngology–Head & Neck Surgery* 141 (11), 990–996.
- Weir, J. P., 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the sem. *The Journal of Strength & Conditioning Research* 19 (1), 231–240.
- Welch, P. D., 1967. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Transactions on audio and electroacoustics* 15 (2), 70–73.
- Witten, I., Frank, E., Hall, M., 2011. *Data Mining*. Morgan Kaufmann.
- Xu, Z., Gutiérrez-Tobal, G. C., Wu, Y., Kheirandish-Gozal, L., Ni, X., Hornero, R., Gozal, D., 2019. Cloud algorithm-driven oximetry-based diagnosis of obstructive sleep apnoea in symptomatic habitually snoring children. *European Respiratory Journal* 53 (2).
- Yu, L., Liu, H., 2004. Efficient Feature Selection via Analysis of Relevance and Redundancy. *Journal of Machine Learning Research* 5, 1205–1224.
- Zhou, S.-M., Gan, J. Q., Sepulveda, F., 2008. Classifying mental tasks based on features of higher-order statistics from eeg signals in brain–computer interface. *Information Sciences* 178 (6), 1629–1640.
- Zweig, M. H., Campbell, G., 1993. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry* 39 (4), 561–577.

# Index

- A**  
additional information 4, 71, 72, 74, 83  
apnea-hypopnea index 11, 21
- B**  
bispectrum 53  
blood oxygen saturation 2, 13, 21
- C**  
clinical characteristics 21, 83  
    age 22–24, 76  
    AHI distribution 22–24, 76  
    population group 76  
    sampling rate 76  
Cohen's kappa 50  
comparison of performance 78  
conclusions 89  
contributions 87  
conventional oximetric indices 14, 80
- D**  
databases 21  
    Burgos university hospital 24, 62,  
    75, 80  
    childhood adenotonsillectomy trial  
    22, 62, 75, 80  
    university of Chicago 23, 53, 56, 60,  
    62, 75, 78  
deep learning 4, 9, 18, 43, 74, 78  
    convolutional neural networks 4, 8,  
    18, 43, 62, 74  
    lack of interpretation 80, 84
- F**  
feature engineering 2, 16, 29, 66, 69, 78,  
81  
    limitations 43  
feature extraction 2, 16, 29  
feature selection 2, 16, 39  
    fast correlation-based filter 39, 55,  
    57, 60  
feature-engineering 53  
frequency domain 16  
frequency domain analysis 29, 32, 34, 36  
    bispectrum 2, 5, 16, 34, 53, 70, 71  
    power spectral density 32, 54, 57,  
    70, 71  
    wavelet analysis 2, 6, 16, 36, 56, 70,  
    71  
frequency resolution 36, 72  
future research lines 90
- G**  
gaussianity 34, 70, 71
- H**  
hypotheses 19

- I**  
intraclass correlation coefficient 51
- L**  
limitations 83  
linearity 34, 70, 71
- N**  
non-stationary 30  
nonlinear analysis 30  
    detrended fluctuation analysis 2, 7,  
    16, 30, 60, 73
- O**  
objectives 20  
obstructive sleep apnea 2  
oximetry 2, 13  
oxygen desaturation index 29, 54, 57, 61,  
66, 70, 71, 74
- P**  
pattern recognition 2, 16, 40  
    logistic regression 40, 59  
    multilayer perceptron 42, 56, 59, 61  
    support vector machines 41, 59  
pediatric obstructive sleep apnea 10, 21  
    diagnosis 11, 12  
polysomnography 2, 11, 21  
    alternatives 12  
    limitations 11
- R**  
redundancy 4, 16, 39
- S**  
scaling behaviour 31, 73  
screening protocol 72, 75  
signal preprocessing 27  
signal processing 2, 9, 16, 27  
state of the art 14, 78, 80  
stationarity 36, 72, 73  
statistical analysis 47  
    agreement 50  
    diagnostic ability 48  
    hypothesis tests 47  
    validation 51, 53, 56, 60, 64, 81, 83  
statistical moments 29, 57, 71
- T**  
thematic consistency 2



Pediatric obstructive sleep apnea (OSA) is a high prevalent disease (1%-5%). It is associated with many negative effects on the overall health and life quality of the affected children when it is untreated, including cardiometabolic malfunctioning and neurobehavioral abnormalities. Overnight polysomnography (PSG) is the gold standard for pediatric OSA diagnosis. Despite its effectiveness, PSG is costly, complex, highly intrusive, and lacks availability. This has prompted the search for simplified screening tests. One of these alternatives tests is the automated analysis of the blood oxygen saturation signal from overnight oximetry due to its easy acquisition and interpretation, as well as its suitability for children. In this context, the present Doctoral Thesis focuses on applying novel signal processing algorithms in order to enhance the diagnostic ability of the oximetry signal in the framework of pediatric OSA. Particularly, three novel feature extraction algorithms (bispectrum, wavelet, and detrended fluctuation analysis), as well as a novel deep-learning architecture based on convolutional neural networks are proposed. In view of the results, we feel that this compendium of publications could contribute to the use of clinical screening tools to diagnose pediatric OSA based on the automated analysis of the oximetry signal, aiming at providing an early and timely diagnosis and treatment of the affected children.

**Doctoral Thesis**

**Compendium of publications**

**International Mention**

**UvA**