



UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

MÁSTER EN INGENIERÍA DE TELECOMUNICACIÓN

**Ayuda al diagnóstico de la apnea obstructiva del
sueño infantil usando técnicas de análisis
automático de señales cardiorrespiratorias basadas
en *Deep Learning***

Autor:

D. Alejandro Martínez de Abajo

Tutores:

Dr. D Roberto Hornero Sánchez

Dr. D Jorge Jiménez García

Valladolid, Junio de 2025

TÍTULO: Ayuda al diagnóstico de la apnea obstructiva del sueño infantil usando técnicas de análisis automático de señales cardiorrespiratorias basadas en *Deep Learning*

AUTOR: D. Alejandro Martínez de Abajo

TUTOR: Dr. D. Roberto Hornero Sánchez
Dr. D. Jorge Jiménez García

DEPARTAMENTO: Teoría de la Señal y Comunicaciones e Ingeniería Telemática

TRIBUNAL

PRESIDENTE: Dra. D^a. María García Gadañón

VOCAL: Dr. D. Luis Miguel San José Revuelta

SECRETARIO: Dra. D^a. Miriam Antón Rodríguez

SUPLENTE 1: Dr. D. Javier Manuel Aguiar Pérez

SUPLENTE 2: Dra. D^a. Beatriz Sainz de Abajo

FECHA: Junio de 2025

CALIFICACIÓN:

Agradecimientos

En especial agradecimiento a los profesores Jorge Jiménez García y Roberto Hornero Sánchez por brindarme la oportunidad de poder realizar y llevar a cabo este proyecto. Su constante orientación, sus valiosos consejos y sus aportaciones tanto en el desarrollo del sistema como en la redacción de este informe han sido fundamentales para el correcto progreso de este trabajo.

Mi más sincero agradecimiento a todos los miembros del grupo de investigación GIB de apnea, Fernando, Gonzalo, Clara, María, Máximo, Adrián, Verónica, Roberto y Jorge por todo su apoyo, su ayuda y sus consejos, el increíble ambiente de trabajo y el fantástico trato recibido.

Finalmente, me gustaría agradecer a mi familia y amigos por apoyarme constantemente y animarme a llevar el proyecto adelante y a mi tutor Jorge y mi amiga Beatriz por ayudarme en todo lo que han podido y más, a todos vosotros, gracias.

Resumen

El sueño constituye un proceso biológico fundamental para el mantenimiento de la salud, al desempeñar un papel clave en funciones cognitivas e inmunológicas. Un descanso adecuado resulta especialmente importante durante la infancia, etapa en la que el cuerpo y el cerebro se encuentran en pleno desarrollo. Alteraciones en el sueño, como los microdespertares o ciertos trastornos, pueden implicar riesgos significativos para la salud. Entre ellos, destaca la apnea obstructiva del sueño, un problema relevante especialmente en edades tempranas.

El diagnóstico de la apnea obstructiva del sueño se realiza habitualmente mediante un estudio del sueño nocturno en laboratorio a través de la polisomnografía, una prueba precisa pero costosa en términos de tiempo y recursos. Por ello, la investigación científica ha centrado esfuerzos en la búsqueda de métodos alternativos más accesibles. Entre estas alternativas, cobran relevancia los enfoques basados en técnicas de inteligencia artificial, en particular el aprendizaje profundo, que permiten estimar de forma precisa el índice de apneas e hipopneas a partir de señales fisiológicas.

Este trabajo se centra en evaluar la viabilidad del uso conjunto de tres señales fisiológicas: saturación de oxígeno en sangre (SpO_2), flujo aéreo (FA) y variabilidad de la frecuencia cardíaca (HRV), extraídas de la base de datos pública “*Childhood Adenotonsillectomy Trial*” (CHAT). Estas señales fueron procesadas mediante un modelo de aprendizaje profundo basado en redes neuronales convolucionales (CNN). Para ello, los registros se segmentaron en intervalos de 20 minutos, introduciendo las señales individualmente, por pares y en conjunto, con el objetivo de estimar el número de eventos respiratorios por registro. Posteriormente, se aplicó una regresión Huber para estimar el índice de apneas e hipopneas y clasificar a los pacientes según la gravedad del trastorno.

Se desarrolló un modelo para cada combinación posible de señales, siendo la combinación de SpO_2 y HRV la que ofreció los mejores resultados. Las métricas obtenidas para esta combinación fueron: sensibilidad del 79.50%, especificidad del 83.58% y exactitud del 80.39% para un umbral de 1 e/h; 82.42%, 98.14% y 93.46% para 5 e/h, respectivamente; y 69.05%, 97.73% y 93.79% para 10 e/h, respectivamente.

Nuestro modelo representa una contribución adicional a los estudios previos sobre la señal SpO_2 y propone un enfoque novedoso para el resto de combinaciones de señales. En el caso de FA, el uso de una CNN con la base de datos CHAT constituye una propuesta original. Para HRV, aunque existen trabajos previos que han empleado CNN, estos se han centrado en la señal ECG, a diferencia de nuestro enfoque basado directamente en HRV. En la combinación SpO_2 -FA se han utilizado variantes de CNN, pero no una CNN unidimensional como en nuestro caso. En cuanto a la combinación SpO_2 -HRV, los estudios previos emplean variantes de HRV, y aquellos que sí usan HRV no aplican metodologías basadas en CNN, por lo que nuestra propuesta también supone una aportación novedosa. Para la combinación FA-HRV se encuentran investigaciones centradas en ECG, pero no directamente en HRV, lo que distingue nuevamente nuestro trabajo. Finalmente, para la combinación SpO_2 -FA-HRV, la literatura existente se basa en comparaciones directas y en el uso de ECG, sin aplicación de CNN ni empleo directo de HRV. Por tanto, puede concluirse que nuestro trabajo introduce, en la mayoría de los casos, un enfoque o metodología nueva que complementa significativamente la literatura científica existente.

Palabras clave:

Inteligencia Artificial, Apnea Obstructiva del Sueño, Pediatría, Aprendizaje profundo.

Abstract

Sleep is a fundamental biological process for maintaining health, playing a key role in cognitive and immunological functions. Adequate rest is especially important during childhood, a stage in which the body and brain are in full development. Sleep disturbances, such as micro-awakenings or certain disorders, can pose significant health risks. Among these, obstructive sleep apnoea stands out as a particularly significant problem in early childhood.

Obstructive sleep apnoea is usually diagnosed by means of a nocturnal sleep study in a laboratory using polysomnography, a test that is accurate but costly in terms of time and resources. For this reason, scientific research has focused its efforts on finding more accessible alternative methods. Among these alternatives, approaches based on artificial intelligence techniques, particularly deep learning, are gaining prominence, as they allow for the accurate estimation of the apnoea-hypopnoea index from physiological signals.

This work focuses on evaluating the feasibility of using three physiological signals together: blood oxygen saturation (SpO₂), airflow (FA) and heart rate variability (HRV), extracted from the public database such as the Childhood Adenotonsillectomy Trial (CHAT). These signals were processed using a deep learning model based on convolutional neural networks (CNN). To do this, the records were segmented into 20-minute intervals, introducing the signals individually, in pairs and together, with the aim of estimating the number of respiratory events per record. Subsequently, a Huber regression was applied to estimate the apnoea-hypopnoea index and classify patients according to the severity of the disorder.

A model was developed for each possible combination of signals, with the combination of SpO₂ and HRV offering the best results. The metrics obtained for this combination were: sensitivity of 79.50%, specificity of 83.58% and accuracy of 80.39% for a threshold of 1 e/h; 82.42%, 98.14% and 93.46% for 5 e/h respectively; and 69.05%, 97.73% and 93.79% for 10 e/h respectively.

Our model represents an additional contribution to previous studies on the SpO₂ signal and proposes a novel approach for the other signal combinations. In the case of AF, the use of a CNN with the CHAT database is an original proposal. For HRV, although there are previous studies that have used CNN, these have focused on the ECG signal, unlike our approach based directly on HRV. In the SpO₂–AF combination, variants of CNN have been used, but not a one-dimensional CNN as in our case. As for the SpO₂–HRV combination, previous studies use HRV variants, and those that do use HRV do not apply CNN-based methodologies, so our proposal also represents a novel contribution. For the FA–HRV combination, there is research focused on ECG, but not directly on HRV, which again distinguishes our work. Finally, for the SpO₂–AF–HRV combination, the existing literature is based on direct comparisons and the use of ECG, without the application of CNN or direct use of HRV. Therefore, it can be concluded that our work introduces, in most cases, a new approach or methodology that significantly complements the existing scientific literature.

Keywords:

Artificial Intelligence, Obstructive Sleep Apnoea, Paediatrics, Deep Learning.

ÍNDICE GENERAL

ÍNDICE GENERAL	XI
Capítulo 1: Introducción	1
1.1. Características de la Apnea Obstructiva del Sueño.....	3
1.2. Diagnóstico de la Apnea Obstructiva del Sueño.....	4
1.3. Tratamientos de la Apnea Obstructiva del Sueño	7
1.4. Apnea Obstructiva del Sueño Pediátrica.....	10
1.5. Deep Learning.....	13
1.6. Hipótesis y objetivos.....	16
1.7. Estructura del Trabajo Fin de Máster.....	16
Capítulo 2: Estado de la Técnica	19
2.1. Contexto del caso de estudio.....	19
2.2. Señales elegidas	20
2.3. Metodología empleada.....	20
2.4. Estudios de estimación de la severidad.....	23
2.5. Aportaciones realizadas	24
Capítulo 3: Señales y Bases de datos.....	31
3.1. Base de datos CHAT.....	31
3.2. Señales empleadas	32
Capítulo 4: Metodología	37
4.1. Preprocesado	37
4.2. Segmentación de señales.....	38
4.3. Generador de datos	40
4.4. Modelo desarrollado	41
4.4.1. Estructura básica	41
4.4.2. Optimización de los hiperparámetros	45
4.4.3. Estimación del IAH.....	49
4.5. Análisis estadístico.....	50
4.6. Especificaciones técnicas.....	53
Capítulo 5: Resultados	56
5.1. Optimización del modelo	56
5.2. Rendimiento diagnóstico del modelo.....	59
5.2.1. Regresión del IAH	59
5.2.2. Resultados de clasificación de 4 clases.....	64
5.2.3. Resultados de clasificación binaria	68
Capítulo 6: Discusión	73
6.1. Introducción	73
6.2. Comparativa de señales.....	73
6.3. Comparativa con estudios previos	76
Capítulo 7: Conclusiones y Líneas futuras	85

7.1.	Contribuciones realizadas	85
7.2.	Conclusiones	85
7.3.	Limitaciones y líneas futuras	86
Bibliografía		89

Índice de figuras

FIGURA 1-1. ZONAS DE COLOCACIÓN DE ELECTRODOS VISTAS LATERAL (IZQUIERDA) Y SUPERIOR (DERECHA): C-CENTRAL, F-FRONTAL, O-OCCIPITAL, A-AURICULAR Y P-PARIETAL (DEPARTAMENTO DE FISIOLÓGIA, FACULTAD DE MEDICINA UNAM).....	5
FIGURA 1-2. DIBUJO ESQUEMÁTICO DE UN TRATAMIENTO CON CPAP (AUTORÍA PROPIA, GENERADA CON IA).....	7
FIGURA 1-3. DIAGRAMA DE FLUJO REPRESENTATIVO DE POSIBLES PROCEDIMIENTOS QUIRÚRGICOS Y TRATAMIENTOS NO INVASIVOS PARA PACIENTES CON AOS (XIA & SAWAN, 2021).....	9
FIGURA 1-4. PUNTUACIÓN BRODSKY GRADOS 0-4. (ADAPTADO DE CLÍNICA OTORRINOLARINGOLÓGICA SANTA CATHARINA, 2020).....	10
FIGURA 1-5. TABLA DE CÁLCULO IMC/EDAD PARA NIÑOS ENTRE 5 Y 19 AÑOS, PROPORCIONADA POR LA OMS (OMS, 2007).....	11
FIGURA 1-6. TABLA DE CÁLCULO IMC/EDAD PARA NIÑAS ENTRE 5 Y 19 AÑOS, PROPORCIONADA POR LA OMS (OMS, 2007).....	11
FIGURA 1-7. DIAGRAMA DE VENN DE LOS CONCEPTOS: INTELIGENCIA ARTIFICIAL, MACHINE LEARNING Y DEEP LEARNING (ADAPTADA DE KELLEHER, 2019).	14
FIGURA 1-8. SIMILITUDES ENTRE UNA NEURONA DEL CEREBRO HUMANO (SUPERIOR) Y UNA NEURONA ARTIFICIAL DE UNA RED NEURONAL (INFERIOR) (AUTORÍA PROPIA, BASADA EN KELLEHER, 2019).....	15
FIGURA 3-1. EJEMPLO DE FRAGMENTO DE SEÑAL SpO_2 FILTRADA (GENERADA CON MATLAB).	33
FIGURA 3-2. EJEMPLO DE FRAGMENTO DE SEÑAL FA FILTRADA (GENERADA CON MATLAB).	34
FIGURA 3-3. SEGMENTO ECG CON COMPLEJO PQRST. (AUTORÍA PROPIA, INSPIRADO DE WAGNER & STRAUSS, 2014).	35
FIGURA 3-4. EJEMPLO DE FRAGMENTO DE SEÑAL HRV FILTRADA (GENERADA CON MATLAB).....	35
FIGURA 4-1. SEÑAL SpO_2 COMPLETA (SUPERIOR) VS SEÑAL SpO_2 FRAGMENTADA (INFERIOR) (AUTORÍA PROPIA, ELABORADA CON MATLAB).....	38
FIGURA 4-2. SEÑAL FA COMPLETA (SUPERIOR) VS SEÑAL FA FRAGMENTADA (INFERIOR) (AUTORÍA PROPIA, ELABORADA CON MATLAB).....	39
FIGURA 4-3. SEÑAL HRV COMPLETA (SUPERIOR) VS SEÑAL HRV FRAGMENTADA (INFERIOR) (AUTORÍA PROPIA, ELABORADA CON MATLAB).....	39
FIGURA 4-4. ESQUEMA DEL PROCESO DE UMBRALIZACIÓN DE LOS EVENTOS (AUTORÍA PROPIA).....	40
FIGURA 4-5. ESQUEMA DE FUNCIONAMIENTO DEL GENERADOR DE DATOS (AUTORÍA PROPIA).	41
FIGURA 4-6. ESQUEMA DE UNA RED NEURONAL MULTICAPA CON CAPAS ETIQUETADAS (ADAPTADA DE KELLEHER, 2019).....	42
FIGURA 4-7. ESQUEMA GENÉRICO DEL MODELO CNN INICIAL (AUTORÍA PROPIA).	43
FIGURA 4-8. ESQUEMA DE UNA ETAPA DE ENTRENAMIENTO DE UN MODELO, INDICANDO LAS FUNCIONES FEEDFORWARD O FORWARD PROPAGATION, LOSS FUNCTION Y BACK PROPAGATION (AUTORÍA PROPIA).	44
FIGURA 4-9. EJEMPLO DE UNA CURVA DE APRENDIZAJE DEL MODELO CNN (AUTORÍA PROPIA).....	48
FIGURA 4-10. EJEMPLO DE GRÁFICA BLAND-ALTMAN (IZQUIERDA) Y GRÁFICA SCATTER (DERECHA) DEL MODELO CNN (AUTORÍA PROPIA).	48
FIGURA 4-11. PRIMER FRAGMENTO DE UN SUJETO, SEÑAL SpO_2 (AUTORÍA PROPIA, ELABORADO CON MATLAB).....	49
FIGURA 4-12. EJEMPLO DE REGRESIÓN HUBER CON LA SEÑAL SpO_2	50
FIGURA 4-13. EJEMPLO DE MATRIZ DE CONFUSIÓN DE VALORES DE IAH DE LA SEÑAL SpO_2 EN TEST TRAS SOMETERLA A LA REGRESIÓN HUBER.	51
FIGURA 4-14. ESQUEMA UMBRALIZACIÓN PROGRESIVA DE LA MATRIZ DE CONFUSIÓN (AUTORÍA PROPIA).	53
FIGURA 5-1. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL NÚMERO DE EVENTOS PARA LA SEÑAL SpO_2 EN EL CONJUNTO DE TEST.	59
FIGURA 5-2. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL NÚMERO DE EVENTOS PARA LA SEÑAL FA EN EL CONJUNTO DE TEST.....	60
FIGURA 5-3. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL NÚMERO DE EVENTOS PARA LA SEÑAL HRV EN EL CONJUNTO DE TEST.	60
FIGURA 5-4. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL NÚMERO DE EVENTOS PARA LAS SEÑALES SpO_2 -FA EN EL CONJUNTO DE TEST.....	60
FIGURA 5-5. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL NÚMERO DE EVENTOS PARA LAS SEÑALES SpO_2 -HRV EN EL CONJUNTO DE TEST.....	61
FIGURA 5-6. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL NÚMERO DE EVENTOS PARA LAS SEÑALES FA-HRV EN EL CONJUNTO DE TEST.	61
FIGURA 5-7. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL NÚMERO DE EVENTOS PARA LAS SEÑALES SpO_2 -FA-HRV EN EL CONJUNTO DE TEST.....	61
FIGURA 5-8. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL IAH PARA LA SEÑAL SpO_2 EN EL CONJUNTO DE TEST.	62
FIGURA 5-9. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL IAH PARA LA SEÑAL FA EN EL CONJUNTO DE TEST.	62

FIGURA 5-10. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL IAH PARA LA SEÑAL HRV EN EL CONJUNTO DE TEST.....	62
FIGURA 5-11. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL IAH PARA LAS SEÑALES SpO ₂ -FA EN EL CONJUNTO DE TEST.....	63
FIGURA 5-12. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL IAH PARA LAS SEÑALES SpO ₂ -HRV EN EL CONJUNTO DE TEST.....	63
FIGURA 5-13. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL IAH PARA LAS SEÑALES FA-HRV EN EL CONJUNTO DE TEST.....	63
FIGURA 5-14. GRÁFICAS BLAND-ALTMAN (IZQUIERDA) Y SCATTER PLOT (DERECHA) DE LA ESTIMACIÓN DEL IAH PARA LAS SEÑALES SpO ₂ -FA-HRV EN EL CONJUNTO DE TEST.	64
FIGURA 5-15. MATRIZ DE CONFUSIÓN DEL VALOR DE IAH POSTERIOR A LA REGRESIÓN EN EL CONJUNTO DE TEST DE LA SEÑAL SpO ₂	64
FIGURA 5-16. MATRIZ DE CONFUSIÓN DEL VALOR DE IAH POSTERIOR A LA REGRESIÓN EN EL CONJUNTO DE TEST DE LA SEÑAL FA.....	65
FIGURA 5-17. MATRIZ DE CONFUSIÓN DEL VALOR DE IAH POSTERIOR A LA REGRESIÓN EN EL CONJUNTO DE TEST DE LA SEÑAL HRV.	65
FIGURA 5-18. MATRIZ DE CONFUSIÓN DEL VALOR DE IAH POSTERIOR A LA REGRESIÓN EN EL CONJUNTO DE TEST DE LAS SEÑALES SpO ₂ -FA.....	66
FIGURA 5-19. MATRIZ DE CONFUSIÓN DEL VALOR DE IAH POSTERIOR A LA REGRESIÓN EN EL CONJUNTO DE TEST DE LAS SEÑALES SpO ₂ -HRV.....	66
FIGURA 5-20. MATRIZ DE CONFUSIÓN DEL VALOR DE IAH POSTERIOR A LA REGRESIÓN EN EL CONJUNTO DE TEST DE LAS SEÑALES FA-HRV.....	67
FIGURA 5-21. MATRIZ DE CONFUSIÓN DEL VALOR DE IAH POSTERIOR A LA REGRESIÓN EN EL CONJUNTO DE TEST DE LAS SEÑALES SpO ₂ -FA-HRV.	67

Índice de tablas

TABLA 1-1. FACTORES DE RIESGO QUE FAVORECEN LA POSIBILIDAD DE PADECER AOS (BITNERS & ARENS, 2020A).	2
TABLA 2-1. TABLA RESUMEN REVISIÓN BIBLIOGRÁFICA CON LA SEÑAL SpO_2 (PRIMERA PARTE) EN COLOR AZUL OSCURO.	25
TABLA 2-2. TABLA RESUMEN REVISIÓN BIBLIOGRÁFICA CON LA SEÑAL SpO_2 EN COLOR AZUL OSCURO (SEGUNDA PARTE) Y LA SEÑAL FA EN COLOR NARANJA.	26
TABLA 2-3. TABLA RESUMEN REVISIÓN BIBLIOGRÁFICA CON LA SEÑAL HRV (PRIMERA PARTE) Y DERIVADOS EN COLOR MORADO.	27
TABLA 2-4. TABLA RESUMEN REVISIÓN BIBLIOGRÁFICA CON LA SEÑAL HRV (SEGUNDA PARTE) Y DERIVADOS EN COLOR MORADO Y LAS SEÑALES SpO_2 -FA EN COLOR VERDE.	28
TABLA 2-5. TABLA RESUMEN REVISIÓN BIBLIOGRÁFICA CON LAS SEÑALES SpO_2 -HRV (Y DERIVADAS) EN ROJO, LAS SEÑALES FA-HRV (ECG) EN AZUL CLARO Y LAS SEÑALES SpO_2 -HRV(ECG)-FA EN AMARILLO.	29
TABLA 3-1. TABLA RESUMEN DE LOS DATOS CONTENIDOS EN LA BASE DE DATOS CHAT (JIMÉNEZ-GARCÍA, 2022).	32
TABLA 4-1. ESTRUCTURA GENERAL DE UNA MATRIZ DE CONFUSIÓN.	50
TABLA 4-2. LIBRERÍAS EMPLEADAS Y SUS VERSIONES.	53
TABLA 4-3. CARACTERÍSTICAS DEL HARDWARE EMPLEADO.	54
TABLA 5-1. TABLA DE RESULTADOS DE OPTIMIZACIÓN DEL HIPERPARÁMETRO NUM_FILTERS PARA LA SEÑAL SpO_2 (R2_SCORE).	56
TABLA 5-2. TABLA DE RESULTADOS DE OPTIMIZACIÓN DEL HIPERPARÁMETRO NUM_FILTERS PARA LA SEÑAL SpO_2 (VAL_LOSS).	56
TABLA 5-3. RANGO DE VALORES EMPLEADOS PARA LA OPTIMIZACIÓN DE LOS DISTINTOS HIPERPARÁMETROS PARA CADA UNA DE LAS COMBINACIONES DE SEÑALES.	57
TABLA 5-4. VALORES DE HIPERPARÁMETROS ÓPTIMOS PARA LAS SEÑALES INDIVIDUALES SpO_2 , FA Y HRV.	57
TABLA 5-5. VALORES DE HIPERPARÁMETROS ÓPTIMOS PARA LAS SEÑALES EMPAREJADAS 2 A 2 SpO_2 -FA, SpO_2 -HRV Y FA-HRV.	58
TABLA 5-6. VALORES DE HIPERPARÁMETROS ÓPTIMOS PARA TODAS SEÑALES EMPAREJADAS SpO_2 -FA-HRV.	58
TABLA 5-7. RESULTADOS DEL PROCESO DE OPTIMIZACIÓN EN LAS VARIABLES VAL_LOSS Y R2_SCORE.	58
TABLA 5-8. RESULTADOS DE VAL_LOSS Y R2_SCORE OBTENIDOS EN EL MEJOR MODELO DE CADA UNA DE LAS SEÑALES PARA LOS VALORES DE DROPOUT 0.1 Y 0.2.	59
TABLA 5-9. COMPARATIVA DE LOS RESULTADOS ESTADÍSTICOS (PRECISION, RECALL, F1-SCORE Y KAPPA) DE LAS MATRICES DE CONFUSIÓN TRAS LA REGRESIÓN DE TODAS LAS COMBINACIONES DE SEÑALES PARA EL CONJUNTO DE TEST.	68
TABLA 5-10. RESULTADOS UMBRALIZACIÓN BINARIA MATRIZ DE CONFUSIÓN DE LA SEÑAL SpO_2 PARA LOS CONJUNTOS DE ENTRENAMIENTO, VALIDACIÓN Y TEST.	69
TABLA 5-11. RESULTADOS UMBRALIZACIÓN BINARIA MATRIZ DE CONFUSIÓN DE LA SEÑAL FA PARA LOS CONJUNTOS DE ENTRENAMIENTO, VALIDACIÓN Y TEST.	69
TABLA 5-12. RESULTADOS UMBRALIZACIÓN BINARIA MATRIZ DE CONFUSIÓN DE LA SEÑAL HRV PARA LOS CONJUNTOS DE ENTRENAMIENTO, VALIDACIÓN Y TEST.	69
TABLA 5-13. RESULTADOS UMBRALIZACIÓN BINARIA MATRIZ DE CONFUSIÓN DE LA SEÑAL SpO_2 -FA PARA LOS CONJUNTOS DE ENTRENAMIENTO, VALIDACIÓN Y TEST.	70
TABLA 5-14. RESULTADOS UMBRALIZACIÓN BINARIA MATRIZ DE CONFUSIÓN DE LA SEÑAL SpO_2 -HRV PARA LOS CONJUNTOS DE ENTRENAMIENTO, VALIDACIÓN Y TEST.	70
TABLA 5-15. RESULTADOS UMBRALIZACIÓN BINARIA MATRIZ DE CONFUSIÓN DE LA SEÑAL FA-HRV PARA LOS CONJUNTOS DE ENTRENAMIENTO, VALIDACIÓN Y TEST.	70
TABLA 5-16. RESULTADOS UMBRALIZACIÓN BINARIA MATRIZ DE CONFUSIÓN DE LA SEÑAL SpO_2 -FA-HRV PARA LOS CONJUNTOS DE ENTRENAMIENTO, VALIDACIÓN Y TEST.	71
TABLA 6-1. COMPARATIVA ENTRE LAS SEÑALES SpO_2 , FA Y SpO_2 -FA EN EL CONJUNTO DE TEST.	74
TABLA 6-2. COMPARATIVA ENTRE LAS SEÑALES SpO_2 , HRV Y SpO_2 -HRV EN EL CONJUNTO DE TEST.	74
TABLA 6-3. COMPARATIVA ENTRE LAS SEÑALES FA, HRV Y FA-HRV EN EL CONJUNTO DE TEST.	74
TABLA 6-4. COMPARATIVA ENTRE LAS SEÑALES SpO_2 , FA, HRV Y SpO_2 -FA-HRV EN EL CONJUNTO DE TEST.	75
TABLA 6-5. COMPARATIVA ENTRE LAS SEÑALES SpO_2 -FA, SpO_2 -HRV, FA-HRV Y SpO_2 -FA-HRV EN EL CONJUNTO DE TEST.	75
TABLA 6-6. TABLA COMPARATIVA DE ESTUDIOS PREVIOS CON NUESTRO ESTUDIO PARA LA SEÑAL SpO_2 Y UMBRALES DE 1, 5 Y 10E/H.	77
TABLA 6-7. TABLA COMPARATIVA DE ESTUDIOS PREVIOS CON NUESTRO ESTUDIO PARA LA SEÑAL FA Y UMBRALES DE 1, 5 Y 10E/H.	78
TABLA 6-8. TABLA COMPARATIVA DE ESTUDIOS PREVIOS CON NUESTRO ESTUDIO PARA LA SEÑAL HRV Y UMBRALES DE 1, 5 Y 10E/H.	79
TABLA 6-9. TABLA COMPARATIVA DE ESTUDIOS PREVIOS CON NUESTRO ESTUDIO PARA LA SEÑAL SpO_2 -FA Y	

UMBRALES DE 1, 5 Y 10E/H.....	80
TABLA 6-10. TABLA COMPARATIVA DE ESTUDIOS PREVIOS CON NUESTRO ESTUDIO PARA LA SEÑAL SpO_2 -HRV Y UMBRALES DE 1, 5 Y 10E/H.....	81
TABLA 6-11. TABLA COMPARATIVA DE ESTUDIOS PREVIOS CON NUESTRO ESTUDIO PARA LA SEÑAL FA-HRV Y UMBRALES DE 1, 5 Y 10E/H.....	81
TABLA 6-12. TABLA COMPARATIVA DE ESTUDIOS PREVIOS CON NUESTRO ESTUDIO PARA LA SEÑAL SpO_2 -FA-HRV Y UMBRALES DE 1, 5 Y 10E/H.....	82

Índice de siglas y acrónimos

AASM	<i>American Academy of Sleep Medicine</i>
Acc	<i>Accuracy</i>
ACS	<i>Apnea Central del Sueño</i>
AMS	<i>Apnea Mixta del Sueño</i>
ANN	<i>Artificial Neural Networks</i>
AOS	<i>Apnea Obstructiva del Sueño</i>
CHAT	<i>Childhood Adenotonsillectomy Trial</i>
CNN	<i>Convolutional Neural Networks</i>
CO ₂	<i>Dióxido de Carbono</i>
CPAP	<i>Continuous Positive Airway Pressure</i>
DL	<i>Deep Learning</i>
ECG	<i>Electrocardiograma</i>
EEG	<i>Electroencefalograma</i>
EMG	<i>Electromiograma</i>
EOG	<i>Electro-oculograma</i>
FA	<i>Flujo Aéreo</i>
FFT	<i>Fast Fourier Transformation</i>
FT	<i>Fourier Transformation</i>
HRV	<i>Heart Rate Variability</i>
HSAT	<i>Home Sleep Apnea Testing</i>
IA	<i>Inteligencia Artificial</i>
IAH	<i>Índice de Apnea-Hipopnea</i>
ICC	<i>Intraclass Correlation Coefficient</i>
IMC	<i>Índice Masa Corporal</i>
IoT	<i>Internet of Things</i>
IRM	<i>Imagen por Resonancia Magnética</i>
LSTM	<i>Long Short-Term Memory</i>
ML	<i>Machine Learning</i>
OMS	<i>Organización Mundial de la Salud</i>
ORP	<i>Odds Ratio Product</i>
OSA	<i>Obstructive Sleep Apnea</i>
PLN	<i>Procesado de Lenguaje Natural</i>
PPG	<i>Photoplethysmography</i>
PRV	<i>Pulse Rate Variability</i>
PSG	<i>Polisomnografía</i>
PTT	<i>Pulse Transit Time</i>
ReLU	<i>Rectified Linear Unit</i>
REM	<i>Rapid Eye Movement</i>
RNN	<i>Recurrent Neural Networks</i>
Se	<i>Sensibilidad</i>
Sp	<i>Especificidad</i>
SpO ₂	<i>Saturación de Oxígeno en Sangre</i>
WCOB	<i>Weighted Center of Bispectrum</i>
WWSC	<i>Watchful Waiting with Supportive Care</i>
XAI	<i>Explainable Artificial Intelligence</i>

Capítulo 1: INTRODUCCIÓN

La apnea obstructiva del sueño (AOS), conocida en la literatura anglosajona como *Obstructive Sleep Apnea* (OSA), constituye una de las afecciones de mayor prevalencia dentro del ámbito de las enfermedades respiratorias. Esta patología se caracteriza por la presencia de una serie de episodios de carácter recurrente durante la fase del sueño en los cuales tiene lugar una obstrucción total o parcial de las vías respiratorias superiores. Durante el transcurso de dichos episodios, se generan una serie de alteraciones respiratorias que causan eventos de respiración anormales. Estas anomalías pueden derivar en una serie de efectos adversos para la salud en los pacientes que los padecen (Bitners & Arens, 2020a; Moffa et al., 2020). Entre los motivos principales causantes de estos episodios nocturnos destacan: fallos en las funciones de ciertos componentes anatómicos del cuerpo humano, como son los tejidos blandos y los músculos dilatadores presentes en la faringe; defectos anómalos en la estructura craneofacial, debidos a problemas hereditarios o enfermedades degenerativas; obstrucción de las vías buconasales, causado por un exceso de mucosa o tejido blando; o defectos en piezas dentales y la propia composición ósea de la boca (Malhotra & White, 2002). No obstante, las alteraciones neurológicas también pueden desenvolver un papel fundamental en la aparición de estos episodios. Esto se debe a que la obstrucción de las vías respiratorias puede derivar de la ausencia o insuficiencia de reflejos nerviosos encargados de regular el flujo aéreo durante la etapa de sueño, es decir, la falta de estímulo suficiente impide que los músculos de la garganta se mantengan en la posición adecuada para permitir una respiración normal (Bitners & Arens, 2020a).

Se estima que esta patología tan solo en España afecta a un 6-8% de la población española¹, es decir, en torno a 3-4 millones de personas (Álvarez-Sala et al., 2015) y en Estados Unidos afecta aproximadamente al 12% de la población², es decir, más de 40 millones de personas (Ling & Wu, 2024). A nivel global, la AOS afecta a mil millones de personas, convirtiéndola en el trastorno orgánico del sueño más común en todo el mundo (Acevedo et al., 2018). Estas cifras son excesivamente preocupantes teniendo en cuenta que muchos de los casos permanecen sin ser diagnosticados y pueden suponer un problema grave de salud para los pacientes que la padecen (Bitners & Arens, 2020a).

Aunque la aparición de los primeros textos académicos con una definición oficial del síndrome de la apnea del sueño data de los años 70 (Guilleminault et al., 1977), años atrás ya se podían observar documentos que mencionaban este trastorno de manera indirecta y sin percatarse de ello. Concretamente, si nos remontamos a los años 30 del siglo XIX, en torno a 1835-1837, se publicaría la obra del científico escritor Charles Darwin, conocida como *The Posthumous Papers of the Pickwick club* (Dickens, 2008). Esta novela describe con detalle a un sujeto obeso conocido como *Joe* pareciente de una patología denominada por el propio autor como Somnolencia Diurna Excesiva (SDE) (Dickens, 1837; Guilleminault & Abad, 2004). En dicha novela se describe a *Joe* como: “un niño maravilloso, gordo [...] está de pie sobre la alfombra, con los ojos cerrados como si estuviera dormido”. Como se puede apreciar, el escritor inglés acababa de relatar un caso de apnea del sueño en un sujeto con obesidad sin ser consciente de ello. Posteriormente, surgirían los primeros artículos con un enfoque más riguroso y científico que tratarían de evaluar este fenómeno del SDE. Concretamente, un estudio publicado en 1956 narraría de nuevo la historia de una persona con obesidad desde un enfoque más analista y detallado. En dicho informe se destaca la importancia de una correcta ventilación alveolar, y se propone la posibilidad de evaluar la hipoventilación del paciente mediante el análisis de la respuesta respiratoria a variaciones de dióxido de carbono (CO_2), pudiendo realizar así un diagnóstico adecuado (Burwell et al., 1994).

¹ La población española a fecha de Octubre de 2024 es de 48.797.875 personas, según datos oficiales extraídos directamente del Instituto Nacional de Estadística o INE. Disponible públicamente a través del siguiente enlace: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177095&menu=ultiDatos&idp=1254735572981

² La población de Estados Unidos o EEUU a fecha de Octubre de 2024 es de 335.135.000 personas, según datos oficiales extraídos directamente de la base de datos Datosmacro. Disponible públicamente a través del siguiente enlace: <https://datosmacro.expansion.com/paises/usa>

FACTOR DE RIESGO	GRADO	MOTIVOS
Sexo masculino	MEDIO	Factores anatómicos
Menopausia	BAJO	Desconocido, posiblemente anatómicos
Edad	MEDIO	Factores anatómicos y reflejos
Obesidad	ALTO	Factores anatómicos
Alcohol	MEDIO	Afección a los reflejos musculares
Tabaco	BAJO	Inflamación del tracto respiratorio
Raza	BAJO	Desconocido, posiblemente anatómicos

Tabla 1-1. Factores de riesgo que favorecen la posibilidad de padecer AOS (Bitners & Arens, 2020a).

Debemos tener en cuenta la existencia de ciertos factores de riesgo como la obesidad, las enfermedades degenerativas o incluso el género de los pacientes, que pueden inducir en una mayor probabilidad de desarrollar AOS (Malhotra & White, 2002). En la Tabla 1-1 se recogen los factores de riesgo más relevantes junto al grado de peligrosidad que implica.

Es crucial destacar el significativo impacto que tiene la obesidad en la posibilidad de desarrollar AOS. De hecho, se estima que el 70% de pacientes que padecen esta patología presentan además síntomas de obesidad (Malhotra & White, 2002). Empleando el Índice de Masa Corporal (IMC) como parámetro de referencia, podemos clasificar la relación peso/altura de una persona como: infra peso ($IMC < 18.5 [Kg/m^2]$), peso normal ($18.5 [Kg/m^2] \leq IMC < 24.9 [Kg/m^2]$), sobrepeso ($25 [Kg/m^2] \leq IMC < 29.9 [Kg/m^2]$), obesidad leve o tipo I ($30 [Kg/m^2] \leq IMC < 34.9 [Kg/m^2]$), obesidad moderada o tipo II ($35 [Kg/m^2] \leq IMC < 39.9 [Kg/m^2]$) u obesidad mórbida o tipo III ($IMC \geq 40 [Kg/m^2]$) (Okunogbe et al., 2022). Es importante tener en cuenta que en el caso de los niños, la forma de calcular el grado de obesidad o el nivel de peso es diferente a la utilizada con los adultos.

Otro de los factores determinantes, aunque no tan importantes como la obesidad, es el género o sexo de los pacientes. Los hombres tienen un mayor riesgo de sufrir AOS que las mujeres, esto se debe a diferencias en su condición anatómica como: la estructura y tamaño de los músculos, una distribución de grasa distinta, diferencias hormonales y otros factores que difieren de un sexo a otro que podrían ser la explicación a este fenómeno (Malhotra & White, 2002). Sin embargo, ciertos estudios argumentan que esta diferencia de posibilidades de padecer AOS viene sesgada por la desigualdad entre la cantidad de investigaciones realizadas en hombres y en mujeres, y en las diferencias que presentan ambos sexos en los síntomas, como por ejemplo los ronquidos (Malhotra & White, 2002). Es decir, parte de esa disparidad en sufrir o no de este trastorno puede darse por el infradiagnóstico de la AOS en mujeres (Malhotra & White, 2002). Aun así, independientemente de esta teoría las estadísticas entre hombres y mujeres se igualan en dos etapas de la vida: la primera, en las edades más tempranas hasta alcanzar la pubertad; y la segunda, durante la etapa menopaúsica, aumentando en las mujeres las posibilidades de sufrir AOS y llegando a igualar la de los hombres (Ho & Brass, 2011). Además, resulta interesante destacar como la raza puede afectar en la posibilidad de desarrollar este trastorno. A pesar de no tener una respuesta concreta, la hipótesis más acertada indica que las diferencias en la composición y distribución de los tejidos blandos y las estructuras óseas pueden presentar ciertas variaciones en las distintas razas ante este trastorno (Malhotra & White, 2002). En un estudio realizado en 1997 se evaluó a un total de 225 afroamericanos y 622 caucásicos de entre 2 y 86 años. El estudio concluyó que los jóvenes afroamericanos tienen un mayor riesgo de padecer apnea del sueño que aquellos que no lo son (Redline et al., 1997). Otros factores como la edad y el consumo de alcohol o tabaco también puede afectar significativamente, aunque estos factores no tienen un impacto tan crítico comparados con la obesidad y el sexo citados previamente (Bitners & Arens, 2020a).

1.1. Características de la Apnea Obstructiva del Sueño

Antes de comenzar a hablar de las principales características, síntomas y formas de diagnóstico de este trastorno, debemos diferenciar los distintos tipos de apnea del sueño que existen. Para ello, podemos separarlas en función del origen o motivo que las causa (Xia & Sawan, 2021):

- **Apnea Obstructiva del Sueño (AOS):** es el tipo de apnea del sueño más frecuente, relacionada con la obstrucción de la vía aérea superior. Mediante procesos biomecánicos como la relajación de los músculos, ocurre una obstrucción parcial o total de la cavidad traqueal dificultando el flujo de aire.
- **Apnea Central del Sueño (ACS):** este tipo de apnea del sueño es originada por problemas neurológicos. La interrupción del flujo aéreo se origina por una alteración en la actividad neuronal responsable de mantenerlo, ya que no se generan adecuadamente los estímulos necesarios para activar los músculos implicados.
- **Apnea Mixta del Sueño (AMS):** la menos frecuente de las 3 ya que combina ambas mencionadas previamente, este tipo de apnea es la más peligrosa y compleja al incluir problemas otorrinolaringológicos y neuronales.

En este caso de estudio nos enfocaremos principalmente en la AOS. Este tipo de apnea viene caracterizada por dos procesos fisiológicos definidos en la *American Academy of Sleep Medicine* (AASM) de la siguiente forma (Berry et al., 2012; Eguía et al., 2007; Malhotra et al., 2021):

- **Apnea:** se conoce como apnea al evento en el que ocurre una disminución del flujo aéreo de al menos el 90% durante más de 10 segundos en adultos y de 2 ciclos respiratorios en niños (aproximadamente 6 segundos).
- **Hipopnea:** en este caso se define como hipopnea al evento en el que ocurre una disminución del flujo aéreo pero en este caso de entre el 30% y el 90% durante más de 10 segundos en adultos y de 2 ciclos respiratorios en niños (6 segundos aproximadamente). Además, se experimenta también una caída del nivel de saturación de oxígeno en sangre (SpO_2) del 3% o un microdespertar.

La aparición recurrente de episodios de apnea e hipopnea durante el sueño provoca un flujo de aire irregular e insuficiente en el sistema respiratorio, lo que conlleva una disminución de la concentración de oxígeno en la sangre arterial. Este fenómeno se conoce como hipoxemia o desaturación de oxígeno en sangre (Eguía et al., 2007). Debido a la necesidad de las células de nuestro organismo de disponer de oxígeno con el que poder generar energía, la pérdida porcentual de este en el torrente sanguíneo puede afectar y dañar células e incluso órganos, siendo los más críticos el sistema neuronal con el cerebro y el cardiovascular con el corazón (Terán Santos et al., 2006). Por esa razón este trastorno conlleva graves problemas para la salud de quien lo padece, tanto a corto como a largo plazo.

Ante una disminución de oxígeno en el torrente sanguíneo, nuestro cuerpo trata de compensar esa pérdida enviando estímulos al sistema respiratorio para que intente volver a un estado de respiración normal, lo que puede generar un esfuerzo respiratorio aumentado, obligando a ciertos músculos a trabajar más para compensar esta falta de oxígeno (National Heart, 2016; Terán Santos et al., 2006). Además, el corazón también se ve perjudicado al verse forzado a aumentar su carga de trabajo con el fin de bombear más sangre para que a los órganos vitales no les afecte dicha disminución de oxígeno, provocando así incrementos en la frecuencia cardíaca y la presión arterial no deseados. En algunos casos, si no se logra compensar la falta de oxígeno el cerebro puede activar el estado de alarma y generar un microdespertar, para intentar así restablecer el correcto flujo de aire en nuestro cuerpo y recuperar los niveles de oxígeno en sangre normales (Terán Santos et al., 2006).

Una vez el paciente se despierta, se recupera el control sobre las vías respiratorias y generalmente las funciones corporales se restablecen a la normalidad. Pero estos eventos de apnea-hipopnea pueden darse varias veces en una noche, generando desaturaciones y microdespertares continuos y evitando que el organismo entre en la fase de sueño profundo. Esta fase es esencial para poder descansar correctamente y así poder funcionar adecuadamente durante el día, por lo que no alcanzarla de forma recurrente puede acabar desencadenando problemas de salud graves a largo plazo (Terán Santos et al., 2006).

El trastorno de la AOS presenta una serie de síntomas para los pacientes que lo sufren tanto diurnos como nocturnos. Por la noche, el sujeto puede experimentar síntomas como: asfixias, ronquidos, pesadillas, congestión nasal, insomnio o nocturia entre otros (Olivi, 2013). Por otro lado, sufrir eventos de apnea obstructiva por la noche, también tiene repercusión en el día a día. Pudiendo presentar síntomas como: cansancio, fatiga, irritabilidad, apatía, depresión, dificultad para memorizar y concentrarse, problemas cardiovasculares e incluso una mayor probabilidad de sufrir accidentes de tráfico (Olivi, 2013). Según la Sociedad Española de Neumología y Cirugía Torácica, se descubrió en 1998 que la AOS era la segunda causa de muerte en accidentes de tráfico en España, siendo los pacientes con un grado elevado de AOS hasta 7 veces más propensos a verse involucrados en accidentes de tráfico que aquellos que no presentan este trastorno (Terán-Santos et al., 1999). Esta cifra, puede llegar a incrementarse hasta 11 si se consume alcohol y drogas a mayores (Ana et al., 2006).

1.2. Diagnóstico de la Apnea Obstructiva del Sueño

En el contexto del diagnóstico de la AOS, la polisomnografía (*polysomnography*, PSG) es considerada el estándar de referencia tanto en adultos como en niños. Esta prueba debe realizarse en unidades del sueño especializadas, adecuadamente equipadas y por un equipo de profesionales, garantizando así unos resultados precisos y confiables (Bitners & Arens, 2020a; Mazzotti et al., 2018). Realizar esta prueba, especialmente en niños, presenta una dificultad añadida debido a la posibilidad de no poder conciliar el sueño correctamente en un entorno extraño durante la noche, lo que puede alterar los resultados. Durante esta prueba el paciente debe dormir en la unidad del sueño con varios sensores y electrodos conectados a su cuerpo con el objetivo de registrar las señales biomédicas de interés y así poder diagnosticar el grado de AOS (Becerra et al., 1991; Behera et al., 2021; Budhiraja et al., 2005; Leske, 2014; Mazzotti et al., 2018; Mildenhall, 2008). Algunas de las señales típicamente recogidas son:

- **Electro-oculograma (EOG):** permite registrar la actividad eléctrica relacionada con los movimientos oculares lentos de la fase no REM y los rápidos de la fase REM además del potencial eléctrico del ojo. Para poder registrar estas señales se colocan electrodos en varios puntos cercanos a los músculos del ojo por la parte externa.
- **Electrocardiograma (ECG):** permite registrar la actividad eléctrica del corazón y calcular el ritmo cardíaco, permitiendo detectar anomalías asignadas a los efectos de los trastornos respiratorios producidos por patologías como la AOS. En este caso los electrodos se colocan a lo largo de la piel del paciente, concretamente por los brazos, piernas y pecho.
- **Electromiograma mentoniano (EMGm):** permite registrar la actividad eléctrica de los músculos del mentón para ayudar a identificar la fase REM. De igual manera que en los casos previos, se emplean electrodos colocados a lo largo del mentón.
- **Electroencefalograma (EEG):** permite registrar la actividad eléctrica del cerebro de manera no invasiva mediante una serie de electrodos colocados en el cuero cabelludo alrededor del cráneo. Estos electrodos se deben colocar en ciertas regiones de interés según el sistema internacional 10-20 y es recomendable registrar como mínimo 6 canales en las zonas frontal, central y occipital (F1, F2, C3, C4, O1 y O2) para poder realizar correctamente la prueba (Figura 1-1).

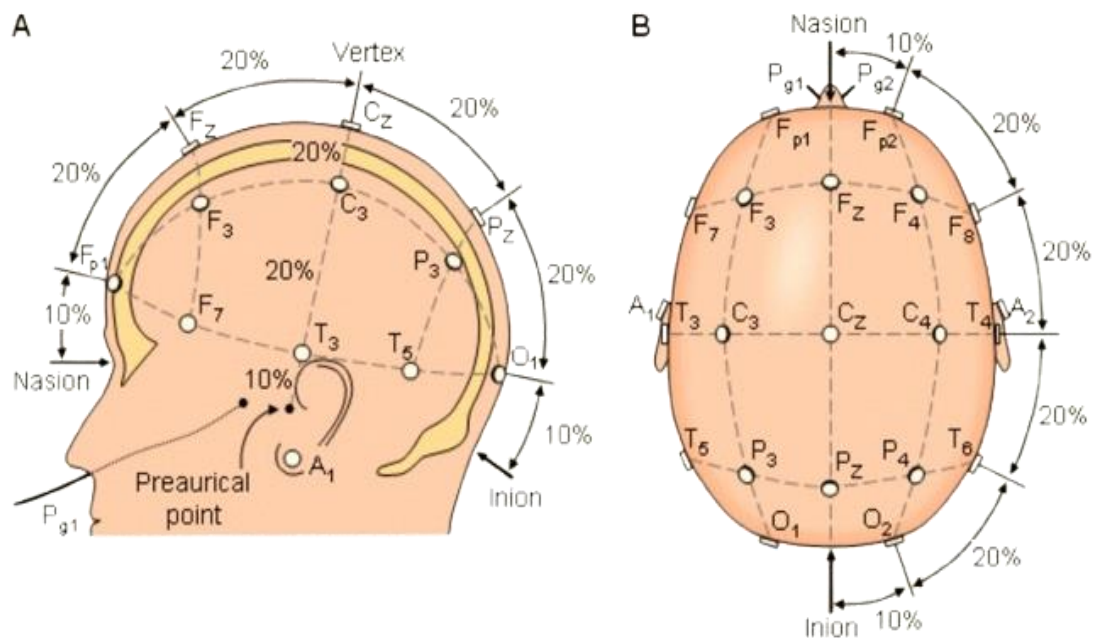


Figura 1-1. Zonas de colocación de electrodos vistas lateral (izquierda) y superior (derecha): C-central, F-frontal, O-Occipital, A-Auricular y P-Parietal (Departamento de fisiología, facultad de medicina UNAM).

- **Flujo aéreo (FA) respiratorio:** sensores como las cánulas de presión³, permiten detectar el flujo aéreo al registrar variaciones de temperatura entre el aire que el paciente inhala y exhala durante la fase del sueño. Estos dispositivos se colocan alrededor de la boca y la nariz para poder detectar los cambios de temperatura.
- **Movimiento/esfuerzo respiratorio:** sensores capaces de registrar el movimiento torácico y abdominal. Para ello, se emplean bandas toracoabdominales que permiten realizar un seguimiento de los movimientos realizados al inspirar y expirar durante el sueño.
- **Pulsioximetría:** empleando un pulsioxímetro, se puede registrar la señal SpO₂ y la frecuencia del pulso. Generalmente, este dispositivo se presenta en forma de pinza que, al colocarse en el dedo, utiliza diodos emisores de luz (LED) y fotodiodos para emitir y captar luz roja e infrarroja, permitiendo así medir la cantidad de luz absorbida por la sangre en dos longitudes de onda diferentes.
- **Nivel de dióxido de carbono:** la medición se realiza a través de sensores que detectan la concentración de CO₂ en el aire espirado, ya sea mediante capnografía o monitorización transcutánea. Estos dispositivos emplean tecnologías ópticas que miden la absorción de luz por parte del CO₂, lo que permite estimar con precisión sus niveles durante la respiración.
- **Presión esofágica:** permite medir el esfuerzo inspiratorio del paciente mediante cambios de presión sobre un globo hinchado con aire. Este dispositivo consiste en un catéter pequeño con un globo inflable en su extremo que se introduce por la nariz o boca hasta el esófago. Al ser tan invasivo, su uso es limitado.
- **Material audiovisual:** permite realizar un seguimiento del estado del paciente de forma audible y visual. Consiste en realizar una grabación en video del sujeto para poder controlar sus movimientos y una grabación de audio para poder registrar sonidos como ronquidos o respiraciones forzosas.

³ Cánula de presión: compuesta por un tubo flexible colocado en las fosas nasales (conocido como tubo nasal) y un transductor de presión que convierte las variaciones de presiones del aire en señales medibles (Wegener, 2017).

La PSG presenta una elevada complejidad, un alto costo de realización y un posible sesgo en los resultados (Leske, 2014). Este sesgo se debe principalmente a que el paciente se encuentra en un entorno desconocido como es una unidad del sueño, y tiene gran cantidad de aparataje y sensores colocados por todo su cuerpo, lo que puede ocasionar efectos no deseados como no poder conciliar el sueño adecuadamente. Por ello a lo largo de los últimos años se han investigado una gran variedad de alternativas para reducir costes y tiempo de diagnóstico, facilitando la prueba tanto al médico como al paciente. La opción más prometedora y elegida de forma frecuente por los especialistas médicos es la realización de una prueba simplificada en el domicilio de la persona (*Home Sleep Apnea Test*, HSAT). Esta prueba permite registrar algunas de las señales citadas anteriormente en el propio hogar del paciente, logrando así abaratar costes tanto en recursos como en tiempo, y permitiendo una mayor comodidad para el sujeto gracias a la posibilidad de poder dormir y descansar correctamente en un entorno conocido como es su propia habitación. En el HSAT se registran generalmente señales como: la frecuencia cardíaca, la SpO₂ mediante pulsioximetría, grabación de audio para detectar ronquidos y el flujo aéreo, entre otros. No obstante, las señales recogidas pueden variar entre pacientes, médicos e incluso dispositivos (Bitners & Arens, 2020a; Caples et al., 2021).

Blumenthal y McGinnis afirmaron: “*if something cannot be measured... it cannot be improved*” (Malhotra et al., 2021). Esta celebre cita hace referencia a la importancia de cuantificar todo aquello que nos rodea con el fin de poder analizarlo y mejorarlo. Aplicándolo a nuestro caso de estudio, el objetivo principal de tanto la PSG como el HSAT no es otro que diagnosticar a los pacientes de AOS cuantizando el grado de severidad que presentan de dicho trastorno, pudiendo así aplicar el mejor tratamiento a cada uno de ellos en función del grado diagnosticado. Para ello, como en todo campo de la medicina se emplean puntuaciones o métricas con las que poder clasificar y evaluar las diferentes enfermedades o patologías que pueda presentar un paciente. Sin embargo, la métrica por excelencia y, por tanto, la que más se emplea para diagnosticar los distintos grados de AOS es el Índice de Apnea-Hipopnea (IAH). La primera mención de esta métrica, se remonta a 1976, siendo Guilleminault el primero en emplear el término “*apnea index*” en su publicación “*The Sleep Apnea Syndromes*” (Guilleminault et al., 1977; Pevernagie et al., 2020). Este índice representa la relación entre la cantidad de eventos dados en un periodo de sueño y el tiempo del dicho periodo, es decir, podemos darle forma con la siguiente expresión (Ecuación 1.1):

$$IAH = \frac{\text{número de apneas e hipopneas [eventos]}}{\text{tiempo de sueño [h]}} \quad (1.1)$$

En función del IAH calculado en la prueba se pueden clasificar a los pacientes adultos que sufren de este trastorno de 4 maneras distintas (Berry et al., 2012; Guilleminault, 2018; Malhotra et al., 2021; Mediano et al., 2022; Pevernagie et al., 2020):

- **Sin AOS:** en el caso de que el $IAH < 5$ [eventos/hora], se considerará que el paciente no tiene o tiene eventos de apnea-hipopnea dentro de lo normal, y por lo tanto no sufre de la patología de AOS, por lo que no hace falta ningún tratamiento.
- **AOS leve:** el grado de apnea menos peligroso, pero no por ello menos importante. En este caso, el $IAH \sim \{5 - 14\}$ [eventos/hora], se caracteriza por pocos eventos de apnea durante la etapa de sueño. A pesar de no ser peligroso, puede indicar una fase temprana de AOS más severa y debe ser vigilado regularmente y tratado.
- **AOS moderada:** en este caso, el $IAH \sim \{15 - 30\}$ [eventos/hora], lo que implica eventos de carácter frecuente dificultando bastante el descanso adecuado, sobre todo si dichos eventos van acompañados de un microdespertar.
- **AOS severa:** este es el caso más grave de todos, en este caso el $IAH > 30$ [eventos/hora], lo que implica eventos de carácter muy frecuente durante la etapa de sueño, imposibilitando el correcto descanso del paciente y desencadenando problemas de salud graves además de otras patologías.

A pesar de que el IAH es la métrica más empleada, existen otras alternativas como son por ejemplo (Malhotra et al., 2021; Torre-Bouscoulet et al., 2007): la carga hipóxica, haciendo referencia a la cantidad de tiempo que el paciente se encuentra en niveles muy bajos de oxígeno en sangre durante un periodo de tiempo; la intensidad del microdespertar, como su propio nombre indica, mide la fuerza con la que la persona se despierta, clasificando como muy fuerte al microdespertar que logra despertar por completo al sujeto, y como débil al microdespertar del cual el paciente no es consciente; el *Odds Ratio Product* (ORP), empleado para cuantificar la profundidad y calidad de sueño del paciente, usando para ello el EEG; el Índice de Desaturación de Oxígeno al 3% (*Oxygen Desaturation Index 3%*, ODI3), indicador empleado para medir cuantas veces por unidad de tiempo (generalmente 1 hora) el nivel de oxígeno en sangre cae un cierto umbral (frecuentemente 3%); o la duración del evento de apnea-hipopnea, similar al IAH, pero midiendo los lapsos temporales de los eventos en lugar de la cantidad de ellos (Malhotra et al., 2021; Torre-Bouscoulet et al., 2007). Estas son solo algunas de las métricas más relevantes discutidas en los estudios. No obstante, la elección de la métrica más adecuada en cada caso y sujeto, quedará determinada por el contexto de la situación, las características de la persona y la experiencia o preferencias del profesional de la salud encargado. Ya que, no en todos los casos están presentes los mismo síntomas o se tiene la misma estructura anatómica.

1.3. Tratamientos de la Apnea Obstructiva del Sueño

Como la gran mayoría de los trastornos y enfermedades comúnmente conocidas, existe uno o varios tratamientos con los que poder eliminar o reducir los síntomas que presentan. En el caso de la AOS hay una gran variedad de opciones disponibles, que dependiendo del paciente y del grado de AOS, pueden ser más o menos invasivos, complejos o efectivos. Al igual que ocurría con el IAH en la parte de diagnóstico con la existencia de una métrica estándar para clasificar los grados de AOS, ocurre con el tratamiento en adultos. El tratamiento más frecuente y efectivo empleado para abordar la AOS en adultos, es la presión positiva continua en vías respiratorias, o CPAP por sus siglas en inglés *Continuous Positive Airway Pressure* (Bitners & Arens, 2020b).

La CPAP consiste en un tipo de tratamiento donde se intenta mantener un flujo de aire continuo gracias a una presión positiva constante ofrecida mediante una mascarilla y un aparato regulador de presión (Figura 1-2). Esta presión positiva constante logra mantener las vías respiratorias en un estado adecuado para que haya un correcto flujo aéreo y el paciente pueda respirar con normalidad. Por lo general, la máquina es ajustable en función de la presión u otras necesidades fisiológicas del sujeto y puede emplearse una mascarilla nasal, oronasal u oral con distintas formas, tamaños y áreas de cobertura en función de los gustos de cada persona (H. Chang et al., 2020).



Figura 1-2. Dibujo esquemático de un tratamiento con CPAP (Autoría propia, generada con IA).

Para conseguir que el tratamiento logre surtir efecto, la CPAP debe de usarse como mínimo 4 horas por día (depende del grado de la AOS), 5 días a la semana (Xia & Sawan, 2021). Aunque se ha notado una mejoría considerable en la variabilidad de la frecuencia cardíaca de los pacientes durante la noche, simplemente por comenzar a aplicar el tratamiento (Xia & Sawan, 2021). El seguimiento de estas directrices permitirá al paciente disminuir la posibilidad de sufrir enfermedades cardiovasculares, diabetes o hipertensión, entre otras. Además, mejorará la calidad de vida tanto física como psicológica y mental al permitir al paciente tener un correcto sueño y por consiguiente un descanso adecuado, permitiendo al cerebro y resto de órganos repararse y funcionar correctamente (Xia & Sawan, 2021). Tras 3 meses, se ha comprobado que los pacientes experimentan mejoría en la presión sanguínea, capacidad pulmonar e incluso mejoras a nivel ocular tanto en la retina como en la sensibilidad de la vista (Spannella et al., 2017).

Sin embargo, como cualquier otra intervención o tratamiento, la terapia CPAP tiene una serie de desventajas que a pesar de no estar presentes en todos los casos, deben ser consideradas. Alguno de los problemas o efectos secundarios que puede presentar un sujeto tras el uso de la CPAP pueden ser: sequedad, congestión, hemorragias y otras molestias nasales, distensión abdominal y otros problemas gastrointestinales o problemas dentales y mandibulares (Buela-Casal G, 1999). Además, se pueden originar otros problemas como incomodidad al llevar la mascarilla o despertares por el ruido generado por la máquina, aunque estos últimos pueden ser subsanados cambiando el equipo para que se ajuste a las necesidades del usuario.

Aunque la CPAP es el tratamiento más empleado en adultos, existen otros dependiendo de las necesidades y condiciones médicas de la persona, pudiendo ser más o menos invasivos, largos o agresivos. Una de las alternativas menos invasivas consiste en dispositivos colocados en la boca para corregir la forma de la mandíbula o la colocación de la lengua en el paciente. Pero se debe tener en cuenta que este tipo de tratamientos solo funciona con aquellos sujetos que sufran de malformaciones o anomalías bucales o mandibulares que afecten directa o indirectamente a la AOS (Konrad, 2006; Malhotra & White, 2002). Por otro lado, un cambio de hábitos en la vida del usuario puede ser también un tratamiento efectivo en casos más leves. Estos hábitos incluyen el cese de consumo de sustancias nocivas tales como alcohol o tabaco, establecer un horario de sueño constante y equilibrado a lo largo de la semana respetando los ciclos adecuadamente, aumentar la cantidad de ejercicio y actividad física realizado diariamente, o realizar un cambio a una dieta más balanceada y adecuada con el objetivo de bajar un posible exceso de peso presente en el paciente (B. Gil et al., 2005).

Otro tipo de tratamientos existentes son los quirúrgicos, aunque estos no se emplean frecuentemente salvo que el beneficio sea considerable o no haya otras alternativas debido a que son los más invasivos, complejos y requieren de un mayor tiempo de recuperación. Los tratamientos quirúrgicos son muy variados e incluyen operaciones bucales o maxilares; operaciones nasales como pueden ser la colocación de una tabique desviado o corrección de hipertrofia de cornetes; o incluso operaciones para reposicionar o retirar parte de los tejidos como la úvula, el paladar o la faringe (uvulopalatofaringoplastia) (Konrad, 2006; Malhotra & White, 2002; Won et al., 2008). Finalmente, debemos mencionar la existencia de otras alternativas como tratamientos con medicamentos y drogas o tratamientos mediante ultrasonidos. Aunque estos últimos no son empleados de manera frecuente debido a su bajo índice de éxito en comparación con el resto de tratamientos previamente mencionados (Konrad, 2006; Xia & Sawan, 2021).

En la Figura 1-3 se puede observar una herramienta empleada frecuentemente por profesionales de la salud para evaluar las alternativas y posibilidades quirúrgicas a la hora de preparar un tratamiento para subsanar el trastorno de AOS presente en un paciente. En dicha figura, se puede ver un complejo diagrama de flujo donde se recogen una serie de posibilidades tanto de operación como de tratamientos no invasivos disponibles en función de las variables de entrada como son la edad, los síntomas o las deformaciones presentes en el sujeto (Xia & Sawan, 2021).

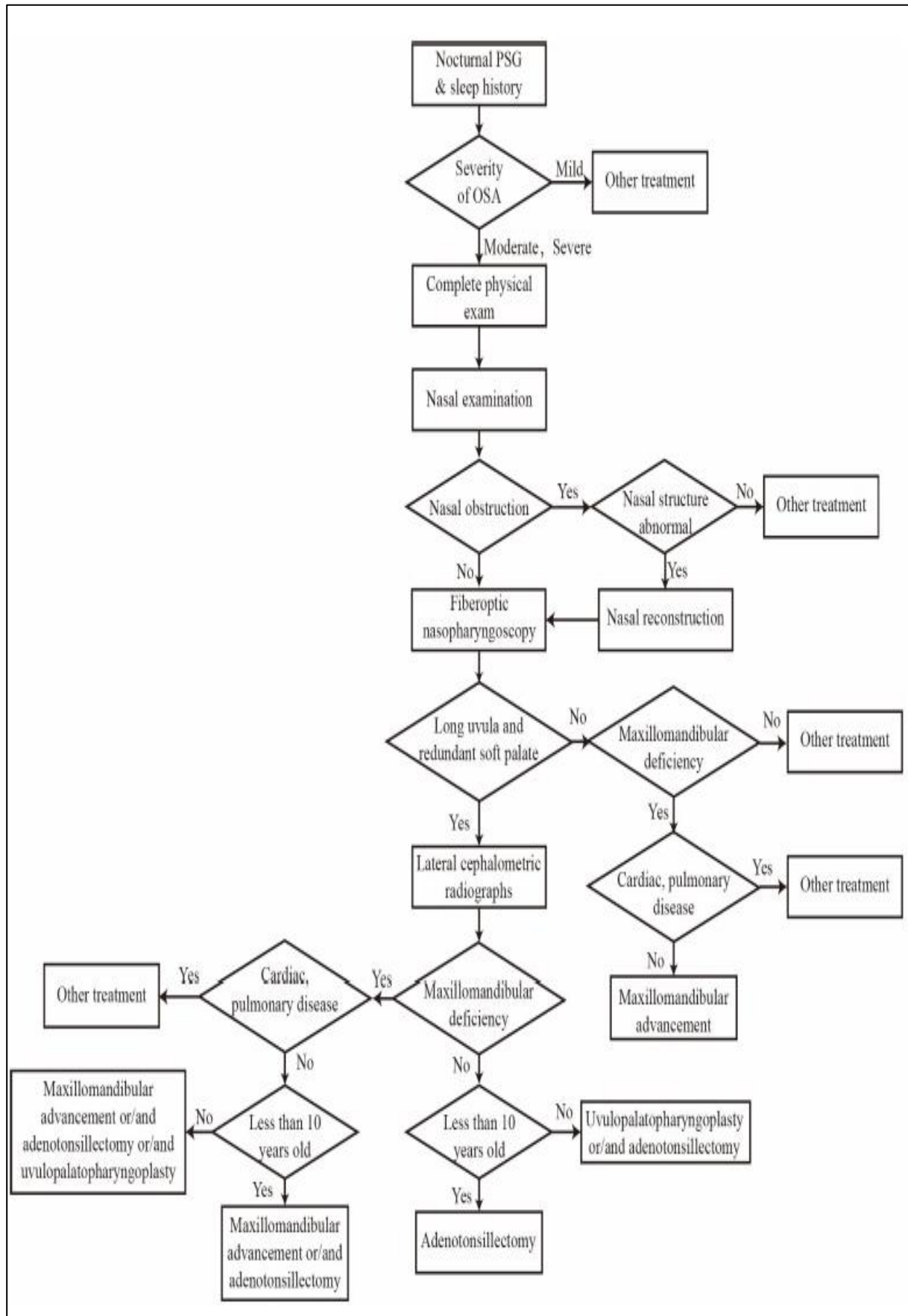


Figura 1-3. Diagrama de flujo representativo de posibles procedimientos quirúrgicos y tratamientos no invasivos para pacientes con AOS (Xia & Sawan, 2021).

1.4. Apnea Obstructiva del Sueño Pediátrica

El síndrome de la AOS no es exclusivo de adultos, si bien son el grupo más favorable a padecer dicho trastorno debido al envejecimiento natural del cuerpo humano, los niños también pueden sufrir este síndrome con consecuencias más graves (Bitners & Arens, 2020a). Consecuencias que afectan negativamente a su desarrollo tanto físico como mental, perjudicando por lo tanto su capacidad de aprendizaje (Bitners & Arens, 2020a). Por lo tanto, destinares este subapartado a resaltar las particularidades más relevantes del síndrome de la AOS en niños.

Debemos tener en cuenta que en la mayoría de los casos de AOS infantil, a diferencia de los adultos, la principal causa del problema son las amígdalas y las vegetaciones adenoideas. Debido a un tamaño o crecimiento inusual de estos tejidos se produce un bloqueo de forma parcial o completa del paso del aire a través del tracto respiratorio, produciendo los eventos conocidos como apnea o hipopnea durante la fase del sueño. Para poder evaluar y clasificar el tamaño de las amígdalas se emplea frecuentemente la puntuación de Brodsky. Este sistema empleado tanto en niños como en adultos, permite clasificar el grado de obstrucción de las vías respiratorias en función de la presencia excesiva de las amígdalas (Brodsky, 1989). La escala incluye los siguientes niveles:

- **Grado 0:** paciente completamente sano. Las amígdalas se encuentran en su posición y tienen un tamaño correcto, sin causar ningún problema al tracto respiratorio ni al flujo de aire. Este es el grado más común, presente en toda persona que no tenga ninguna patología respiratoria relacionada en la garganta.
- **Grado 1:** paciente con problema leve. Las amígdalas ocupan un 25% del espacio entre los pilares amigdalinos, es decir, las paredes de la faringe, limitando ligeramente el flujo de aire.
- **Grado 2:** paciente con problema medio. Las amígdalas ocupan entre un 25% y un 50% del espacio de la faringe, limitando más el flujo aéreo que el caso previo.
- **Grado 3:** paciente con problema severo. Las amígdalas ocupan entre un 50% y un 75% del espacio de la faringe, dificultando en gran medida el flujo aéreo.
- **Grado 4:** paciente crítico. Las amígdalas ocupan más del 75% del espacio de la faringe, siendo el grado más severo y llegando a los casos más extremos donde las amígdalas pueden hacer contacto entre sí imposibilitando la respiración del sujeto con normalidad.

La Figura 1-4 ilustra los diferentes grados definidos en la Escala de Brodsky mencionados anteriormente:



Figura 1-4. Puntuación Brodsky Grados 0-4. (Adaptado de Clínica Otorrinolaringológica Santa Catharina, 2020).

Al igual que ocurría con los adultos, no hay un único factor de riesgo que conlleve la aparición de este trastorno. Existen otros factores clave como malformaciones, deformidades u obesidad excesiva (Moffa et al., 2020). La Organización Mundial de la Salud (OMS) define el sobrepeso en niños en dos rangos de edad. El primer rango, en niños menores de 5 años se define como sobrepeso aquel peso dos veces superior a dos desviaciones típicas por encima de los patrones; y la obesidad aquel peso tres veces superior a dicha desviación típica. Para niños entre 5 y 19 años se define el sobrepeso como un peso superior a una desviación típica por encima de la mediana y la obesidad mórbida como un peso dos veces superior a dicha desviación típica (OMS, 2024). En las Figuras 1-5 y 1-6 se muestran las distintas gráficas empleadas por la OMS para medir el grado de obesidad a partir del índice de masa corporal (normalizado por edad y sexo) en niños y niñas entre 5 y 19 años, respectivamente:

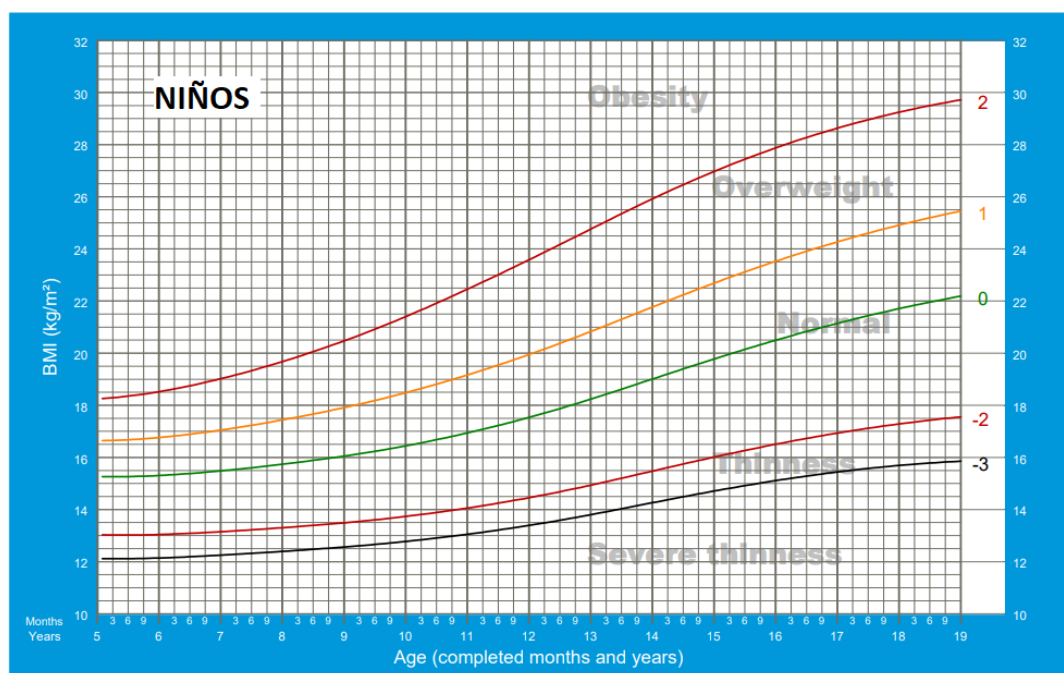


Figura 1-5. Tabla de cálculo IMC/Edad para niños entre 5 y 19 años, proporcionada por la OMS (OMS, 2007).

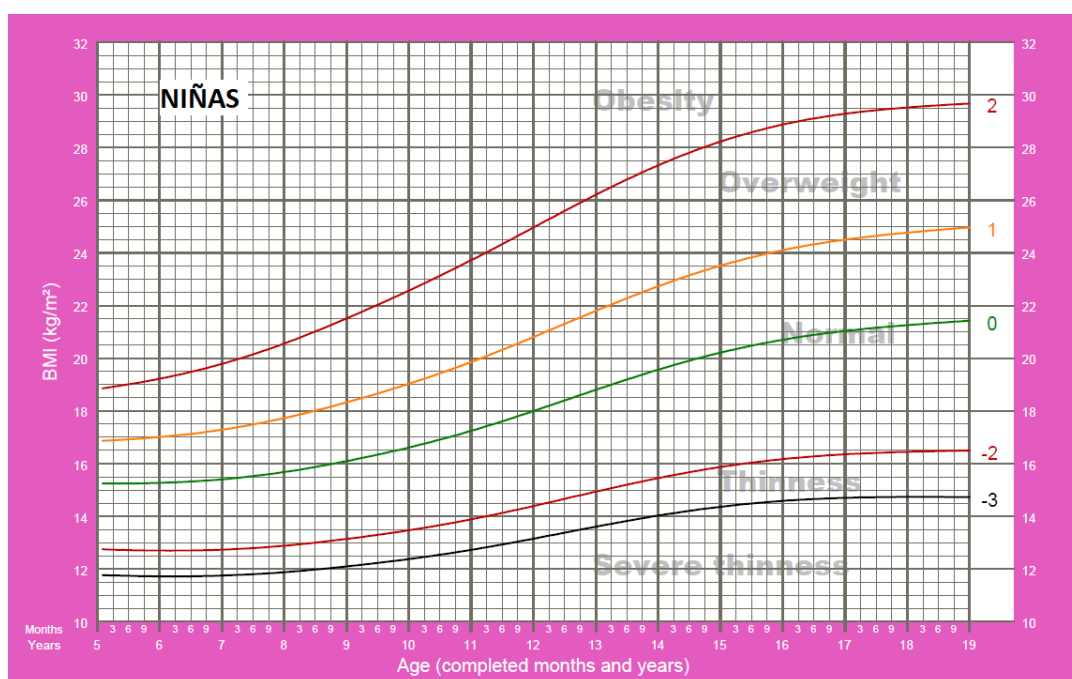


Figura 1-6. Tabla de cálculo IMC/Edad para niñas entre 5 y 19 años, proporcionada por la OMS (OMS, 2007).

Además, debemos destacar que los grados de AOS en niños no son clasificados de la misma forma, ni son diagnosticados de igual manera que en los adultos. Es cierto que se emplea el mismo índice, el IAH, pero los valores para diagnosticar y clasificar los distintos grados de AOS y la propia definición de la métrica son distintos (Bitners & Arens, 2020a; Moffa et al., 2020):

- **Sin AOS:** en el caso de que el $IAH < 1$ [eventos/hora], al igual que los adultos, el paciente no presenta síntomas de AOS.
- **AOS leve:** el caso de apnea menos peligroso, en este caso, el IAH está entre 1 y 5 e/h. Los umbrales del IAH empleados en niños son más restrictivos que en el caso de los adultos ya que se considera AOS leve un valor de IAH que para adultos sería un caso sin AOS (Rosen et al., 1992).
- **AOS moderada:** en este caso, el IAH es superior a 5 e/h pero inferior a 10 e/h, es decir, eventos más frecuentes que en la AOS leve. En estos casos, es necesaria atención inmediata para evitar daños mayores a largo plazo (Kaditis et al., 2016).
- **AOS crítica:** este es el caso más grave de todos, en este caso el $IAH \geq 10$ e/h, lo que implica eventos de manera muy frecuente, imposibilitando el descanso correctamente. Este grado al igual que el caso moderado, requiere de tratamiento inmediato para no desencadenar secuelas más graves u otros problemas de salud (Kaditis et al., 2016).

Otra de las grandes diferencias entre adultos y niños es la forma del diagnóstico. A pesar de seguir siendo la PSG el estándar más empleado para diagnosticar la AOS tanto en adultos como en niños, no se obtiene el mismo resultado en ambos grupos. Esto se debe en parte a que las señales registradas en niños a lo largo de la noche en el laboratorio del sueño no logran alcanzar la precisión deseada, lo cual es uno de los principales motivos por el que se buscan alternativas a este método de diagnóstico (Bitners & Arens, 2020a). Esto se debe a que es mucho más difícil que un niño en temprana edad, en un entorno no familiar como es una unidad del sueño y lejos de sus padres, mantenga la quietud y tranquilidad necesarias para conciliar el sueño. Por lo que las horas de sueño profundo se ven reducidas y los resultados pueden ser insatisfactorios (Alonso Álvarez et al., 2008). Por ejemplo, al estar nervioso no podrá dormir, aumentará su ritmo cardíaco, variará su EEG y probablemente no alcance apenas la fase de sueño profundo. Estos efectos pueden ser aún mayores cuanto más pequeño sea el paciente, llevando a la comunidad científica a investigar y desarrollar otra serie de sistemas alternativos a la PSG para diagnosticar clínicamente la AOS en niños (Bitners & Arens, 2020a).

Algunos de los métodos alternativos empleados para realizar un filtrado inicial de pacientes con sospechas de AOS son (Bitners & Arens, 2020a; Leske, 2014): la exploración física, realizada mediante la observación por parte de un especialista de la respiración por nariz y boca, el tamaño de los tejidos de la garganta (empleando la escala Brodsky mencionada anteriormente), calidad de la voz o ronquidos entre otros; la realización de cuestionarios sobre su día a día tanto al usuario como a sus padres o tutores legales, con el fin de identificar comportamientos que la propia persona no pueda notar al estar dormido, tales como ronquidos; el uso de versiones adaptadas de HSAT para niños, con el fin de que puedan realizar una prueba similar a la PSG desde su propia casa; uso de radiografías o Imagen por Resonancia Magnética (IRM) dinámicas para poder observar la estructura interna de la zona de la garganta; o endoscopias con o sin sueño inducido, observando directamente el estado de la garganta mediante pequeñas cámaras, siendo este último un método que no se utiliza apenas debido a que es muy invasivo (Bitners & Arens, 2020a; Leske, 2014).

Los tratamientos en niños también son diferentes con respecto a los tratamientos que se aplican a los adultos. En el caso de pacientes pediátricos el principal problema que desencadena el trastorno de la AOS es el crecimiento anormal de las amígdalas o del tejido adenoideo de la garganta. Por lo que, un procedimiento quirúrgico como puede ser la adenoamigdalectomía soluciona o reduce en gran medida el problema de obstrucción (Bitners & Arens, 2020a). Este procedimiento puede ser realizado mediante electrocirugía, ablación mediante técnicas de plasma o la más frecuente, eliminación del tejido o de la capsula de las amígdalas entera, con técnicas basadas en frío (Bitners & Arens, 2020a).

Debemos tener en cuenta que existen otros tratamientos diferentes como pueden ser (Bitners & Arens, 2020a; Konrad, 2006): alternativas quirúrgicas como la traqueotomía, destinadas a pacientes con malformaciones o anomalías estructurales; cirugías para el paladar, los laterales de la faringe o la nariz, destinadas a mejorar el flujo aéreo en el tracto respiratorio; otros procedimientos no invasivos como son la pérdida de peso corporal o el cambio de hábitos de vida; tratamientos bucodentales como colocación de aparatos para reestructurar ciertos componentes de la dentadura, la lengua o la mandíbula; o el empleo de medicamentos especiales (Bitners & Arens, 2020a; Konrad, 2006). Se demostró en un experimento con 464 niños de entre 5 a 9 años con AOS, que aquellos que eran sometidos a la adenoamigdalectomía presentaban una mejoría en la calidad de vida y respiración, frente a aquellos que se sometían a otros tratamientos. Concretamente el 79% de los niños sometidos a adenoamigdalectomía mejoraron frente a un 46% de pacientes que mejoraron al ser sometidos a otros tratamientos alternativos (Padilla et al., 2013).

1.5. Deep Learning

La humanidad ha experimentado cambios y evoluciones en las bases de la civilización desde sus remotos inicios, algunos de estos eventos incluso han llevado a los historiadores a definir “cambios de época” repartidos a lo largo de la historia. Estos cambios, generalmente están marcados por grandes descubrimientos o revoluciones, como por ejemplo las distintas revoluciones industriales que han ocurrido a lo largo de la historia (Gonzalez-Hernandez et al., 2021). La primera revolución industrial data de 1780 a 1840, época en la que surgieron nuevas herramientas como la máquina de vapor, nuevos combustibles como el carbón y tuvo lugar la evolución de ciertas industrias, como la textil. La segunda revolución industrial, también conocida como “La Revolución Tecnológica”, data de 1870 a 1914, y se caracteriza por la aparición de nuevas fuentes de energía como la electricidad y el petróleo, la aparición de nuevas industrias como la química o la eléctrica, y la instauración de la producción en cadena de productos en masa. La tercera revolución industrial, también conocida como “La Revolución Digital”, data de mediados del siglo XX, y se caracteriza por la implementación y mejora de ordenadores, robótica e internet en todos los aspectos de la vida cotidiana y la industria. Finalmente, la cuarta revolución industrial es conocida como “Industria 4.0” o “La Revolución de la Inteligencia Artificial (IA)”. Actualmente estamos experimentando los inicios de esta cuarta revolución; revolución que promete innovar en los campos de la IA, el *Big Data* y el Internet de las cosas (*Internet of Things*, IoT), permitiendo así la creación de industrias y ciudades inteligentes (Ynzunza Cortés et al., 2017). En este apartado, nos centraremos en esta última revolución, concretamente en la parte de la IA, ya que nos servirá como herramienta más adelante en el caso de estudio.

La IA puede definirse como la capacidad de los ordenadores para ejecutar, de la forma más similar posible, tareas que habitualmente son realizadas por seres humanos, así como aprender a partir de la experiencia. En este sentido, desarrollar IA implica crear algoritmos y sistemas capaces de emular con precisión el comportamiento o pensamiento humano (Rouhiainen, 2018). No obstante, el término IA engloba gran cantidad de técnicas y campos distintos. Entre los que podemos distinguir: Aprendizaje Profundo o *Deep Learning* (DL) y Aprendizaje Máquina, Aprendizaje Automático o *Machine Learning* (ML). El término ML hace referencia a la capacidad de los ordenadores de realizar una cierta tarea y adquirir experiencia resolviéndola sin ser programados explícitamente para ello; esto se consigue empleando una serie de algoritmos que les brindan la posibilidad de aprender relaciones entre estructuras de información. Por otro lado, el término DL es un subconjunto del ML que emplea estructuras con múltiples capas como son redes neuronales de distintos tipos con las que son capaces de extraer características más complejas de los datos de entrada y aprender de ellas, pudiendo elaborar conceptos mucho más complejos (Kelleher, 2019; Voulodimos et al., 2018a). En la Figura 1-7 podemos ver representados estos conceptos mencionados previamente en un diagrama de Venn:

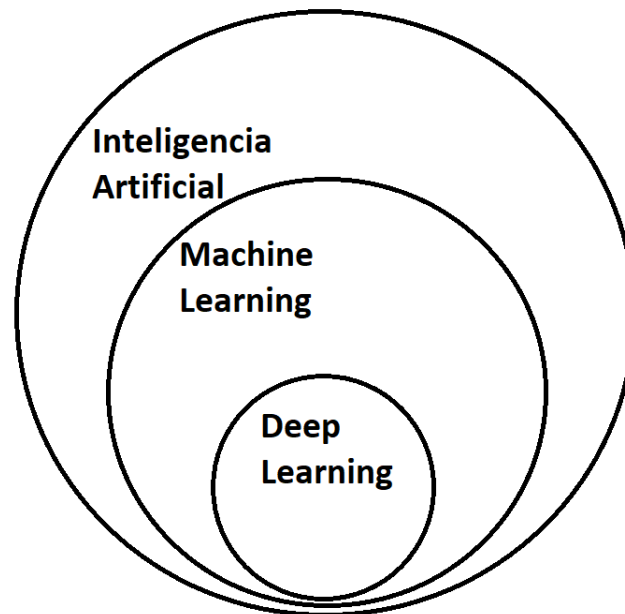


Figura 1-7. Diagrama de Venn de los conceptos: Inteligencia Artificial, Machine Learning y Deep Learning (Adaptada de Kelleher, 2019).

En el campo del análisis y procesado automático de señales existen una gran cantidad de estudios enfocados al uso de técnicas de ML y DL para procesar estas señales, de hecho más adelante en el capítulo de la revisión del estado de la técnica haremos hincapié en estudios que emplean este tipo de herramientas aplicadas al diagnóstico de la AOS. Conforme avanzan los estudios centrados en la comparativa entre el uso de métodos de ML y DL, crece la popularidad del uso del DL, ya que cada vez se está demostrando más su superioridad frente a los modelos de ML para este tipo de tareas (Gutiérrez-Tobal et al., 2022; İŞCAN, 2022; Shlezinger & Eldar, 2023). Por lo tanto y para el caso de estudio que nos concierne, nos centraremos en los algoritmos y técnicas de DL, debido a que trataremos de crear un sistema capaz de extraer las características de los datos por sí mismo de manera automática. Con dicho sistema, intentaremos reconocer patrones a detectar para poder clasificar distintos sujetos en los distintos grados de AOS a partir de una serie de señales biomédicas. Sin embargo, debemos de tener en cuenta que el nivel de precisión de los modelos de DL aún no ha alcanzado su punto óptimo y puede continuar mejorando en gran medida. Finalmente, y teniendo en cuenta todo lo mencionado anteriormente, no profundizaremos en los términos de IA y ML más allá de lo estrictamente necesario.

Lo primero que debemos tener en cuenta es que los sistemas de DL están basados generalmente en una serie de estructuras denominadas redes neuronales. Estas redes neuronales consisten en modelos de computación que tratan de emular la estructura del cerebro humano y están compuestas de una serie de capas donde se procesa la información, que a su vez están compuestas por unidades básicas de procesamiento conocidas como neuronas artificiales (Kelleher, 2019). Las neuronas del cerebro presentan 3 elementos diferentes: el cuerpo o núcleo celular, las dendritas y una fibra conocida como axón (Ostrosky, 2010). Tratando de recrear esa estructura en las redes neuronales artificiales, cada neurona artificial de la red se compone de un elemento central de procesamiento, que tratará de emular el cuerpo celular. Si dicha neurona, recibe un estímulo lo suficientemente fuerte a través de las dendritas (o en el caso de nuestra red neuronal, los canales de entrada), transmitirá un impulso a otras neuronas a través del axón (o canales de salida a otras neuronas artificiales en el caso de nuestra red). Debemos de tener en cuenta que el cerebro humano tiene alrededor de 86.000 millones de neuronas, por lo que haría falta una red neuronal de un tamaño descomunal para emular completamente su comportamiento (Kelleher, 2019; Ostrosky, 2010). Sin embargo, los avances tecnológicos actuales permiten realizar redes neuronales con una gran densidad de neuronas, pudiendo realizar así tareas realmente complejas. A continuación en la Figura 1-8 podemos ver un esquema de las similitudes entre una neurona real y una artificial:

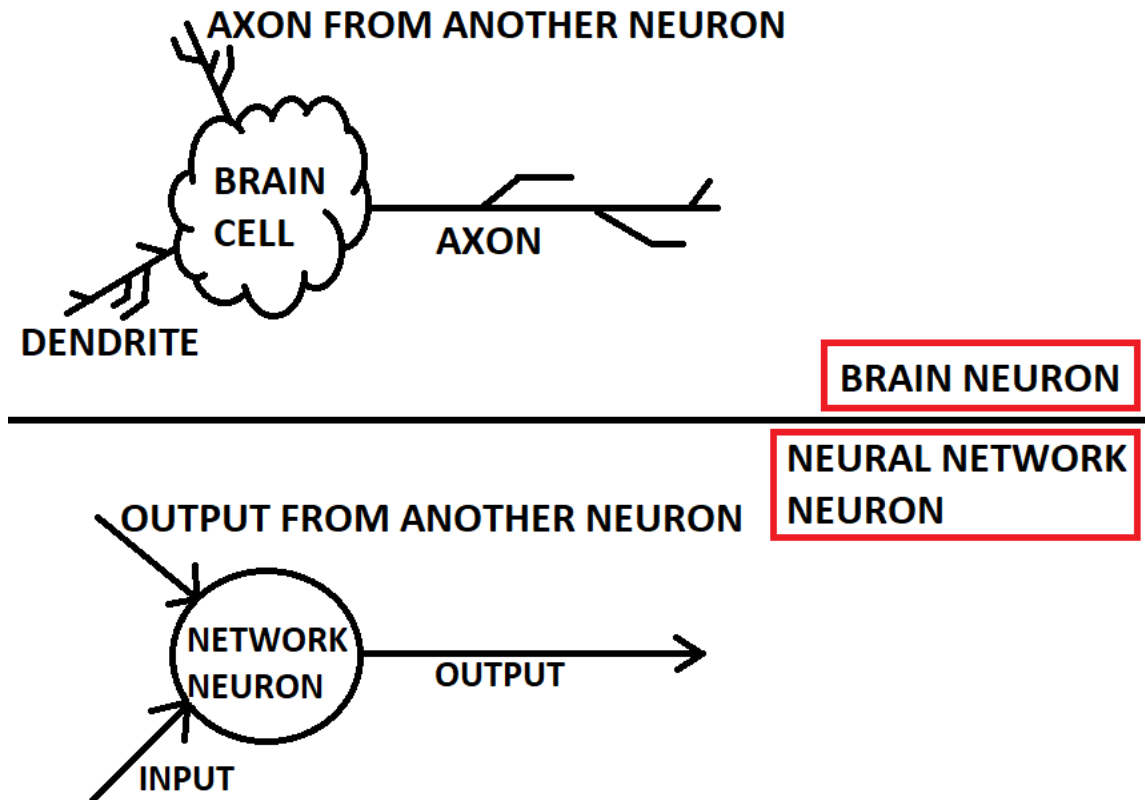


Figura 1-8. Similitudes entre una neurona del cerebro humano (superior) y una neurona artificial de una red neuronal (inferior) (Autoría propia, basada en Kelleher, 2019).

Como ya hemos mencionado, las redes neuronales representan las bases de los sistemas DL. Pero debemos tener en cuenta que existen varios tipos de redes neuronales que pueden ser más o menos efectivas en función de la tarea a la que son destinadas, algunos de estos ejemplos son (Kelleher, 2019):

- **Redes Neuronales Artificiales:** también conocidas como ANN, son el tipo de red neuronal más básico de todas, destinadas a problemas de clasificación o regresiones sencillas. Este tipo de modelos son muy útiles para desarrollar sistemas pequeños debido a su simplicidad.
- **Redes Neuronales Convolucionales:** también conocidas como CNN, destinadas a procesar datos en formato matricial como imágenes o señales, permiten reconocer patrones en los datos, generar conceptos complejos y clasificarlos gracias a aplicar la operación matemática de convolución. Son frecuentemente utilizadas en detección de objetos o patrones y en el reconocimiento y generación de imágenes.
- **Redes Neuronales Recurrentes:** también conocidas como RNN, destinadas a datos que tienen una componente temporal como sonidos o lenguaje hablado, este tipo de redes contienen conexiones de retroalimentación que las permiten recordar información. Son ampliamente utilizadas en Procesado de Lenguaje Natural (PLN), como traductores, generadores de texto a partir de imágenes o creadores de contenido escrito.
- **Redes Neuronales de Memoria a Largo y Corto plazo:** también LSTM (del inglés, *Long Short-Term Memory*). Representan una variante o subconjunto de las RNN, pero con una mayor capacidad gracias a su habilidad de poder recordar u olvidar conceptos, lo que las permite, mantener una memoria, de ciertas ideas o palabras a lo largo del tiempo.

Debemos tener en cuenta que existen muchas más variaciones y versiones de estas redes adaptadas a las necesidades de cada proyecto o estudio (Perumal et al., 2024).

1.6. Hipótesis y objetivos

En los apartados anteriores se ha mencionado que la PSG es una prueba costosa de realizar, por lo que se han buscado otras alternativas como el uso de técnicas de procesamiento automático de señales basadas en DL para poder realizar una estimación empleando una menor cantidad de recursos. Esto permitirá reducir el coste económico, computacional y temporal de realizar un diagnóstico de AOS.

La hipótesis planteada en este caso de estudio se basa en determinar si es posible estimar el IAH con la suficiente precisión como para permitir realizar un diagnóstico fiable de la AOS infantil, empleando para ello distintas combinaciones de un conjunto reducido de señales (como pueden ser SpO₂, FA y HRV) procesadas mediante modelos automáticos basados en DL.

Para ello, el objetivo general de este Trabajo Fin de Máster es desarrollar un modelo de DL diseñado específicamente para el procesamiento de cualquiera de las combinaciones de las señales SpO₂, FA y HRV extraídas de bases de datos especializadas en la AOS infantil. Posteriormente, se realizará un análisis detallado del rendimiento del sistema y se optimizarán los puntos indicados en función de los resultados del análisis. Finalmente, se evaluará la viabilidad clínica de cada una de las combinaciones en función de las necesidades de cada paciente, los recursos disponibles y las decisiones y criterio del equipo médico especializado. A continuación listamos los objetivos específicos a conseguir:

- Obtener una base de datos de registros nocturnos apropiada con información sociodemográfica y clínica de pacientes pediátricos para poder desarrollar nuestro modelo.
- Desarrollar modelos de DL para el análisis de señales de SpO₂, FA y HRV tanto de manera individual como de manera conjunta.
- Comparar y valorar resultados obtenidos y extraer las conclusiones pertinentes.

1.7. Estructura del Trabajo Fin de Máster

Durante el transcurso del proyecto, se ha recogido y registrado toda la información empleada y obtenida en esta memoria. Para poder estructurar correctamente todos los datos, y así facilitar el trabajo al lector, hemos optado por dividir el proyecto en una serie de capítulos temáticos. Este documento, por lo tanto, constará de 7 capítulos y una bibliografía, donde se ha ido narrando paso a paso todo el trabajo realizado, entre ellos: la revisión del estado de la técnica, obtención y preparación de los datos, desarrollo de un modelo, evaluación y comparativa de resultados, extracción de conclusiones y líneas futuras.

- **Capítulo 1. Introducción:** en este primer capítulo hemos realizado una breve introducción sobre el caso de estudio, mencionando las principales características de la ASO, su diagnóstico, síntomas y tratamiento tanto para adultos como para niños. Además, también hemos visto información general acerca de las técnicas de DL, la formulación de una hipótesis y sus respectivos objetivos y por último una detallada descripción del documento y su división en distintos apartados.
- **Capítulo 2. Estado de la técnica:** a lo largo de este capítulo, hemos analizado todo el trabajo previo realizado por expertos en este campo en concreto. Hemos tratado de extraer todos los objetivos y resultados logrados y a partir de ellos, desarrollamos nuestra hipótesis con la que complementar dichos estudios.
- **Capítulo 3. Señales y bases de datos:** en este tercer capítulo, hemos revisado las componentes de las 3 señales, sus características, las posibles herramientas con las que podemos obtenerlas y su viabilidad dentro de nuestro proyecto. Además, hemos analizado varias bases de datos como CHAT para elegir una candidata a ser empleada en nuestro sistema.

- **Capítulo 4. Metodología:** una vez elegidas las señales y la base de datos, hemos realizado una breve explicación de los conceptos referentes a la metodología empleada y comentado punto por punto el desarrollo del modelo, el generador empleado, el estimador y el análisis estadístico de los resultados.
- **Capítulo 5. Resultados:** con el modelo desarrollado, hemos analizado los resultados obtenidos tanto en lo referente a la optimización de los hiperparámetros como a los propios resultados generados al realizar la estimación del número de eventos por segmento y el grado de severidad de la AOS.
- **Capítulo 6. Discusión:** en este penúltimo capítulo se realizó una discusión de los resultados obtenidos con sus posibles aportaciones al campo de estudio y una comparativa con todos los estudios anteriores comentados en la revisión del estado de la técnica.
- **Capítulo 7. Conclusiones y líneas futuras:** a modo de broche final, se han expuesto las conclusiones finales extraídas del caso de estudio que ha supuesto este proyecto, junto a una recopilación de las limitaciones y líneas futuras de cara a mejorar o complementar este estudio en futuras investigaciones.
- **Bibliografía:** de manera adicional, se aporta la bibliografía consultada durante el desarrollo del proyecto y la elaboración de este documento.

Capítulo 2: ESTADO DE LA TÉCNICA

En este segundo capítulo realizaremos una revisión de los estudios existentes en este ámbito de trabajo. Para ello, citaremos los trabajos más relevantes en el campo del diagnóstico de la AOS pediátrica junto a los resultados y conclusiones derivados de los mismos que puedan resultar de interés para elegir la dirección en la que enfocar el desarrollo de nuestro propio proyecto. A lo largo del capítulo se incluirán referencias de todos los documentos e investigaciones citados para poder realizar un mejor seguimiento por parte del lector.

Comenzaremos con una breve contextualización, seguida de un análisis de una serie de tablas localizadas al final del capítulo donde se resumen todos los trabajos de investigación analizados para revisar el estado de la técnica en este campo de estudio. A continuación, trataremos con un poco más de detalle algunos puntos de interés como pueden ser: la metodología elegida, las señales empleadas o las predicciones realizadas. Para finalmente poder extraer una conclusión sobre la que encaminaremos nuestro propio trabajo.

2.1. Contexto del caso de estudio

Como ya se hemos podido observar la AOS es una patología prevalente, de hecho una gran parte de la población la padece y solo una porción reducida ha sido diagnosticada. Parte de este infradiagnóstico se debe al elevado coste y complejidad de las pruebas basadas en PSG. Por ello, gran parte de la comunidad científica se ha volcado en la búsqueda de alternativas más sencillas a este sistema de diagnosis (Mazzotti et al., 2018).

Para poder simplificar el proceso de diagnóstico, varios autores han optado por tratar de realizar una estimación del grado de AOS a partir de una cantidad de señales reducida; señales que pueden obtenerse mediante una versión de PSG más sencilla, o mediante otras alternativas menos costosas. Gracias a esta revisión podremos disponer de una visión más general del estado de gran parte de las investigaciones y los avances conseguidos, y por lo tanto, podremos identificar zonas que aún no han sido exploradas donde poder realizar nuestra aportación.

En las Tablas 2-1, 2-2, 2-3, 2-4 y 2-5 se han agrupado diferentes estudios referentes a la aplicación de técnicas de ML/DL con las que poder realizar una estimación de la presencia de la AOS en niños. De esta manera, hemos agrupado los datos de cada investigación en función de: los detalles sobre las bases de datos, las señales elegidas, la metodología empleada y los resultados de clasificación obtenidos de sensibilidad (*sensitivity*, Se), exactitud (*accuracy*, Acc) y especificidad (*specificity*, Sp) entre otras cosas. Además, se ha agregado una indicación sobre si el estudio trató de estimar los grados de AOS mediante el IAH o solamente la presencia o no de AOS (añadiendo a mayores el umbral empleado en este caso).

Los estudios analizados datan de entre 2003 y 2024 y han sido agrupados en 7 grupos distintos en función de las señales empleadas para realizar la predicción de la OSA. El grupo de estudios que emplean solamente señales de SpO₂ (Álvarez et al., 2017; L. Chang et al., 2013; Hornero et al., 2017; Jiménez-García et al., 2020; Kirk et al., 2003; Mortazavi et al., 2024; Tsai et al., 2013; Vaquerizo-Villar et al., 2018a, 2018b, 2021; Villa et al., 2015; Xu et al., 2019); los que emplean la señal de FA únicamente (Barroso-García et al., 2017; Barroso-García, Gutiérrez-Tobal, Gozal, et al., 2021; Barroso-García, Gutiérrez-Tobal, Kheirandish-Gozal, et al., 2021; Jiménez-García et al., 2020); los que emplean solo el HRV, ECG o derivadas (Dehkordi et al., 2016; García-Vicente et al., 2023; E. Gil et al., 2009, 2010; Lazaro et al., 2014; Martín-Montero et al., 2021, 2022, 2023; Shouldice et al., 2004; Ye et al., 2023); los grupos que emplean dos de las señales citadas previamente combinadas, como puede ser SpO₂-FA (Barroso-García et al., 2020; Gutiérrez-Tobal et al., 2015; Jiménez-García et al., 2020, 2022, 2024), SpO₂-HRV (Garde et al., 2014, 2019; Ye et al., 2023), o HRV - FA (Sturludóttir et al., 2023); y finalmente, los estudios que emplean las 3 señales de manera simultánea (Tan et al., 2014).

2.2. Señales elegidas

Hay una gran cantidad de estudios que emplean las señales SpO₂, SpO₂-FA y HRV, pero no es tan frecuente encontrar investigaciones basadas en el resto de combinaciones entre ellas. Como se puede observar en las tablas (Tablas 2-1, 2-2, 2-3, 2-4 y 2-5), una gran cantidad de trabajos se enfocan en el uso de la señal SpO₂ de forma aislada o combinada con la señal FA (Álvarez et al., 2017; Barroso-García et al., 2017, 2020; Barroso-García, Gutiérrez-Tobal, Gozal, et al., 2021; Barroso-García, Gutiérrez-Tobal, Kheirandish-Gozal, et al., 2021; L. Chang et al., 2013; Gozal et al., 2018; Gutiérrez-Tobal et al., 2015; Hornero et al., 2017; Jiménez-García et al., 2020, 2022, 2024; Kirk et al., 2003; Mortazavi et al., 2024; Tsai et al., 2013; Vaquerizo-Villar et al., 2018b, 2021; Villa et al., 2015; Xu et al., 2019). Esto se debe a que el FA y la SpO₂ son las señales que se encuentran más estrechamente ligadas a la descripción de los eventos de apnea e hipopnea, según los estándares de la AASM (Marcus et al., 2012). La desaturación de oxígeno en sangre y la reducción del flujo aéreo son dos de los hallazgos más importantes en las PSG de pacientes que padecen este trastorno y por lo tanto candidatos ideales para intentar detectar la presencia o no de la AOS (Olivi, 2013). Por otro lado, señales cardíacas como el HRV o el PRV son relativamente fáciles de obtener mediante dispositivos no invasivos, como el ECG o el pulsioxímetro respectivamente (Martín Montero, 2024). Estas señales pueden aportar información relevante para caracterizar la respuesta del sistema cardiovascular y del sistema nervioso frente a los eventos apneicos. En particular, diversos estudios han demostrado que la PRV, extraída a partir de la señal de fotopletimografía (*photoplethysmography*, PPG) registrada por pulsioxímetros, refleja patrones similares a los observados en la HRV (Martín Montero, 2024). Además, su análisis ha mostrado su utilidad como complemento a la oximetría convencional para mejorar la detección de la apnea del sueño en entornos domiciliarios y su fácil extracción a partir del ECG, una señal sencilla de obtener e interpretar en comparación con otras señales más complicadas de registrar y analizar cómo puede ser el EEG (D’Rozario et al., 2017). Por otro lado, el PRV también puede ser obtenido a partir del pulsioxímetro de forma rápida y simple (Gutiérrez-Tobal et al., 2019). Aunque estas señales por sí solas pueden no ser suficiente para estimar el grado de AOS con una precisión suficiente, cabe la posibilidad de que permitan aportar más información a la hora de realizar la estimación, a cambio de aumentar ligeramente la carga computacional necesaria. Es por esto por lo que algunos autores han optado por investigar la viabilidad de emplear las señales HRV y PRV en sus investigaciones (García-Vicente et al., 2023; E. Gil et al., 2010; Martín-Montero et al., 2021, 2022, 2023; Shouldice et al., 2004; Sturludóttir et al., 2023; Tan et al., 2014; Ye et al., 2023).

Por lo tanto, debido a que existen varios estudios que han evaluado la capacidad diagnóstica de alguna de estas tres señales, pero apenas existen estudios que empleen estas señales de manera combinada o conjunta, resulta de interés centrar nuestra línea de investigación en este ámbito. Por ello, el enfoque principal de este trabajo consistirá en analizar los resultados obtenidos en la predicción de la AOS, empleando distintas combinaciones de señales, con el objetivo de valorar si supone una mejora o no el hecho de añadir algunas señales a los estudios ya realizados o probar combinaciones de ellas que no hayan sido previamente empleadas por otros autores.

2.3. Metodología empleada

Podemos observar la gran diversidad de metodologías empleadas, desde la comparativa directa como es el caso de Tan et al. (2014) hasta el uso de modelos como DL, ML u otro tipo de algoritmos. En los trabajos de Jiménez-García et al. (2020) o Martín Montero et al. (2023) entre otros, emplean alguna de las versiones del algoritmo AdaBoost (*Adaptive Boosting*). En todos los casos, AdaBoost se utiliza como clasificador supervisado, aprovechando su capacidad para combinar múltiples clasificadores débiles, como árboles de decisión simples, y generar un modelo robusto que mejore progresivamente su rendimiento con cada iteración (Barroso-García et al., 2020; Jiménez-García et al., 2020; Martín-Montero et al., 2023). Por otra parte, en otros estudios como los de Lázaro et al. (2014), Gil et al. (2009) y Garde

et al. (2014), se destaca el uso de otro algoritmo como es el LDA (*Linear Discriminant Analysis*) para generar una proyección lineal capaz de separar de la mejor manera las distintas clases involucradas. En estos casos se puede ver cómo este modelo ha sido empleado para separar solamente aquellos casos de pacientes con valores de IAH superior a 5 e/h de aquellos que tienen un valor inferior. Debemos tener en cuenta que los datos deben ser linealmente separables para poder realizar la clasificación (Garde et al., 2014; E. Gil et al., 2009; Lazaro et al., 2014).

También se han empleado algunas aproximaciones basadas en ML como es el caso de los textos de Gutiérrez-Tobal et al. (2015), Martín-Montero et al. (2022), Ye et al. (2023) o Barroso García et al. (2017) entre muchos otros. En ellos se opta por un enfoque clásico de ML que se divide en dos fases: extracción y selección de características, y reconocimiento de patrones (clasificación/regresión). Estos métodos de ML pueden ser: la Regresión Logística (*Logistic Regression*, LR), los Perceptrones Multicapa de poca profundidad (*Multilayer Perceptron*, MLP) o variantes de estos modelos (Barroso-García et al., 2017, 2020; Barroso-García, Gutiérrez-Tobal, Gozal, et al., 2021; Barroso-García, Gutiérrez-Tobal, Kheirandish-Gozal, et al., 2021; L. Chang et al., 2013; Garde et al., 2019; Gutiérrez-Tobal et al., 2015; Hornero et al., 2017; Martín-Montero et al., 2021, 2022; Vaquerizo-Villar et al., 2018b, 2018a; Villa et al., 2015; Xu et al., 2019; Ye et al., 2023). Aunque este tipo de modelos, requieren de una menor cantidad de datos y son mucho más simples y económicos de implementar que modelos de DL más complejos, tienen una capacidad limitada a la hora de establecer relaciones profundas en imágenes o señales. Por ende, las LR y MLP son empleadas en sistemas donde se requiera una clasificación binaria sencilla, siempre y cuando se dispongan de datos tabulados; y estructuras más complejas como las CNN, en análisis espacial como pueden ser en este caso imágenes para el diagnóstico o señales biomédicas (Assi et al., 2018; Gutiérrez-Tobal et al., 2022; Kelleher, 2019).

En el ámbito del DL pueden distinguirse dos enfoques principales en función de la arquitectura neuronal empleada. Por un lado, algunos estudios han optado por utilizar redes neuronales del tipo MLP profundas como modelo base, permitiendo así disponer de una red neuronal completamente conectada que permita emplear un menor número de capas que otro tipo de redes (Barroso-García et al., 2020; Barroso-García, Gutiérrez-Tobal, Kheirandish-Gozal, et al., 2021; Hornero et al., 2017; Martín-Montero et al., 2022; Vaquerizo-Villar et al., 2018b). Por otro lado, una parte significativa de la literatura que también aplica técnicas de DL ha preferido emplear arquitecturas basadas en redes neuronales convolucionales (CNN) (García-Vicente et al., 2023; Jiménez-García et al., 2022, 2024; Mortazavi et al., 2024; Sturludóttir et al., 2023; Vaquerizo-Villar et al., 2021). En estos últimos trabajos, las CNN desarrolladas se han implementado tanto de forma aislada, como en combinación con otras arquitecturas o variantes derivadas. Entre estas combinaciones destacan enfoques como CNN + RNN, CNN-BiGRU, redes tipo ResNet, o el uso de CNN en dos dimensiones (2D CNN), con el objetivo de aprovechar mejor la estructura espacial y secuencial de los datos fisiológicos procesados (Jiménez-García et al., 2022, 2024; Mortazavi et al., 2024).

La estructura de la CNN elegida es diferente en cada estudio, aunque generalmente la mayoría ha optado por emplear una estructura de red multicapa de entre 16 y 64 filtros. Es importante destacar que el número de capas no puede ser excesivamente grande para evitar que la red requiera mucho tiempo de entrenamiento o se produzca sobreajuste; pero por otro lado, tampoco puede ser excesivamente pequeño ya que la red podría no aprender las relaciones entre los datos y por lo tanto no ser capaz de realizar estimaciones adecuadamente o generar resultados erróneos (García-Vicente et al., 2023; Mortazavi et al., 2024). La elección de los diferentes valores para los hiperparámetros tales como: el número de capas o bloques, el número de filtros por capa, las tasas de aprendizaje y *dropout* del modelo o el tamaño de cada uno de los filtros entre otros, dependerá del contexto del estudio y quedará a elección del propio autor; ya que dichos valores pueden cambiar en función de: la cantidad de datos disponibles, el objetivo del proyecto, los recursos computacionales disponibles y la complejidad del problema a abordar durante el trabajo.

Estos hiperparámetros mencionados previamente pueden suponer una gran diferencia en el desempeño del modelo. Sin embargo, no todos tienen el mismo impacto, ya que mientras que algunos pueden variar los resultados enormemente, otros simplemente permiten incrementar unas décimas la precisión del modelo. Pero casi todos los trabajos evaluados emplean la estructura de bloques convolucionales y capas para construir su modelo. Algunas de las capas empleadas para conformar los bloques convolucionales son: capas de convolución de 1 o 2 dimensiones, para poder extraer las relaciones entre los datos; capas de *batch normalization*, con las que poder normalizar los datos y así aumentar la estabilidad y rapidez del entrenamiento del modelo; capas de reducción de dimensiones como las capas *max pooling*, que permiten condensar la información contenida en los datos que se usan para alimentar el modelo pero sin afectar a las características; capas *dropout*, empleadas para desactivar ciertas neuronas aleatoriamente a lo largo de la red durante la fase de entrenamiento para prevenir el sobreajuste u *overfitting*; y finalmente capas *flattenning* y *fully-connected* situadas justo antes de la salida del modelo para generar la decisión final a partir de las relaciones establecidas a lo largo de toda la red (Kelleher, 2019; O'Shea & Nash, 2015). Si bien la composición y cantidad de estos mencionados bloques convolucionales depende de cada estudio, por lo general existe una estrategia predominante en la mayoría de los trabajos a la hora de estructurar cada bloque. Dicha estrategia consistente en la generación de los bloques a partir de una o dos capas de convolución, seguidas de capas de normalización y finalmente capas de reducción de dimensiones (García-Vicente et al., 2023; Jiménez-García et al., 2022, 2024; Mortazavi et al., 2024; Sturludóttir et al., 2023; Vaquerizo-Villar et al., 2021).

Los estudios de Jiménez-García et al. (2022) o Mortazavi et al. (2024) se enfocaron en maximizar la eficiencia durante la fase de entrenamiento, empleando algoritmos que permitan optimizar la etapa de descenso de gradiente, como es el caso del algoritmo “Adam” (*Adaptive Moment Estimation*) (Jiménez-García et al., 2022; Mortazavi et al., 2024). Este algoritmo es una combinación de otros más sencillos, como lo son el algoritmo “Momentum” y el algoritmo “RMSprop” (Kingma & Ba, 2014). Todos ellos son ampliamente empleados para optimizar el descenso de gradiente, permitiendo un movimiento lo más reducido posible en el eje perpendicular al mínimo de la función; a la vez, que maximiza el movimiento en la dirección del mínimo (Kingma & Ba, 2014). El algoritmo de optimización “Adam”, permite un ajuste dinámico de ciertos hiperparámetros con los que modificar la convergencia y así situarse en una zona cercana al mínimo global de forma más rápida y eficiente; permitiendo así realizar los ciclos de entrenamiento de forma más rápida (Ruder, 2016).

En cuanto a las funciones de activación empleadas, la función ReLU (o *Rectified Linear Unit*) es la que más frecuentemente es utilizada en los modelos basados en redes neuronales, debido a las ventajas que aporta con el descenso de gradiente. No obstante, esta función no es tan empleada en las capas finales de la red, en las cuales tal y como han elegido autores como García-Vicente et al. (2023), Jiménez García et al. (2022) o Mortazavi et al. (2024) entre otros, se emplean funciones como *softmax* en clasificación multiclase o funciones sigmoide en caso de ser una clasificación binaria (García-Vicente et al., 2023; Jiménez-García et al., 2022, 2024; Mortazavi et al., 2024; Sturludóttir et al., 2023; Vaquerizo-Villar et al., 2021).

Además, estudios como el de Mortazavi et al. (2024) o el de Vaquerizo-Villar et al. (2021) emplean otras técnicas de prevención del sobreajuste durante la etapa de entrenamiento como pueden ser el *Early Stopping*, o el *Batch Shuffling*; evitando así que la red se ciña en exceso a los datos proporcionados (Mortazavi et al., 2024; Vaquerizo-Villar et al., 2021). Otra forma ampliamente empleada a lo largo de los distintos estudios para evitar el sobreajuste, consiste en emplear las ya mencionadas capas de *dropout*; estas capas permiten la desactivación de cierto número de neuronas dentro de la red, evitando así que la red profundice demasiado en los datos durante las fases de entrenamiento (García-Vicente et al., 2023; Jiménez-García et al., 2022).

Resulta de interés destacar las variantes de CNN empleadas y citadas previamente; por ejemplo, Mortazavi et al. (2024), emplea variantes de CNN o RNN a lo largo de su investigación. En este estudio, se comparó el rendimiento de dos redes diferentes, una basada en ResNet y otra en BiGRU (Mortazavi et

al., 2024). Las ResNet (*Residual Networks*), consisten en una tipo de CNN que permite crear redes neuronales profundas, evitando o minimizando la posibilidad de que el gradiente se dispare o se desvanezca (es decir, se haga demasiado grande o pequeño) durante la fase de entrenamiento. Para ello emplea el concepto de bloques residuales, estructuras que permiten crear caminos adicionales alternativos al camino principal denominados atajos. La red neuronal puede emplear dichos atajos para acceder directamente a capas mucho más profundas, permitiendo así la posibilidad de crear redes mucho más complejas sin tener problemas con los gradientes (He et al., 2015a). Por otro lado, también emplea la combinación CNN-BiGRU (*Bidirectional Gated Recurrent Unit Network*). Este modelo de redes denominadas BiGRU, son a su vez una variante de redes RNN que permite establecer relaciones entre datos teniendo en cuenta la serie temporal. Para ello, se emplean compuertas (o *gates*) para controlar el flujo de información, permitiendo así seleccionar la información que se transmite hacia delante (recordándola) y la que es descartada (olvidándola) (She & Jia, 2021). Esta combinación permite al autor emplear una CNN para la extracción de las características y la red BiGRU para establecer las relaciones entre dichas características. Podemos observar como el modelo combinado CNN + BiGRU superó al ResNet (Mortazavi et al., 2024). Otros autores como Jiménez García et al. (2024), también emplean una combinaciones de CNN+RNN y añade a mayores un modelo de IA explicativa (*Explainable Artificial Intelligence*, XAI) como es Grad-CAM (*Gradient-weighted Class Activation Map*) (Jiménez-García et al., 2024). Este tipo de técnicas permiten mostrar y visualizar fácilmente sobre los datos aquellas zonas o características en las que se ha centrado el algoritmo para realizar la predicción y dar un resultado. En el caso de imágenes para el diagnóstico o señales, frecuentemente se resalta mediante un mapa de color, que va variando de intensidad conforme el algoritmo le haya dado mayor o menor importancia (Selvaraju et al., 2017).

Tras este estudio, queda clara la viabilidad de los modelos de DL basados en CNN o variantes del mismo en sistemas de detección de objetos o patrones en imágenes y señales. Es por ello, por lo que se ha optado por elegir las redes CNN como base inicial para el desarrollo de nuestro propio sistema. Además, la elección de este tipo de redes permitirá realizar una comparación más precisa y directa con los modelos desarrollados por otros autores que han optado por aportar nuevas investigaciones centrándose en un enfoque de DL.

2.4. Estudios de estimación de la severidad

Es importante destacar que no todos los sistemas desarrollados han tratado de estimar los mismos parámetros, pueden apreciarse dos grandes ramas que han seguido los diferentes autores de las investigaciones. En primer lugar, tenemos aquellos que han tratado de determinar simplemente la existencia o no de la AOS, sin importar el grado de dicha patología dado por el IAH. En estos casos, se ha dado un valor umbral para el cual se decide si hay o no presencia de AOS en el paciente, en función del valor estimado por el propio modelo. Generalmente el valor umbral que ha sido elegido es de 5 e/h, aunque algunos autores como Gutiérrez-Tobal et al. (2015) o Chang L et al. (2013) han optado por un valor umbral de 3 e/h. El valor de dicho umbral es por tanto elección del autor que desarrolla el estudio (Álvarez et al., 2017; L. Chang et al., 2013; Dehkordi et al., 2016; Garde et al., 2014; E. Gil et al., 2009, 2010; Gutiérrez-Tobal et al., 2015; Hornero et al., 2017; Kirk et al., 2003; Lazaro et al., 2014; Mortazavi et al., 2024; Shouldice et al., 2004; Sturludóttir et al., 2023; Tsai et al., 2013; Vaquerizo-Villar et al., 2018b, 2018a; Villa et al., 2015).

Por otro lado, los estudios restantes han tratado de estimar el valor del IAH, clasificando por lo tanto la AOS en leve, moderada o crítica, fijando los umbrales en 1, 5 y 10 e/h. Para ello han obtenido 3 resultados, relacionados con cada uno de los grados de severidad de la AOS respectivamente (Barroso-García et al., 2017, 2020; Barroso-García, Gutiérrez-Tobal, Gozal, et al., 2021; Barroso-García, Gutiérrez-Tobal, Kheirandish-Gozal, et al., 2021; García-Vicente et al., 2023; Jiménez-García et al., 2020, 2022, 2024; Martín-Montero et al., 2021, 2022, 2023; Tan et al., 2014; Vaquerizo-Villar et al., 2021; Xu et al., 2019; Ye et al., 2023).

Podemos observar que pese a existir esta diferencia entre investigaciones, todos han optado por representar los resultados principales en términos de sensibilidad, especificidad y exactitud (*Se Sp* y *Acc*), independientemente de si se ha tratado de estimar el grado de severidad o solo la existencia o no de la AOS.

En nuestro caso, y debido a que se considera que puede resultar de mayor utilidad, se tratará de seguir el segundo camino; desarrollando un sistema enfocado en la predicción del valor de la severidad de la AOS en función del valor de IAH en los pacientes.

2.5. Aportaciones realizadas

Tras analizar los estudios previos en este campo de investigación, resulta fundamental evaluar aquellos aspectos en los que podamos aportar un enfoque novedoso. Para ello, trataremos de explorar nuevas metodologías, combinaciones con las señales o emplear otro tipo de datos que permitan ampliar el conocimiento en esta disciplina.

En primer lugar, nuestro enfoque se basa en el uso de una CNN como modelo principal junto a una regresión lineal de Huber. Esto representa una innovación significativa en algunas de las señales, ya que, hasta la fecha, en el caso del FA los estudios previos se han centrado en la aplicación de otro tipo de metodologías como AdaBoost, LR o MLP. Por ende, nuestro proyecto aportará una nueva perspectiva en el estudio de las señales de FA empleando metodologías basadas en CNN. En el caso de la señal HRV, el concepto es prácticamente idéntico, si bien las señales HRV han sido previamente analizadas con enfoques como LR, MLP, LSBoost o QDA, aquellos estudios que han empleado CNN han optado por usar la señal ECG que es ligeramente distinta a la señal HRV. Dado que la señal ECG y la HRV son distintas, nuestro enfoque basado en CNN y HRV representa una contribución novedosa en el análisis de esta señal. Finalmente para el caso de la señal SpO₂ sí que existen estudios empleando la misma metodología, por lo que nuestro enfoque aportará un nuevo ejemplo de estructura de CNN y por lo tanto se reforzarán los resultados con este tipo de señal.

Por otro lado, en el caso de la combinación de las señales por parejas el hecho de realizar la combinación de FA con HRV, ya representa un avance claro nunca antes realizado en ningún estudio, ya que los únicos estudios referentes a este tema han empleado PRV o ECG en lugar de HRV. Es por ello, que emplear CNN con esas combinaciones de señales representa una nueva aportación en materia al campo de estudio. Por otro lado con el caso de SpO₂ con HRV, se ha realizado un estudio pero no se ha empleado CNN como metodología, por lo que nuestro estudio supondría también un avance novedoso.

Finalmente debemos destacar que en el caso del análisis conjunto de las 3 señales, ocurre algo similar, los estudios previos realizados emplean ECG en lugar de HRV, además de emplear como metodología la comparación directa. Debido a este elemento diferenciador, en nuestro caso, también estaríamos añadiendo una interpretación completamente diferente, al combinar las 3 señales sustituyendo el ECG por el HRV y emplear una CNN como sistema de análisis en lugar de la comparación directa.

En conclusión, la comparación de las señales SpO₂, FA y HRV tanto de manera individual como en combinaciones por parejas o en conjunto, representa un enfoque completamente innovador en la literatura científica. Al aplicar CNN para comparar directamente los resultados de analizar todas las señales y sus combinaciones aportamos un nuevo sistema con el que mejorar la detección y el diagnóstico de la AOS infantil.

AUTORES	SEÑALES	BASE DE DATOS	SUJETOS	METODOLOGÍA	ESTIMACIÓN	RESULTADOS
Vaquerizo Villar et al. (2018)	SpO ₂	Chicago University	981 entre 2 y 10 años	FCBF + SVM & MLP & LR	IAH (Umbral 5 e/h)	Acc: 83.7±4.9% (LR), 84.0±5.2%(SVM), 83.2±5.2% (MLP) Se: 72.6±4.7% (LR), 71.9±4.4%(SVM), 73.3±6.6% (MLP) Sp: 90.2±6.2% (LR), 91.1±7.2%(SVM), 89.0±6.9% (MLP)
Kirk et al. (2003)	SpO ₂	Alberta Chlidren's Hospital (Priv.)	58 entre 4 y 18 años	SnoreSat	IAH (Umbral 5 e/h)	Acc: 64.0% Se: 67.0% Sp: 60.0%
Tsai et al. (2013)	SpO ₂	Chang Gung Hospital Taiwan (Priv.)	148 entre 3 y 12 años	Correlación Pearson y Análisis curvas ROC	IAH (Umbral 1 e/h)	Acc: 85.1% Se: 83.8% Sp: 86.5%
Chang et al. (2013)	SpO ₂	Base privada	141 niños entre 2 y 13 años	Binary LR	IAH (Umbral 3 e/h)	Acc: 72.0% Se: 60.0% Sp: 86.0%
Villa et al. (2015)	SpO ₂	University of Rome (Priv.)	268 niños entre 0.9 y 17.2 años	Univariate LR	IAH (Umbral 5 e/h)	Acc: 69.4% Se (≤ 5 e/h > 5 e/h): 96.9% 36.0% Sp (≤ 5 e/h > 5 e/h): 39.2% 97.4%
Jiménez García et al. (2020)	SpO ₂	University of Chicago	974 niños entre 0 y 13 años	FCBF + AdaBoost.M2	IAH (e/h) y grado de AOS	Acc: 78.21% / 77.44% / 85.90% Se: 91.43% / 66.44% / 40.96% Sp: 22.67% / 84.02% / 98.05%
Vaquerizo et al. (2021)	SpO ₂	CHAT + University of Chicago + Hospital of Burgos (Priv.)	3196 niños entre 0 y 18 años	CNN	IAH (e/h) y grado de AOS	Acc: CHAT: 77.6% / 97.4% / 97.8% UofC&B: 80.1% / 83.9% / 92.3% Se: CHAT: 71.2% / 83.7% / 83.9% UofC&B: 90.8% / 76.0% / 79.5% Sp: CHAT: 81.8% / 100.0% / 99.3% UofC&B: 36.4% / 88.6% / 95.8%

Tabla 2-1. Tabla resumen revisión bibliográfica con la señal SpO₂ (primera parte) en color azul oscuro.

AUTORES	SEÑALES	BASE DE DATOS	SUJETOS	METODOLOGÍA	ESTIMACIÓN	RESULTADOS
Xu et al. (2019)	SpO ₂	University of Beijing, China (Priv.)	432 niños entre 2 y 15 años	MLP de 3 capas	IAH (e/h) y grado de AOS	Acc: 79.6% / 79.4% / 88.2% Se: 95.3% / 77.8% / 73.5% Sp: 19.1% / 80.5% / 92.7%
Hornero et al. (2017)	SpO ₂	Base privada multicéntrica	4191 niños entre 2 y 18 años	MLP	IAH (e/h) y Grado de AOS	Acc (1/5/10 e/h): 75.2% / 81.7% / 90.2% Se (1/5/10 e/h): 84.0% / 68.2% / 68.7% Sp (1/5/10 e/h): 53.2% / 87.2% / 94.1%
Mortazavi et al. (2024)	SpO ₂	CHAT	844 niños entre 5 y 9.9 años	ResNet (CNN) & CNN-BiGru (CNN + RNN)	IAH (Umbral de 1, 5 y 10)	Acc: ResNet: 72.80% / 91.67% / 95.90% CNN-BiGRU: 86.53% / 91.54% / 96.17% Se: ResNet: 96.80% / 77.13% / 73.25% CNN-BiGRU: 96.27% / 77.79% / 74.58% Sp: ResNet: 48.75% / 97.58% / 98.65% CNN-BiGRU: 61.27% / 97.23% / 99.81%
Jiménez García et al. (2020)	FA	University of Chicago	974 niños entre 0 y 13 años	FCBF + AdaBoost.M2	IAH (e/h) y grado de AOS	Acc: 80.51% / 62.82% / 78.97% Se: 99.37% / 62.33% / 39.77% Sp: 1.33% / 63.11% / 89.58%
Barroso García et al. (2021a)	FA	University of Chicago	946 niños entre 0 y 13 años	FCBF + AdaBoost & Bayesian MLP	IAH (e/h) y grado de AOS	Acc: AdaBoost: 73.61% / 57.46% / 76.07% MLP: 80.85% / 57.14% / 70.47% Se: AdaBoost: 79.89% / 74.43% / 41.06% MLP: 100.00% / 77.25% / 50.00% Sp: AdaBoost: 47.24% / 47.18% / 89.58% MLP: 0.00% / 45.05% / 75.96%
Barroso García et al. (2021b)	FA	University of Chicago	946 niños entre 0 y 13 años	FCBF + MLP	IAH (e/h) y grado de AOS	Acc: 78.14% / 61.20% / 77.35% Se: 94.07% / 78.66% / 55.85% Sp: 11.16% / 50.61% / 83.16%
Barroso García et al. (2017)	FA	University of Chicago	501 niños entre 3 y 9 años	FSLR + LR	IAH (e/h) y grado de AOS	Acc: 60.00% / 76.00% / 80.00% Se: 60.50% / 65.00% / 83.80% Sp: 58.60% / 80.60% / 79.00%

Tabla 2-2. Tabla resumen revisión bibliográfica con la señal SpO₂ en color azul oscuro (segunda parte) y la señal FA en color naranja.

AUTORES	SEÑALES	BASE DE DATOS	SUJETOS	METODOLOGÍA	ESTIMACIÓN	RESULTADOS
Dehkordi et al. (2016)	PPG	British Columbia Children's Hospital in Vancouver (Priv.)	160 niños entre 4 y 13 años	LASSO	IAH (Umbral de 5 e/h)	Acc: 71.00% Se: 76.00% Sp: 68.00%
Shouldice et al (2004)	HRV a partir del ECG	Sleep Medicine Center of Kosair, Louisville	25 niños entre 7 y 18 años	QD (Quadratic Discriminant)	IAH (Umbral 12.5 e/h)	Acc: 84.0% Se: 85.7% Sp: 81.8%
Gil et al. (2010)	HRV	Miguel Servet Hospital in Zaragoza	21 niños entre 2 y 6 años	QDA (Quadratic Discriminant Analysis)	IAH (Umbral 5 e/h)	Acc: 80.00% Se: 87.50% Sp: 71.40%
Lázaro et al. (2014)	PPG Y ECG	Miguel Servet Hospital in Zaragoza	21 niños entre 2 y 6 años	LDA (Linear Discriminant Analysis)	IAH (Umbral 18 e/h)	Acc: 86.70% Se: 100.00% Sp: 71.40%
Martín Montero et al. (2021)	HRV	CHAT y Univ. of Chicago	1738 niños entre 5 y 9.9 años	FCBF + MLP	IAH (e/h) y grado de AOS	Acc: 63.40% / 81.00% / 89.30% Se: 76.30% / 62.50% / 66.70% Sp: 38.30% / 84.20% / 91.60%
Martín Montero et al. (2022)	HRV	CHAT y Univ. of Chicago	1738 niños entre 5 y 9.9 años	MLP	IAH (e/h) y grado de AOS	Acc: 74.58% / 84.95% / 91.64% Se: 85.47% / 64.44% / 53.66% Sp: 35.38% / 93.78% / 97.67%
Martín Montero et al. (2023)	HRV	CHAT	1610 niños entre 5 y 9.9 años	LSBoost & AdaBoost	IAH (e/h) y grado de AOS	Acc: 80.07% / 63.18% / 84.12% Se: 90.76% / 66.67% / 40.00% Sp: 23.40% / 61.17% / 92.03%
Gil et al. (2009)	HRV (del ECG) y PRV a partir de la PPG	Hospital de Zaragoza (Priv.)	21 niños entre 2.5 y 5.4 años	LDA	IAH (Umbral de 5 e/h)	Acc: 80.00% Se: 87.50% Sp: 71.40%

Tabla 2-3. Tabla resumen revisión bibliográfica con la señal HRV (primera parte) y derivados en color morado.

AUTORES	SEÑALES	BASE DE DATOS	SUJETOS	METODOLOGÍA	ESTIMACIÓN	RESULTADOS
García Vicente et al. (2023)	ECG	CHAT	1610 niños entre 5 y 9.9 años	CNN	IAH (e/h) y grado de AOS	Acc: 75.92% / 86.96% / 91.97% Se: 84.19% / 76.67% / 53.66% Sp: 46.15% / 91.39% / 98.06%
Jiménez García et al. (2020)	SpO ₂ y FA	University of Chicago	974 niños entre 0 y 13 años	FCBF + AdaBoost.M2	IAH (e/h) y grado de AOS	Acc: 78.21% / 77.95% / 85.90% Se: 90.79% / 72.60% / 44.58% Sp: 25.33% / 81.15% / 97.07%
Barroso García et al. (2020)	SpO ₂ y FA	University of Chicago	946 niños entre 0 y 13 años	FCBF + Bayesian MLP	IAH (e/h) y grado de AOS	Acc: 81.1% / 60.9% / 80.6% Se: 99.3% / 80.9% / 63.8% Sp: 4.2% / 48.9% / 85.1%
Jiménez García et al. (2024)	SpO ₂ y FA	CHAT + University of Chicago	2512 niños entre 0 y 13 años	CNN + RNN y Grad-CAM	IAH (e/h) y grado de AOS	Acc: CHAT: 87.3% / 93.5% / 93.5% UofC: 84.1% / 84.6% / 90.5% Se: CHAT: 87.0% / 80.2% / 71.4% UofC: 96.8% / 82.9% / 78.3% Sp: CHAT: 88.1% / 99.1% / 97.0% UofC: 30.7% / 85.7% / 93.8%
Jiménez García et al. (2022)	SpO ₂ y FA	CHAT + University of Chicago	2512 niños entre 0 y 13 años	2D CNN	IAH (e/h) y grado de AOS	Acc: CHAT: 84.6% / 93.5% / 94.4% UofC: 84.1% / 84.1% / 90.3% Se: CHAT: 82.4% / 80.2% / 71.4% UofC: 95.2% / 82.2% / 78.3% Sp: CHAT: 92.5% / 99.1% / 98.1% UofC: 37.3% / 85.3% / 93.5%
Gutiérrez-Tobal et al. (2015)	SpO ₂ y FA	University Hospital of Burgos (Priv.)	50 niños entre 3 y 13 años	FSLR y LR	IAH (Umbral de 3 e/h)	Acc: 86.30% Se: 85.90% Sp: 87.40%

Tabla 2-4. Tabla resumen revisión bibliográfica con la señal HRV (segunda parte) y derivados en color morado y las señales SpO₂-FA en color verde.

AUTORES	SEÑALES	BASE DE DATOS	SUJETOS	METODOLOGÍA	ESTIMACIÓN	RESULTADOS
Garde et al. (2014)	SpO ₂ y PRV	British Columbia Children's Hospital in Vancouver (Priv.)	146 niños entre 4 y 13 años	AROC + LDA	IAH (Umbral de 5 e/h)	Acc: 84.90% Se: 88.40% Sp: 83.60%
Ye et al. (2023)	SpO ₂ y HR	University of Beijing, China (Priv.)	3139 niños entre 2 y 18 años	XGBoost & LR	IAH (e/h) y grado de AOS	Acc: XGBoost: 90.45% / 85.67% / 89.81% LR: 79.30% / 83.92% / 87.44% Se: XGBoost: 90.26% / 82.07% / 84.77% LR: 78.90% / 80.00% / 85.79% Sp: XGBoost: 100.0% / 93.78% / 92.11% LR: 100.0% / 92.75% / 89.10%
Garde et al. (2019)	SpO ₂ y HRV	British Columbia Children's Hospital in Vancouver (Priv.)	235 niños entre 6 y 10 años	LR	IAH (e/h) y grado de AOS	Acc: 71.0% / 78.0% / 88.0% Se: 68.0% / 58.0% / 90.0% Sp: 86.0% / 89.0% / 87.0%
Sturludóttir et al. (2023)	FA y ECG	Children's Hospital in Reykjavik (Priv.)	20 niños entre 10 y 13 años	CNN	IAH (Umbral de 5 e/h)	Acc: 95.85% Se: 75.15% Sp: 96.20%
Tan et al. (2014)	SpO ₂ , ECG y FA	University of Chicago (Priv.)	100 niños entre 2 y 16 años	Comparación directa PSG vs Resp. PG	IAH (e/h) y grado de AOS	Acc: 86.0% / 85.0% / 93.0% Se: 82.5% / 62.5% / 65.0% Sp: 90.0% / 100.0% / 100.00%

Tabla 2-5. Tabla resumen revisión bibliográfica con las señales SpO₂-HRV (y derivadas) en rojo, las señales FA-HRV (ECG) en azul claro y las señales SpO₂-HRV(ECG)-FA en amarillo.

Capítulo 3: SEÑALES Y BASES DE DATOS

Para la realización de este proyecto se ha requerido de una gran cantidad de datos y señales, con los que poder entrenar, validar y evaluar un modelo de IA enfocado en la detección de AOS pediátrica. Actualmente, existen varias opciones con las que poder obtener la información requerida como son la recogida de datos en centros especializados como hospitales universitarios o el uso de bases de datos públicas disponibles a través de internet. Para poder disponer de esa gran cantidad de datos requerida, se ha optado por recurrir a bases de datos públicas especializadas en AOS infantil como es el caso de CHAT (*Childhood Adenotonsillectomy Trial*) (Marcus et al., 2013).

3.1. Base de datos CHAT

Con el propósito de analizar los efectos y la evolución de diversos enfoques terapéuticos en pacientes pediátricos con diagnóstico de AOS, en 2007 un grupo de expertos en medicina del sueño y otorrinolaringología iniciaron un estudio en múltiples centros de Estados Unidos (Marcus et al., 2013). Este ensayo clínico centrado en la investigación de la AOS infantil comparó en pacientes de entre 5 y 9 años de edad la eficacia de la adenotonsilectomía temprana contra aquellos que en su lugar se sometían a la observación vigilante con atención de apoyo (*Watchful Waiting with Supportive Care*, WWSC). La principal diferencia entre estos métodos radicaba en el periodo de realización de la cirugía como tratamiento: mientras que algunos niños, se sometían a la cirugía de forma temprana, en otros casos (WWSC) se optó por someter al paciente a un monitoreo prolongado para comprobar si los síntomas se agravaban, mejoraban o remitían con la aplicación de otros tratamientos alternativos. El objetivo de este estudio por tanto, era comprobar si la realización del tratamiento quirúrgico de forma temprana era siempre necesario, o por el contrario, existía la posibilidad de no realizar dicho tratamiento y aplicar en su lugar otro tipo de métodos como son: mejores hábitos de vida, medicamentos o la pérdida de peso; permitiendo así que algunos pacientes pudieran experimentar mejoría o incluso una completa recuperación sin la necesidad de realizar la cirugía (Marcus et al., 2013).

La aparición de este estudio nace de la creciente preocupación por las repercusiones neurocognitivas causadas por la presencia de AOS en niños y sus posibles limitaciones de cara a su futuro desarrollo. Siendo la cirugía de amígdalas el tratamiento más común en niños, se planteó la hipótesis de tomar alternativas menos invasivas en función del paciente para tratar este trastorno. Por ende, se eligieron 464 niños de manera aleatoria de entre 5 y 9 años con presencia de AOS. Algunos de los sujetos fueron sometidos a la cirugía, mientras que otros solo fueron sometidos a un proceso de vigilancia sin cirugía. A ambos grupos se les realizó un control exhaustivo mediante pruebas cognitivas, tests y polisomnografías a lo largo de un periodo de 7 meses. Los resultados indicaban una clara mejoría tanto en síntomas como en calidad de vida de aquellos que fueron sometidos al procedimiento quirúrgico de forma temprana (79% de los pacientes mejoraron). No obstante, aquellos que no fueron sometidos al tratamiento sino a una vigilancia activa, mejoraron en un 46% de los casos gracias a los tratamientos alternativos. En conclusión, el estudio demostró que si bien la cirugía temprana ayuda a reducir los efectos y subsanar el trastorno en una gran mayoría de los casos, existen también un gran porcentaje de sujetos con AOS moderada o leve que pueden presentar mejoría o incluso recuperación sin recurrir a la opción quirúrgica. Los resultados fueron publicados en *The New England Journal of Medicine* en 2013 (Marcus et al., 2013).

Del resultado de este estudio, no solo surgió información crucial sobre la AOS infantil, sino que también dio lugar a la creación de una base de datos con gran cantidad de información sobre este trastorno que ha sido frecuentemente empleada en investigaciones actuales tales como Jiménez García et al. (2022, 2024), Martín-Montero et al. (2021, 2022, 2023), Mortazavi et al. (2024), Vaquerizo-Villar et al. (2021), García Vicente et al. (2023) o Calderón et al. (2020) entre muchos otros.

	CHAT-Entrenamiento	CHAT-Validación	CHAT-Testeo
Sujetos (n)	1006 (61.42%)	326 (19.90%)	306 (18.68%)
Edad (años)	7 [6; 8]	7 [6; 8]	6.9 [6; 8]
Niñas (n)	520 (51.7%)	168 (51.5%)	168 (54.9%)
Niños (n)	471 (46.8%)	156 (47.9%)	134 (43.8%)
IMC (kg/m²)	17.4 [15.6; 21.7]	17.1 [15.4; 21.8]	17.6 [15.7; 21.7]
Sin AOS (n)	219	69	67
AOS Leve (n)	496	168	148
AOS Moderada (n)	160	44	49
AOS Severa (n)	131	45	42

Tabla 3-1. Tabla resumen de los datos contenidos en la base de datos CHAT (Jiménez-García, 2022).

Como ya se ha mencionado previamente, esta base de datos comprende registros de 464 pacientes pediátricos, con edades comprendidas entre los 5 los 9 años de edad de los cuales 49.06% eran niños y 50.94% eran niñas. Además, existe una diversidad étnica en su composición, siendo un 52.85% de la base de datos niños de ascendencia afroamericana, un 36.80% de origen caucásico y un 10.35% de otras etnias, permitiendo así cubrir cualquier posible sesgo de los resultados causado por el factor étnico. Aparte de los datos característicos de los pacientes, también se añaden datos sobre las evaluaciones neurocognitivas, parámetros polisomnográficos, resultados de calidad de vida o incluso las estadísticas de los efectos adversos en ambos grupos. Entre estos efectos adversos registrados podemos observar algunos como: hemorragias en las amígdalas, dolor postoperatorio (solo presente en los sujetos sometidos a la cirugía), asma, catarro, deshidratación, hipersomnolencia o infecciones variadas entre otros (Marcus et al., 2013). Podemos observar el resumen de los datos de la base CHAT en la Tabla 3-1.

La elección de esta base de datos para el desarrollo de nuestro proyecto viene incentivada por la gran cantidad de información contenida gracias al elevado número de sujetos participantes y la gran variedad de datos recogidos de todos ellos. Además al ser pública puede ser solicitada mediante una serie de formularios, lo que permite comparar los resultados obtenidos con otros estudios realizados sobre la misma base de una manera mucho más directa y equitativa.

3.2. Señales empleadas

Para el caso de estudio que nos concierne, se han obtenido las señales de SpO₂, FA y HRV registradas durante las PSG realizadas a los pacientes. A continuación, realizaremos una breve descripción sobre cada una de las 3 señales, los procedimientos para poder obtenerlas y la interpretación de sus valores.

- **Saturación de oxígeno en sangre (SpO_2):**

La saturación de oxígeno en sangre es una métrica empleada para medir la cantidad de oxígeno transportado a los diferentes tejidos del cuerpo humano a través de moléculas de hemoglobina presentes en la sangre arterial. La hemoglobina es la encargada de combinarse con hasta 4 moléculas de oxígeno en los pulmones para repartirlas al resto del cuerpo. Además, también transporta dióxido de carbono desde los tejidos de regreso a los pulmones para su eliminación. Aunque existen tanto sangre arterial como venosa, la medición de la saturación de oxígeno se realiza preferentemente en la sangre arterial. Esto se debe a que la sangre venosa presenta niveles de oxígeno más estables y menos sensibles a cambios inmediatos, lo que limita su utilidad para evaluar la oxigenación en tiempo real. De hecho, en muchos casos, las variaciones en la saturación de oxígeno venosa son tan bajas que pueden resultar mínimas o incluso prácticamente nulas. En cambio, la sangre arterial refleja con mayor precisión y rapidez las variaciones en la oxigenación, ya que recibe directamente el oxígeno de los pulmones. Estos cambios pueden detectarse mediante la absorción de luz, principio en el que se basa la técnica de la pulsioximetría (Nitzan & Taitelbaum, 2008).

Para medir la saturación de oxígeno en sangre disponemos de dos métodos principalmente: la pulsioximetría y el análisis gaseoso de la sangre arterial (Rauniyar et al., 2020). La pulsioximetría consiste en la colocación de un dispositivo conocido como pulsioxímetro en zonas como los dedos, el lóbulo de la oreja o la frente. Este aparato permite medir la frecuencia cardíaca y la SpO_2 utilizando luz roja e infrarroja para detectar la cantidad de oxígeno unido a la hemoglobina en la sangre. Emplea dos longitudes de onda específicas: luz roja (660 nm) y luz infrarroja (940 nm). Parte de la luz es absorbida por la oxihemoglobina y otra parte por la desoxihemoglobina. Gracias a este fenómeno, y mediante un fotodetector ubicado en el extremo opuesto del emisor, es posible registrar las diferencias en la intensidad de la señal continua y alterna en ambas longitudes de onda, lo que permite calcular los valores deseados (Mildenhall, 2008). Podemos ver un fragmento de la señal SpO_2 en la Figura 3-1.

Los factores de SpO_2 dentro de la normalidad son del: 95-100%, valores inferiores a estos pueden suponer una indicación de que existe algún tipo de anomalía o problema en nuestro cuerpo. Concretamente valores de 90-94% supone hipoxemia leve, 85-89% hipoxemia moderada y valores inferiores al 85% hipoxemia grave, lo que supone un problema de salud muy grave (Jubran, 2015). Debemos tener en cuenta que existen factores que pueden afectar al SpO_2 como pueden ser: la altitud, la presencia de enfermedades respiratorias como la neumonía, la AOS o el COVID-19, problemas cardiovasculares como anemia o insuficiencia cardíaca o la calidad del aire presente en el ambiente entre otros (Álvarez et al., 2010).

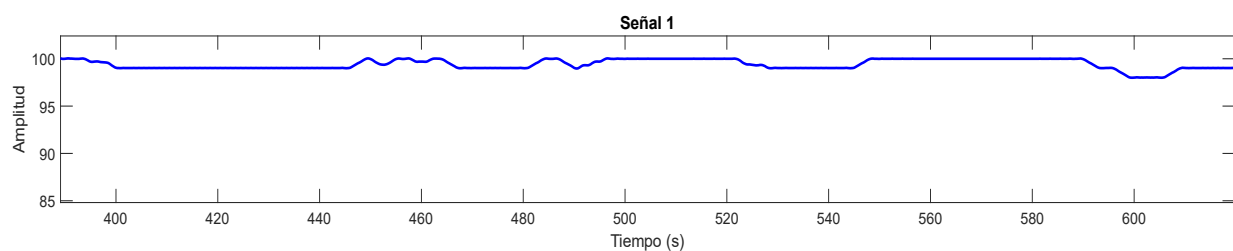


Figura 3-1. Ejemplo de fragmento de señal SpO_2 filtrada (Generada con Matlab).

- **Flujo Aéreo (FA):**

El flujo aéreo, o también conocido como flujo de respiración, consiste en una métrica con la que poder medir la cantidad de aire que entra y sale de los pulmones. La respiración permite un intercambio de gases con los que poder obtener energía para todos los tejidos del cuerpo a partir de moléculas de oxígeno y deshacerse de los componentes residuales como son el dióxido de carbono. Ambas métricas están estrechamente ligadas, ya que si el FA se interrumpe o reduce afecta al SpO_2 debido a que la hemoglobina no puede transportar oxígeno a lo largo del cuerpo si no entra suficiente del exterior (Figura 3-2).

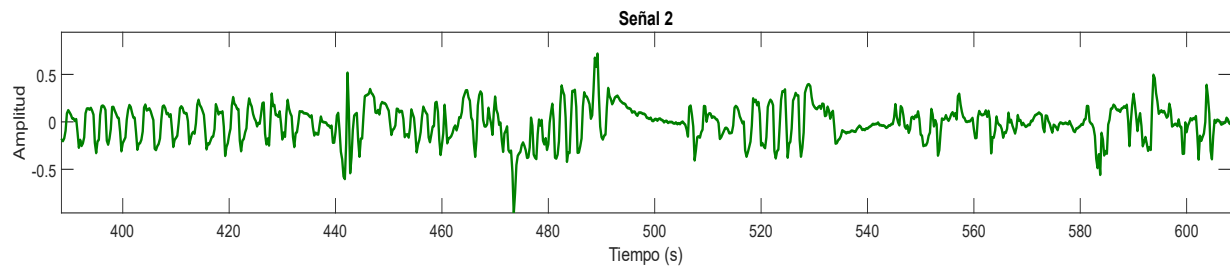


Figura 3-2. Ejemplo de fragmento de señal FA filtrada (Generada con Matlab).

El FA puede medirse de distintas maneras: mediante el volumen del aire inhalado y exhalado, a través de la velocidad del flujo de aire en las vías respiratorias o evaluando la capacidad pulmonar del sujeto. La neumotacografía es una de las técnicas de medición del flujo aéreo existentes, no obstante al ser tan invasiva se opta por emplear alternativas como la estimación directa a partir de sensores de presión o temperatura o la pletismografía pulmonar, que permite medir el volumen de aire dentro de los pulmones gracias a una cabina que mide los cambios de presión del tórax (Grimaldi et al., 2011; Miquel-gomara Perelló et al., 2002; Pellegrino et al., 2005). Para el caso que nos concierne, nos centraremos en los métodos de medida empleados en la PSG: las cánulas nasales o termistores y de manera indirecta las bandas torácicas y abdominales. Las cánulas nasales consisten en una serie de tubos que permiten monitorear los cambios de presión en el flujo que entra o sale del cuerpo. Por otro lado, los termistores son pequeños dispositivos que permiten medir la cantidad de aire que se inhala o exhala gracias a la diferencia de temperatura entre estos dos (Wegener, 2017). Las bandas abdominales o torácicas no representan una medida directa del flujo aéreo, pero permiten obtener parámetros relacionados al aportar información sobre cuánto se expande el tórax al inspirar e espirar.

En circunstancias normales el valor del FA debe tener una amplitud aproximadamente regular sin caídas muy notables, en el caso de que se pierda esa continuidad por un lapso de tiempo prolongado, se puede considerar como un evento anómalo. Concretamente, si la reducción del flujo es de al menos el 90% durante 10 segundos estamos ante un caso de apnea y si la reducción es de al menos el 30% un evento de hipopnea (Caples et al., 2021).

• Heart Rate Variability (HRV):

La variabilidad del ritmo cardíaco consiste en una métrica que permite evaluar las fluctuaciones en los intervalos consecutivos de los latidos del corazón, es decir, la variación en la distancia entre picos R-R en un ECG. Debemos tener en cuenta que una señal de ECG está compuesta por repeticiones de un patrón concreto, determinado como “complejo PQRST”. Este complejo representa la actividad eléctrica del corazón y está compuesto de 5 elementos que se pueden ver en la Figura 3-3 (Wagner & Strauss, 2014) :

- **Intervalo PR:** etapa comprendida entre la onda P y el complejo QRS. En este lapso de tiempo, se mide el retardo entre el inicio de la activación del miocardio ventricular y auricular hasta la respuesta del miocardio.
- **Intervalo QRS:** etapa comprendida entre el inicio y fin del complejo QRS. En este periodo, se mide el tiempo desde que se inicia la activación ventricular hasta su final. Dado que la activación de la pared ventricular izquierda y el tabique intraventricular requiere de un mayor tiempo de activación que el lado derecho, se requiere de una fuerza de compensación adecuada con la que poder equilibrar esto, dicha fuerza queda representada en el complejo QRS.
- **Segmento ST:** intervalo que mide el tiempo entre la activación ventricular y su posterior recuperación.
- **Intervalo QT:** etapa que mide el tiempo desde la activación ventricular hasta el final de su recuperación.

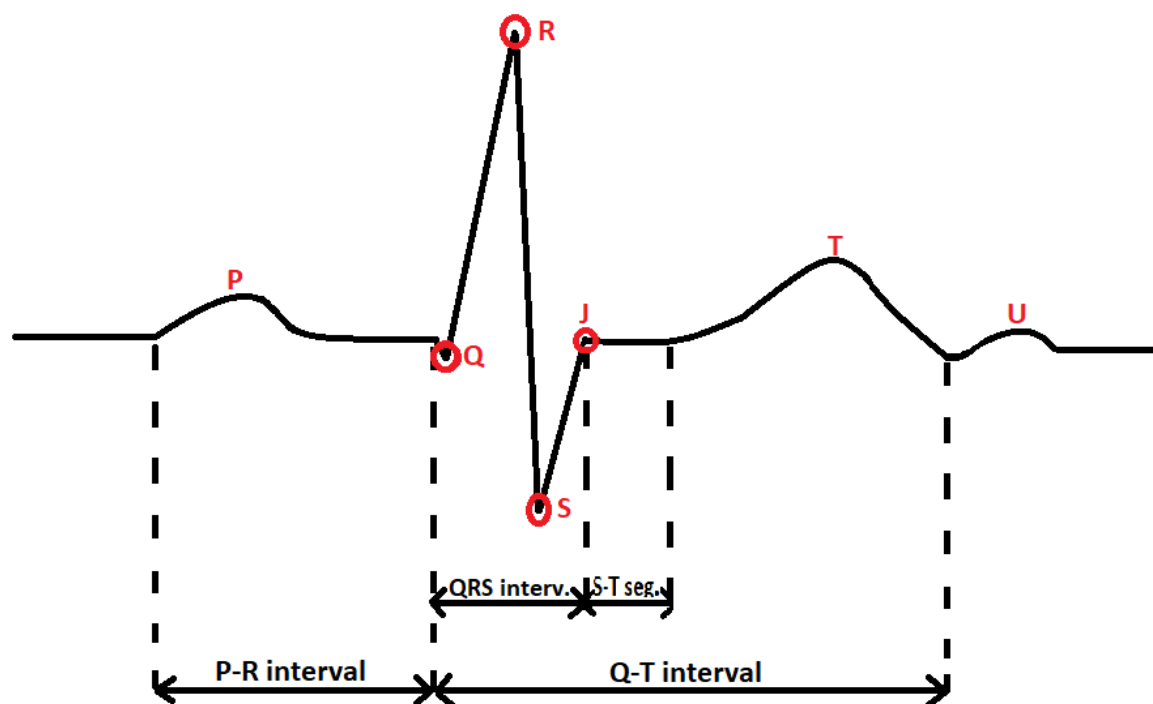


Figura 3-3. Segmento ECG con complejo PQRST. (Autoría propia, inspirado de Wagner & Strauss, 2014).

El HRV puede ser obtenido a partir del propio ECG, simplemente debemos analizar las variaciones o fluctuaciones de los intervalos RR entre latidos consecutivos. Para ello, debemos obtener un ECG e identificar los picos R del complejo QRS; una vez localizados, se debe medir la variabilidad respecto al siguiente punto y al anterior y repetir a lo largo de todo el ECG; finalmente, se analiza dicha variabilidad y se obtiene una medida del HRV (Malik et al., 1996). No obstante, existen otras maneras de obtener una medida similar al HRV sin tener que recurrir al ECG, empleando herramientas como el pulsioxímetro podemos medir el valor del *Pulse Rate Variability* (PRV) y emplearlo como una estimación del HRV para extraer la información deseada.

Los valores normales entre intervalo RR para un ser humano de entre 20 y 40 años sin ninguna patología cardíaca y en reposo oscilan los 600-1000 milisegundos. No obstante, existen factores naturales que modifican estos valores como pueden ser la edad (el valor va decreciendo conforme somos más mayores), la condición física, el nivel de estrés (a mayor cantidad de estrés más se reduce el valor) o la cantidad de sueño acumulada por el sujeto (Zahn et al., 2016). Durante el sueño, el HRV tiende a aumentar en personas sanas en comparación con los niveles registrados durante el día cuando el sujeto está despierto. Sin embargo, en individuos con AOS ocurre lo contrario, el HRV se reduce de forma significativa debido a las alteraciones respiratorias ocasionadas por el trastorno (Qin et al., 2021). Podemos ver un ejemplo en la Figura 3-4.

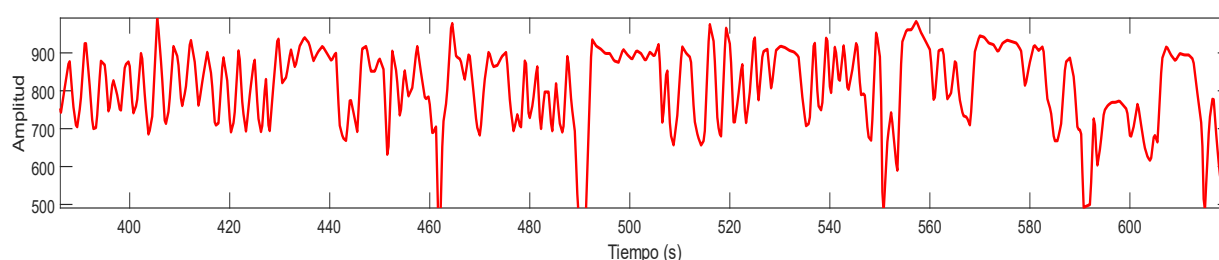


Figura 3-4. Ejemplo de fragmento de señal HRV filtrada (Generada con Matlab).

Capítulo 4: METODOLOGÍA

En este capítulo, se detallaremos la metodología empleada a lo largo del desarrollo del proyecto para su correcta ejecución. En cada punto, explicaremos la forma de implementar nuestro modelo por pasos. En primer lugar, se explicará el proceso de preprocesado realizado seguido del segmentado de las señales. A continuación, se explorará la red CNN desarrollada, describiendo su funcionamiento, estructura, optimización y aplicaciones en este campo. Posteriormente, se presentará la fase de optimización y pruebas realizadas. Finalmente, se comentará el proceso de obtención de resultados y la extracción de conclusiones de todo el trabajo realizado.

4.1. Preprocesado

Cualquier señal existente en un medio real se ve afectada por efectos no deseados como puede ser: el ruido, las interferencias, información redundante no deseada o artefactos aleatorios dañinos para la parte importante de la señal (Jaime & Elizondo, 2002; Najarian & Splinter, 2012). Por ello, en todo aquel campo que implique trabajar con señales, desde la bioingeniería hasta las telecomunicaciones, se requiere la aplicación de una serie de pasos de limpieza y preprocesado con los que poder eliminar la mayor cantidad de elementos no deseados posible, y así poder trabajar más fácil y eficientemente con toda la información útil (Akdemir Akar et al., 2013). Estas técnicas de limpieza y preprocesado, no solo aumentan y mejoran la calidad de la señal, sino que también eliminan componentes no deseadas que pueden suponer un mayor gasto computacional y temporal si estuvieran presentes a la hora de ser procesadas. Para ello, existen diversas técnicas como pueden ser: el filtrado, la interpolación y el remuestreo, el inventanado, la normalización y el escalado o la transformación de dominio entre otras (Najarian & Splinter, 2012).

Para poder trabajar adecuadamente con las señales se deben tener en cuenta una serie de puntos. En primer lugar, dado que se ha realizado un remuestreo, se aplicó un filtro antialiasing con el fin de eliminar o atenuar las componentes de alta frecuencia que podrían inducir aliasing durante el proceso de remuestreo. Es importante tener en cuenta que estos filtros no eliminan completamente las componentes no deseadas sino que solo las atenúan en gran medida; además, la frecuencia de corte elegida para el filtro tiene una cierta tolerancia que impide que sea un corte exacto a la frecuencia deseada (Najarian & Splinter, 2012; Semeria, 2015). Estos fenómenos deben ser tenidos en cuenta a la hora de preprocesar correctamente una señal. Una vez las señales están limpias, nos encontramos con que cada una de ellas tiene una frecuencia de muestreo distinta por lo que trabajar con todas de forma simultánea en un modelo puede ser más complicado. Para solucionar esto, se ha realizado un remuestreo de todas las señales a una frecuencia de 4 Hz, permitiendo así disponer de un dominio común y poder trabajar mejor con el modelo.

Teniendo la señal limpia y remuestreada, el siguiente paso consiste en realizar la normalización y el escalado. Esta etapa es fundamental en casos donde se trabaja con más de una señal distinta, ya que cada una puede estar en un rango completamente distinto al del resto. Esta etapa, permite que todas las señales empleadas se encuentren dentro de un mismo rango absoluto, facilitando así el procesamiento y manipulación de las mismas. Para poder realizar dicha normalización y escalado, emplearemos métricas como la media (μ) y la desviación estándar (σ) permitiendo ajustar la señal a una distribución con $\mu=0$ y $\sigma=1$ mediante la normalización Z-score. La normalización Z-score es ampliamente empleada para normalizar señales gracias a su sencillez y a la utilidad que supone eliminar sesgos debido a la diferencia entre los rangos de valores de las señales (Al-Faiz et al., 2019). Para aplicar esta transformación basta con emplear la Ecuación 4.1, donde X es el valor original de la señal, y μ , σ se obtienen a partir de los datos de entrenamiento:

$$X' = \frac{X - \mu}{\sigma} \quad (4.1)$$

4.2. Segmentación de señales

Una vez las señales de la base de datos CHAT han sido correctamente preprocesadas, disponemos de 1638 registros nocturnos de sujetos pediátricos diagnosticados con distintos grados de AOS. Estos registros contienen: las señales SpO₂, FA y HRV, entre otras, y las marcas de los eventos de apnea/hipopnea etiquetadas por expertos. Los registros nocturnos duran en torno a 8-10 horas y como fueron remuestreados a 4 Hz, se ha optado por segmentar cada uno de los registros en fragmentos de 20 minutos de duración, eliminando la parte residual del registro en caso de no llegar a formar un fragmento completo. Cada fragmento contendrá por lo tanto las 3 señales mencionadas y las marcas que localizan los eventos de apnea/hipopnea en esos periodos de 20 min. El resultado de realizar el segmentado son 47.951 nuevos registros de 20 min de duración, separados entre entrenamiento (*train*), validación (*val*) y testeo (*test*) siguiendo la regla del 60-20-20 (60% de los datos a *train*, 20% a *val* y 20% a *test*).

Todos estos fragmentos extraídos de los registros originales se han empleado como dataset para el modelo de DL a desarrollar. Cada ejemplo del dataset tendrá 2 componentes:

- Una primera matriz de 4.800 x 3, compuesta por el número total de muestras del fragmento (20 min * 60 s/min * 4 muestras/s = 4.800 muestras) y las 3 señales registradas (SpO₂, FA y HRV) de forma que se pueda extraer los datos de la señal deseada simplemente eligiendo una de las 3 componentes de la matriz.
- Una segunda variable escalar que representa la cantidad de eventos de apnea/hipopnea presentes en ese fragmento etiquetados por el experto médico. Esta variable es un número entero.

En las Figuras 4-1, 4-2 y 4-3 podemos observar el resultado de segmentar los registros para cada una de las 3 señales. En primer lugar el primer registro de la señal SpO₂ frente al primer fragmento de ese propio registro. Además, se han añadido una serie de líneas verticales punteadas que indican cada uno de los fragmentos. Como podemos apreciar, hay una serie de muestras (en especial al comienzo) donde la señal de SpO₂ se reduce a amplitud 0, esto se debe a fallos en el sensor o falsos contactos puntuales con el paciente al colocarle los sensores. No obstante, se puede observar cómo esos puntos no han sido etiquetados por los expertos ya que no son considerados eventos de apnea/hipopnea.

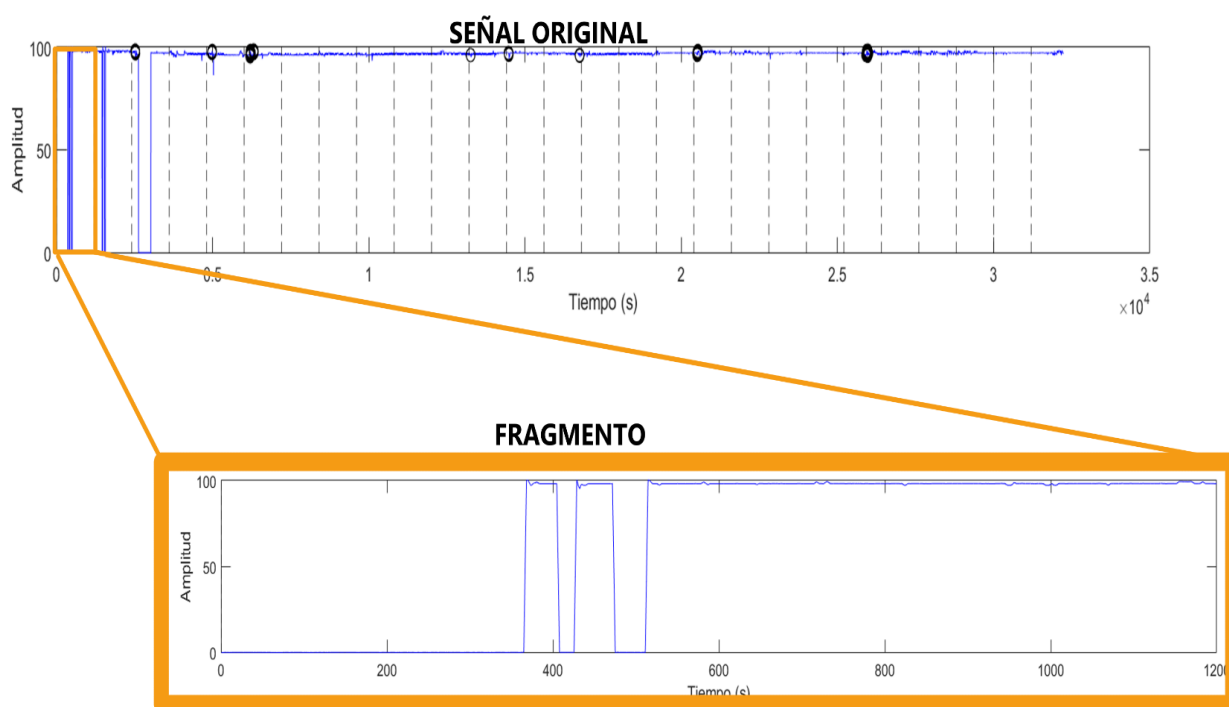


Figura 4-1. Señal SpO₂ completa (superior) vs señal SpO₂ fragmentada (inferior) (Autoría propia, elaborada con Matlab).

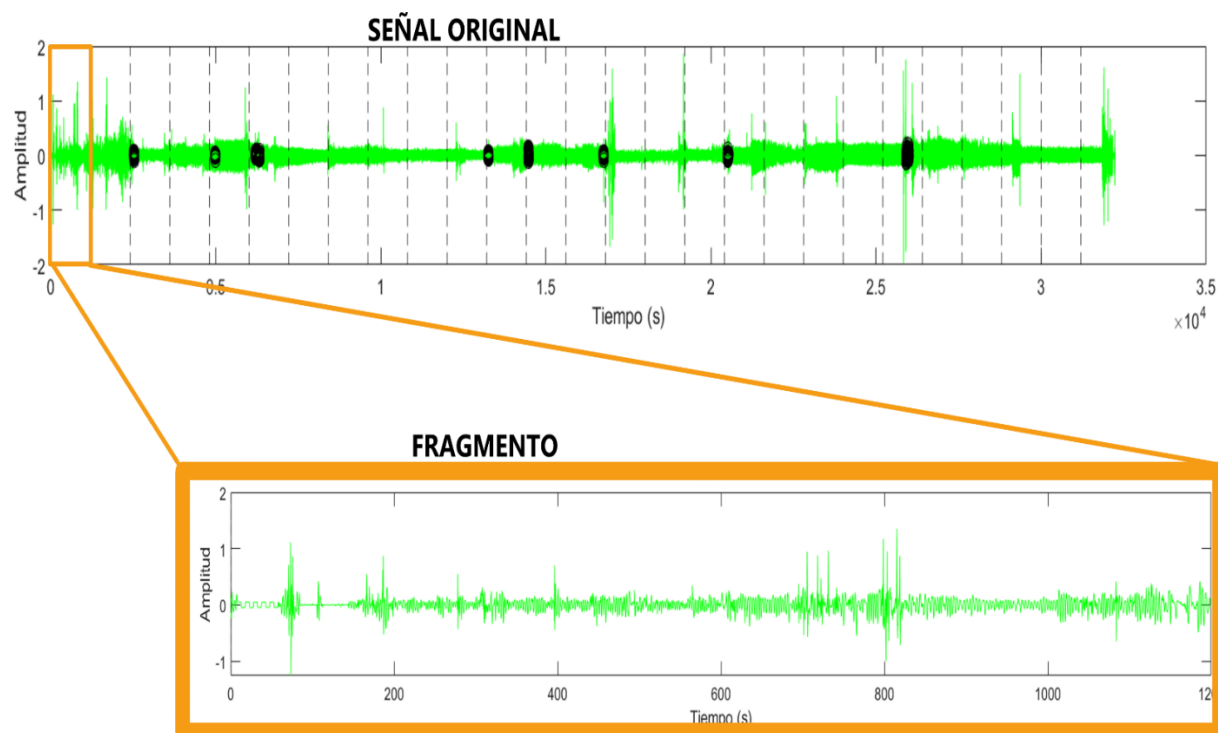


Figura 4-2. Señal FA completa (superior) vs señal FA fragmentada (inferior) (Autoría propia, elaborada con Matlab).

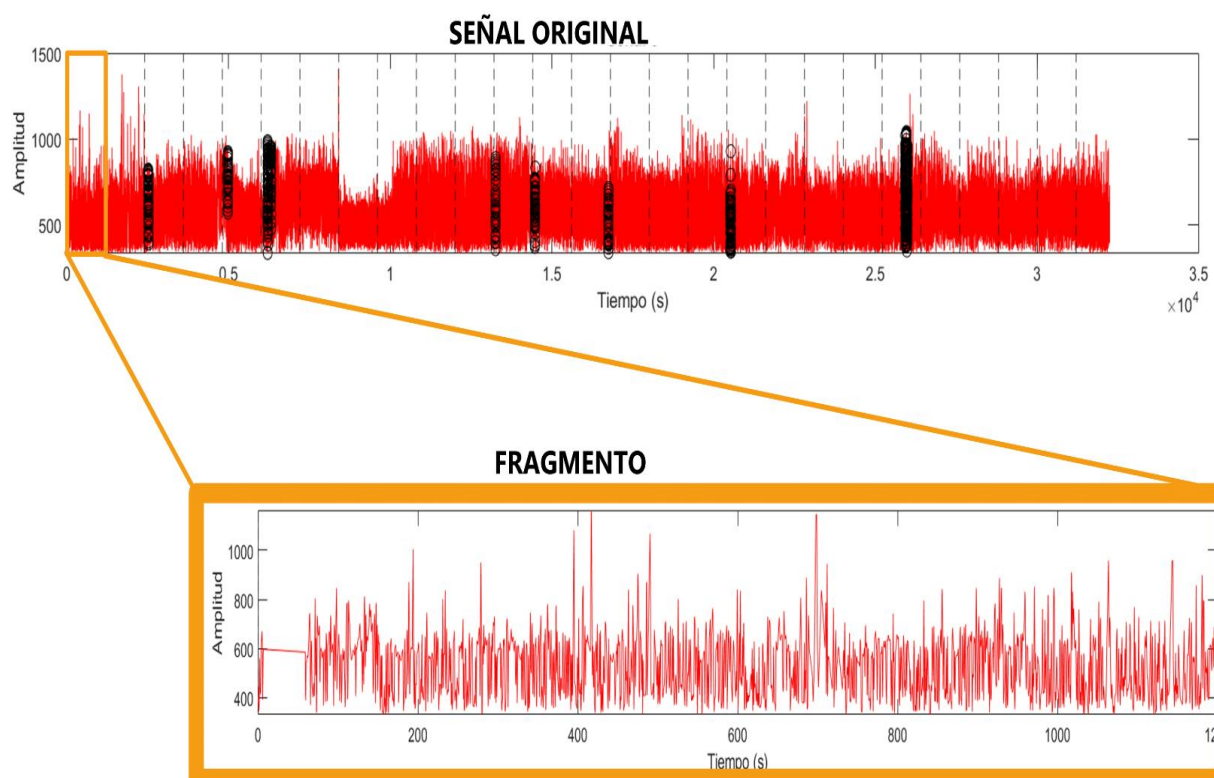


Figura 4-3. Señal HRV completa (superior) vs señal HRV fragmentada (inferior) (Autoría propia, elaborada con Matlab).

En las marcas de las señales registradas se pueden apreciar los intervalos correspondientes a los eventos de apnea/hipopnea, que se extienden durante varios segundos. Por ende, debemos contabilizar cada intervalo como un único evento, es decir, fusionar todas esas marcas como si fuesen una única marca. Esto permitirá a la red aprender correctamente, ya que podrá comparar las estimaciones realizadas con el número de eventos presentes en cada segmento (Figura 4-4).

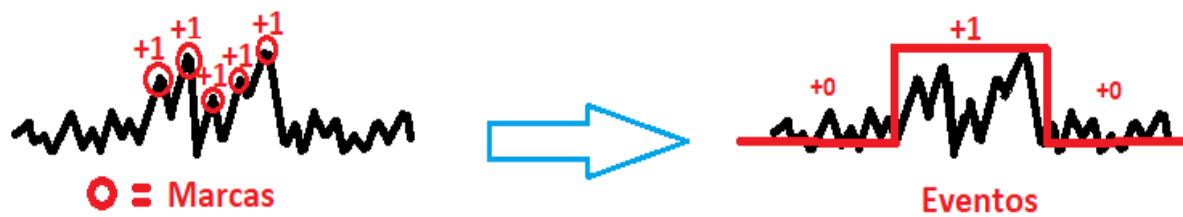


Figura 4-4. Esquema del proceso de umbralización de los eventos (Autoría propia).

4.3. Generador de datos

Lo primero que debemos tener en cuenta a la hora de desarrollar cualquier algoritmo de IA es la estructura que se dará a los datos de entrada. Generalmente, se emplea una división de todos los datos disponibles en distintos grupos: un primer grupo de entrenamiento (*train*), un grupo de validación (*val*) y un grupo de testeo/evaluación (*test*). Para realizar esta división existen distintas estrategias que se pueden seguir. Aunque la metodología puede variar según el caso, la técnica más empleada es 60-20-20, haciendo referencia al porcentaje de datos que se introducen en *train*, *val* y *test* respectivamente (Aggarwal, 2018a; Kelleher, 2019).

- **Datos de entrenamiento (*train*):** el grupo de datos de entrenamiento se emplea para desarrollar el modelo y optimizar los pesos o parámetros y que así aprenda las características y patrones presentes en los datos.
- **Datos de validación (*val*):** el grupo de validación se emplea para evaluar el rendimiento del modelo durante el entrenamiento pero sin ajustar de forma directa sus parámetros. Su función principal es controlar que no haya una gran diferencia entre los resultados obtenidos durante el entrenamiento y la validación, para evitar así posibles sobreajustes con los datos.
- **Datos de testeo (*test*):** el grupo de test, se emplea para realizar una última verificación del modelo después de entrenarlo y validarlo, empleando datos que nunca haya visto para observar si el comportamiento es el deseado y el modelo consigue generalizar y aprender de la forma adecuada.

Con la información preparada, disponemos de un gran volumen de señales con los que poder entrenar el modelo. Sin embargo, emplear simultáneamente todas las muestras disponibles podría suponer una cantidad de datos excesiva para los recursos computacionales del equipo de procesamiento desbordando su memoria y comprometiendo el rendimiento y la funcionalidad del algoritmo. Para evitar este problema, es necesario implementar un mecanismo que permita gestionar la carga de datos de manera eficiente, introduciendo los datos en fracciones de manera controlada y escalonada.

Para poder introducir los datos en el sistema y no saturarlo, se ha establecido una división en los datos en función del valor de *batch_size*, permitiendo así que los datos no se introduzcan en el modelo todos a la vez. Para poder ir introduciendo estos lotes, una de las soluciones ampliamente empleada son los generadores de datos. Estos sistemas permiten ir alimentando a los modelos con datos de forma continua y progresiva, de forma que el modelo disponga a la larga de todos los datos posibles pero sin saturarse al tener que almacenar y procesar toda la información a la vez. Este generador extrae los datos en lotes del tamaño que fije el hiperparámetro *batch_size* y los introduce en el modelo secuencialmente durante la fase de entrenamiento. Una vez el modelo termine de procesar ese lote de datos, el generador le introducirá más lotes de forma continua hasta terminar con todos los datos disponibles. Cuando se hayan terminado todos los lotes al final de la época, se reorganizan aleatoriamente los datos y se volverán a introducir por lotes a lo largo de la siguiente época. El ciclo se repetirá hasta llegar al número máximo del número de épocas o el procedimiento de *Early Stopping* pare el entrenamiento. En la Figura 4-5 podemos ver un esquema de su funcionamiento.

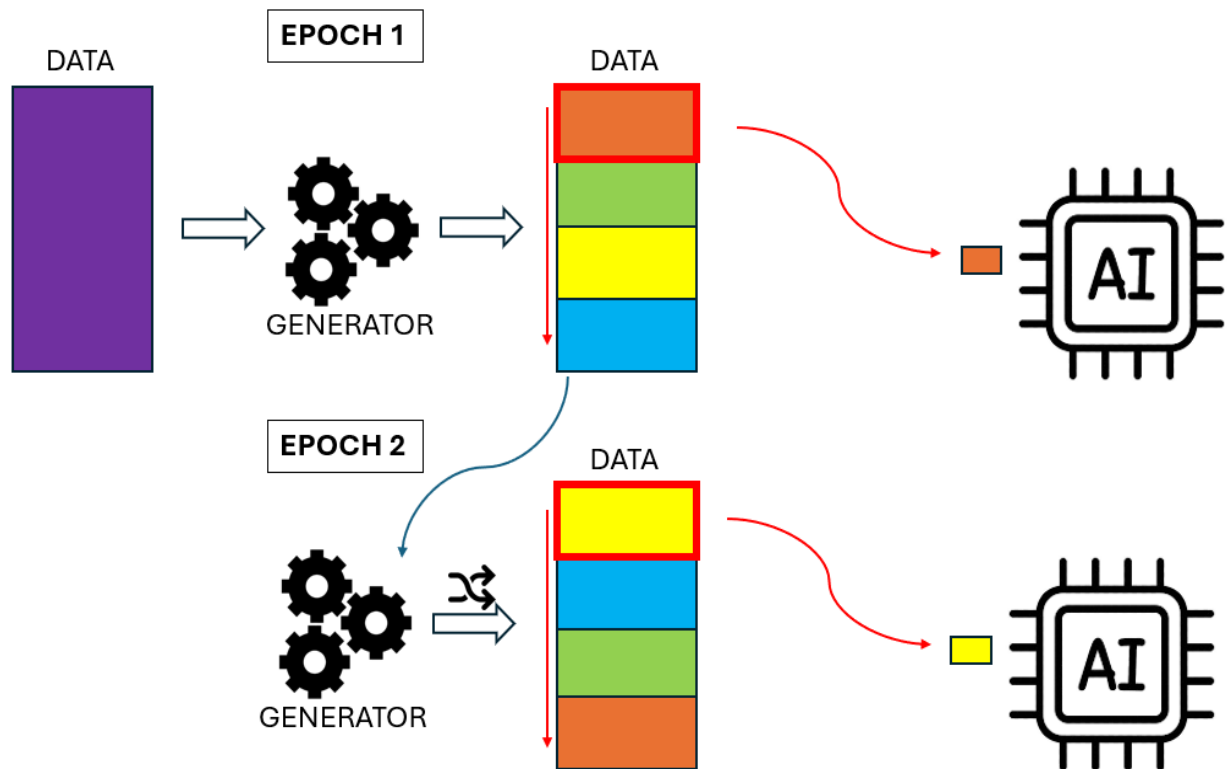


Figura 4-5. Esquema de funcionamiento del generador de datos (Autoría propia).

Debemos tener en cuenta que cuando se emplea más de una señal, es recomendable normalizar dichas señales mediante su media y desviación estándar como se ha visto en la parte de preprocesado. Es decir, en nuestro caso al emplear las señales de 1 en 1, no las hemos transformado; pero, al emplear 2 o incluso las 3, debemos de normalizarlas, ya que si no la diferencia entre las escalas de cada una de las señales puede introducir un sesgo considerable dentro del sistema. Cabe destacar que si se normalizan las señales para entrenar el modelo, se debe de normalizar de la misma manera cuando se emplee el modelo para realizar las estimaciones.

4.4. Modelo desarrollado

Las redes neuronales constituyen un pilar básico de los sistemas DL, esto se debe principalmente a su eficiencia y utilidad en el reconocimiento e identificación de características en distintos tipos de datos. Existen una gran cantidad de tipos de redes diferentes, lo que permite disponer de un amplio abanico de posibilidades a la hora de realizar una elección, dependiendo de la tarea que se pretenda desempeñar. En este trabajo nos hemos centrado en las redes CNN, ya que son especialmente útiles a la hora de reconocer características en datos espaciales como imágenes o señales lo que las hace ideales para la búsqueda de patrones, anomalías o artefactos.

4.4.1. Estructura básica

Las CNN son modelos de DL especializados en la extracción de características a partir de datos con estructura espacial, como imágenes o señales. Como ya se ha explicado previamente, este tipo de modelos tratan de recrear de la manera más precisa posible el comportamiento del cerebro humano a la hora de analizar patrones, mediante la agrupación de neuronas artificiales en estructuras concretas. Originalmente, estas redes fueron diseñadas para el reconocimiento de dígitos escritos de manera manual. La idea principal de dicho proyecto consistía en desarrollar una red en la cual las neuronas de las capas iniciales pudieran extraer patrones visuales sencillos, que pudieran ser combinados en capas más profundas para

permitir crear conceptos más complejos con los que poder generar una respuesta final (Aggarwal, 2018a; Kelleher, 2019; Szeliski, 2011; Voulodimos et al., 2018b). Este principio sigue vigente en la actualidad: las redes extraen características básicas de los datos en las primeras capas, como puede ser el color de los píxeles en una imagen; combinan dichas características en representaciones más complejas, como puede ser la creación de texturas o contornos; y finalmente, elaboran una respuesta final como puede ser el reconocimiento de objetos en las imágenes introducidos como datos de entrada a la red (O'Shea & Nash, 2015).

Para lograr esto las CNN emplean un sistema de bloques secuenciales, que a su vez, están compuestos por una serie de capas encargadas de realizar las funciones de extracción de características, adaptación de las dimensiones o normalización de resultados entre otras funciones (He et al., 2015b; Kelleher, 2019). La realización de estos pasos sobre los datos de entrada depende de la elección, colocación y orden de las distintas capas que conforman los bloques de la red. Generalmente, los bloques de estas CNN son denominados como “*Conv Blocks*”, ya que suelen estar compuestos por un número variable de capas de convolución, capas de activación y capas de normalización, aunque, también pueden aplicarse bloques de reducción de dimensionalidad. Tras estos bloques se suelen añadir bloques densos (también conocidos como *fully connected*) empleados para terminar de establecer las dimensiones red (O'Shea & Nash, 2015). No obstante, la elección de la composición de los bloques depende en gran medida del objetivo de la CNN y la estructura de los datos que se someten al análisis. En la Figura 4-6 se puede observar un esquema de una red neuronal sencilla a modo ilustrativo.

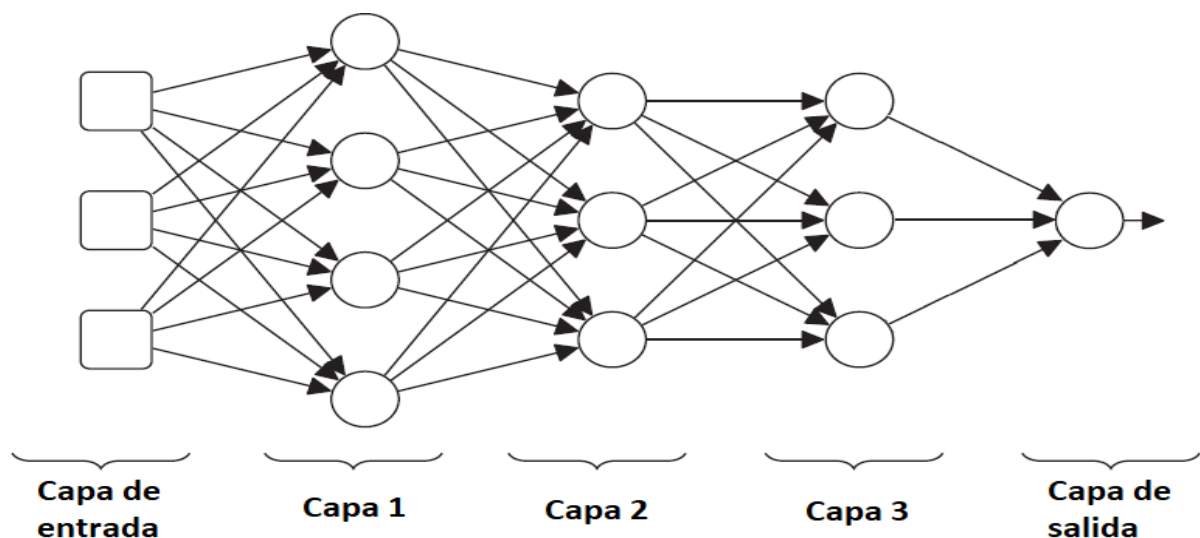


Figura 4-6. Esquema de una red neuronal multicapa con capas etiquetadas (Adaptada de Kelleher, 2019).

Para lograr crear conceptos elaborados, dentro de cada capa convolucional se encuentran unas estructuras encargadas de realizar una serie de convoluciones sobre los datos. Estas estructuras se llaman filtros o *kernels*, y consisten en pequeñas matrices compuestas por una serie de pesos que son multiplicados por los datos de entrada, generando una nueva matriz denominada *feature map*, que representa la información extraída de los datos originales. Basándonos en la revisión del estado de la técnica previamente realizada, se ha desarrollado un modelo CNN basado en 5 bloques idénticos conformados por una serie de capas. La estructura de dichos bloques está conformada por las siguientes capas en este orden:

- **Capa convolucional:** en primer lugar, una capa de convolución 1D con 64 filtros de tamaño 6 y una inicialización de coeficientes del filtro “*he_normal*”. Este método asigna valores aleatorios a los pesos de las capas siguiendo una distribución gaussiana con media 0 y desviación estándar 1 basándose en la cantidad de neuronas de la capa (He et al., 2015b). La tarea de cada capa consiste generar los mapas de características que se procesaran en las siguientes capas de forma que se puedan ir estableciendo relaciones entre las características.

- **Capa *batch normalization*:** esta capa trata de normalizar las activaciones de la capa anterior con cada lote (*batch*) durante el entrenamiento calculando la media y reescalando los valores. Se ha añadido una capa *batch normalization* después de cada capa convolucional.
- **Capa de activación:** en esta capa se aplica la función de activación *ReLU* para todas las neuronas (Ecuación 4.2). La elección de esta función frente a otras alternativas como *ELU* o *Leaky ReLU* se debe a su simplicidad computacional, su buen rendimiento empírico y su efectividad a la hora de evitar el problema del gradiente desvanecido en valores positivos. Aunque funciones como *ELU* o *Leaky ReLU* pueden resolver limitaciones específicas de *ReLU* como el caso de neuronas “muertas”, *ReLU* sigue siendo la opción más común debido a que permite un entrenamiento más rápido y ha demostrado funcionar correctamente en una amplia gama de tareas (Clevert et al., 2015; Glorot et al., 2011; Maas et al., 2013).

$$ReLU(x) = \max(0, x) \quad (4.2)$$

- **Capa *MaxPooling1D*:** finalmente, aquí se reduce la dimensionalidad de los datos para eliminar información redundante reteniendo la activación máxima entre muestras adyacentes, en nuestro caso aplicaremos una ventana de tamaño $K = 2$.

Esta estructura se repite tantas veces como quede fijado en el hiperparámetro *num_blocks* de forma secuencial hasta llegar a un último conjunto de capas finales que se encargaran de tomar la decisión final (Kelleher, 2019). Dichas capas son:

- **Capa *Flatten*:** convierte el array multidimensional con los mapas de características generados en las capas convolucionales en un vector unidimensional apto para analizar los patrones extraídos en las capas finales del modelo.
- **Capa *fully connected*:** en esta capa hemos definido una única neurona a la que se conectan todas las de la capa anterior, hemos fijado la activación a tipo “linear” y de nuevo como en las capas convolucionales la inicialización es “*he-normal*”.

Por lo que el resultado de la composición inicial del modelo sería (Figura 4-7):

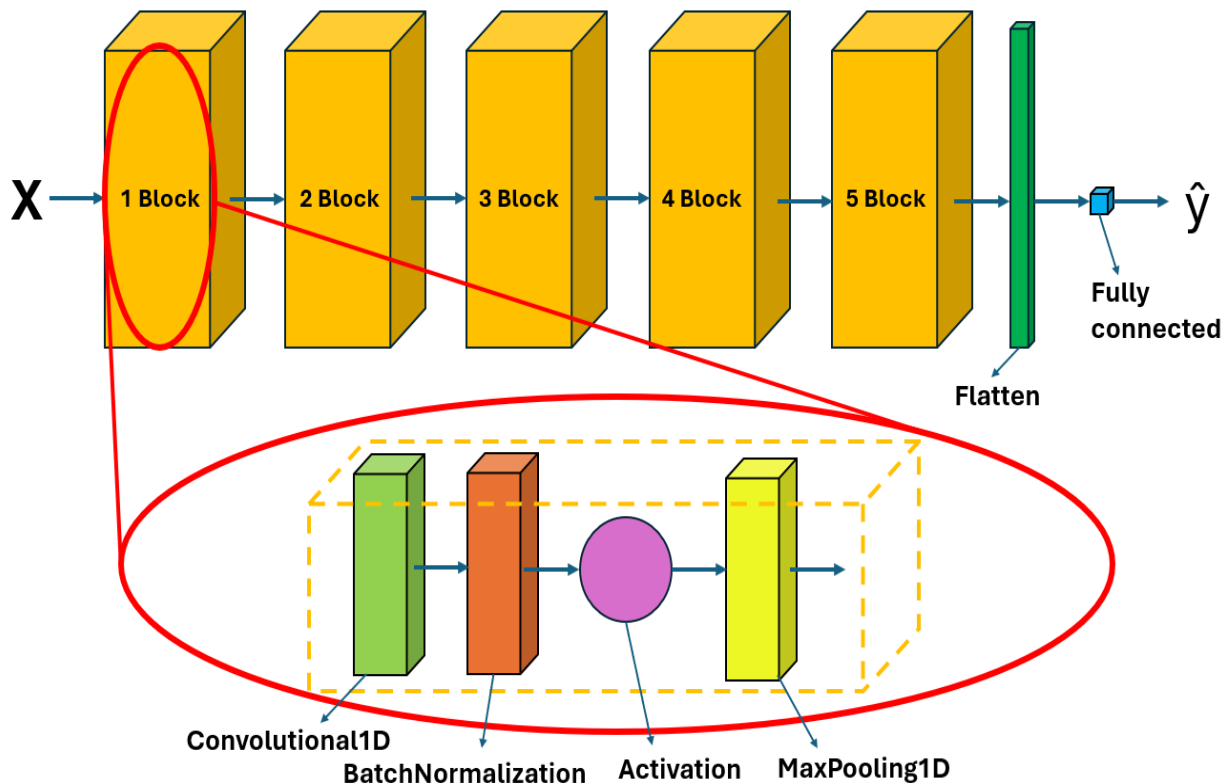


Figura 4-7. Esquema genérico del modelo CNN inicial (Autoría propia).

Durante el proceso de entrenamiento del modelo, los parámetros de las capas (convolucionales, *fully connected*, etc.) deben ajustarse. Estos parámetros incluyen los pesos y el sesgo, parámetros presentes en cada neurona, pudiendo coincidir los valores de estos entre ellas. Para poder ajustar estos valores adecuadamente, en primer lugar se realiza el proceso de inicialización, proceso en el cual se inicializa los valores de sesgo a 0 y los valores de los pesos a números aleatorios (Kelleher, 2019; O'Shea & Nash, 2015). Con los parámetros ya inicializados, se introducen los datos en la red neuronal, se analizan y procesan y se genera una salida. Esta etapa se denomina como *feedforward* o *forward propagation* y consiste en el proceso de introducir datos por la entrada hasta generar una respuesta (O'Shea & Nash, 2015).

Una vez el modelo genera una predicción, se evalúa la diferencia entre el valor estimado y el valor real mediante una función de pérdida. Esta evaluación permite determinar si el error supera un umbral aceptable y, en caso afirmativo, se inicia una nueva etapa de ajuste del modelo para mejorar su precisión. En esta segunda etapa, se realiza el mismo camino que en la etapa anterior pero en sentido opuesto: se inicia en las capas finales y se van realizando las derivadas hacia atrás en cada neurona hasta llegar a las capas iniciales con el objetivo de tratar de minimizar los gradientes con respecto a los parámetros de la red. Esta etapa es conocida como *back propagation* y su objetivo principal consiste en realizar un proceso de optimización con el objetivo principal de minimizar una función de error generada a partir de los datos reales y las estimaciones del modelo. Para ello, la fase de optimización tratará de indicar los pesos a ajustar y la dirección en la que hacerlo acorde a una función matemática que establece cuanto debe afectar el nuevo valor calculado. Esta función se conoce como función de aprendizaje y va determinado por la tasa de aprendizaje (*learning rate*), elemento que fija el nivel de atención que se debe dar a la corrección de los parámetros de una neurona frente a los anteriores. Actualmente, existen algoritmos capaces de ajustar dinámicamente la tasa de aprendizaje para que la red funcione de una mejor manera, adaptándose a cada momento del entrenamiento. En la Figura 4-8 se representa un esquema de este proceso.

El objetivo de realizar estos pasos es poder alcanzar el mínimo global de la función de error a optimizar. Para ello, al calcular las derivadas se procura determinar la dirección en la que debemos ajustar los parámetros de la red para reducir el valor de la función de perdidas. Es decir, intentar que los resultados de nuestro modelo se asemejen cada vez más a los resultados reales esperados.

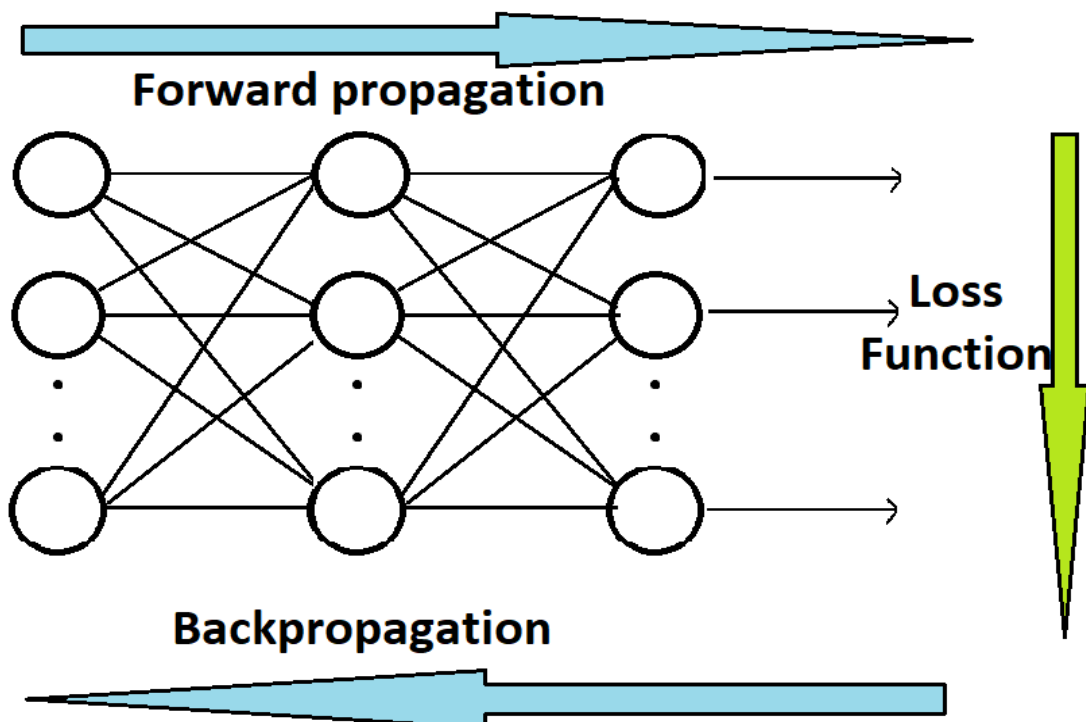


Figura 4-8. Esquema de una etapa de entrenamiento de un modelo, indicando las funciones *feedforward* o *forward propagation*, *loss function* y *back propagation* (Autoría propia).

El modelo debe ser compilado antes de poder emplearse en el sistema. A la hora de realizar la compilación es necesario añadir una serie de parámetros con los que poder configurar adecuadamente dicho proceso. En primer lugar, debemos fijar el algoritmo de optimización; en vez de emplear el predeterminado hemos optado por seleccionar el algoritmo “*Adam*”. Este algoritmo, es ampliamente utilizado dada su capacidad de poder ajustar de manera adaptativa la tasa de aprendizaje inicial proporcionada. Por otro lado, también debemos de fijar una función de pérdidas; en nuestro caso, emplearemos la función de pérdida Huber que permite combinar características del MSE (*Mean Square Error*) y del MAE (*Mean Absolute Error*). La función Huber funciona de forma que si el error entre la predicción y el valor real es pequeño, se emplea una función cuadrática como el MSE y en caso opuesto, una función lineal como el MAE (Huber, 1963).

Una vez desarrollado el modelo, se han aplicado una serie de mejoras para aumentar su eficiencia y mejorar los resultados durante la etapa de entrenamiento mediante una serie de *callbacks* que permitan implementar estas mejoras. Se han añadido los siguientes *callbacks* al modelo (TensorFlow Team, 2024):

- ***Model Checkpoint***: este *callback* permite crear puntos de progreso para guardar el modelo en un punto determinado. En nuestro caso, guarda el modelo cada vez que se obtengan mejores resultados en la variable de monitoreo que se le indique (en nuestro caso el error de validación), junto a la época en la que se obtuvo.
- ***Early Stopping***: este *callback* reduce el tiempo de entrenamiento a la vez que previene el sobreajuste, ya que permite monitorizar el error del modelo a lo largo de las épocas de entrenamiento. En nuestro caso, permite que si no se mejora dicho valor en 30 épocas el modelo termine la etapa de entrenamiento y restaure los pesos a los obtenidos en la época en la que el error de validación fue el mínimo.
- ***Reduce LR On Plateau***: este *callback* permite reducir de forma dinámica el valor de la tasa de aprendizaje cuando el modelo se estanca, favoreciendo así una mejor convergencia. Al igual que con *EarlyStopping*, es necesario establecer un valor para la variable *patience_reduce_learn_rate*, que define cuántas épocas deben transcurrir sin mejora en el error antes de aplicar la reducción. En nuestro caso, se ha asignado un valor de 0.5 a la variable factor, lo que implica que la tasa de aprendizaje se reducirá a la mitad cada vez que se active este *callback*.

Dado que trabajamos con 3 señales distintas (SpO₂, FA y HRV), y nuestro objetivo es compararlas entre sí, ya sea de forma individual, en grupos de 2 o en conjunto; puede resultar de interés desarrollar algún mecanismo que permita elegir las señales que se quieren introducir en el modelo. Para ello, se ha diseñado un selector de señales que organiza los datos de los registros en matrices tridimensionales. En esta estructura, cada fila representa un fragmento, cada columna corresponde a las muestras de los fragmentos y cada capa de la matriz (dimensión de profundidad) alberga cada una de las señales seleccionadas. Este enfoque permite, mediante la modificación de un único parámetro (*signals_selected* en nuestro caso), elegir cualquier combinación de señales, generando así matrices de 1, 2 o 3 columnas en función de lo requerido. Cabe destacar que las marcas de los eventos siempre se cargan, dado que su valor permanece invariable independientemente del número y combinación de las señales seleccionadas. Además, el modelo se ha diseñado con variables dinámicas para permitir una mejor adaptación en función de las señales cargadas.

4.4.2. Optimización de los hiperparámetros

Aunque lograr un ajuste óptimo de los parámetros o pesos de una red neuronal es fundamental, esto por sí solo no garantiza un rendimiento satisfactorio del modelo. Es igualmente importante ajustar los hiperparámetros, que son configuraciones establecidas antes del proceso de entrenamiento y que no se actualizan durante el mismo. Estos hiperparámetros determinan aspectos clave de la arquitectura, dimensiones y el proceso de aprendizaje del modelo. En nuestro caso, los hiperparámetros optimizados han sido (Goodfellow et al., 2016; Kelleher, 2019):

- **Número de bloques (*num_blocks*):** este valor establece el número de bloques o capas convolucionales que empleará el modelo, a mayor valor más complejos serán los conceptos creados por la red y por lo tanto más aprenderá la red. Como ya se ha mencionado previamente, debemos tener en cuenta que si este valor es elevado la red puede volverse demasiado profunda, por lo que el modelo podría memorizar en lugar de aprender de los datos o en caso contrario no aprender correctamente.
- **Número de filtros por capa (*num_filters*):** este valor fija el número de filtros o *kernels* por capa convolucional, debemos tener en cuenta que emplear un valor alto tiene consecuencias similares al hiperparámetro *num_blocks*. Fijar un valor alto supone un mayor tiempo de computación y aumentar la posibilidad de causar sobreajuste, mientras que un valor bajo causaría subajuste.
- **Tamaño de los filtros (*filter_size*):** con este valor, se fija el número de coeficientes (parámetros) de los filtros. Debemos tener en cuenta que en nuestro caso se han usado filtros unidimensionales al ser señales y no imágenes, es decir, los filtros se aplican sobre la dimensión temporal. Debemos tener en cuenta que los filtros pueden reducir las dimensiones de los datos (dependiendo de la configuración del *padding* o del paso del filtro), puede darse el caso de que al fijar un valor demasiado grande las dimensiones se reduzcan tanto que el modelo no pueda trabajar con ellos y acabe fallando.
- **Tamaño de lote (*batch_size*):** en los modelos de DL se requieren gran cantidad de datos, por lo que en algunos casos meter todos a la vez no supone una buena idea. Por lo tanto se suelen dividir en lotes o *batches*, y así poder introducirlos por partes dentro del modelo, permitiendo focalizar los datos y que el algoritmo pueda trabajar de forma más eficiente. Es importante optimizar este hiperparámetro ya que pueden existir relaciones entre los datos que no sean visible en caso de que estén en lotes distintos, por lo que, debemos elegir un valor que permita al algoritmo procesar los datos eficientemente pero sin excedernos o perderemos parte de la información referente a patrones entre los datos.
- **Tasa de aprendizaje inicial (*learning_rate*):** en un algoritmo de optimización de descenso de gradiente, la tasa de aprendizaje permite al modelo prestar más o menos atención a las correcciones de los parámetros en cada época regulando así la variación de estos entre épocas. Si bien se suele ajustar de manera dinámica, el valor inicial sobre el que se comenzará a realizar dicho ajuste también puede ser optimizado como un hiperparámetro. Debemos tener en cuenta que un valor inicial elevado puede impedir la obtención de un mínimo error y un valor reducido puede incrementar el número de épocas necesarias de manera no deseada además de obtener un modelo entrenado de manera subóptima debido a la convergencia hacia mínimos locales de la función de error.
- **Tasa de Dropout (*dropout_rate*):** el *dropout* consiste en una estrategia de prevención del sobreajuste que implica desactivar un porcentaje de neuronas para que así el modelo no memorice los datos. Esta tasa también puede suponer un hiperparámetro a optimizar.

A pesar de que el número de épocas del modelo también puede ser un hiperparámetro a ajustar, no suele aplicarse, ya que se suelen usar otros como el *EarlyStopping*. Esta técnica permite fijar un valor de paciencia para el cual el algoritmo debe seguir probando épocas, de forma que si el modelo mejora en ese periodo se vuelve a reiniciar la paciencia, y en caso contrario, se termine el entrenamiento evitando así gastar recursos innecesarios (O'Shea & Nash, 2015; Prechelt, 1998).

Debemos tener en cuenta que al tener varias combinaciones de señales posibles, los hiperparámetros que son óptimos para una combinación pueden no serlo para las demás. Debido a esto, tendremos que ajustar los hiperparámetros de cada posible combinación de señales por separado, y para ello, realizaremos la etapa de entrenamiento con cada una de las señales dando un rango de valores a cada hiperparámetro para observar cual es el que mejor se comporta. El procedimiento a seguir será el siguiente: en primer lugar, emplearemos los hiperparámetros definidos inicialmente para todas las señales. Después, se ajustará uno de ellos, entrenando el modelo con distintos valores dentro de un rango específico durante 4 iteraciones, mientras que los demás permanecerán constantes. Cada uno de estos valores se probará en la

fase de entrenamiento cada una de las iteraciones, con el objetivo de descartar posibles resultados atípicos. Una vez evaluadas todas las posibilidades, se analizarán las pérdidas de validación obtenidas. A partir de este análisis, se seleccionará el valor que haya mostrado el mejor desempeño en términos de pérdidas de validación, y se fijará para esa señal antes de pasar a ajustar el siguiente hiperparámetro. Se repetirá el procedimiento con el resto de hiperparámetros hasta tener el valor óptimo de todos ellos para la señal elegida. En el apartado de resultados se mostrarán los valores óptimos para todos los hiperparámetros de cada señal y combinación de señales.

Una vez que se disponen de los hiperparámetros óptimos para cada una de las posibles combinaciones, debemos elegir el mejor modelo posible para cada una de ellas. Para poder elegirlo adecuadamente, crearemos una variable de monitoreo de las pérdidas de validación que fijaremos a infinito; y someteremos al modelo, ya con los hiperparámetros optimizados, a una gran cantidad de ciclos de entrenamiento. En cada uno de estos ciclos, compararemos el valor de las pérdidas de validación con el de la variable de monitoreo, en caso de ser mejor el valor obtenido, se actualiza el valor de la variable de monitoreo a ese nuevo valor y se guarda el modelo como el mejor modelo. Tras una serie de ciclos, obtendremos el mejor modelo posible de todos ellos para cada una de las señales.

Una vez tengamos el mejor modelo de cada una de las señales, podemos guardarlo y emplearlo cuando se necesite. Además, cargar un fragmento en un modelo ya entrenado es mucho más rápido y consume muchos menos recursos que la etapa de entrenamiento realizada con anterioridad.

Con el modelo inicial y el generador de datos, la siguiente etapa consiste en implementar un sistema con el que recoger los resultados y analizarlos para comprobar la evolución del modelo. Para ello, se han utilizado gráficas *scatter plot*, Bland-Altman y de la curva de aprendizaje del modelo (Bland & Altman, 1986; Hicks et al., 2022). Además, las librerías empleadas para el desarrollo del propio modelo permiten la obtención de resultados de la fase de entrenamiento mediante un historial de monitoreo de las variables. Las distintas gráficas permitirán comprobar de forma directa si el modelo está comportándose de la manera esperada o por el contrario está cayendo en el subajuste o sobreajuste. Por otro lado, los historiales permitirán analizar más en profundidad los resultados del entrenamiento y buscar posibles fallos o puntos de mejora en el modelo.

En primer lugar, las curvas de aprendizaje son gráficas que permiten realizar un seguimiento del rendimiento del modelo durante el entrenamiento (Aggarwal, 2018b; Kelleher, 2019). Para ello, se representan las curvas de pérdidas en entrenamiento (*train_loss*) y en validación (*val_loss*) de manera simultánea. Estas gráficas por lo general tienen una tendencia decreciente ya que salvo que el modelo esté mal diseñado, conforme avanzan las épocas del entrenamiento el modelo debe disminuir el valor de las pérdidas (Aggarwal, 2018b; Goodfellow et al., 2016; Kelleher, 2019).

No obstante, debemos tener en cuenta que la relación entre ambas curvas también nos puede dar mucha información, la separación entre ellas nos puede indicar si el modelo está comportándose adecuadamente o no. En caso de que las curvas estén muy separadas, hay claros indicios de que el modelo está cometiendo sobreajuste ya que los errores cometidos con los datos de entrenamiento son significativamente inferiores a los cometidos en los datos de validación (Goodfellow et al., 2016; Kelleher, 2019). Por otro lado, en caso de que las curvas sean similares pero los errores cometidos sean muy elevados (es decir, muy separadas del eje de origen), es indicador de que existe un problema de subajuste, es decir, el modelo no es capaz de aprender a partir de los datos proporcionados (Goodfellow et al., 2016; Kelleher, 2019). En un caso ideal, ambas curvas son próximas, disminuyen de igual manera y están próximas al eje de origen. Además, también nos pueden indicar si el modelo se ha entrenado durante un número suficiente de épocas. Para ello, debemos de observar la tendencia de las curvas: si vemos que decrecen pero no llegan a estabilizarse el modelo requiere de un mayor número de épocas, ya que el hecho de no terminar de alcanzar una etapa más plana indica que aún puede haber margen de mejora; en el caso contrario, si se mantiene plano mucho tiempo, la cantidad de épocas es excesiva, consumiendo recursos innecesariamente (Aggarwal, 2018b; Kelleher, 2019). En la Figura 4-9 vemos un ejemplo de estas curvas.

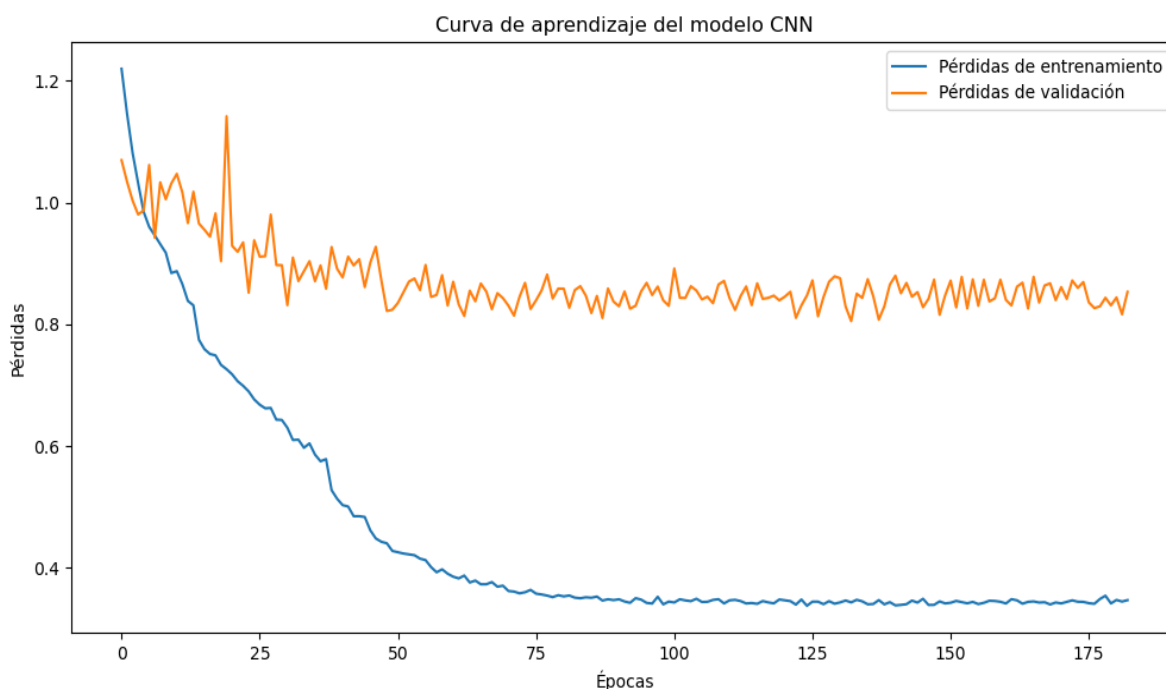


Figura 4-9. Ejemplo de una curva de aprendizaje del modelo CNN (Autoría propia).

Por otra parte, también resulta interesante emplear las gráficas *scatter plot* y Bland-Altman para observar el comportamiento del modelo desde otra perspectiva distinta. En el caso de las gráficas *scatter plot*, simplemente hemos enfrentado los puntos reales (y) y estimados (x) en color rojo frente a una recta $x = y$ en color azul, siendo las zonas donde mayor superposición entre estos indicadores de buen comportamiento del modelo. Si bien la idea es similar, en los gráficos de Bland-Altman existen una serie de diferencias. En primer lugar, en lugar de situar los valores reales y estimados en los ejes X e Y , se fija el promedio de las mediciones entre los puntos en el eje X y la diferencia en el eje Y . A continuación, se calcula la media y los límites de acuerdo ($\text{media} \pm 1.96 * \text{std}$) y se añaden al gráfico de forma horizontal. Si las diferencias están cercanas a 0 significa que el modelo tiene una elevada exactitud en las predicciones, si por el contrario existe una tendencia de crecimiento o decrecimiento, puede existir la posibilidad de un sesgo y por lo tanto podemos detectar si el modelo infraestima o sobreestima el número de eventos o el IAH (Altman & Bland, 1983; Bland & Altman, 1986). En la Figura 4-10 podemos ver un ejemplo de este tipo de gráficas.

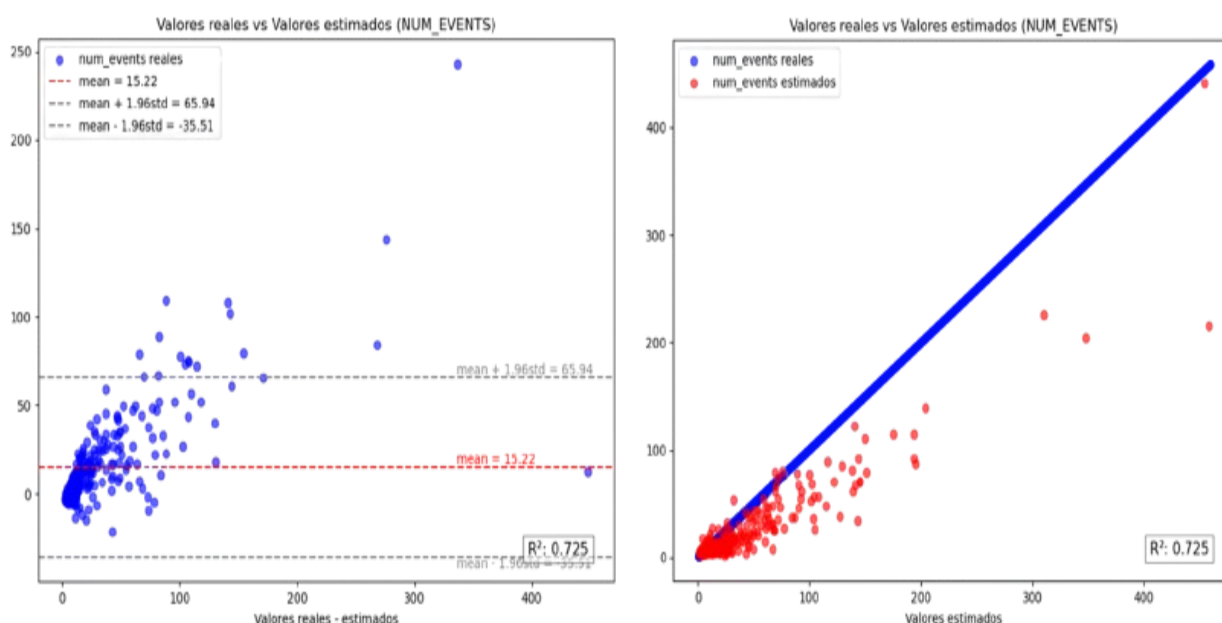


Figura 4-10. Ejemplo de gráfica Bland-Altman (izquierda) y gráfica Scatter (derecha) del modelo CNN (Autoría propia).

Además, se ha empleado coeficiente de determinación R^2_score (Ecuación 4.3) para poder evaluar la concordancia de los resultados obtenidos con las medidas reales. Debemos tener en cuenta que cuanto más próximo sea el valor del R^2_score a 1, mejor será la exactitud del modelo. Esta métrica se ha añadido a todas las gráficas con el fin de aumentar la cantidad de información que proporcionan. La razón por la que se utiliza el R^2_score en lugar de otras métricas como el *Intraclass Correlation Coefficient* (ICC) se debe, en parte, a la mayor simplicidad en el cálculo e interpretación del R^2_score . Además, el ICC está diseñado para evaluar la fiabilidad entre observadores o entre múltiples mediciones de un mismo sujeto, por lo que no resulta tan adecuado en contextos de regresión supervisada estándar. En este caso, el uso del R^2_score es más apropiado y relevante (Chicco & Jurman, 2020; Koo & Li, 2016).

$$R^2_{score} = 1 - \frac{Residual_{square}}{Total_{square}} \quad (4.3)$$

Finalmente, hemos configurado el historial del modelo para que nos permita obtener información de las variables: época del modelo, MAE, MSE, pérdidas de entrenamiento y pérdidas de validación. Gracias a la información proporcionada por dicho fichero, podemos observar con más precisión lo que ocurre en el modelo, identificar si hay fallos en la estructura general, si se han dado comportamientos anormales en alguna época o comparar si el desempeño durante el entrenamiento y la validación funciona acorde a lo esperado.

4.4.3. Estimación del IAH

Con el sistema previamente desarrollado y habiendo elegido el mejor modelo para cada señal, es posible estimar la cantidad de eventos de apnea/hipopnea presentes en cada fragmento. No obstante, nuestro objetivo es poder estimar el valor del IAH en cada sujeto a través del número total de eventos en el registro nocturno. Para ello, a partir de la estimación del número de eventos por cada segmento de 20 min y la duración del registro nocturno del paciente debemos trabajar en calcular el valor del IAH. Lo primero que debemos tener en cuenta, es que el valor del IAH será asignado a un único sujeto, por lo que se han analizado con el modelo todos los fragmentos de 20 min de un mismo sujeto y se han sumado los valores de número de eventos estimados obtenidos en cada uno de los fragmentos. De esta manera, obtendremos el número de eventos totales estimado para un único usuario y a partir de ahí podremos obtener el IAH correspondiente a ese sujeto.

Otro de los factores que debemos tener en cuenta es la diferencia entre el tiempo de grabación y el tiempo de sueño del paciente. El tiempo de sueño corresponde al período en el que la persona está dormida y es calculado durante la PSG mediante la contabilización de los ciclos de sueño a través del EEG, mientras que el tiempo de grabación por otro lado, abarca desde la colocación de los sensores en el paciente durante la PSG hasta su retirada al final de la noche. Sin embargo, el marcaje de las fases del sueño y el posterior cómputo del tiempo total de sueño no está disponible empleando las señales y el modelo de este trabajo. Deberemos utilizar el tiempo total de grabación, teniendo en cuenta que la duración del registro es mayor que el tiempo total de sueño (al abarcar periodos en los que el sujeto está despierto y periodos en los que está durmiendo). En la Figura 4-11, podemos observar cómo varían los valores de las señales en los primeros momentos correspondiente a la colocación de los sensores, en los que el paciente presumiblemente continúa despierto.

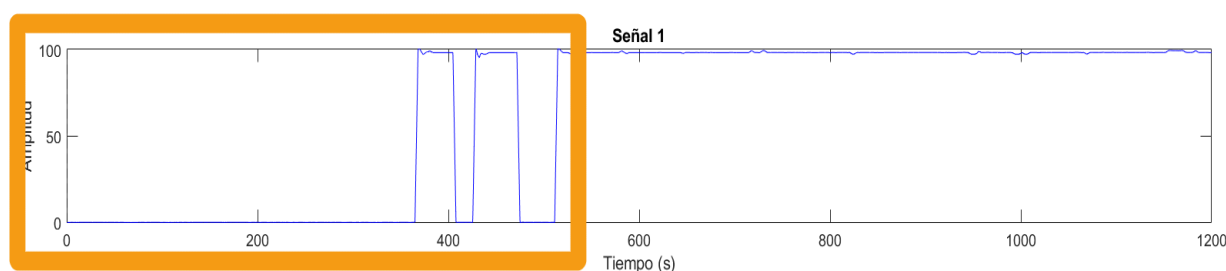


Figura 4-11. Primer fragmento de un sujeto, señal SpO_2 (Autoría propia, elaborado con Matlab).



Figura 4-12. Ejemplo de regresión Huber con la señal SpO_2 .

Por lo tanto, debemos estimar el IAH teniendo en cuenta esta limitación. En primer lugar, se ha calculado el tiempo de grabación por sujeto simplemente empleando el número de fragmentos de 20 min convertidos a eventos por hora, pero como ya se ha mencionado, este valor incluye tiempo en el que el sujeto permanece despierto. Así que se ha calculado la tasa de eventos detectados durante el tiempo de grabación, para luego poder aplicar un sistema de regresión del IAH basado en Huber destinado a obtener una estimación precisa del IAH teniendo en cuenta las limitaciones mencionadas anteriormente (Figura 4-12). De esta manera, podemos tratar de corregir la infraestimación del IAH originado al emplear el tiempo de grabación en lugar del tiempo de sueño del paciente. Esto permite obtener una estimación más precisa del valor de IAH y por lo tanto, permite realizar una mejor umbralización para clasificar a los sujetos en los distintos grados de AOS infantil.

4.5. Análisis estadístico

Para evaluar el rendimiento del modelo de predicción, es fundamental llevar a cabo un análisis estadístico que permita evaluar el desempeño de nuestro sistema. Para ello, se han empleado matrices de confusión. La matriz de confusión es una tabla que permite evaluar el rendimiento de modelos de clasificación, separando aquellas predicciones que se hayan realizado correctamente de aquellas que son erróneas. Para realizar esta separación, la matriz de confusión se fundamenta en el uso de la estructura de la Tabla 4-1 empleando los valores reales y estimados.

En nuestro caso obtendremos 6 matrices de confusión para cada una de las posibles combinaciones de señales: dos matrices para el conjunto de datos de entrenamiento, dos para el de validación y dos para el de test, podemos ver un ejemplo de una de ellas en la Figura 4-13. Para cada uno de los conjuntos de señales, hemos calculado la matriz de confusión de los 4 niveles de severidad obtenidos a partir del IAH final estimado. Generar estas matrices permite realizar un mejor seguimiento del comportamiento del algoritmo, ya que podemos observar los resultados en los 3 subconjuntos de datos (entrenamiento, validación y test) para controlar que no exista subajuste ni sobreajuste en el modelo. No obstante, los resultados que mayor importancia tienen son aquellos referentes al conjunto de test, ya que son aquellos que determinan cómo de bien se comporta el modelo con datos nuevos en una prueba final.

	ESTIMADO POSITIVO	ESTIMADO NEGATIVO
REAL POSITIVO	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
REAL NEGATIVO	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

Tabla 4-1. Estructura general de una matriz de confusión.

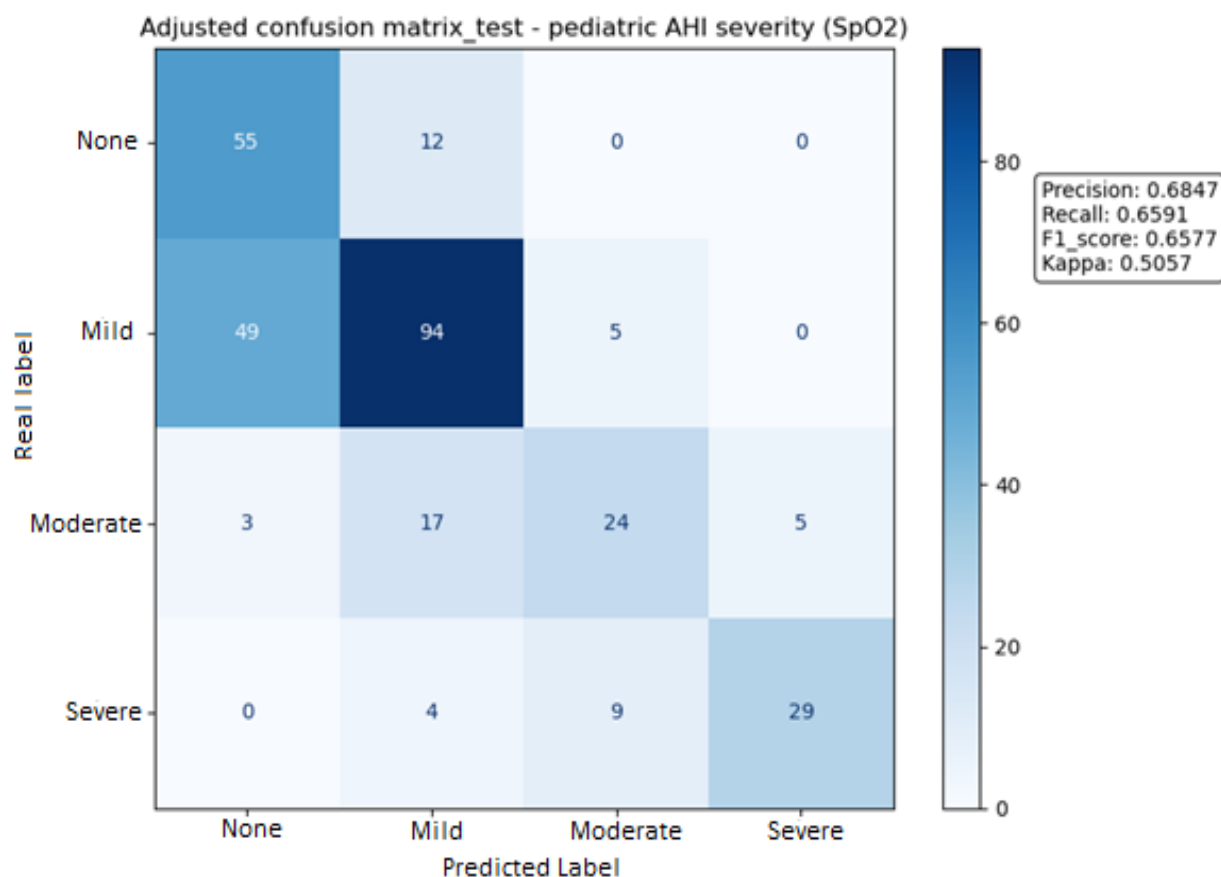


Figura 4-13. Ejemplo de matriz de confusión de valores de IAH de la señal SpO₂ en test tras someterla a la regresión Huber.

Debemos tener en cuenta que cada sujeto puede estar clasificado en 1 de las 4 clases existentes: Sin AOS, AOS Leve, Moderada y Grave. Las matrices de confusión, además de aportar una visión directa del funcionamiento del modelo, permiten generar y obtener métricas de gran utilidad a la hora de realizar el análisis estadístico. Por lo que, aprovechando las matrices de confusión obtenidas, vamos a calcular las siguientes métricas a partir de los datos proporcionados:

- **Exactitud (*Acc*):** también conocida como *Accuracy*, esta métrica permite calcular la relación entre los sujetos adecuadamente clasificados y el total, comparando los dos casos correctamente clasificados con todas las clases clasificadas de la matriz (Ecuación 4.2):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.2)$$

- **Sensibilidad (*Se*):** también conocida como *Recall*, esta métrica permite medir la capacidad del modelo para detectar correctamente los casos positivos. Para ello, compara las estimaciones clasificadas correctamente frente a todos los casos positivos (Ecuación 4.3):

$$Se = \frac{TP}{TP + FN} \quad (4.3)$$

- **Especificidad (*Sp*):** de manera opuesta a la sensibilidad, esta métrica permite medir la capacidad del modelo para detectar correctamente los casos negativos (Ecuación 4.4):

$$Sp = \frac{TN}{TN + FP} \quad (4.4)$$

- **Valor predictivo positivo (PPV):** indica la probabilidad de que un caso detectado como positivo sea realmente positivo (en textos en inglés también se conoce como *precision*) (Ecuación 4.5):

$$PPV(\%) = \frac{TP}{TP + FP} * 100 \quad (4.5)$$

- **Valor predictivo negativo (NPV):** de manera similar al caso anterior, indica la probabilidad de que un caso detectado como negativo sea realmente negativo (Ecuación 4.6):

$$NPV(\%) = \frac{TN}{TN + FN} * 100 \quad (4.6)$$

- **Likelihood Ratio Positivo (LR+):** consiste en un ratio entre las probabilidades de que un caso detectado como positivo sea realmente positivo, frente a que un caso detectado como negativo sea en realidad un caso positivo (Ecuación 4.7):

$$LR+ = \frac{Se}{1 - Sp} \quad (4.7)$$

- **Likelihood Ratio Negativo (LR-):** consiste en un ratio entre las probabilidades de que un caso detectado como negativo sea realmente negativo, frente a que un caso detectado como positivo sea en realidad un caso negativo (Ecuación 4.8):

$$LR- = \frac{1 - Se}{Sp} \quad (4.8)$$

- **Puntuación F1 (F1_score):** esta métrica es ampliamente utilizada ya que permite considerar tanto el PPV como la Se, se calcula mediante la media armónica de ambas (Ecuación 4.9):

$$F1_{score} = 2 * \frac{VPP * Se}{VPP + Se} \quad (4.9)$$

- **Cohen's Kappa (Kappa):** evalúa la concordancia entre las distintas clasificaciones y ajusta la probabilidad de coincidencia aleatoria, es realmente útil en problemas con varias clases ya que tiene en cuenta el factor de coincidencia al azar (McHugh, 2012) (Nota: α_{ij} implica el número de veces que la clase real i fue estimada por la clase j) (Ecuación 4.10):

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \quad (4.10)$$

$$P_0 = \frac{\sum_{i=1}^{n_{casos}} \alpha_{ii}}{Total\ de\ casos} \quad (4.11)$$

$$P_e = \sum_{i=1}^4 \left(\frac{\sum_{j=1}^{n_{casos}} \alpha_{ij}}{Total\ de\ casos} * \frac{\sum_{j=1}^{n_{casos}} \alpha_{ji}}{Total\ de\ casos} \right) \quad (4.12)$$

Estas métricas permiten evaluar el comportamiento del modelo. Aplicándolas a los resultados obtenidos con cada una de las señales (individuales o combinadas), podemos realizar una comparación directa del desempeño de cada una y clasificarlas por lo tanto en función de su rendimiento; permitiendo así elegir la mejor combinación de cara a realizar una predicción lo más precisa posible de la AOS infantil.

En este estudio, el objetivo consiste en catalogar a los sujetos pediátricos en función de la presencia o no del trastorno de AOS y en caso de padecerla, clasificarla según sus distintos grados de severidad. Por lo que redefiniremos los valores de la matriz (TN, TP, FP y FN) según los 3 umbrales establecidos para la AOS (leve, moderada y severa).

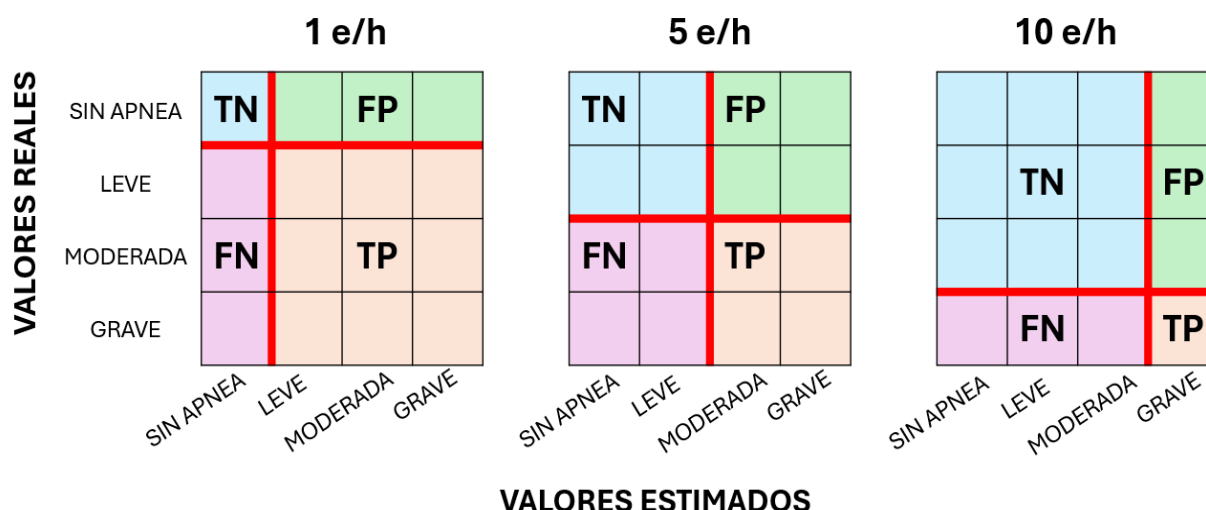


Figura 4-14. Esquema umbralización progresiva de la matriz de confusión (Autoría propia).

Este proceso se conoce como umbralizado progresivo, y su propósito es reducir la matriz de confusión de 4x4 para convertir un problema de clasificación multiclase en un problema de clasificación binaria con una matriz de confusión 2x2. Para poder realizar esta clasificación, agruparemos todos los casos en los que el IAH real y estimado es menor o mayor que cierto umbral, es decir, trataremos de trabajar con 3 versiones de una misma matriz en función del umbral que se desea evaluar. En la Figura 4-14, se puede ver un esquema del procedimiento.

4.6. Especificaciones técnicas

A lo largo de este apartado, resumiremos todas las especificaciones de los recursos computacionales empleados tanto a nivel de hardware como software. En primer lugar, debemos resaltar que el sistema ha sido desarrollado por completo en el lenguaje Python. Este lenguaje de programación de alto nivel permite el uso de una sintaxis sencilla y gran versatilidad. Python es ampliamente utilizado en el desarrollo de sistemas de IA debido a la gran cantidad de librerías que existen relacionadas con ese campo. En nuestro caso, hemos empleado el entorno de desarrollo integrado PyCharm (JetBrains, s.r.o) para la elaboración de nuestro sistema, ya que ofrece una gran cantidad de herramientas tanto para el desarrollo como la depuración. Además, hemos optado por elegir el gestor de entornos y paquetes de Anaconda (Anaconda Inc.). El sistema operativo del ordenador es Microsoft Windows 11 (Tabla 4-2).

Programa/Librería	Versión
PyCharm	2024.3.2
Anaconda Navigator	2.5.1
keras	2.10.0
matplotlib	3.8.2
numpy	1.26.4
pandas	2.2.1
plotly	5.24.1
scikit-learn	1.4.2
scipy	1.13.0
tensorflow	2.10.1

Tabla 4-2. Librerías empleadas y sus versiones.

En cuanto a los recursos de computación del ordenador, en la Tabla 4-3 se resumen las características del equipo:

Componente	Características
CPU	13th Gen Intel® Core™ i9-13900KF
GPU	NVIDIA GeForce RTX 4090 (24 GB)
RAM	64 GB DDR4
SSD	4 TB SSD Kingston
Sistema Operativo	Microsoft Windows 11 Pro (x64)

Tabla 4-3. Características del hardware empleado.

Capítulo 5: RESULTADOS

En este quinto capítulo, revisaremos y analizaremos los resultados obtenidos a lo largo del desarrollo del sistema descrito en el capítulo previo. Para ello, presentaremos los resultados del proceso de optimización de los hiperparámetros del modelo para cada una de las combinaciones de señales, de las estimaciones del IAH realizadas por el propio modelo, de las predicciones de los 4 niveles de severidad de la AOS y finalmente de la capacidad de diagnóstico de la AOS en cada uno de los umbrales (1, 5 y 10 e/h).

5.1. Optimización del modelo

En primer lugar, revisaremos los resultados del proceso de optimización de los hiperparámetros del modelo para cada una de las señales. Debemos recordar que para cada señal y sus posibles combinaciones, se ha realizado un proceso de optimización para todos los hiperparámetros existentes. Para cada uno de estos hiperparámetros, recogeremos los valores de *Val_loss* y *R2_score* y elegiremos aquel que tenga los mejores resultados. A continuación, mostramos una de las tablas empleadas para elegir el mejor resultado de un hiperparámetro (resaltado en negrita) a modo de ejemplo, debemos tener en cuenta que por cada señal e hiperparámetro se ha realizado una tabla similar y en algunos casos 2 para poder obtener un valor de hiperparámetro más preciso (Tablas 5-1 y 5-2):

<i>R2_score</i>	8	16	24	32	64	128
1. Iteración	0.7412	0.7538	0.7770	0.7681	0.7953	0.7936
2. Iteración	0.7557	0.7651	0.7853	0.7934	0.7833	0.7945
3. Iteración	0.7629	0.7827	0.7544	0.7778	0.7935	0.7904
4. Iteración	0.7365	0.7788	0.7822	0.7761	0.7923	0.7807

Tabla 5-1. Tabla de resultados de optimización del hiperparámetro *num_filters* para la señal *SpO₂* (*R2_score*).

<i>Val_loss</i>	8	16	24	32	64	128
1. Iteración	0.5175	0.4963	0.4839	0.4783	0.4618	0.4645
2. Iteración	0.5044	0.4896	0.4715	0.4669	0.4717	0.4724
3. Iteración	0.4809	0.4744	0.4985	0.4784	0.4692	0.4769
4. Iteración	0.4966	0.4781	0.4728	0.4844	0.4700	0.4709

Tabla 5-2. Tabla de resultados de optimización del hiperparámetro *num_filters* para la señal *SpO₂* (*Val_loss*).

RANGO DE VALORES						
<i>num_filters</i>	8	16	24	32	64	128
<i>num_blocks</i>	3	4	5	6	7	8
<i>size_filters</i>	5	7	9	17	21	27
<i>batch_size</i>	64	128	256	512	1024	2048
<i>batch_size</i> (iter. 2)	18	24	32	48	64	80
<i>learning_rate</i>	0.00001	0.0005	0.0001	0.005	0.001	0.01
<i>learning_rate</i> (iter. 2)	0.001	0.002	0.004	0.006	0.008	0.01

Tabla 5-3. Rango de valores empleados para la optimización de los distintos hiperparámetros para cada una de las combinaciones de señales.

Como se puede apreciar, de toda la tabla se ha elegido el mejor valor (resaltado en negrita). En el caso de la tabla de ejemplo se optó por elegir el valor de 64 para el hiperparámetro *num_filters* en la señal SpO₂ debido a que es el valor que ha obtenido en mayor número de ocasiones el mejor valor de toda la iteración. Para todos los hiperparámetros, analizaremos el siguiente rango de valores con todas las señales. En algunos casos, los rangos iniciales de los hiperparámetros *learning_rate* y *batch_size* no cubrían el valor deseado y se realizó una segunda etapa de optimización (iter. 2 en tablas) con otro rango distinto para encontrar un valor mucho más preciso para ambos hiperparámetros (Tabla 5-3).

Para las señales elegidas de manera individual (SpO₂, FA y HRV), el resultado de la optimización de los hiperparámetros es el siguiente (Tabla 5-4):

	SpO ₂	FA	HRV
<i>num_filters</i>	64	32	32
<i>num_blocks</i>	7	8	8
<i>size_filters</i>	17	9	9
<i>batch_size</i>	512	24	256
<i>learning_rate</i>	0.008	0.004	0.008

Tabla 5-4. Valores de hiperparámetros óptimos para las señales individuales SpO₂, FA y HRV.

Por otro lado, en el caso de las señales emparejadas 2 a 2 (SpO₂-FA, SpO₂-HRV y FA-HRV), el resultado de la optimización de los valores de los hiperparámetros es el siguiente (Tabla 5-5):

	SpO ₂ -FA	SpO ₂ -HRV	FA-HRV
<i>num_filters</i>	32	64	128
<i>num_blocks</i>	8	7	7
<i>size_filters</i>	17	13	15
<i>batch_size</i>	18	128	28
<i>learning_rate</i>	0.001	0.002	0.006

Tabla 5-5. Valores de hiperparámetros óptimos para las señales emparejadas 2 a 2 SpO₂-FA, SpO₂-HRV y FA-HRV.

Finalmente, para las 3 señales combinadas (SpO₂-FA-HRV), el resultado de la optimización de los valores de los hiperparámetros es la siguiente (Tabla 5-6):

SpO ₂ -FA-HRV	
<i>num_filters</i>	32
<i>num_blocks</i>	6
<i>size_filters</i>	17
<i>batch_size</i>	64
<i>learning_rate</i>	0.004

Tabla 5-6. Valores de hiperparámetros óptimos para todas señales emparejadas SpO₂-FA-HRV.

El resultado de la etapa de optimización permitió obtener mejoras significativas en el entrenamiento del modelo CNN tanto en las pérdidas de validación como en el valor de *R2_score*. A continuación, en la Tabla 5-7 se recoge el valor medio de las 4 iteraciones realizadas previas a la etapa de optimización y tras dicho proceso, teniendo en cuenta que se han elegido los valores de los hiperparámetros previamente mostrados en las Tablas 5-4, 5-5 y 5-6:

	Inicio optimización (valor medio)		Final optimización (valor medio)		Final optimización (mejor resultado)	
	<i>Val_loss</i>	<i>R2_score</i>	<i>Val_loss</i>	<i>R2_score</i>	<i>Val_loss</i>	<i>R2_score</i>
SpO ₂	0.4670	0.7939	0.4539	0.7912	0.4513	0.7928
FA	1.0298	0.0371	0.7731	0.4466	0.7668	0.4795
HRV	0.8553	0.4536	0.7647	0.5509	0.7909	0.5578
SpO ₂ -FA	0.6428	0.5356	0.5397	0.6747	0.5278	0.6925
SpO ₂ -HRV	0.4356	0.8256	0.4214	0.8328	0.4196	0.8410
FA-HRV	0.8232	0.4542	0.7376	0.5466	0.7287	0.5573
SpO ₂ -FA-HRV	0.4587	0.7713	0.4285	0.8156	0.4219	0.8239

Tabla 5-7. Resultados del proceso de optimización en las variables *Val_loss* y *R2_score*.

Una vez realizado el proceso de optimización, iteraremos una gran cantidad de veces (entre 20 y 40) el entrenamiento del modelo e iremos guardando aquellos que consigan mejorar el valor *Val_loss*. Para todos los modelos se ha realizado el procedimiento con el valor de 0.1 para la tasa de *Dropout*, aunque también se ha probado con un valor de 0.2 en algunos casos, buscando así un ajuste más fino (Tabla 5-8).

	<i>Val_loss</i>		<i>R2_score</i>	
	<i>Dropout 0.1</i>	<i>Dropout 0.2</i>	<i>Dropout 0.1</i>	<i>Dropout 0.2</i>
SpO ₂	0.4527	0.4542	0.7915	0.7879
FA	0.7547	0.7791	0.4901	0.4236
HRV	0.7603	0.7622	0.5288	0.5447
SpO ₂ -FA	0.5143	0.5266	0.7003	0.6952
SpO ₂ -HRV	0.4123	0.4091	0.8331	0.8373
FA-HRV	0.7236	0.7249	0.5503	0.5602
SpO ₂ -FA-HRV	0.4182	0.4256	0.8094	0.8054

Tabla 5-8. Resultados de *Val_loss* y *R2_score* obtenidos en el mejor modelo de cada una de las señales para los valores de *Dropout* 0.1 y 0.2.

5.2. Rendimiento diagnóstico del modelo

En este apartado, presentaremos los resultados obtenidos de emplear el modelo desarrollado para estimar el número de eventos y con ello, el valor del IAH y su respectiva severidad en la AOS infantil. En primer lugar, se presentan los resultados de realizar la regresión del IAH a partir del número de eventos en cada segmento. A continuación, se muestran los resultados de la clasificación de los IAH obtenidos en cada una de las 4 clases de severidad de la AOS mediante matrices de confusión. Finalmente, mostraremos la capacidad diagnóstica de nuestros algoritmos de DL para detectar la AOS infantil en cada umbral.

5.2.1. Regresión del IAH

Tras obtener el modelo entrenado, validado y testeado con los datos de los registros, se extrajeron 2 gráficas por cada combinación de señales: una gráfica *scatter plot* y una gráfica Bland-Altman. En las Figuras 5-1, 5-2, 5-3, 5-4, 5-5, 5-6 y 5-7 podemos ver todas las gráficas extraídas de estimar el número de eventos.

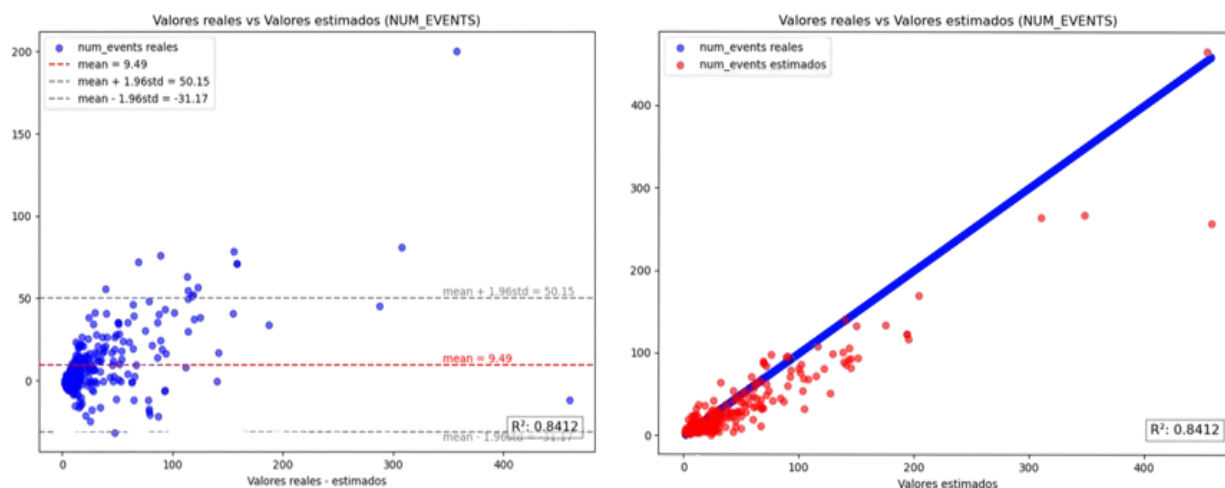


Figura 5-1. Gráficas Bland-Altman (izquierda) y *scatter plot* (derecha) de la estimación del número de eventos para la señal SpO₂ en el conjunto de test.

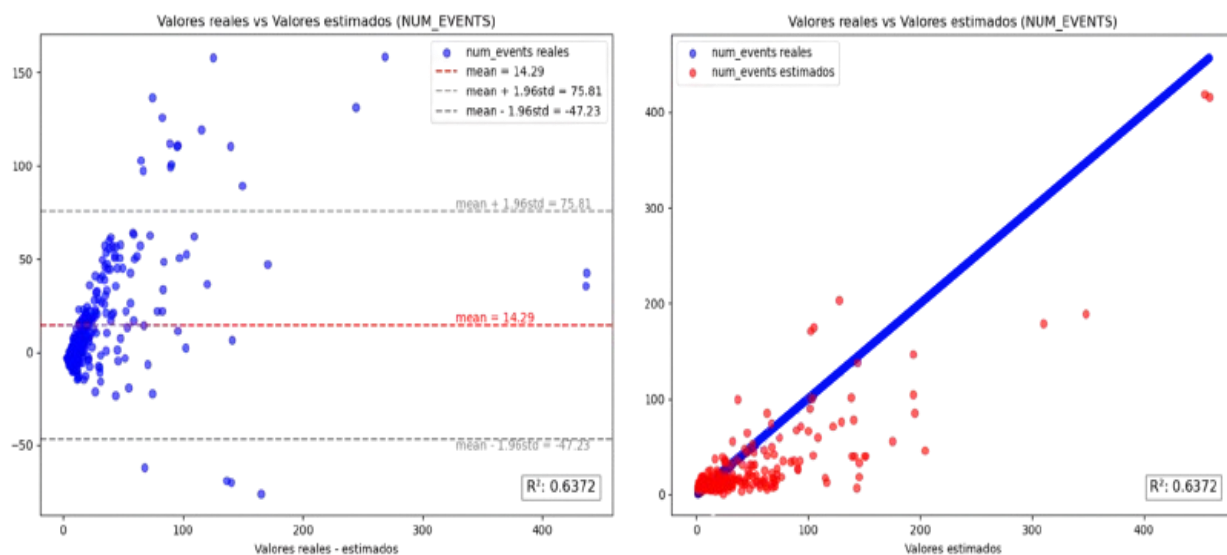


Figura 5-2. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del número de eventos para la señal FA en el conjunto de test.

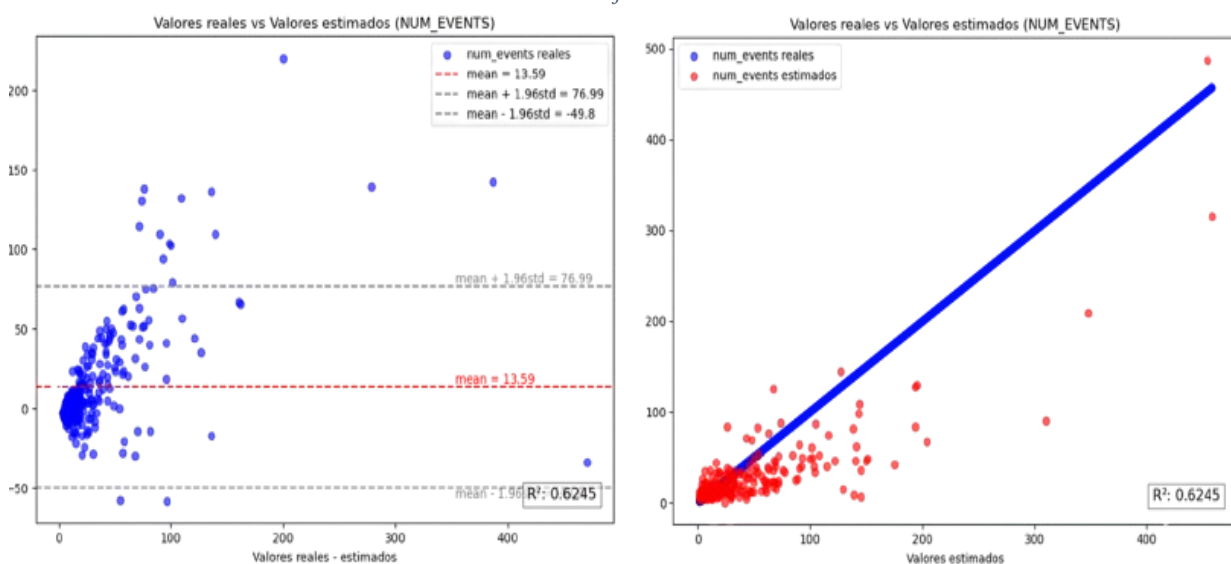


Figura 5-3. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del número de eventos para la señal HRV en el conjunto de test.

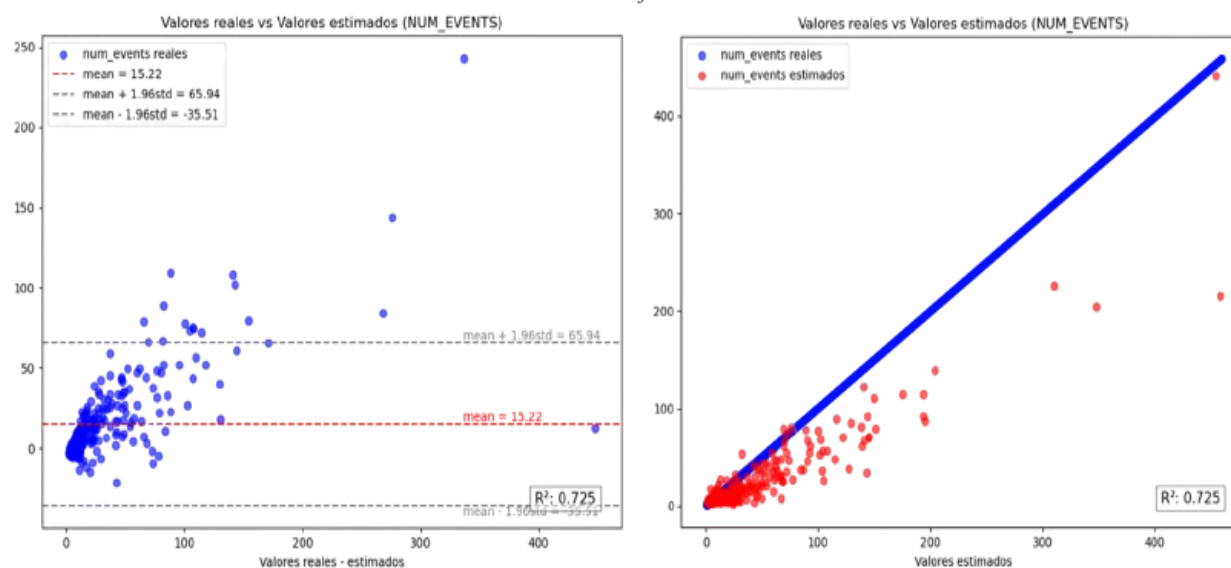


Figura 5-4. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del número de eventos para las señales SpO₂-FA en el conjunto de test.

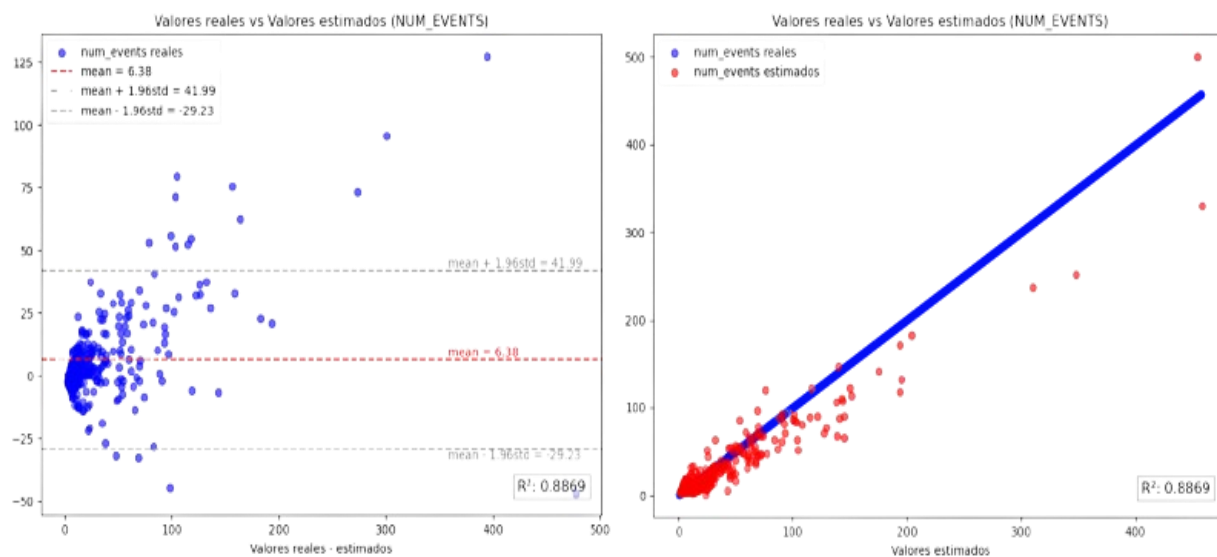


Figura 5-5. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del número de eventos para las señales SpO_2 -HRV en el conjunto de test.

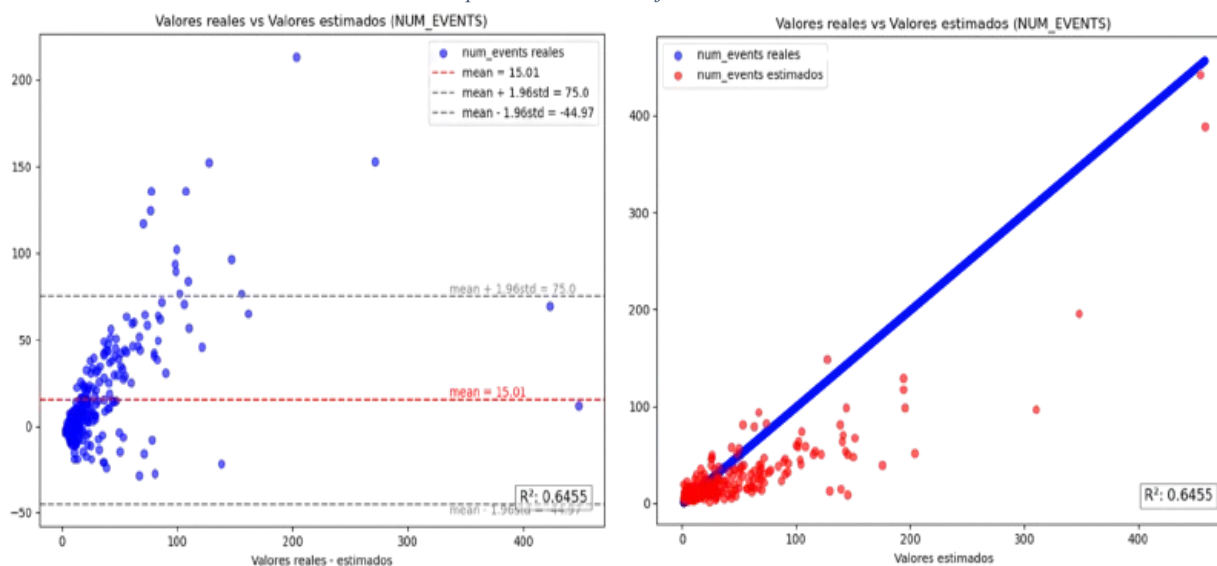


Figura 5-6. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del número de eventos para las señales FA-HRV en el conjunto de test.

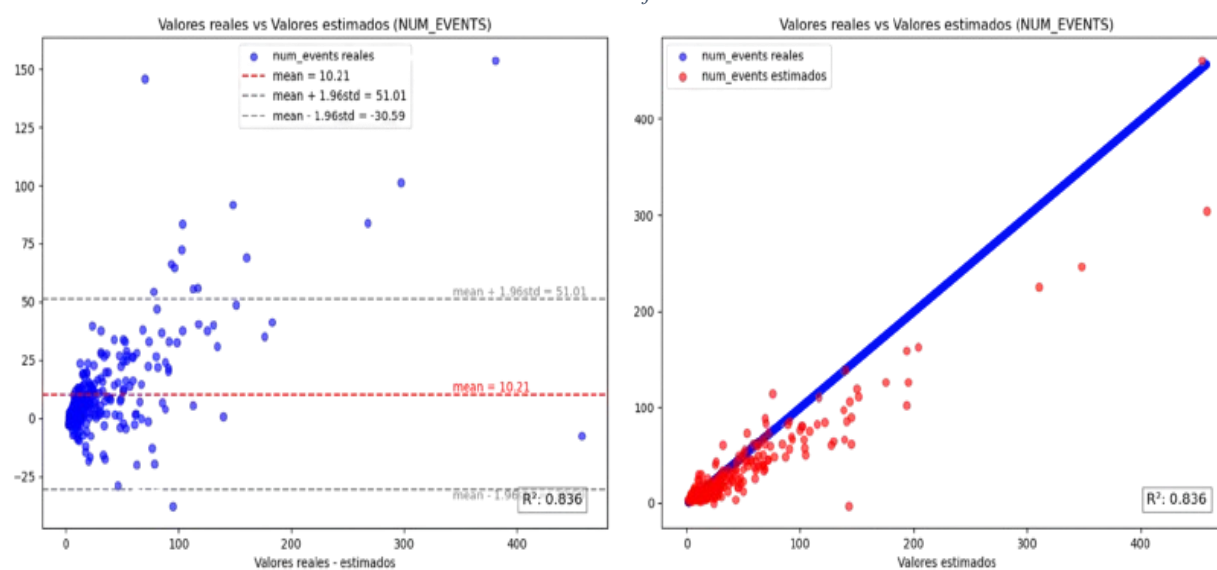


Figura 5-7. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del número de eventos para las señales SpO_2 -FA-HRV en el conjunto de test.

Estas gráficas permiten observar el resultado de las predicciones realizadas por el modelo. A continuación, mostraremos las gráficas de *scatter plot* y Bland-Altman para las predicciones del IAH en el conjunto de test, es decir, la regresión del IAH a partir del número de eventos detectados y el tiempo total de registro, realizada mediante el método de Huber (Figuras 5-8, 5-9, 5-10, 5-11, 5-12, 5-13 y 5-14).

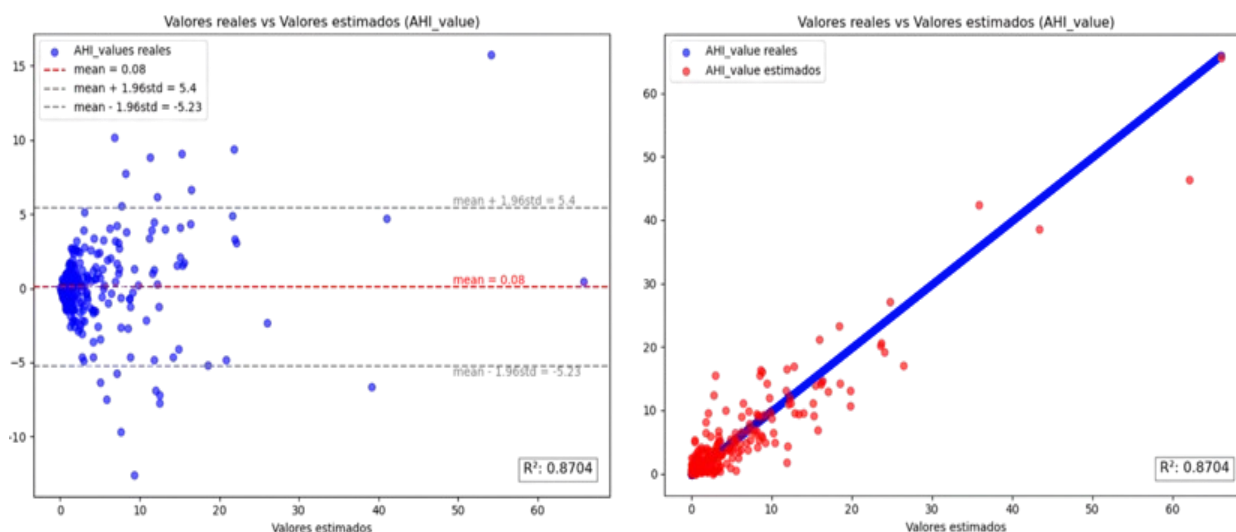


Figura 5-8. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del IAH para la señal SpO₂ en el conjunto de test.

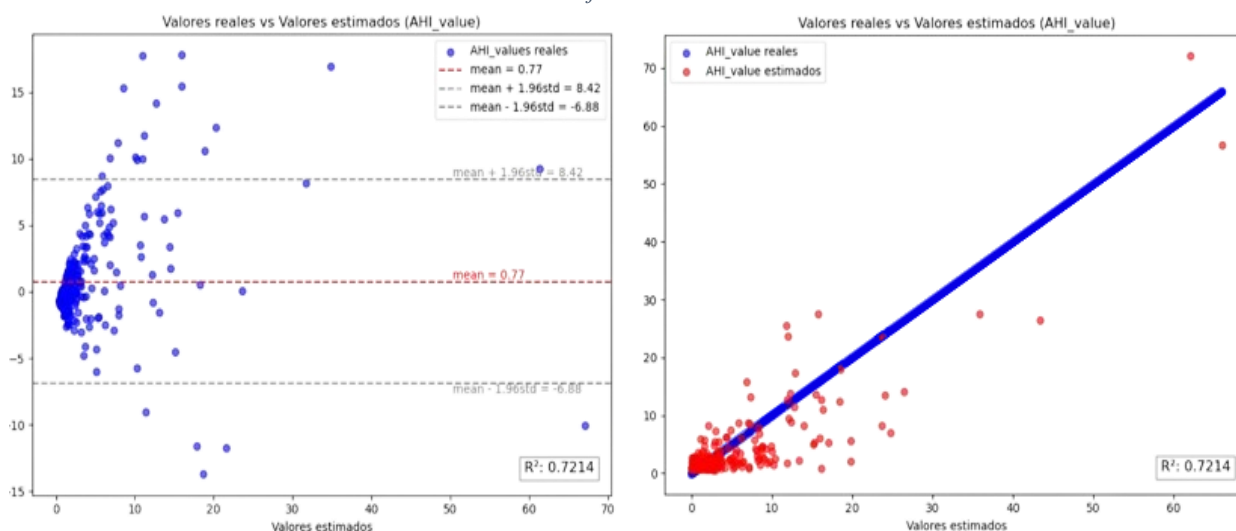


Figura 5-9. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del IAH para la señal FA en el conjunto de test.

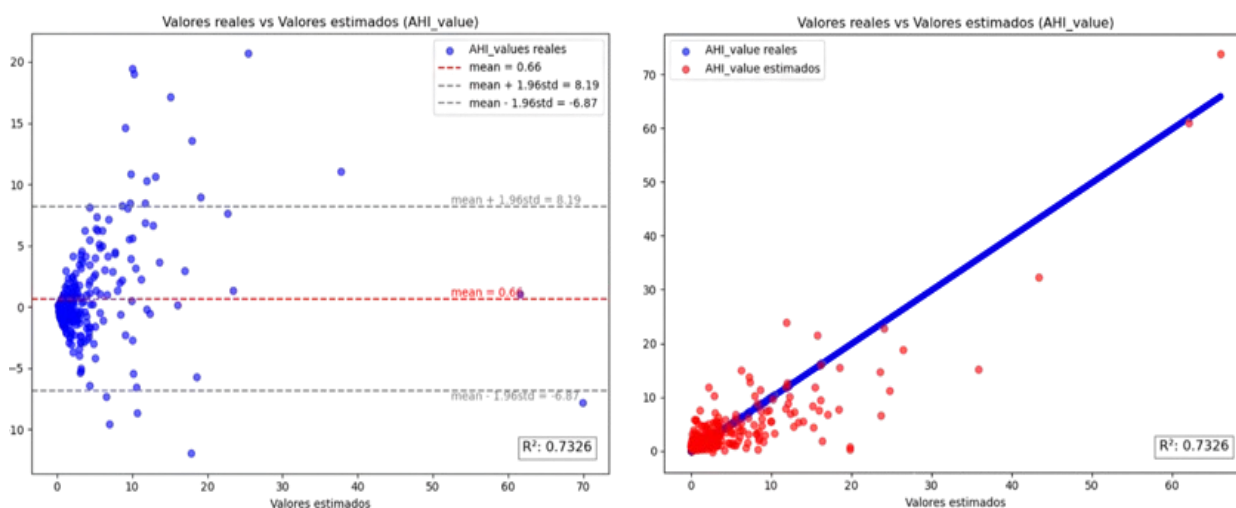


Figura 5-10. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del IAH para la señal HRV en el conjunto de test.

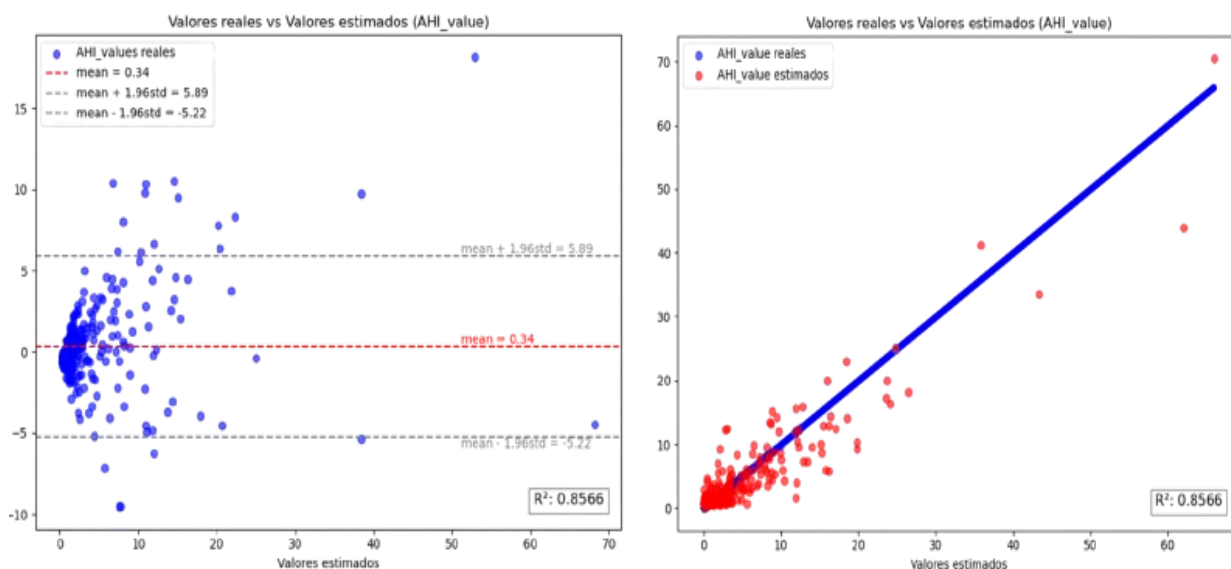


Figura 5-11. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del IAH para las señales SpO₂-FA en el conjunto de test.

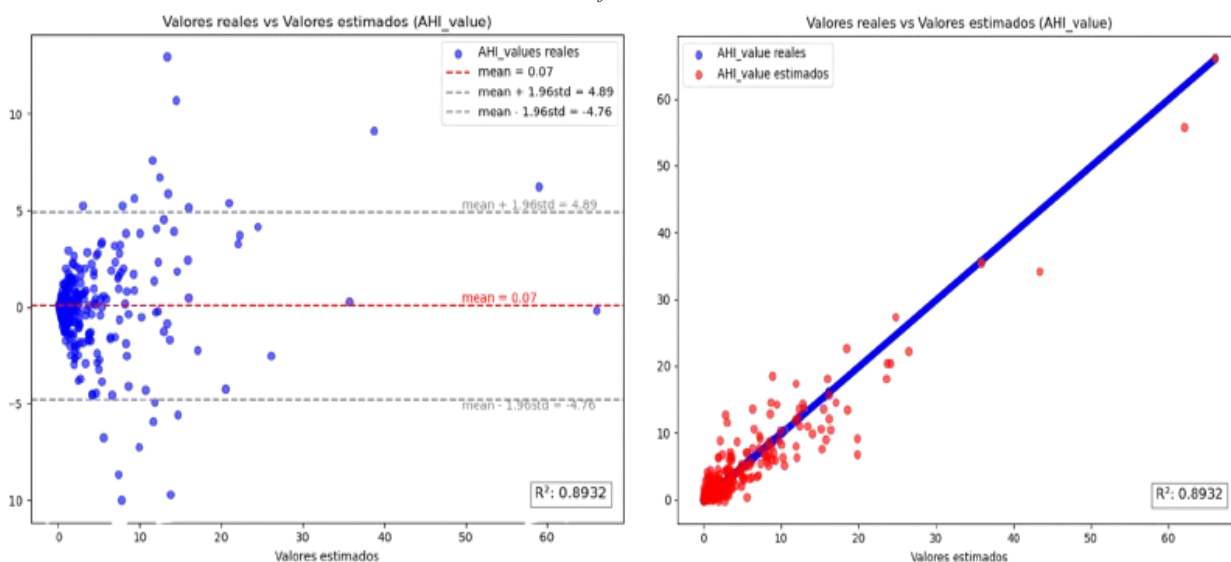


Figura 5-12. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del IAH para las señales SpO₂-HRV en el conjunto de test.

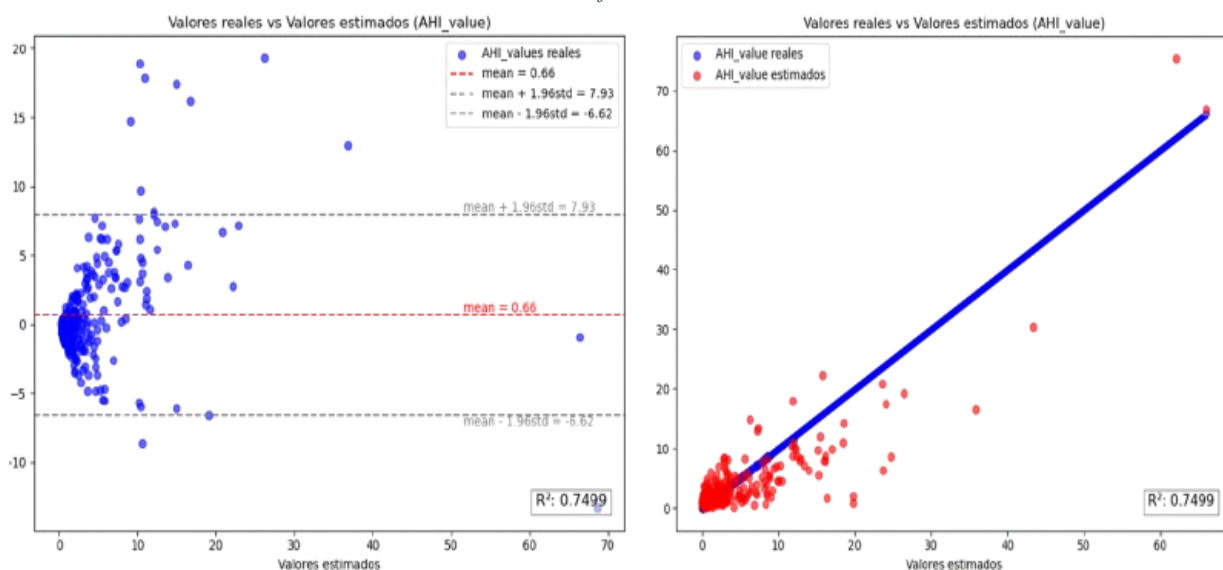


Figura 5-13. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del IAH para las señales FA-HRV en el conjunto de test.

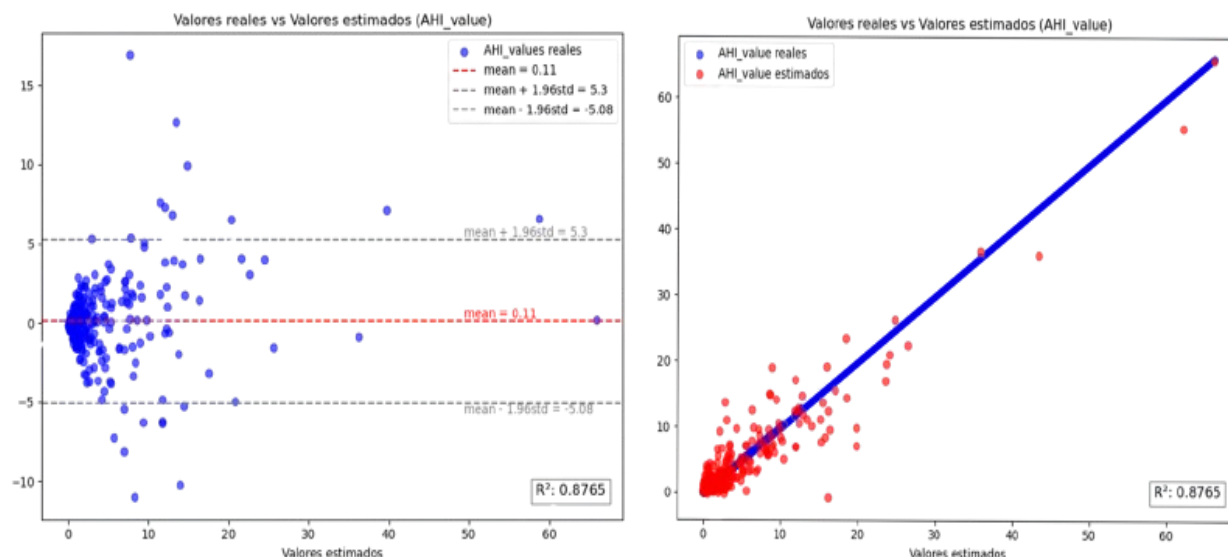


Figura 5-14. Gráficas Bland-Altman (izquierda) y scatter plot (derecha) de la estimación del IAH para las señales SpO_2 -FA-HRV en el conjunto de test.

5.2.2. Resultados de clasificación de 4 clases

Tras obtener la estimación del IAH final de todos los sujetos, podemos realizar una etapa de clasificación de cada uno de ellos en los distintos grados de severidad de la AOS: sin AOS, leve, moderado y grave. Para ello, estableceremos los umbrales (o puntos de corte) del IAH de 1, 5 y 10 e/h y etiquetaremos cada paciente acorde al valor obtenido. A continuación, mostramos las matrices de confusión resultantes de realizar dicha clasificación en el subconjunto de test con el sesgo corregido para cada una de las combinaciones (Figuras 5-15, 5-16, 5-17, 5-18, 5-19, 5-20 y 5-21):

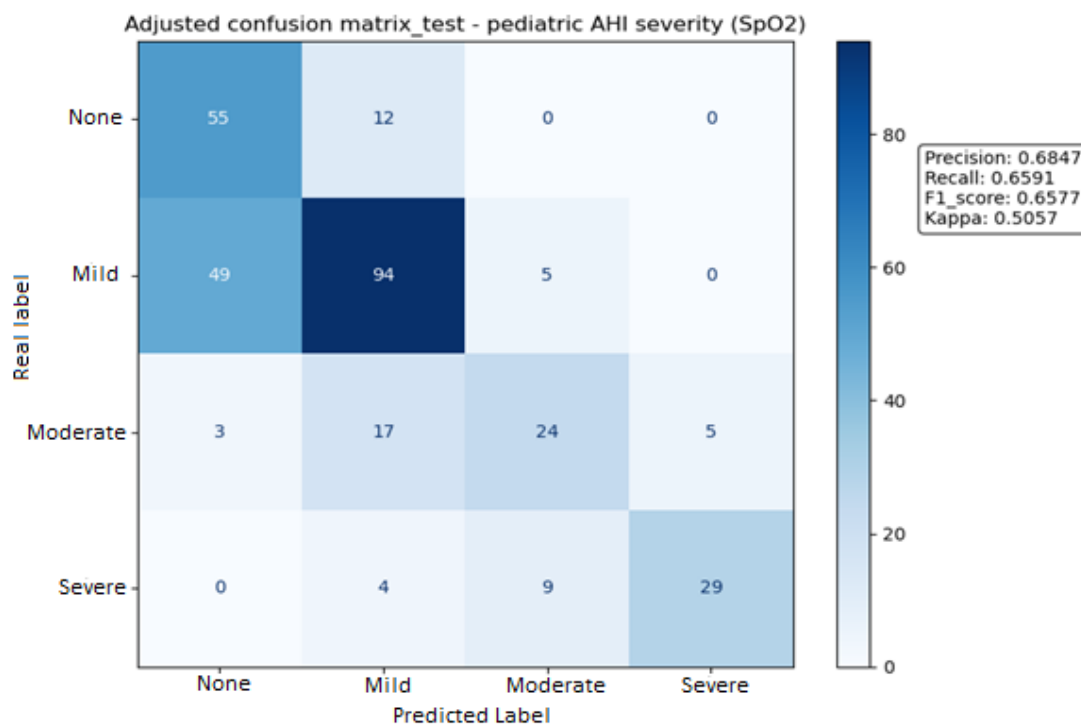


Figura 5-15. Matriz de confusión del valor de IAH posterior a la regresión en el conjunto de test de la señal SpO_2 .

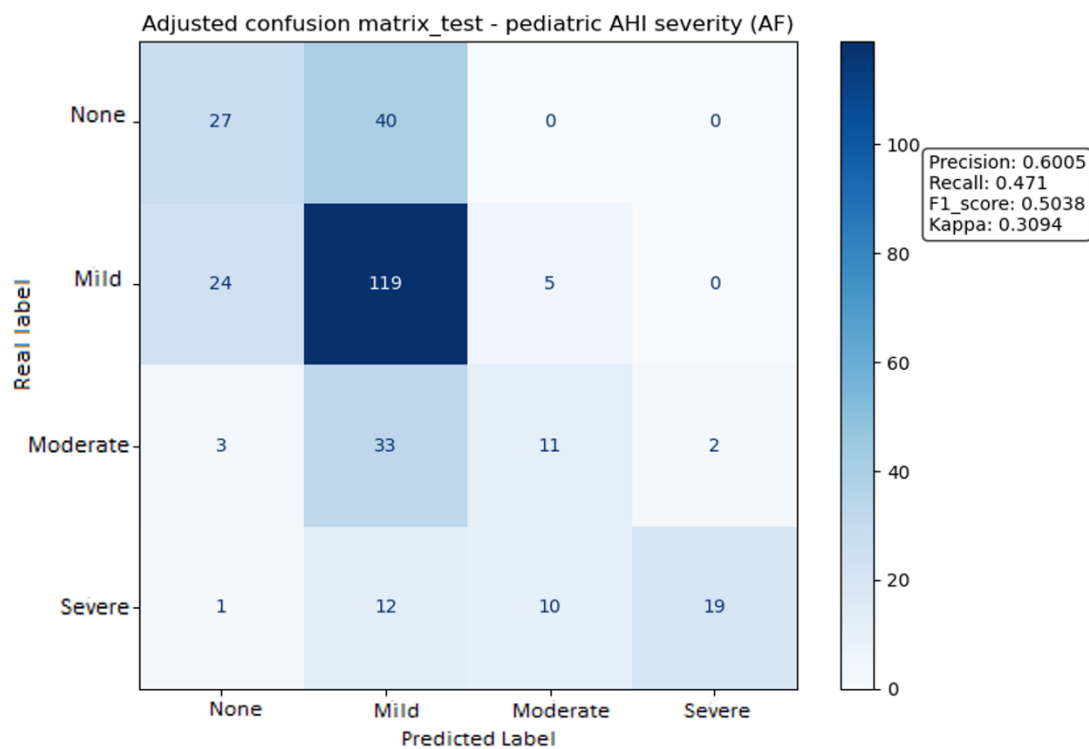


Figura 5-16. Matriz de confusión del valor de IAH posterior a la regresión en el conjunto de test de la señal FA.

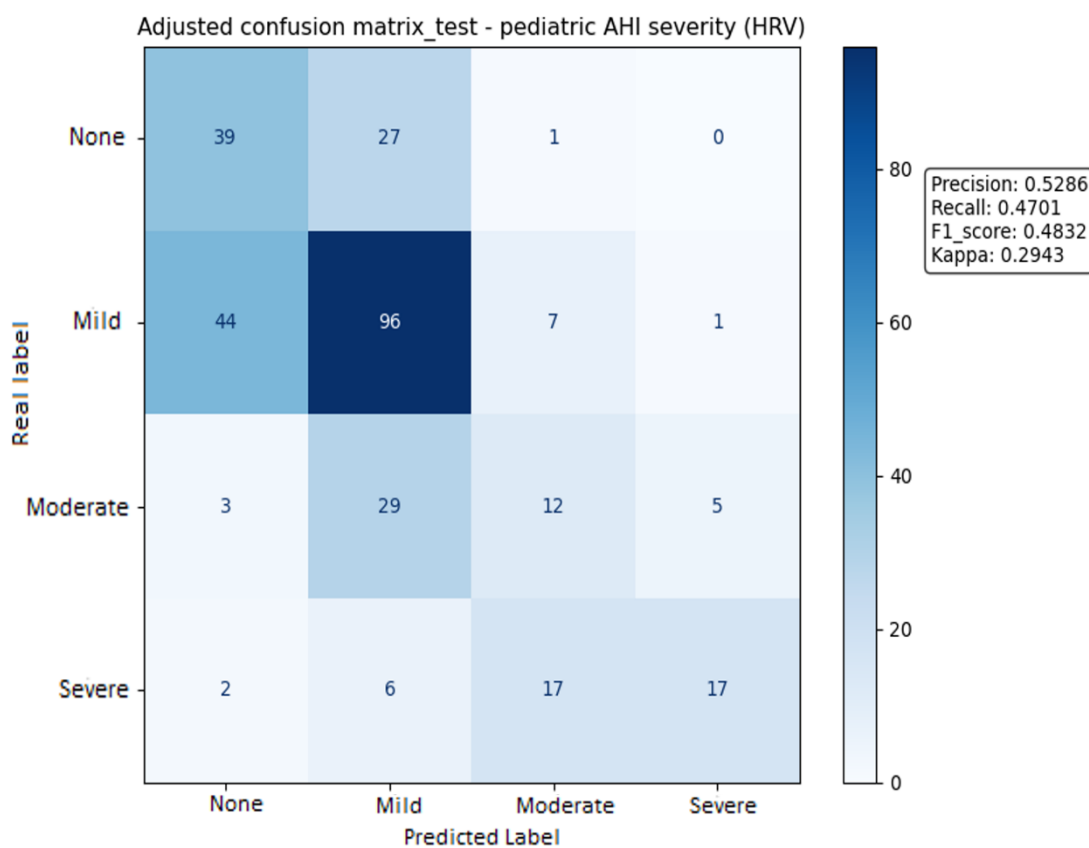


Figura 5-17. Matriz de confusión del valor de IAH posterior a la regresión en el conjunto de test de la señal HRV.

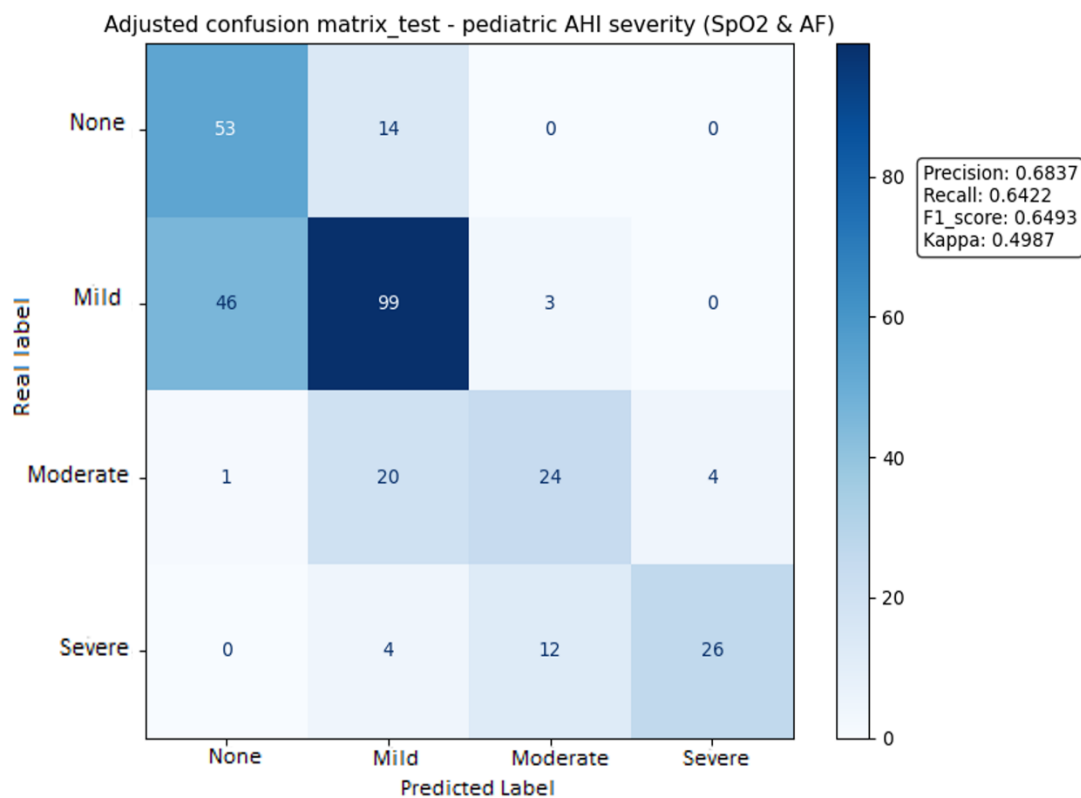


Figura 5-18. Matriz de confusión del valor de IAH posterior a la regresión en el conjunto de test de las señales SpO₂-FA.

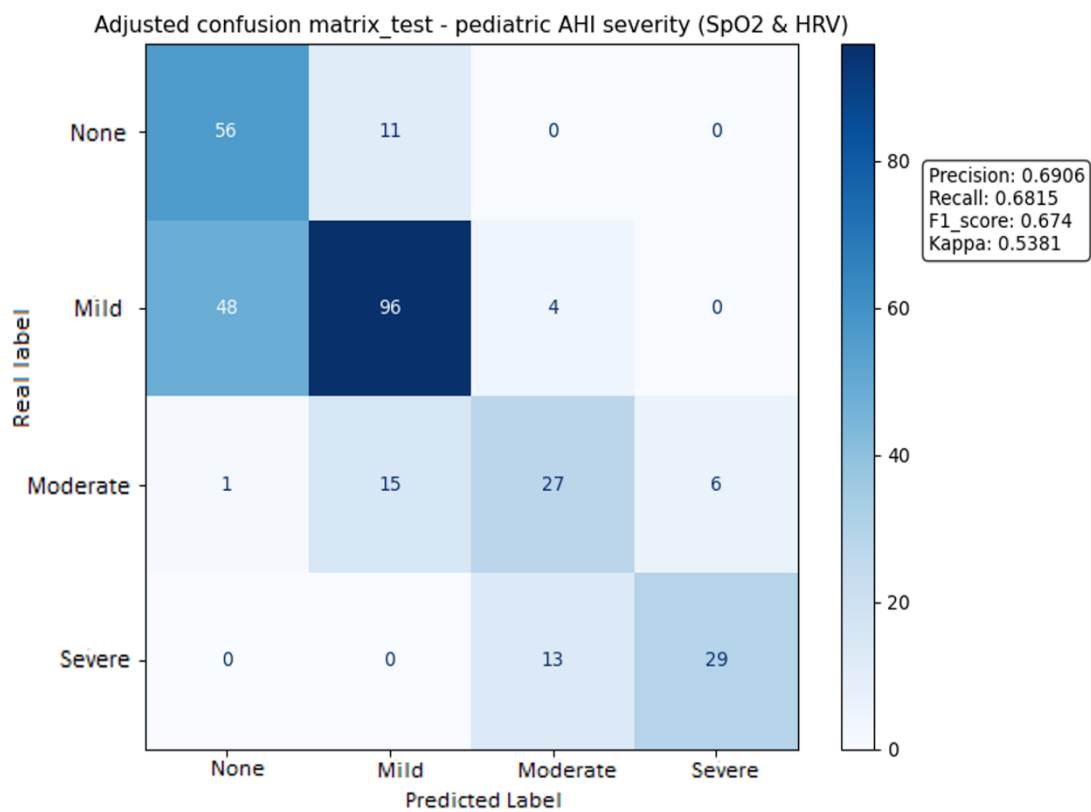


Figura 5-19. Matriz de confusión del valor de IAH posterior a la regresión en el conjunto de test de las señales SpO₂-HRV.

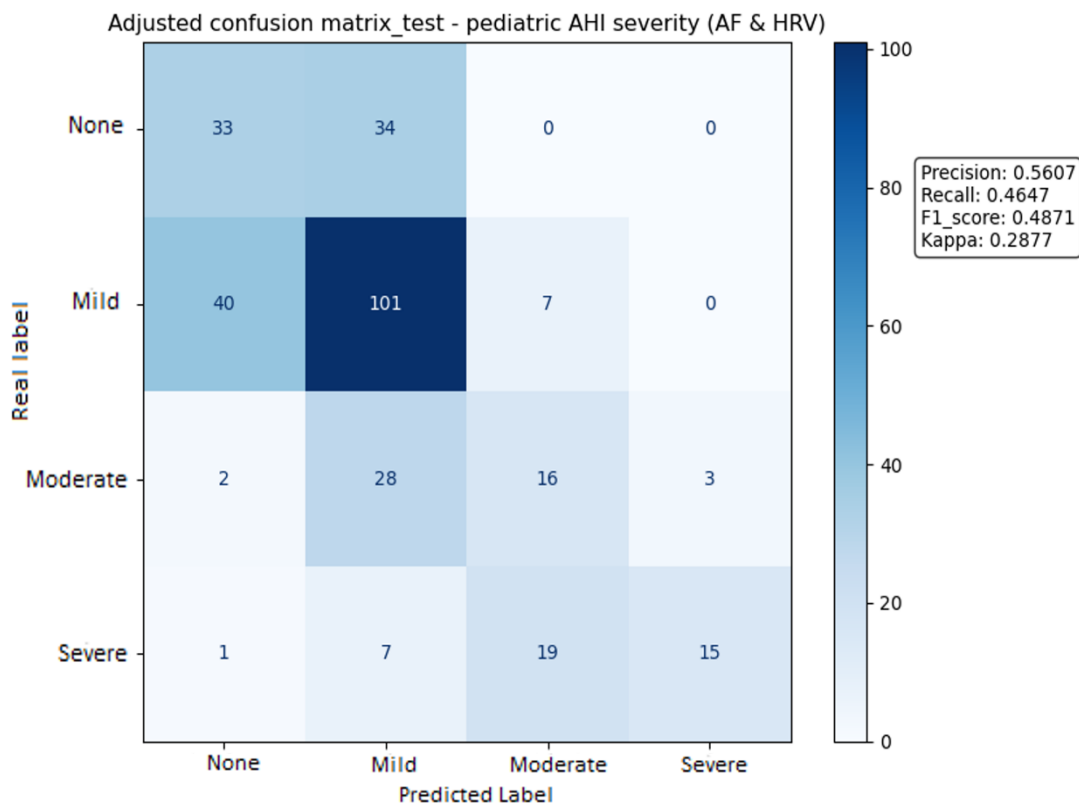


Figura 5-20. Matriz de confusión del valor de IAH posterior a la regresión en el conjunto de test de las señales FA-HRV.

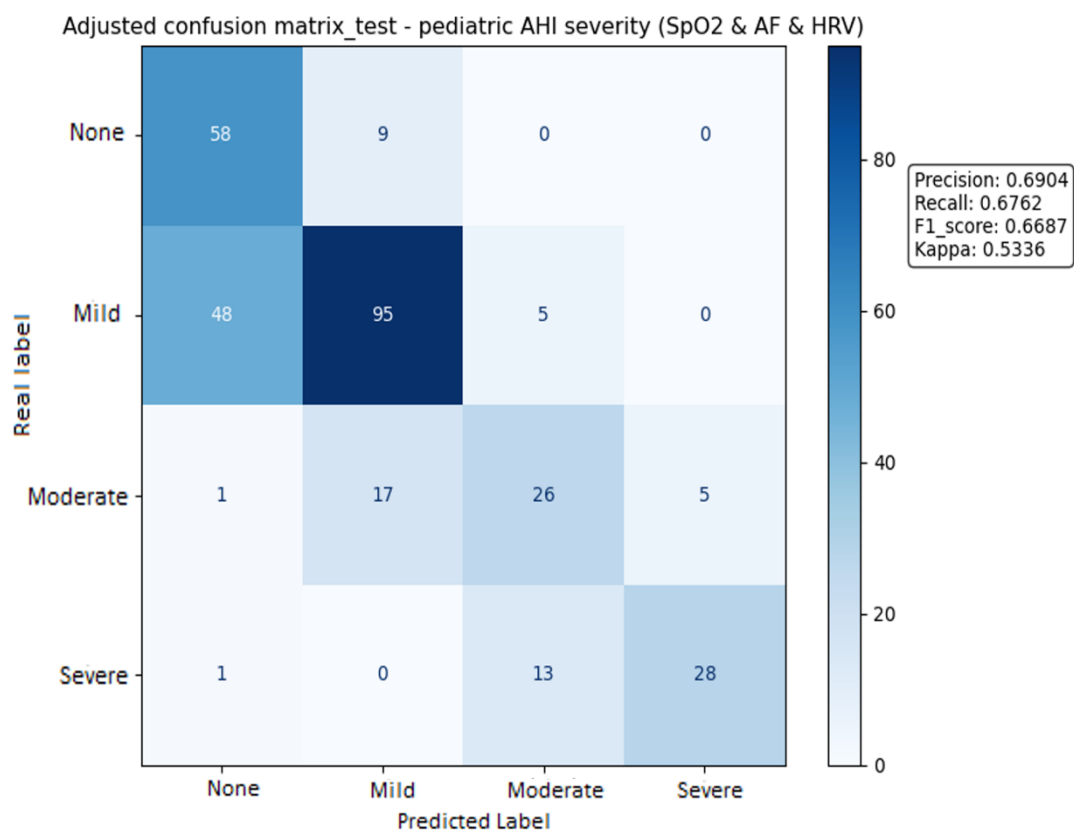


Figura 5-21. Matriz de confusión del valor de IAH posterior a la regresión en el conjunto de test de las señales SpO₂-FA-HRV.

En todas las matrices se han empleado las herramientas proporcionadas por *scikit-learn* para calcular los parámetros: *precision*, *recall*, *F1_score* y *Kappa*. En la siguiente tabla, hemos recopilado todos los valores extraídos de las matrices de confusión tras la regresión (Tabla 5-9):

	<i>Precision</i>	<i>Recall</i>	<i>F1_score</i>	<i>Kappa</i>
SpO ₂	0.6847	0.6591	0.6577	0.5057
FA	0.6005	0.4710	0.5038	0.3094
HRV	0.5286	0.4701	0.4832	0.2943
SpO ₂ -FA	0.6837	0.6422	0.6493	0.4987
SpO ₂ -HRV	0.6906	0.6815	0.6740	0.5381
FA-HRV	0.5607	0.4647	0.4871	0.2877
SpO ₂ -FA-HRV	0.6904	0.6762	0.6687	0.5336

Tabla 5-9. Comparativa de los resultados estadísticos (*precision*, *recall*, *F1-score* y *Kappa*) de las matrices de confusión tras la regresión de todas las combinaciones de señales para el conjunto de test.

En aquellos casos en los que el valor de *Kappa* sea superior al 0.4 implica que el modelo presenta un rendimiento superior al mero azar (Abraira, 2001; McHugh, 2012). En los casos en que *Kappa* sea superior a 0.5, el modelo puede considerarse útil para la asignación de clases. Cuando *Kappa* supera el valor de 0.6, el desempeño del modelo es significativamente mejor que el azar ($\kappa \approx 0$), aunque aún presenta margen de mejora. Como se puede apreciar, en algunos casos el valor de *Kappa* es superior a 0.5 como es el caso de las señales SpO₂, SpO₂-HRV y SpO₂-FA-HRV. Sin embargo, en otros casos como FA, HRV y FA-HRV no alcanza el 0.4, por lo que el desempeño no es todo lo preciso que se desearía. Por otro lado, el valor de *F1-score* nos indica el balance entre *precision* y *recall*, es decir, la capacidad del modelo de detectar correctamente cada clase minimizando los falsos positivos. Para valores elevados como SpO₂-HRV o SpO₂-FA-HRV, el modelo tiene un buen equilibrio entre la detección de casos positivos (*Se*) y la fiabilidad de dichas detecciones (*PPV*).

5.2.3. Resultados de clasificación binaria

Una vez realizado el análisis de la clasificación multiclase en las distintas severidades de la AOS, procederemos con la etapa de umbralización previamente descrita. Este paso es fundamental para poder transformar el problema de clasificación multiclase en un problema de clasificación binaria, permitiendo así un mejor análisis. A partir de cada una de las matrices de confusión obtenidas, extraeremos diversas métricas en distintos formatos. En particular, se calcularán en formato porcentual las métricas: *Se*, *Sp*, *Acc*, *PPV* y *NPV*. Y de forma no porcentual: *LR+* y *LR-*. Todas estas métricas se calcularán para cada uno de los 3 umbrales de la severidad de la AOS para los conjuntos de entrenamiento, validación y test de todas las combinaciones de señales.

A continuación, en las Tablas 5-10, 5-11, 5-12, 5-13, 5-14, 5-15 y 5-16 mostramos agrupados todos los resultados de las distintas combinaciones de señales posibles para todos los conjuntos de datos. En dichas tablas podemos ver como en todos los casos tanto la *Sp* como la *Acc* tienden a aumentar conforme sube el valor del umbral, siendo el valor más alto correspondiente a 10 e/h. Por otro lado, la *Se* decrece al pasar de 1 e/h a 5 e/h pero vuelve a aumentar con 10 e/h. Finalmente, mientras el *NPV* tiende a aumentar conforme incrementa el umbral el *PPV* tiene un comportamiento menos predecible.

SpO ₂		Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	LR+	LR-
1 e/h	Train	82.97	83.11	83.00	94.64	57.59	4.9111	0.2049
	Val	81.32	86.96	82.52	95.87	55.56	6.2348	0.2148
	Test	78.24	82.09	79.08	93.97	51.40	4.3685	0.2650
5 e/h	Train	66.67	98.60	89.36	95.10	87.91	47.6667	0.3381
	Val	71.91	98.73	91.41	95.52	90.35	56.8090	0.2845
	Test	73.63	97.67	90.52	93.06	89.74	31.6593	0.2700
10 e/h	Train	72.52	99.31	95.83	94.06	96.02	105.7570	0.2767
	Val	64.44	98.58	93.87	87.88	94.54	45.2722	0.3607
	Test	69.05	98.11	94.12	85.29	95.22	36.4571	0.3155

Tabla 5-10. Resultados umbralización binaria matriz de confusión de la señal SpO₂ para los conjuntos de entrenamiento, validación y test.

FA		Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	LR+	LR-
1 e/h	Train	85.51	44.75	76.64	84.76	46.23	1.5477	0.3237
	Val	72.76	57.97	69.63	86.57	36.36	1.7312	0.4698
	Test	88.28	40.30	77.78	84.06	49.09	1.4788	0.2907
5 e/h	Train	55.33	98.46	85.98	93.60	84.41	35.9622	0.4537
	Val	64.04	94.94	86.50	82.61	87.55	12.6489	0.3787
	Test	46.15	97.67	82.35	89.36	81.08	19.8462	0.5513
10 e/h	Train	58.02	98.86	93.54	88.37	94.02	50.7634	0.4247
	Val	48.89	98.22	91.41	81.48	92.31	27.4756	0.5204
	Test	45.24	99.24	91.83	90.48	91.93	59.7143	0.5518

Tabla 5-11. Resultados umbralización binaria matriz de confusión de la señal FA para los conjuntos de entrenamiento, validación y test.

HRV		Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	LR+	LR-
1 e/h	Train	81.07	69.41	78.53	90.50	50.50	2.6498	0.2728
	Val	77.04	62.32	73.93	88.39	42.16	2.0446	0.3684
	Test	79.50	58.21	74.84	87.16	44.32	1.9023	0.3522
5 e/h	Train	62.20	99.58	88.77	98.37	86.62	148.2417	0.3796
	Val	51.69	94.94	83.13	79.31	83.96	10.2079	0.5089
	Test	56.04	95.81	83.99	85.00	83.74	13.3883	0.4588
10 e/h	Train	70.23	99.66	95.83	96.84	95.72	204.8346	0.2987
	Val	40.00	99.29	91.10	90.00	91.18	56.2000	0.6043
	Test	40.48	97.73	89.87	73.91	91.17	17.8095	0.6091

Tabla 5-12. Resultados umbralización binaria matriz de confusión de la señal HRV para los conjuntos de entrenamiento, validación y test.

SpO ₂ -FA		Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	LR+	LR-
1 e/h	Train	85.51	79.45	84.19	93.73	60.42	4.1617	0.1823
	Val	82.88	81.16	82.52	94.25	56.00	4.3990	0.2110
	Test	80.33	79.10	80.07	93.20	53.00	3.8446	0.2486
5 e/h	Train	61.86	99.16	88.37	96.77	86.46	73.7113	0.3847
	Val	69.66	98.73	90.80	95.38	89.66	55.0337	0.3073
	Test	72.53	98.60	90.85	95.65	89.45	51.9780	0.2786
10 e/h	Train	71.76	99.43	95.83	94.95	95.92	125.5725	0.2841
	Val	60.00	98.93	93.56	90.00	93.92	56.2000	0.4043
	Test	61.90	98.48	93.46	86.67	94.20	40.8571	0.3868

Tabla 5-13. Resultados umbralización binaria matriz de confusión de la señal SpO₂-FA para los conjuntos de entrenamiento, validación y test.

SpO ₂ -HRV		Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	LR+	LR-
1 e/h	Train	82.85	91.32	84.69	97.17	59.70	9.5491	0.1878
	Val	80.54	92.75	83.13	97.64	56.14	11.1152	0.2098
	Test	79.50	83.58	80.39	94.53	53.33	4.8421	0.2453
5 e/h	Train	74.57	99.44	92.25	98.19	90.57	133.2947	0.2557
	Val	74.16	99.16	92.33	97.06	91.09	87.8764	0.2606
	Test	82.42	98.14	93.46	94.94	92.95	44.2995	0.1792
10 e/h	Train	83.21	99.43	97.32	95.61	97.53	145.6107	0.1689
	Val	71.11	98.22	94.48	86.49	95.50	39.9644	0.2941
	Test	69.05	97.73	93.79	82.86	95.20	30.381	0.3167

Tabla 5-14. Resultados umbralización binaria matriz de confusión de la señal SpO₂-HRV para los conjuntos de entrenamiento, validación y test.

FA-HRV		Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	LR+	LR-
1 e/h	Train	81.07	81.28	81.11	93.96	54.43	4.3302	0.2329
	Val	75.49	73.91	75.15	91.51	44.74	2.8936	0.3317
	Test	82.01	49.25	74.84	85.22	43.42	1.6160	0.3653
5 e/h	Train	66.32	98.88	89.46	96.02	87.83	59.2762	0.3406
	Val	62.92	96.20	87.12	86.15	87.36	16.5693	0.3854
	Test	58.24	96.74	85.29	88.33	84.55	17.8885	0.4316
10 e/h	Train	73.28	99.43	96.02	95.05	96.13	128.2443	0.2687
	Val	44.44	98.93	91.41	86.96	91.75	41.6296	0.5616
	Test	35.71	98.86	90.20	83.33	90.62	31.4286	0.6502

Tabla 5-15. Resultados umbralización binaria matriz de confusión de la señal FA-HRV para los conjuntos de entrenamiento, validación y test.

SpO ₂ -FA-HRV		<i>Se (%)</i>	<i>Sp (%)</i>	<i>Acc (%)</i>	<i>PPV (%)</i>	<i>NPV (%)</i>	<i>LR+</i>	<i>LR-</i>
1 e/h	<i>Train</i>	83.10	88.58	84.29	96.32	59.33	7.2796	0.1908
	<i>Val</i>	81.32	91.30	83.44	97.21	56.76	9.3521	0.2046
	<i>Test</i>	79.08	86.57	80.72	95.45	53.70	5.8870	0.2417
5 e/h	<i>Train</i>	71.82	99.02	91.15	96.76	89.62	73.3603	0.2846
	<i>Val</i>	73.03	98.73	91.72	95.59	90.70	57.6966	0.2731
	<i>Test</i>	79.12	97.67	92.16	93.51	91.70	34.0220	0.2138
10 e/h	<i>Train</i>	79.39	99.43	96.82	95.41	96.99	138.9313	0.2073
	<i>Val</i>	68.89	98.22	94.17	86.11	95.17	38.7156	0.3167
	<i>Test</i>	66.67	98.11	93.79	84.85	94.87	35.2000	0.3398

Tabla 5-16. Resultados umbralización binaria matriz de confusión de la señal SpO₂-FA-HRV para los conjuntos de entrenamiento, validación y test.

Capítulo 6: DISCUSIÓN

En este penúltimo capítulo, realizaremos una discusión final analizando todo el proceso de desarrollo realizado y los resultados obtenidos al trabajar con las distintas señales y el modelo de DL. Además, compararemos los resultados obtenidos en nuestro proyecto con los obtenidos por otros estudios.

6.1. Introducción

En este proyecto se propone una metodología de DL basada en CNN que busca avanzar en el diagnóstico del AOS en población pediátrica utilizando un conjunto reducido de señales. A diferencia de trabajos previos que han abordado esta problemática mediante alternativas a la PSG o técnicas de ML (Alonso Álvarez et al., 2008; L. Chang et al., 2013; Gutiérrez-Tobal et al., 2015; Villa et al., 2015), nuestro enfoque se sitúa dentro de una línea más reciente que explora el uso de DL con señales reducidas. Aunque algunos estudios han considerado emplear tanto individualmente como en ciertas combinaciones las señales SpO₂, FA y HRV en contextos pediátricos, no se ha investigado sistemáticamente el potencial diagnóstico de otras combinaciones específicas entre ellas. En este sentido, la propuesta metodológica de este trabajo ofrece un enfoque novedoso que no solo amplía el uso de DL en este ámbito, sino que también contribuye a identificar qué configuraciones de señales podrían ser más informativas y prácticas en comparación con las estrategias utilizadas anteriormente.

Esta capacidad de comparar de forma controlada distintas combinaciones de señales permite, además, establecer cuál de ellas ofrece el mejor rendimiento diagnóstico dentro del conjunto evaluado. Así, el presente estudio no solo aporta valor en términos metodológicos, sino que también sienta una base sólida para futuras investigaciones, al señalar de forma objetiva la combinación de señales óptima. De este modo, se facilita el diseño de modelos más eficientes y se favorece el desarrollo de herramientas de diagnóstico más accesibles, precisas y adaptadas a las necesidades clínicas reales.

Gracias a los resultados obtenidos en el capítulo anterior, podemos analizar en profundidad el comportamiento del modelo con las distintas señales. Como podemos observar, en la amplia mayoría de los casos los valores de especificidad son muy elevados, llegando a alcanzar valores como 99.66% en algunos casos. Sin embargo, los valores de sensibilidad son bastante más bajos en comparación de los de especificidad, indicando que el modelo tiene más problemas clasificando a los pacientes enfermos que a los sanos. En otras palabras, en términos generales hay una gran cantidad de falsos negativos o sujetos que son clasificados como pacientes sanos cuando en realidad sufren AOS. Finalmente, podemos observar cómo los valores de *accuracy* son elevados, lo que implica que hay una gran cantidad de casos correctamente clasificados como positivos y una gran cantidad de controles clasificados como negativos. Además, podemos notar que esta última métrica tiende a crecer conforme aumenta el umbral. Por lo tanto, la conclusión que podemos extraer es que: el modelo clasifica correctamente la gran mayoría de los sujetos, pero aquellos que no son correctamente clasificados tienden a ser infraestimados, es decir, clasificados como sanos o como un grado de severidad menor al correspondiente.

6.2. Comparativa de señales

Gracias a todas las tablas de resultados extraídas previamente, podemos realizar una comparativa directa entre todas las posibles combinaciones de señales para poder decantarse por aquella o aquellas combinaciones que presenten una estimación más fiable y precisa de la AOS pediátrica. Para poder realizar esta comparativa, extraeremos los valores estadísticos calculados previamente y los compararemos entre sí de la siguiente manera: en primer lugar, compararemos cada una de las señales agrupadas de 2 en 2 con las señales que la componen de manera individual; a continuación, emplearemos las 3 señales individuales frente a la agrupación de las 3 de forma simultánea; y finalmente, evaluaremos las diferencias entre las señales agrupadas de 2 en 2 y las señales agrupadas de 3 en 3. Las Tablas 6-1, 6-2 y 6-3 recogen todos los casos previamente descritos para los conjuntos de test.

	SpO ₂	SpO ₂ -FA	FA
<i>PRECISION</i>	0.6944	0.6837	0.6005
<i>RECALL</i>	0.6722	0.6422	0.4710
<i>F1_SCORE</i>	0.6688	0.6493	0.5038
<i>KAPPA</i>	0.5166	0.4987	0.3094
<i>R2_SCORE</i>	0.8710	0.8566	0.7214

Tabla 6-1. Comparativa entre las señales SpO₂, FA y SpO₂-FA en el conjunto de test.

	SpO ₂	SpO ₂ -HRV	HRV
<i>PRECISION</i>	0.6944	0.6906	0.5286
<i>RECALL</i>	0.6722	0.6815	0.4701
<i>F1_SCORE</i>	0.6688	0.6740	0.4832
<i>KAPPA</i>	0.5166	0.5381	0.2943
<i>R2_SCORE</i>	0.8710	0.8932	0.7326

Tabla 6-2. Comparativa entre las señales SpO₂, HRV y SpO₂-HRV en el conjunto de test.

	FA	FA-HRV	HRV
<i>PRECISION</i>	0.6005	0.5607	0.5286
<i>RECALL</i>	0.4710	0.4647	0.4701
<i>F1_SCORE</i>	0.5038	0.4871	0.4832
<i>KAPPA</i>	0.3094	0.2877	0.2943
<i>R2_SCORE</i>	0.7214	0.7499	0.7326

Tabla 6-3. Comparativa entre las señales FA, HRV y FA-HRV en el conjunto de test.

Hemos resaltado en negrita los mejores resultados de la señal o combinación de señales de la comparativa realizada en cada caso. Ahora, compraremos las 3 señales individuales con la agrupación de las 3 señales simultáneamente. En la siguiente tabla (Tabla 6-4) se muestran los resultados.

	SpO ₂	FA	HRV	SpO ₂ -FA-HRV
PRECISION	0.6944	0.6005	0.5286	0.6904
RECALL	0.6722	0.4710	0.4701	0.6762
F1_SCORE	0.6688	0.5038	0.4832	0.6687
KAPPA	0.5166	0.3094	0.2943	0.5336
R2_SCORE	0.8710	0.7214	0.7326	0.8765

Tabla 6-4. Comparativa entre las señales SpO₂, FA, HRV y SpO₂-FA-HRV en el conjunto de test.

En este caso, como podemos observar la combinación de las 3 señales tiene un *Kappa* superior ligeramente a la señal SpO₂ y muy superior a las señales FA y HRV. Ahora, compararemos las señales 2 a 2 con las 3 señales combinadas y analizaremos los resultados (Tabla 6-5).

	SpO ₂ -FA	SpO ₂ -HRV	FA-HRV	SpO ₂ -FA-HRV
PRECISION	0.6837	0.6906	0.5607	0.6904
RECALL	0.6422	0.6815	0.4647	0.6762
F1_SCORE	0.6493	0.6740	0.4871	0.6687
KAPPA	0.4987	0.5381	0.2877	0.5336
R2_SCORE	0.8566	0.8932	0.7499	0.8765

Tabla 6-5. Comparativa entre las señales SpO₂-FA, SpO₂-HRV, FA-HRV y SpO₂-FA-HRV en el conjunto de test.

A partir de los resultados obtenidos, se observa que, de las 3 señales, SpO₂ es la más eficaz para detectar la AOS de manera individual. Esta señal alcanzó un valor de *Kappa* de 0.4357, superando claramente a los valores obtenidos con FA (*Kappa*=0.2304) y HRV (*Kappa*=0.2943). Asimismo, los valores de *precision* (0.6244) y *recall* (0.6161) de SpO₂ también se sitúan por encima de los alcanzados por FA (*precision*: 0.5617; *recall*: 0.4457) y HRV (*precision*: 0.5286; *recall*: 0.4701). Esta superioridad se refleja igualmente en el *F1_score*, donde SpO₂ obtuvo 0.6196 frente a 0.4624 y 0.4832 para FA y HRV respectivamente. Finalmente, el *R2_score* refuerza esta tendencia, con un valor de 0.8702 para SpO₂, en comparación con 0.7214 y 0.7326 para FA y HRV. Estos resultados confirman que la señal SpO₂ proporciona un rendimiento notablemente superior en la detección del AOS respecto a las otras señales evaluadas.

Por otro lado, los hallazgos realizados demuestran que algunas combinaciones obtienen unos resultados mejores que otras. Concretamente, empleando el mismo modelo en todas ellas las señales FA, HRV y FA-HRV obtienen un valor de *Kappa* de alrededor de 0.3, indicando que el estimador es una clasificación débil pero que apenas supera a la clasificación por azar (porque al azar es $\kappa \approx 0$) (Abraira, 2001; McHugh, 2012). La combinación SpO₂-FA tienen un valor de *Kappa* más elevado, alcanzando casi el valor 0.5, indicando que en este caso, las estimaciones son considerablemente mejores que con las otras combinaciones. Finalmente, los mejores resultados se obtienen de las combinaciones de señales SpO₂-HRV y SpO₂-FA-HRV, que sobrepasan el valor de *Kappa* de 0.5, aunque podrían ser mejores. Estos resultados están en línea con los logrados en estudios similares como el de Jiménez-García et al. (2022, 2024) que obtiene 0.6011 y 0.6231 respectivamente, el de Vaquerizo-Villar et al. (2021) que obtiene 0.515, el de Mortazavi et al. (2024) que obtiene 0.57 para ResNet y 0.63 para CNN-BiGRU o el de Xu Z et al. (2019) que obtiene 0.339 (Jiménez-García et al., 2022, 2024; Mortazavi et al., 2024; Vaquerizo-Villar et al., 2021; Xu et al., 2019).

En el caso del resto de métricas, los valores de *precision* son más elevados en la combinación SpO₂-HRV (0.6906) que en el resto de combinaciones de dos señales como son SpO₂-FA (0.6837) y FA-HRV (0.5607). Los valores de *recall*, *F1_score* y *R2_score* tienen el mismo comportamiento, siendo los valores de SpO₂-HRV (0.6815, 0.6740 y 0.8932 respectivamente) superiores a los valores obtenidos con SpO₂-FA (0.6422, 0.6493 y 0.8566 respectivamente) y FA-HRV (0.6762, 0.6687 y 0.8765 respectivamente).

Finalmente, a la hora de elegir una opción u otra para la versión final del sistema la decisión queda entre las señales SpO₂-HRV y SpO₂-FA-HRV. No obstante, debido a que SpO₂-HRV logró unos valores de *precision* (0.6906), *recall* (0.6815), *F1_score* (0.6740), *R2_score* (0.8932) y *Kappa* (0.5381) superiores a los valores de *precision* (0.6904), *recall* (0.6762), *F1_score* (0.6687), *R2_score* (0.8765) y *Kappa* (0.5336) de SpO₂-FA-HRV; y que además está compuesta de dos señales en lugar de tres, por lo que su extracción es mucho más sencilla tanto en recursos como en tiempo, se ha optado por elegirla como la mejor candidata para el sistema. Además, estas dos señales pueden obtenerse mediante el uso de un pulsioxímetro, por lo que, el proceso de extracción de las señales es mucho más sencillo.

Por lo tanto, eligiendo la señal SpO₂-HRV, tendremos un sistema capaz de obtener valores de *precision* cercanos al 0.7. Concretamente, los resultados obtenidos promediando las 4 clases muestran una *precision* de 0.6906, lo que indica que un 69.06% de las predicciones positivas realizadas por el modelo son correctas. Por otro lado, el valor obtenido para el *recall* de 0.6815, lo cual indica que un 68.15% de los casos reales positivos fueron correctamente identificados. Estos dos valores indican que el modelo acierta en más de la mitad de las predicciones realizadas, pero aún tiene margen de mejora de entorno al 30-35%. Finalmente, el valor de *F1_score* de 0.6740, representa un buen balance del modelo pero indica una mayor predisposición a comprometer la sensibilidad en favor de mayor *precision* (Cifci et al., 2023; Hicks et al., 2022).

La combinación de las señales SpO₂ con HRV es ligeramente superior a la misma combinación añadiendo la señal FA. Sin embargo, esta combinación es considerablemente superior en términos de *Kappa* a las señales SpO₂ con FA y HRV con FA (Abraira, 2001; McHugh, 2012). Esto parece indicar, que la señal de FA no termina de aportar la información necesaria para realizar la predicción de forma correcta. Por el contrario, la señal SpO₂ parece aportar un gran valor al modelo. Como hemos podido observar, las señales SpO₂ y FA son de las 3 señales las que están más estrechamente relacionadas con la AOS, por lo que tiene sentido que sea la que más información aporte a nuestro modelo para realizar la estimación correcta (Jubran, 2015; Mildenhall, 2008). Aunque debemos tener en cuenta que es más complicado extraer y analizar información de la señal FA para detectar la AOS (Jiménez-García et al., 2020, 2022). Finalmente, y observando los resultados obtenidos, podemos concluir que en nuestro modelo y de la forma en la que se han preprocesado los datos, la señal SpO₂ es la señal que mayor información nos proporciona, la señal de HRV aporta más información complementaria al SpO₂ y finalmente el FA no ayudó a mejorar los resultados al combinarla con las otras dos. A pesar de que como ya se ha mencionado la FA es una de las señales más estrechamente ligadas a la AOS, debemos tener en cuenta que factores como el preprocesado o la forma de extraer la información pueden afectar en la calidad de la aportación de la propia señal al modelo.

6.3. Comparativa con estudios previos

Una vez habiendo discutido todos los resultados obtenidos por nuestro modelo, realizaremos una comparativa de dichos resultados frente a aquellos aportados por otros autores. Debemos considerar que solo lo compararemos con los estudios donde se ha estimado la AOS infantil para los umbrales de 1, 5 y 10 e/h. De esta manera podremos realizar una comparación directa, ya que carece de sentido comparar nuestro estudio con aquellos que tratan de estimar la AOS para otros umbrales distintos (p.ej., 3 e/h) (Álvarez et al., 2017); y tiene un valor añadido al emplear 3 umbrales distintos frente a aquellos que solo emplean uno (p.ej., 5 e/h) (Sturludóttir J et al. 2023). Para ello, dividiremos los resultados en distintas tablas agrupadas en función de las señales empleadas (Tablas 6-6, 6-7, 6-8, 6-9, 6-10, 6-11, 6-12 y 6-13).

Estudios (SpO ₂)	Umbral	Base de Datos	Sujetos	Metodología	Resultados		
					Acc.	Se.	Sp.
Vaquerizo Villar et al. (2021)	1 e/h	University of Chicago	981	Segmentation (20min) ↓ CNN ↓ Bayesian Optimization	80.1%	90.8%	36.4%
	5 e/h				83.9%	76.0%	88.6%
	10 e/h				92.3%	79.5%	95.8%
	1 e/h	CHAT			77.6%	71.2%	81.8%
	5 e/h				97.4%	83.7%	100.0%
	10 e/h				97.8%	83.9%	99.3%
Jiménez García et al. (2020)	1 e/h	University of Chicago	974	Time and frec. domain analysis ↓ FCBF ↓ AdaBoost.M2	78.21%	91.43%	22.67%
	5 e/h				77.44%	66.44%	84.02%
	10 e/h				85.90%	40.96%	98.05%
Xu Z et al. (2019)	1 e/h	University of Beijing	432	MLP de 3 capas	79.60%	95.30%	19.10%
	5 e/h				79.40%	77.80%	80.50%
	10 e/h				88.20%	73.50%	92.70%
Hornero R et al. (2017)	1 e/h	Base privada	4191	SPSS Statistics (v20) ↓ MLP	75.20%	84.00%	53.20%
	5 e/h				81.07%	68.20%	87.20%
	10 e/h				90.20%	68.70%	94.10%
Mortazavi et al. (2024)	1 e/h	CHAT	844	Segmentation (20min) ↓ 1D ResNet	72.80%	96.80%	48.75%
	5 e/h				91.67%	77.13%	97.58%
	10 e/h				95.90%	73.25%	98.65%
	1 e/h			Segmentation (20min) ↓ CNN-BiGru	86.53%	96.27%	61.27%
	5 e/h				91.54%	77.79%	97.23%
	10 e/h				96.17%	74.58%	99.81%
Nuestro estudio	1 e/h	CHAT	1639	Segmentation (20min) ↓ CNN	79.08%	78.24%	82.09%
	5 e/h				90.52%	73.63%	97.67%
	10 e/h				95.83%	69.05%	98.11%

Tabla 6-6. Tabla comparativa de estudios previos con nuestro estudio para la señal SpO₂ y umbrales de 1, 5 y 10e/h.

Estudios (FA)	Umbral	Base de Datos	Sujetos	Metodología	Resultados		
					Acc.	Se.	Sp.
Barroso García et al. (2021a)	1 e/h	University of Chicago	946	Wavelet decomposition	73.61%	79.89%	47.24%
	5 e/h			↓ FCBF	57.46%	74.43%	47.18%
	10 e/h			↓ AdaBoost.M2	76.07%	41.06%	89.58%
	1 e/h			Wavelet decomposition	80.85%	100.0%	0.00%
	5 e/h			↓ FCBF	57.14%	77.25%	45.05%
	10 e/h			↓ Bayesian MLP	70.47%	50.00%	75.96%
Jiménez García et al. (2020)	1 e/h	University of Chicago	974	Time and freq. domain analysis	80.51%	99.37%	1.33%
	5 e/h			↓ FCBF	62.82%	62.33%	63.11%
	10 e/h			↓ AdaBoost.M2	78.97%	39.77%	89.58%
Barroso García et al. (2021b)	1 e/h	University of Chicago	946	Bispectrum estimation	78.14%	94.07%	11.16%
	5 e/h			↓ Bispectral band	61.20%	78.66%	50.61%
	10 e/h			↓ FCBF ↓ MLP	77.35%	55.85%	83.16%
Barroso García et al. (2017)	1 e/h	University of Chicago	501	Welch method (FFT)	60.00%	60.50%	58.60%
	5 e/h			↓ FSLR	76.00%	65.00%	80.60%
	10 e/h			↓ LR	80.00%	83.80%	79.00%
Nuestro estudio	1 e/h	CHAT	1639	Segmentation (20min)	77.78%	88.28%	40.30%
	5 e/h			↓ CNN	82.35%	46.15%	97.67%
	10 e/h				91.83%	45.24%	99.24%

Tabla 6-7. Tabla comparativa de estudios previos con nuestro estudio para la señal FA y umbrales de 1, 5 y 10e/h.

Estudios (HRV)	Umbral	Base de Datos	Sujetos	Metodología	Resultados		
					Acc.	Se.	Sp.
Martín Montero et al. (2021)	1 e/h	CHAT y University of Chicago	1738	Bispectrum Estimation	63.40%	76.30%	38.30%
	5 e/h			↓ Bispectral features ⁴	81.00%	62.50%	84.20%
	10 e/h			↓ FCBF ↓ MLP	89.30%	66.70%	91.60%
Martín Montero et al. (2022)	1 e/h	CHAT y University of Chicago	1738	MLP	74.58%	85.47%	35.38%
	5 e/h				84.95%	64.44%	93.78%
	10 e/h				91.64%	53.66%	97.67%
Martín Montero et al. (2023)	1 e/h	CHAT	1610	Segmentation (10min)	80.07%	90.76%	23.40%
	5 e/h			↓ Time and freq. domain analysis	63.18%	66.67%	61.17%
	10 e/h			↓ LSBoost	84.12%	40.00%	92.03%
	1 e/h			Segmentation (10min)	-	66.94%	-
	5 e/h			↓ Time and freq. domain analysis	-	72.08%	-
	10 e/h			↓ AdaBoost	-	77.13%	-
García Vicente et al. (2023) ⁵	1 e/h	CHAT	1610	Segmentation (10min)	75.92%	84.19%	46.15%
	5 e/h			↓ CNN	86.96%	76.67%	91.39%
	10 e/h				91.97%	53.66%	98.06%
Nuestro estudio	1 e/h	CHAT	1639	Segmentation (20min)	74.84%	79.50%	58.21%
	5 e/h			↓ CNN	83.99%	56.04%	95.81%
	10 e/h				89.87%	40.48%	97.73%

Tabla 6-8. Tabla comparativa de estudios previos con nuestro estudio para la señal HRV y umbrales de 1, 5 y 10e/h.

⁴ Bispectral region amplitude, bispectral region entropy, bispectral region momento and weighted center of bispectrum.

⁵ El trabajo de García Vicente et al. (2023), no emplea HRV sino ECG.

Estudios (SpO ₂ -FA)	Umbral	Base de Datos	Sujetos	Metodología	Resultados		
					Acc.	Se.	Sp.
Jiménez García et al. (2024)	1 e/h	CHAT	2612	Segmentation (30min) ↓ CNN + RNN ↓ GradCAM	87.30%	87.00%	88.10%
	5 e/h				93.50%	80.20%	99.10%
	10 e/h				93.50%	71.40%	97.00%
	1 e/h	University of Chicago			84.10%	96.80%	30.70%
	5 e/h				84.60%	82.90%	85.70%
	10 e/h				90.50%	78.30%	93.80%
Jiménez García et al. (2020)	1 e/h	University of Chicago	974	Time and frec. domain analysis ↓ FCBF ↓ AdaBoost.M2	78.21%	90.79%	25.33%
	5 e/h				77.95%	72.60%	81.15%
	10 e/h				85.90%	44.58%	97.07%
Barroso García et al. (2020)	1 e/h	University of Chicago	946	Recurrence plot analysis ↓ FCBF ↓ Bayesian MLP	81.10%	99.30%	4.20%
	5 e/h				60.90%	80.90%	48.90%
	10 e/h				80.60%	63.80%	85.10%
Jiménez García et al. (2022)	1 e/h	CHAT	2612	Segmentation (5min) ↓ 2D CNN	84.60%	82.40%	92.50%
	5 e/h				93.50%	80.20%	99.10%
	10 e/h				94.40%	71.40%	98.10%
	1 e/h	University of Chicago			84.10%	95.20%	37.30%
	5 e/h				84.10%	82.20%	85.30%
	10 e/h				90.30%	78.30%	93.50%
Nuestro estudio	1 e/h	CHAT	1639	Segmentation (20min) ↓ CNN	80.07%	80.33%	79.10%
	5 e/h				90.85%	72.53%	98.60%
	10 e/h				93.46%	61.90%	98.48%

Tabla 6-9. Tabla comparativa de estudios previos con nuestro estudio para la señal SpO₂-FA y umbrales de 1, 5 y 10e/h.

Estudios (SpO ₂ -HRV)	Umbral	Base de Datos	Sujetos	Metodología	Resultados		
					Acc.	Se.	Sp.
Ye P et al. (2023)	1 e/h	University of Beijing	3139	XGBoost	90.45%	90.26%	100.0%
	5 e/h				85.67%	82.07%	93.78%
	10 e/h				89.81%	84.77%	92.11%
	1 e/h			LR	79.30%	78.90%	100.0%
	5 e/h				83.92%	80.00%	92.75%
	10 e/h				87.44%	85.79%	89.10%
Garde et al. (2014)	5 e/h	British Columbia Children’s Hospital	146	AROC ↓ LDA	84.90%	88.40%	83.60%
Garde et al. (2019)	1 e/h	British Columbia Children’s Hospital	235	Segmentation (2min) ↓ Time and freq. domain analysis ↓ LR	71.00%	68.00%	86.00%
	5 e/h			78.00%	58.00%	89.00%	
	10 e/h			88.00%	90.00%	87.00%	
Nuestro estudio	1 e/h	CHAT	1639	Segmentation (20min) ↓ CNN	80.39%	79.50%	83.58%
	5 e/h				93.46%	82.42%	98.14%
	10 e/h				93.79%	69.05%	97.73%

Tabla 6-10. Tabla comparativa de estudios previos con nuestro estudio para la señal SpO₂-HRV y umbrales de 1, 5 y 10e/h.

Debido a que no existen por el momento estudios que traten de estimar el grado de AOS con los umbrales de 1, 5 y 10 e/h para la combinación de señales FA-HRV; compararemos nuestro estudio con los trabajos existentes aunque no se hayan basado en la misma metodología (Tabla 6-11).

Estudios (FA-HRV)	Umbral	Base de Datos	Sujetos	Metodología	Resultados		
					Acc.	Se.	Sp.
Sturludóttir J et al. (2023)	5 e/h	Hospital in Reykjavik	20	Segmentation (10seg) ↓ CNN	95.85%	75.15%	96.20%
Nuestro estudio	1 e/h	CHAT	1639	Segmentation (20min) ↓ CNN	74.84%	82.01%	49.25%
	5 e/h				85.29%	58.24%	96.74%
	10 e/h				90.20%	35.71%	98.86%

Tabla 6-11. Tabla comparativa de estudios previos con nuestro estudio para la señal FA-HRV y umbrales de 1, 5 y 10e/h.

Estudios (SpO ₂ -FA- HRV)	Umbral	Base de Datos	Sujetos	Metodología	Resultados		
					Acc.	Se.	Sp.
Tan H et al. (2014)	1 e/h	University of Chicago	100	Direct comparison	86.0%	82.5%	90.0%
	5 e/h				85.0%	62.5%	100.0%
	10 e/h				93.0%	65.0%	100.0%
Nuestro estudio	1 e/h	CHAT	1639	Segmentation (20min) ↓ CNN	80.72%	79.08%	86.57%
	5 e/h				92.16%	79.12%	97.67%
	10 e/h				93.79%	66.67%	98.11%

Tabla 6-12. Tabla comparativa de estudios previos con nuestro estudio para la señal SpO₂-FA-HRV y umbrales de 1, 5 y 10e/h.

Si analizamos la comparativa de las señales individuales, en el caso de la señal SpO₂ nuestro proyecto contrastado con los autores que emplean CNN como Vaquerizo et al. (2021) o variantes de CNN como Mortazavi et al. (2024) presenta unos resultados similares o ligeramente inferiores. Sin embargo, los resultados de *accuracy* y especificidad obtenidos son superiores a todos los otros estudios realizados previamente que no emplean CNN o variantes; y los valores de sensibilidad, son superiores a una gran mayoría de estos. Por otro lado, para la señal FA no existe ningún trabajo que emplee la metodología CNN, por lo que nuestro enfoque ya presenta una proposición novedosa a este campo de estudio. Además, los valores de *accuracy* y especificidad obtenidos son superior a todas las investigaciones anteriores en un promedio de un 20%, mientras que para el caso de la sensibilidad es completamente opuesto. Finalmente, en el caso de la señal HRV, podemos observar que nuestro modelo logra unos valores superiores o cercanos a la mayoría de los estudios previos realizados. No obstante, si comparamos nuestros resultados con los de García Vicente et al. (2023) que emplea la misma metodología y base de datos aunque no exactamente la misma señal (dicho estudio emplea la señal ECG), nuestro resultados son inferiores, lo cual puede deberse a las técnicas de preprocesado realizadas (a pesar de que se ha empleado la misma base de datos, el número de sujetos descartados es mayor), una estructura de CNN mucho más compleja o al uso de una señal más compleja como es el ECG en lugar del HRV. De nuevo, conseguimos demostrar la utilidad de las redes CNN para el análisis de esta señal en comparación con el resto de trabajos que emplean otro tipo de metodologías. Sin embargo, debemos destacar que el valor de 56.0% obtenido en la sensibilidad es el valor más bajo de todos los estudios. Y sin embargo, el valor de especificidad de 95.8% es el más alto de todos ellos, por lo que significa que el modelo infraestima la gravedad del trastorno y tiene a clasificar los casos de AOS como un grado inferior al real.

Por otra parte, en el caso de las señales SpO₂-FA, aquellas publicaciones que emplean una metodología de 2D CNN o CNN + RNN, han demostrado un rendimiento superior en todos los aspectos que nuestro estudio empleando 1D CNN. Comparando ambas metodologías para la base de datos CHAT para hacerlo más equitativo, podemos observar como en los estudios previamente mencionados el valor de *accuracy* es superior en un 2.6%, el valor de sensibilidad es superior en un 7.7% y el valor de especificidad es superior en un 0.5% respecto a nuestro estudio. Por otro lado, el resto de autores que emplean metodologías distintas a los modelos basados en CNN, obtienen resultados inferiores en comparación con nuestro estudio para las métricas de *accuracy* y especificidad, llegando a ser casi el doble que otros estudios (Barroso García, 2020). Sin embargo, de nuevo la sensibilidad obtenida es menor en nuestro caso que en los valores obtenidos por el resto. Por lo tanto, nuestro proyecto demuestra que emplear CNN o sus variantes logra una mejoría en los resultados en comparación con otras metodologías. Con la señal FA-HRV, se ha empleado la misma metodología en nuestro caso que en el resto, un modelo CNN. Sin embargo, los resultados son inferiores al estudio de (Sturludóttir, 2023), esto se puede deber al uso de otras bases de datos, un preprocesado de los datos distinto o incluso una estructura diferente para el modelo.

Las combinaciones de señales como SpO₂-HRV y SpO₂-FA-HRV representan una contribución completamente novedosa dentro del campo de estudio, ya que, hasta la fecha, no se han encontrado investigaciones previas que utilicen estas configuraciones aplicando metodologías basadas en DL. En ambos casos, los resultados obtenidos en cuanto a *accuracy* han sido elevados en comparación con los reportados por estudios anteriores. En el caso específico de la combinación SpO₂-FA-HRV, nuestros resultados muestran una mejora considerable en la sensibilidad respecto a los valores reportados por Tan H et al. (2014) para los umbrales de 5 y 10 e/h (79.12% y 66.67% frente a 62.5% y 65.0%, respectivamente). No obstante, la especificidad en nuestro proyecto resulta ligeramente inferior, ya que Tan H et al. reporta valores entre 90% y 100%, mientras que los nuestros oscilan entre 86% y 98%. A pesar de ello, la *accuracy* obtenida supera también la de Tan H et al. para los umbrales de 5 y 10 e/h, al igual que ocurre con la sensibilidad. Por otro lado, en la combinación SpO₂-HRV, los resultados de *accuracy* de nuestro modelo son superiores a los reportados por Ye P et al. (2023) y Garde et al. (2019) en todos los umbrales, salvo en el caso de 1 e/h, donde el modelo XGBoost de Ye P et al. supera al nuestro por un 10%. En cuanto a la sensibilidad, nuestros valores para 1 y 5 e/h son comparables a los obtenidos por Ye P et al. utilizando LR, y muestran una clara mejoría frente a los de Garde et al. (2019). Por el contrario, la especificidad presenta una tendencia inversa: nuestros resultados son superiores para 5 y 10 e/h, con mejoras de hasta un 10% sobre Garde et al. y Ye P et al. (XGBoost), y un 5% sobre Ye P et al. (LR) en 10 e/h. El valor más alto de especificidad, sin embargo, lo obtiene Ye P et al. (2023) con XGBoost en el umbral de 1 e/h, alcanzando un 100%, aunque el número de sujetos de control era muy reducido en la base de datos empleada en dicho estudio.

Por lo tanto, debemos destacar que en el caso de las otras señales, el uso de metodologías basadas en CNN o variantes obtienen mejores resultados en todas las métricas que aquellos estudios que emplean otro tipo de sistemas o modelos. Destacando así la superioridad de metodologías basadas en DL frente a aquellas basadas en ML (Gutiérrez-Tobal et al., 2022). A mayores, nuestro estudio permite realizar una comparativa entre todas las combinaciones de señales previamente mencionadas, por lo que permite añadir un enfoque complementario a los ya aportados por las tablas previas al no solamente evaluar las señales y sus combinaciones de manera aislada si no permitir una comparativa directa entre ellas empleando una misma metodología común para todas.

Para el caso del FA y tal y como se ha mencionado previamente, obtenemos unos valores ligeramente menores que en el resto de casos empleando la misma base de datos. Por lo tanto, si los datos empleados son iguales pero los resultados son más bajos puede deberse principalmente a dos motivos: el modelo empleado o el preprocesado de las señales de FA. Si los resultados no son tan buenos como en el resto de los proyectos, puede implicar que la metodología empleada no está tan optimizada como aquellas metodologías y modelos empleados por otros autores. En este caso, deberíamos valorar el uso de este modelo u otros para trabajar con la señal FA. Por otro lado, también debemos tener en cuenta el factor del preprocesado, ya que factores como el filtrado realizado o la frecuencia de muestreo elegida podrían estar afectando negativamente a la calidad de los datos. En cualquiera de los casos, los resultados obtenidos indicarían que es necesario mejorar el procesamiento de la señal de FA para incrementar su utilidad diagnóstica junto con otras señales.

Capítulo 7: CONCLUSIONES Y LÍNEAS FUTURAS

En este último capítulo, se resumirán todos los conceptos y puntos desarrollados y analizados a lo largo de todo el documento. Para ello, se presentarán las conclusiones derivadas de este estudio, recapitularemos los mejores hallazgos y sus implicaciones a la hora de realizar el diagnóstico de la AOS infantil. A mayores, comentaremos las limitaciones de este estudio y las posibles líneas futuras que podrán ser exploradas para mejorar y aumentar las posibilidades de nuestro proyecto.

7.1. Contribuciones realizadas

A lo largo de todo el documento se han ido presentando los resultados del proceso de optimización de hiperparámetros, los resultados del modelo encargado de realizar las predicciones del número de eventos y los resultados del sistema final encargado de predecir el valor final del IAH y obtener el grado de severidad de la AOS pediátrica. Por lo tanto, el análisis de estos resultados permiten sintetizar una serie de contribuciones aportadas por este estudio, dichos puntos se describen a continuación.

- En primer lugar, este trabajo ha permitido realizar una comparativa entre todas las combinaciones posibles de las señales SpO₂, FA y HRV. Gracias a nuestro estudio, hemos podido extraer una serie de métricas estadísticas con las que comparar de manera directa cada una de las combinaciones. Gracias a esta comparativa, hemos podido elegir la combinación SpO₂-HRV como la agrupación de señales con mejores resultados a la hora de realizar la predicción.
- En segundo lugar, hemos podido desarrollar un sistema basado en una CNN capaz de realizar estimaciones de la severidad de la AOS infantil en base a una, dos o tres señales y sus combinaciones. Este estudio permite disponer de una herramienta más con la que los equipos médicos especializados puedan complementar sus decisiones a la hora de realizar el diagnóstico.
- Hemos podido realizar una aportación a la literatura completamente novedosa para combinaciones de señales SpO₂-FA-HRV y SpO₂-HRV añadiendo un nuevo enfoque y nuevas posibilidades. Para las señales SpO₂ y HRV y la combinación SpO₂-FA, hemos aportado otro estudio más empleando distintas técnicas de preprocesado y distintas bases de datos para comparar sus resultados.
- Finalmente, para la señal FA, hemos aportado una nueva metodología basada en CNN con la que complementar los estudios ya existentes basados en otro tipo de técnicas.

7.2. Conclusiones

Las conclusiones extraídas a partir del estudio y sistema desarrollados se presentan enumeradas a continuación.

- Los sistemas basados en CNN, concretamente DL, son realmente útiles a la hora de realizar estimaciones a partir de datos sin etiquetar como las señales SpO₂, FA y HRV. Este tipo de técnicas permiten poder emplear datos con un mínimo de preprocesado sin etapas específicas de extracción.
- De todas las combinaciones de señales empleadas, la combinación de SpO₂-HRV presenta los mejores resultados como son: 69.06% de *precision*, 68.15% de *recall*, 67.40% de *F1_score*, 0.5381 de *Kappa* y 0.8932 de *R2_Score* para el conjunto de test para fragmentos de 20 min.
- Los modelos CNN óptimos difieren para cada una de las señales y sus combinaciones debido al ajuste de sus hiperparámetros y las particularidades de cada una de estas señales.

- Podemos observar que las señales SpO₂ y HRV mejoran sus resultados cuando se combinan juntas, mientras que las señales SpO₂ y FA empeoran en comparación con SpO₂ y mejoran en comparación con FA. Finalmente, las señales HRV y FA empeoran al combinarse en comparación a FA y mejoran en comparación a HRV.
- Observando los valores de sensibilidad, exactitud y especificidad obtenidos podemos asegurar que nuestro modelo infraestima ligeramente el IAH en cualquiera de las señales y por tanto puede clasificar incorrectamente su nivel de severidad en algunos casos.
- Debido a que los valores de sensibilidad, exactitud y especificidad obtenidos para FA son ligeramente menores que otros casos empleando la misma base de datos, podemos observar que existen factores como el preprocesado realizado en la señal que puede ser un claro condicionante a la hora de obtener mejores resultados.

7.3. Limitaciones y líneas futuras

Finalmente, se exponen a continuación las limitaciones observadas al finalizar el proyecto y las posibles líneas futuras que podrían implementarse a partir de nuestro modelo. A continuación, comentaremos las limitaciones experimentadas y las posibles líneas futuras de nuestro proyecto:

- Hemos desarrollado un sistema solo aplicable a los pacientes pediátricos, de manera que no podrá emplearse con pacientes adultos.
- Nos hemos centrado exclusivamente en las 3 señales SpO₂, FA y HRV y sus combinaciones sin explorar otras posibles señales o combinaciones.
- Nos hemos centrado exclusivamente en redes CNN como metodología, sin explorar otros modelos o sus combinaciones.
- Al no haber realizado ningún tipo de preprocesado las señales han sido empleadas en crudo, lo que implica presencia de artefactos no deseados que pueden afectar negativamente a los resultados.
- El uso de la base de datos CHAT nos ha permitido disponer de una de las mayores bases de datos de AOS pediátrica disponibles. No obstante, no se ha explorado la posibilidad de combinarlas con otras bases de datos para aumentar así la cantidad de datos disponibles.
- Se ha empleado solamente el algoritmo de optimización “Adam” sin evaluar el desempeño y rendimiento de otras alternativas.
- Se la utilizado solamente la regresión Huber como algoritmo de regresión sin evaluar la posibilidad de usar otras alternativas.
- No se han aplicado algoritmos de interpretabilidad por parte de los profesionales de la salud, al no haber implementado ninguna técnica de XAI.

Una vez comentadas las limitaciones, ahora mostraremos las posibles líneas futuras que pueden ser abordadas:

- Extrapolar nuestro sistema a los adultos, bien de forma equivalente empleando solo sujetos adultos o bien de manera conjunta empleando ambos grupos.
- Probar el modelo con otro tipo de señales como puede ser un ECG, un EEG o señales acústicas referentes a la respiración o ronquidos del paciente. Además de combinar estas nuevas señales con las señales ya analizadas.
- Explorar la posibilidad de utilizar otro tipo de modelos de DL, como pueden ser: una RNN, una red ResNet o una combinación de nuestra CNN con otro tipo de red o algoritmo.

- Realizar un preprocesado y limpieza de los datos de los sujetos, eliminando artefactos no deseados.
- Aumentar la cantidad de datos disponible, ya sea combinándola con otras bases de datos u obteniendo otras nuevas a partir de estudios del sueño en sujetos pediátricos.
- Buscar nuevos algoritmos de optimización alternativos al algoritmo “*Adam*”, nuevas variables de monitoreo de pérdidas y otros métodos de regresión para lograr una estimación del IAH más precisa, tratando de estimar con mayor exactitud el tiempo de sueño.
- Añadir alguna herramienta de XAI con la que poder permitir a los especialistas en la salud que empleen nuestro sistema comprender como ha decidido nuestro sistema.

Con estas líneas futuras se coloca el broche final a este TFM elaborado de forma conjunta con el Grupo de Ingeniería Biomédica (GIB) durante el transcurso del Máster de Ingeniero de Telecomunicaciones en la Universidad de Valladolid.

BIBLIOGRAFÍA

- Abraira, V. (2001). El índice kappa. *SEMERGEN - Medicina de Familia*, 27(5), 247–249. [https://doi.org/10.1016/S1138-3593\(01\)73955-X](https://doi.org/10.1016/S1138-3593(01)73955-X)
- Acevedo, E., Poladian, A., De Tomaszewski, S. M. M., Peña, C., & Pardo, L. Di. (2018). *Círculo Argentino de Odontología Staff Consejo asesor Diseño e impresión*.
- Aggarwal, C. C. (2018a). *Neural Networks and Deep Learning*.
- Aggarwal, C. C. (2018b). *Neural Networks and Deep Learning* (1st ed.). Springer International Publishing AG.
- Akdemir Akar, S., Kara, S., Latifoğlu, F., & Bilgiç, V. (2013). Spectral analysis of photoplethysmographic signals: The importance of preprocessing. *Biomedical Signal Processing and Control*, 8(1), 16–22. <https://doi.org/10.1016/j.bspc.2012.04.002>
- Al-Faiz, M. Z., Ibrahim, A. A., & Hadi, S. M. (2019). The effect of Z-Score standardization (normalization) on binary input due the speed of learning in back-propagation neural network. *Iraqi Journal of Information & Communications Technology*, 1(3), 42–48. <https://doi.org/10.31987/ijict.1.3.41>
- Alonso Álvarez, D. M., Luz Alonso Álvarez, M., Terán Santos, J., Aurelio Cordero Guevara, J., Isabel Navazo Egüia, A., Ordax Carbajo, E., & Fernando Masa Jiménez Rafael Pelayo, J. (2008). *Reliability of Respiratory Polygraphy for the Diagnosis of Sleep Apnea-Hypopnea Syndrome in Children* *Fiabilidad de la poligrafía respiratoria para el diagnóstico del síndrome de apneas-hipopneas durante el sueño en niños*.
- Altman, D. G., & Bland, J. M. (1983). Measurement in Medicine: the Analysis of Method Comparison Studies †. In *The Statistician* (Vol. 32).
- Álvarez, D., Alonso-Álvarez, M. L., Gutiérrez-Tobal, G. C., Crespo, A., Kheirandish-Gozal, L., Hornero, R., Gozal, D., Terán-Santos, J., & Del Campo, F. (2017). Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal Oximetry: An Alternative to Respiratory Polygraphy in Unattended Settings. *Journal of Clinical Sleep Medicine*, 13(05), 693–702. <https://doi.org/10.5664/jcsm.6586>
- Álvarez, D., Hornero, R., Marcos, J. V., & del Campo, F. (2010). Multivariate Analysis of Blood Oxygen Saturation Recordings in Obstructive Sleep Apnea Diagnosis. *IEEE Transactions on Biomedical Engineering*, 57(12), 2816–2824. <https://doi.org/10.1109/TBME.2010.2056924>
- Álvarez-Sala, J. L., García Mochón, L., González Núñez, J., Martínez Hernández, D., Orero González, A., Rodríguez de la Pinta, M. L., & Sordo del Castillo, L. (2015). *Evaluación sanitaria y socioeconómica del síndrome de apneas e hipopneas del sueño (SAHS) en España*. www.imc-sa.es
- Ana, D., Sánchez, I., Miró, D. E., María, D., Martínez, P., & De Vaca, N. (2006). *Síndrome de apnea obstructiva del sueño y accidentes de tráfico*.
- Assi, K. J., Nahiduzzaman, K. M., Ratrou, N. T., & Aldosary, A. S. (2018). Mode choice behavior of high school goers: Evaluating logistic regression and MLP neural networks. *Case Studies on Transport Policy*, 6(2), 225–230. <https://doi.org/10.1016/j.cstp.2018.04.006>
- Barroso-García, V., Gutiérrez-Tobal, G. C., Gozal, D., Vaquerizo-Villar, F., Álvarez, D., del Campo, F., Kheirandish-Gozal, L., & Hornero, R. (2021). Wavelet Analysis of Overnight Airflow to Detect Obstructive Sleep Apnea in Children. *Sensors*, 21(4), 1491. <https://doi.org/10.3390/s21041491>

- Barroso-García, V., Gutiérrez-Tobal, G. C., Kheirandish-Gozal, L., Álvarez, D., Vaquerizo-Villar, F., Núñez, P., 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. <https://doi.org/10.1016/j.cmpb.2019.105083>
- 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. <https://doi.org/10.1016/j.combiomed.2020.104167>
- Barroso-García, V., Gutiérrez-Tobal, G., Kheirandish-Gozal, L., Álvarez, D., Vaquerizo-Villar, F., Crespo, A., Del Campo, F., Gozal, D., & Hornero, R. (2017). Irregularity and Variability Analysis of Airflow Recordings to Facilitate the Diagnosis of Paediatric Sleep Apnoea-Hypopnoea Syndrome. *Entropy*, 19(9), 447. <https://doi.org/10.3390/e19090447>
- Becerra, E. R., Botebol Benhamou, G., & Muñoz Villa, C. (1991). *ESTUDIO DE LOS TRASTORNOS RESPIRATORIOS DEL SUEÑO (POLISOMNOGRAFÍA)*.
- Behera, D., Suri, J. C., Ish, P., Behera, R. R., Chakrabarti, S., & Balasubramanian, V. (2021). Role of esophageal manometry in polysomnography of patients with chronic respiratory failure. *Monaldi Archives for Chest Disease*, 91(4). <https://doi.org/10.4081/monaldi.2021.1756>
- 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., Ward, S. L. D., & 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. *Journal of Clinical Sleep Medicine*, 08(05), 597–619. <https://doi.org/10.5664/jcsm.2172>
- Bitners, A. C., & Arens, R. (2020a). Evaluation and Management of Children with Obstructive Sleep Apnea Syndrome. *Lung*, 198(2), 257–270. <https://doi.org/10.1007/s00408-020-00342-5>
- Bitners, A. C., & Arens, R. (2020b). Evaluation and Management of Children with Obstructive Sleep Apnea Syndrome. In *Lung* (Vol. 198, Issue 2, pp. 257–270). Springer. <https://doi.org/10.1007/s00408-020-00342-5>
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England)*, 1(8476), 307–310. <http://www.ncbi.nlm.nih.gov/pubmed/2868172>
- Brodsky, L. (1989). Modern Assessment of Tonsils and Adenoids. *Pediatric Clinics of North America*, 36(6), 1551–1569. [https://doi.org/10.1016/S0031-3955\(16\)36806-7](https://doi.org/10.1016/S0031-3955(16)36806-7)
- Budhiraja, R., Goodwin, J. L., Parthasarathy, S., & Quan, S. F. (2005). Comparison of Nasal Pressure Transducer and Thermistor for Detection of Respiratory Events During Polysomnography in Children. *Sleep*, 28(9), 1117–1121. <https://doi.org/10.1093/sleep/28.9.1117>
- Buela-Casal G, B. M. (1999). EFECTOS DE LA ADHERENCIA AL TRATAMIENTO CON CPAP EN LA APNEA DEL SUEÑO SOBRE VARIABLES NEUROPSICOLÓGICAS Y PSICOFISIOLÓGICAS. In *Journal of Health Psychology* (Vol. 11, Issue 2).
- Burwell, C. S., Robin, E. D., Whaley, R. D., & Bickelmann, A. G. (1994). Extreme Obesity Associated with Alveolar Hypoventilation — A Pickwickian Syndrome*. *Obesity Research*, 2(4), 390–397. <https://doi.org/10.1002/j.1550-8528.1994.tb00084.x>
- Caples, S. M., Anderson, W. M., Calero, K., Howell, M., & Hashmi, S. D. (2021). Use of polysomnography and home sleep apnea tests for the longitudinal management of obstructive sleep apnea in adults: an American Academy of Sleep Medicine clinical guidance statement. *Journal of Clinical Sleep Medicine*, 17(6), 1287–1293. <https://doi.org/10.5664/jcsm.9240>

- Chang, H., Chen, Y., & Du, J. (2020). Obstructive sleep apnea treatment in adults. *The Kaohsiung Journal of Medical Sciences*, 36(1), 7–12. <https://doi.org/10.1002/kjm2.12130>
- 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. <https://doi.org/10.1016/j.ijporl.2012.11.028>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Cifci, M. A., Hussain, S., & Canatalay, P. J. (2023). Hybrid Deep Learning Approach for Accurate Tumor Detection in Medical Imaging Data. *Diagnostics*, 13(6), 1025. <https://doi.org/10.3390/diagnostics13061025>
- Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2015). *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. <http://arxiv.org/abs/1511.07289>
- Dehkordi, P. K., 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*. <https://doi.org/10.3929/ethz-b-000107756>
- Dickens, C. (1837). *The Posthumous Papers of the Pickwick club* (Boz, Ed.; 1st Edition). Chapman & Hall.
- Dickens, C. (2008). *The Pickwick Papers* (J. Kinsley, Ed.). Oxford University Press. <https://doi.org/10.1093/owc/9780199536245.001.0001>
- D’Rozario, A. L., Cross, N. E., Vakulin, A., Bartlett, D. J., Wong, K. K. H., Wang, D., & Grunstein, R. R. (2017). Quantitative electroencephalogram measures in adult obstructive sleep apnea – Potential biomarkers of neurobehavioural functioning. *Sleep Medicine Reviews*, 36, 29–42. <https://doi.org/10.1016/j.smrv.2016.10.003>
- Eguía, V. M., Cascante, J. A., Manuel, V., & Astibia, E. (2007). Síndrome de apnea-hipopnea del sueño. Concepto, diagnóstico y tratamiento médico Sleep apnea-hypopnea syndrome. Concept, diagnosis and medical treatment. In *An. Sist. Sanit. Navar* (Vol. 30, Issue 1).
- García-Vicente, C., Gutiérrez-Tobal, G. C., Jiménez-García, J., Martín-Montero, A., Gozal, D., & Hornero, R. (2023). ECG-based convolutional neural network in pediatric obstructive sleep apnea diagnosis. *Computers in Biology and Medicine*, 167, 107628. <https://doi.org/10.1016/j.compbiomed.2023.107628>
- 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. <https://doi.org/10.1371/journal.pone.0112959>
- Garde, A., Hoppenbrouwer, X., Dehkordi, P., Zhou, G., Rollinson, A. U., Wensley, D., Dumont, G. A., & Ansermino, J. M. (2019). Pediatric pulse oximetry-based OSA screening at different thresholds of the apnea-hypopnea index with an expression of uncertainty for inconclusive classifications. *Sleep Medicine*, 60, 45–52. <https://doi.org/10.1016/j.sleep.2018.08.027>
- Gil, B., Gómez Jiménez, F. J., & Soto Más, J. A. (2005). *Hipertensión refractaria. Obesidad y síndrome de apnea del sueño* (Vol. 22, Issue 3).
- Gil, E., Bailon, 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. <https://doi.org/10.1109/TBME.2009.2037734>
- Gil, E., Mendez, M., Vergara, J. M., Cerutti, S., Bianchi, A. M., & Laguna, P. (2009).

- Discrimination of Sleep-Apnea-Related Decreases in the Amplitude Fluctuations of PPG Signal in Children by HRV Analysis. *IEEE Transactions on Biomedical Engineering*, 56(4), 1005–1014. <https://doi.org/10.1109/TBME.2008.2009340>
- Glorot, X., Bordes, A., & Bengio, Y. (2011). *Deep Sparse Rectifier Neural Networks*.
- Gonzalez-Hernandez, J., Armas-Álvarez, I., Coronel-Lazcano, B., Vergara-Martínez, M., Maldonado-López, O., & Granillo-Macías, N. (2021). El desarrollo tecnológico en las revoluciones industriales Technological development in industrial revolutions. In *Publicación semestral* (Vol. 8, Issue 16). <https://repository.uaeh.edu.mx/revistas/index.php/sahagun/issue/archive>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*.
- Gozal, D., Hornero, R., Vaquerizo-Villar, F., Álvarez, D., Kheirandish-Gozal, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., & del Campo, F. (2018). Wavelet analysis of oximetry recordings to assist in the automated detection of moderate-to-severe pediatric sleep apnea-hypopnea syndrome. *PLoS ONE*, 13(12). <https://doi.org/10.1371/journal.pone.0208502>
- Grimaldi, F. P., Cabrera Galán, C., Carmona García, R., & Lorenzo, A. C. M. (2011). *Exploración funcional respiratoria: aplicación clínica*.
- Guilleminault, C. (2018). *Clinical overview of the sleep apnea syndrome*. <https://www.researchgate.net/publication/262188956>
- Guilleminault, C., & Abad, V. C. (2004). Obstructive sleep apnea syndromes. *Medical Clinics of North America*, 88(3), 611–630. <https://doi.org/10.1016/j.mcna.2004.01.002>
- Guilleminault, C., Tilkian, A., & Dement, W. C. (1977). The sleep apnea syndromes. *Annu Rev Med*.
- 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. <https://doi.org/10.1016/j.bspc.2015.02.014>
- Gutiérrez-Tobal, G. C., Álvarez, D., Kheirandish-Gozal, L., del Campo, F., Gozal, D., & Hornero, R. (2022). Reliability of machine learning to diagnose pediatric obstructive sleep apnea: Systematic review and meta-analysis. *Pediatric Pulmonology*, 57(8), 1931–1943. <https://doi.org/10.1002/ppul.25423>
- Gutiérrez-Tobal, G. C., Álvarez, D., Vaquerizo-Villar, F., Barroso-García, V., Martín-Montero, A., Crespo, A., del Campo, F., & Hornero, R. (2019). Pulse Rate Variability Analysis to Enhance Oximetry as at-Home Alternative for Sleep Apnea Diagnosing. In *IFMBE Proceedings* (Vol. 68, Issue 2, pp. 213–217). Springer Verlag. https://doi.org/10.1007/978-981-10-9038-7_39
- He, K., Zhang, X., Ren, S., & Sun, J. (2015a). *Deep Residual Learning for Image Recognition*. <http://arxiv.org/abs/1512.03385>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015b). *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. <http://arxiv.org/abs/1502.01852>
- Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, 12(1), 5979. <https://doi.org/10.1038/s41598-022-09954-8>
- Ho, M. L., & Brass, S. D. (2011). Obstructive sleep apnea. *Neurology International*, 3(3), 15. <https://doi.org/10.4081/ni.2011.e15>
- 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., ... Gozal, D. (2017). Nocturnal Oximetry-based Evaluation of Habitually Snoring Children. *American Journal of Respiratory and Critical Care Medicine*, 196(12), 1591–1598. <https://doi.org/10.1164/rccm.201705-0930OC>
- Huber, P. J. (1963). *Robust Estimation of a Location Parameter*. <https://doi.org/10.1214/aoms/1177703707>
- İŞCAN, Z. (2022). Comparison of Deep Learning and Traditional Machine Learning Classification Performance in a SSVEP Based Brain Computer Interface. *Balkan Journal of Electrical and Computer Engineering*, 10(3), 347–355. <https://doi.org/10.17694/bajece.1088353>
- Jaime, M. C. J., & Elizondo, E. (2002). *Fundamentos de Procesamiento de Imágenes INSTITUTO TECNOLÓGICO DE CIUDAD MADERO*.
- Jiménez-García, J., García, M., Gutiérrez-Tobal, G. C., Kheirandish-Gozal, L., Vaquerizo-Villar, F., Álvarez, D., del Campo, F., Gozal, D., & Hornero, R. (2022). A 2D convolutional neural network to detect sleep apnea in children using airflow and oximetry. *Computers in Biology and Medicine*, 147, 105784. <https://doi.org/10.1016/j.compbiomed.2022.105784>
- Jiménez-García, J., García, M., Gutiérrez-Tobal, G. C., Kheirandish-Gozal, L., Vaquerizo-Villar, F., Álvarez, D., del Campo, F., Gozal, D., & Hornero, R. (2024). An explainable deep-learning architecture for pediatric sleep apnea identification from overnight airflow and oximetry signals. *Biomedical Signal Processing and Control*, 87, 105490. <https://doi.org/10.1016/j.bspc.2023.105490>
- 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. <https://doi.org/10.3390/e22060670>
- Jubran, A. (2015). Pulse oximetry. *Critical Care*, 19(1). <https://doi.org/10.1186/s13054-015-0984-8>
- Kaditis, A. G., Alonso Alvarez, M. L., Boudewyns, A., Alexopoulos, E. I., Ersu, R., Joosten, K., Larramona, H., Miano, S., Narang, I., Trang, H., Tsaoussoglou, M., Vandenbussche, N., Villa, M. P., Van Waardenburg, D., Weber, S., & Verhulst, S. (2016). Obstructive sleep disordered breathing in 2- to 18-year-old children: diagnosis and management. *European Respiratory Journal*, 47(1), 69–94. <https://doi.org/10.1183/13993003.00385-2015>
- Kelleher, J. D. (2019). *DEEP LEARNING*. Massachusetts Institute of Technology.
- Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. <http://arxiv.org/abs/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. <https://doi.org/10.1378/chest.124.5.1702>
- Konrad, B. (2006). *Alternatives to CPAP in the treatment of the obstructive sleep apnoea syndrome*.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lazaro, J., Gil, E., Vergara, J. M., & Laguna, P. (2014). Pulse Rate Variability Analysis for Discrimination of Sleep-Apnea-Related Decreases in the Amplitude Fluctuations of Pulse Photoplethysmographic Signal in Children. *IEEE Journal of Biomedical and Health*

- Informatics*, 18(1), 240–246. <https://doi.org/10.1109/JBHI.2013.2267096>
- Leske, V. (2014). Polisonmnografía en pediatría. *FASO AÑO 21 - Nº 2*.
- Ling, V., & Wu, C. (2024, May). *Sleep Apnea Statistic and Facts You Should Know*. National Council On Aging (NCOA).
- Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier Nonlinearities Improve Neural Network Acoustic Models*.
- Malhotra, A., Ayappa, I., Ayas, N., Collop, N., Kirsch, D., Mcardle, N., Mehra, R., Pack, A. I., Punjabi, N., White, D. P., & Gottlieb, D. J. (2021). Metrics of sleep apnea severity: beyond the apnea-hypopnea index. *Sleep*, 44(7). <https://doi.org/10.1093/sleep/zsab030>
- Malhotra, A., & White, D. P. (2002). Obstructive sleep apnoea. *The Lancet*, 360(9328), 237–245. [https://doi.org/10.1016/S0140-6736\(02\)09464-3](https://doi.org/10.1016/S0140-6736(02)09464-3)
- Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3), 354–381. <https://academic.oup.com/eurheartj/article/17/3/354/485572>
- Marcus, C. L., Brooks, L. J., Draper, K. A., Gozal, D., Halbower, A. C., Jones, J., Schechter, M. S., Sheldon, S. H., Spruyt, K., Ward, S. D., Lehmann, C., & Shiffman, R. N. (2012). Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics*, 130(3), 576–584. <https://doi.org/10.1542/peds.2012-1671>
- 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., ... Redline, S. (2013). A Randomized Trial of Adenotonsillectomy for Childhood Sleep Apnea. *New England Journal of Medicine*, 368(25), 2366–2376. <https://doi.org/10.1056/NEJMoa1215881>
- Martín Montero, A. (2024). *Heart rate variability analysis in pediatric obstructive sleep apnea: Automated signal processing to identify biomarkers and help in the diagnosis* [Universidad de Valladolid]. <https://doi.org/10.35376/10324/66849>
- Martín-Montero, A., Armañac-Julián, P., Gil, E., Kheirandish-Gozal, L., Álvarez, D., Lázaro, J., Bailón, R., Gozal, D., Laguna, P., Hornero, R., & Gutiérrez-Tobal, G. C. (2023). Pediatric sleep apnea: Characterization of apneic events and sleep stages using heart rate variability. *Computers in Biology and Medicine*, 154, 106549. <https://doi.org/10.1016/j.compbiomed.2023.106549>
- Martín-Montero, A., Gutiérrez-Tobal, G. C., Gozal, D., Barroso-García, V., Álvarez, D., del Campo, F., Kheirandish-Gozal, L., & Hornero, R. (2021). Bispectral Analysis of Heart Rate Variability to Characterize and Help Diagnose Pediatric Sleep Apnea. *Entropy*, 23(8), 1016. <https://doi.org/10.3390/e23081016>
- Martín-Montero, A., Gutiérrez-Tobal, G. C., Kheirandish-Gozal, L., Vaquerizo-Villar, F., Álvarez, D., del Campo, F., Gozal, D., & Hornero, R. (2022). Heart rate variability as a potential biomarker of pediatric obstructive sleep apnea resolution. *Sleep*, 45(2). <https://doi.org/10.1093/sleep/zsab214>
- Mazzotti, D. R., Lim, D. C., Sutherland, K., Bittencourt, L., Mindel, J. W., Magalang, U., Pack, A. I., de Chazal, P., & Penzel, T. (2018). Opportunities for utilizing polysomnography signals to characterize obstructive sleep apnea subtypes and severity. *Physiological Measurement*, 39(9), 09TR01. <https://doi.org/10.1088/1361-6579/aad5fe>
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 276–282. <https://doi.org/10.11613/BM.2012.031>

- Mediano, O., González Mangado, N., Montserrat, J. M., Alonso-Álvarez, M. L., Almendros, I., Alonso-Fernández, A., Barbé, F., Borsini, E., Caballero-Eraso, C., Cano-Pumarega, I., de Carlos Villafranca, F., Carmona-Bernal, C., Carrillo Alduenda, J. L., Chiner, E., Cordero Guevara, J. A., de Manuel, L., Durán-Cantolla, J., Farré, R., Franceschini, C., ... Egea, C. (2022). Documento internacional de consenso sobre apnea obstructiva del sueño. *Archivos de Bronconeumología*, 58(1), 52–68. <https://doi.org/10.1016/j.arbres.2021.03.017>
- Mildenhall, J. L. (2008). The theory and application of pulse oximetry. *Journal of Paramedic Practice*, 1(2), 52–58. <https://doi.org/10.12968/jpar.2008.1.2.42016>
- Miquel-gomara Perelló, J., Román Rodríguez, M., & Grupo Respiratorio De La Societat Balear De Medicina Familiar I Comunitaria, Y. DE. (2002). Peak-flow meter: technique and utilities in Pri-mary Health Care TÉCNICAS Y PROCEDIMIENTOS. In *MEDIFAM* (Vol. 12).
- Moffa, A., Rinaldi, V., Costantino, A., Cassano, M., Gelardi, M., Fiore, V., Lopez, M. A., Baptista, P., Campisi, P., & Casale, M. (2020). Childhood Obstructive Sleep Apnea: from Diagnosis to Therapy—an Update. *Current Sleep Medicine Reports*, 6(3), 157–162. <https://doi.org/10.1007/s40675-020-00182-y>
- Mortazavi, E., Tarvirdizadeh, B., Alipour, K., & Ghamari, M. (2024). Deep learning approaches for assessing pediatric sleep apnea severity through SpO2 signals. *Scientific Reports*, 14(1), 22696. <https://doi.org/10.1038/s41598-024-67729-9>
- Najarian, K., & Splinter, R. (2012). *Second Edition Biomedical Signal and Image Processing*.
- National Heart, L. and B. Institute. (2016). *What's Sleep apnea?*. National Heart, Lung and Blood Institute.
- Nitzan, M., & Taitelbaum, H. (2008). *IEEE Instrumentation & Measurement Magazine The Measurement of Oxygen Saturation in Arterial and Venous Blood*.
- Okunogbe, A., Nugent, R., Spencer, G., Powis, J., Ralston, J., & Wilding, J. (2022). Economic impacts of overweight and obesity: current and future estimates for 161 countries. *BMJ Global Health*, 7(9), e009773. <https://doi.org/10.1136/bmjgh-2022-009773>
- Olivi, H. (2013). *APNEA DEL SUEÑO: CUADRO CLÍNICO Y ESTUDIO DIAGNÓSTICO SLEEP APNOEA: CLINICAL PRESENTATION AND DIAGNOSTIC ALGORITHMS*.
- OMS. (2024, March). *OMS: Obesidad y Sobrepeso*.
- O'Shea, K., & Nash, R. (2015). *An Introduction to Convolutional Neural Networks*. <https://arxiv.org/abs/1511.08458>
- Ostrosky, F. (2010). *Desarrollo del cerebro*.
- Padilla, J., Uceda, M., Ziegler, O., Lindo, F., Herrera-Pérez, E., & Huicho, L. (2013). Association between Allergic Rhinitis and Asthma Control in Peruvian School Children: A Cross-Sectional Study. *BioMed Research International*, 2013, 1–7. <https://doi.org/10.1155/2013/861213>
- Pellegrino, R., Viegi, G., Brusasco, V., Crapo, R. O., Burgos, F., Casaburi, R., Coates, A., van der Grinten, C. P. M., Gustafsson, P., Hankinson, J., Jensen, R., Johnson, D. C., MacIntyre, N., McKay, R., Miller, M. R., Navajas, D., Pedersen, O. F., & Wanger, J. (2005). Interpretative strategies for lung function tests. *European Respiratory Journal*, 26(5), 948–968. <https://doi.org/10.1183/09031936.05.00035205>
- Perumal, T., Mustapha, N., Mohamed, R., & Shiri, F. M. (2024). A Comprehensive Overview and Comparative Analysis on Deep Learning Models. *Journal on Artificial Intelligence*, 6(1), 301–360. <https://doi.org/10.32604/jai.2024.054314>
- Pevernagie, D. A., Gnidovec-Strazisar, B., Grote, L., Heinzer, R., McNicholas, W. T., Penzel, T., Randerath, W., Schiza, S., Verbraecken, J., & Arnardottir, E. S. (2020). On the rise and fall

- of the apnea–hypopnea index: A historical review and critical appraisal. *Journal of Sleep Research*, 29(4). <https://doi.org/10.1111/jsr.13066>
- Prechelt, L. (1998). *Early Stopping - But When?* (pp. 55–69). https://doi.org/10.1007/3-540-49430-8_3
- Qin, H., Steenbergen, N., Glos, M., Wessel, N., Kraemer, J. F., Vaquerizo-Villar, F., & Penzel, T. (2021). The Different Facets of Heart Rate Variability in Obstructive Sleep Apnea. *Frontiers in Psychiatry*, 12. <https://doi.org/10.3389/fpsyt.2021.642333>
- Rauniyar, N., Pujari, S., & Shrestha, P. (2020). Study of Oxygen Saturation by Pulse Oximetry and Arterial Blood Gas in ICU Patients: A Descriptive Cross-sectional Study. *Journal of Nepal Medical Association*, 58(230), 789–793. <https://doi.org/10.31729/jnma.5536>
- Redline, S., Tishler, P., Hans, M., Tosteson, T., Strohl, K., & Spry, K. (1997). Racial differences in sleep-disordered breathing in African-Americans. *Respir Crit Care Med*.
- Rosen, C. L., D'andrea, L., & Haddad, G. G. (1992). Adult Criteria for Obstructive Sleep Apnea Do Not Identify Children with Serious Obstruction 1-3. In *AM REV RESPIR DIS* (Vol. 146).
- Rouhiainen, L. (2018). *INTELIGENCIA ARTIFICIAL 101 COSAS QUE DEBES SABER HOY SOBRE NUESTRO FUTURO INTELIGENCIA ARTIFICIAL*. www.planetadelibros.com
- Ruder, S. (2016). *An overview of gradient descent optimization algorithms*. <http://arxiv.org/abs/1609.04747>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. <http://gradcam.cloudcv.org>
- Semeria, M. (2015). *Los tres teoremas fundamentales: Fourier - Nyquist y Shannon*. www.cema.edu.ar/publicaciones/doc_trabajo.html
- She, D., & Jia, M. (2021). A BiGRU method for remaining useful life prediction of machinery. *Measurement*, 167, 108277. <https://doi.org/10.1016/j.measurement.2020.108277>
- Shlezinger, N., & Eldar, Y. C. (2023). *Model-Based Deep Learning*. <http://arxiv.org/abs/2306.04469>
- 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. <https://doi.org/10.1093/sleep/27.4.784>
- Spannella, F., Giulietti, F., Di Pentima, C., Lombardi, F. E., Borioni, E., & Sarzani, R. (2017). Blood Pressure and Metabolic Changes After 3-Month CPAP Therapy in a Very Elderly Obese with Severe Obstructive Sleep Apnea: A Case Report and Review of the Literature. *High Blood Pressure & Cardiovascular Prevention*, 24(3), 341–346. <https://doi.org/10.1007/s40292-017-0190-7>
- Sturludóttir, J. E., Sigurðardóttir, S., Serwatko, M., Arnardóttir, E. S., Hrubos-Strøm, H., Clausen, M. V., Sigurðardóttir, S., Óskarsdóttir, M., & Islind, A. S. (2023). Deep learning for sleep analysis on children with sleep-disordered breathing: Automatic detection of mouth breathing events. *Frontiers in Sleep*, 2. <https://doi.org/10.3389/frsle.2023.1082996>
- Szeliski, R. (2011). *Computer Vision: Algorithms and Applications*. www.springer.com/series/3191
- 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. <https://doi.org/10.5665/sleep.3392>
- TensorFlow Team. (2024). *TensorFlow Callbacks* - v2.16.1.

- https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/Callback
- Terán Santos, J., Luz Alonso Álvarez, M., Cordero Guevara, J., Ayuela Azcárate, J. M., & Monserrat Canal, J. M. (2006). Síndrome de apneas-hipopneas durante el sueño y corazón. *Revista Española de Cardiología*, 59(7), 718–724. <https://doi.org/10.1157/13091373>
- Terán-Santos, J., Jimenez-Gomez, A., & Cordero-Guevara, J. (1999). The Association between Sleep Apnea and the Risk of Traffic Accidents. *New England Journal of Medicine*, 340(11), 847–851. <https://doi.org/10.1056/NEJM199903183401104>
- Torre-Bouscoulet, L., Castorena-Maldonado, A., Baños-Flores, R., Vázquez-García, J. C., Sonia Meza-Vargas, M., & Pérez-Padilla, R. (2007). Índice de desaturación de oxígeno frente a índice de apneas-hipopneas en adultos con sospecha de apnea obstructiva durante el sueño a 2.240 m de altitud. *Archivos de Bronconeumología*, 43(12), 649–654. <https://doi.org/10.1157/13112962>
- 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. <https://doi.org/10.1016/j.ijporl.2013.05.011>
- 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). Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings. *Computer Methods and Programs in Biomedicine*, 156, 141–149. <https://doi.org/10.1016/j.cmpb.2017.12.020>
- 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). 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. <https://doi.org/10.1371/journal.pone.0208502>
- Vaquerizo-Villar, F., Alvarez, D., Kheirandish-Gozal, L., Gutierrez-Tobal, G. C., Barroso-Garcia, V., Santamaria-Vazquez, E., Campo, F. del, 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*, 25(8), 2906–2916. <https://doi.org/10.1109/JBHI.2020.3048901>
- 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. <https://doi.org/10.1001/jamaoto.2015.2354>
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018a). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018, 1–13. <https://doi.org/10.1155/2018/7068349>
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018b). Deep Learning for Computer Vision: A Brief Review. In *Computational Intelligence and Neuroscience* (Vol. 2018). Hindawi Limited. <https://doi.org/10.1155/2018/7068349>
- Wagner, G. S. ., & Strauss, D. G. . (2014). *Marriott's practical electrocardiography*. Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Wegener, A. (2017). Canula nasal de alto flujo en pediatría. *Neuromological Pediatrics* .
- Won, C. H. J., Li, K. K., & Guilleminault, C. (2008). Surgical Treatment of Obstructive Sleep Apnea: Upper Airway and Maxillomandibular Surgery. *Proceedings of the American Thoracic Society*, 5(2), 193–199. <https://doi.org/10.1513/pats.200708-121MG>

- Xia, F., & Sawan, M. (2021). Clinical and Research Solutions to Manage Obstructive Sleep Apnea: A Review. *Sensors*, 21(5), 1784. <https://doi.org/10.3390/s21051784>
- 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), 1801788. <https://doi.org/10.1183/13993003.01788-2018>
- Ye, P., Qin, H., Zhan, X., Wang, Z., Liu, C., Song, B., Kong, Y., Jia, X., Qi, Y., Ji, J., Chang, L., Ni, X., & Tai, J. (2023). Diagnosis of obstructive sleep apnea in children based on the XGBoost algorithm using nocturnal heart rate and blood oxygen feature. *American Journal of Otolaryngology*, 44(2), 103714. <https://doi.org/10.1016/j.amjoto.2022.103714>
- Ynzunza Cortés, C. B., Izar Landeta, J. M., Bocarando Chacón, J. G., Aguilar Pereyra, F., & Larios Osorio, M. (2017). *El Entorno de la Industria 4.0: Implicaciones y Perspectivas Futuras*. <https://www.redalyc.org/articulo.oa?id=94454631006>
- Zahn, D., Adams, J., Krohn, J., Wenzel, M., Mann, C. G., Gomille, L. K., Jacobi-Scherbening, V., & Kubiak, T. (2016). Heart rate variability and self-control—A meta-analysis. *Biological Psychology*, 115, 9–26. <https://doi.org/10.1016/J.BIOPSYCHO.2015.12.007>