

DetECCIÓN AUTOMÁTICA DE LA APNEA DEL SUEÑO INFANTIL UTILIZANDO TÉCNICAS DE *DEEP LEARNING* Y *EXPLAINABLE ARTIFICIAL INTELLIGENCE* EN SEÑALES DE FLUJO AÉREO

V. Barroso García^{1,2}, F. Vaquerizo Villar^{1,2}, G.C. Gutiérrez Tobal^{1,2}, C. García Vicente¹, D. Álvarez González^{1,2}, D. Gozal³, R. Hornero Sánchez^{1,2}

¹ Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España, veronica.barroso@gib.tel.uva.es

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), España

³ Office of The Dean, Joan C. Edwards School of Medicine, Marshall University, Huntington, Virginia Occidental, EEUU, gozal@marshall.edu

Resumen

La alta prevalencia de la apnea obstructiva del sueño (AOS) pediátrica y las limitaciones de la prueba diagnóstica estándar han fomentado el estudio de estrategias alternativas que ayuden a su diagnóstico automático. Los métodos propuestos suelen basarse en técnicas de *feature engineering*, lo que implica una complejidad y subjetividad inherente. Otros utilizan técnicas de *deep learning*, que mejoran el rendimiento diagnóstico pero carecen de transparencia e interpretabilidad. En este trabajo proponemos evaluar un modelo explicable basado en redes neuronales convolucionales (CNN) para estimar la severidad de la AOS infantil utilizando la señal de flujo (FA). Para ello, se analizaron 1638 registros de FA, que fueron divididos en segmentos de 10 minutos. El modelo CNN estimó el número de eventos apnéicos por segmento. Después, se aplicó el algoritmo Grad-CAM para identificar las regiones de FA en las que se fija la CNN al hacer su predicción. El modelo propuesto mostró una alta concordancia entre el índice de apnea-hipopnea estimado y el real (coeficiente de correlación intraclase = 0.87 en el grupo de test), así como un alto rendimiento diagnóstico (κ para 4 clases = 0.38 y precisiones del 81.05%, 85.62% y 92.81% para 1, 5 y 10 eventos/h en el grupo de test). Grad-CAM reveló que la CNN se centra en el comienzo y el final del evento apnéico, es decir, donde la señal FA cambia bruscamente de amplitud. Así, nuestra propuesta sería muy útil para identificar automáticamente los patrones respiratorios asociados con la AOS infantil y ayudar en su diagnóstico.

1. Introducción

La apnea obstructiva del sueño (AOS) es un trastorno respiratorio muy común y altamente prevalente. Se estima que entre el 1% y el 5.7% de los niños sufren algún grado de la enfermedad [1], aunque el porcentaje exacto puede variar según otros trastornos coexistentes [2], [3]. Los niños con AOS experimentan pausas recurrentes (apneas) y/o reducciones del flujo aéreo (hipopneas) mientras duermen, que pueden durar desde 6 segundos (2 ciclos respiratorios) hasta varios minutos cada una [4]. La presencia de estos eventos conduce a hipoxemia, hipercapnia, micro despertares y, en general, a un sueño fragmentado y no reparador que predispone a los sujetos pediátricos afectados a sufrir serias comorbilidades, como disfunciones cardiovasculares, metabólicas y

neurodegenerativas [1], [5]. Por eso, es de suma importancia que estos niños sean diagnosticados y tratados a tiempo.

El método de referencia para su diagnóstico es la polisomnografía nocturna (PSG), una prueba en la que se registra la actividad cerebral, la respiración, la frecuencia cardíaca y otros signos vitales durante el sueño [6]. Si bien su eficacia y fiabilidad son ampliamente conocidas, la PSG también es compleja, costosa, incómoda para los niños y de disponibilidad muy limitada [7]. Este hecho, unido a la alta prevalencia de la enfermedad, ha fomentado el estudio de otras estrategias que ayuden a diagnosticar la AOS automáticamente.

El procesado de señales cardiorrespiratorias, como la señal de flujo aéreo (FA), ha demostrado ser especialmente útil para este propósito [8]–[12]. Esta señal es capaz de reflejar la actividad respiratoria del niño mientras duerme y, por ende, los ceses respiratorios asociados con la AOS [4]. Es por ello que los métodos de *feature engineering* han resultado ser tan eficaces a la hora de caracterizar los cambios que experimenta esta señal en el dominio del tiempo y de la frecuencia cuando se producen los eventos apnéicos [8]–[12]. No obstante, es importante destacar que estos métodos requieren un conocimiento previo y exhaustivo de las características a extraer [13]. Además, al depender de características diseñadas manualmente, podría haber dificultades para identificar las relaciones complejas y no lineales presentes en los datos, perdiendo así información relevante [13]. Por el contrario, los enfoques basados en *deep learning* (DL) podrían superar estas limitaciones y mejorar la capacidad diagnóstica del FA. Así, las redes neuronales convolucionales (CNN) han mostrado un alto rendimiento diagnosticando la AOS infantil a partir de señales cardiorrespiratorias [14], [15]. Si bien estos modelos son capaces de identificar automáticamente las características útiles de la señal en crudo, hay que tener en cuenta que pierden interpretabilidad y transparencia con respecto a aquellos basados en *feature engineering* [16]. Para lidiar con esta limitación, en este estudio proponemos aplicar un modelo CNN a la señal de FA junto con técnicas de *explainable artificial intelligence* (XAI) que permitan

comprender los resultados reportados y mejorar la fiabilidad de dicho modelo. Concretamente, aplicamos el algoritmo *Gradient-weighted Class Activation Mapping* (Grad-CAM) dado su éxito dotando de interpretabilidad a este tipo de redes [16]. Así, los objetivos de nuestro estudio fueron: (i) evaluar la utilidad de un modelo CNN para determinar la severidad de AOS pediátrica a partir del FA e (ii) identificar las características de esta señal en las que se fija la CNN para hacer su predicción.

2. Sujetos y señales

Nuestra investigación involucró 1638 registros polisomnográficos procedentes de la base de datos pública *Childhood Adenotonsillectomy Trial* (CHAT, #NCT00560859), la cual está disponible bajo petición en <https://sleepdata.org/datasets/chat>. El Comité Ético aprobó el protocolo de estudio y todos los tutores legales de los niños dieron su consentimiento informado.

Los sujetos fueron diagnosticados siguiendo las reglas de la *American Academy of Sleep Medicine* (AASM) [4]. A partir del índice de apnea-hipopnea (IAH) obtenido, los especialistas médicos determinaron si estos niños estaban libres de AOS (1 evento/hora (e/h) > IAH), o si padecían AOS leve (1 e/h ≤ IAH < 5 e/h), moderada (5 e/h ≤ IAH < 10 e/h) o severa (IAH ≥ 10 e/h). La Tabla 1 muestra sus datos demográficos y clínicos. La población bajo estudio fue distribuida aleatoriamente en tres grupos: entrenamiento (61.4%), validación (19.9%) y test (18.7%). No se encontraron diferencias estadísticamente significativas (*p*-valor > 0.01 al aplicar el test de Mann-Whitney *U* con corrección de Bonferroni) en edad, sexo, índice de masa corporal (IMC) ni IAH entre los grupos establecidos.

Los registros de FA utilizados se adquirieron con un termistor a frecuencias de muestreo de entre 20 y 512 Hz durante la PSG. Después, éstos fueron remuestreados a 4 Hz y estandarizados como en estudios previos [14]. Finalmente, se segmentaron en tramas de 10 minutos para alimentar la CNN propuesta en este estudio.

3. Métodos

3.1. Arquitectura del modelo CNN

La arquitectura propuesta en nuestro estudio está inspirada en el modelo CNN desarrollado en trabajos previos para estimar la severidad de AOS infantil a partir de la señal de oximetría y/o FA [14], [15]. Así, la CNN recibe como entrada los segmentos de 10 minutos de la señal de FA previamente estandarizados [14], [15]. Después, los datos son procesados mediante *N_c* bloques convolucionales, cada uno de ellos formado por las siguientes capas consecutivas: convolución 1D, *batch normalization*, *Rectified Linear Unit* (ReLU), *max pooling* y *dropout* [14], [15]. Así, en cada bloque se aplican *N_f* filtros a los datos de entrada para extraer los mapas de características (convolución), seguido de la normalización de estos (*batch normalization*) y de una activación no lineal para identificar las relaciones complejas existentes en ellos (ReLU) [13]. Después, se

| | Entrenamiento | Validación | Test |
|--------------------------------|---------------|-------------|-------------|
| Sujetos (<i>n</i>) | 1006 (61.4%) | 326 (19.9%) | 306 (18.7%) |
| Edad (años) | 7.0 [2.0] | 7.0 [2.0] | 6.9 [2.0] |
| Varones (<i>n</i>) | 471 (46.8%) | 156 (47.9%) | 134 (43.8%) |
| IMC (kg/m²) | 17.4 [6.1] | 17.1 [6.4] | 17.6 [6.0] |
| IAH (e/h) | 2.6 [4.8] | 2.4 [4.6] | 2.3 [5.1] |
| No AOS (<i>n</i>) | 219 (21.8%) | 69 (21.2%) | 67 (21.9%) |
| AOS Leve (<i>n</i>) | 496 (49.3%) | 168 (51.5%) | 148 (48.4%) |
| AOS Moderado (<i>n</i>) | 160 (15.9%) | 44 (13.5%) | 49 (16.0%) |
| AOS Severo (<i>n</i>) | 131 (13.0%) | 45 (13.8%) | 42 (13.7%) |

IMC: índice de masa corporal; IAH: índice de apnea-hipopnea. Datos presentados como número de sujetos (%) o mediana [rango intercuartil].

Tabla 1. Datos demográficos y clínicos de los sujetos.

reduce la dimensionalidad del mapa de características extrayendo aquellas que son dominantes (*max pooling*) y se reduce la interdependencia entre neuronas para prevenir el sobreajuste eliminando aleatoriamente una proporción *p_{drop}* de activaciones (*dropout*) [13].

Tras los *N_c* bloques de convolución, una capa de *flattening* concatena los mapas de características en un único vector unidimensional y, finalmente, una capa de activación lineal realiza la combinación lineal de sus elementos para estimar el número de eventos apneicos contenidos en cada segmento de 10 minutos [14], [15]. El IAH de cada sujeto se obtuvo como el promedio de los valores estimados por la CNN para todos los segmentos contenidos en su señal de FA, seguido de un proceso de regresión lineal para corregir posibles subestimaciones.

Para obtener la configuración óptima de los hiperparámetros *N_c*, *N_f* y *p_{drop}*, se utilizó el algoritmo de estimación del momento adaptativo (método Adam) con una tasa de aprendizaje inicial de 0.001 [15]. Al igual que en estudios previos, se empleó la función *Huber loss* con el parámetro *delta* = 1, dada su idoneidad en distribuciones aleatorias [14].

3.2. Evaluación del modelo CNN

Para evaluar la concordancia entre los IAH estimados por la CNN y los valores reales obtenidos mediante la PSG, se empleó el coeficiente de correlación intraclase (CCI) y el error cuadrático medio (*mean-squared error*, MSE). El rendimiento diagnóstico se evaluó empleando el *kappa* de 4 clases, la matriz de confusión y la precisión de 4 clases. También se obtuvo la sensibilidad (S), especificidad (E), precisión (P), valor predictivo positivo y negativo (VPP y VPN) y razón de verosimilitud positiva y negativa (LR+ y LR-) para los puntos de corte del IAH 1, 5 y 10 e/h.

3.3. Interpretabilidad del modelo CNN

Se aplicó Grad-CAM para dotar de interpretabilidad a nuestro modelo CNN y comprender cómo éste realiza sus predicciones [16]. Esta técnica de XAI proporciona una visualización intuitiva y contextual de las regiones de la señal de FA que son más influyentes en la estimación de la CNN. Para ello, calcula los gradientes de la predicción con respecto a los mapas de características presentes en cada capa convolucional de la CNN [16]. A partir de estos gradientes, se obtiene un mapa de activación por cada

capa convolucional, es decir, un mapa de calor que resalta con valores más elevados (colores más cálidos) las regiones del FA que más contribuyen a la predicción final de la red. El mapa de calor final se calcula promediando los mapas de activación obtenidos en cada capa.

4. Resultados

4.1. Hiperparámetros óptimos del modelo CNN

Los valores de N_c , N_f y p_{drop} variaron entre [6:1:10], [32, 64, 128, 256] y [0.0, 0.1, 0.2], respectivamente. Cada uno de los modelos configurados con estos valores fueron entrenados en el grupo de entrenamiento. La configuración que obtuvo mayor índice $kappa$ en el grupo de validación fue establecida como óptima: $N_c = 9$, $N_f = 256$ y $p_{drop} = 0.01$, con un $kappa = 0.42$.

4.2. Evaluación del modelo CNN en test

Las Tablas 2 y 3 muestran la matriz de confusión obtenida por la CNN en test, así como las métricas estadísticas derivadas de ella, respectivamente. La arquitectura propuesta alcanzó un CCI de 0.87 y un MSE de 7.90 e/h en test. Además, obtuvo una precisión global del 61.76% (189/306 sujetos correctamente clasificados) y un $kappa$ de 0.38 en test con cuatro clases. Aplicando los umbrales diagnósticos convencionales, nuestro modelo logró precisiones del 81.05%, 85.62% y 92.81% para 1, 5 y 10 e/h en test, respectivamente. También destaca la elevada LR+ alcanzada para 10 e/h (31.43).

| | | CNN | | | |
|-----|----------|--------|------|----------|--------|
| | | No AOS | Leve | Moderado | Severo |
| PSC | No AOS | 19 | 48 | 0 | 0 |
| | Leve | 10 | 127 | 11 | 0 |
| | Moderado | 0 | 26 | 18 | 5 |
| | Severo | 0 | 7 | 10 | 25 |

Tabla 2. Matriz de confusión obtenida por el modelo CNN en el grupo de test.

| IAH | S (%) | E (%) | P (%) | VPP (%) | VPN (%) | LR+ | LR- |
|--------|-------|-------|-------|---------|---------|-------|------|
| 1 e/h | 95.82 | 28.36 | 81.05 | 82.67 | 65.52 | 1.34 | 0.15 |
| 5 e/h | 63.74 | 94.88 | 85.62 | 84.06 | 86.08 | 12.46 | 0.38 |
| 10 e/h | 59.52 | 98.11 | 92.81 | 83.33 | 93.84 | 31.43 | 0.41 |

Tabla 3. Rendimiento diagnóstico obtenido para 1, 5 y 10 e/h por el modelo CNN en el grupo de test.

4.3. Interpretabilidad del modelo CNN en test

La Figura 1 muestra cuatro ejemplos de los resultados obtenidos con Grad-CAM en segmentos donde la predicción del modelo CNN fue correcta: (a) respiración normal y (b) eventos apneicos, y en aquellos donde la predicción fue errónea: (c) sobrestimación y (d) infraestimación. Como puede observarse en (b), para identificar el evento de apnea el modelo se centra principalmente en el comienzo y el final del mismo.

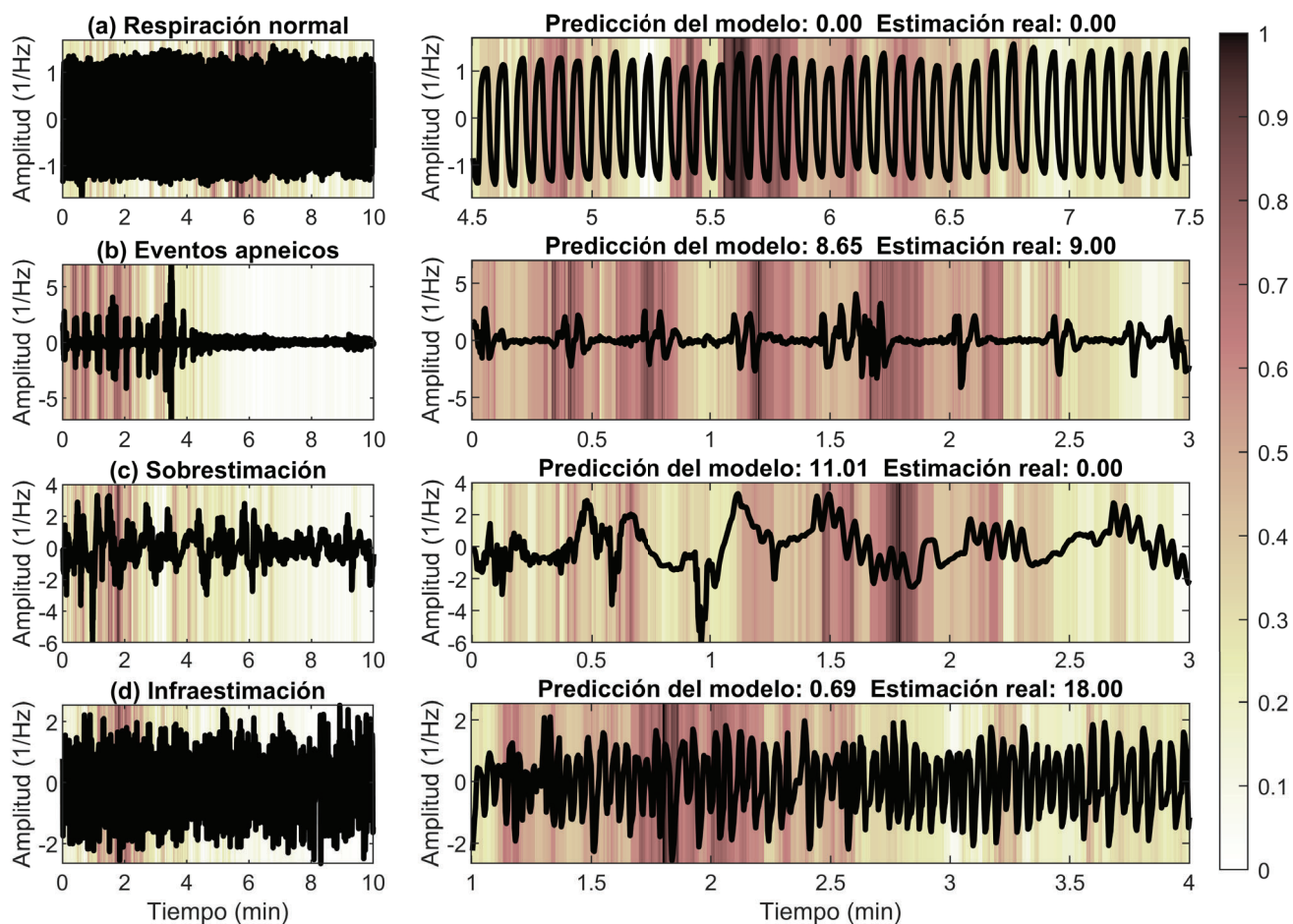


Figura 1. Resultados de Grad-CAM en (a) respiración normal, (b) eventos apneicos, (c) sobrestimación y (d) infraestimación.

5. Discusión y conclusiones

En este estudio se ha desarrollado y testado un modelo interpretable basado en CNN para estimar la severidad de la AOS infantil a partir de la señal de FA. Así, el modelo propuesto alcanzó una alta precisión diagnóstica (81.05%, 85.62% y 92.81% para 1, 5 y 10 e/h, respectivamente), especialmente en los casos de AOS moderada y severa, donde obtuvimos un robusto marcador para determinar su presencia ($LR+ \geq 10$) [11]. Además, Grad-CAM nos permitió identificar los patrones respiratorios en los que se fija la CNN y comprender los resultados que reporta. Tal y como muestra la Figura 1(a), cuando la respiración es normal el mapa de calor alcanza unos valores inferiores, por lo que la CNN se estaría centrando menos en estas regiones. No obstante, aparece un foco de atención en torno al minuto 5.6. La ausencia de cambios de amplitud en esta parte de la señal indicaría que la CNN se fija en otras propiedades del FA, como la distancia entre ciclos respiratorios. En el caso de los eventos apneicos (Figura 1(b)), podemos observar que los valores del mapa de calor se incrementan. Este aumento se mantiene a lo largo de todo el evento, aunque es particularmente más elevado al comienzo y al final. Este hecho revela que, para identificar el patrón respiratorio asociado al evento de apnea, el modelo se fija principalmente en aquellas regiones donde la señal de FA cambia bruscamente de amplitud. Sin embargo, los cambios repentinos de amplitud también pueden ser ocasionados por la presencia de ciertos artefactos, como el ruido fisiológico producido por el movimiento del niño. Como se muestra en la Figura 1(c), estos artefactos podrían conducir a una sobreestimación de los eventos apneicos. Finalmente, Grad-CAM nos permitió comprender los casos de infraestimación del modelo (Figura 1(d)). Así, los resultados obtenidos indicaron que, sin tener información sobre los *arousals* o las desaturaciones, las reducciones parciales del FA que se producen por la ocurrencia de hipopneas podrían pasar desapercibidas para la CNN.

Estudios previos de nuestro grupo analizaron la señal de FA para estimar automáticamente la severidad de AOS en niños [8]–[11]. La mayoría de ellos utilizaron métodos convencionales que alcanzaron precisiones de 73.61%–81.10% para 1 e/h, 57.14%–62.82% para 5 e/h y 70.47%–80.60% para 10 e/h. En este trabajo, se han alcanzado precisiones dentro de este rango para 1 e/h, así como notablemente superiores para 5 y 10 e/h. Otro estudio analizó conjuntamente FA con la señal de oximetría mediante CNN [14], obteniendo 84.10%–84.64%, 84.10%–93.46% y 90.26%–94.44% de precisión en 1, 5 y 10 e/h, respectivamente. Con sólo un canal, en el presente estudio se ha alcanzado un rendimiento sólo ligeramente inferior usando los mismos registros de CHAT.

Este estudio presenta algunas limitaciones que deben ser consideradas. El modelo CNN fue testado únicamente con CHAT, por lo que podría ser validado en otras bases de datos, e incluso con registros domiciliarios de FA, para aumentar la solidez de nuestros resultados. También se podría estimar el IAH utilizando otros métodos de DL, como redes neuronales recurrentes o *transformers*, y otras

técnicas de XAI, como LIME o SHAP, lo que constituye una interesante línea de investigación futura.

En conclusión, el modelo propuesto mostró un alto rendimiento estimando la severidad de la AOS infantil. La mejora obtenida respecto a las técnicas convencionales, así como la identificación de las regiones del FA que más contribuyen a la predicción, pone de manifiesto la eficacia y fiabilidad de usar conjuntamente CNN y FA para simplificar el diagnóstico de AOS en niños. Así, nuestra propuesta sería una herramienta útil para ello.

Agradecimientos

Este estudio ha sido financiado por el Ministerio de Ciencia e Innovación - AEI y ERDF (PID2020-115468RB-I00 y PDC2021-120775-I00) y por el CIBER-BBN (CB19/01/00012). G.C. Gutiérrez Tobal ha recibido una ayuda postdoctoral de la Universidad de Valladolid. D. Álvarez es beneficiario de una ayuda Ramón y Cajal (RYC2019-028566-I) del Ministerio de Ciencia e Innovación - AEI, cofinanciada por el FSE. D. Gozal ha recibido financiación del NIH (AG061824 y HL166617).

Referencias

- [1] C. L. Marcus *et al.*, “Diagnosis and management of childhood obstructive sleep apnea syndrome,” *Pediatrics*, vol. 130, 2012.
- [2] A. Moffa *et al.*, “Childhood Obstructive Sleep Apnea: from Diagnosis to Therapy—an Update,” *Current Sleep Medicine Reports*, vol. 6, no. 3, pp. 157–162, 2020.
- [3] A. C. Bitners *et al.*, “Evaluation and Management of Children with Obstructive Sleep Apnea Syndrome,” *Lung*, vol. 198, no. 2, pp. 257–270, 2020.
- [4] R. B. Berry *et al.*, “Rules for Scoring Respiratory Events in Sleep: Update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events,” *J. Clin. Sleep Med.*, vol. 08, no. 05, pp. 597–619, 2012.
- [5] M. Luz Alonso-Alvarez *et al.*, “Consensus document on sleep apnea-hypopnea syndrome in children,” *Arch. Bronconeumol.*, vol. 47, no. SUPPL. 5, pp. 2–18, 2011.
- [6] C. Jon, “Polysomnography in Children,” in *Pediatric Otolaryngology for the Clinician*, Totowa, NJ: Humana Press, 2009, pp. 35–47.
- [7] P. J. Ryan *et al.*, “Validation of British Thoracic Society guidelines for the diagnosis of the sleep apnoea/hypopnoea syndrome: can polysomnography be avoided?,” *Thorax*, vol. 50, no. 9, pp. 972–5, 1995.
- [8] J. Jiménez-García *et al.*, “Assessment of Airflow and Oximetry Signals to Detect Pediatric Sleep Apnea-Hypopnea Syndrome Using AdaBoost,” *Entropy*, vol. 22, no. 6, pp. 670, 2020.
- [9] V. Barroso-García *et al.*, “Wavelet Analysis of Overnight Airflow to Detect Obstructive Sleep Apnea in Children,” *Sensors*, vol. 21, no. 4, pp. 1491, 2021.
- [10] V. Barroso-García *et al.*, “Bispectral analysis of overnight airflow to improve the pediatric sleep apnea diagnosis,” *Comput. Biol. Med.*, vol. 129, 2021.
- [11] V. Barroso-García *et al.*, “Usefulness of recurrence plots from airflow recordings to aid in paediatric sleep apnoea diagnosis,” *Comput. Meth. Prog. Bio.*, vol. 183, pp. 105083, 2020.
- [12] G. C. Gutiérrez-Tobal *et al.*, “Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients’ home,” *Biomed. Signal Process. Control*, vol. 18, pp. 401–7, 2015.
- [13] I. Goodfellow *et al.*, *Deep learning*. MIT press, 2016.
- [14] J. Jiménez-García *et al.*, “A 2D convolutional neural network to detect sleep apnea in children using airflow and oximetry,” *Comput. Biol. Med.*, vol. 147, pp. 105784, 2022.
- [15] F. Vaquerizo-Villar *et al.*, “A Convolutional Neural Network Architecture to Enhance Oximetry Ability to Diagnose Pediatric Obstructive Sleep Apnea,” *IEEE J. Biomed. Heal. Inform.*, vol. 25, no. 8, pp. 2906–2916, 2021.
- [16] R. R. Selvaraju *et al.*, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Oct., pp. 618–626, 2017.