

Estimación de la severidad de apnea del sueño pediátrica a partir de la señal de flujo aéreo usando redes neuronales convolucionales

V. Barroso García^{1,2}, F. Vaquerizo Villar^{1,2}, G.C. Gutiérrez Tobal^{1,2}, D. Álvarez González^{2,3}, F. del Campo^{2,3}, L. Kheirandish Gozal⁴, D. Gozal⁴, 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 Servicio de Neumología, Hospital Universitario Río Hortega, Valladolid, España, fsas@telefonica.net

⁴ Department of Child Health, The University of Missouri School of Medicine, Columbia, Missouri, EEUU, gozald@health.missouri.edu

Resumen

El objetivo de este estudio fue evaluar la utilidad de un modelo de deep learning basado en redes neuronales convolucionales (convolutional neural networks, CNN) para estimar automáticamente la severidad de la apnea obstructiva del sueño (AOS) infantil a partir de la señal de flujo aéreo monocal (FA). Para ello, se analizaron 3158 registros pediátricos de FA: 1639 usados para entrenar el modelo (entrenamiento), 759 para optimizar sus hiperparámetros (validación) y 760 para evaluar su capacidad diagnóstica (test). Todos los registros fueron divididos en segmentos de 5 minutos. Para cada uno de estos segmentos, el modelo CNN estimó su correspondiente índice de apnea-hipopnea (IAH). Después, se obtuvo el IAH de cada sujeto pediátrico como el promedio de los valores estimados por la CNN para los segmentos contenidos en la señal de FA. El modelo propuesto alcanzó una precisión global del 53.42% y un Cohen's kappa de 0.31 en test con cuatro clases. Respecto a la evaluación del rendimiento diagnóstico por grupos de severidad, la CNN obtuvo una precisión del 76.18% para 1 evento/h, 81.45% para 5 eventos/h y 90.66% para 10 eventos/h en test. Además, también destaca el alto ratio de verosimilitud positiva alcanzado para 10 eventos/h (22.16). Así, nuestra propuesta mejoró la capacidad diagnóstica obtenida por el FA mediante métodos convencionales de feature engineering. Estos resultados sugieren que la información extraída de la señal de FA mediante técnicas de deep learning puede ser una potente herramienta para ayudar a diagnosticar la AOS pediátrica, especialmente en los niños más severamente afectados.

1. Introducción

La apnea obstructiva del sueño (AOS) es un trastorno crónico caracterizado por presentar recurrentes ceses respiratorios durante el sueño [1]. Dependiendo del tipo de evento, estos ceses pueden ser totales (eventos de apnea) o parciales (eventos de hipopnea) [1]. La presencia de estos eventos conducen a anomalías en el intercambio de gases, hipoxia, micro despertares y, por tanto, a un sueño fragmentado y no reparador [2]. Consecuentemente, los niños afectados por AOS pueden sufrir graves consecuencias para su salud y desarrollo, como alteraciones neuroconductuales, cardiovasculares, metabólicas y endocrinas [2], [3].

Esta enfermedad afecta hasta al 5.7% de la población infantil [3]. Sin embargo, a pesar de su alta prevalencia y de las graves consecuencias que conlleva, se estima que el 90% de los niños con AOS aún no cuentan con un diagnóstico objetivo [4]. Dicho diagnóstico se lleva a cabo mediante la polisomnografía nocturna (PSG), una prueba que consiste en registrar simultáneamente numerosas señales neurofisiológicas y cardiorrespiratorias mientras el niño duerme [5]. Posteriormente, los especialistas médicos inspeccionan visualmente todos registros y derivan el índice de apnea-hipopnea (IAH), el parámetro clínico usado para determinar la presencia y severidad de AOS [2]. Aunque la PSG es la técnica de referencia para diagnosticar la enfermedad, esta técnica es compleja, costosa, incómoda para los niños y de disponibilidad muy limitada [6]. Por ello, es necesario desarrollar pruebas diagnósticas más sencillas.

Una alternativa es el análisis automático de un conjunto reducido de registros cardiorrespiratorios, como la señal de flujo aéreo (FA). Esta señal es capaz de reflejar la actividad respiratoria del niño mientras duerme, incluidos los ceses respiratorios asociados con la AOS [1]. Además, puede ser fácilmente adquirida en la casa del paciente mediante un dispositivo de monitorización portátil con termistor incorporado. Así, diversos estudios ya han mostrado la utilidad del análisis automático del FA para diagnosticar la AOS pediátrica [7]–[11]. Sin embargo, estos estudios están basados en técnicas de *feature engineering*, por lo que presentan algunas limitaciones [12]. Estos enfoques requieren una caracterización exhaustiva de la señal, por lo que debe conocerse *a priori* la idoneidad de las características que van a ser extraídas. Además, proporcionan un nivel de abstracción bajo, lo que podría limitar su capacidad para detectar patrones complejos. Consecuentemente, podría haber información relevante del FA que pasase desapercibida en estos análisis. Por el contrario, las técnicas de *deep learning* permiten identificar automáticamente las características útiles y encontrar los patrones existentes a partir de la señal en crudo [12]. En este sentido, las redes neuronales convolucionales (convolutional neural networks, CNN)

son la técnica de *deep learning* más ampliamente utilizada dada su idoneidad para procesar matrices multidimensionales [13]. En el contexto de la AOS infantil, sólo dos estudios recientes de nuestro grupo de investigación han propuesto modelos CNN para estimar el IAH a partir de la señal de oximetría sola y combinada con FA [14], [15]. Sin embargo, y hasta donde sabemos, este es el primer estudio que evalúa individualmente el FA pediátrico mediante técnicas de *deep learning*.

De acuerdo con las consideraciones anteriores, el presente estudio parte de la hipótesis de que una arquitectura CNN podría identificar automáticamente la información relevante que está asociada con la presencia de eventos apneicos en FA. Por tanto, nuestro objetivo es evaluar la utilidad de un modelo CNN para estimar el IAH y determinar la severidad de AOS pediátrica a partir del FA.

2. Sujetos y señales

Nuestro estudio ha empleado tres bases de datos de FA pediátrico: (i) the *Childhood Adenotonsillectomy Trial* (CHAT), una base de datos pública compuesta por 1639 registros y obtenida de la web <https://sleepdata.org>, (ii) una base de datos privada procedente del *Comer Children's Hospital* de la Universidad de Chicago (UofC) que incluía 974 registros, y (iii) una base de datos privada procedente del *LeBonheur Children's Hospital* de la Universidad de Tennessee (UofT), formada por 545 registros pediátricos. El estudio fue aprobado por el Comité Ético de los centros participantes y los tutores legales de los niños dieron su consentimiento informado.

Todos los niños involucrados fueron diagnosticados por especialistas médicos siguiendo las reglas de la *American Academy of Sleep Medicine* [1]. Para determinar la severidad de cada sujeto se usaron los umbrales de IAH 1, 5 y 10 eventos/hora (e/h) [2]. La Tabla 1 muestra los datos demográficos y clínicos de los sujetos bajo estudio. Cabe señalar que sólo los registros de la base de datos CHAT incluían marcajes de eventos apneicos. Consecuentemente, ésta fue usada para entrenar el modelo de *deep learning*. Respecto a UofC y UofT, éstas fueron distribuidas aleatoriamente en dos grupos: validación (50%) y test (50%), para optimizar los hiperparámetros y evaluar la capacidad diagnóstica del modelo propuesto, respectivamente. Se encontraron diferencias estadísticamente significativas (p -valor < 0.01 al aplicar el test de Mann-Whitney U y la corrección de Bonferroni) en edad, sexo, índice de masa corporal (IMC) e IAH entre entrenamiento-validación y entrenamiento-test, pero no entre validación-test.

Las señales de FA fueron obtenidas durante la PSG. Estas se registraron utilizando un termistor oronasal a frecuencias de muestreo que oscilaban entre 20 y 512 Hz.

3. Métodos

3.1. Preprocesado y segmentación

Para que todas las señales tuvieran la misma frecuencia de muestro, estas fueron remuestreadas a 10 Hz, frecuencia que ha demostrado ser adecuada para caracterizar la señal de FA y reducir el coste computacional [14]. Después, se

	Entrenamiento (CHAT)	Validación (UofC y UofT)	Test (UofC y UofT)
Sujetos (n)	1639 (100.0%)	759 (24.0%)	760 (24.1%)
Edad (años)	7.0 [2.0]	6.0 [6.9]	6.00 [6.2]
Varones (n)	777 (47.4%)	437 (57.6%)	455 (59.9%)
IMC (kg/m²)	17.3 [6.2]	18.4 [8.0]	18.4 [7.9]
IAH (e/h)	2.5 [4.8]	3.1 [6.7]	3.3 [7.3]
No AOS (n)	355 (21.7%)	173 (22.8%)	174 (22.9%)
AOS Leve (n)	814 (49.6%)	303 (39.9%)	302 (39.8%)
AOS Moderado (n)	252 (15.4%)	127 (16.7%)	128 (16.8%)
AOS Severo (n)	218 (13.3%)	156 (20.6%)	156 (20.5%)

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

Tabla 1. Datos demográficos y clínicos de los sujetos.

sometieron a un proceso de estandarización para normalizar los niveles de referencia de FA entre los diferentes sujetos pediátricos. Finalmente, los registros fueron divididos en segmentos de 5 minutos como en estudios previos [14]. En la etapa de entrenamiento y validación, esta segmentación se realizó con un solapamiento del 50% para incrementar el número de segmentos disponibles.

3.2. Arquitectura CNN

Como puede observarse en la Figura 1, la arquitectura CNN que proponemos está compuesta de tres módulos. El primero de ellos corresponde al módulo de entrada, que recibe los segmentos de 5 minutos de FA.

El siguiente es el módulo CNN, el cual está constituido por λ bloques convolucionales. Cada uno de estos bloques está formado por cinco capas consecutivas: convolucional, *batch normalization*, activación, *pooling* y *dropout* [12], [13]. La capa convolucional es la encargada de extraer los mapas de características mediante filtros de convolución [12], [13]. En nuestro estudio, cada capa convolucional tenía 64 filtros de tamaño = 5×1 , paso = 1 y *padding* = 0 [15]. Después, se aplica un *batch normalization* para normalizar los mapas de características [12]. En la siguiente capa, se usa una función no lineal para determinar cuáles de esos mapas de características se activan [12], [13]. Así, la función de activación empleada en este estudio fue la *Rectified Linear Unit* (ReLU) [14], [15]. Respecto al *pooling*, éste permite reducir la dimensionalidad fusionando patrones similares [12]. En concreto, nosotros utilizamos un *max pooling* con factor = 2, el más utilizado en las CNN [14], [15]. Cada bloque convolucional finaliza con una capa de *dropout*. Ésta elimina aleatoriamente activaciones con probabilidad p_{drop} para reducir el sobreajuste durante el entrenamiento [12], [13].

Por último, el tercer módulo es el de salida. Este se compone de una capa de *flattening*, cuyo objetivo es transformar los mapas de característica en vectores unidimensionales. También incluye una capa de activación lineal, la cual es utilizada para obtener el IAH estimado de cada segmento de FA de entrada.

3.3. Entrenamiento y optimización de la CNN

El proceso de entrenamiento de la CNN comienza con la inicialización de los pesos y sesgos de la red. Para ello, utilizamos el inicializador He-normal [15]. Después, los

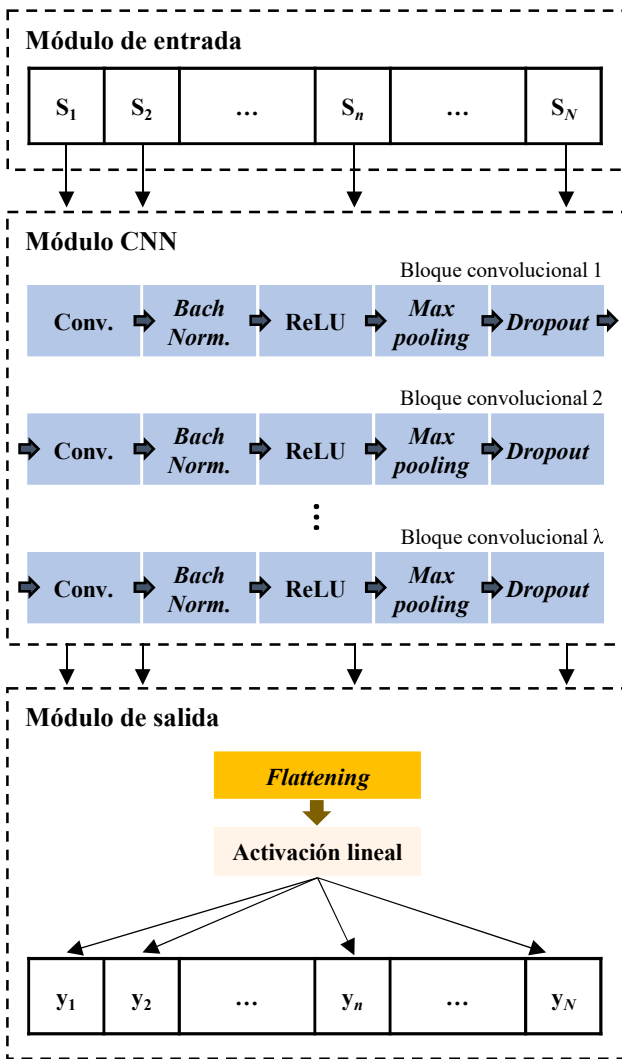


Figura 1. Arquitectura CNN propuesta

pesos y sesgos son actualizados en cada iteración mediante la estimación del momento adaptativo (algoritmo de Adam) [12]. En este estudio, el algoritmo de Adam utilizó una tasa de aprendizaje inicial de 0.001 y la función de error *Huber loss* [15]. Esta función contiene el hiperparámetro delta (δ), el cual debe ser ajustado para controlar el efecto de los valores atípicos e incrementar la robustez [16]. Así, la función *Huber loss* fue calculada en el conjunto de validación para monitorear la convergencia de la CNN durante su entrenamiento. En este sentido, la tasa de aprendizaje se redujo en un factor de 2 durante el entrenamiento cuando la pérdida en el conjunto de validación no mejoró en las últimas 10 épocas [14], [15]. Además, se aplicó un criterio de parada temprana de 30 épocas sin mejoras para evitar el sobreajuste.

3.4. Estimación del IAH

Para obtener el IAH de cada sujeto pediátrico, se han utilizado los valores reportados por el módulo de salida de la CNN. En primer lugar, se calcula el promedio de los valores de IAH estimados para los N segmentos de 5 minutos contenidos en la señal de FA correspondiente [15]:

$$\bar{y} = \frac{\sum_{n=1}^N y_n}{N}, \quad (1)$$

donde y es la salida de la red obtenida para el segmento n .

Después, el IAH de cada sujeto es calculado mediante un modelo regresión lineal para corregir la posible subestimación causada por la CNN [15]:

$$\text{IAH} = (\beta \cdot \bar{y}) + \varepsilon, \quad (2)$$

siendo β la pendiente y ε el sesgo del modelo. Dichos coeficientes fueron ajustados en el conjunto de validación.

3.5. Análisis estadístico

El rendimiento diagnóstico del modelo CNN se evaluó globalmente mediante la matriz de confusión (4 clases), la precisión global y el Cohen's kappa (κ). También se evaluó por grados de severidad mediante las siguientes métricas: sensibilidad (S), especificidad (E), precisión (P), valor predictivo positivo y negativo (VPP y VPN) y razón de verosimilitud positiva y negativa (LR+ y LR-).

4. Resultados

4.1. Optimización de la CNN

Los hiperparámetros a optimizar fueron el número de bloques convolucionales λ , la probabilidad p_{drop} y el δ de la función *Huber loss*. El valor de estos hiperparámetros fue variando para configurar distintos modelos CNN que fueron entrenados en el grupo de entrenamiento. Para cada uno de estos modelos, el índice kappa fue obtenido en el grupo de validación. Así, aquella configuración que obtuvo el mayor kappa fue establecida como óptima. En este estudio, los valores óptimos fueron $\lambda = 7$, $p_{drop} = 0.075$ y $\delta = 5.5$, con un kappa de 0.29 en el grupo de validación.

4.2. Rendimiento diagnóstico de la CNN

La Tabla 2 muestra la matriz de confusión obtenida por el modelo CNN en el grupo de test. Éste alcanzó una precisión global del 53.42% (406/760 sujetos correctamente clasificados) y un kappa de 0.31 en test con cuatro clases. A partir de la matriz de confusión, se obtuvo el rendimiento diagnóstico alcanzado por el modelo CNN para 1, 5 y 10 e/h en el grupo de test. Los resultados obtenidos son mostrados en la Tabla 3. Como puede observarse, el modelo CNN alcanzó una alta precisión diagnóstica en 5 y 10 e/h (81.45% y 90.66%, respectivamente), así como un alto LR+ en 10 e/h (22.16).

		CNN			
		No AOS	Leve	Moderado	Severo
PSC	No AOS	11	155	8	0
	Leve	13	246	38	5
	Moderado	3	66	46	13
	Severo	2	19	32	103

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	96.93	6.32	76.18	77.70	37.93	1.03	0.49
5 e/h	68.31	89.29	81.45	79.18	82.52	6.38	0.35
10 e/h	66.03	97.02	90.66	85.12	91.71	22.16	0.35

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

5. Discusión y conclusiones

En este trabajo se ha evaluado la utilidad de analizar la señal de FA mediante un modelo de *deep learning* basado en CNN para simplificar el diagnóstico de la AOS infantil. Así, el modelo propuesto logró estimar el IAH de los sujetos pediátricos con una alta precisión diagnóstica (76.18%, 81.45% y 90.66% para 1, 5 y 10 e/h, respectivamente), especialmente en los casos con AOS moderada y severa. Además, destaca el alto LR+ alcanzado por la CNN en 10 e/h (22.16). Ya que un valor superior a 10 en esta métrica es un robusto indicador para establecer la presencia de una enfermedad [17], la metodología propuesta podría ser una útil herramienta para determinar automáticamente la presencia de AOS severa en niños.

En los últimos años, diversos estudios de nuestro grupo de investigación se han centrado en el análisis automático del FA como ayuda al diagnóstico de la AOS pediátrica [7]–[11], [14]. La mayoría de estos estudios se basaron en la caracterización de la señal de FA mediante técnicas convencionales de *feature engineering*, con las cuales alcanzaron precisiones que oscilan entre el 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 [8]–[11]. En el presente trabajo, alcanzamos precisiones dentro de este rango para 1e/h, así como notablemente superiores para 5 y 10 e/h. Otro estudio analizó conjuntamente el FA y la señal de oximetría mediante CNN [14]. Éste alcanzó altas precisiones en CHAT (84.64%, 93.46% y 94.44% para 1, 5 y 10 e/h, respectivamente), que se redujeron notablemente cuando la propuesta se evaluó en UofC (84.10%, 84.10% y 90.26% para 1, 5 y 10 e/h, respectivamente). Analizando un único canal, nuestro modelo obtuvo una precisión ligeramente inferior en 1 y 5 e/h, y superior en 10 e/h. Además, también lo superó en S en 1 e/h y en E, VPP y LR+ en 5 y 10 e/h.

Este estudio presenta algunas limitaciones que deben ser consideradas. En primer lugar, el modelo CNN fue entrenado únicamente con la base de datos CHAT, ya que los registros de UofC y UofT no contenían anotaciones sobre el comienzo y finalización de los eventos apnéicos. Además, se observaron diferencias entre estas bases de datos, particularmente entre CHAT y UofC y entre CHAT y UofT, lo que podría haber afectado a la capacidad de generalización de nuestro enfoque. En este sentido, el modelo propuesto podría ser validado en otras bases de datos, e incluso con registros domiciliarios de FA, lo que aumentaría la solidez de nuestros resultados. También se podría estimar el IAH mediante la aplicación y/o combinación de otros métodos de *deep learning*, como las redes *long short-term memory* (LSTM), lo que constituye una interesante línea de investigación futura. Finalmente, sería interesante utilizar técnicas de *explainable artificial intelligence* (XAI). Esto nos permitirían identificar y comprender qué características del FA extrae el modelo CNN para derivar su predicción.

En conclusión, el modelo CNN mostró un alto rendimiento diagnóstico estimando el IAH pediátrico. La mejora obtenida respecto a las técnicas convencionales pone de manifiesto la utilidad de los métodos de *deep learning*, y en particular de las arquitecturas CNN, para incrementar la

capacidad diagnóstica del FA. Estos resultados sugieren que la información extraída de la señal de FA mediante técnicas de *deep learning* puede ser una potente herramienta para ayudar a diagnosticar la AOS infantil, especialmente en los niños más severamente afectados.

Agradecimientos

Este estudio ha sido financiado por el Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación y ERDF (PID2020-115468RB-I00 y PDC2021-120775-I00), por el CIBER-BBN (CB19/01/00012), por la Sociedad Española de Neumología y Cirugía Torácica (SEPAR 649/2018) y por la Sociedad Española de Sueño (Beca de Investigación SES 2019). D. Álvarez es beneficiario de una ayuda Ramón y Cajal (RYC2019-028566-I) del Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación, cofinanciada por el Fondo Social Europeo.

Referencias

- [1] 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.
- [2] M. Luz Alonso-Álvarez *et al.*, “Consensus document on sleep apnea-hypopnea syndrome in children,” *Arch. Bronconeumol.*, vol. 47, no. SUPPL. 5, pp. 2–18, 2011.
- [3] C. L. Marcus *et al.*, “Diagnosis and management of childhood obstructive sleep apnea syndrome,” *Pediatrics*, vol. 130, no. 3, 2012.
- [4] L. Kheirandish-Gozal, “What Is ‘Abnormal’ in Pediatric Sleep?,” *Respir. Care*, vol. 55, no. 10, pp. 1366–1376, 2010.
- [5] C. Jon, “Polysomnography in Children,” in *Pediatric Otolaryngology for the Clinician*, Totowa, NJ: Humana Press, 2009, pp. 35–47.
- [6] 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. 972–975, 1995.
- [7] 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. 401–407, 2015.
- [8] V. Barroso-García *et al.*, “Usefulness of recurrence plots from airflow recordings to aid in paediatric sleep apnoea diagnosis,” *Comput. Methods Programs Biomed.*, vol. 183, pp. 105083, 2020.
- [9] V. Barroso-García *et al.*, “Bispectral analysis of overnight airflow to improve the pediatric sleep apnea diagnosis,” *Comput. Biol. Med.*, vol. 129, pp. 104167, 2021.
- [10] 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.
- [11] 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.
- [12] I. Goodfellow *et al.*, *Deep learning*. MIT press, 2016.
- [13] Y. Lecun *et al.*, “Deep learning,” *Nat.*, vol. 521, no. 7553, pp. 436–444, 2015.
- [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. Informatics*, vol. 25, no. 8, pp. 2906–2916, 2021.
- [16] P. J. Huber, “Robust Estimation of a Location Parameter,” *Ann. Math. Stat.*, vol. 35, no. 1, pp. 73–101, 1964.
- [17] J. J. Deeks and D. G. Altman, “Diagnostic tests 4: likelihood ratios,” *BMJ*, vol. 329, pp. 168–169, 2004.