Cribado no supervisado de la apnea del sueño usando la PPG de un dispositivo de muñeca: un estudio piloto

Diego Cajal[1,2], Rodrigo Lozano-Puñet[1,2], Ana Victoria Gil[3,4], José Pablo Cubero[3,4], Jose María Marín[3,4], Eduardo Gil[1,2], Jesús Lázaro[1,2], Raquel Bailón[1,2]

[1] Instituto de Investigación en Ingeniería de Aragón (I3A), IIS Aragón, Universidad de Zaragoza

[2] Centro de Investigación Biomédica En Red – Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)

[3] Hospital Universitario Miguel Servet, Zaragoza

[4] Instituto de Investigación Sanitaria de Aragón

Resumen

En un trabajo previo se presentó un método para detectar apnea del sueño usando señales habituales en dispositivos vestibles (wearables): la saturación periférica de oxígeno (SpO_2) y la fotopletismografía de pulso (PPG). Sin embargo, dichas señales se obtuvieron con un pulsioxímetro convencional de dedo. Este artículo describe un estudio piloto que aplica la misma metodología a señales adquiridas por un wearable de muñeca en un entorno domiciliario no supervisado ($n = 20$). Se aplicó un clasificador para diferenciar segmentos de respiración normal y anómala. Posteriormente, se calculó el Índice de Variación Cíclica de la Frecuencia Cardiaca (CVHRI) dentro de los segmentos con respiración anómala. El clasificador alcanzó una exactitud del 62.0 % con los datos del wearable y el CVHRI tuvo una correlación media-alta con el AHI ($r = 0,66$, $p < 0,001$), lo que sugiere que su potencial para la estratificación de pacientes sigue siendo viable.

1. Introducción

El síndrome de apnea-hipopnea del sueño (SAHS) se caracteriza por episodios recurrentes de interrupción de la respiración o de reducción significativa del flujo aéreo, con posibles consecuencias graves para la salud. Aunque el tratamiento adecuado mitiga estos efectos y su impacto socioeconómico, el SAHS sigue infradiagnosticado, en parte por las limitaciones de la polisomnografía, el estándar de referencia [1]. La falta de diagnóstico se asocia a mayor riesgo de hipertensión, enfermedad cardiovascular, ictus, somnolencia diurna, accidentes de tráfico y menor calidad de vida [2].

Los dispositivos wearables ofrecen una vía prometedora para el cribado accesible en el hogar. En un trabajo previo usando sensores no vestibles, mostramos un método basado en SpO_2 y en intervalos pulso-a-pulso (PPI) derivados de PPG para clasificar segmentos en *respiración normal* y *anómala* [3]. El PPI de los segmentos anómalos se usó para calcular el CVHRI, índice introducido en ese estudio, que mide la frecuencia del patrón bradicardia-taquicardia. Este patrón es típico cuando ocurren eventos apneicos de forma repetida. El índice propuesto tuvo una correlación fuerte ($r = 0{,}94$, $p < 0{,}001$) con el Índice de Apnea-Hipopnea (AHI).

En el presente trabajo, evaluamos la traslación de aquel método a un escenario real con datos de un wearable de muñeca operado por pacientes y controles sin supervisión en sus domicilios. Para mantener la consistencia de las covariables, se ha preservado la metodología original, incluyendo el etiquetado de eventos y conversión a etiquetas por segmentos, sustituyendo únicamente el dispositivo con el que se obtienen las señales. Como clasificador, se ha utilizado el modelo entrenado durante el trabajo anterior, que toma como entradas los parámetros de Hjorth de las señales PPI y SpO_2 [4]. El modelo no se reentrenó con las nuevas señales. De la misma forma, el CVHRI se ha calculado a partir de la PPI del wearable en los segmentos con *respiración anómala*.

2. Métodos

2.1. Experimento

Se analizaron 20 participantes (edad 47 ± 15 años; 12 hombres) de un estudio mayor en curso, aprobado por el comité de ética de Aragón (CEICA, PI23/336): 10 sospechosos de SAHS y 10 controles. Las grabaciones se realizaron sin supervisión, durante una noche, en sus respectivos domicilios, activando los equipos al acostarse y desactivándolos al despertar.

Se emplearon:

- Polígrafo ApneaLink Air (ResMed Inc., USA). Este dispositivo obtuvo la SpO_2 muestreada a 1 Hz (denotada $x_{\mathrm{SpO_2}}[n]$ en este artículo) y la presión nasal a 100 Hz, ambas usadas como referencia.
- Maxim MAXREFDES103 (Analog Devices Inc., USA). Se trata de un wearable de muñeca capaz de grabar PPG con canales rojo, verde e infrarrojo, muestreados a 256 Hz (denotadas $x^{r}_{\mathrm{PPG}}[n]$, $x^{g}_{\mathrm{PPG}}[n]$ y $x^{ir}_{\mathrm{PPG}}[n]$, respectivamente).

Las señales de referencia del ApneaLink se usaron exclusivamente para etiquetar eventos apneicos e hipopneicos conforme a las reglas de la American Academy of Sleep Medicine (AASM) [5] mediante un algoritmo automático descrito en [3]. Esas anotaciones de eventos se agruparon por segmentos

ISBN: 978-84-09-80259-3

para alimentar el clasificador, y se compararon con la salida del clasificador cuando las señales de entrada provenían del wearable. Además, la referencia sirvió para calcular el AHI, que se comparó con el CVHRI. El cálculo del AHI fue realizado por profesionales médicos cualificados.

2.2. Estimación de SpO_2

A partir de la PPG del wearable se estimaron PPI y SpO_2, denotados $d_{\text{PPI}}[n]$ y $d_{\text{SpO}_2}[n]$. La $d_{\text{PPI}}[n]$ se extrajo del canal verde de la PPG, $x^g_{\text{PPG}}[n]$, identificando los puntos de máxima pendiente ascendente, siguiendo el algoritmo descrito en [6].

Para la estimación de $d_{\text{SpO}_2}[n]$, se calculó el cociente $d_{\text{R}}[n]$ a partir de las componentes de continua (DC) y alterna (AC) de $x^r_{\text{PPG}}[n]$ (d^{r}_{AC} y d^{r}_{DC}) y $x^{ir}_{\text{PPG}}[n]$ ($d^{\text{ir}}_{\text{AC}}$ y $d^{\text{ir}}_{\text{DC}}$):

$$d_{\text{R}}[n] = \frac{\frac{d^{\text{r}}_{\text{AC}}[n]}{d^{\text{r}}_{\text{DC}}[n]}}{\frac{d^{\text{ir}}_{\text{AC}}[n]}{d^{\text{ir}}_{\text{DC}}[n]} + \epsilon}. \quad (1)$$

Se añade $\epsilon = 0{,}01$ al denominador porque la amplitud puede tomar valores cercanos a cero temporalmente por movimientos que llevan a posicionamientos subóptimos del sensor.

Las componentes AC se obtienen como la diferencia entre las envolventes superior e inferior de cada canal $x^r_{\text{PPG}}[n]$ y $x^i_{\text{PPG}}r[n]$ (ver Figura 1). La envolvente superior fue calculada por interpolación de los puntos de máxima amplitud de cada pulso, mientras que la envolvente inferior fue calculada por interpolación de los inicios de pulso, definidos en [6], ambas a una frecuencia de muestreo de 256 Hz. La DC es la envolvente inferior.

Para depurar errores se aplicó un umbral al cuadrado de la derivada de $d_{\text{R}}[n]$. Este umbral se ajustó de forma empírica a 10^{-8}, eliminando valores atípicos causados principalmente por errores en la estimación de las envolventes. A continuación, se aplicó un segundo umbral para $d_{\text{R}}[n] > 4$. Después, $d_{\text{R}}[n]$ se filtró paso-bajo a 0,1 Hz para eliminar variaciones rápidas debidas al ruido.

Según la documentación del wearable, la saturación se debería calcular como:

$$d_{\text{SpO}_2}[n] = -16{,}6(d_{\text{R}}[n])^2 + 8{,}3(d_{\text{R}}[n]) + 100. \quad (2)$$

No obstante, esta fórmula no dio buenos resultados con nuestros datos. En su lugar se realizó una calibración específica por sujeto, transformando linealmente $d_{\text{R}}[n]$ para igualar media y desviación típica de $x_{\text{SpO}_2}[n]$. Finalmente, se remuestreó a 25 Hz según recomendaciones de la AASM [5] y se redondeó al entero más cercano (ver Figura 2).

2.3. Clasificación de segmentos y CVHRI

Con $d_{\text{PPI}}[n]$ y $d_{\text{SpO}_2}[n]$ del wearable se identificaron segmentos de *respiración normal* y *respiración anómala*. Las señales se dividieron en segmentos de 180 s con solape de 150 s. De cada segmento se extrajeron los parámetros de Hjorth (Actividad, Movilidad y Complejidad) tanto de $d_{\text{PPI}}[n]$ como de $d_{\text{SpO}_2}[n]$ (6 rasgos por segmento), alimentando un clasificador previamente entrenado con señales no wearables [3]. Dado que aquí $d_{\text{SpO}_2}[n]$ es estimada a partir de la PPG del wearable, el clasificador se evaluó dos veces: (i) usando la estimación $d_{\text{SpO}_2}[n]$ y (ii) usando $x_{\text{SpO}_2}[n]$ del dispositivo de referencia. Esta doble estimación nos permitió cuantificar el impacto de la estimación de $d_{\text{SpO}_2}[n]$.

Luego se calculó el CVHRI a partir de $d_{\text{PPI}}[n]$ solo en los segmentos clasificados como *respiración anómala*, donde se espera el patrón taquicardia-bradicardia. El CVHRI se define como la suma de las frecuencias de los picos espectrales de cada segmento, dividida por el número total de segmentos analizados. Esto resulta en un parámetro por sesión para cada paciente, análogo al AHI.

3. Resultados

El etiquetado con señales de referencia produjo 6578 segmentos de *respiración normal* (88.4 %) y 859 de *respiración anómala* (11.6 %). El AHI medio fue de 10.4 eventos/hora, con una desviación estándar de 13.2 eventos/hora. La distribución de los sujetos fue la siguiente: 8 sujetos con AHI < 5 (40 %), 7 sujetos con $5 \leq$ AHI < 15 (35 %) y 5 sujetos con AHI ≥ 15 (25 %).

Los resultados de clasificación se resumen en la Tabla 1. Las columnas Acc, P_n, R_n, $F1_n$, P_{ab}, R_{ab} y $F1_{ab}$ son, respectivamente: exactitud balanceada, precisión de la clase *respiración normal*, recall de la clase *respiración normal*, F1-score de la clase *respiración normal*, precisión de la clase *respiración anómala*, recall de la clase *respiración anómala* y F1-score de la clase *respiración anómala*. Para evaluar el impacto de la estimación de $d_{\text{SpO}_2}[n]$, se obtuvieron resultados de la clasificación tanto con la señal estimada del wearable, $d_{\text{SpO}_2}[n]$, como con la señal del dispositivo de referencia $x_{\text{SpO}_2}[n]$.

El CVHRI correlacionó con el AHI con un coeficiente de correlación de Pearson de $r = 0{,}66$ ($p < 0{,}001$). La Figura 3 presenta un diagrama de dispersión del AHI contra el CVHRI. También se recalculó el CVHRI a partir de la clasificación de segmentos con la señal de referencia, obteniendo $r = 0{,}77$ ($p < 0{,}001$) (ver Figura 4).

Tabla 1: Resultados de clasificación de segmentos (%).

SpO_2	Acc	P_n	R_n	$F1_n$	P_{ab}	R_{ab}	$F1_{ab}$
$d_{\text{SpO}_2}[n]$	62.0	91.3	84.1	0.88	25.1	39.9	0.31
$x_{\text{SpO}_2}[n]$	73.9	94.4	88.0	0.91	39.4	59.8	0.48

4. Discusión

Los resultados con señales del wearable muestran degradación frente a los obtenidos con el sistema de polisomnografía en [3], algo esperable por: (i) una PPG más ruidosa y con artefactos; (ii) la estimación de $d_{\text{SpO}_2}[n]$ a partir de dicha señal; y (iii) el uso de un modelo preentrenado en otro tipo de señales. La menor robustez de PPG se asocia al registro do-

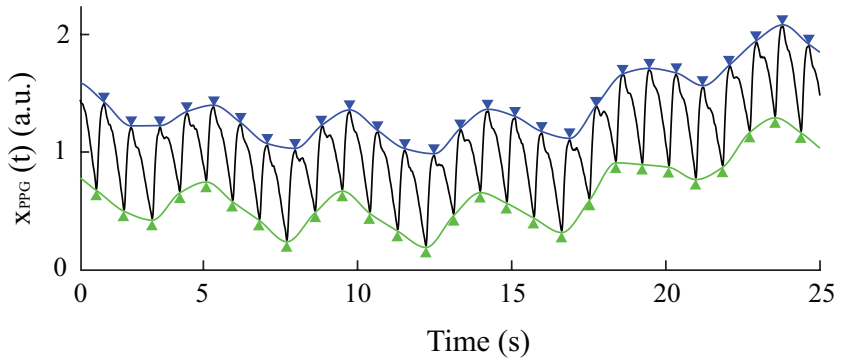


Figura 1: Envolventes superior (azul) e inferior (verde) de la PPG (negro). Los triángulos azules representan amplitudes máximas de pulso, mientras que los triángulos verdes representan puntos basales de pulso.

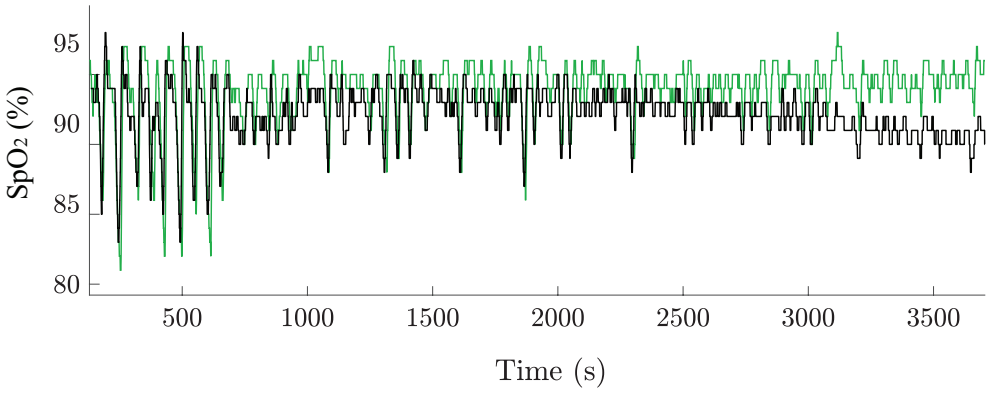


Figura 2: Estimación de SpO_2 desde el dispositivo Maxim. Referencia en verde, estimación en negro.

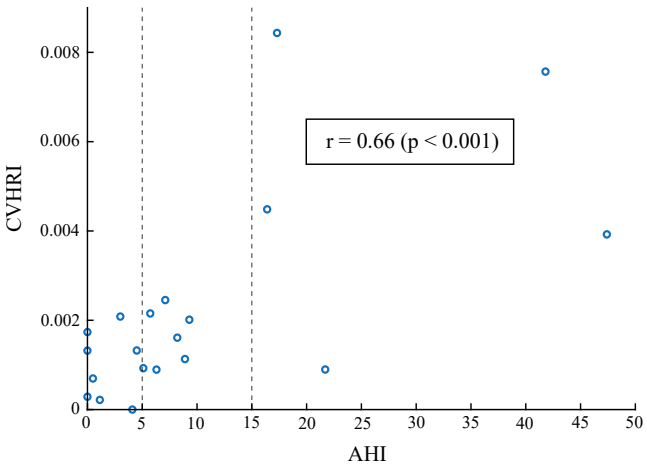


Figura 3: AHI vs CVHRI con $d_{SpO_2}[n]$ estimada del wearable. Líneas verticales separan AHI $<$ 5, 5$\leq$AHI $<$ 15 y AHI $\geq$ 15.

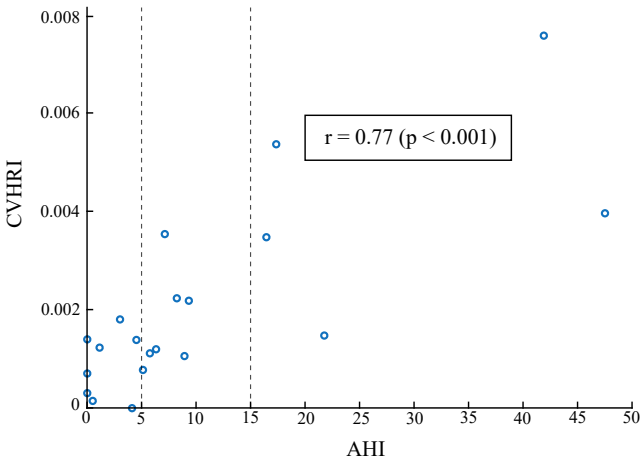


Figura 4: AHI vs CVHRI usando $x_{SpO_2}[n]$ de referencia. Líneas verticales separan AHI $<$ 5, 5$\leq$AHI $<$ 15 y AHI $\geq$ 15.

miciliario no supervisado. Con más datos, el modelo podría reentrenarse para estas señales, dejando margen de mejora.

La estimación de $d_{SpO_2}[n]$ no se hizo con la fórmula del fabricante, sino con calibración por sujeto, con buenos resultados cualitativos, si bien debe evaluarse y optimizarse cuantitativamente. Aunque no ideal, la calibración previa por usuario es común en otras medidas (p. ej., tensión arterial).

Sin embargo, la aproximación llevada a cabo en este estudio puede mejorarse, potencialmente habilitando una mejora en los resultados de clasificación, como evidencian los resultados usando la señal de referencia $x_{SpO_2}[n]$. Alternativamente, podría reentrenarse el clasificador usando los parámetros de Hjorth de $d_R[n]$ en lugar de $d_{SpO_2}[n]$: al ser operaciones de normalización y escalado, las métricas (varianza, frecuencia

ISBN: 978-84-09-80259-3

central y ancho de banda) deberían conservar proporcionalidad.

El modelo de clasificación debería ser reentrenado con las nuevas señales cuando haya más datos disponibles. Las diferencias entre las señales $d_{\mathrm{SpO}_2}[n]$ y $x_{\mathrm{SpO}_2}[n]$ sugieren que una mejora en la clasificación es esperable.

También debería aumentarse el tamaño muestral, especialmente el subgrupo con AHI ≥ 15, ya que aquí solo hubo 5 casos, lo que limita el análisis y nos impide hacer un estudio de estratificación entre sujetos con o sin SAHS, como sí se hizo en [3]. Además, el desbalance de clases es evidente, como cabía esperar (11.6 % de la clase *respiración anómala*). Sumado al pequeño tamaño muestral, deja únicamente 859 segmentos de la clase objetivo. Las métricas de evaluación del clasificador elegidas, como la exactitud balanceada, tratan de tener esto en cuenta, pero la cantidad de datos limita la potencia del estudio.

Aun así, los resultados son prometedores. La caída de la correlación AHI-CVHRI (de $r = 0{,}94$ en [3] a $r = 0{,}66$) se debe sobre todo a la bajada de exactitud del clasificador (de 86.3 % en [3] a 62.0 %). Aquí debemos hacer la advertencia de que en el anterior estudio no se reportó la exactitud balanceada, sino que se hizo un submuestreo de la clase mayoritaria en el test, por lo que cabe esperar que las diferencias en realidad sean menores. La razón del cambio de métrica utilizada es que la exactitud balanceada es un mejor indicador de cómo se comporta el clasificador en un caso real. Pese a ello, la correlación AHI–CVHRI sigue siendo media-alta, lo que sugiere que el CVHRI es robusto frente a errores del clasificador. Además, al usar $x_{\mathrm{SpO}_2}[n]$ de referencia, la exactitud asciende a 73.9 % y la correlación a $r = 0{,}77$. Aunque estos valores no son tan altos como los de [3], supera la exactitud obtenida únicamente con señales de wearables. Esto es coherente con [3], donde los parámetros de Hjorth de $x_{\mathrm{SpO}_2}[n]$ fueron los predictores principales.

5. Conclusiones

El método de [3] puede aplicarse a wearables usados en casa y sin supervisión, con buen rendimiento: el clasificador logra 62.0 % de exactitud balanceada y la correlación AHI–CVHRI es de $r = 0{,}66$ ($p < 0{,}001$), lo que sugiere que la estratificación entre sujetos con y sin SAHS sigue siendo factible. sin embargo, es necesario ampliar la muestra, sobre todo para el grupo con AHI ≥ 15, para realizar un estudio más exhaustivo. Además, se deberán poner esfuerzos en mejorar la estimación de $d_{\mathrm{SpO}_2}[n]$ a partir de $d_R[n]$ o usar directamente $d_R[n]$ como entrada del clasificador. Junto con el reentrenamiento del modelo para las nuevas señales, se esperan mejoras adicionales.

Agradecimientos

Trabajo financiado en parte por el proyecto PID2021-126734OB-C21, financiado por MCIN/ AEI /10.13039/501100011033/ y “ERDF A way of making Europe”; en parte por el proyecto TED2021-131106B-I00, cofinanciado por CIBER (Instituto de Salud Carlos III), MCIN/ AEI /10.13039/501100011033/ y “European Union NextGenerationEU/PRTR”; y por el Gobierno de Aragón (Grupo de Referencia BSICoS T39-23R). Cálculos realizados en ICTS NANBIOSIS (Unidad HPC, Universidad de Zaragoza).

Referencias

[1] Y. Yeghiazarians, H. Jneid, J. R. Tietjens, S. Redline, D. L. Brown, N. El-Sherif, R. Mehra, B. Bozkurt, C. E. Ndumele, and V. K. Somers, “Obstructive Sleep Apnea and Cardiovascular Disease: A Scientific Statement From the American Heart Association,” *Circulation*, vol. 144, 7 2021.

[2] T. Young, P. E. Peppard, and D. J. Gottlieb, “Epidemiology of Obstructive Sleep Apnea: A population health perspective,” *American Journal of Respiratory and Critical Care Medicine*, vol. 165, pp. 1217–1239, 5 2002.

[3] D. Cajal, E. Gil, P. Laguna, C. Varon, D. Testelmans, B. Buyse, C. Jensen, R. Hoare, R. Bailon, and J. Lazaro, “Obstructive Sleep Apnea Screening by Joint Saturation Signal Analysis and PPG-derived Pulse Rate Oscillations,” *IEEE Journal of Biomedical and Health Informatics*, 2023.

[4] B. Hjorth, “EEG analysis based on time domain properties,” *Electroencephalography and Clinical Neurophysiology*, vol. 29, pp. 306–310, 1 1970.

[5] R. Berry, R. Brooks, C. Gamaldo, Harding SM, Marcus CL, Vaughn BV, and Tangredi MM, “The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2,” *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 2012.

[6] J. Lázaro, E. Gil, J. M. Vergara, and P. Laguna, “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*, vol. 18, no. 1, pp. 240–246, 2014.