

Detección de episodios de apnea del sueño por medio de un modelo de aprendizaje automático con datos de PSG

Liliam Zamanta Santos Estupiñan
Bioingeniería
Universidad de Antioquia
Medellín, Colombia
zamanta.estupinan@udea.edu.co

Laura Esperanza Vargas Patarroyo
Bioingeniería
Universidad de Antioquia
Medellín, Colombia
laurae.vargas@udea.edu.co

Sebastián García Morales
Ingeniería de Telecomunicaciones
Universidad de Antioquia
Medellín, Colombia
sebastian.garciam1@udea.edu.co

Resumen—La apnea del sueño es un trastorno común que se diagnostica mediante estudios de sueño que registran múltiples señales fisiológicas durante la noche. Este trabajo propone un sistema automático para la detección de apnea del sueño utilizando tres modelos de inteligencia artificial: una red neuronal multicapa, una red neuronal convolucional unidimensional y una máquina de soporte vectorial. Los resultados mostraron que la red neuronal multicapa presentó sobreajuste, mientras que la red convolucional logró una mejor capacidad de generalización, aunque con cierta inestabilidad. El modelo basado en vectores de soporte obtuvo el mejor rendimiento, con alta precisión y estabilidad durante la validación. Este estudio demuestra que los métodos de inteligencia artificial pueden ser herramientas útiles para apoyar el análisis automático de señales fisiológicas del sueño, permitiendo una detección más rápida y eficiente de la apnea del sueño y contribuyendo a mejorar el proceso diagnóstico en entornos clínicos.

Keywords—Polisomnografía, red neuronal multicapa, red neuronal convolucional unidimensional, máquina de soporte vectorial, apnea del sueño.

I. PLANTEAMIENTO DEL PROBLEMA

Dormir es una actividad fundamental para la salud humana porque, durante este proceso, el cuerpo lleva a cabo funciones esenciales como la reparación celular, la regulación hormonal y la consolidación de la memoria. Además, reduce el riesgo de desarrollar enfermedades crónicas, como la diabetes [1]. Por lo tanto, la baja calidad del sueño es considerada un problema de salud pública. Según la OMS, el 40% de la población mundial sufre trastornos del sueño, lo que puede desencadenar enfermedades graves y generar altos costos sanitarios [2].

Para un adulto promedio, se recomienda tener entre 4 y 6 ciclos de sueño por noche, lo que equivale a aproximadamente 7-9 horas de sueño. Cada ciclo de sueño, que dura entre 90 y 120 minutos, incluye fases de sueño no REM y REM, y es crucial para la recuperación física y mental. Durante el sueño profundo, el cuerpo realiza funciones de reparación y crecimiento, mientras que en el sueño REM se procesan las emociones y se consolidan los recuerdos. La interrupción de estos ciclos puede afectar negativamente la salud y el bienestar general [3].

Durante el ciclo del sueño, el cerebro pasa por cinco fases que se pueden dividir en dos grandes grupos: las fases 1 a 4 corresponden al sueño no REM (sueño de movimientos oculares no rápidos) y la fase 5 corresponde al sueño REM.

Cada fase tiene características específicas y una duración promedio mostradas en la **Tabla 1** [4].

TABLA I. FASES DEL SUEÑO

Fase del sueño	Duración promedio (min/ciclo)	Características
<i>Fase 1 (Transición vigilia – sueño)</i>	5 – 10	Sueño superficial. Las interrupciones en esta fase pueden causar fatiga crónica y bajo rendimiento cognitivo.
<i>Fase 2 (Sueño ligero)</i>	20 minutos por ciclo, ocupando aproximadamente el 50% del ciclo total de sueño	Los husos del sueño protegen contra estímulos externos, y su falta puede afectar la consolidación de habilidades motoras.
<i>Fase 3 (Sueño profundo inicial)</i>	20 - 40	Proporciona una restauración física moderada. La interrupción de esta fase aumenta el riesgo de inflamación sistémica.
<i>Fase 4 (Sueño profundo delta)</i>	20 – 40	Fundamental para la reparación celular y la liberación de la hormona de crecimiento. La deficiencia en esta fase se ha vinculado al envejecimiento prematuro y a la diabetes.
<i>Fase REM (Sueño paradójico)</i>	10 – 20 aumentando en duración con cada ciclo de sueño	Es indispensable para la memoria emocional y la creatividad. La privación de esta fase puede resultar en ansiedad y dificultades para resolver problemas.

Un estudio PSG, o polisomnografía, es una prueba que mide la actividad cerebral y corporal durante el sueño. Se utiliza para diagnosticar trastornos del sueño, como apnea, sonambulismo, o movimientos de piernas durante la noche [5].

Al analizar datos polisomnográficos (PSG) durante el sueño mediante algoritmos de aprendizaje automático, permiten procesar grandes volúmenes de datos en tiempo real, identificando patrones anómalos no detectables con métodos convencionales como trastornos del sueño [6].

La apnea del sueño es un trastorno en el que la respiración se interrumpe o se vuelve muy superficial durante el descanso. Estas interrupciones pueden durar desde unos segundos hasta

varios minutos y pueden ocurrir hasta 30 veces por hora. El tipo más común es la apnea obstructiva del sueño, la cual se produce cuando los músculos de la garganta se relajan en exceso durante el sueño, provocando que las vías respiratorias se estrechen o se bloqueen. Esto genera un colapso o una obstrucción temporal que impide el paso normal del aire. Factores como el sobrepeso, ciertas características anatómicas como amígdalas grandes o una mandíbula pequeña, el consumo de alcohol o sedantes, la edad avanzada y dormir boca arriba pueden aumentar el riesgo de padecer esta condición. Cuando la respiración se reanuda, suele hacerlo con un ronquido fuerte o un resoplido. Las personas que padecen apnea suelen roncar intensamente y, debido a la fragmentación constante del sueño, experimentan somnolencia diurna. Esto aumenta considerablemente el riesgo de sufrir accidentes de tránsito, accidentes laborales y diversos problemas de salud, como enfermedades cardiovasculares, hipertensión y diabetes tipo 2, debido a la baja calidad del sueño [6].

Entre los algoritmos de aprendizaje automático se han visto enfoques que incluyen Máquinas de Soporte Vectorial (SVM – por sus siglas en inglés) y el uso de Análisis Lineal Discriminado y técnicas seleccionadas. Además, también se han aplicado técnicas de aprendizaje profundo como Long Short-Term Memory (LSTMs) y Redes Neurales Convolucionales (CNN – por sus siglas en inglés) que han mostrado diferentes grados de éxito [7].

Se han empleado, a veces en combinación con otros modelos como las cadenas de Markov, para clasificar eventos de apnea e hipopnea. Un modelo CNN bimodal de fusión de características combinado con un mecanismo de atención de canal logró una precisión del 95.91%. También se utilizó una CNN – 1 D (una dimensión) para clasificar la gravedad de la apnea (normal, leve, moderada, grave). Un enfoque, SomnNET, basado en CNN 1-D para la detección de apnea por segundo utilizando señales de SpO₂, logró una precisión del 97.08% [8]. Se han empleado, a veces en combinación con otros modelos como las cadenas de Markov, para clasificar eventos de apnea e hipopnea. Un modelo CNN bimodal de fusión de características combinado con un mecanismo de atención de canal logró una precisión del 95.91%. También se utilizó una CNN – 1 D (una dimensión) para clasificar la gravedad de la apnea (normal, leve, moderada, grave). Un enfoque, SomnNET, basado en CNN 1-D para la detección de apnea por segundo utilizando señales de SpO₂, logró una precisión del 97.08% [8]. Se han empleado, a veces en combinación con otros modelos como las cadenas de Markov, para clasificar eventos de apnea e hipopnea. Un modelo CNN bimodal de fusión de características combinado con un mecanismo de atención de canal logró una precisión del 95.91%. También se utilizó una CNN – 1 D (una dimensión) para clasificar la gravedad de la apnea (normal, leve, moderada, grave). Un enfoque, SomnNET, basado en CNN 1-D para la detección de apnea por segundo utilizando señales de SpO₂, logró una precisión del 97.08% [8].

Un modelo Perceptrón multicapa (MLP- por sus siglas en inglés) demostró alta precisión en el diagnóstico de OSA (Obstructive sleep apnea) moderada a grave [8]. Un modelo Perceptrón multicapa (MLP- por sus siglas en inglés) demostró alta precisión en el diagnóstico de OSA (Obstructive sleep apnea) moderada a grave [8]. Un modelo Perceptrón multicapa

(MLP- por sus siglas en inglés) demostró alta precisión en el diagnóstico de OSA (Obstructive sleep apnea) moderada a grave [8].

También se han empleado Máquina de Soporte Vectorial (SVM por sus siglas en inglés) en un sistema de detección de apnea-hipopnea basado en SVM que utiliza características de intervalos R-R del ECG y una característica estadística de SpO₂ mostró un buen rendimiento con una precisión del 99.1% [8]. También se han empleado Máquina de Soporte Vectorial (SVM por sus siglas en inglés) en un sistema de detección de apnea-hipopnea basado en SVM que utiliza características de intervalos R-R del ECG y una característica estadística de SpO₂ mostró un buen rendimiento con una precisión del 99.1% [8]. También se han empleado Máquina de Soporte Vectorial (SVM por sus siglas en inglés) en un sistema de detección de apnea-hipopnea basado en SVM que utiliza características de intervalos R-R del ECG y una característica estadística de SpO₂ mostró un buen rendimiento con una precisión del 99.1% [8].

II. OBJETIVOS

A. Objetivo general

Detectar episodios de apnea del sueño en adultos por medio de un modelo de aprendizaje automático con datos de polisomnografía.

B. Objetivos específicos

- Seleccionar características de los datos de ISRUC-Sleep dataset por medio de estadística descriptiva.
- Desarrollar un modelo de aprendizaje automático, a partir de lo encontrado en la literatura, que detecte eventos de apnea del sueño.
- Medir la precisión y la capacidad de generalización del modelo utilizando métricas adecuadas, como la precisión, la sensibilidad y la especificidad.

III. METODOLOGÍA

A. Preprocesamiento de los datos

A partir de la base de datos pública ISRUC-Sleep dataset preprocesada de la Universidad de Coimbra Portugal-Instituto de sistemas y robótica, la cual se puede encontrar como [ISRUC Processed](#) en la plataforma Kaggle. Se tomaron los datos para el entrenamiento y prueba de los diferentes modelos. Este conjunto de datos se creó para apoyar la investigación en la clasificación de eventos de apnea mediante técnicas de aprendizaje profundo. Contiene datos clasificados en dos grupos principales: Eventos y No-Eventos, cada uno almacenado en directorios estructurados. Los archivos de datos representan segmentos de grabaciones polisomnográficas (PSG), listas para ser preprocesadas y analizadas [9].

Incluye datos obtenidos de participantes adultos, tanto sanos como con trastornos del sueño. El conjunto de datos está estructurado en tres grupos:

- Datos de 100 sujetos, cada uno con una sesión de grabación.
- Datos de 8 sujetos, cada uno con dos sesiones de grabación para apoyar estudios longitudinales.

- Datos de 10 sujetos sanos, cada uno con una sesión de grabación, para facilitar los estudios comparativos.

Las grabaciones de PSG fueron puntuadas visualmente por dos expertos en sueño e incluían información electrofisiológica, neurológica y contextual [10].

A partir de lo anterior, se realizó un preprocesamiento adicional debido a que no se utilizaron todos los grupos, solo el primero y el tercero. Además de que la carpeta de Eventos se filtró solo para obtener los datos de apnea central (AC) y de apnea obstructiva (OA), pero estos datos se encontraban desbalanceados por lo que se optó por tomar las zonas de No eventos de los sujetos con patologías y combinarlos con los sanos. Así, se solucionó el desbalanceo de datos, quedando con 456 archivos .csv de No Eventos y 454 archivos de Eventos.

Adicionalmente, se cortaron los canales de interés de las señales PSG los cuales se describen en la Figura #. Estos canales son las características para la entrada del modelo. Se escogieron a partir de la naturaleza de la enfermedad.

Channel Number	Type of Signal	Label	Frequency Rate (Hz)	Butterworth Notch Filter	Description
1	EOG	LOC-A2	200	0.3 Hz–35 Hz, 50 Hz	Left eye movements
2	EOG	ROC-A1	200	0.3 Hz–35 Hz, 50 Hz	Right eye movements
3	EEG	F3-A2	200	0.3 Hz–35 Hz, 50 Hz	Brain channels with references A1 and A2, placed in the left and right ear lobes
4	EEG	C3-A2	200	0.3 Hz–35 Hz, 50 Hz	Brain channels with references A1 and A2, placed in the left and right ear lobes
5	EEG	O1-A2	200	0.3 Hz–35 Hz, 50 Hz	Brain channels with references A1 and A2, placed in the left and right ear lobes
6	EEG	F4-A1	200	0.3 Hz–35 Hz, 50 Hz	Brain channels with references A1 and A2, placed in the left and right ear lobes
7	EEG	C4-A1	200	0.3 Hz–35 Hz, 50 Hz	Brain channels with references A1 and A2, placed in the left and right ear lobes
8	EEG	O2-A1	200	0.3 Hz–35 Hz, 50 Hz	Brain channels with references A1 and A2, placed in the left and right ear lobes
9	Chin EMG	X1	200	10 Hz–70 Hz, 50 Hz	Chin EMG, placed between the chin and the lower lip
10	ECG (EKG)	X2	200	50 Hz	Electrocardiographic
11	Leg-1 EMG	X3	200	10 Hz–70 Hz, 50 Hz	Left leg movement
12	Leg-2 EMG	X4	200	10 Hz–70 Hz, 50 Hz	Right leg movement
13	Snore	X5	200	10 Hz–70 Hz, 50 Hz	Snore (derived)
14	Flow-1	X6	12.5		Airflow (pressure based)
15	Flow-2	DC3	25		Airflow (pressure based)
16	Abdominal Effort	X7	25		Abdominal efforts
17	Abdominal Effort	X8	25		Abdominal efforts
18	Pulse Oximetry	SaO2	12.5		Pulse oximetry (SaO2)
19	Body Position	DC8	25		Body position (BPOS)

Fig. 1. Detalles de las señales grabadas en el dataset ISRUC-Sleep resaltando los canales seleccionados adaptado de [10].

B. Modelos desarrollados

Se realizaron tres modelos con estructuras diferentes descritas en la teoría, como lo fue:

1) Modelo de red neuronal convolucional de 1 dimensión (CNN – 1D)

Se desarrolló una red neuronal convolucional unidimensional (CNN 1D) compuesta por dos capas convolucionales con 32 y 64 filtros, y tamaños de kernel de 5 y 3 respectivamente. Cada capa fue seguida por batch normalization y reducción de dimensionalidad mediante MaxPooling. La salida fue aplanada y pasada por una capa densa de 64 neuronas con activación ReLU, seguida de una capa de dropout con una tasa de 0.3. La capa final consistió en una neurona con activación sigmoide para realizar una clasificación binaria. El modelo se entrenó durante 30 épocas con un tamaño de batch de 32, utilizando el optimizador Adam y la función de pérdida de entropía cruzada binaria.

2) Modelo con Máquina de Soporte Vectorial (SVM)

Este modelo se implementó utilizando el clasificador SVC de Scikit-learn. Se realizó una búsqueda de hiperparámetros mediante GridSearchCV, explorando distintas combinaciones de C (parámetro de regularización), kernel (tipo de función del núcleo) y gamma (coeficiente para núcleos no lineales).

Se probaron los siguientes valores:

- 'C': valores entre 0.01 y 100 (con np.logspace(-2, 2, 10))
- 'kernel': ['linear', 'rbf', 'poly', 'sigmoid']
- 'gamma': ['scale', 'auto']

La métrica usada para la validación cruzada fue el F1-Score con cv=5 y ejecución paralela (n_jobs=-1) para eficiencia.

Se estandarizaron los datos antes de entrenar el modelo. Finalmente, se seleccionó la mejor combinación de hiperparámetros que optimizaba la métrica F1.

3) Modelo de perceptrón multicapa (MLP)

Se implementó un modelo de red neuronal multicapa utilizando la clase MLPClassifier de scikit-learn. La arquitectura consistió en dos capas ocultas con 32 y 16 neuronas respectivamente, empleando la función de activación ReLU y el optimizador Adam. Se utilizó una tasa de aprendizaje inicial de 0.001, regularización L2 con un valor de 0.001, y un tamaño de batch de 32. El modelo fue entrenado manualmente durante 30 épocas, controlando el número de iteraciones por época mediante max_iter=1 y fijando la semilla aleatoria en 42 para garantizar la reproducibilidad.

TABLA II. MÉTRICAS ENCONTRADAS DE LOS MODELOS

Métrica		Precisión	Recall	F1-Score	Accuracy
CNN	Entrenamiento	0,997	0,997	0,997	0,997
	Test	0,829	0,828	0,828	0,828
SVM	Entrenamiento	<u>0,9421</u>	<u>0,9419</u>	<u>0,9419</u>	<u>0,9419</u>
	Test	<u>0,8581</u>	<u>0,8571</u>	<u>0,8571</u>	<u>0,8571</u>
MLP	Entrenamiento	0,9984	0,9984	0,9984	0,9984
	Test	0,7253	0,7253	0,7253	0,7253

Los modelos se encuentran en el repositorio [Aruallo/Apnea_identification](#) público en Github además de los diferentes Notebooks utilizados en el estudio.

IV. RESULTADOS

Se desarrollaron y compararon tres modelos de clasificación: una Red Neuronal Convolucional 1D (CNN - 1D), un Perceptrón Multicapa (MLP) y una Máquina de Vectores de Soporte (SVM) que tuvieron sus respectivas curvas de aprendizaje respecto a la precisión de clasificación. Estas se denotan en las Figuras 2, 3 y 4 respectivamente.

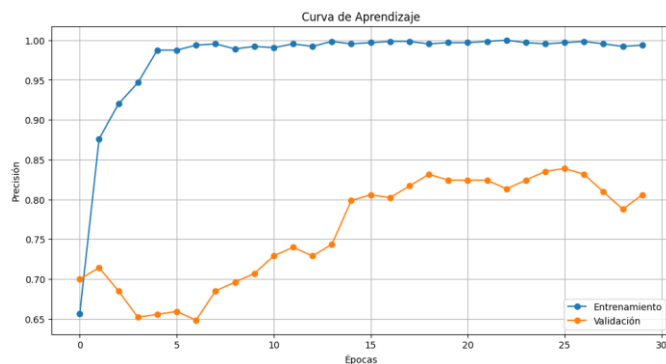


Fig. 2. Curva de aprendizaje de CNN 1D respecto a la precisión.

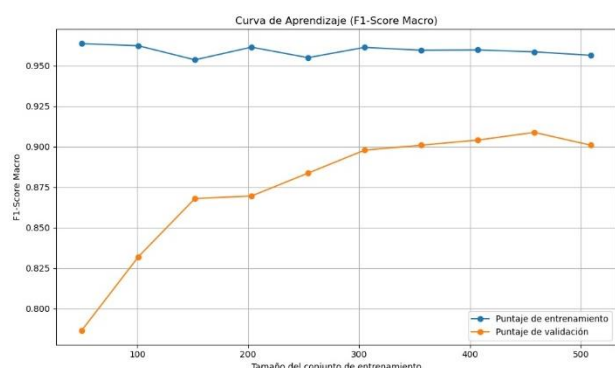


Fig. 3. Curva de aprendizaje de SVM.

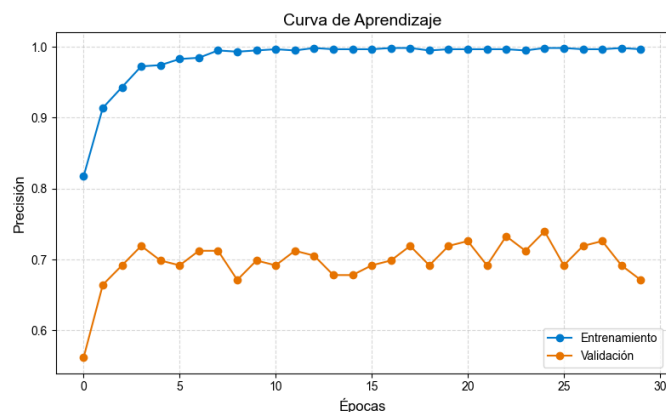


Fig. 4. Curva de aprendizaje de MLP respecto a la precisión.

El modelo MLP mostró un comportamiento característico de sobreajuste: alcanzó rápidamente una precisión cercana al 100% en los datos de entrenamiento, pero su desempeño en validación se mantuvo estancado alrededor del 70%. Esto indica que el modelo aprendió bien los datos de entrenamiento, pero no logró generalizar adecuadamente a nuevos datos, posiblemente debido a una arquitectura poco adecuada o a la falta de regularización.

Por su parte, la CNN - 1D presentó un mejor equilibrio entre entrenamiento y validación. Si bien también alcanzó una alta precisión en entrenamiento, la precisión en validación fue

superior a la del MLP, llegando a valores cercanos al 80%. Esto sugiere que la red fue capaz de capturar características relevantes de las señales PSG, especialmente aquellas de tipo temporal, propias de las series fisiológicas. No obstante, la curva de validación presentó cierta inestabilidad, lo cual indica que aún hay espacio para mejorar mediante técnicas como regularización, ajuste de hiperparámetros o aumento de datos.

Finalmente, el modelo SVM fue el que presentó los mejores resultados generales. Las curvas de aprendizaje mostraron un F1-Score macro alto y estable tanto en entrenamiento como en validación, sin señales evidentes de sobreajuste. A medida que se incrementó el tamaño del conjunto de entrenamiento, el rendimiento en validación mejoró de forma consistente, alcanzando valores cercanos al 90%. Esto demuestra la capacidad del modelo para generalizar bien con volúmenes moderados de datos y con un costo computacional relativamente bajo en comparación con las redes neuronales.

En resumen, si bien los modelos basados en redes neuronales tienen potencial para tareas más complejas o cuando se dispone de grandes volúmenes de datos, en este caso particular la SVM se posicionó como el modelo más robusto y efectivo para la tarea de clasificación de apnea del sueño con señales PSG.

CONCLUSIONES

En este trabajo se desarrollaron y evaluaron distintos modelos de aprendizaje automático para la detección de apnea del sueño a partir de señales fisiológicas del conjunto de datos ISRUC-Sleep. Se implementaron modelos como redes neuronales convolucionales (CNN), perceptrones multicapa (MLP) y máquinas de soporte vectorial (SVM), aplicando técnicas de preprocesamiento, normalización y validación cruzada.

Entre los modelos comparados, la SVM mostró el mejor rendimiento en datos de prueba, con un equilibrio destacado entre precisión y capacidad de generalización, superando ligeramente a la CNN y con un margen más amplio respecto al MLP. Esto resalta la importancia de la elección adecuada del modelo según la naturaleza del problema y las características de los datos.

En conjunto, los resultados obtenidos evidencian que el aprendizaje automático puede ser una herramienta poderosa y confiable para asistir en el diagnóstico de trastornos respiratorios del sueño como la apnea, siempre que se acompañe de un tratamiento cuidadoso de los datos y una evaluación rigurosa de los modelos.

Una de las principales limitaciones durante el entrenamiento de los modelos fue la disponibilidad restringida de datos, así como la selección adecuada de características relevantes para el proceso de clasificación. A pesar de estas limitaciones, se demostró que es posible clasificar eventos de apnea del sueño utilizando modelos de aprendizaje automático.

REFERENCES

- [1] "Los beneficios de dormir", Los Institutos Nacionales de Salud. Consultado: el 26 de marzo de 2025. [En línea]. Disponible en: <https://salud.nih.gov/recursos-de-salud/nih-noticias-de-salud/los-beneficios-de-dormir>

- [2] “‘Dormir no es descansar’: los trastornos del sueño y sus consecuencias : Prensa Gobierno de Mendoza”. Consultado: el 26 de marzo de 2025. [En línea]. Disponible en:
<https://www.mendoza.gov.ar/prensa/trastornos-del-sueno-y-sus-consecuencias-dormir-no-es-descansar/>
- [3] A. K. Patel, V. Reddy, K. R. Shumway, y J. F. Araujo, “Physiology, Sleep Stages”, en *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. Consultado: el 17 de julio de 2025. [En línea]. Disponible en:
<http://www.ncbi.nlm.nih.gov/books/NBK526132/>
- [4] A. Pouliou *et al.*, “A New Approach for Sleep Stage Identification Combining Hidden Markov Models and EEG Signal Processing”, *J. Med. Biol. Eng.*, vol. 45, núm. 1, pp. 1–12, feb. 2025, doi: 10.1007/s40846-025-00928-5.
- [5] S. M. Valiensi, A. Folgueira, M. Ponce de León, y J. Alfie, “Análisis de movimientos periódicos de piernas durante el sueño, presión arterial de 24 horas y frecuencia cardíaca”, *Rev. Fac. Cienc. Medicas Cordoba Argent.*, vol. 80, núm. 1, pp. 3–10, mar. 2023, doi: 10.31053/1853.0605.v80.n1.36765.