

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO**



**Segmentación no supervisada de señales
electroencefalográficas comparada
con la etapificación visual estándar para estudios de sueño**

Lucas Javier Rivera Clasing

TESIS PARA OPTAR AL GRADO DE MAGISTER EN INFORMÁTICA MÉDICA

**Director de Tesis: Prof. Dr. Adrián Ocampo Garcés
Co-Director de Tesis: Prof. Dr. Alejandro Bassi Acuña**

2025

*A mi amada esposa Belén.
A mi hijo Alonso y a los que vendrán.*

Agradecimientos

A mi Dios.

A mi esposa, por ser quien es.

A mi madre, por amarme no importando nada.

A mi padre, por inspirarme a aspirar a más, incluso en un contexto donde soñar no siempre parecía posible.

A mi hermana, porque no me imagino un mundo sin ella.

A Don Andrónico Luksic Craig, por apoyarme en este sueño de seguir estudiando.

A mi director de tesis Dr. Adrián Ocampo, por creer en mí.

Al Dr. Alejandro Bassi, mi codirector de tesis, por su paciencia inagotable y por acompañarme siempre con respeto, incluso ante mis mil preguntas.

A la Universidad de Chile y su cultura universitaria.

A mi amigo y compañero de cohorte del Magíster en Informática Médica, Richard von Moltke, por su constante apoyo académico y su amistad incondicional en los momentos en que más lo necesité.

A la comisión evaluadora, encabezada por su presidente Dr. Steffen Härtel, y nuevamente a mis directores de tesis, por su comprensión y empatía durante la etapa final de este magíster, en la que me tocó atravesar la experiencia más difícil de mi vida.

Y nuevamente, gracias.

Tesis financiada por:

*Fundación Guillermo Puelma, Universidad de Chile
Proyecto FONDECYT (ANID) 1231901*

INDICE

1.-RESUMEN.....	1
2.-ABSTRACT.....	3
3.-INTRODUCCION	5
3.1.- Ciclo sueño-vigilia.....	5
3.2.- Polisomnografía.....	5
3.3.- Etapificación del ciclo sueño-vigilia	6
3.4.- Electroencefalograma	7
3.5.- Análisis de señales.....	9
3.5.1.-Tipo de señales.....	10
3.5.2.-Transformada de Fourier.....	11
3.5.3.-Transformada de Fourier de tiempo corto.....	11
3.5.4.-Espectrograma.....	12
3.5.5.-Teorema de Muestreo de Nyquist-Shannon	14
3.6.- Inteligencia Artificial.....	14
3.7.- Estrategias para evaluar correlación	17
4.-PROBLEMA	19
5.-HIPÓTESIS.....	20
6.-OBJETIVOS.....	21
6.1.- Objetivo general	21
6.2.- Objetivos específicos.....	21
7.-MATERIAL Y MÉTODOS.....	22
7.1.- Material.....	22
7.2.- Método.....	23

7.2.1.-Objetivo específico 1.....	24
7.2.2.-Objetivo específico 2.....	28
7.2.3.-Objetivo específico 3.....	30
7.2.4.-Objetivo específico 4.....	31
8.-RESULTADOS	33
8.1.- Objetivo específico 1	33
8.2.- Objetivo específico 2	35
8.3.- Objetivo específico 3	40
8.4.- Objetivo específico 4	42
9.-DISCUSIÓN.....	46
10.-CONCLUSIÓN	50
11.-ANEXO	52
12.-REFERENCIAS	65

1.- RESUMEN

Introducción -problema

La etapificación correcta de los estados del ciclo sueño-vigilia (Estados del CSV) —que comprende la vigilia, las etapas N1, N2 y N3 del sueño No REM, y el sueño REM— es un requisito esencial en los estudios clínicos del sueño. El *gold standard* para la etapificación del sueño es la opinión experta, realizada mediante el diagnóstico visual de los registros electrofisiológicos presentes en la técnica de Polisomnografía (*Polysomnography*, PSG). En este proceso, los registros se categorizan siguiendo reglas establecidas por la *American Academy of Sleep Medicine* (AASM).

El procedimiento está orientado al diagnóstico clínico mediante la detección visual de elementos de interés, capturando un repertorio reducido de características presentes en las señales, en ventanas temporales arbitrarias. Una debilidad notable en las normas de evaluación actuales radica en la carencia de estrategias analíticas avanzadas para llevar a cabo una etapificación más precisa y detallada del ciclo sueño-vigilia, más allá de la limitada capacidad visual humana.

Solución

Esta tesis se enfocó en aplicar métodos de extracción de características sobre la señal de electroencefalogramas (*Electroencephalogram*, EEG) de registro de sueño, realizando un análisis no supervisado mediante técnicas informáticas de *clustering*, con el objetivo de estudiar el agrupamiento natural de los datos registrados.

Metodología

Se utilizaron registros polisomnográficos de adultos sanos sin trastornos del sueño, extraídos desde la base de datos pública ISRUC-Sleep y desde un conjunto recolectado en el marco del proyecto FONDECYT 123190. En total se analizaron 16 pacientes, cuyas señales EEG fueron adquiridas a una frecuencia de muestreo de 200 Hz

Las señales EEG fueron preprocesadas, filtradas, segmentadas y reducidas mediante PCA. Se aplicó *K-means* y *Agglomerative Clustering* para refinar la excesiva fragmentación inicial de la señal, de este modo llevando una cantidad de alrededor de 25.000 etiquetas distintas a un número más manejable que pudiese ser óptimo para lograr una comparación eficiente con la etapificación tradicional.

Métrica

Como métrica se utilizó la Información Mutua Ajustada (AMI) y Kappa de Cohen, permitiendo evaluar la correspondencia entre la etapificación tradicional versus la no supervisada.

Hipótesis

El análisis no supervisado del EEG del sueño, utilizando técnicas de agrupamiento (*clustering*), permite obtener una segmentación temporal de los estados del sueño que muestra una alta concordancia estadística con la etapificación visual polisomnográfica realizada por expertos.

Objetivo general

Obtener segmentaciones estables de señales del EEG del sueño utilizando técnicas no supervisadas de agrupamiento (*clustering*) y evaluar su correspondencia con la etapificación visual polisomnográfica realizada por expertos, aplicando criterios para reducir la fragmentación.

Objetivos específicos

1. Seleccionar una batería métodos para extraer características de las señales de EEG.
2. Determinar las características más relevantes de la señal de EEG para la construcción del vector de características.
3. Refinar algorítmicamente los agrupamientos para evitar fragmentación exagerada de la segmentación no supervisada.
4. Evaluar correspondencia entre resultados de métodos de análisis no supervisado con la etapificación de experto basada en las reglas de la AASM.

Resultados

La evaluación estadística mostró un acuerdo moderado entre el agrupamiento no supervisado y la clasificación manual ($AMI \approx 0.2952$ y $Kappa \approx 0.4397$), lo que indica que, si bien el método captura dinámicas clave del sueño, no reproduce completamente la etapificación manual.

Discusión

Los hallazgos sugieren que la arquitectura del sueño podría ser más compleja que el modelo convencional de cinco etapas. El enfoque no supervisado propuesto proporciona una representación más detallada y dinámica del sueño y podría servir como base para modelos de etapificación basados en datos. En el futuro, este método podría contribuir a la

etapificación del sueño basada exclusivamente en EEG, reduciendo la dependencia de otros canales PSG. Además, la re-evaluación del ciclo sueño-vigilia a través de métodos basados en datos podría conducir a una comprensión más fisiológicamente fundamentada de los estados cerebrales durante el sueño

2.- ABSTRACT

Introduction – Problem

Accurate staging of the sleep-wake cycle states—including wakefulness, the N1, N2, and N3 stages of non-REM sleep, and REM sleep—is essential in clinical sleep studies. The gold standard for sleep staging is expert assessment, based on the visual interpretation of electrophysiological signals recorded through polysomnography (PSG). During this process, recordings are categorized following rules established by the American Academy of Sleep Medicine (AASM).

This procedure is oriented toward clinical diagnosis through the visual detection of relevant elements, capturing a limited set of signal characteristics within arbitrarily defined time windows. A notable limitation of the current evaluation guidelines lies in the absence of advanced analytical strategies to enable a more precise and detailed staging of the sleep-wake cycle, beyond the inherent limitations of human visual analysis.

Solution

This thesis focused on the application of feature extraction methods on sleep EEG (electroencephalogram) signals, implementing an unsupervised analysis using clustering techniques in order to investigate the natural grouping structure of the recorded data.

Methodology

Polysomnographic recordings from healthy adults without sleep disorders were used, sourced from the ISRUC-Sleep public dataset and from recordings collected under the FONDECYT project 123190. In total, 16 patients were analyzed, with EEG signals sampled at 200 Hz.

EEG signals were preprocessed, filtered, segmented, and reduced via principal component analysis (PCA). K-means and Agglomerative Clustering were applied to refine the initial

over-fragmentation of the signal, reducing approximately 25,000 unique labels to a more manageable number suitable for effective comparison with traditional sleep staging.

Metrics

Adjusted Mutual Information (AMI) and Cohen's Kappa were used as evaluation metrics to assess the correspondence between traditional expert-based staging and the unsupervised segmentation.

Hypothesis

Unsupervised analysis of sleep EEG using clustering techniques enables a temporal segmentation of sleep states that demonstrates high statistical agreement with the expert visual staging derived from PSG.

General Objective

To obtain stable segmentations of sleep EEG signals using unsupervised clustering techniques and evaluate their correspondence with expert visual staging, applying criteria to reduce over-fragmentation.

Specific Objectives

1. Select a set of methods for extracting features from EEG signals.
2. Determine the most relevant features of EEG signals for the construction of the feature vector.
3. Algorithmically refine the clusters to prevent excessive fragmentation in unsupervised segmentation.
4. Evaluate the correspondence between unsupervised analysis results and expert staging based on AASM guidelines.

Results

The statistical evaluation showed moderate agreement between unsupervised clustering and manual classification ($AMI \approx 0.2952$, $Kappa \approx 0.4397$), indicating that while the method captures key sleep dynamics, it does not fully replicate the manual staging.

Discussion

The findings suggest that sleep architecture may be more complex than the conventional five-stage model. The proposed unsupervised approach offers a more detailed and dynamic representation of sleep and may serve as a foundation for data-driven staging models. In the future, this method could support EEG-only staging strategies, reducing reliance on additional PSG channels. Furthermore, re-evaluating the sleep-wake cycle through data-driven methods may lead to a more physiologically grounded understanding of brain states during sleep.

3.- INTRODUCCION

3.1.- Ciclo sueño-vigilia

El ciclo sueño-vigilia se compone de tres estados conductuales bien definidos y distintos: la vigilia (W, *Wake*), el sueño de movimientos oculares no rápidos (NREM, *Non Rapid Eye Movement*) y el sueño de movimientos oculares rápidos (REM, *Rapid Eye Movement*).

Los estados de sueño son formas especializadas de reposo caracterizados por el desacople cognitivo respecto de la entrada sensorial. El desenganche sensorial durante el sueño está determinado por un complejo sustrato de circuitos neurales que modifican la actividad del cerebro, sus entradas sensoriales y sus salidas motoras, permitiendo su funcionamiento “fuera de línea”. El funcionamiento “fuera de línea” del cerebro durante el sueño cumple un rol central en las propiedades restauradoras del dormir (1, 2).

El sueño es un estado activamente regulado por el cerebro y no se limita simplemente a la ausencia de vigilia (3). En humanos, los estados de sueño NREM y REM se alternan de manera continua y regular durante la noche, formando los llamados ciclos ultradianos de sueño, que ocurren entre 4 y 6 veces por noche. Un adulto típico experimenta ciclos de sueño completos que duran entre 90 y 120 minutos. Cada ciclo comienza con el sueño NREM, que progresa a través de sus tres etapas, y finaliza con el sueño REM. El primer período de sueño NREM es el más largo, mientras que el primer episodio de sueño REM es el más corto. Sin embargo, a medida que avanza la noche, los episodios de sueño REM se vuelven progresivamente más largos, representando entre el 20% y el 25% del tiempo total de sueño (4).

3.2.- Polisomnografía

El método diagnóstico utilizado para estudiar el ciclo sueño-vigilia tiene por nombre Polisomnografía (PSG). Se realiza en un entorno controlado, generalmente en un laboratorio acondicionado, y consiste en registrar diversas señales de manera no invasiva permitiendo así tener mayor certeza al momento de realizar la etapificación del ciclo sueño-vigilia. Las principales variables que registra una PSG son el Electroencefalograma (EEG), mide la actividad eléctrica cerebral, el Electrooculograma (EOG), registra los movimientos oculares, y el Electromiograma (EMG), que detecta la actividad muscular. Se pueden registrar variables dependiendo del análisis final que se quiera obtener, como por ejemplo señales cardíacas utilizando el electrocardiograma (ECG), señales de respiración, oximetría de pulso y/o cámaras de video (3, 4, 9).

3.3.- Etapificación del ciclo sueño-vigilia

La etapificación o *scoring* de las etapas del sueño es el *gold standard* para el estudio del ciclo sueño-vigilia en humanos (5, 6). Es el proceso de clasificación y asignación de las diferentes etapas de sueño a intervalos de tiempo específicos de la señal registrada, generalmente en períodos de 30 segundos llamados "épocas". La asignación de las etapas del sueño se basa tradicionalmente en la evaluación visual de señales electrofisiológicas adquiridas a través de una polisomnografía (PSG) (7), teniendo una representación gráfica mediante un hipnograma (Figura 1).

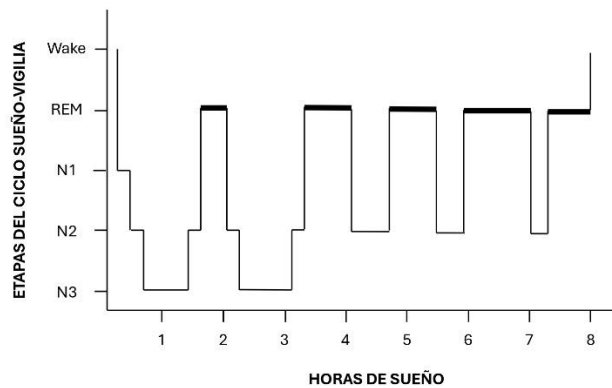


Figura 1. Hipnograma clásico. Representación gráfica del sueño que muestra cómo cambian los Estados del CSV a lo largo del tiempo durante una noche.

Esta clasificación se basa en las pautas desarrolladas por Rechtschaffen y Kales (R&K) en 1968 (8). La etapificación general del ciclo sueño vigilia, es dividida en tres estados: W, NREM y REM, y el estado NREM se subdivide en tres etapas: N1, N2, N3. Por lo tanto, la etapificación consiste en asignar cada época a una de cinco clases: W, N1, N2, N3, REM, teniendo cada una de estas etapas características distintivas:

- Etapa W (vigilia): Se caracteriza por la presencia del ritmo alfa en la señal del EEG, generalmente en la región occipital, con frecuencia de señal de 8-13 Hz y alrededor de 50 mV de amplitud (6).
- Etapa N1: es una fase de transición entre la vigilia y el sueño profundo predominando ondas theta (3,5-7,5 Hz). Durante esta etapa la persona puede despertarse fácilmente. La actividad cerebral medida por el electroencefalograma (EEG) muestra una amplitud limitada de hasta 200 mV.

- Etapa N2: Pueden aparecer husos del sueño y complejos K (grafo elemento del registro del EEG que consta de una onda bifásica de voltaje elevado). Se debe clasificar como N2 si durante la última mitad de la época anterior o la primera mitad de la actual hay uno o más complejos K o uno o más husos de sueño, y se debe seguir clasificando como N2 (incluso sin husos del sueño ni complejos K) hasta que aparezca una nueva etapa. N2 generalmente abarca alrededor del 50% del tiempo total de sueño (6).
- Etapa N3: Esta etapa se conoce como sueño profundo y se caracteriza por la presencia predominante de ondas delta en el EEG. También pueden observarse algunos husos del sueño y complejos K aislados. Para clasificar una época como N3, al menos el 20% de ella debe estar compuesta por ondas delta. En promedio, la etapa N3 representa alrededor del 20% del tiempo total de sueño (6).
- Etapa REM: Los movimientos oculares son rápidos. Los microdespertares pueden ocurrir con mayor facilidad. El EEG muestra baja amplitud, frecuencia mixta y posibles ondas en forma de sierra. El EMG se encuentra en su nivel más bajo. Al despertar de sueño REM, las personas reportan muy frecuentemente actividad onírica (sueños). Una etapa debe seguir clasificándose como REM hasta que ocurra una o más de las siguientes condiciones: una transición a la etapa W o N3, aumento del tono muscular EMG del mentón o aparición de un complejo K sin despertar. Esta etapa normalmente abarca entre el 20% y el 25% del tiempo total de sueño (6).

3.4.- Electroencefalograma

El EEG es una de las señales biomédicas más importantes y valiosas en el campo de la neurología clínica, debido a su relevancia, utilidad práctica y bajo costo. Se trata de un método no invasivo que permite medir la actividad eléctrica de la corteza cerebral. Desde su descubrimiento por Berger en 1929, el EEG ha sido ampliamente utilizado, especialmente en el estudio del sueño y los procesos cognitivos (10). Para su registro, se emplean múltiples electrodos que se colocan en diversas ubicaciones del cuero cabelludo, siguiendo el sistema de colocación internacional 10-20 (Figura 2).

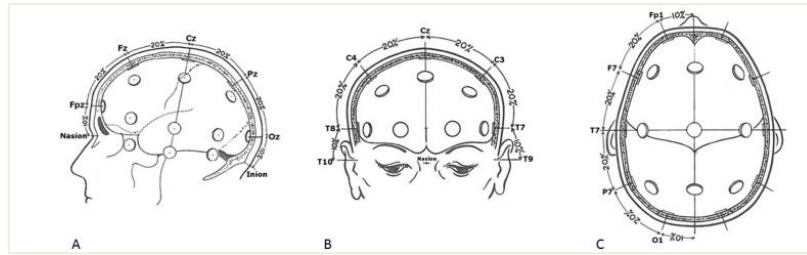


Figura 2: Colocación de electrodos estándar del sistema 10–20. A: vista lateral, B: frontal, C: desde arriba (42).

El EEG mide las variaciones en los potenciales eléctricos generados por la actividad neuronal sincronizada en la corteza cerebral. Estas variaciones se reflejan en diferentes tipos de ondas cerebrales que se clasifican según su frecuencia: delta (≤ 3 Hz), theta (3,5-7,5 Hz), alfa (8-13 Hz), beta (13-30 Hz) y gamma (>30 Hz) (Figura 3) (11).

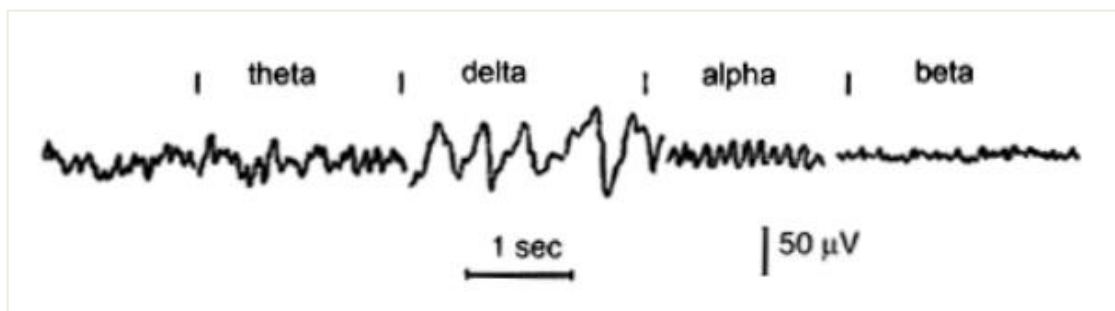


Figura 3. Representación de las ondas cerebrales registradas en un electroencefalograma (EEG). La figura muestra un ejemplo representativo de las principales ondas cerebrales y sus diferencias morfológicas capturadas en un registro de EEG, clasificadas según su frecuencia y amplitud. De izquierda a derecha, se pueden observar las ondas Theta (3.5 - 7.5 Hz), Delta (≤ 3 Hz), Alpha (8 - 13 Hz) y Beta (≥ 14 Hz). Esta figura destaca las diferencias morfológicas de las ondas cerebrales (43).

El EEG es fundamental para estudiar el ciclo sueño-vigilia, ya que permite identificar y clasificar las diferentes etapas del sueño basándose en patrones eléctricos específicos, particularmente los subestados del sueño NREM. Mientras el sueño REM y la vigilia se caracterizan por una actividad eléctrica rápida y de baja amplitud, indistinguibles a la observación visual (4), durante el sueño NREM el EEG exhibe una gradiente continua de

elementos de baja frecuencia (husos de sueño, complejos K, ondas delta) que progresan en amplitud desde la etapa más superficial (N1) hasta la etapa profunda (N3), en la que predominan las ondas delta. Los mecanismos generadores de estos grafoelementos han sido relacionados a procesos restauradores de la integridad tisular del cerebro y el cuerpo, y a procesos cognitivos de la mayor importancia tales como consolidación de la memoria, regulación emocional y plasticidad cerebral (12).

3.5.- Análisis de señales

El análisis de señales es un campo multidisciplinario que implica la extracción, procesamiento y evaluación de información contenida en las señales, las cuales pueden ser eléctricas, biológicas, acústicas o de otro tipo. El análisis de señales emplea técnicas matemáticas y computacionales para representar y descomponer señales complejas en componentes fundamentales, como frecuencia, amplitud y fase, dependiendo del método utilizado. Esto permite interpretar la estructura de la señal y extraer información relevante en diferentes dominios. Para ello se utilizan distintos tipos de transformadas matemáticas, como la Transformada de Fourier (FT), la Transformada de Wavelets, entre otras.

El análisis de señales implica una serie de pasos estructurados que permiten procesar, interpretar y extraer información relevante. A continuación, se detallan los principales pasos involucrados en el análisis de señales:

- Adquisición de la señal (15): el proceso comienza con la recolección de datos mediante sensores o dispositivos de medición. La calidad de los datos adquiridos es fundamental para garantizar resultados precisos.
- Preprocesamiento de la señal (15): en esta etapa, la señal es preparada para su análisis incluyendo técnicas como:
 - Filtrado: para eliminar ruido e interferencias.
 - Normalización: proceso matemático mediante el cual se ajustan los valores de una señal para que cumplan con un rango o escala estándar. Este procedimiento se utiliza para mejorar la comparabilidad entre diferentes señales o para facilitar su análisis en algoritmos que son sensibles a la escala de los datos.
 - Segmentación: división de la señal en segmentos específicos para facilitar su análisis.

- Transformación de la señal (14): permite analizar la señal en diferentes dominios. Algunas de estas transformaciones son la FT (estudia el contenido frecuencial de señales estacionarias), transformada Wavelet (permite analizar la señal en el dominio temporal y frecuencial especialmente para señales no estacionarias), transformada de Hilbert (para calcular la señal analítica y la envolvente de una señal), entre otras.
- Extracción de características (15): este paso implica identificar y medir características relevantes de la señal, como por ejemplo frecuencias dominantes, amplitud y patrones temporales.
- Clasificación y detección de patrones (17): Los algoritmos de aprendizaje automático se utilizan para clasificar señales y detectar patrones complejos.
- Interpretación de resultados (15): Los datos procesados se interpretan en términos específicos para el análisis en cuestión.

3.5.1.-Tipo de señales

En el ámbito de la física teórica y la ingeniería, una señal se define como una función matemática que representa la variación de una magnitud física con respecto al tiempo, el espacio o cualquier otra variable independiente (13). Las señales pueden presentarse en diversas formas (eléctricas, acústicas, ópticas o fisiológicas) dependiendo del medio o sistema del que provienen, y se utilizan para transmitir información en sistemas de comunicación, control, análisis biomédico, entre otros (15).

Desde el punto de vista matemático, una señal puede expresarse como una función continua o discreta (47).

Una clasificación fundamental en el análisis de señales es la distinción entre señales estacionarias y no estacionarias.

Señales Estacionarias

Una señal es estacionaria si sus propiedades estadísticas, como la media, la varianza y la función de autocorrelación, permanecen constantes en el tiempo. Esto implica que el comportamiento de la señal no cambia, y sus características son predecibles. Ejemplos comunes incluyen señales sinusoidales puras y ruido blanco (41).

Señales No Estacionarias

Una señal es no estacionaria cuando sus propiedades estadísticas cambian con el tiempo. La mayoría de las señales del mundo, como el habla, los datos financieros y los registros EEG, son no estacionarias debido a la variación en sus patrones (41).

3.5.2.-Transformada de Fourier

La FT permite convertir una señal del dominio temporal al dominio frecuencial. Fue desarrollada por el matemático francés Jean-Baptiste Fourier, quien demostró que cualquier señal periódica puede representarse como una suma infinita de senos y cosenos con diferentes frecuencias, amplitudes y fases (13).

Matemáticamente, la FT de una función continua $x(t)$ está dada por:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt$$

Donde:

- $X(f)$: Representa la señal en el dominio frecuencial.
- $x(t)$: Señal en el dominio temporal.
- f : Frecuencia.
- $e^{-j2\pi ft}$: Función exponencial compleja que representa ondas sinusoidales.

La representación resultante se conoce como el espectro de frecuencia (14).

Tipos de FT:

- Transformada de Fourier Continua (FT): Se aplica a señales infinitas y continuas.
- Transformada de Fourier Discreta (DFT): Se aplica a señales discretas y finitas, siendo la versión digital de la FT (18).
- Transformada Rápida de Fourier (FFT): Es un algoritmo eficiente para calcular la DFT de manera rápida (41).
- Transformada de Fourier de Tiempo Corto (STFT): Extensión de la FT que permite analizar señales no estacionarias, dividiéndolas en segmentos de tiempo.

3.5.3.-Transformada de Fourier de tiempo corto

La Transformada de Fourier de Tiempo Corto (STFT) es una técnica matemática utilizada para analizar señales no estacionarias (13). A diferencia de la FT tradicional, que otorga una representación global del dominio de frecuencias de una señal, la STFT permite obtener información tanto del dominio temporal como del dominio frecuencial (14).

Fundamento Matemático de la STFT.

La STFT se basa en dividir una señal continua o discreta en segmentos de tiempo, aplicando la FT a cada segmento utilizando una función ventana deslizante. Matemáticamente, la STFT en tiempo discreto se expresa como:

$$X(m, \omega) = \sum_{\{n=-\infty\}}^{\{\infty\}} x[n] \omega[n - m] e^{-j\omega n}$$

Donde:

- Se aplica a señales discretas $x[n]$, es decir, muestras de una señal continua.
- $\omega[n - m]$: Función ventana centrada en m .
- $e^{-j\omega n}$: Función exponencial compleja que representa la FT (16).

Esta ecuación permite generar un espectrograma, que es una representación visual de la señal, con el tiempo en el eje horizontal, la frecuencia en el eje vertical y la magnitud como la intensidad del color (47).

Elección de la Ventana en la STFT.

La selección de la ventana es crucial para determinar la precisión del análisis. Ventanas comunes incluyen:

- Ventana de Hamming: Suaviza los bordes y reduce artefactos en el espectro.
- Ventana de Hann: Minimiza la fuga espectral con transiciones suaves.
- Ventana Rectangular: Sin atenuación en los bordes, pero con mayor fuga espectral (41).

Existe un compromiso entre la resolución temporal y frecuencial:

- Ventanas estrechas: buena resolución temporal, pero baja resolución en frecuencia.
- Ventanas amplias: buena resolución en frecuencia, pero baja precisión temporal (18).

3.5.4.-Espectrograma

Un espectrograma es una representación visual del contenido frecuencial de una señal en función del tiempo. Muestra cómo las diferentes frecuencias presentes en una señal varían en intensidad a lo largo del tiempo, permitiendo una interpretación conjunta de ambos dominios (13).

Matemáticamente, un espectrograma se genera mediante la aplicación de la Transformada de Fourier de Tiempo Corto (STFT), que descompone una señal en componentes

frecuenciales dentro de intervalos de tiempo definidos por una función ventana (47). Matemáticamente se obtiene calculando la magnitud al cuadrado de la STFT:

$$|X(m, \omega)|^2$$

El proceso para generar un espectrograma implica aplicar la STFT a una señal discreta mediante una ventana deslizante. La señal se divide en segmentos temporales y, en cada segmento, se calcula la Transformada de Fourier. El resultado es un conjunto de espectros de frecuencia a lo largo del tiempo, visualizados como una imagen bidimensional (Figura 4) (47).

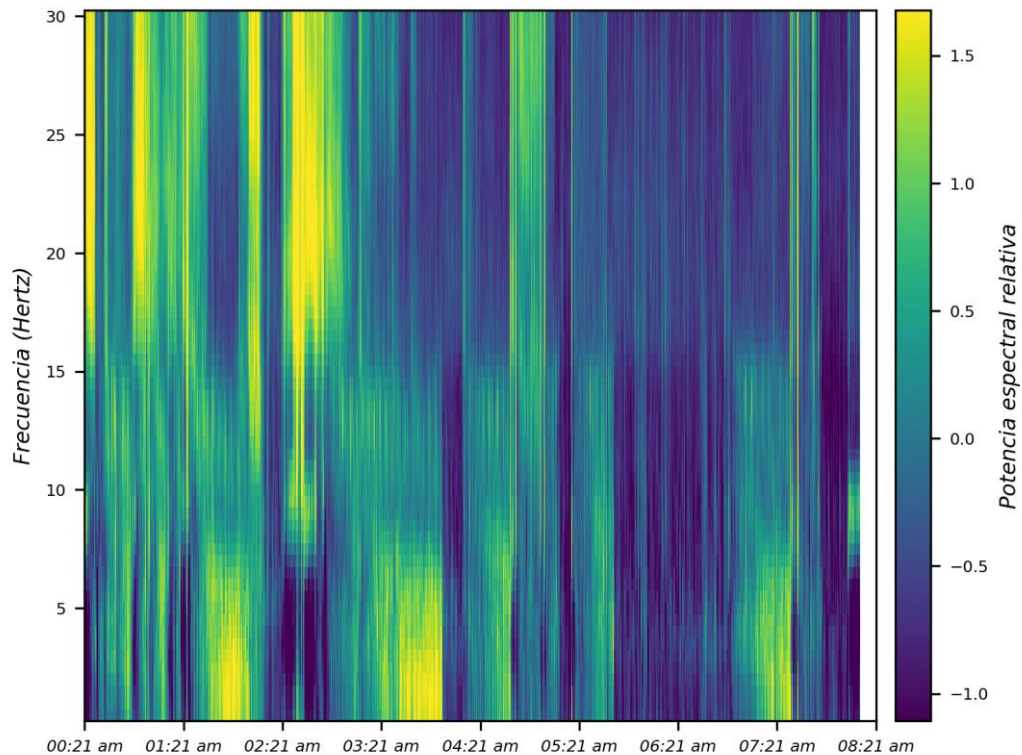


Figura 4. Espectrograma de una noche de sueño paciente normal. Representación visual del contenido frecuencial de una señal en función del tiempo. Muestra cómo las diferentes frecuencias presentes en una señal varían en intensidad a lo largo del tiempo. La abscisa representa el tiempo. La ordenada indica frecuencia. La intensidad del color muestra la potencia de cada frecuencia.

La precisión de un espectrograma depende de varios factores técnicos como son la función de ventana, el tamaño de ventana y la tasa de muestreo (debe cumplir con el Teorema de Muestreo de Nyquist para evitar el *aliasing*) (19).

3.5.5.-Teorema de Muestreo de Nyquist-Shannon

El teorema de muestreo de Nyquist-Shannon es un principio fundamental en el procesamiento de señales digitales y establece que una señal analógica de banda limitada puede ser reconstruida perfectamente a partir de sus muestras discretas, siempre que la frecuencia de muestreo sea al menos el doble de la frecuencia máxima contenida en la señal. Esta frecuencia mínima de muestreo se conoce como la frecuencia de Nyquist (46).

Matemáticamente, si una señal $x(t)$ contiene componentes de frecuencia hasta un límite B , la frecuencia de muestreo f_s debe cumplir la siguiente condición: $f_s \geq 2B$. Esta restricción evita el fenómeno de aliasing, que ocurre cuando una señal es muestreada a una frecuencia menor a la de Nyquist, lo que provoca que diferentes componentes espectrales se superpongan y se vuelvan indistinguibles en el dominio digital, generando distorsión (13). El teorema de Nyquist-Shannon no solo es esencial en telecomunicaciones y procesamiento de audio, sino que también tiene aplicaciones críticas en la resonancia magnética, la compresión de imágenes y la transmisión digital de información. Su correcta aplicación garantiza una conversión precisa entre los dominios analógico y digital, evitando pérdidas de información y errores en la reconstrucción.

3.6.- Inteligencia Artificial

La Inteligencia Artificial (IA) es un campo de estudio dentro de las ciencias computacionales que se enfoca en desarrollar sistemas y algoritmos capaces de realizar tareas que requieren inteligencia humana, tales como el aprendizaje, la percepción, la toma de decisiones, entre otros. El objetivo principal de la IA es imitar procesos cognitivos complejos mediante métodos computacionales avanzados (20). Dentro de la IA, existen subcampos como el *Machine Learning* (ML) y el *Deep Learning* (DL), que proporcionan modelos y técnicas específicas para procesar grandes volúmenes de datos, aprender patrones y hacer predicciones sin intervención humana explícita (17).

El *Machine Learning* es un subcampo de la IA que permite a los sistemas aprender de los datos mediante algoritmos estadísticos, mejorando su rendimiento con la experiencia. Se divide en tres categorías principales: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo (21).

El *Deep Learning*, por su parte, es una extensión avanzada del ML que utiliza redes neuronales profundas, compuestas por múltiples capas de procesamiento. Estas capas permiten extraer características complejas de los datos mediante representaciones jerárquicas, siendo especialmente útiles en tareas como el reconocimiento de imágenes, el procesamiento del lenguaje natural y el análisis de señales biomédicas (22).

Aprendizaje Supervisado

El aprendizaje supervisado es una técnica fundamental en el campo del aprendizaje automático que consiste en entrenar un modelo a partir de un conjunto de datos etiquetados, donde cada entrada está asociada a una salida deseada. Este enfoque permite a los modelos aprender a mapear nuevas entradas a salidas correspondientes, lo que resulta invaluable para tareas de clasificación y regresión. Las aplicaciones del aprendizaje supervisado son vastas y abarcan desde la clasificación de imágenes y textos hasta la predicción de series temporales y la detección de fraudes (23).

Aprendizaje No Supervisado

El aprendizaje no supervisado es una rama del aprendizaje automático que se enfoca en descubrir patrones y estructuras latentes en datos sin etiquetas. A diferencia del aprendizaje supervisado, no requiere de un conjunto de entrenamiento con salidas conocidas. Esta característica lo hace ideal para explorar grandes conjuntos de datos y descubrir relaciones desconocidas, lo que lo convierte en una herramienta valiosa para el descubrimiento de patrones, la detección de anomalías y la reducción de dimensionalidad (24).

El *clustering* es un método de aprendizaje no supervisado que organiza los datos en grupos llamados *clusters* o agrupamientos. Su objetivo es que los elementos dentro de cada grupo sean más similares entre sí que con los de otros grupos (25). Los algoritmos de *clustering* se pueden dividir en: jerárquicos y particionales.

Clustering Jerárquico

El *clustering* jerárquico es una técnica que genera una estructura de árbol o dendrograma, donde los datos se agrupan en niveles jerárquicos. Este método no requiere especificar el número de agrupaciones de antemano y se realiza mediante dos enfoques principales (27,28):

- Modo aglomerativo (*bottom-up*): Comienza con cada punto de datos como un *cluster* individual. Luego, los pares de *clusters* más similares se fusionan sucesivamente hasta formar una jerarquía completa.
- Modo divisivo (*top-down*): Comienza con todos los datos en un solo *cluster*. Luego, el algoritmo divide cada *cluster* en *clusters* más pequeños de manera recursiva.

La distancia entre *clusters* se mide utilizando métodos como:

- Distancia de enlace completo: Considera la distancia máxima entre puntos de dos *clusters*.
- Distancia de enlace simple: Considera la distancia mínima entre puntos de dos *clusters*.
- Distancia de promedio de grupo: Considera el promedio de todas las distancias entre puntos de dos *clusters* (28).

El *clustering* jerárquico es interpretativamente potente debido a su estructura visual clara, aunque su complejidad computacional puede volverse un desafío para grandes conjuntos de datos.

Clustering Particional

A diferencia del *clustering* jerárquico, el *clustering* particional encuentra todas las agrupaciones al mismo tiempo, dividiendo los datos en subconjuntos sin imponer una estructura jerárquica. No construye una jerarquía, sino que genera una partición fija de los datos en un número determinado de *clusters*. A continuación, se presenta uno de los algoritmos de *clustering* particional más famosos y característicos de este tipo de análisis:

K-means: El algoritmo *K-means* es un método de *clustering* basado en la partición de datos. Divide un conjunto de datos en *K clusters*, donde *K* es un número definido previamente. El proceso sigue estos pasos:

- Se seleccionan aleatoriamente *K* centroides iniciales.
- Cada punto de datos se asigna al *cluster* cuyo centroide está más cercano según una medida de distancia (generalmente la distancia euclidiana).

- Se recalculan los centroides como el promedio de los puntos asignados a cada *cluster*.
- El proceso se repite hasta que los centroides ya no cambian significativamente (26).

Matemáticamente, el criterio de minimización está dado por:

$$J = \sum_{\{i=1\}}^{\{K\}} \sum_{\{x \in C_i\}}^{\{n\}} ||x - \mu_i||^2$$

Donde:

- J: Función de costo (distancia total).
- C_i : Conjunto de puntos en el *cluster* i
- μ_i : Centroide del *cluster* i

El algoritmo es eficiente y fácil de implementar, pero su principal desventaja es que es sensible a la inicialización de los centroides (26).

3.7.- Estrategias para evaluar correlación

En el análisis estadístico, existen diversas métricas diseñadas para evaluar relaciones, concordancias y dependencias entre conjuntos de datos. Entre estas métricas destacan el Kappa de Cohen (k) y la Información Mutua Ajustada (AMI), cada una de las cuales aborda un aspecto específico de la relación entre variables. A continuación, se desarrollan sus definiciones, fundamentos y aplicaciones.

Kappa de Cohen

El Kappa de Cohen es una métrica que evalúa la concordancia entre dos observadores o clasificadores, corrigiendo el acuerdo esperado por azar. Fue introducida por Cohen (1960) como una forma de mejorar la interpretación de la concordancia cuando las categorías tienen una distribución desigual (30).

Matemáticamente, se define como:

$$k = \frac{P_o - P_e}{1 - P_e}$$

Donde:

- P_o : Proporción de acuerdo observado.
- P_e : Proporción de acuerdo esperado por azar.

Un valor de $k = 1$ indica una concordancia perfecta, mientras que $k = 0$ implica que la concordancia no es mejor que la esperada por azar. Valores negativos sugieren una discordancia sistemática. Esta métrica es ampliamente utilizada en ciencias sociales, medicina y aprendizaje automático (31).

Información Mutua Ajustada

La Información Mutua Ajustada (AMI, por sus siglas en inglés) es una métrica que mide la similitud entre dos particiones, ajustada para tener en cuenta el acuerdo esperado por azar. Se basa en el concepto de Información Mutua (MI), que evalúa cuánta información sobre una variable se puede obtener al conocer otra. La AMI corrige la MI bruta para evitar sesgos en conjuntos de datos con números de clases desbalanceados (32).

Se define como:

$$AMI(U, V) = \frac{MI(U, V) - \mathbb{E}[MI(U, V)]}{\max(H(U), H(V)) - \mathbb{E}[MI(U, V)]}$$

Donde:

- $MI(U, V)$: Información mutua entre las particiones U y V.
- $H(U)$: Entropía de la partición U.
- $\mathbb{E}[MI(U, V)]$: Valor esperado de la información mutua por azar.

La AMI es una métrica común en análisis de *clustering*, ya que permite comparar particiones de datos independientemente del número de clases o tamaños de los *clusters* (33).

4.- PROBLEMA

Las pautas que establecen la clasificación tradicional del estudio del sueño han sufrido escasas revisiones desde su establecimiento en 1968 (34), de hecho, las características de la PSG utilizadas para etapificar el sueño fueron seleccionadas por su facilidad de identificación visual, sin aplicar métodos avanzados de análisis de datos que validaran su relevancia de manera objetiva (29). En efecto, 26 de las 29 reglas de etapificación establecidas por la AASM se determinaron mediante consenso entre evaluadores humanos y se propusieron sin una validación completa de su relevancia biológica para el sueño (35). Estas pautas en la actualidad se siguen sustentado en la inspección visual de las lecturas electrofisiológicas y se limitan a capturar una gama reducida de características, solo aquellas perceptibles por el ojo humano. Por otra parte, la contribución que realizan los métodos de análisis modernos a esta clasificación es mínima, infiriendo con ello la gran cantidad de información no utilizada que podría servir para caracterizar más profundamente la dinámica del sueño mediante la sofisticada gama de algoritmos de análisis de series temporales que actualmente existen (29). Por tal motivo, se considera una debilidad en las reglas de etapificación actualmente vigentes, la exclusión de técnicas actualizadas de procesamiento analítico de las señales neurofisiológicas, que podrían incorporar elementos diagnósticos complementarios.

Hay que reconocer que el esfuerzo por incorporar técnicas analíticas modernas en el estudio del ciclo sueño-vigilia, con el objetivo de lograr una etapificación automática, ha sido considerable. Sin embargo, estos intentos no han dado frutos debido a la alta variabilidad entre evaluadores (acuerdo de alrededor del 70-80%) y el bajo acuerdo intraevaluador (alrededor del 90%) (36). Uno de los principales motivos por el cual se ha incursionado en esta área, ha sido la gran cantidad de tiempo utilizado para etapificar las señales, además de ser una tarea tediosa y propensa al error humano. Es con ello que estos esfuerzos se han depositado principalmente en el desarrollo de algoritmos de aprendizaje supervisado intentado emular y mejorar la capacidad humana para realizar etapificaciones, pudiendo evaluar esto en dos extensas revisiones bibliográficas (6, 36). Con respecto a aprendizaje no supervisado, el principal enfoque que se le da es la selección de características y la reducción de dimensionalidad utilizando técnicas como *Principal Components Analysis* (PCA), esto para luego utilizarlas en métodos de clasificación supervisada. Otro enfoque no supervisado es el estudio del agrupamiento natural de los datos registrados en la lectura del sueño, mediante técnicas de *clustering*. El uso de *clustering* para la creación de modelos de

predicción no es muy común en esta área del conocimiento (6), sin embargo, hay ciertos estudios que demuestran la trascendencia que esta técnica pudiese tener en el entendimiento más profundo de la arquitectura y dinámica del sueño (29).

Este estudio será hecho con un enfoque impulsado por datos (*data-driven*) sacando ventajas de estrategias automatizadas en el procesamiento del alto volumen de datos. Esta automatización no solo garantiza celeridad en el proceso, sino que también contribuye a minimizar los errores (por ejemplo, errores en la identificación de patrones ocultos, errores en la reproducibilidad del análisis, errores de inconsistencia en la categorización, entre otros), una hazaña prácticamente inalcanzable si se abordara de forma manual. Además, este enfoque no solo permite extraer información clave para la clasificación de los estados de sueño, sino que también lo hace sin introducir sesgos humanos. Al aplicarlo, se abre la posibilidad de establecer una referencia objetiva del ciclo sueño-vigilia basada únicamente en las características intrínsecas de los datos.

5.- HIPÓTESIS

El análisis no supervisado del EEG del sueño, utilizando técnicas de agrupamiento (*clustering*), permite obtener una segmentación temporal de los estados del sueño que muestra una alta concordancia estadística con la etapificación visual polisomnográfica realizada por expertos.

6.- OBJETIVOS

6.1.- Objetivo general

Obtener segmentaciones estables de señales del EEG del sueño utilizando técnicas no supervisadas de agrupamiento (*clustering*) y evaluar su correspondencia con la etapificación visual polisomnográfica realizada por expertos, aplicando criterios para reducir la fragmentación.

6.2.- Objetivos específicos

1. Seleccionar una batería métodos para extraer características de las señales de EEG.
2. Determinar las características más relevantes de la señal de EEG para la construcción del vector de características.
3. Refinar algorítmicamente los agrupamientos para evitar fragmentación exagerada de la segmentación no supervisada.
4. Evaluar correspondencia entre resultados de métodos de análisis no supervisado con la etapificación de experto basada en las reglas de la AASM.

7.- MATERIAL Y MÉTODOS

7.1.- Material

La elección del conjunto de datos adecuado desempeña un papel fundamental en la calidad y validez de los resultados obtenidos. Se utilizaron datos de pacientes provenientes de la base de datos ISRUC-Sleep y del proyecto FONDECYT 1231901 (*LONG COVID IN PATIENTS WITH CARDIOVASCULAR RISK FACTORS*). Específicamente se utilizaron solo las señales de EEG, y de ellas solo electrodos centrales (C4-M1 del sistema 10-20).

ISRUC-Sleep es una base de datos pública de polisomnografía (PSG) diseñada para apoyar la investigación en el campo del sueño. Esta base de datos fue desarrollada por el Instituto de Sistemas y Robótica de la Universidad de Coimbra (ISR-UC) y proporciona una colección comprensiva de registros de sueño obtenidos de adultos, incluyendo tanto sujetos sanos como sujetos con trastornos del sueño, algunos de los cuales estaban bajo el efecto de medicación para el sueño. El proyecto fue aprobado por el *Sleep Medicine Centre, The Central Hospital of University of Coimbra* (CHUC), Portugal (41). Se utilizaron datos de 10 pacientes sin trastorno de sueño. La selección de la Base de Datos ISRUC-Sleep se hizo en base a los siguientes criterios:

Calidad de los Datos:

Resolución Temporal y Espacial: La base de datos ISRUC-Sleep ofrece grabaciones con una adecuada resolución temporal y espacial, cruciales para el análisis detallado del EEG, segmentación a 30 s con un muestreo de 200 Hz y la utilización de 8 electrodos para el estudio de EEG.

Puntuación Visual por Expertos: Las grabaciones PSG han sido puntuadas visualmente por dos expertos humanos, lo que asegura una referencia de alta calidad para validar nuestros algoritmos.

Confiabilidad y Validez:

Revisada por Pares: La base de datos ha sido utilizada y referenciada en múltiples publicaciones, lo que respalda su confiabilidad y validez en la comunidad científica.

Consistencia en las Grabaciones: Las grabaciones se realizaron bajo condiciones controladas en un entorno clínico y/o en un laboratorio del sueño, lo que minimiza las variaciones no deseadas en los datos.

Accesibilidad y Documentación:

Acceso Público y Transparente: La base de datos está disponible públicamente para la comunidad investigadora, lo que facilita la reproducibilidad de los estudios.

Documentación Completa: ISRUC-Sleep proporciona una documentación detallada sobre los protocolos de adquisición de datos, las características de los sujetos y las metodologías de puntuación, permitiendo una comprensión completa de los datos.

El proyecto FONDECYT 1231901, investiga los efectos a largo plazo del contagio por COVID-19 con un enfoque particular en las alteraciones que podrían manifestarse en el ciclo sueño-vigilia. El estudio analiza datos del ciclo sueño-vigilia de pacientes que reportan síndrome de COVID prolongado (*Long-COVID*), utilizándose seis de estos pacientes cuyo registro PSG exhibe una arquitectura compatible con la ausencia de trastornos del sueño. En total, para este estudio se utilizaron 16 pacientes sin trastorno de sueño.

7.2.- Método

El diseño metodológico de este estudio no se construyó a partir de decisiones predefinidas, sino que fue desarrollándose a través de una serie de evaluaciones guiadas directamente por los objetivos específicos. A diferencia de metodologías en las que las técnicas se definen completamente antes del análisis, en este caso los objetivos específicos 1, 2 y 3 requerían seleccionar métodos y características como parte del proceso analítico mismo, no como prerequisites previos. Por lo tanto, el enfoque adoptado es comparativo y fundamentado en evidencia. Por ello, decisiones como el tipo de extracción espectral, las características utilizadas o el algoritmo de agrupamiento no fueron asumidas de antemano, sino que surgieron de un proceso evaluativo guiado por criterios empíricos, técnicos y bibliográficos. En consecuencia, estos elementos se presentan en la sección de Resultados como hallazgos del análisis, y no como simples decisiones metodológicas previas.

7.2.1.-Objetivo específico 1

Seleccionar una batería métodos para extraer características de las señales de EEG.

Extracción de características primarias

Dominio de extracción

Se eligió utilizar un método para extraer características desde el dominio de tiempo-frecuencia debido a que las señales EEG son intrínsecamente no estacionarias. Los métodos tiempo-frecuencia permiten analizar esta no estacionariedad de forma efectiva, capturando tanto las características temporales como frecuenciales. También permiten analizar las características en eventos transitorios. La información combinada que proveen estos métodos es crucial para entender la dinámica de la señal y las interacciones entre diferentes componentes de frecuencia a lo largo del tiempo.

Batería de métodos

Los métodos más utilizados para la extracción de características en la clasificación automática de etapas del sueño en el dominio tiempo-frecuencia según evidencia (6) son Wavelets, STFT, Descomposición Empírica en Modos - Transformada Hilbert-Huang (EMD-HHT), Wigner-Ville y Choi-Williams. La Tabla I presenta un análisis comparativo de estos métodos en función de las variables consideradas para la correspondiente selección de los métodos en cuestión.

	Wavelets	STFT	EMD-HHT	Wigner - Ville	Choi-Williams
Porcentaje de uso en el dominio	60	10	17.5	10	2.5
Experiencia previa del investigador en la utilización del método	NO	SI	NO	NO	NO

Tabla I. Comparación de métodos según criterios de selección aplicados en este estudio.

En función del método con mayor porcentaje de uso en el dominio seleccionado, y considerando el conocimiento previo del investigador sobre las técnicas disponibles, se aplicó un primer filtro de selección. Este permitió enfocar el análisis en dos métodos: STFT y Wavelets. Para determinar cuál de ellos sería utilizado como primer método de extracción de características primarias, se aplicó un segundo y definitivo filtro, basado en el análisis de

distintos criterios y atributos relevantes de ambas técnicas. La Tabla II presenta esta comparación, sustentada en las referencias 37, 38, 39 y 40.

Todo esto condujo a la segmentación de la señal utilizando STFT con ventanas multitaper de 2 s con un traslape de 1 s.

Criterio	STFT	Wavelets	Comentario
Resolución adaptable	✗	✓	STFT ofrece una resolución temporal y frecuencia fija.
Implementación simple	✓	✗	STFT es más directa y rápida de programar.
Bajo costo computacional	✓	✗	STFT con FFT es más eficiente.
Visualización detallada	✗	✓	El espectrograma dado por STFT proporciona una visualización clara de cómo varían las frecuencias a lo largo del tiempo.
Base teórica sólida	✓	✓	Ambas están bien respaldadas, aunque wavelets son más recientes.

Tabla II. Criterios para la elección de características primarias.

Extracción de características secundarias

Una segunda estrategia de extracción de características se definió a partir de un enfoque exploratorio y comparativo, luego de aplicar técnicas clásicas de segmentación y análisis espectral sobre ventanas de 2 s con un traslape de 1 s. Este tipo de segmentación generó una secuencia con alta variabilidad entre segmentos consecutivos, lo cual dificultaba establecer correlaciones estadísticamente significativas con la etapificación tradicional del sueño. Por ello, la búsqueda de un proceso capaz de identificar “estados” (concepto que hace alusión al convencional estudio que se hace del ciclo sueño vigilia mediante “estados de sueño”, en donde la palabra “estado” implica algo constante en el tiempo como mínimo durante 30 segundos) permitiría abordar esta problemática a una escala en la que sería posible identificar configuraciones estables caracterizadas por una variabilidad dinámica recurrente.

Se exploraron métodos que permitieran caracterizar cada segmento no solo por sus propiedades individuales, sino también en función de su entorno temporal inmediato. Para ello, se desarrolló un proceso destinado a construir características que reflejaran relaciones de proximidad dinámica. Este procedimiento incluyó el uso de un algoritmo de *clustering* para discretizar el espacio espectral extraído (obtenido con el método de extracción de características primarias), y una función (Anexo 8) en lenguaje Python diseñada para estimar la probabilidad de ocurrencia de cada categoría dentro de una vecindad temporal estrecha. La elección del algoritmo de *clustering* se fundamentó exclusivamente en su capacidad para realizar una partición simple del espacio espectral, sin imponer suposiciones adicionales sobre la estructura de los datos. En este contexto, se optó por un algoritmo que cumpliera con criterios básicos de simplicidad, eficiencia e interpretabilidad. La selección del algoritmo *K-means* se realizó en base a un análisis de la literatura que se resume en la Tabla III. La Tabla IV muestra la justificación del criterio de puntuación.

Criterio / Algoritmo	<i>K-means</i>	<i>DBSCAN</i>	<i>GMM</i>	<i>Jerárquico</i>	<i>Spectral Clustering</i>
Más utilizado según evidencia	✓✓✓✓ [49][50][51]	✓✓ [50]	✓✓✓ [49]	✓✓✓ [50]	✓✓ [50]
Simplicidad / facilidad de uso	✓✓✓✓ [51]	✓✓✓ [50]	✓✓ [49]	✓✓ [49]	✓✓ [50]
Eficiencia computacional	✓✓✓✓ [51]	✓✓ [50]	✓✓ [49]	✓ [49]	✓ [50]
Interpretabilidad	✓✓✓✓ [50][51]	✓✓ [50]	✓✓ [49]	✓✓✓✓ [49]	✓✓ [50]
Soporte en bibliotecas	✓✓✓✓ [49][50][51]	✓✓✓✓ [50]	✓✓✓✓ [49]	✓✓✓✓ [49]	✓✓✓✓ [50]

Tabla III. Comparación de algoritmos de *clustering*.

Criterio	Parámetro de evaluación	Escala de ✓
Más utilizado según evidencia	Número de veces citado, uso en la industria, aparición en revisiones sistemáticas.	✓ = poco utilizado ✓✓✓✓ = uso dominante o por defecto
Simplicidad / facilidad de uso	Número de parámetros, complejidad matemática, facilidad de implementación.	✓ = difícil y técnico ✓✓✓✓ = muy fácil de usar
Eficiencia computacional	Complejidad algorítmica, escalabilidad, consumo de memoria.	✓ = muy costoso ✓✓✓✓ = altamente eficiente
Interpretabilidad	Facilidad de explicar los resultados, representación visual simple.	✓ = difícil de explicar ✓✓✓✓ = altamente intuitivo
Soporte en bibliotecas	Disponibilidad en bibliotecas estándar (scikit-learn, R, MATLAB), documentación.	✓ = escaso soporte ✓✓✓✓ = soporte amplio y universal

Tabla IV. Justificación de criterios de puntuación.

La construcción de una función para estimar la probabilidad de ocurrencia requirió identificar la duración de ventana temporal óptima para la segmentación dinámica de señales EEG, mediante un enfoque sistemático que evalúa y compara el rendimiento de cuatro duraciones de ventana: 3, 5, 7 y 9 segundos. Para ello, se implementó una estrategia cuantitativa orientada a maximizar la correspondencia con la etapificación tradicional del sueño, minimizando al mismo tiempo la variabilidad del desempeño entre pacientes.

La elección de la mejor ventana temporal se fundamentó en los siguientes tres criterios:

- **Concordancia promedio con la etapificación tradicional:** Se calcularon las métricas de AMI y Kappa de Cohen para cada paciente y ventana temporal. Los valores se promediaron entre pacientes para cada configuración, con el fin de estimar su capacidad de representar fielmente las transiciones de estados del sueño.
- **Estabilidad intersujeto:** Para cada ventana, se estimó la varianza de los valores de AMI y Kappa entre pacientes, lo que permite evaluar la consistencia del rendimiento de cada configuración temporal a nivel poblacional.

- **Compromiso entre concordancia y estabilidad:** Se analizó conjuntamente la magnitud de las métricas y su dispersión para identificar la configuración que ofreciera el equilibrio más favorable entre precisión y robustez metodológica.

Este análisis permitió observar no solo qué ventana entrega los valores de concordancia más altos, sino también cuál mantiene una menor dispersión entre sujetos, permitiendo así inferir qué configuración entrega resultados más estables y generalizables.

Procedimiento estadístico:

- Para cada ventana y paciente, se almacenaron los valores de AMI y Kappa.
- Posteriormente, se agruparon los resultados por ventana temporal y se calcularon los estadísticos descriptivos (media, varianza).
- Finalmente, se realizó una clasificación cualitativa de la estabilidad de cada configuración en cuatro niveles: muy baja, baja, moderada y alta, para facilitar la interpretación práctica de las diferencias encontradas.

7.2.2.-Objetivo específico 2

Determinar las características más relevantes de la señal de EEG para la construcción del vector de características.

La selección de la característica primaria se fundamentó en distintos criterios luego de realizar un análisis de la literatura que se resume en la Tabla V. La Tabla VI muestra la justificación del criterio de puntuación.

Criterio / Característica	Espectrograma	Energía en bandas	Frecuencia dominante	Ancho de banda	Centroides espectrales	Variación temporal
Riqueza de información multiescala	✓✓✓✓ [52][53][54]	✓✓ [54][55]	✓ [56]	✓✓ [56]	✓✓ [57]	✓✓✓ [58]
Uso clínico / interpretabilidad en bioseñales	✓✓✓✓ [52][54]	✓✓✓ [54]	✓ [56]	✓ [56]	✓✓ [57]	✓✓ [58]
Resolución espectral-temporal completa	✓✓✓✓ [52][54]	✓ [54]	✓ [56]	✓ [56]	✓✓ [57]	✓✓✓ [58]
Visualización para análisis exploratorio	✓✓✓✓ [52][54]	✓✓ [54]	✓ [56]	✓ [56]	✓ [57]	✓✓ [58]

Tabla V. Evaluación de características EEG

Criterio	Parámetro evaluado	Escala de ✓
Riqueza de información multiescala	Capacidad del descriptor para reflejar la dinámica en múltiples escalas temporales y frecuencias	✓ = baja capacidad multiescala ✓✓✓✓ = máxima riqueza de contenido temporal/frecuencial
Uso clínico / interpretabilidad en bioseñales	Facilidad para ser interpretado por clínicos o especialistas en señales biomédicas	✓ = poco interpretativo en el contexto clínico ✓✓✓✓ = altamente interpretable por expertos médicos
Resolución espectral-temporal completa	Precisión en representar simultáneamente el contenido espectral y su evolución en el tiempo	✓ = pobre resolución conjunta tiempo-frecuencia ✓✓✓✓ = resolución densa y detallada
Visualización para análisis exploratorio	Grado en que la característica facilita representaciones visuales útiles para el análisis preliminar	✓ = pobre valor visual/exploratorio ✓✓✓✓ = facilita visualización clara e intuitiva

Tabla VI. Justificación de criterios de puntuación.

Utilizando el método secundario de extracción, se construyeron características que reflejaran relaciones de proximidad dinámica entregando información para identificar patrones estables a pesa de la variabilidad a escala temporal corta.

7.2.3.-Objetivo específico 3

Refinar algorítmicamente los agrupamientos para evitar fragmentación exagerada de la segmentación no supervisada.

Para abordar el problema de la fragmentación excesiva en la segmentación no supervisada del sueño, se implementó un proceso de refinamiento en dos etapas utilizando *K-means* y *Agglomerative Clustering*. El objetivo fue reducir el número de estados o etiquetas distintas obtenidos a partir de las extracciones de características, de modo que la segmentación final que se obtuviese fuera más comparables con la etapificación tradicional del sueño, la cual se basa solo en cinco estados bien definidos.

Primera etapa de refinamiento: Discretización del espacio espectral con *K-means*

Se aplicó el algoritmo *K-means* para realizar una primera agrupación sobre los vectores característicos extraídos del espectrograma (características primarias). Este procedimiento tuvo como propósito:

- Discretizar el espacio en función de la similitud espectral, utilizando la distancia euclidiana para agrupar segmentos de señal con características frecuenciales similares.
- Establecer una base sólida para el siguiente proceso de refinamiento, asegurando que las agrupaciones capturen la variabilidad espectral sin provocar una fragmentación excesiva de la señal.

Segunda etapa de refinamiento: Agrupamiento de características contextuales con *Agglomerative Clustering*

Tras la primera agrupación, y posterior a la construcción de las características secundarias, se aplicó *Agglomerative Clustering* sobre las características dinámicas de cada agrupación espectral, con el fin de estructurar la señal en estados más estables y representativos.

La medida de similitud utilizada fue la divergencia de Kullback-Leibler (KL) simetrizada, porque el vector característico con el que se trabajó es una distribución y la divergencia KL es un método estándar para comparar distribuciones. La simetrización de esta divergencia

fue deseable para cumplir con una condición básica de métrica. En este caso, la simetrización se realizó promediando ambas direcciones de la divergencia KL.

El uso de este algoritmo se justifica porque permite definir múltiples niveles de corte, lo que facilita la selección óptima mediante un criterio de optimización. En este caso, se empleó la entropía de la distribución de *clusters* para determinar el nivel de corte más adecuado en el *clustering* jerárquico, logrando reducir la fragmentación y preservar un alto nivel informativo. Con un nivel 3 de entropía, se redujo el número de etiquetas optimizando la segmentación sin perder la complejidad intrínseca de los datos. Estas nuevas agrupaciones fueron denominadas "Estados Dinámicos", ya que representan patrones temporales estructurados en la señal EEG, permitiendo su comparación con la etapificación tradicional del sueño.

7.2.4.-Objetivo específico 4

Evaluar correspondencia entre resultados de métodos de análisis no supervisado con la etapificación de experto basada en las reglas de la AASM.

Para comparar los resultados de los métodos de agrupamiento no supervisado con la etapificación tradicional del sueño basada en las reglas de la AASM, se realizaron los siguientes pasos:

1. Cálculo de métricas de comparación:
 - Información Mutua Ajustada (AMI): Para evaluar la similitud estructural entre las agrupaciones no supervisadas (estados dinámicos) y la etapificación tradicional a escala de 1 s.
 - Kappa de Cohen: Para estimar el nivel de concordancia entre ambas segmentaciones a escala de 1 segundo, se asignó a cada estado dinámico la etapa de sueño tradicional con la que presentó mayor coincidencia temporal. Esta asignación se justificó en que los estados dinámicos no poseen una categoría predefinida, por lo que fue necesario vincularlos a la etapa convencional que predominaba durante su ocurrencia. A partir de esta correspondencia, se construyeron dos secuencias paralelas de etiquetas —una proveniente de la segmentación tradicional y otra de la segmentación dinámica reinterpretada—, ambas con resolución temporal de 1 segundo.

Sobre estas secuencias se aplicó el coeficiente kappa de Cohen, con el fin de cuantificar la concordancia global entre ambos esquemas de segmentación.

2. Cálculo de métricas:

- Se calculó varianza, media y desviación estándar de la AMI y Kappa de Cohen de todos los pacientes, asegurando la consistencia y generalización de los métodos implementados.

La Tabla VII muestra los rangos y la interpretación de las métricas.

Información Mutua Ajustada	
Valores	Interpretación
0	Independencia completa
1	Concordancia Perfecta
Kappa de Cohen	
Valores	Interpretación
< 0.0	Concordancia pobre
0.01 - 0.20	Concordancia ligera
0.21 - 0.40	Concordancia justa
0.41 - 0.60	Concordancia moderada
0.61 - 0.80	Concordancia sustancial
0.81 - 1.00	Concordancia perfecta

Tabla VII. Rango de valores e interpretación Kappa de Cohen

8.- RESULTADOS

8.1.- Objetivo específico 1

Seleccionar una batería métodos para extraer características de las señales de EEG.

Para cumplir este objetivo, se aplicó un procedimiento comparativo sobre múltiples métodos candidatos de extracción primaria, evaluando su aplicabilidad en función de criterios técnicos y bibliográficos (Tabla I y II en apartado 7.2.1).

La STFT fue seleccionada como el método de extracción de características primarias con mejor puntuación, dado su bajo costo computacional, compatibilidad con ventanas temporales fijas y alta capacidad de visualización mediante espectrogramas.

En cuanto al proceso de agrupamiento espectral (extracción secundaria), se evaluaron distintas técnicas de *clustering* (Tabla III en apartado 7.2.1), siendo *K-means* el algoritmo que obtuvo la mayor puntuación general, destacando por su simplicidad, escalabilidad y buen desempeño en la discretización de espacios espectrales.

Asimismo, se evaluaron diferentes tamaños de ventana temporal (3, 5, 7 y 9 s), considerando métricas de estabilidad interpaciente y concordancia con la etapificación tradicional en los 16 casos analizados en este estudio. Los resultados (Tablas VIII y IX) indicaron que:

- La ventana de 3 segundos presentó la menor varianza en ambas métricas (AMI = 0.00122, categorizada como "muy baja"; Kappa = 0.00517, "baja"), indicando una alta estabilidad.
- Por otro lado, la ventana de 9 segundos mostró el mayor valor promedio de AMI (0.2982) y un valor elevado de Kappa (0.4369), pero con una varianza moderada-alta en ambos casos.
- La ventana de 5 segundos obtuvo un valor de AMI de 0.2952 con varianza "baja", y un valor de Kappa de 0.4397 con varianza "moderada", lo que la posiciona como un punto de equilibrio entre concordancia y estabilidad siendo la elegida para proceder con el estudio.

En conjunto, este objetivo concluyó con la selección de STFT como método de extracción primaria, K-means como técnica de discretización, y una ventana de 5 segundos como la más adecuada para el análisis posterior.

		3 s	5 s	7 s	9 s	Promedio
Paciente_1	AMI	0.29042106	0.27779717	0.28753125	0.29996213	0.2889279
	k	0.4528111	0.46923831	0.45783949	0.48524663	0.46628389
Paciente_2	AMI	0.26126595	0.27208836	0.26415901	0.26451645	0.26550744
	k	0.4293188	0.42502919	0.44047349	0.42949714	0.43107966
Paciente_3	AMI	0.29002847	0.31484674	0.28541389	0.31341626	0.30092634
	k	0.50577525	0.45165905	0.39162709	0.50068121	0.46243565
Paciente_4	AMI	0.27221224	0.30327102	0.28850145	0.30602849	0.2925033
	k	0.42683682	0.31434421	0.31216693	0.34226965	0.3489044
Paciente_5	AMI	0.30097357	0.30675907	0.33032194	0.30209039	0.31003624
	k	0.57621484	0.53419929	0.53826116	0.5458283	0.5486259
Paciente_6	AMI	0.25996056	0.27247488	0.28184832	0.27722869	0.27287811
	k	0.33506706	0.43160975	0.35941802	0.35293863	0.36975836
Paciente_7	AMI	0.32452644	0.344137	0.35193738	0.34307151	0.34091808
	k	0.51020119	0.51645258	0.51872322	0.51714193	0.51562973
Paciente_8	AMI	0.35362424	0.38043528	0.38084943	0.38199813	0.37422677
	k	0.54024529	0.53717302	0.51450222	0.48521831	0.51928471
Paciente_9	AMI	0.23882011	0.24688582	0.2455514	0.25532577	0.24664577
	k	0.3819806	0.39500273	0.40330929	0.41210942	0.39810051
Paciente_10	AMI	0.30503528	0.30177811	0.30023368	0.30584318	0.30322256
	k	0.44994313	0.47019693	0.34539028	0.43225572	0.42444652
Paciente_11	AMI	0.22151097	0.25272259	0.23914977	0.23753213	0.23772886
	k	0.36750589	0.41175536	0.37581511	0.37994206	0.38375461
Paciente_12	AMI	0.28203909	0.2670728	0.2712405	0.26366987	0.27100556
	k	0.49790103	0.32329262	0.33446389	0.34029104	0.37398715
Paciente_13	AMI	0.26738434	0.26664165	0.27734526	0.29464025	0.27650287
	k	0.49833632	0.41319283	0.47024261	0.38557617	0.44183698
Paciente_14	AMI	0.30264856	0.32020889	0.32586165	0.33478025	0.32087484
	k	0.54126099	0.60485603	0.5905606	0.56538303	0.57551516
Paciente_15	AMI	0.24748935	0.26808751	0.25162757	0.27882186	0.26150657
	k	0.3546717	0.37839846	0.21019554	0.19452312	0.2844472
Paciente_16	AMI	0.23517043	0.25947228	0.24791103	0.25551294	0.24951667
	k	0.45798784	0.47510215	0.4147389	0.4539928	0.45045542

Tabla VIII. Resultados de AMI y Kappa por ventana temporal

Ventana	AMI (promedio)	AMI (varianza)	AMI (varianza)	Kappa (promedio)	Kappa (varianza)	Kappa (varianza)
3 s	0.2835	0.001220441	muy baja	0.4437	0.005169381	baja
5 s	0.2952	0.001315347	baja	0.4397	0.006049119	moderada
7 s	0.2923	0.001612039	alta	0.4256	0.009418658	alta
9 s	0.2982	0.001402511	moderada	0.4369	0.008849615	alta

Tabla IX. Concordancia promedio y variabilidad por ventana temporal.

8.2.- Objetivo específico 2

Determinar las características más relevantes de la señal de EEG para la construcción del vector de características.

Representación espectral mediante el espectrograma

Para dar respuesta a este objetivo, se compararon seis características extraídas del análisis espectral en función de cuatro criterios: riqueza multiescala, resolución espectro-temporal, visualización exploratoria e interpretabilidad clínica (Tabla V en apartado 7.2.2).

La evaluación arrojó que el espectrograma fue la característica con mejor puntuación general, destacando por su alta riqueza informativa y capacidad de representar de forma detallada las dinámicas espectrales del sueño. A diferencia de otras características derivadas de la STFT, el espectrograma preserva completamente la resolución espectral y temporal de la señal, lo que permite visualizar dinámicamente los cambios asociados a las distintas etapas del ciclo sueño-vigilia como se muestra en la Figura 5. Además, posteriormente, el uso del espectrograma permitió realizar análisis visuales entre las bandas de frecuencia y los estados dinámicos encontrados (Figura 6).

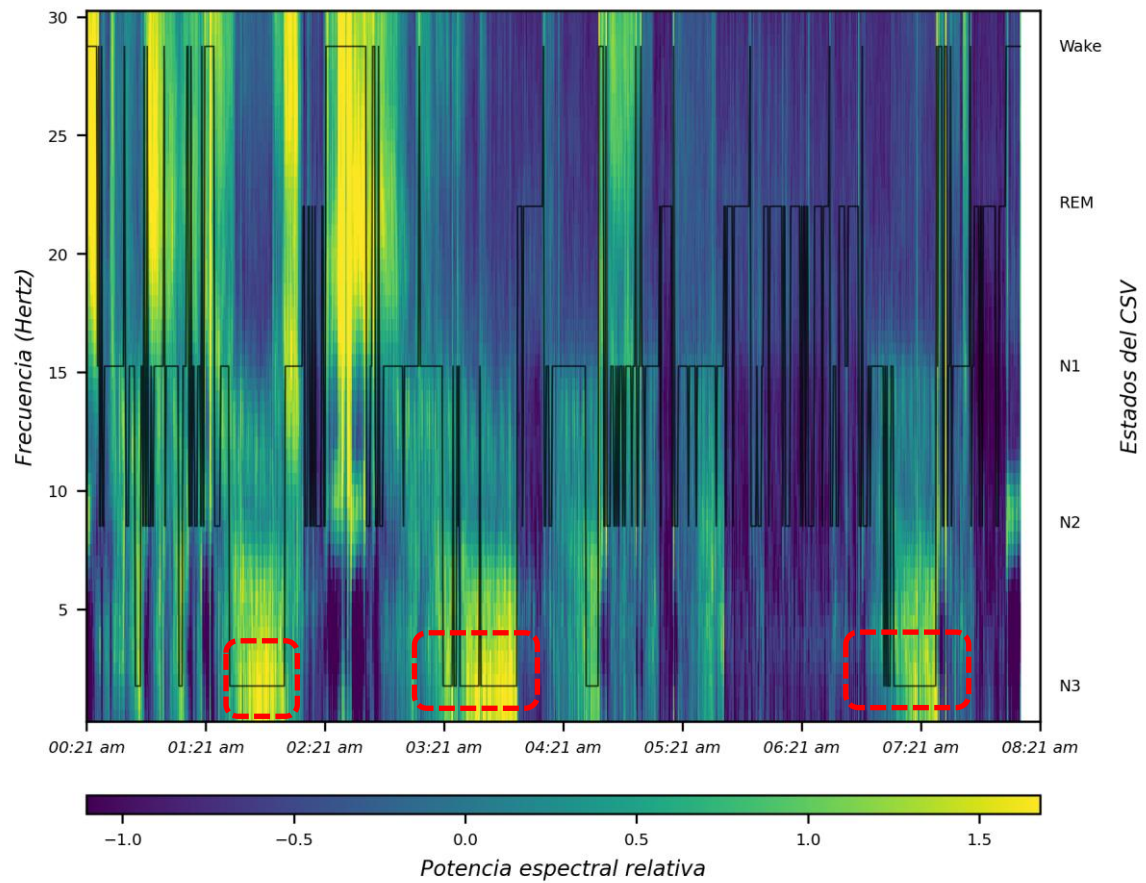


Figura 5. Visualización de la dinámica espectral versus el scoring de sueño. Se superpone el hipnograma (línea negra) sobre el espectrograma, ofreciendo una fácil lectura entre el estado de sueño que se desea analizar en función del rango de frecuencias asociado. Se aprecia claramente que cuando el hipnograma está en N3 (cuadrado rojo punteado), hay un aumento de intensidad en banda delta. También se observan patrones diferenciados para REM u otros estados.

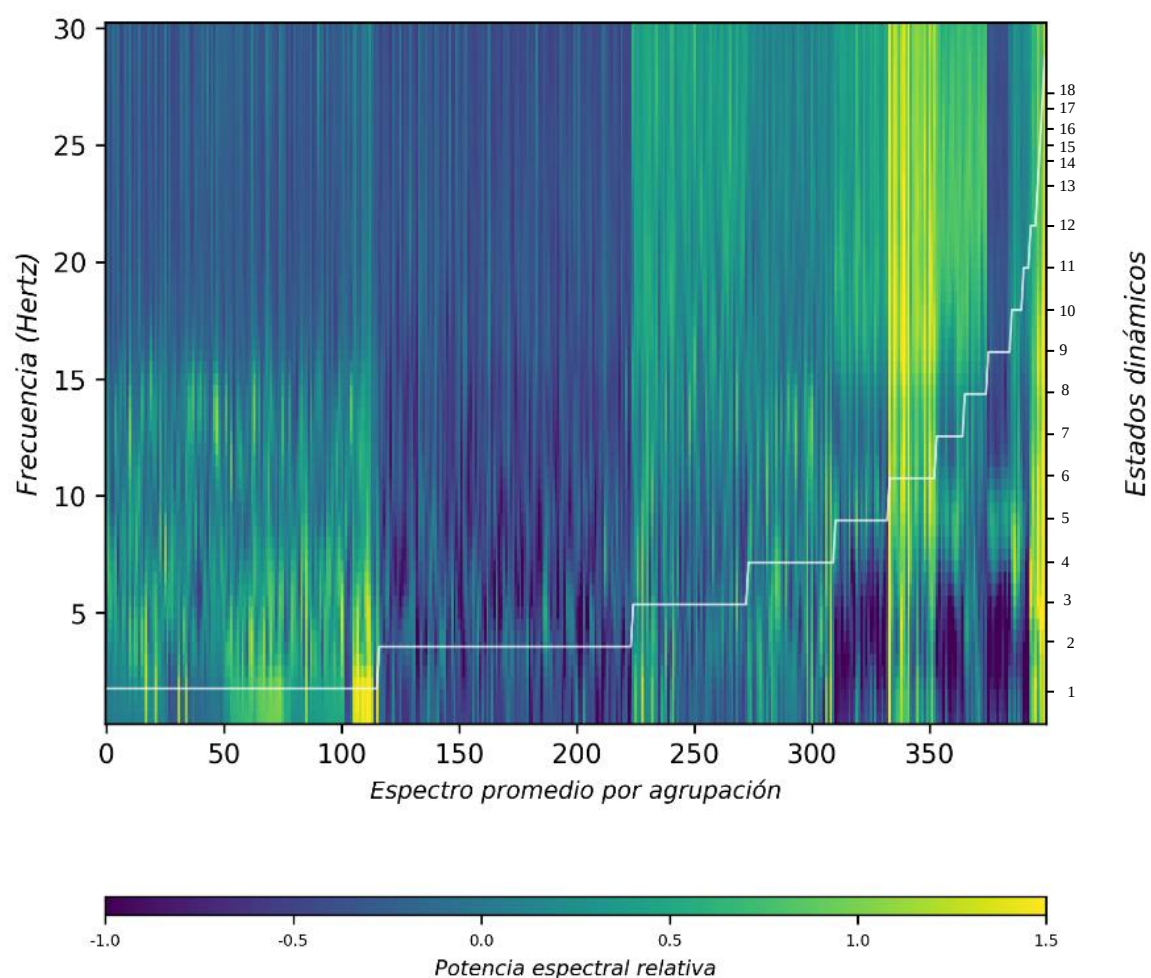


Figura 6. Características espectrales promedio de cada agrupación tras el primer refinamiento, ordenadas según su dinámica temporal. La figura permite visualizar simultáneamente la descripción espectral de las agrupaciones obtenidas tras el primer refinamiento, junto con la caracterización dinámica de cada una de ellas a partir de su agrupamiento según patrones de dinámica temporal. La visualización permite evaluar directamente las bandas de frecuencia que predominan por cada estado dinámico (cada escalón de la línea blanca superpuesta representa un estado dinámico diferente numerados de izquierda a derecha a partir del 1), de este modo se puede inferir qué estado de sueño tradicional corresponde a cada estado dinámico. El estado dinámico 1 podría corresponderse al estado tradicional N2 debido al rango de frecuencia que prevalece.

Para el desarrollo del espectrograma se utilizó un espectro multitaper en ventanas de 2 segundos con traslape de 1 segundo, obteniendo una muestra por segundo (asegurando una resolución temporal adecuada para capturar eventos fásicos sin perder información relevante sobre la estructura del sueño), con una resolución espectral de 0.5 Hz (que ofrece un buen detalle espectral) y un rango de frecuencias entre 0.5 y 30 Hz (considerando que frecuencias más altas suelen contener más ruido). El espectrograma fue normalizado para destacar

elementos informativos, utilizando la potencia espectral en escala logarítmica y aplicando una normalización a promedio 0 y varianza 1 (z-score) en cada canal espectral. Además, se aplicó una reducción dimensional mediante análisis de componentes principales (10 componentes de caracterización) para disminuir el ruido y acelerar el proceso (Anexo 4)

Representación dinámica mediante la matriz de transiciones

Para construir una característica que permitiese reflejar relaciones de proximidad dinámica entre agrupaciones espectrales, se construyó una matriz de transiciones (Figura 7), basada en la dinámica de ocurrencias entre los estados definidos por *K-means*. Esta matriz se convirtió en una nueva característica secundaria que reflejó la estabilidad y conectividad entre estados, y permitió definir un nuevo espacio de representación denominado “espacio dinámico”

En resumen, este objetivo permitió identificar al espectrograma como característica primaria óptima, y a la matriz de transiciones como característica secundaria.

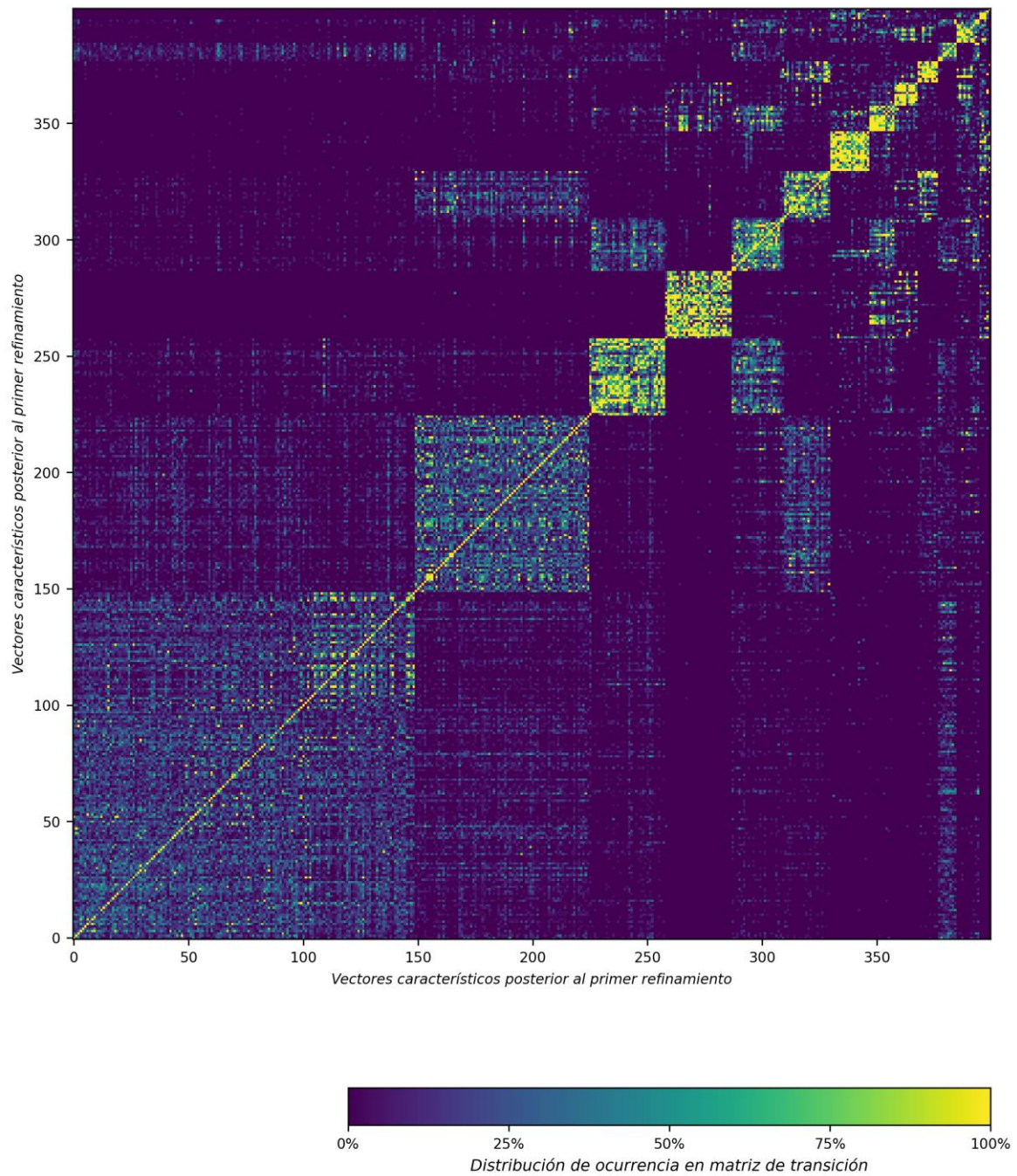


Figura 7. Matriz de características dinámicas (vector columna). Muestra la matriz de transiciones dinámicas de los vectores característicos ordenada en función del estado dinámico resultante del segundo proceso de agrupamiento. En esta matriz, cada celda representa la probabilidad de transición entre los agrupamientos del primer refinamiento. Los valores en la diagonal principal reflejan la probabilidad de que un agrupamiento permanezca constante a lo largo del tiempo. Los valores fuera de la diagonal muestran las transiciones entre agrupamientos diferentes. Una celda con un valor alto (más amarillo) sugiere que el agrupamiento fila tiende a aparecer frecuentemente en el contexto del agrupamiento columna correspondiente.

8.3.- Objetivo específico 3

Refinar algorítmicamente los agrupamientos para evitar fragmentación exagerada de la segmentación no supervisada.

La implementación del refinamiento en dos etapas permitió reducir en un 99.9% la fragmentación inicial de la señal pasando de alrededor de 25.000 etiquetas distintas a un rango de 14-24, logrando una representación con menor variabilidad y más comparable con la clasificación tradicional del sueño. A continuación, se detallan los principales resultados:

Reducción de etiquetas distintas: con *K-means*, se redujo la cantidad de etiquetas de alrededor de 25.000 a 400, asegurando una segmentación más manejable basada en la similitud espectral. Con *Agglomerative Clustering*, el número de etiquetas disminuyó aún más, alcanzando un rango entre 14 y 24 utilizando un nivel de entropía de 3 bits.

Comparabilidad con la etapificación tradicional: la disminución de la variabilidad inicial permitió mejorar la correspondencia entre la segmentación no supervisada y la tradicional. Pasó de tener un $AMI = 3.79 \times 10^{-12}$ a un $AMI \approx 0.2952$

La Figura 8 presenta y resume la cadena de procesos que se desarrollaron para completar los Objetivos Específicos 1, 2 y 3.

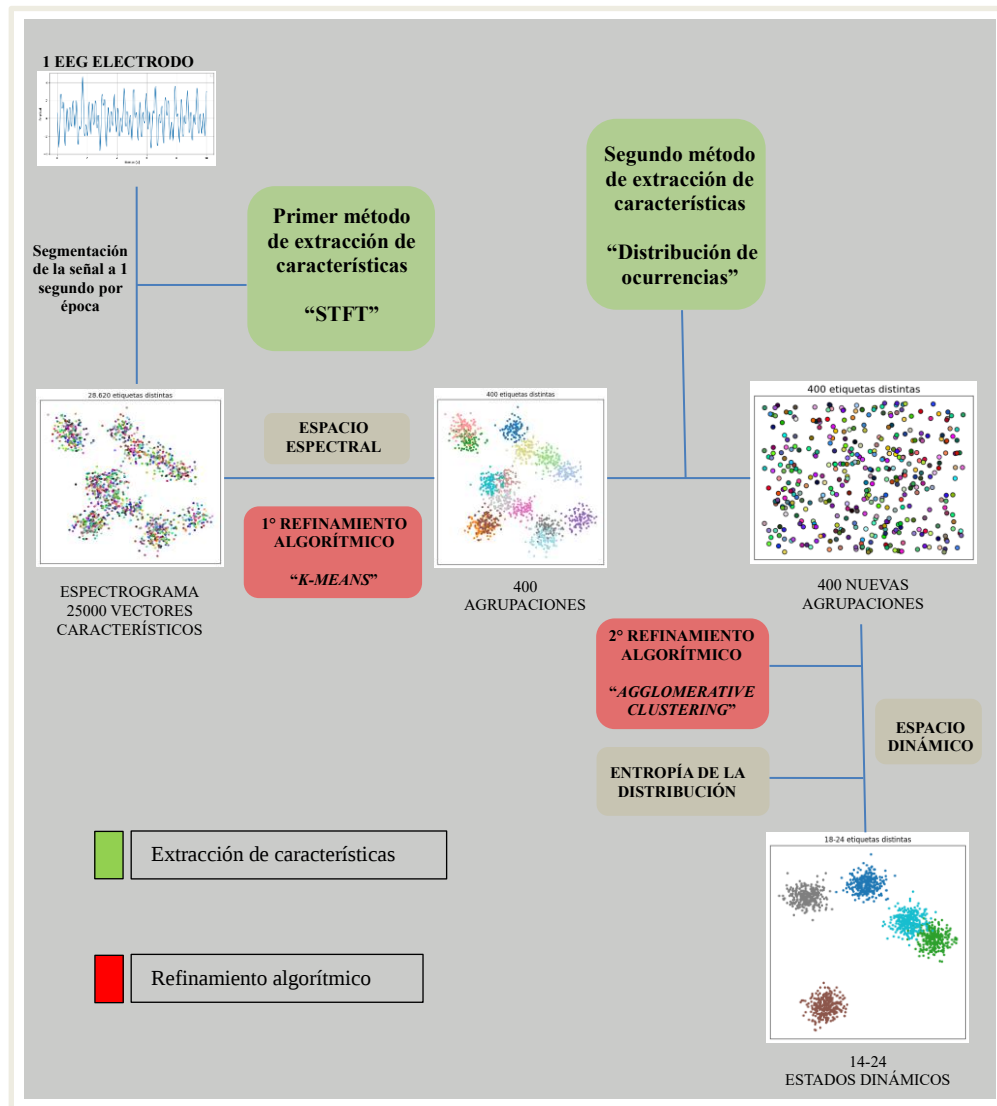


Figura 8. Esquema de la arquitectura del sistema de caracterización y refinamiento EEG. La imagen representa los dos procesos principales utilizados para alcanzar los objetivos del estudio: extracción de características y refinamiento algorítmico. En una primera etapa, la señal recibida desde un electrodo fue caracterizada mediante un análisis espectral (STFT), generando aproximadamente 25.000 vectores característicos que conforman el espacio espectral inicial. A partir de este espacio, se realizó un primer refinamiento utilizando el algoritmo *K-means*, reduciendo su dimensionalidad a 400 agrupaciones. Posteriormente, se aplicó un segundo método de extracción de características basado en la obtención de una distribución de ocurrencia en función de las agrupaciones obtenidas con *K-means*. Este procedimiento generó un nuevo conjunto de 400 vectores característicos que conforman un nuevo espacio denominado espacio dinámico. Finalmente, sobre este espacio se aplicó un segundo refinamiento mediante *clustering* aglomerativo, obteniendo entre 14 y 24 agrupaciones. El nivel de corte de este refinamiento fue ajustado utilizando la entropía de la distribución de ocurrencia.

8.4.- Objetivo específico 4

Evaluar correspondencia entre resultados de métodos de análisis no supervisado con la etapificación de experto basada en las reglas de la AASM.

La evaluación de correspondencia fue resuelta contabilizando la cantidad de información compartida entre la etapificación tradicional y la no supervisada mediante la métrica de Información Mutua Ajustada y evaluando el grado de acuerdo entre estas dos técnicas mediante la medida estadística Kappa de Cohen.

La Información Mutua Ajustada y el Kappa de Cohen para los 16 pacientes están detallados en la Figura 9 a y b. Tanto para los valores de Información Mutua Ajustada y Kappa de Cohen se realizó un análisis estadístico calculando la varianza, la media y la desviación estándar, estos resultados se presentan en la Tabla X.

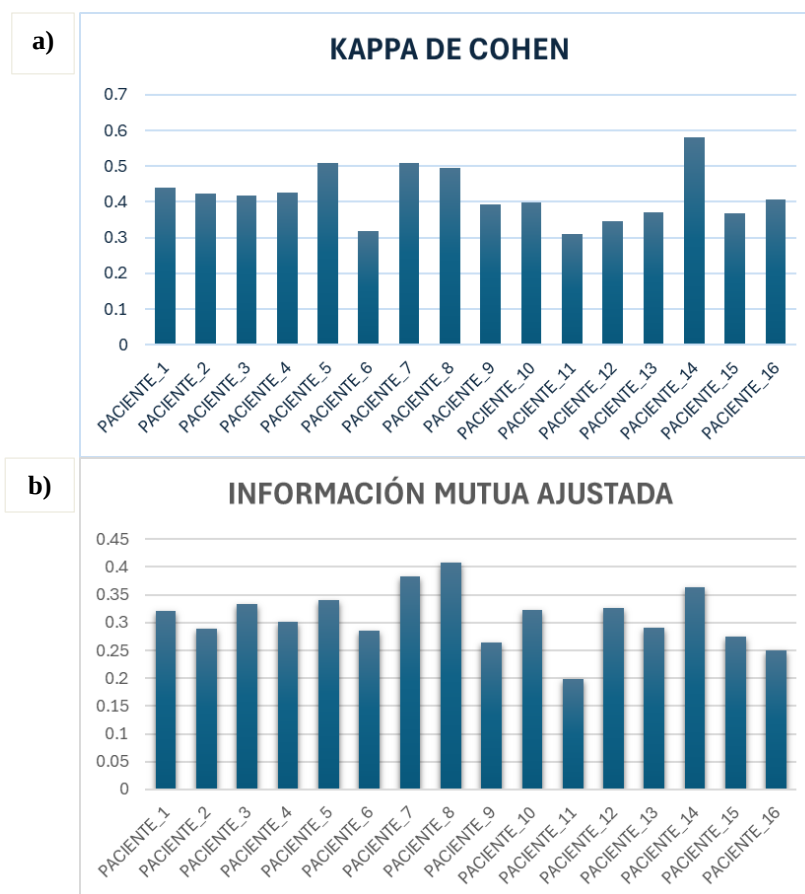


Figura 9. Concordancia entre etapificación tradicional y segmentación no supervisada.

- a) Coeficiente Kappa de Cohen calculado entre la etapificación tradicional del sueño y la segmentación no supervisada, para cada paciente analizado.
- b) Información Mutua Ajustada calculada entre la etapificación tradicional del sueño y la segmentación no supervisada, para cada paciente analizado.

	AMI	Kappa de Cohen
Varianza	0.002717931	0.005449767
Media	0.308621917	0.419342641
Desviación estándar	0.052133778	0.073822535

Tabla X. Cálculo de Kappa de Cohen e Información Mutua Ajustada. Con los datos de la Figura 9 se calculó estadísticos para Kappa de Cohen e Información Mutua Ajustada

Visualmente, esta correspondencia puede apreciarse en la Figura 10, la cual presenta el análisis a lo largo de toda una noche de sueño, permitiendo evidenciar la correspondencia que existe entre un estado dinámico y el estado de sueño tradicional por épocas de 30 s (para facilitar la comparación con la etapificación tradicional, se agruparon bloques de 30 segundos a partir de las etiquetas dinámicas de 1 segundo, calculando en cada bloque la proporción relativa de ocurrencia de cada estado dinámico). Para evidenciar esta coincidencia moderada con mayor detalle, la Figura 11 muestra un segmento específico de 45 minutos. Y para profundizar en la construcción de este análisis es posible revisar la Figura 12.

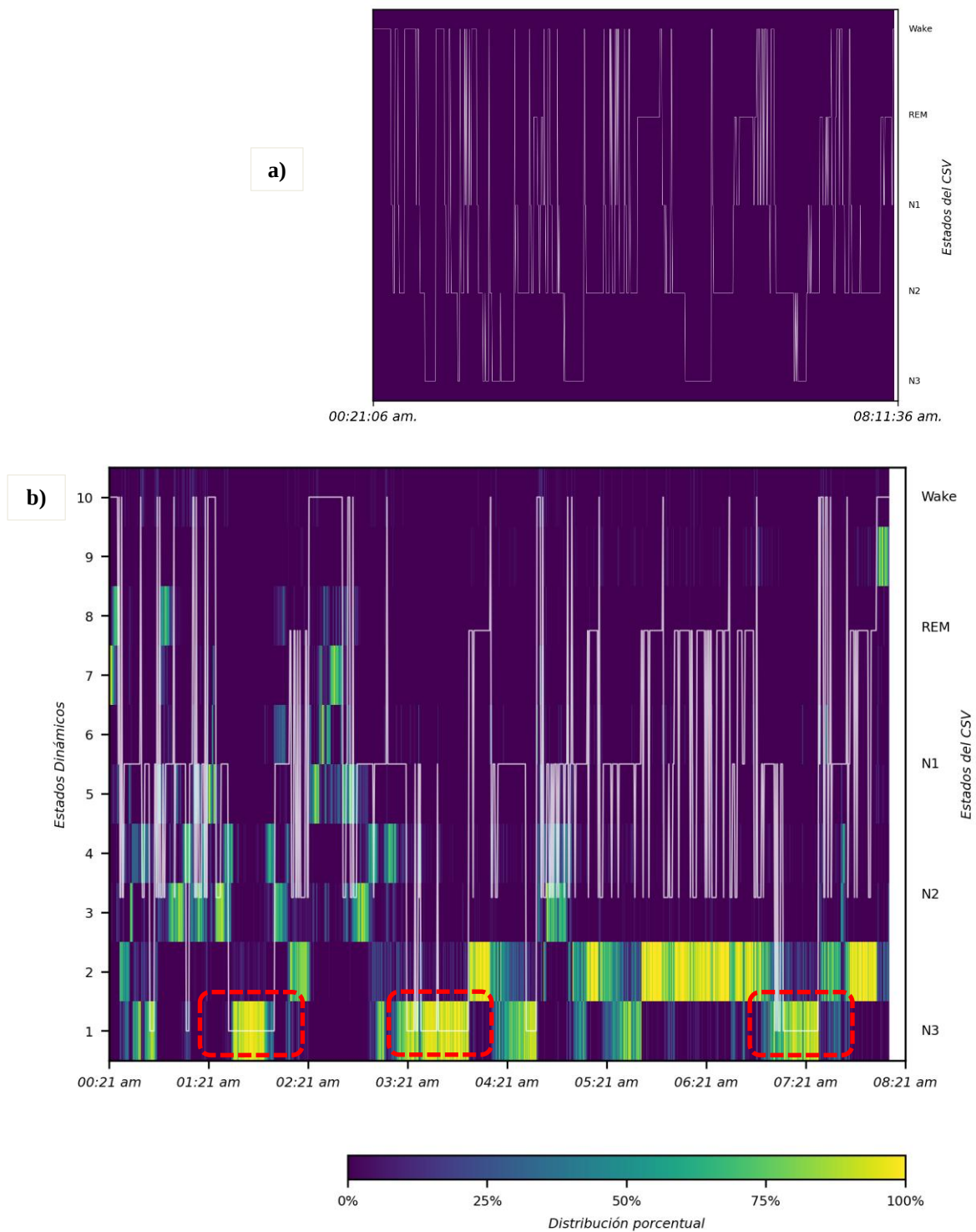


Figura 10. Evaluación de Correspondencia. **a)** Etapificación tradicional (hipnograma) del caso de estudio número 2. La ordenada derecha corresponde a los estados del CSV. **b)** Sobre la figura A, se superpone la etapificación no supervisada, presentando la activación y desactivación de los estados dinámicos de la ordenada izquierda. El color representa la distribución porcentual de los estados dinámicos en las épocas de 30 segundos de acuerdo con la barra de colores adyacente. Por ejemplo, cuando el hipnograma está en N3 (cuadrado rojo punteado), se activa principalmente el estado dinámico 1. Así mismo ocurre cuando el hipnograma está en REM, se activa principalmente el estado dinámico 2.

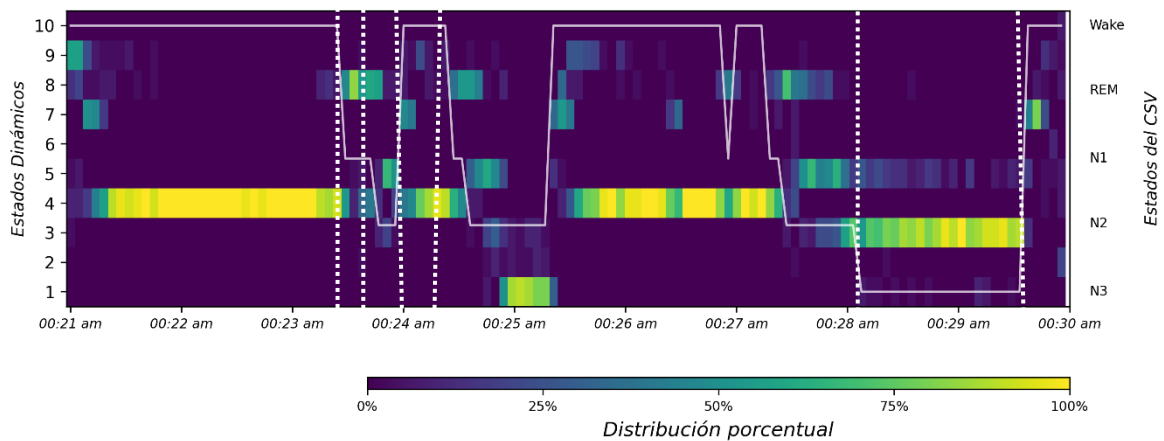


Figura 11. Detalle de la evaluación de correspondencia. La abscisa muestra los primeros 45 minutos de sueño del caso de estudio número 1. Se evidencia visualmente la coincidencia moderada entre las transiciones del hipnograma al cambiar de estado (línea vertical punteada), y las transiciones de estados dinámicos. En los límites temporales marcados por los cambios de estado en el hipnograma (línea punteada blanca), se observan cambios abruptos en la activación y desactivación de los estados dinámicos.

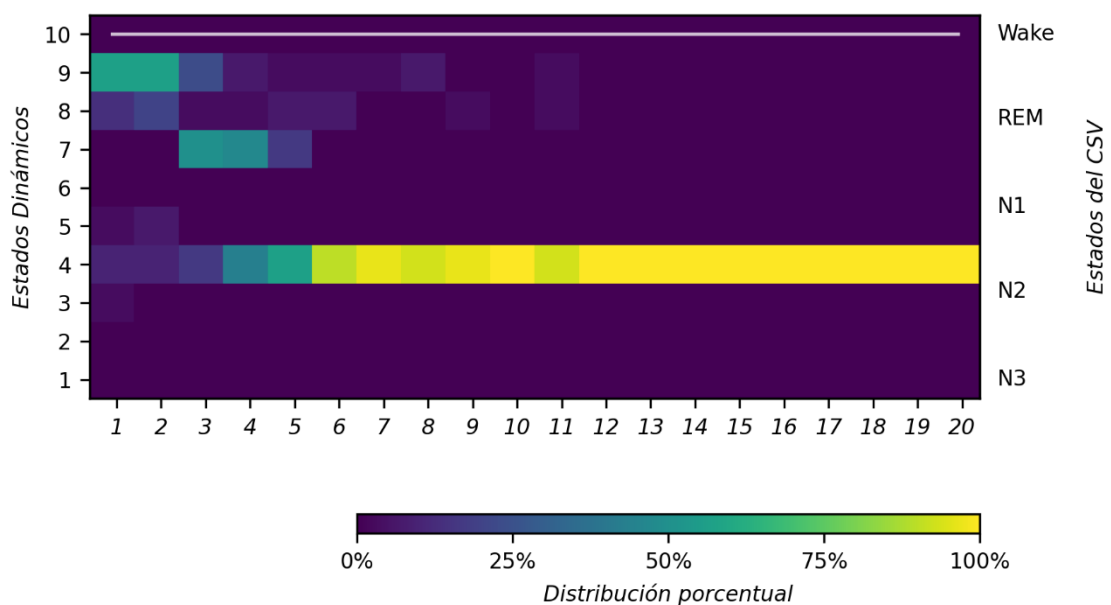


Figura 12. Descomposición de estados dinámicos por época tradicional. Esta figura ilustra las primeras 20 épocas de 30 segundos presentadas en la Figura 11 (00:21:06 am – 00:31:06 am). En ella se muestra la descomposición de cada época en sus respectivos estados dinámicos, los cuales presentan una distribución de ocurrencia particular en cada caso. La época 20 revela una predominancia del estado dinámico 4 a lo largo de sus 30 sub-épocas de 1 segundo, lo que indica una alta estabilidad de ese estado en dicho intervalo. En contraste, la época 3 exhibe una distribución más heterogénea, con una mayor presencia de los estados dinámicos 4 y 7, pero también con una participación menor de los estados 8 y 9.

9.- DISCUSIÓN

Los resultados de esta investigación muestran que es posible utilizar métodos de análisis no supervisado para segmentar señales EEG de sueño y obtener agrupaciones temporales que presentan un nivel de concordancia estadísticamente significativa —aunque no alta— con la etapificación visual tradicional establecida por la AASM. Este hallazgo proporciona evidencia empírica del potencial que tienen las estrategias de segmentación basadas en datos para capturar, de forma autónoma, la dinámica cerebral durante el sueño, abriendo paso a nuevos enfoques que podrían complementar o reevaluar las categorías actuales del ciclo sueño-vigilia.

La concordancia observada, con valores de $AMI \approx 0.2952$ y $Kappa \approx 0.4397$, se considera moderada y consistente con estudios que muestran una variabilidad significativa incluso entre expertos humanos. Estudios han reportado niveles de acuerdo interevaluador que rara vez superan el 85%, mientras que el acuerdo intraevaluador se sitúa entre un 80% y 90%, dependiendo de la etapa evaluada (59). Estos márgenes de error han sido reiteradamente observados en la literatura, y reflejan la complejidad subjetiva del *scoring* tradicional.

Cabe destacar que la mayoría de los estudios recientes en análisis automático del sueño se han basado en modelos supervisados (6, 36), e incluso varios de ellos han reportado concordancias sustanciales, con valores de *accuracy* y Kappa propios del rango alto, utilizando únicamente un canal de EEG (65, 66, 67, 68, 69). Naturalmente, estos enfoques tienden a reportar niveles de concordancia más elevados en comparación con los métodos no supervisados (29, 70, 71, 72, 73), lo cual es esperable dada su dependencia de etiquetas humanas durante el entrenamiento. Sin embargo, los resultados obtenidos en esta investigación se encuentran en un rango comparable al de varios trabajos no supervisados. Por ejemplo, algunos estudios reportaron niveles de acuerdo moderado-alto con coeficientes Kappa cercanos a 0.6 (70), otros alcanzaron valores entre 0.6 y 0.7 (71), y otro informó un Kappa de aproximadamente 0.51 (72). En el caso de Decat et al. (2022), aunque no se utilizó explícitamente el coeficiente de Kappa, el análisis de superposición entre *clusters* y etapas AASM mediante matrices de confusión permitió evidenciar una correspondencia moderada en etapas como N3, a pesar de discrepancias importantes en otras fases del sueño.

En este contexto, resulta especialmente relevante destacar que el método desarrollado se basa exclusivamente en un único canal de EEG, lo cual representa una simplificación sustantiva respecto a la polisomnografía tradicional, que requiere múltiples señales

complementarias como los movimientos oculares (EOG) y el tono muscular (EMG) para diferenciar los estados conductuales del sueño. A pesar de esta simplificación, el enfoque propuesto logra identificar y caracterizar estados cerebrales distintos mediante patrones dinámicos emergentes en la propia actividad eléctrica cortical, lo que no solo demuestra la robustez del análisis orientado por datos, sino que también sugiere la posibilidad de capturar elementos fisiológicamente relevantes que trascienden la estructura taxonómica clásica del sueño

El enfoque propuesto en esta investigación representa un cambio de paradigma respecto a la clasificación tradicional del sueño. A diferencia de metodologías que asumen desde el inicio un modelo predefinido de cinco etapas —por ejemplo, forzando cinco agrupaciones y luego evaluando su correspondencia con la clasificación estándar (como ocurre en las referencias 29, 70, 71, 72 y 73)—, esta propuesta adopta una perspectiva guiada por los datos. En lugar de imponer una estructura taxonómica fija, permite que los patrones dinámicos emerjan de manera no supervisada, capturando así una mayor riqueza en la variabilidad temporal de la señal. Esta idea encuentra respaldo en estudios que han observado cómo la dinámica cerebral durante el sueño presenta transiciones graduales y cambios continuos en la conectividad, lo cual desafía implícitamente la rigidez de las categorías tradicionales (60, 61). Bajo esta lógica, los estados dinámicos identificados en esta investigación no necesariamente equivalen a las etapas tradicionales, sino que representan configuraciones funcionales del cerebro identificadas a través de la variabilidad espectral y contextual de la señal. La segmentación resultante puede entenderse como una forma de autoestructuración de la señal en función de sus propiedades frecuenciales y contextuales, reflejadas en dinámicas temporales específicas.

Desde el punto de vista neurofisiológico, los estados dinámicos identificados en este estudio podrían estar capturando transiciones microestructurales, microdespertares o configuraciones funcionales transitorias que no son fácilmente detectables mediante la segmentación clásica en ventanas de 30 segundos. Esta perspectiva encuentra respaldo en estudios que han demostrado que las transiciones entre estados de vigilia y sueño están precedidas por cambios graduales en la actividad cerebral, lo que sugiere que dichos estados no son discretos, sino parte de un continuo dinámico (60). De forma complementaria, se ha evidenciado mediante modelamiento computacional del cerebro completo que la estabilidad y la conectividad funcional varían a lo largo del sueño de manera no necesariamente alineada con las categorías tradicionales (61). Estos hallazgos apoyan la idea de que los estados dinámicos extraídos a partir del EEG podrían representar subestados fisiológicamente

significativos, revelando una organización del sueño más rica y fluida de lo que permite observar la etapificación clásica. Desde esta mirada, pensar en un análisis o *scoring* dinámico del ciclo sueño vigilia no es descabellado, considerando que, desde una perspectiva de redes, el cerebro puede entenderse como un sistema dinámico complejo donde emergen patrones globales de organización funcional (44).

La arquitectura metodológica implementada incluyó un refinamiento en dos etapas: un primer *clustering* espectral mediante *K-means* y luego un refinamiento utilizando *Agglomerative Clustering*. Esta estrategia se alinea conceptualmente con trabajos que combinan información espectral y temporal para una clasificación más robusta del sueño, demostrando que el uso conjunto de características multiescala y métodos de clasificación estructurados mejora la precisión en la segmentación automática de señales EEG durante el sueño (62).

También es importante subrayar que, a pesar del creciente interés en la etapificación automática del sueño, la mayoría de los estudios previos se han basado en aprendizaje supervisado y han conservado el marco de referencia de las cinco etapas del sueño (63, 64). Pocos trabajos, como el presente, han buscado una segmentación verdaderamente autónoma que no esté condicionada al momento de la clusterización por el modelo tradicional. Esto representa un aporte novedoso, ya que permite observar patrones emergentes de forma exploratoria y sin sesgos de categorización.

Uno de los aportes metodológicos más relevantes de esta investigación es la incorporación de la AMI como métrica principal para evaluar correspondencia. A diferencia de métricas como la *accuracy* o Kappa de Cohen, cuya aplicabilidad es más directa en contextos supervisados, la AMI se fundamenta en principios de la teoría de la información de Shannon, lo que le otorga una solidez conceptual particularmente adecuada para comparar estructuras emergentes en ausencia de una correspondencia directa entre etiquetas.

Desde una perspectiva matemática, la AMI mide la cantidad de información compartida entre dos particiones, corrigiendo explícitamente por el valor esperado bajo asignaciones aleatorias. A diferencia de métricas como la *accuracy* o el Kappa de Cohen —las cuales requieren una correspondencia directa entre etiquetas—, la AMI corrige por azar de forma robusta y es invariante a permutaciones de clases. Esto la convierte en una métrica particularmente adecuada para evaluar la similitud entre particiones en contextos no supervisados, donde las etiquetas de agrupamiento no están previamente definidas ni asociadas a categorías fijas.

En el contexto específico de la neurociencia computacional, donde las etiquetas funcionales del cerebro no siempre están claramente definidas y los estados pueden emerger como estructuras dinámicas y continuas, el uso de la AMI permite establecer relaciones estructurales entre agrupamientos sin imponer un marco categórico externo. De este modo, la evaluación de concordancia no se limita a una comparación taxonómica, sino que se convierte en una validación de la estructura relacional del fenómeno analizado.

Cabe destacar que se observa que ninguno de los estudios ya mencionados (29, 70, 71, 72, 73) ha incorporado la AMI como métrica de validación. La mayoría de estos trabajos utilizan métricas tradicionales como el índice de Kappa, el F1-score o el porcentaje de solapamiento, sin aprovechar las ventajas conceptuales y estadísticas que ofrece la AMI en contextos no supervisados.

Por tanto, el presente trabajo no solo propone una nueva perspectiva para abordar el análisis del sueño, sino que además introduce una herramienta de validación más coherente con los principios del análisis no supervisado, contribuyendo a mejorar la calidad metodológica de futuras investigaciones en segmentación automática del sueño y caracterización de estados cerebrales.

En definitiva, esta investigación no busca reemplazar el modelo tradicional, sino demostrar que existen otras formas posibles de representar el sueño. Los hallazgos no confirman completamente la hipótesis de una alta concordancia, pero sí evidencian una correspondencia estadísticamente moderada, lo cual justifica continuar con esta línea de trabajo. Se propone así una nueva forma de entender el sueño: no como un conjunto fijo de etapas, sino como un continuo de estados dinámicos, estructurados por la propia señal cerebral y su variabilidad contextual.

Este nuevo enfoque expande el lenguaje científico para describir el dormir humano, integrando elementos de análisis multiescala, continuidad temporal y dinámica cerebral. A futuro, estos avances podrían conducir al desarrollo de una etapificación más fisiológicamente fundada, menos dependiente de la interpretación visual humana y más cercana a los procesos reales que ocurren en el cerebro durante el sueño.

10.- CONCLUSIÓN

La presente investigación abordó el desafío de redefinir la segmentación del sueño mediante técnicas de análisis no supervisado, con énfasis en la señal EEG como fuente primaria de información. El objetivo central consistió en explorar si es posible identificar configuraciones en la arquitectura del sueño que mantengan una correspondencia estadísticamente significativa con la etapificación visual tradicional establecida por la AASM. Para ello, se aplicaron algoritmos de *clustering* sobre representaciones espectrales y dinámicas de la señal EEG, evaluando la concordancia de los resultados con métricas robustas como AMI y Kappa de Cohen.

Los resultados obtenidos, si bien no mostraron una concordancia perfecta con la segmentación manual, sí evidenciaron un acuerdo moderado ($AMI \approx 0.2952$ y $Kappa \approx 0.4397$) entre ambos enfoques. Estos valores son consistentes con hallazgos previos en la literatura que cuestionan la estabilidad del *scoring* visual incluso entre expertos. La concordancia moderada observada, si bien no permite afirmar que el análisis no supervisado replique completamente la clasificación tradicional, sí entrega evidencia a favor de su validez como un método alternativo que captura aspectos relevantes de la dinámica cerebral durante el sueño.

Dado lo anterior, la hipótesis planteada —que el análisis no supervisado del EEG permite obtener segmentaciones con alta concordancia estadística respecto a la etapificación visual— no puede ser plenamente aceptada bajo los criterios convencionales de "alta" concordancia. No obstante, tampoco se rechaza en términos absolutos, ya que se observa una tendencia significativa hacia el acuerdo, lo cual valida la dirección metodológica de este trabajo y sugiere un prometedor potencial de mejora.

Es importante destacar que este estudio no buscaba necesariamente reemplazar la etapificación tradicional, sino más bien explorar su complementariedad o incluso superar sus limitaciones a través de una mirada basada únicamente en datos. En este sentido, los estados dinámicos identificados podrían ofrecer una resolución más fina y continua de la actividad cerebral, permitiendo detectar transiciones sutiles que la etapificación a 30 segundos no logra capturar.

Además, la metodología desarrollada constituye una contribución metodológica valiosa, especialmente al incorporar evaluaciones de estabilidad intersujeto y concordancia multiescala en la elección de la ventana temporal de análisis.

En resumen, la hipótesis no se confirma plenamente, pero tampoco se refuta, y los hallazgos respaldan la utilidad del enfoque propuesto. Los resultados abren camino a futuras investigaciones que profundicen en la relación entre los estados dinámicos derivados de técnicas no supervisadas y los procesos neurofisiológicos subyacentes al sueño. Esta aproximación, con refinamientos adicionales y mayor cantidad de datos, podría contribuir a una etapificación del sueño más fisiológicamente fundada y menos dependiente de la interpretación visual humana.

11.- ANEXO

Este anexo se corresponde a la referencia número 48.

Anexo 1. Script para refinar algorítmicamente los datos

```
1 c,w,t = clust_connect(p,400,z=[-2,-1,1,2])
2 cm = np.apply_along_axis(lambda i: np.mean(nx[:,c==i],1),
3     0,np.arange(w.shape[0])[None,:])
4 dcm = distance_matrix(cm.T,cm.T)
5 clust = AgglomerativeClustering(n_clusters=11, metric='precomputed',
6     linkage='average',compute_full_tree=True)
7 pp = kldiv(t); pp = (pp + pp.T)/2
8 l = clust.fit_predict(pp.T)
9 clm = clust_matrix(clust)
10 if True:
11     hclm = np.apply_along_axis(ventropy,0,clm)
12     l = category(clm[:,np.argwhere(hclm<3)[0][0]])
13 else:
14     def f_nmi(v): return adjusted_mutual_info_score(v[c],ss,
15         average_method='max')
16     nmi = np.apply_along_axis(f_nmi,0,clm)
17     l = category(clm[:,np.argmax(nmi)])
18     plt.plot(nmi); plt.suptitle("Información mutua ajustada");
19     plt.show()
20 lw = np.unique(l,return_counts=True)[1]
21 ncl = np.max(l)+1
22 v = segapply(np.arange(ncl),1,func=lambda i: np.mean(nx[:,
23     l[c]==i],1))
24 pc1 = m_pca(nx[:,nonoutl(np.mean(nx,0))])[0,:];
25 pc1 = -np.sign(pc1[0])*pc1
26 o = -lw
27 l = order(o)[1]; v = v[:,np.argsort(o)]
28 i = np.argsort(l)
```

Anexo 2. Función “Segmentador de índices ajustable”

```
1 def segslice(N, width, extra=0, reextra=None):
2     width = max(int(width), 1)
3     extra = max(int(extra), 0)
4     if reextra == None:
5         reextra = extra
6     else:
7         reextra = max(int(reextra), 0)
8     l = extra + width + reextra
9     v = np.zeros(l, int)
10    f = range(-extra, int(N/width)*width-extra, width)
11    for i in f:
12        v[:] = np.arange(i, i+l)
13        if i < 0:
14            v[v < 0] = 0
15        if i+l > N:
16            v[v >= N] = N-1
17    yield v
```

Anexo 3. Función “Aplicador de funciones por segmento”

```
1 def segapply(s, width, extra=0, reextra=None, func=lambda x: x):
2     if type(s) != np.ndarray:
3         raise RuntimeError('object is not a numpy.ndarray')
4     seg = segslice(s.shape[-1], width, extra, reextra)
5     x = func(s[..., next(seg)])
6     m = np.zeros(x.shape + (int(s.shape[-1]/max(int(width), 1)),))
7     m[..., 0] = x
8     for k, sl in enumerate(seg):
9         m[..., k+1] = func(s[..., sl])
10    return m
```

Anexo 4. Función “Generador de espectrogramas multitaper”

```
1 def mtxpq(s, fs, width, extra=0, r=None, NW=2.5, Kmax=5):
2     width = max(int(width), 1)
3     extra = max(int(extra), 0)
4     l = width + 2*extra
5     if r != None:
6         r = tuple(i*l/fs for i in r)
7         li = math.floor(min(max(r[0], 0), l))
8         hi = math.ceil(min(max(r[1], li), l)) + 1
9     else:
10        li = 0
11        hi = l//2
12    if extra > 0:
13        s = np.concatenate((np.full(extra, s[0]), s,
14                               np.full(extra, s[-1])))
15    w = signal.windows.dpss(l, NW, Kmax, return_ratios=True)
16    if type(w) != tuple:
17        w = (w[None, :], (1,))
18    w, p = w
19    p = p/sum(p)
20    x = signal.spectrogram(
21        s, fs, window=w[0, :], nperseg=l, noverlap=2*extra,
22        detrend='linear')[2][li:hi]*p[0]
23    if len(p) > 1:
24        for i in range(1, len(p)):
25            x[:, :] = x + signal.spectrogram(s, fs,
26                window=w[i, :], nperseg=l, noverlap=2*extra,
27                detrend='linear')[2][li:hi]*p[i]
28    return x
```

Anexo 5. Función “Mapeo de categorías únicas”

```
1 def category(v):
2     u = np.unique(v)
3     t = np.zeros(np.max(u)+1)
4     t[u] = np.arange(u.shape[0])
5     return np.int32(t[v])
```

Anexo 6. Función “Generador de orden relativo”

```
1 def order(v):
2     i = np.argsort(v)
3     n = v.shape[0]
4     r = np.zeros(n)
5     r[i] = np.arange(n)
6     return np.int32(r)
```

Anexo 7. Función “Análisis de componentes principales”

```
1 def m_pca(x):
2     mcov = np.cov(x, rowvar=True, bias=True)
3     w, v = linalg.eig(mcov)
4     return (v.T[np.flip(np.argsort(w)), :])
```

Anexo 8. Función “Secuencias probabilísticas condicionales”

```
1 def seqp(s, c=None, z=[-2, -1, 1, 2]):
2     if c is None: c = s
3     n = np.max(s)+1; m = np.max(c)+1
4     t = np.zeros((m, n))
5     for d in z:
6         for i in range(n):
7             if d<=0: k = np.argwhere(s[-d:] == i)
8             else: k = np.argwhere(s[:-d] == i) + d
9             if k.shape[0]>0:
10                u = np.unique(c[k], return_counts=True)
11                t[u[0],i] += u[1]/k.shape[0]
12     return t/len(z)
```

Anexo 9. Función “Cálculo de ganancia logarítmica”

```
1 def lgain(w, t):
2
3     wn = w/np.sum(w)
4
5     a = 1/wn.shape[0]/100
6     return np.log(t+a) - np.log(wn[:,None]+a)
```

Anexo 10. Función “Matriz de divergencia de Kullback-Leibler”

```
1 def kldiv(t, e=1e-4): # kullback-leibler
2     n = t.shape[0]
3     m = np.zeros(t.shape)
4     p = (t+e/n)/(1+e)
5     lp = np.log2(p)
6     for i in range(n):
7         for j in range(n):
8             if i==j: next
9
10            m[i,j] = ( np.sum(p[:,i]*(lp[:,i]-lp[:,j]))
11                      - p[i,i]*(lp[i,i]-lp[i,j]) - p[j,i]*
12                        (lp[j,i]-lp[j,j])
13                      + (p[i,i]+p[j,i])*np.log2((p[i,i]+
14                        p[j,i])/(p[i,j]+p[j,j])) )
15     return m
```

Anexo 11. Función “Clusterización inicial (K-means) y Cálculo de transiciones dinámicas)

```
1 def clust_connect(m, n=20, z=[2,1,-1,-2]):
2     clust = KMeans(n_clusters=n, n_init=5, max_iter=600)
3     clust.fit(m.T)
4     c = clust.labels_
5     cc = clust.cluster_centers_
6     i = np.argsort(cc[:,0])
7     j = np.zeros(n); j[i] = np.arange(n); c[:] = j[c]
8     w = np.unique(c, return_counts=True)[1]
9     t = seqp(c, z=z)
10    return c, w, t
```

Anexo 12. Función “Matriz de conexiones jerárquicas”

```
1 def clust_matrix(c):
2     c = c.children_
3     l = c.shape[0] + 1
4     m = np.tile(np.arange(l)[:,None], (1,l))
5     for i in range(l-1):
6         k = np.argwhere((m[:,i]==c[i,0]) | (m[:,i]==c[i,1]))
7         m[:,i+1] = m[:,i]
8         m[k,i+1] = l + i
9     return m
```

Anexo 13. Función “Cálculo de entropía de distribuciones”

```
1 def ventropy(v):
2     c = np.unique(v, return_counts=True)[1]
3     c = c/sum(c)
4     return -sum(np.log2(c)*c)
```

Anexo 14. Función “Árbol de relaciones jerárquicas”

```
1 def clust_tree(c): # c = clust.children_
2     l = c.shape[0]
3
4     def c_t(i):
5         if i <= l:
6             return [i]
7         else:
8             return [i, c_t(c[i-l-1, 0]), c_t(c[i-l-1, 1])]
9     return c_t(2*l)
```



UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
INSTITUTO DE CIENCIAS BIOMÉDICAS
PROGRAMA DISCIPLINARIO DE INMUNOLOGÍA



CONSENTIMIENTO INFORMADO

Estudio Clínico “COVID prolongado en pacientes con factores de riesgo cardiovascular”

Investigadores Responsables: Mercedes López Nitsche/Adrián Ocampo Gárces.

Lugar de la Investigación: Programa de Inmunología. Instituto de Ciencias Biomédicas. Facultad de Medicina. Universidad de Chile. Dirección: Avda. Independencia 1027. Comuna de Independencia. Santiago. Teléfono: 2978 6725.

1. INFORMACION SOBRE EL ESTUDIO DE INVESTIGACIÓN

1.1. ¿Cuál es el propósito del estudio?

El COVID prolongado es una enfermedad producida posterior a la infección con el virus SARS-CoV-2. En esta condición, los pacientes presentan una variedad de síntomas que aparecen o se mantienen más de dos meses después de contraer el virus. En algunos casos, el síndrome de COVID prolongado dura meses o años y puede causar distintos grados de discapacidad. Entre los síntomas más frecuentes encontramos cansancio, palpitaciones, mareos, dificultad para respirar o falta de aliento y tos. El objetivo de este estudio es describir y caracterizar clínica, inmunológica y virológicamente el síndrome de COVID prolongado en pacientes que tienen factores de riesgo cardiovascular, como sobrepeso, hipertensión arterial, tabaquismo o diabetes.

1.2. ¿Por qué debiera yo considerar mi participación como sujeto de investigación en este estudio y cuales son los objetivos de este?

Usted contribuirá al conocimiento de la relación entre el sistema inmune y virus SARS-CoV-2. Los datos que aquí se obtengan nos ayudarán a describir las características clínicas y la respuesta inmune durante el desarrollo del síndrome de COVID prolongado. Nuestro objetivo es contribuir a mejorar y optimizar terapias para esta condición e incidir en políticas de salud y rehabilitación.

1.3. ¿Quién puede participar en este estudio?

Pueden participar personas con factores de riesgo cardiovascular, que hayan contraído COVID-19 confirmado por PCR y que hayan respondido la encuesta de salud que forma parte de este estudio.

1.4. ¿Tengo necesariamente que participar en este estudio? ¿Si acepto participar, puedo cambiar de opinión o retirarme?

Usted no tiene ninguna obligación de participar en el estudio. Si acepta participar puede retirarse en cualquier momento del estudio, sin ningún perjuicio para usted.





1.5. ¿Si decido participar en el estudio, en que consisten precisamente las evaluaciones, y que tipo de tratamientos o procedimientos me van a practicar?

Usted deberá responder una encuesta de salud simple, que no le tomará más de 10 minutos y asistir a una entrevista médica de alrededor de 30 minutos. Si usted cumple con los criterios para ingresar al estudio, se le solicitará leer este consentimiento informado y si está de acuerdo se le invitará a firmarlo. Una vez ingresado/a al estudio, se le tomará una muestra de sangre venosa equivalente a tres cucharadas soperas, que serán utilizada para análisis inmunológicos. La toma de muestra se realizará en las unidades de toma de muestra del Hospital del Salvador y del Hospital José Joaquín Aguirre, por personal entrenado. Las muestras serán transportadas y almacenadas en el Programa de Inmunología de la Facultad de Medicina de la Universidad de Chile hasta su procesamiento, bajo condiciones de refrigeración adecuadas.

Para evaluar la persistencia del virus SARS-CoV-2 en su cuerpo, se le solicitará una muestra de deposiciones. Para esto le solicitaremos que usted recolecte una muestra de sus deposiciones en un contenedor especial y que las traiga a las unidades de toma de muestra del Hospital del Salvador o del Hospital José Joaquín Aguirre según corresponda. Las muestras de deposiciones se refrigerarán y almacenarán hasta su posterior procesamiento, en el Programa de Inmunología de la Facultad de Medicina de la Universidad de Chile.

Para evaluar la respiración durante el dormir realizaremos un examen denominado polisomnografía. Este examen es un procedimiento rutinario en la práctica médica que se realiza sin provocar dolor ni riesgo de daño físico, y sirve para definir la calidad y cantidad de las fases de sueño. Consiste en la medición de la actividad eléctrica cerebral, ritmo respiratorio y cardíaco, y movimientos de las piernas durante el sueño. Con ese propósito se instalan una serie de sensores en la cabeza y extremidades que son conectados mediante cables flexibles a un dispositivo portátil del tamaño un poco mayor a un celular que se sostiene entre las ropas del paciente. Usted podrá deambular sin dificultad mientras es registrado. Mientras duerme, será filmado mediante una cámara que permite registrar en completa oscuridad, y será asistido por dos investigadores durante toda la noche de registro. La toma del examen puede realizarse en domicilio, no requiere de premedicación u otros procedimientos adicionales, siendo sólo recomendado un horario de permanencia en cama equivalente a una noche de sueño normal.

Por último, la asistencia a controles se reforzará con un llamado telefónico previo al control. En caso de que usted presente alguna novedad en su estado de salud, se agendará un control médico extra y será derivado a su sistema de salud según corresponda.

1.6. ¿Cuál es el destino final de la sangre que no se utilice en el estudio? .

Su sangre se procesará en el *Laboratorio de Regulación de la Respuesta Inmune antitumoral*, del Instituto de Ciencias Biomédicas, de la Facultad de Medicina de la Universidad de Chile, a cargo de la responsable del estudio Dra. Mercedes López Nitsche, quien oficiará como responsable de la muestra, mientras dure el proyecto. Al finalizar el proyecto, parte de su muestra se eliminará según



las normas de Bioseguridad autorizadas por la Unidad de Bioseguridad de la Facultad de Medicina de la Universidad de Chile y otra parte quedará alojada en el Biorepositorio de la Facultad de Medicina de nuestra universidad, para ser usada en otros estudios similares a este.

1.7. ¿Qué peligros podría experimentar en este estudio, y que harán los investigadores para reducir el riesgo de que éstos se presenten?

Esta investigación no representa ningún riesgo previsible para los pacientes. No se realizará ningún procedimiento invasivo aparte de la toma de muestra de sangre. De la punción venosa que se realizará para la extracción de sangre, no se espera que aparezca ninguna reacción salvo el dolor habitual y algunas veces un pequeño hematoma (moretón). De la punción venosa que se realizará para la extracción de sangre, no se espera que aparezca ninguna reacción salvo el dolor habitual y algunas veces un pequeño hematoma (moretón).

1.8. ¿Qué harán los investigadores para asegurar que la información que recolectarán sobre mí, no caerá en manos equivocadas?

Usted será identificado con un código de ingreso al estudio, siendo su identidad sólo conocida por el investigador/a responsable del proyecto. Todos los datos recolectados en el proyecto se mantendrán en la plataforma REDCap, alojada en nuestra universidad. REDCap es una plataforma digital que permite que los datos recolectados sean almacenados en forma muy segura.

1.9. ¿Qué beneficios personales puedo yo esperar al participar en este estudio?

Su participación le permite acceder a un diagnóstico completo de su condición de COVID persistente y los resultados de las pruebas serán comunicados y discutidos con los participantes. Este estudio permite a los y las pacientes acceder a un seguimiento clínico por parte de un equipo compuesto por expertos en diversas áreas de la medicina, lo que nos permitirá hacer un diagnóstico integral y ofrecerle recomendaciones para el manejo de su condición y la derivación a especialista según corresponda.

1.10. ¿En qué podría este estudio beneficiar a otros?

Este estudio permitir a través de un enfoque multidisciplinar evaluar mejores estrategias epidemiológicas, clínicas y terapéuticas, efectos de las inmunizaciones y prevención del síndrome de COVID prolongado.

1.11. ¿Recibiré algún pago por participar en este estudio?

Su participación es voluntaria en este estudio y no percibirá ningún pago.

09/MAY/2023





1.12. ¿Se cobrará a mí, a mi Isapre o compañía de seguros de salud el costo de algunos de estos estudios?

Este estudio no tiene costo alguno para Ud. durante el desarrollo de este proyecto. Los tratamientos y complicaciones propias de su enfermedad y de su curso natural, las hospitalizaciones, honorarios, exámenes y tratamientos habituales deberán ser costeados por su sistema de salud (FONASA O ISAPRE).

1.14. ¿Una vez que yo haya ingresado como sujeto de estudio, a quien tendría que dirigirme para averiguar más acerca del estudio o para hacer llegar algún reclamo respecto al trato que hubiese recibido?

Derechos del participante: Usted recibirá una copia íntegra y escrita de este documento firmado. Si usted requiere cualquier otra información sobre su participación en este estudio puede comunicarse con:

Investigadores: Dra. Mercedes López Nitsche/Adrián Ocampo Gárce

Autoridad de la Institución: Programa Disciplinario de Inmunología. Instituto de Ciencias Biomédicas. Facultad de Medicina. Universidad de Chile.

Otros Derechos del participante

En caso de duda sobre sus derechos comunicarse con la presidenta(s) del Comité de Ética de Investigación en Seres Humanos Dra. Lucía Cifuentes. Teléfono: 2 - 978 95 36. Email: comiteceish@med.uchile.cl. Ubicación: a un costado de la Biblioteca Central de la Facultad de Medicina, Universidad de Chile en Av. Independencia 1027, Comuna de Independencia.

09/MAY/2023



ESTUDIO CLÍNICO "COVID prolongado en pacientes con factores de riesgo cardiovascular"

Investigadores: Mercedes López Nitsche/Adrian Ocampo Gárce



2. DOCUMENTACIÓN DEL CONSENTIMIENTO

2.1. Consentimiento Informado.

Declaro haber leído la información descrita y mis preguntas acerca del estudio han sido respondidas satisfactoriamente. Al firmar este documento recibiré una copia del mismo. Finalmente, indico que entiendo el propósito de mi participación en este estudio, además de las condiciones en las cuales se me ha invitado a participar.

Nombre del Participante:

RUT:

Fecha:

Firma:

SI ACEPTO

NO ACEPTO

09/MAY/2023



ESTUDIO CLÍNICO "COVID prolongado en pacientes con factores de riesgo cardiovascular"

Investigadores: Mercedes López Nitsche; Adrian Ocampo Gárce

Av. Independencia 1027. Fono: (56 2) 978 6347 Fax: (56 2) 978 6979. Santiago - Chile

11



2.3. Consentimiento del investigador

Al sujeto de investigación he entregado información sobre el estudio, y en mi opinión esta información es precisa y suficiente para que el sujeto entienda completamente la naturaleza, los riesgos y beneficios del estudio, y los derechos que tiene en tanto sujeto de investigación. No ha existido coerción ni ha actuado bajo influencia alguna. He sido testigo que el sujeto firmó el documento.

Nombre del Investigador que toma el consentimiento:

RUT:

Fecha:

Firma:

Nombre del representante de la Dirección del Establecimiento de Salud:

RUT:

Fecha:

Firma:

09|MAY|2023





UNIVERSIDAD DE CHILE - FACULTAD DE MEDICINA
COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES HUMANOS

ACTA DE APROBACIÓN DE PROYECTO

(Documento en versión 3.1 corregida 17.11.2020)

Con fecha 09 de mayo de 2023, el Comité de Ética de Investigación en Seres Humanos de la Facultad de Medicina, Universidad de Chile, integrado por los siguientes miembros:

Dra. Lucia Cifuentes O., Médico Genetista, Profesor Titular, Presidenta Subrogante
Dra. María Angela Delucchi B., Médico Pediatra Nefrólogo
Sra. Claudia Marshall F., Educadora, Representante de la comunidad
Dra. Grisel Orellana V., Médico Neuropsiquiatra, Profesor Asociado
Prof. Julieta González B., Bióloga Celular, Profesor Asociado
Dra. María Luz Bascuñán R., Psicóloga PhD, Profesor Asociado
Sra. Karima Yarmuch G., Abogada, Dirección Jurídica, Facultad de Medicina
Srta. Javiera Cobo R., Nutricionista, Secretaria Ejecutiva
Prof. Verónica Aliaga C., Kinesióloga, Magíster en Bioética, Profesor Asociado
Dr. Dante Cáceres L., Médico Veterinario, Doctor en Salud Pública, Profesor Asociado

Ha revisado el Proyecto de Investigación titulado: **"LONG COVID IN PATIENTS WITH CARDIOVASCULAR RISK FACTORS"**. Cuyo investigador Responsable es la Dra. Mercedes López, quien desarrolla labores en el Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile.

El Comité revisó los siguientes documentos del estudio:

- Proyecto Fondecyt Regular
- Curriculum Vitae de los investigadores
- Carta Compromiso del investigador
- Consentimiento Informado

El proyecto y los documentos señalados en el párrafo precedente han sido analizados a la luz de los postulados de la Declaración de Helsinki, de las Pautas Éticas Internacionales para la Investigación Biomédica en Seres Humanos CIOMS 2016, y de las Guías de Buena Práctica Clínica de ICH 1996.

09/MAY/2023





UNIVERSIDAD DE CHILE - FACULTAD DE MEDICINA
COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES HUMANOS

Sobre la base de esta información el Comité de Ética de Investigación en Seres Humanos de la Facultad de Medicina de la Universidad de Chile se ha pronunciado de la siguiente manera sobre los aspectos del proyecto que a continuación se señalan:

- a) Carácter de la población a estudiar: Pacientes adultos con COVID, que se clasificaran en aquellos con y sin Long COVID.
- b) Utilidad del proyecto: Proyecto muy interesante que abarca una temática desconocida e importante en la población, como son las personas con COVID prolongado, evaluándose aquellos pacientes con riesgo cardiovascular.
- c) Riesgos y beneficios: riesgos mínimos; el principal es la obtención de sangre venosa.
- d) Protección de los participantes (asegurada por el consentimiento informado): Adecuada
- e) Notificación oportuna de reacciones adversas: No aplica
- f) Compromiso del investigador responsable en la notificación de los resultados del estudio al finalizar el proyecto: Si-
- g) Requiere seguimiento o visita en terreno: Si ____ No X
N.º de vistas: ____

Por lo tanto, el comité estima que el estudio propuesto está bien justificado y que no significa para los sujetos involucrados riesgos físicos, psíquicos o sociales mayores que mínimos.

Este comité también analizó y aprobó los correspondientes documentos de Consentimiento Informado en su versión modificada el 08 de mayo de 2023, que se adjunta firmado, fechado y timbrado por este CEISH.

Sin perjuicio de lo anterior, según lo establecido en el artículo 10 bis del D.S N° 114 de 2011, del Ministerio de Salud que aprueba el reglamento de la ley N° 20.120; es preciso recordar que toda investigación científica en seres humanos deberá contar con la autorización expresa del o de los directores de los establecimientos dentro de los cuales se efectúe, la que deberá ser evacuada dentro del plazo de 20 días hábiles contados desde la evaluación conforme del CEISH, siendo de responsabilidad del investigador enviar a este Comité una copia de la misma dentro del plazo señalado.

09/MAY/2023





UNIVERSIDAD DE CHILE - FACULTAD DE MEDICINA
COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES HUMANOS

En virtud de las consideraciones anteriores el Comité otorga la aprobación ética para la realización del estudio propuesto, dentro de las especificaciones del protocolo.

Se extiende este documento por el periodo de **04 años** a contar desde la fecha de aprobación prorrogable según informe de avance y seguimiento bioético.

Lugar de realización del estudio:

- Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile.

JAVIERA DE LOS
ANDES COBO
RIVEROS

Nta. Javiera Cobo Riveros
Secretaria Ejecutiva CEISH

Santiago, 09 de mayo 2023

Proyecto: N° 011-2023
Archivo acta: N° 008

09/MAY/2023



12.-REFERENCIAS

1. Scammell, T. E.; Arrigoni, E.; Lipton, J. O. “Neural Circuitry of Wakefulness and Sleep”. *Neuron* 93: 747–765, 2017.
2. Li, J.; Vitiello, M. V.; Gooneratne, N. S. “Sleep in Normal Aging”. *Sleep Medicine Clinics* 13: 1–11, 2018.
3. Hobson, J. A. “Sleep is of the brain, by the brain and for the brain”. *Nature* 437: 1254-1256, 2005.
4. Carskadon, M. A.; Dement, W. C. “Normal human sleep: An overview”. En: *Principles and Practice of Sleep Medicine*, 5ª ed.; Kryger, M. H.; Roth, T.; Dement, W. C., Eds.; Elsevier, 2011; pp 16-26.
5. Lan, K. C.; Chang, D. W.; Kuo, C. E.; Wei, M. Z.; Li, Y. H.; Shaw, F. Z.; Liang, S. F. “Using off-the-shelf lossy compression for wireless home sleep staging”. *Journal of Neuroscience Methods* 246: 142–152, 2015.
6. Aboalayon, K.; Faezipour, M.; Almuhammadi, W.; Moslehpour, S. “Sleep stage classification using EEG signal analysis: A comprehensive survey and new investigation”. *Entropy* 18: 272, 2016.
7. Rodríguez-Sotelo, J.; Osorio-Forero, A.; Jiménez-Rodríguez, A.; Cuesta-Frau, D.; Cirugeda-Roldán, E.; Peluffo, D. “Automatic Sleep Stages Classification Using EEG Entropy Features and Unsupervised Pattern Analysis Techniques”. *Entropy* 16: 6573–6589, 2014.
8. Gudmundsson, S.; Runarsson, T. P.; Sigurdsson, S. “Automatic sleep staging using support vector machines with posterior probability estimates”. *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)* 366–372, IEEE, 2005.

9. Iber, C.; Ancoli-Israel, S.; Chesson, A. L.; Quan, S. F. The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications; American Academy of Sleep Medicine, 2007.
10. Nunez, P. L. “Electroencephalography (EEG)”. En Encyclopedia of the Human Brain; Ramachandran, V. S., Ed.; Academic Press, 2002; pp 169–179.
11. Buzsáki, G. Rhythms of the Brain; Oxford University Press, 2006.
12. Walker, M. P. “The role of sleep in cognition and emotion”. Annals of the New York Academy of Sciences 1156: 168–197, 2009.
13. Oppenheim, A. V.; Schafer, R. W. Discrete-time signal processing, 3^a ed.; Prentice Hall, 2010.
14. Mallat, S. A wavelet tour of signal processing: The sparse way, 3^a ed.; Academic Press, 2009.
15. Bashar, S. K.; Ding, E.; Walkey, A. J.; McManus, D. D. “ECG signal feature extraction trends in methods and applications”. BioMedical Engineering OnLine 22: 1–22, 2023.
16. Cohen, L. Time-frequency analysis; Prentice Hall, 1995.
17. Goodfellow, I.; Bengio, Y.; Courville, A. Deep learning; MIT Press, 2016.
18. Rabiner, L. R.; Schafer, R. W. Digital processing of speech signals; Prentice Hall, 1978.
19. Bracewell, R. N. The Fourier transform and its applications, 3^a ed.; McGraw-Hill, 1999.
20. Russell, S.; Norvig, P. Artificial Intelligence: A Modern Approach, 4^a ed.; Pearson, 2020.

21. Bishop, C. M. Pattern Recognition and Machine Learning; Springer, 2006.
22. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2^a ed.; Springer, 2009.
23. Murphy, K. P. Machine Learning: A Probabilistic Perspective; MIT Press, 2012.
24. Jain, A. K.; Murty, M. N.; Flynn, P. J. "Data clustering: A review". ACM Computing Surveys 31, 264-323, 1999.
25. Kaufman, L.; Rousseeuw, P. J. Finding Groups in Data: An Introduction to Cluster Analysis; Wiley, 2009.
26. Ikotun, A. M.; Ezugwu, A. E.; Abualigah, L.; Abuhaija, B.; Heming, J. "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data". Information Sciences 622: 178–210, 2023.
27. Everitt, B.; Landau, S.; Leese, M.; Stahl, D. Cluster Analysis, 5^a ed.; Wiley, 2011.
28. Jain, A. K. "Data clustering: 50 years beyond K-means". Pattern Recognition Letters 31: 651-666, 2010.
29. Decat, N.; Walter, J.; Koh, Z. H.; et al. "Beyond traditional sleep scoring: Massive feature extraction and data-driven clustering of sleep time series". Sleep Medicine 98: 39–52, 2022.
30. Li, M.; Gao, Q.; Yu, T. "Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters". BMC Cancer 23: 799, 2023.
31. Cohen, J. "A coefficient of agreement for nominal scales". Educ. Psychol. Meas. 20: 37–46, 1960.

32. Vinh, N. X.; Epps, J.; Bailey, J. “Information theoretic measures for clusterings comparison: Variants, properties, normalization, and correction for chance”. *Journal of Machine Learning Research* 11: 2837-2854, 2010.
33. Vinh, N. X.; Epps, J.; Bailey, J. “Information theoretic measures for clusterings comparison: Is a correction for chance necessary?”. En *Proceedings of the 26th Annual International Conference on Machine Learning*; ACM, 2009; pp 1073–1080.
34. Himanen, S. L.; Hasan, J. “Limitations of Rechtschaffen and Kales”. *Sleep Medicine Reviews* 4: 149–167, 2000.
35. Parrino, L.; Ferri, R.; Zucconi, M.; Fanfulla, F. “Commentary from the Italian Association of Sleep Medicine on the AASM manual for the scoring of sleep and associated events: For debate and discussion”. *Sleep Medicine* 10: 799–808, 2009.
36. Fiorillo, L.; Puiatti, A.; Papandrea, M.; Ratti, P.-L.; Favaro, P.; Roth, C.; Bargiotas, P.; Bassetti, C. L.; Faraci, F. D. “Automated sleep scoring: A review of the latest approaches”. *Sleep Medicine Reviews* 48: 101204, 2019.
37. Ramos, J. A.; Olmos, I.; Olvera-López, J. A. “Analysis of EEG Signal Processing Techniques based on Spectrograms”. *Research in Computing Science* 145: 151-162, 2017.
38. Shukla, P. D. *Complex wavelet transforms and their applications*; Tesis de Maestría, Department of Electronic and Electrical Engineering, University of Strathclyde, 2003.
39. Arulampalam, A.; Jenkins, N.; Ekanayake, J. “Power quality and harmonic analysis of future grid-connected power systems with high penetration of distributed energy resources”. *International Journal of Electrical Power & Energy Systems* 29: 1287-1296, 2007.
40. Babiloni, C.; Barry, R.; Basar, E.; Blinowska, K.; Cichocki, A.; Drinkenburg, W.; Klimesch, W.; Knight, R.; Lopes da Silva, F.; Nunez, P.; Oostenveld, R.; Jeong, J.;

- Pascual-Marqui, R.; Valdés-Sosa, P.; Hallett, M. “International Federation of Clinical Neurophysiology (IFCN) – EEG research workgroup: Recommendations on frequency and topographic analysis of resting state EEG rhythms. Part 1: Applications in clinical research studies”. *Clinical Neurophysiology* 131: 285-307, 2020.
41. Khalighi, S.; Sousa, T.; Santos, J. M.; Nunes, U. “ISRUC-Sleep: A comprehensive public dataset for sleep researchers”. *Computer Methods and Programs in Biomedicine* 124: 180–192, 2016.
 42. Seeck, M.; Koessler, L.; Bast, T.; Leijten, F.; Michel, C.; Baumgartner, C.; He, B.; Beniczky, S. “The standardized EEG electrode array of the IFCN”. *Clinical Neurophysiology* 128: 2070–2077, 2017.
 43. McGill University. “Electroencephalography (EEG)”. Recuperado de McGill University.
 44. Sporns, O. *Networks of the Brain*; MIT Press, 2011.
 45. Hipp, J. F.; Hawellek, D. J.; Corbetta, M.; Siegel, M.; Engel, A. K. “Large-scale cortical correlation structure of spontaneous oscillatory activity”. *Nature Neuroscience* 15: 884–890, 2012.
 46. Shannon, C. E. “Communication in the presence of noise”. *Proceedings of the IRE* 37: 10-21, 1949.
 47. Huijse, P. [INFO183] Análisis de sistemas lineales 2020 [Lista de reproducción en YouTube]. Universidad Austral de Chile. Disponible en: YouTube, 2020.
 48. Biblioteca perteneciente al “Laboratorio de cronobiología del sueño”, Universidad de Chile.
 49. Xu, D. and Tian, Y. “A Comprehensive Survey of Clustering Algorithms”. *Ann. Data Sci.* 2: 165–193, 2015.

50. Rodríguez, M.Z.; Comin, C.H.; Casanova, D.; Bruno, O.M.; Amancio, D.R.; Costa, L. da F. and Rodrigues, F.A. “Clustering Algorithms: A Comparative Approach”. *PLOS ONE* 14: e0210236, 2019.
51. Ikotun, A.M.; Abdullahi, M.; Alkasassbeh, M.; Misra, S. and Maskeliūnas, R. “K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data”. *Inf. Sci.* 634: 119–153, 2023.
52. Yin, C.; Yang, Z.; Du, M.; Li, H. and Li, T. “A Deep Learning Method Approach for Sleep Stage Classification Using EEG Spectrogram”. *Int. J. Environ. Res. Public Health* 19(10): 6322, 2022.
53. Tsiouris, K.M.; Pezoulas, V.C.; Kartsakoulis, P.; Zervakis, M.; Konitsiotis, S. and Giannakeas, N. “Feature extraction from EEG spectrograms for epileptic seizure detection”. *Pattern Recognit. Lett.* 139: 112–118, 2020.
54. Acharya, U.R.; Oh, S.L.; Hagiwara, Y.; Tan, J.H. and Adeli, H. “Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Review”. *Expert Syst. Appl.* 134: 112–129, 2019.
55. Hassan, A.R. and Bhuiyan, M.I.H. “Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting”. *Comput. Methods Programs Biomed.* 140: 201–210, 2017.
56. Wang, Y.; Zhang, J.; Li, J.; Sun, J. and Liu, W. “Feature extraction using dominant frequency bands and time-frequency image analysis for chatter detection in milling”. *Comput. Ind.* 103: 34–43, 2018.
57. Ghahremani, P.; Kumar, N.; Povey, D. and Khudanpur, S. “Robust feature extraction using subband spectral centroid histograms”. *ICASSP Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* 1: 4002–4006, 2014.
58. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H. and Faubert, J. “Trends in EEG signal feature extraction applications”. *Front. Artif. Intell.* 5: 1072801, 2022.

59. Danker-Hopfe, H.; Anderer, P.; Zeitlhofer, J.; Boeck, M.; Dorn, H.; Gruber, G.; Heller, E.; Loretz, E.; Moser, D.; Parapatics, S.; Saletu, B. “Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard”. *Journal of Sleep Research*. 18: 74–84, 2009.
60. Gervasoni, D.; Lin, S.C.; Ribeiro, S.; Soares, E.S.; Pantoja, J.; Nicolelis, M.A.L. “Global forebrain dynamics predict rat behavioral states and their transitions”. *Journal of Neuroscience*. 24(49): 11137–11147, 2004.
61. Jobst, B.M.; Hindriks, R.; Laufs, H.; Tagliazucchi, E.; Hahn, G.; Ponce-Alvarez, A.; Deco, G. “Increased stability and breakdown of brain effective connectivity during slow-wave sleep: Mechanistic insights from whole-brain computational modelling”. *Scientific Reports*. 7: 4634, 2017.
62. Lajnef, T.; Chaibi, S.; Ruby, P.; Aguera, P.E.; Eichenlaub, J.B.; Samet, M.; Kachouri, A.; Jerbi, K. “Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines”. *Journal of Neuroscience Methods*. 250: 94–105, 2015.
63. Biswal, S.; Sun, H.; Goparaju, B.; Westover, M.B.; Sun, J.; Bianchi, M.T. “Expert-level sleep scoring with deep neural networks”. *Journal of the American Medical Informatics Association*. 25: 1643–1650, 2018.
64. Supratak, A.; Dong, H.; Wu, C.; Guo, Y. “DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 25(11): 1998–2008, 2017.
65. Zhao, S.; Long, F.; Wei, X.; Ni, X.; Wang, H. and Wei, B. “Evaluation of a Single-Channel EEG-Based Sleep Staging Algorithm”. *Int. J. Environ. Res. Public Health*. 19: 2845, 2022.

66. Phan, H.; Andreotti, F.; Cooray, N.; Chén, O.Y. and De Vos, M. “Deep Learning in Automatic Sleep Staging With a Single Channel Electroencephalography”. *IEEE J. Biomed. Health Inform.* 23: 2085–2096, 2019.
67. Yildirim, M.; Tuncer, T. and Dogan, S. “A Multi Constrained Transformer-BiLSTM Guided Network for Automated Sleep Stage Classification from Single-Channel EEG”. *Biomed. Signal Process. Control.* 76: 103649, 2022.
68. Li, Y.; Song, Y.; Su, Y.; Cao, J. and Yang, Y. “Automatic Sleep Staging Based on Single-Channel EEG Signal Using Null Space Pursuit Decomposition Algorithm”. *IEEE Access.* 9: 46007–46017, 2021.
69. Supratak, A.; Dong, H.; Wu, C. and Guo, Y. “DeepSleepNet: a Model for Automatic Sleep Stage Scoring based on Raw Single-Channel EEG”. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25: 1998–2008, 2017.
70. Rodríguez-Sotelo, J.L.; Osorio-Forero, A.; Jiménez-Rodríguez, A.; Peluffo-Ordoñez, D.H.; Castellanos-Dominguez, G. and Murcia-Correa, L.S. “Automatic Sleep Stages Classification Using EEG Entropy Features and Unsupervised Pattern Analysis Techniques”. *Entropy.* 16: 6573–6589, 2014.
71. Rodríguez-Sotelo, J.L.; Osorio-Forero, A.; Jiménez-Rodríguez, A.; Restrepo-de-Mejía, F.; Peluffo-Ordóñez, D.H. and Serrano, J. “Sleep stages clustering using time and spectral features of EEG signals”. *Lect Notes Comput Sci.* 10337: 444–455, 2017.
72. Swarnkar, V. and Abeyratne, U.R. “Bispectral Analysis of Single Channel EEG to Estimate Macro-Sleep-Architecture”. *Comput. Biol. Med.* 55: 201–210, 2014.
73. Cusinato, R.; Gross, S.; Bainier, M.; Janz, P.; Schoenenberger, P. and Redondo, R.L. “Workflow for the unsupervised clustering of sleep stages identifies light and deep sleep in electrophysiological recordings in mice”. *J Neurosci Methods* 408: 110155, 2024. <https://doi.org/10.1016/j.jneumeth.2024.110155>