

UNIVERSIDAD DE MEDELLÍN



PROCESAMIENTO DE SEÑALES DE ELECTROENCEFALOGRAMAS MEDIANTE ESPECTROGRAMAS DE MODULACIÓN Y TRANSFORMADORES

Para optar al título de Magíster en Modelación y
Ciencia Computacional

Este trabajo se ha realizado desde la maestría de modelación y computación científica de la Universidad de Medellín en coordinación de las facultades de ciencias básicas e ingeniería bajo la dirección de los doctores:

Diana Patricia Tobón Vallejo (directora)
Carlos César Piedrahíta Escobar (codirector)

Juan Felipe Vélez Gómez

Agradecimientos

Agradezco a mis asesores Diana Tobón y Carlos Piedrahíta, por su guía y acompañamiento durante el desarrollo del presente trabajo.

A las profesoras Carmen Cecilia Sánchez y Liliana Giraldo, por su acompañamiento durante mi proceso de maestría.

Finalmente, agradezco a mi hermana Sara por su orientación en el contexto de la biología y ciencias de la vida.

Resumen

El presente trabajo busca evaluar los modelos basados en *Transformers* en una tarea de clasificación, utilizando como insumo los espectrogramas de modulación calculados a partir de señales de electroencefalogramas, para este proyecto, se implementó un modelo soportado en la plataforma Tensorflow, al tiempo que se usó una librería para cálculos de los espectrogramas. El modelo fue probado con tres diferentes experimentos basados en recopilación de electroencefalogramas para actividades cerebrales diferentes, las cuales son: elicitación de emociones, cálculos aritméticos, estimulación con luces parpadeantes. El modelo arroja un rendimiento en la exactitud de 42,19%, 66,77% y 52,93% respectivamente. Aunque inicialmente el modelo no refleja grandes resultados, constituye una base para trabajos futuros.

Palabras claves:

Electroencefalograma, Espectrograma de modulación, Transformers, Tensorflow.

Abstract:

The use of electroencephalograms is broadly known in medicine—diagnosis and treatment processes are supported on this technique—diseases such as Alzheimer's and epilepsy also have a wide association with studies based on the analysis of signals from electroencephalograms. However, its use also covers other fields such as: psychology, neuromarketing, bionics, among others. Consequently, regardless of the context, the relevance of the study of techniques oriented to the analysis of electroencephalograms is remarkable, especially if they are supported in a computational environment. On the other hand, Spectral analysis is a starting point for signal processing, even more so when relating with non-periodic signals, such as those originated by electroencephalograms. Modulation spectrograms are a technique for calculating the second-order spectral components of a signal, which makes it possible to access a frequency analysis that would not be easy to perform with the calculation of a conventional spectrogram. The present work evaluates computational models based on Transformers—an evolution of recurrent neural networks—for different classification tasks, using modulation spectrograms calculated from EEG signals as input. For this study, a model supported by the Tensorflow platform was implemented, while a library was used for spectrogram calculations. The model was tested with three different experiments based on the collection of electroencephalograms for different brain activities, which are: elicitation of emotions, arithmetic calculations, stimulation with flashing lights. The model shows an accuracy performance of 42.19%, 66.77% and 52.93% respectively. Although initially the model does not reflect great results, it constitutes a basis for future work.

Keywords:

Electroencephalogram, Modulation Spectrogram, Transformers, Tensorflow.

Contenido

Introducción	6
1 Descripción General	8
1.1 Objetivo general.....	8
1.2 Objetivos específicos.....	8
2 Marco teórico	9
2.1 Introducción	9
2.2 El cerebro	9
2.3 Dispositivos	12
2.4 Procesamiento de señales.....	16
2.4.1 Definiciones.....	16
2.4.2 Transformada de Fourier.....	18
2.4.3 Espectrogramas.....	20
2.5 EEG	23
2.6 Flujo de trabajo para el procesamiento de señales EEG.....	24
2.7 <i>Transformers</i>	27
3 Trabajos relacionados	29
3.1 Espectrogramas.....	29
3.2 Deep Learning	30
4 Metodología	35
4.1 Preprocesamiento.....	35
4.2 Modelo general (MST).....	37
5 Desarrollo	41
5.1 Herramientas de hardware utilizadas.....	41
5.2 Herramientas de software utilizadas	41
5.3 Conjuntos de datos seleccionados.....	41
5.4 Exploración de los conjuntos de datos.....	43
5.5 Implementación del MST	50
5.6 Entrenamiento y resultados.....	52
5.6.1 Clasificación de emociones	52
5.6.2 Clasificación de tarea cognitiva (cálculos).....	54
5.6.3 Clasificación efecto Flicker	56
5.6.4 Desafíos de la investigación	59

6	Discusión	60
7	Conclusiones	62
8	Referencias	63
9	Anexos	67
	Anexo 9.A – Kits BCI	67
	Anexo 9.B – Señales	68
	Anexo 9.C – Preprocesamiento (Visualización EEG)	74
	Anexo 9.D – Preprocesamiento (ERP)	75
	Anexo 9.E – Visualización Topográfica	75
	Anexo 9.F – Preprocesamiento (Visualización EEG completo)	76
	Anexo 9.G – Preprocesamiento (Visualización de espectrogramas wavelet).....	76
	Anexo 9.H – Preprocesamiento (Visualización de espectrogramas de modulación wavelet)..	77
	Anexo 9.I – Glosario	78
	Anexo 9.J – Procesamiento (MST con Fase A).....	80
	Anexo 9.K – Procesamiento (MST con Fase B).....	81

Introducción

Particularmente, el uso de electroencefalogramas es ampliamente conocido en medicina; procesos de diagnóstico y tratamiento están soportados en esta técnica; por otro lado, estudios relacionados con enfermedades como Alzheimer y la epilepsia también tienen una amplia asociación con los análisis de señales provenientes de electroencefalogramas. Sin embargo, el análisis estas señales abarca también otros campos como, la psicología, el neuromarketing, biónica, entre otros.

Con la aparición de dispositivos comerciales como por ejemplo los fabricados por la empresa InteraXon[®], los electroencefalogramas han tomado auge por fuera de los laboratorios y de los procesos del cuidado de la salud; adicionalmente, en el marco de la cuarta revolución industrial es posible que una de las fronteras de la conectividad con los dispositivos digitales resida en *wearables* orientados al registro y análisis de la actividad; tal como ha ocurrido con otros dispositivos en el pasado, como por ejemplo los relojes inteligentes, los cuales tienen acceso a una cantidad importante de bioseñales que pueden ser registradas, catalogadas y analizadas por sistemas de información, lo que también ha dado pie para el desarrollo de nuevas aplicaciones y plataformas de software.

En consecuencia, independiente del contexto, es destacable la relevancia del estudio de técnicas orientadas al análisis de electroencefalogramas, sobre todo si están apoyadas en un entorno computacional; no obstante, con el importante desarrollo que ha tenido el aprendizaje de máquina y la ciencia de datos en la década pasada, existe un corpus importante de estudios relacionados con este tema, por lo que previamente al desarrollo del presente trabajo se analizaron las metodologías que tuvieron mayor acogida en dicha década. En este caso, prevalecen los enfoques que hacen uso de modelos de *ensemble learning*, como los bosques aleatorios y las máquinas de soporte vectorial, más adelante, entran los modelos basados en redes neuronales, específicamente en aprendizaje profundo, lo cual coincidió con el fortalecimiento de plataformas como Tensorflow[®] y Pytorch[®]

Al final de la década pasada apareció una nueva técnica denominada *Transformes*, la cual es una evolución de las redes neuronales recurrentes, esta técnica mostró gran relevancia en el análisis de secuencias y fue pensada originalmente para resolver problemas de procesamiento de lenguaje natural. Sin embargo, su uso se ha extendido hacia otros tipos de aplicaciones. Posteriormente, en el año 2020 apareció el *Vision Transformer*, el cual, realiza clasificación de imágenes utilizando este tipo de arquitectura. Como se mostrará en el presente trabajo, esta implementación del *Transformer* se acopla a la técnica de los espectrogramas de modulación.

El análisis espectral es un punto de partida para el procesamiento de señales, más aún si se trata de señales no periódicas, tales como los electroencefalogramas. En consecuencia, los espectrogramas se han utilizado continuamente en este tipo de enfoques. Por otra parte, los espectrogramas de modulación constituyen una técnica para calcular los componentes espectrales de segundo orden de una señal, lo que hace posible acceder a un análisis de frecuencias que no sería fácil de realizarse con el cálculo de un espectrograma convencional.

Es así como el presente trabajo fusiona los modelos de *Transformers* con la técnica de espectrogramas de modulación, para dar solución al procesamiento de una tarea de clasificación que tiene como insumo señales de electroencefalogramas.

1 Descripción General

1.1 Objetivo general

Evaluar el rendimiento de modelos de Aprendizaje Profundo basados en la arquitectura *Transformers*, en tareas de clasificación con relación a procesos cognitivos, emociones elicítadas y el efecto Flicker, utilizando como insumo las señales de encefalogramas a las cuales se les calculará el espectrograma de modulación previamente.

1.2 Objetivos específicos

- Perfilar y seleccionar, diferentes bases de datos de encefalografías para el presente estudio.
- Implementar una fase de preprocesamiento con el fin de calcular los espectrogramas de modulación provenientes de cada uno de los canales de los electroencefalogramas.
- Implementar un modelo de *Transformer* para cada uno de los casos de estudio definidos.
- Evaluar las diferentes métricas de desempeño obtenidas de las fases de entrenamiento y validación de cada uno de los ensayos implementados.

2 Marco teórico

2.1 Introducción

El presente marco teórico comienza con la descripción de los procesos cerebrales que hacen posible la técnica conocida como electroencefalograma o EEG; también, proporciona una aproximación temática de la segmentación de las funcionalidades cerebrales, que son de importancia en el campo de la neurociencia y por ende en el presente estudio. Más adelante, se pretende exponer teorías que hacen posible la utilización de los registros creados a partir de los EEG. Para lo cual, se hace una exposición de la teoría de señales, realizando énfasis en su definición como una función con respecto al tiempo, ya que de esta forma se adaptarán las técnicas de análisis en los EEG de insumo. En las técnicas de procesamiento de señales con respecto a sus componentes espectrales se utiliza la definición de la transformada de Fourier discreta, la cual se adapta más fácilmente a una implementación computacional. Con respecto al análisis de señales, se omite de las definiciones el tema de series de Fourier ya que, por la naturaleza de los EEG, este estudio se concentrará en señales no periódicas. Por otro lado, el presente estudio se centra en las técnicas FFT (de sus siglas en inglés *Fast Fourier Transform*) y STFT (se sus siglas en inglés *Short-term Fourier Transform*) para el cálculo de espectrogramas trabajando en menor medida las definiciones de otras técnicas como wavelet o la transformada de Hilbert. Con respecto al procesamiento de los EEG, se elige la rama del aprendizaje de maquina o ML (de sus siglas en inglés *Machine Learning*). Sin embargo, aun así, el compendio temático es demasiado amplio, por lo que el presente marco teórico se dedica a mencionar el modelo basado en *Transformers*, sin buscar abarcar los modelos predecesores. En el tema del análisis de los modelos, se descarta el asunto de la complejidad algorítmica, ya que se trabaja sobre modelos estándar donde dichos análisis ya están incluidos en los textos originales.

Por último, se destaca la utilización de algunos extranjerismos y siglas en inglés, en el presente marco teórico, esto con el fin de facilitar la lectura con relación al ambiente informático.

2.2 El cerebro

Las células del sistema nervioso del cuerpo humano están divididas en dos categorías, células gliales y células nerviosas o neuronas [1], donde el cerebro humano está compuesto por millones de neuronas, las cuales se interconectan formando redes y circuitos. En las neuronas se identifican tres componentes principales tales como: las dendritas, el soma y el axón; no obstante, su morfología es variada en comparación con otros tipos de células del cuerpo humano (ver *Figura 1*).

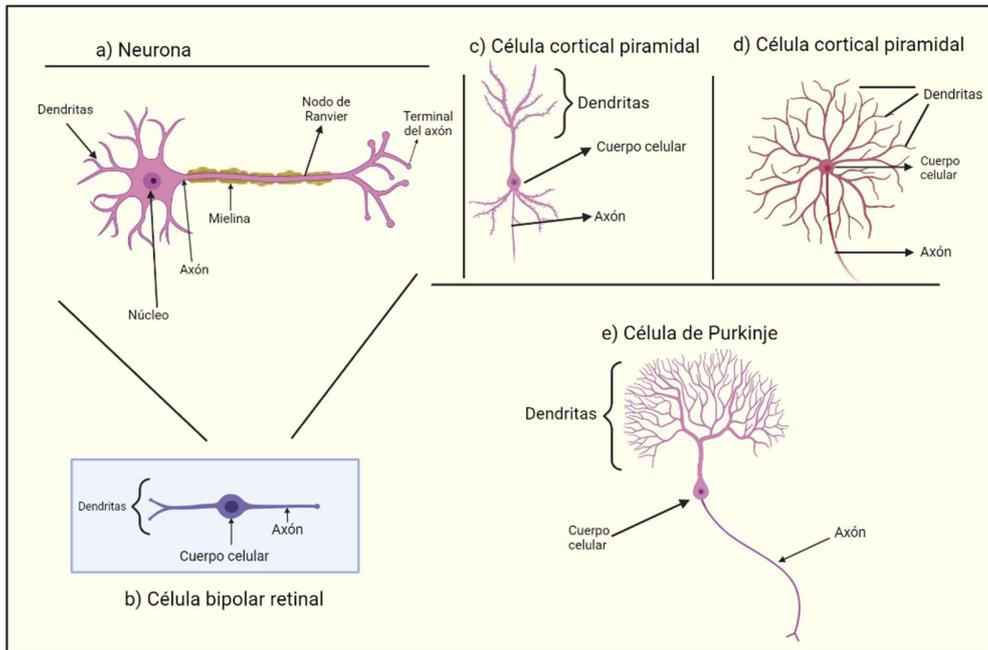


Figura 1: La neurona y algunas de sus diversas morfologías. Las neuronas no tienen forma regular, las longitudes de sus partes son variables al igual que su número de dendritas (figura obtenida y modificada de [1]).

Estas variaciones a nivel morfológico distinguen una especialización a nivel funcional. Los puntos de conexión de las neuronas se dan a través de las dendritas y las terminales de los axones que pertenecen a otras neuronas, hay diferentes configuraciones para las ramificaciones estructurales de las dendritas tales como: Piramidal, Purkinje, Stellate. Su tamaño es variable al igual que el número de conexiones que se llega a establecer entre ellas, se calcula un promedio de 180 conexiones por mm de longitud [2]. A través de las neuronas y otros componentes se forman estructuras complejas de orden funcional (ver *Figura 2*):

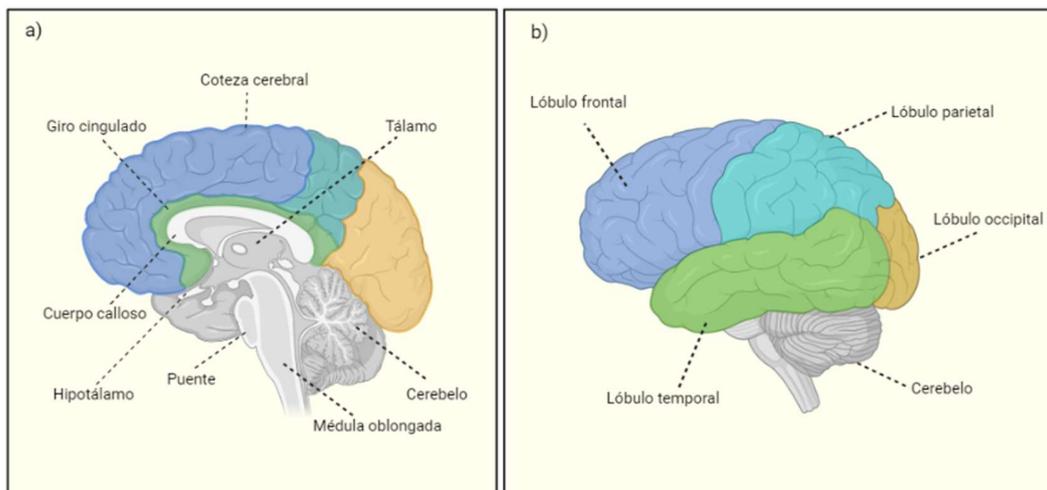


Figura 2: Sección transversal del cerebro humano y descripción de los lóbulos del cerebro. a) Refleja a través de una sección la compleja estructura del cerebro humano, el presente estudio trabajará con la actividad registrada en la parte externa llamada corteza cerebral. b) muestra que el cerebro humano también cuenta con una consideración espacial debido a la diferenciación de los lóbulos (figura obtenida y modificada de [1]).

Cuando se precisa abordar el cerebro humano desde una perspectiva funcional; es útil visualizarlo como un conjunto de componentes o secciones interconectadas que generan varios sistemas y subsistemas, a los cuales se les asigna una función determinada; sin embargo, no todas las funciones se pueden relacionar con un área específica del cerebro como por ejemplo el almacenamiento de los recuerdos [1]. Por otro lado, existe un volumen importante de aportes en el campo de la neurociencia que asocian determinadas tareas a ciertas áreas específicas del cerebro. Por ejemplo, se sabe que ciertas regiones del lado izquierdo, específicamente, en los lóbulos del córtex frontal y temporal son responsables de atender aspectos del lenguaje humano [1].

La corteza cerebral engloba dos hemisferios y cuatro lóbulos. La función del hemisferio derecho es regular las funciones motrices y motoras del lado izquierdo del organismo. Por su parte el hemisferio izquierdo regula del mismo modo el lado derecho. Mientras tanto, los lóbulos se pueden subdividir según su función:

- a) Lóbulo frontal: este genera la capacidad de moverse, de razonar y dar solución de problemas, parte del lenguaje y emociones.
- b) Lóbulo parietal: encargado de las percepciones sensoriales externas, tales como sensibilidad al tacto, presión, temperatura y dolor.
- c) Lóbulo occipital: encargado de la producción de imágenes.
- d) Lóbulo temporal: desempeña un papel importante en tareas visuales complejas como el reconocimiento de rostros, también está encargado de la audición, equilibrio y coordinación; además, regula emociones y motivaciones como la ansiedad, el placer y la ira.

La comunicación entre neuronas y por ende entre las áreas del cerebro destacadas anteriormente, se da a través de un proceso de sinapsis. Este consiste en la propagación de impulsos eléctricos entre las neuronas, los cuales son promovidos por mensajeros químicos llamados neuro transmisores, y canales iónicos de sodio, potasio, calcio y cloruro [2], los cuales están directamente involucrados en la transmisión de iones a través de la membrana celular hacia el interior de la neurona creando una diferencia de potencial entre el interior y el exterior. Esta diferencia de potencial produce diferentes efectos como la hiperpolarización y despolarización; cuando este último ocurre, se puede generar lo que se conoce como un potencial de acción [2], el cual, puede generar efectos de excitación, inhibición o despolarización. Cuando los potenciales de acción generan un efecto de excitación, se origina una reacción en cadena que transmite el potencial a las neuronas vecinas, las cuales experimentarán nuevamente los efectos mencionados anteriormente, logrando transmitir su potencial como una señal a otras neuronas.

El presente trabajo explorará tres aspectos funcionales del cerebro: las emociones, solución de tareas aritméticas y el efecto Flicker. Las emociones se abordan tanto desde el campo de la neurociencia como desde la psicología; los primeros estudios en este campo exponen diversas teorías y modelos para la comprensión de las emociones. Inicia con la teoría William James, la cual basa las emociones en la “sensación corporal” y describe el aspecto fenomenológico de la emoción como la consecuencia de una respuesta fisiológica a un estímulo [3]. Luego aparecen otras teorías que involucran los aspectos cognitivos más allá de los puramente fisiológicos, por ejemplo: la dirección objetiva de las emociones, dicho concepto está enmarcado dentro los modelos llamados evaluativos [4]. Los aspectos evaluativos de la emoción son importantes porque introducen los

conceptos de bueno-malo, agradable-desagradable; atributos que darán pie a los modelos contemporáneos basados en la discretización de emociones, como los expuestos por McDougall y luego por Ekman, el cual describe seis módulos emocionales básicos: alegría, tristeza, ira, disgusto, miedo y sorpresa [3]. Finalmente, estos módulos emocionales discretos se conjugan, con la idea de utilizar los conceptos de valencia y excitación como el núcleo central en la descripción del afecto. Lo anterior da origen al modelo de dos dimensiones de James Russell llamado Circumplex [5] (ver *Figura 3*), el cual es una de las piedras angulares de los métodos de clasificación de emociones a nivel computacional.

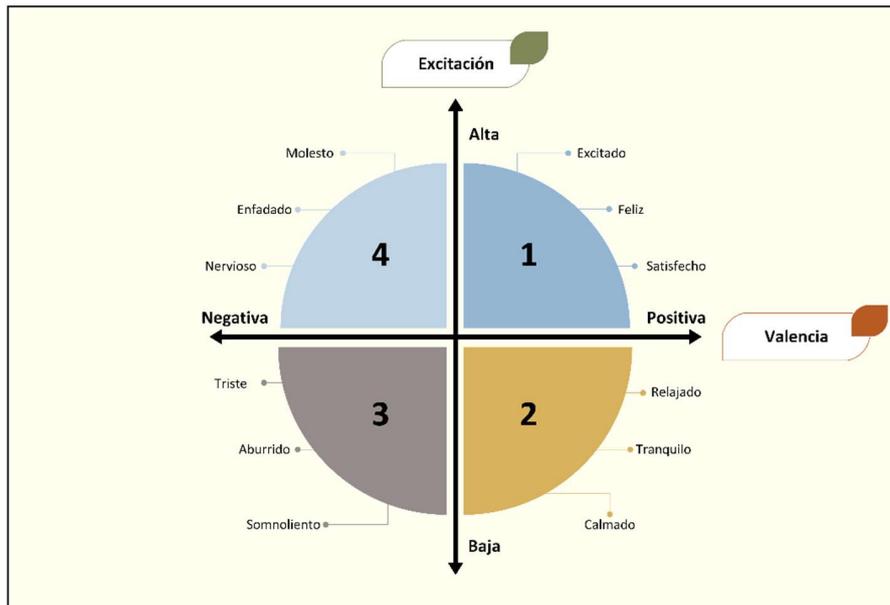


Figura 3: Modelo de circumplex 2D de Russell, este modelo describe los diferentes estados emocionales en términos de dos variables, excitación y valencia. Este enfoque se acopla a las implementaciones informáticas, debido a que se pueden definir dichas variables en escala numérica. (figura obtenida y modificada de [5]).

Realizar cálculos matemáticos es un proceso complejo debido a que concentra diferentes procesos funcionales como: el pensamiento, planeamiento, y toma de decisiones; al tiempo que involucra otros aspectos como la memoria de trabajo [6] y el estado emocional. Neuronalmente se ha detectado un involucramiento de la corteza prefrontal [7] cuando se llevan a cabo cálculos aritméticos.

Finalmente, el efecto Flicker [8] se refiere a la sincronización entre la actividad rítmica de la corteza visual y un estímulo externo a determinada frecuencia. Para efectos de este estudio se analiza el proveniente de una luz estroboscópica.

2.3 Dispositivos

Como se mencionó anteriormente, la sinapsis en fase de excitación produce un impulso eléctrico que viaja de neurona a neurona. Para el estudio de la actividad cerebral se ha desarrollado una técnica conocida como electroencefalografía (EEG) [9], la cual consiste en el registro de la actividad cerebral luego del proceso de sinapsis, la cual se consigue mediante la superposición de electrodos en el cuero cabelludo.

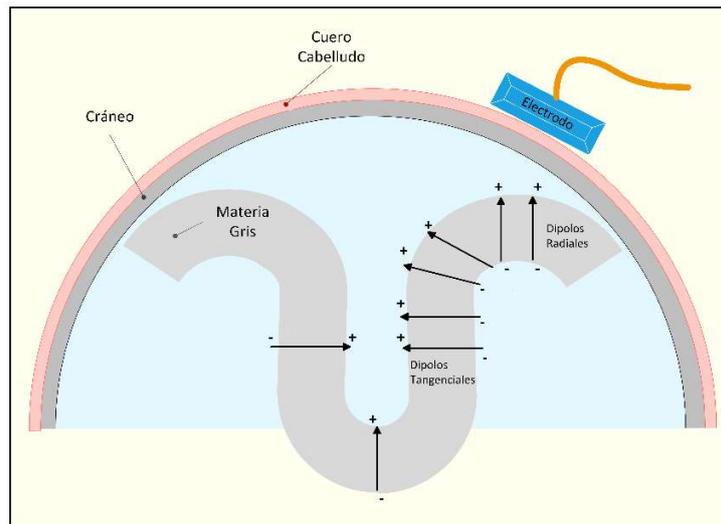


Figura 4: Aspectos biofísicos involucrados en la medición de la señal EEG, en este caso, el EEG se concentra en registrar la actividad del córtex, en la imagen se puede apreciar como la relación de las cargas eléctricas producidas por sinapsis fluyen a través del cráneo y el cuero cabelludo llegando a ser registradas por un electrodo superpuesto. (figura obtenida y modificada de [10]).

Con la evolución de la informática en términos de hardware y software, es posible contar con interfaces cerebro-computador BCI (de sus siglas en inglés *Brain Computer Interface*) de fácil acceso (ver Anexo 9.A); sin embargo, la mayoría de estos sistemas no entran en la categoría de dispositivos médicos, por lo que tienden a contar con prestaciones más bajas y su utilización se orienta a demostraciones, desarrollos de ingeniería e investigaciones que no incluyan estudios clínicos.

Un Kit de captura EEG (ver *Figura 5*) consta de cinco partes principales: (a) Electrodo-terminales, (b) Placa de adquisición, (c) Estructura de montaje, (d) Software; los cuales se describen a continuación:

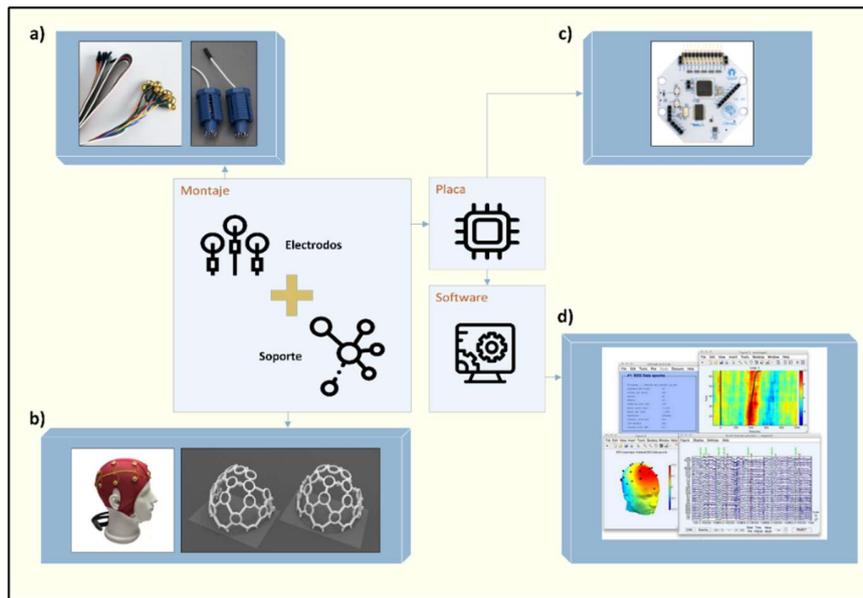


Figura 5: *Arquitectura física de un kit de captura EEG. a) Expone ejemplos de diferentes tipos de electrodos de izquierda a derecha, electrodos pasivos chapados en oro; en este tipo de electrodos es necesario el uso de un medio tipo gel para mejorar la conductividad, luego los electrodos pasivos en seco; los cuales no necesitan del medio tipo gel. b) Estructuras de montaje tipo gorro y estructura fija. c) Ejemplo de una placa Cython® fabricada por OpenBCI® d) Imagen referente al software EEGLAB.*

a) Electrodo-terminales: Dispositivo EEG está provisto de terminales llamadas electrodos, los cuales sirven para capturar las diferencias de potencial, producto de los eventos posinápticos en el cerebro [10].

Estos electrodos pueden ser activos o pasivos, su diferencia radica en el tratamiento de la señal al momento de la captura. Los activos tienen la capacidad de mejorar la señal adquirida en tiempo real, por ejemplo: aplicando funciones de amplificación; mientras que los pasivos son más susceptibles al ruido, y se ven afectados debido a que la conductividad del aire es cercana a cero [11]; por lo tanto algunos de ellos requieren un medio tipo gel para establecer la zona de contacto con el cuero cabelludo, lo que dificulta la preparación del sujeto para la toma de mediciones. También existe una variedad de electrodos pasivos que trabajan en seco (ver *Figura 5*).

Los electrodos van conectados a una placa de adquisición, donde las señales son obtenidas y digitalizadas. Se tomó como referencia la placa *Cython* fabricada por la compañía OpenBCI®, la cual está basada en el chip ADS1299-x. Merece destacarse que dicha placa es multipropósito, ya que no solo sirve para realizar capturas EEG, sino que también se puede extender su uso a otros estudios tales como: Electromiograma (EMG) o Electrocardiograma (ECG).

b) Placa de adquisición: La placa *Cython* cuenta con 8 canales y una posible extensión a 16 para la conexión de electrodos, también permite configurar otros parámetros para la captura de las señales tales como la frecuencia de muestreo. La placa de adquisición EEG necesitará una fuente de alimentación externa y el soporte de un protocolo de transmisión de información; en el caso de la

placa *Cython* se usan baterías convencionales tipo AA y la interfaz de comunicación con el computador es a través de un *dongle* Bluetooth basado en RFduino.

c) Estructura de montaje: existen diferentes tipos como estructuras fijas y gorros elásticos. Su tarea principal es fijar los electrodos sobre regiones específicas del cerebro, estos sistemas de posicionamiento buscan estandarizar la ubicación de los electrodos para diferentes sujetos; esto hace que las señales lleguen a ser más comparables.

El sistema más utilizado para la colocación de los electrodos es el 10-20 [12], este nombre se da por los porcentajes de separación entre electrodos adyacentes basados en las distancias y el número de canales (ver *Figura 6*).

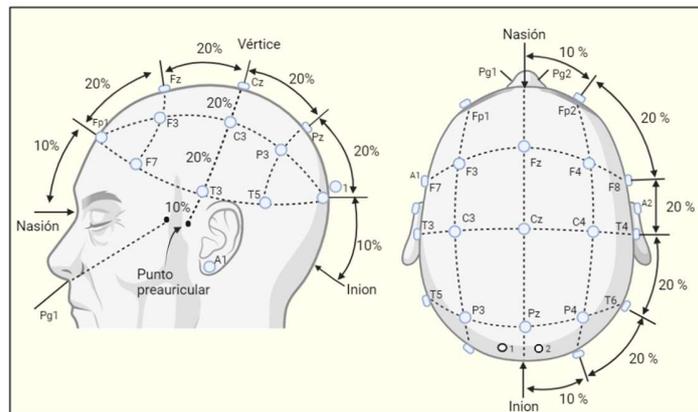


Figura 6: Sistema de posicionamiento 10-20, este sistema busca la estandarización del posicionamiento de los electrodos para que la información proveniente del EEG sea interpretable entre diferentes estudios o laboratorios (figura obtenida y modificada de [12]).

Finalmente, el destino de las señales digitalizadas por la placa, viajan a un dispositivo de procesamiento como un ordenador tipo laptop o de placa reducida como una Raspberry Pi®, esto dependerá del tipo de solución o de estudio en particular. Así mismo, las señales pueden ser almacenadas o procesadas en tiempo real.

d) Software: Existen diferentes componentes de software como las API (de sus siglas en inglés, *Application Programming Interface*), programas, IDE, *toolbox*, entre otros; que están orientados al procesamiento de señales EEG y pueden ser utilizados en diferentes momentos de la fase de experimentación dando soporte a tareas de visualización, aplicación de filtros, cálculos, extracción de componentes de señal, entre otras. Algunos ejemplos de estas API son: MNE[13] y EGGLAB[14].

Dependiendo del enfoque a utilizar en el procesamiento, por ejemplo: el uso de técnicas de ML (de sus siglas en inglés *Machine Learning*), donde es común, el uso de plataformas como TensorFlow® o de lenguajes de programación como Python®

2.4 Procesamiento de señales

2.4.1 Definiciones

El estudio de la actividad eléctrica originada en la sinapsis se puede abordar desde la teoría de señales analógicas. Una señal analógica básicamente es un fenómeno electromagnético que puede ser representado como una función matemática de una o más variables independientes, siendo el tiempo una de las variables más utilizadas [15]. En este caso, para una representación matemática de forma general se elige la función sinusoidal [16], esto se debe que a que dicha función es continua y representa a las señales periódicas en su forma básica, esta representación ofrece la posibilidad de crear modelos para la construcción o descripción de señales más complejas [17]. No obstante, existen otras formas o funciones básicas utilizadas en la representación de las señales, por ejemplo: las funciones de Legendre o las funciones de Hermite; las cuales, estas no serán profundizadas en el presente trabajo. Dentro de su estudio, las señales periódicas tienen ciertas características básicas asociadas, como lo son: la fase, la amplitud, la paridad, la estacionalidad, el periodo y la frecuencia (ver Anexo 9.B). Sin embargo, desde un punto de vista general, las señales se pueden caracterizar a través de otros atributos [18], tales como:

- Energía de la señal
- Energía promedio
- Ancho de Banda
- Duración de tiempo
- Producto tiempo-ancho de banda
- Dimensionalidad de la señal

La energía de una señal está definida por la siguiente ecuación:

$$E = \int_{-\infty}^{+\infty} x^2(t) dt < \infty \quad (2-1)$$

La energía promedio en un intervalo específico define el concepto de la potencia promedio de las señales, matemáticamente la potencia promedio está definida por:

$$0 < \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x^2(t) dt < \infty \quad (2-2)$$

Las señales se dividen en señales de energía y señales de potencia. Una condición básica de las señales de energía es que su amplitud decae a cero mientras que el intervalo de tiempo se acerca a infinito, esto tiene como consecuencia que la potencia promedio de la señal sea igual a 0. Por otra

parte, las señales de potencia se caracterizan porque la potencia promedio es finita, lo que significa que su energía es infinita; esto se puede contemplar solo desde la perspectiva del modelo matemático, debido a que, en el mundo real no es factible observar señales con energía y duración infinitas [19]. Una señal EEG es una señal no periódica, que cuenta con una amplia presencia de ruido debido a sus bajas frecuencias y a los órdenes de sus amplitudes [20]. Adicionalmente, cuentan con la característica de ser señales aleatorias ya que, según lo mencionado, al ser una señal aperiódica y ruidosa, su comportamiento es difícil de predecir tomando como base su estado actual.

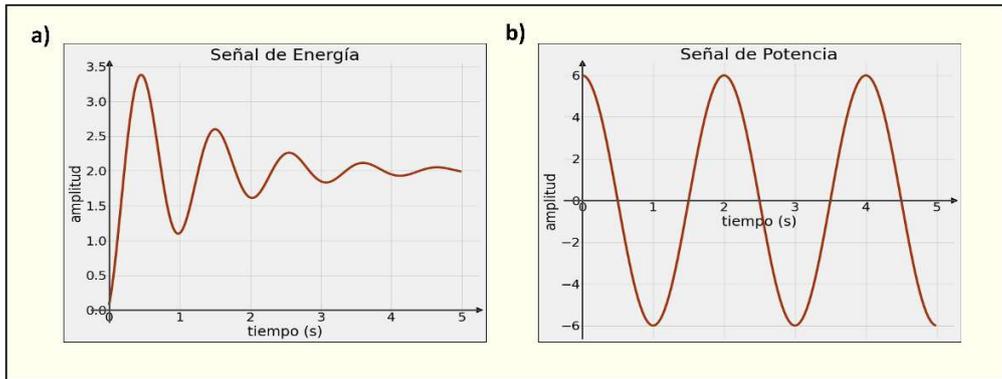


Figura 7: Comparación entre una señal de energía y una señal de potencia, según la teoría revisada en [19], los conceptos de señales de energía y señales de potencia son excluyentes. Lo que significa que una señal no puede ser considerada de energía y de potencia al mismo tiempo. A) muestra como la amplitud de la señal de energía tiende a cero conforme transcurre el tiempo. B) muestra una señal periódica, las cuales son consideradas como señales de potencia. (fuente propia).

El ancho de banda está definido por el rango de frecuencias donde la señal registra la mayor parte de su energía (ver *Figura 8*). En el caso de las señales EEG, los estudios han revelado que la banda de la señal se puede segmentar en sub-rangos, los cuales están asociados a ciertos aspectos funcionales del cerebro [21], [22]. A continuación, se exponen en la siguiente tabla:

Tabla 2-1: Clasificación de las bandas desde un aspecto funcional.

Ondas	Banda (Hz)	Asociación funcional
Delta	[0.1 – 4]	Sueño
Theta	[4 – 8]	Vigilia relajada
Alpha	[8 – 13]	Relajación
Beta	[13 – 31]	Alerta-Vigilia
Gamma	[31 – 60]	Dos sentidos diferentes al mismo tiempo, atención, percepción y cognición

El análisis de la potencia y del espectro de la señal, se conjuga en el concepto del PSD (de sus siglas en ingles *Power Spectral Density*). El cual se define bajo la siguiente expresión:

$$E\{|x(n)|^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) d\omega \tag{2-3}$$

El PSD es de utilidad para un amplio abanico de técnicas de análisis de señales, por ejemplo: modelos estadísticos, modelos basados en aprendizaje de máquina, entre otros.

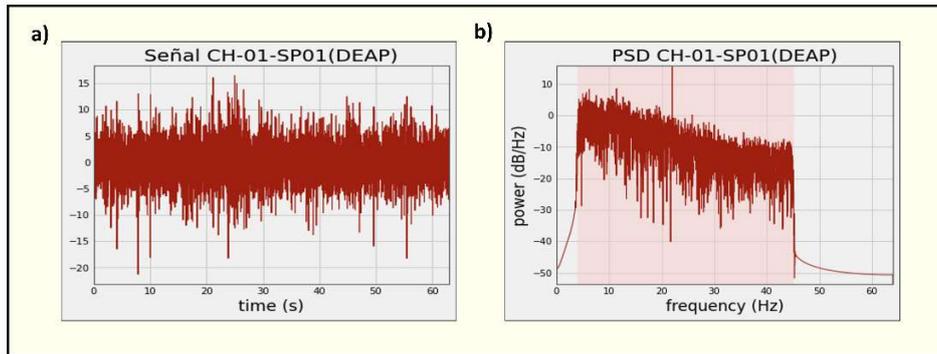


Figura 8: Potencia de una señal EEG con relación a su espectro, a) refleja una señal EEG. b) muestra la Potencia de una señal EEG con relación a su espectro, obsérvese que el rango de frecuencias está entre [4~45], esto se debe a que fue aplicado un filtro previamente (fuente propia).

2.4.2 Transformada de Fourier

La transformada de Fourier FT (de sus siglas en inglés *Fourier Transform*) es una de las herramientas fundamentales al momento de analizar los componentes espectrales de una señal, esta técnica se apoya en dos conceptos claves: el producto punto y la convolución (ver Anexo 9.B). La FT obtiene una representación de los tres componentes básicos de una señal: frecuencia, amplitud y fase; por lo que se considera una representación en tres dimensiones [17]. El funcionamiento de la FT se basa en el cálculo de los productos punto de la señal original con respecto a varias ondas sinusoidales; asumiendo en este caso que, un segmento de la señal EEG se conforma de varios componentes periódicos descritos por varias ondas sinusoidales caracterizadas por una variación en sus frecuencias. De esta manera se consigue una representación de la señal original en el dominio de la frecuencia.

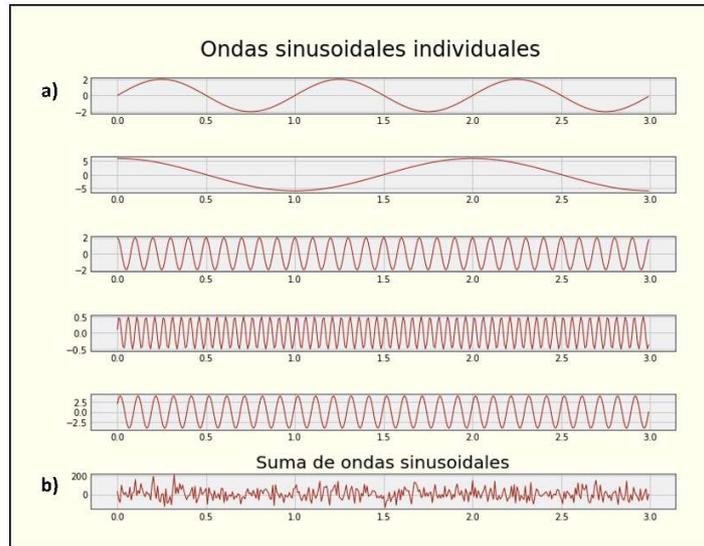


Figura 9: Simulación de una señal complicada. a) muestra la construcción de una señal aparentemente más complicada a partir de la suma de distintas ondas sinusoidales con variaciones de fase, frecuencia y amplitud (a). Adicionalmente a la figura en (b) se le agrega un ruido generado a partir de una distribución normal, dando como resultado una forma de onda parecida a la obtenida desde una señal EEG. (figura obtenida y modificada de [17]).

Para generación de ondas sinusoidales se puede utilizar la ecuación:

$$x(t) = A \text{sen}(2\pi ft + \theta) = A \text{sen}(\omega t + \theta) \quad (2-4)$$

Aunque la función está definida con respecto al tiempo, existen otros parámetros a considerar como: la amplitud máxima definida por A, la frecuencia definida por f y la fase definida por θ ; Por otro lado, ω se conoce como frecuencia angular y equivale a $2\pi f$. Las ondas sinusoidales tienen otras representaciones, por ejemplo, a través de una expresión exponencial compleja:

$$x(t) = C e^{-i\omega t} = C * \cos(\omega t) + C * i * \text{sen}(\omega t) \quad (2-5)$$

La anterior ecuación será la forma más utilizada en la implementación de la transformada de Fourier y como se señalará más adelante también en la transformada wavelet. La transformada de Fourier continua viene definida por la siguiente formula:

$$F(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt \quad (2-6)$$

En su contraparte, la transformada de Fourier discreta DFT (de sus siglas en inglés *Discrete Fourier Transform*) viene definida por la siguiente formula:

$$x(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-i\omega t} \quad (2-7)$$

Se destaca que las frecuencias a utilizar en la transformada de Fourier equivalen a la mitad de los puntos que describen la serie temporal más la frecuencia cero [17]; por lo que si una serie de tiempo es tamaño N, entonces su número máximo de frecuencias individuales será $N/2 + 1$. La fórmula anterior simboliza el cálculo de la transformada para una frecuencia una específica f, con respecto a una señal x; n son los puntos de la serie de tiempo que representa la señal.

Una de las características de la transformada de Fourier es que es invertible, lo que significa que se puede tomar la representación de la señal que está en el dominio de la frecuencia y devolverla a su constitución original en el dominio de tiempo:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\omega)e^{-i\omega t} d\omega \quad (2-8)$$

En este caso, la subversión del proceso obtenido por la transformada se consigue realizando el producto punto por los llamados coeficientes de Fourier (X). Noté que la otra diferencia entre la transformada y la transformada inversa radica en el signo de la función exponencial.

Finalmente, el método de la transformada de Fourier discreta no es práctico para una implementación computacional ya que trae consecuencias negativas a nivel de rendimiento [19]. Por tal motivo, se enriquece el método con técnicas más adaptables a una implementación informática óptima, como es el caso de la estrategia divide y vencerás. Dado es el caso de la transformada rápida de Fourier FFT (de sus siglas en inglés *Fast Fourier Transform*) que consigue mejorar sustancialmente la complejidad algorítmica y en consecuencia los tiempos de ejecución; en este caso, si se habla de la necesidad de realizar $O(N^2)$ operaciones en una DFT; una FFT garantiza un número de operaciones cercano al $O(N \log N)$.

Sin embargo, dentro de la definición de la FT existe necesidad de trabajar con señales estacionarias. En la naturaleza rara vez se presentan señales estacionarias en sentido estricto, por lo cual la suposición de la estacionalidad se limita a pequeños fragmentos de la señal, esto dependerá de su naturaleza, por ejemplo: las señales EEG se suponen estacionarias hasta en un rango de 20 segundos [23].

2.4.3 Espectrogramas

Un espectrograma es el resultado de dividir la señal original en tajadas de tiempo para luego calcular transformaciones en el dominio de la frecuencia de dichos fragmentos de señal; por lo cual, básicamente el espectrograma representa la variación de los componentes de frecuencia a través del tiempo y en consecuencia cuenta con cuatro dimensiones: amplitud, frecuencia, tiempo y fase.

Tal como se expone en [23], existen tres métodos principales para el cálculo de espectrogramas; SFTF (de sus siglas en inglés *Short-Time Fourier Transform*), CWT (de sus siglas en inglés *Complex Wavelet Transform*) y la transformada de Hilbert HT (de sus siglas en *Hilbert Transform*). Tanto SFTF como CWT poseen la misma naturaleza, se le aplica un kernel a la señal original, la cual está en el dominio del tiempo. Una particularidad de estas técnicas es que producen un cambio en el dominio de la señal (en este caso de tiempo a frecuencia). Como se indicó anteriormente, la transformada de Hilbert queda por fuera del presente estudio, ya que una de sus desventajas es su lentitud computacional [24]. A continuación, se presenta un resumen de la técnica.

$$X(t, f) = x(t) \star \lambda_f(t) e^{i2\pi ft} \quad (2-9)$$

La ecuación (2-9) está sustentada en el teorema de convolución y hace referencia a una generalización algorítmica ofrecida por [23]; en este caso, el componente $\lambda_f(t)$, el cual hace referencia a la ventana, homóloga las tres técnicas mencionadas STFT, CWT y HT; por otra parte, la parametrización de $\lambda_f(t)$ influirá la resolución temporal y espacial del espectrograma calculado. A continuación, se exponen algunas pautas para el control de dicha característica:

En los STFT:

- Un segmento de ventana corto ofrece una mayor precisión a nivel temporal, una menor precisión a nivel de frecuencias y una menor resolución.
- Un segmento de ventana largo ofrece una mayor resolución y precisión en frecuencias, pero una baja precisión temporal.

En los CWT:

- Con respecto a la longitud de la wavelet, se debe permitir que esta toque el cero para la frecuencia más baja en ambos extremos, también deberán ir centradas en tiempo igual a 0 para el centro de la wavelet.
- Un número de ciclos largo mejora la precisión a nivel de frecuencias, pero la desmejora en el espacio temporal, por lo que este número se debe elegir según la hipótesis y contexto del experimento.

Los espectrogramas de modulación descritos en [23], *“caracterizan los cambios de amplitud en el marco del análisis de modulación de amplitud, proporcionando una representación de periodicidades de segundo orden, es decir, frecuencias de modulación.”* lo anterior es relevante cuando se desea realizar un análisis de componentes espectrales que no son visibles a simple vista. El espectrograma de modulación se define mediante la siguiente ecuación:

$$X(t, f_{mod}) = \mathcal{F}_t \{|X(t, f)|\} \quad (2-10)$$

El espectrograma de modulación consiste en calcular una doble transformada sobre una señal $x(t)$. Como se detallará más adelante el presente trabajo, utilizará una librería para el calcular los espectrogramas de modulación, la siguiente imagen muestra un resumen de las diferentes técnicas de análisis espectral analizadas (ver *Figura 10*). Más adelante se presentan otros detalles acerca del uso de esta técnica (ver *Figura 27*).

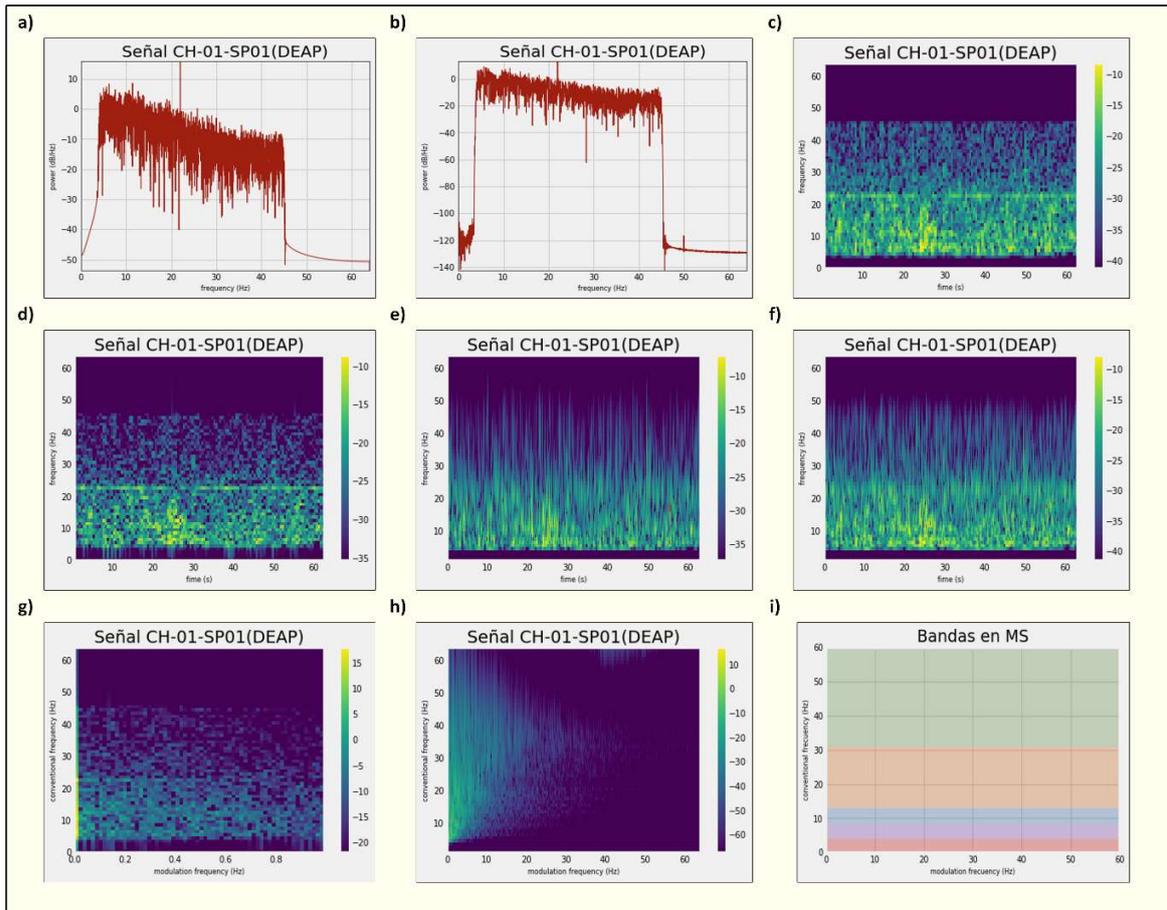


Figura 10: Pruebas técnicas de los espectrogramas. a) muestra es un PSD calculado a partir de una ventana tipo boxcar. b) muestra es un PSD calculado a partir de una ventana tipo Blackman-Harris. c) enseña un espectrograma calculado a partir de STFT con una ventana Hanning. d) muestra un espectrograma calculado utilizando una ventana tipo boxcar. e) y f) muestran espectrogramas calculados con la técnica CWT utilizando 6 y 10 ciclos de wavelet respectivamente, nótese como se mejora la resolución de frecuencias a mayor número de ciclos. g) y h) representan los espectrogramas de modulación, en este caso desaparece la dimensión temporal en el eje horizontal y pasa a ser un gráfico frecuencia-frecuencia. i) muestra la vista lógica del espectrograma de modulación empleada en el presente trabajo, cada color pertenece a una de las bandas descritas anteriormente en su respectivo orden [delta, theta, alpha, beta, gamma]. (fuente propia)

2.5 EEG

Desde el punto de vista estructural y del almacenamiento de datos, el EEG está compuesto por un conjunto de series de tiempo discriminadas según los electrodos activos al momento de la captura, cada serie de electrodos también llamados canales, representa el potencial generado por la actividad cerebral durante el experimento.

Existen diferentes metodologías para la adquisición de las señales a través del dispositivo EEG como, por ejemplo: bipolar y referencial [25], siendo esta última la más utilizada por la mayoría de los dispositivos BCI. Esta técnica adquiere su nombre debido a que utiliza una señal de referencia para calcular el potencial de las lecturas durante el registro de cada uno de los canales. Uno de los orígenes de la señal de referencia más implementado es el lóbulo de la oreja, debido a que dicho punto está aislado de otros artefactos de señales como los originados por el corazón ECG. Sin embargo, este origen también presenta desventajas, ya que se puede incurrir en contaminación por artefactos cuando la oreja elegida está cerca del punto de actividad. Por ejemplo: el punto de referencia en el lóbulo de la oreja derecha queda cerca del canal T4. Para minimizar este inconveniente se implementa una técnica de punto de referencia cruzado.

Una lectura fundamentada en la técnica bipolar se caracteriza porque se calcula el potencial con base en los electrodos cercanos, conformando una especie de par ordenado, de esta forma el potencial leído puede tomar un valor positivo o negativo según el marco de referencia.

Desde el punto de vista informático, los arreglos bidimensionales se acoplan con la estructura de un EEG, donde las filas están constituidas por una combinación entre unidades de tiempo y el respectivo identificador de muestra. Las columnas discriminan los diferentes canales activos al momento de la toma de las señales. Según el dispositivo BCI, también se podría contar con información adicional, originada por giroscopios o acelerómetros.

En relación con el almacenamiento de los datos, existen diferentes formatos de archivos, algunos surgen como estándares, por ejemplo: los EDF [26], otros son específicos según el fabricante de la plataforma BCI, como los BDF [27]; también existe la posibilidad de almacenar los datos en un formato de archivo basado en texto plano. Por otro lado, los registros de un EEG son series de tiempo segmentadas por cada uno de los canales o componentes del BCI (ver *Figura 11*). La serie de tiempo de un canal específico estará definida por la cantidad de segundos de la prueba multiplicada por la frecuencia de muestreo, la cual es configurada previamente al experimento.

Una adquisición EEG se puede componer de varios archivos, por ejemplo: un archivo de datos, tal como el que se describió anteriormente y un archivo de anotaciones de eventos de estimulación, por ejemplo, también se pueden adicionar otros archivos según el contexto del estudio que se esté llevando a cabo; en este caso dicha información es relevante al momento de analizar las señales.

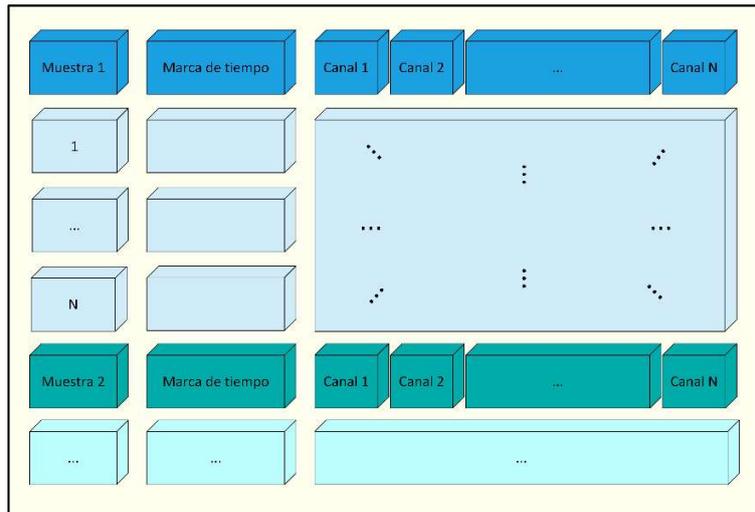


Figura 11: Visión general de la estructura de un EEG. Desde el punto de vista del almacenamiento de los archivos, esta es una de las disposiciones más comunes de los datos provenientes de los EEG, en esta figura no se observa otros aspectos para tener en cuenta, como la codificación y el formato del archivo. (fuente propia).

2.6 Flujo de trabajo para el procesamiento de señales EEG

Dependiendo del objeto de estudio y la aplicación a desarrollar, existen diferentes maneras de abordar el análisis de las señales EEG [28]; se puede definir un flujo de trabajo tomando como base un modelo general de procesamiento de datos, por ejemplo: *machine learning*. (ver *Figura 12*). A continuación, se describen cada una de las fases de un flujo de trabajo para el análisis de señales EEG.

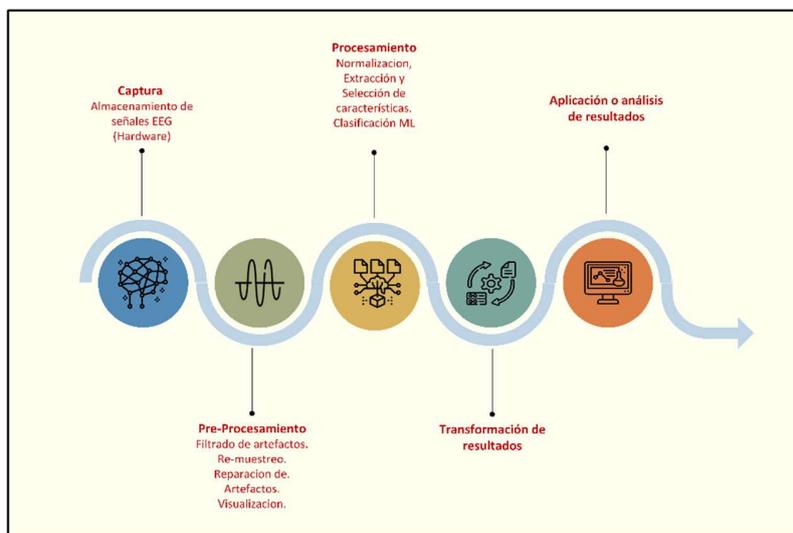


Figura 12: Flujo convencional de trabajo para el procesamiento de señales EEG, los flujos de trabajo dependen del enfoque y de la forma en la que se desean aplicar los resultados, no todos los trabajos referentes al análisis de señales EEG implican el uso de técnicas de ML. (fuente propia).

Captura: En esta fase se realiza una planeación detallada de la ubicación de los electrodos y los estímulos que se van a aplicar, también incluye la preparación y colocación de los diferentes electrodos o la estructura de montaje del BCI dependiendo del caso. Las señales EEG están intervenidas por diversos tipos de ruido, por ejemplo: ruido presente en la impedancia del electrodo (50-60Hz)[9]. El concepto de artefacto categoriza el conjunto de ruidos que afectan las señales EEG, el origen de dichos artefactos es diverso, por ejemplo: el movimiento natural del sujeto da lugar a la generación de una señal EMG, esta señal es detectada por el dispositivo de captura y se mezcla con el registro del EEG, lo que produce variaciones a nivel temporal y espacial [11]. Otras señales también pueden ser captadas por los electrodos, tal es el caso de los latidos del corazón ECG. En síntesis, mientras más cerca sea el origen del artefacto, este afectará en mayor medida las lecturas del EEG [29] por lo que los movimientos de los músculos de la cara o el funcionamiento del aparato oculomotor EOG presentan una gran incidencia en los registros.

El origen de los artefactos se relaciona tanto con el tipo de estímulo como con el objeto de estudio. Se puede inferir que mientras más inmóvil esté el sujeto, se presentan menos artefactos producidos por movimientos[29], sin embargo, esto no es aplicable en todos los casos. Otro tipo de experimentos pueden introducir sesgos de carácter cognitivo, los cuales no necesariamente afectarán la calidad de la señal en sí mismos, pero reducirán la confiabilidad de los análisis posteriores a la captura. Un ejemplo de dichos estudios son los relacionados con el campo de la psicología, una hipótesis relacionada con un proceso cognitivo que dure varios segundos afectará la resolución temporal del análisis EEG [30]. Otras afectaciones pueden provenir desde el mismo protocolo de experimentación, es el caso de los tiempos de espera entre cada uno de los estímulos, la pérdida de concentración por parte del sujeto, la variabilidad de los conjuntos de estímulos, entre otros.

Desde la experimentación, una forma de mejorar la calidad de las capturas EEG es capacitar a los sujetos involucrados en la prueba; tratar de evitar hablar, tocarse la cara, mirar hacia los lados; de esta forma se evita que se generen artefactos en la lectura; no obstante, se debe tener en cuenta que un amplio abanico de restricciones podría generar los sesgos mencionados anteriormente, es decir, el sujeto está tan concentrado en evitar los movimientos que desatiende el estímulo presentado y de paso el objetivo general del estudio[31].

Tabla 2-2: Algunos artefactos y su origen.

Artefacto	Origen
Parpadeo	EOG
Guiño a la izquierda / derecha	EOG
Mirada a la izquierda / derecha	EOG
Surco (ceño fruncido)	EMG
Levanta la ceja (sorpresa)	EMG
Sonreír	EMG
Apretar los dientes (mueca)	EMG
Risa	EMG
Sonrisa izquierda / derecha	EMG
Masticar	EMG
Latido del corazón	ECG
Disparo del electrodo	Eléctrico
Artefacto por sudor	Sudor

Preprocesamiento: Esta es la segunda fase del flujo de trabajo en el análisis de señales EEG, en ella se abarca cualquier tipo de tratamiento previo que se realice sobre los datos arrojados al momento de la captura, por ejemplo: cambios organizacionales relacionados con la disposición inicial de los datos, el filtrado de algunos canales, la aplicación de técnicas de *resampling*, la segregación de los canales en conjuntos según las diferentes zonas del cerebro (frontal, parietal, occipital), la extracción de características de la señal, la eliminación del ruido, entre otros [32].

Aunque las señales EEG son una mezcla entre información y ruido, remover solamente el ruido es particularmente inviable ya que por lo general su eliminación implica la pérdida de información, lo que se pretende es mitigar el ruido con técnicas de filtrado o en casos extremos posiblemente sea preferible eliminar secciones del EEG [33]. Para tal fin, existen una amplia variedad de técnicas (ver *Figura 13*).

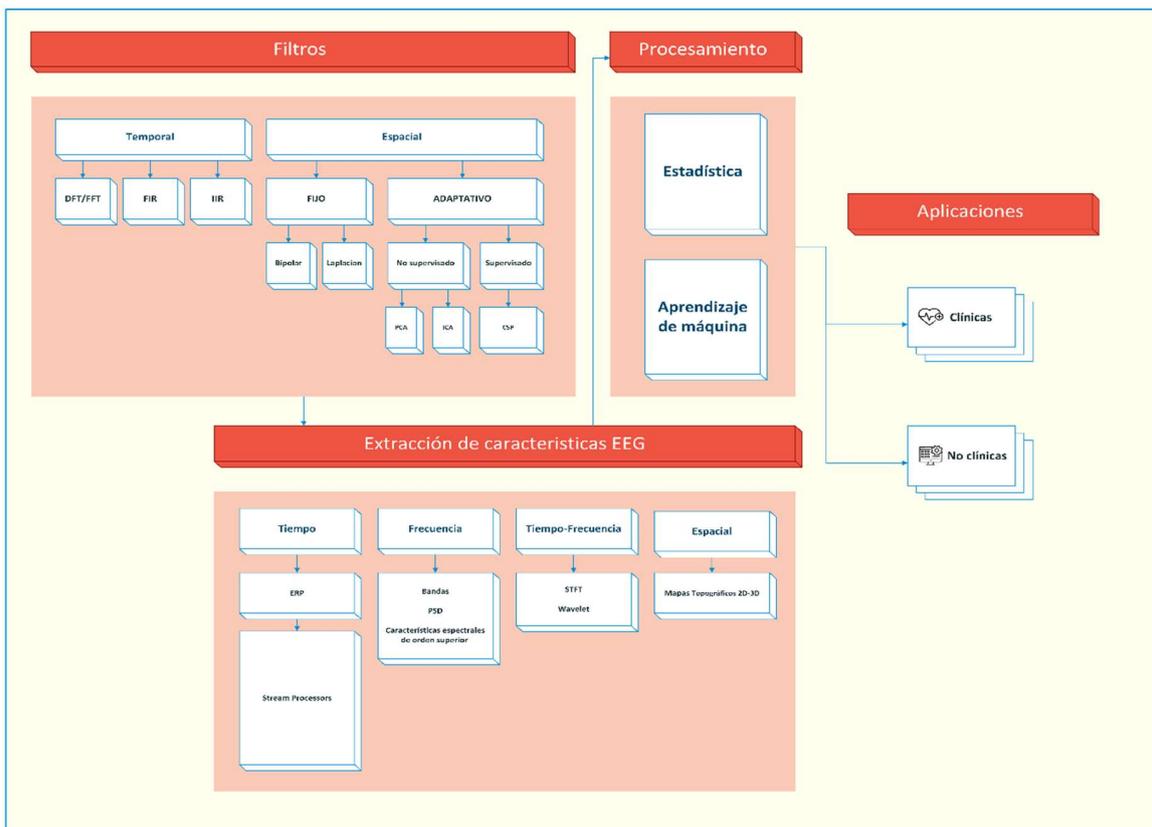


Figura 13: Resumen de técnicas preprocesamiento. Esta figura muestra los diferentes enfoques utilizados en la fase de preprocesamiento, adicionalmente también muestra algunas técnicas para la fase de procesamiento, la aplicación de cada una estas dependerán de las hipótesis y el contexto del estudio. (figura obtenida y modificada de [28]).

Vale la pena destacar la aparición de DFT y FFT entre las técnicas de filtrado, estas técnicas basadas en el análisis espectral son de especial interés sobre todo cuando se intenta explorar las bandas de frecuencias del EEG mencionadas anteriormente (delta, theta, alfa beta, gamma), ya sea por varios

motivos, entre ellos la necesidad de adicionar concepto desde la neurociencia al análisis de las señales. Sin embargo, la extracción de bandas de frecuencias se considera como una tarea de extracción de características [28]. La técnica de espectrograma de modulación descrita en [23] permite segmentar regiones de los componentes espectrales en una banda de frecuencia específica.

Procesamiento: Esta fase depende del enfoque y la tarea de clasificación elegida dentro del objeto de estudio; existe una amplia gama de técnicas, basadas en diferentes disciplinas y conceptos, por ejemplo: enfoques basados en estadística [34], basados en aprendizaje supervisado [35], en aprendizaje no supervisado [36] y el que será esencial para el presente estudio, los basados en aprendizaje profundo que a su vez hacen parte del aprendizaje supervisado [34].

Transformación y aplicación de resultados: Esta fase consiste en adaptar los resultados producto de la fase de procesamiento, para que éstas sean consumidas por una aplicación de software o un dispositivo de hardware; las cuales, podrían hacer parte del BCI directamente, por ejemplo: los componentes biónicos, existen muchas posibles aplicaciones para los resultados de un análisis EEG [28], [37], [38], como se muestra a continuación:

Tabla 2-3: Aplicaciones del procesamiento EEG.

Segmento	Aplicación
Clínico	Dispositivos de comunicación asistida basados en BCI
	Dispositivos de movilidad asistida basados en BCI
	Neuro-rehabilitación
	Diagnóstico médico
	Análisis de estados cognitivos
	Análisis de estados emocionales
No clínicas	Neuro-ergonomía
	Domótica - hogares inteligentes
	Neuro-marketing
	Entretenimiento
	Sistemas de seguridad

2.7 Transformers

Dentro del aprendizaje profundo, existe una técnica basada en *Transformers* [39], la cual surgió en el contexto del procesamiento de lenguaje natural, como una alternativa para resolver el problema de traducción de un texto entre idiomas; una de sus características más relevantes es su capacidad de paralelización, lo que hace que se pueda aumentar el número de parámetros y a la vez reducir el tiempo de entrenamiento. Para entender el concepto de *Transformers* primero es necesario abarcar conceptos como:

- Aprendizaje profundo (DL) por sus siglas en inglés *Deep Learning*.
- Redes neuronales recurrentes (RNN) por sus siglas en inglés *Recurrent Neural Networks*.

- Redes neuronales con propagación hacia adelante (FFNN) por sus siglas en inglés *Feed Forward Neural Networks*.
- Redes neuronales convolucionales (CNN) por sus siglas en inglés *Convolutional Neural Networks*.

Lo anterior se debe a que los *Transformers* son una evolución de las redes neuronales recurrentes. El problema con esta técnica es que para que la recurrencia sea efectiva, el procesamiento de cada una de las capas debe ser secuencial, teniendo como referencia el resultado de la ejecución anterior en el procesamiento actual, de tal forma que el modelo llegue a aprender la relación entre cada una de las entradas. Esto también representa un problema en el aprendizaje por la memoria entre las relaciones de los componentes de entrada, debido a que para secuencias de entrada largas el modelo tiene propensión a olvidar la relación entre el primer y el último elemento. Para el problema de la secuencialidad el *Transformer* encontró una solución, procesando las entradas y las salidas de forma simultánea (ver *Figura 14*). En cuanto a la problemática de la memoria, el *Transformer* implementó un mecanismo llamado *self-attention* [39], el cual consiste en que todos los componentes de las entradas interactúen entre sí; básicamente esto se logra a través de operaciones matemáticas sobre tres estructuras de pesos; en este caso, un arreglo de claves llamado K, un arreglo de valores V, y un arreglo de consultas llamado Q, el resultado obtenido representa la interacción de cada una de las entradas incluyendo consigo misma.

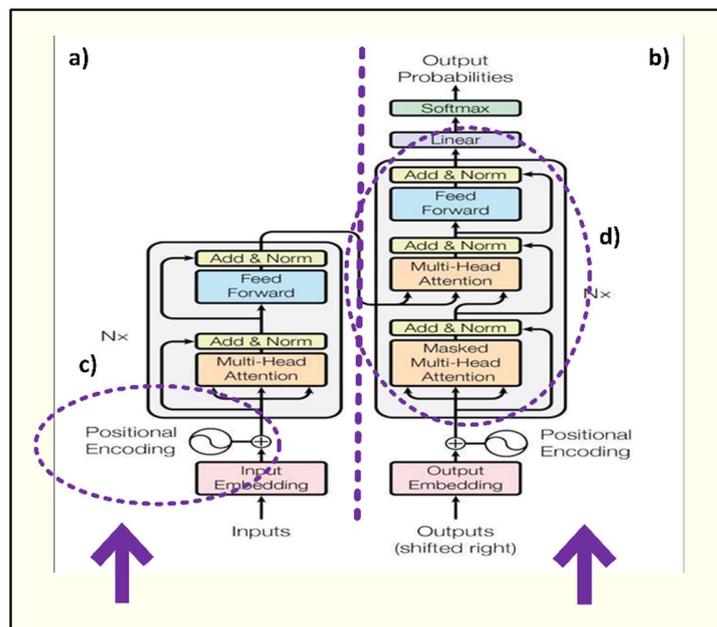


Figura 14: *Arquitectura del Transformer*, se puede apreciar que (a) y (b) son componentes que están divididos: uno para el procesamiento de las entradas y otro para el procesamiento de las salidas. c) muestra el codificador posicional. d) muestra lo que se considera el bloque de Transformer, nótese que internamente cada bloque posee al menos una capa de atención. (figura obtenida y modificada de [39]).

3 Trabajos relacionados

El aspecto central del presente trabajo, el cual consiste en el procesamiento de señales EEG mediante el uso de modelos de aprendizaje profundo basados en Transformer, constituye un enfoque novedoso que está en pleno desarrollo. Si se revisa la documentación disponible (ver *Figura 15*), al momento de la implementación del modelo, solo se contaba con un poco más de cinco referencias, lo cual representó un gran reto tanto en investigación, como en el desarrollo de software. Se puede observar en la siguiente figura que gran parte de referencias adicionales han sido publicadas en el cuarto trimestre del 2022, estas no fueron tomadas en el presente trabajo debido a que no coincidieron con la etapa de levantamiento del estado del arte.

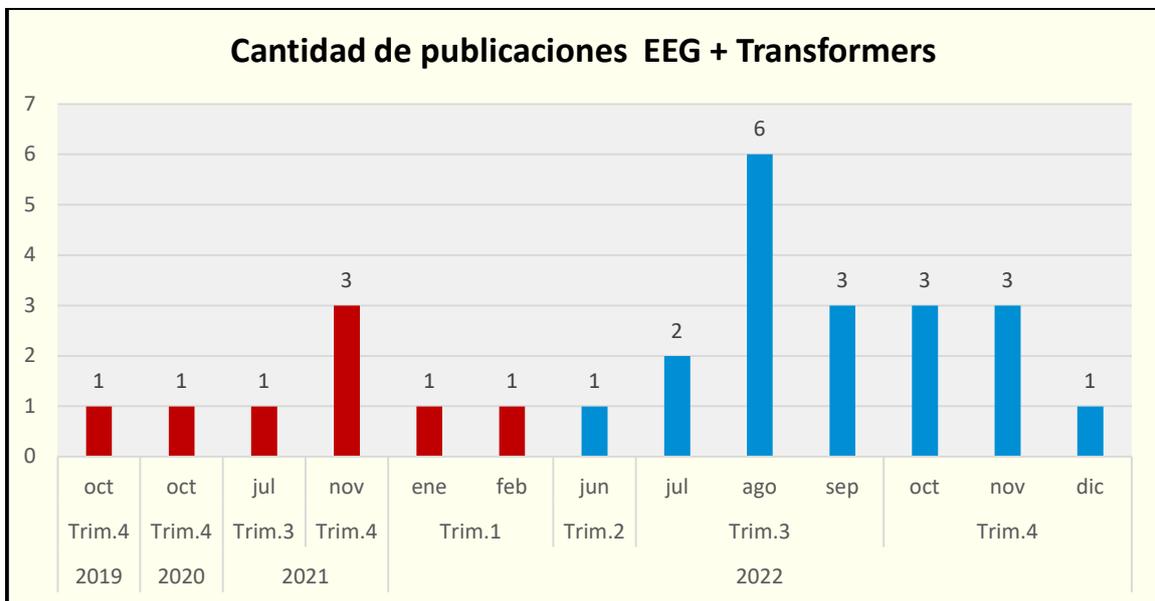


Figura 15: Historial de los trabajos relacionados con EEG y Transformers, estos datos fueron extraídos del portal PubMed®, el cual recopila millones de artículos de biomedicina concentrados en la base de datos MEDLINE que incluye referencias de grandes editores en ciencias de la vida. En rojo se destaca el lapso de tiempo en el cual, se realizó el compendio bibliográfico para el presente trabajo. (figura obtenida y modificada de[40]).

3.1 Espectrogramas

Aunque los espectrogramas aparecen en la década de los 60s, el presente trabajo toma como base lo presentado en [23], en el cual se establece la metodología para el cálculo de los espectrogramas de modulación (SM). También presenta sus principales ventajas y usos incluido el análisis de señales EEG, posteriormente en [41], se describe el SM como un método que facilita el filtrado de características propias de la señal EEG con el propósito de aplicar una análisis posterior, ambos

estudios son de gran relevancia para el presente trabajo ya que darán forma a la fase de procesamiento de las señales EEG.

3.2 Deep Learning

Dejando aparte otros tipos de enfoques como los métodos RF (de sus siglas en inglés *Random Forest*) [42], [43], SVM (de sus siglas en inglés *Support Vector Machine*) [44], K-MEANS [5] de los cuales existe un volumen importante de artículos, cuyo auge se presentó en la década anterior; nuestro interés crece hacia la documentación relacionada con el tema de *Transformers*. Dicho concepto fue presentado en el famoso artículo “*Attention is all you need*” [39], y tal como como se había expuesto anteriormente, su objetivo principal era resolver un problema de procesamiento de lenguaje natural basado en la traducción de texto del inglés al alemán y del inglés al francés. Como conclusión en este artículo, además de dar solución al objetivo propuesto, el método demuestra su principal ventaja, la optimización, ya que permite mejorar sustancialmente el proceso de entrenamiento debido a su alto grado de paralelización.

Al mismo tiempo, al estudio de los *Transformes*, también encuentran otros enfoques; por ejemplo en [45], se especifica que un análisis EEG requiere tener tres aspectos principales en mente, el espacio, el tiempo y el espectro; en este caso, se utilizó un enfoque mixto basado en CNN y LSTM, los cuales logran conjugar estos tres aspectos.

La *Figura 16* muestra como desde el preprocesamiento se separan las diferentes bandas que componen la señal de cada uno de los canales del EEG, luego estos se reorganizan conservando la posición de cada uno de los canales mediante un arreglo 2D; el objetivo de esta disposición es conservar la información espacial dentro del modelo, al final se aplica una segmentación a cada uno de estos arreglos 2D en tajadas de tiempos conservando de esta forma el aspecto temporal.

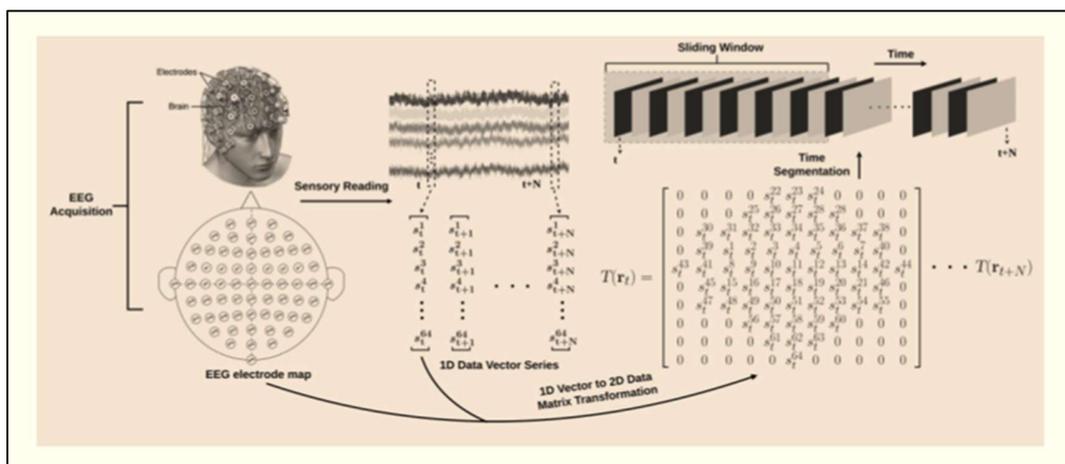


Figura 16: “EEG data acquisition and preprocessing”. En este estudio se puede apreciar cómo se realiza una distribución particular de los valores del EEG en una matriz, conservando la ubicación de cada uno de los canales según el mapa topográfico del cráneo (figura obtenida y modificada de [45]).

Existen otros trabajos basados en CNN [46], [47], los cuales fueron abarcados con una menor profundidad, debido a que el tema de las redes convolucionales se retomará más adelante en el estudio del artículo relacionado con ViT (de sus siglas en inglés *Vision Transformer*). En este trabajo se desarrolla el preprocesamiento de forma interesante; en este caso, los datos de las frecuencias de la EEG son mapeados en una imagen en formato RGB, donde se distinguen aspectos de espacialidad y de frecuencia (ver *Figura 17*)

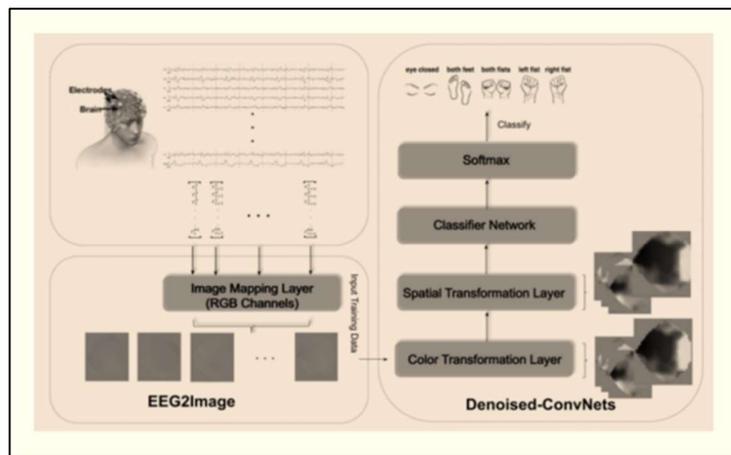


Figura 17: “Workflow of the proposed model EID”, en este estudio los componentes espectrales de la señal son convertidos a imágenes RGB para luego ser procesados. (figura obtenida y modificada de [47]).

Los artículos mencionados anteriormente son de relevancia para la investigación debido a que inspiran de cierta forma la etapa de preprocesamiento.

Con respecto a la utilización de *Transformers* se revisaron diferentes estudios, haciendo una consideración especial en los relacionados con el procesamiento de señales EEG, como por ejemplo [48]; en este estudio se propone una actualización del *Transformer* añadiendo una capa tipo compuerta en su implementación interna (ver *Figura 18*); lo anterior, con el objetivo de mejorar las conexiones residuales y de esta forma estabilizar el entrenamiento; por otra parte, en este trabajo se aplica un preprocesamiento que no va más allá de la aplicación de unos cuantos filtros para retirar artefactos y componentes no deseados, lo anterior se debe a que el modelo utiliza la señal de los canales EEG como entrada. Esto presenta ventajas y desventajas, ya que la señal del canal está descrita en términos de tiempo y al ser procesada por el *Transformer* sin ningún tipo de organización adicional se está discriminando el componente espacial dentro del análisis EEG; en consecuencia, deja desprovisto al modelo del aspecto de la proximidad bajo la perspectiva de un sistema de colocación de electrodos como el 10-20, también deja de lado algunas premisas desde el punto de vista de la neurociencia como es el caso de la lateralización, la cual se puede emplear en determinados análisis. Otro aspecto a denotar en este enfoque es que la naturaleza del *Transformer* obliga a la fase de entrenamiento a buscar una relación de un componente contra sus contrapartes a lo largo de toda la entrada, por lo que aquí entra a jugar el número de electrodos y la frecuencia de muestreo como un parámetro de limitación en arquitecturas de hardware de bajo rendimiento, en este punto no se consigue realizar un análisis más profundo debido a que el estudio no se hace mención al hardware utilizado en la fase entrenamiento.

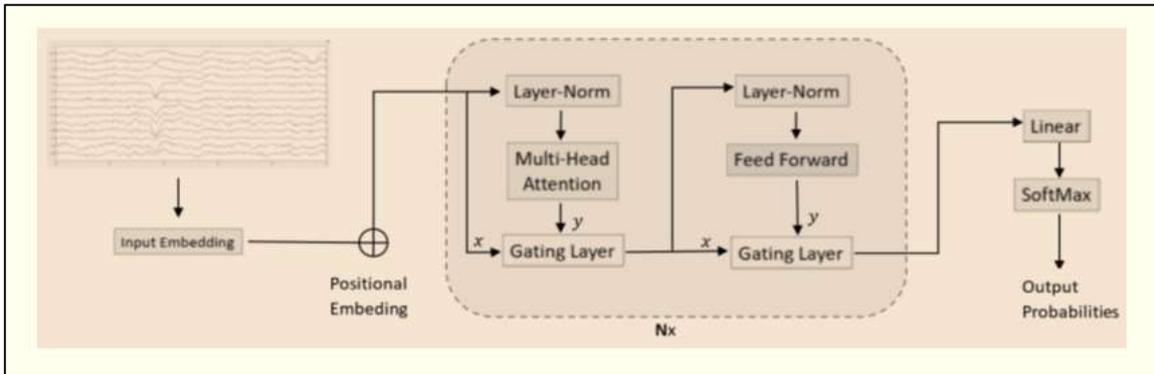


Figura 18: “The graphical illustration of the gated Transformer architecture”, en esta imagen se puede apreciar la arquitectura de esta propuesta que agrega una capa de gating layer a la capa de Transformer. (figura obtenida y modificada de[48]).

Otro trabajo basado en mecanismos de atención como [49] utiliza la codificación espacial descrita en [45], en este caso el desarrollo adiciona una robusta fase de preprocesamiento que involucra una red neuronal convolucional con un kernel 3D, luego utiliza capas sucesivas de atención (ver *Figura 19*), esta idea se utilizará posteriormente en la fase de implementación del presente trabajo.

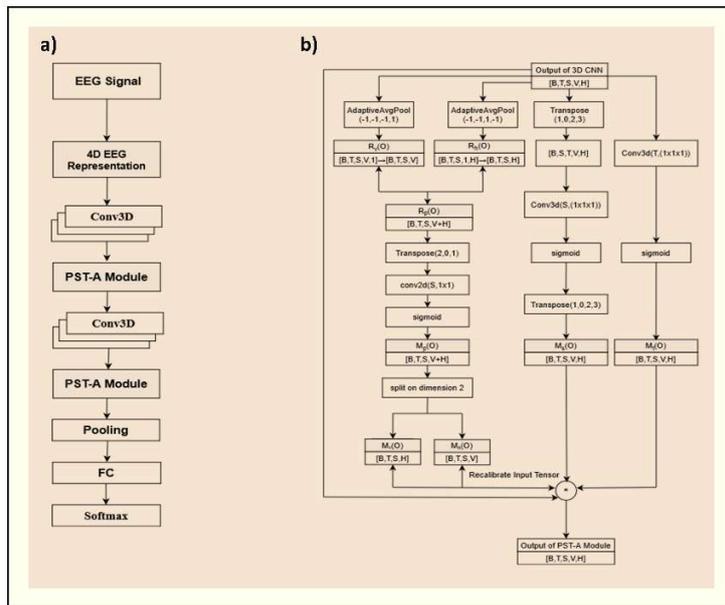


Figura 19: Descripción y uso del módulo PST-A, a) muestra el empleo de la capa PST-A (de sus siglas en inglés, Positional-Spectral-Temporal Attention) en un modelo para el procesamiento de señales EEG. b) expone los componentes de PST-A. (figura obtenida y modificada de[49]).

Como se expuso en el análisis de [39], los *Transformers* se adaptan muy bien al procesamiento de secuencias de elementos en los problemas de análisis de texto; sin embargo, esta técnica no se había desarrollado para el procesamiento de imágenes ya que estas conllevan otros aspectos a tener

en cuenta; en este caso, se analizó un artículo de referencia importante para el presente trabajo [50], el cual expone el concepto de ViT (de sus siglas en inglés, *Vision Transformer*). El ViT le agrega al modelo de *Transformer* descrito en [39] una capa de preprocesamiento, que consiste en la división de la imagen original en sub-imágenes, a las que luego se le aplica una convolución con el fin de extraer los componentes o características principales de dichas sub-imágenes, reduciendo de esta forma la demanda computacional que abarcaría procesar la imagen completa, en este modelo también se modifica el proceso del *positional embedding* expuesto en [39]; con el fin de que el modelo logre aprender la relación de los fragmentos de imágenes colindantes.

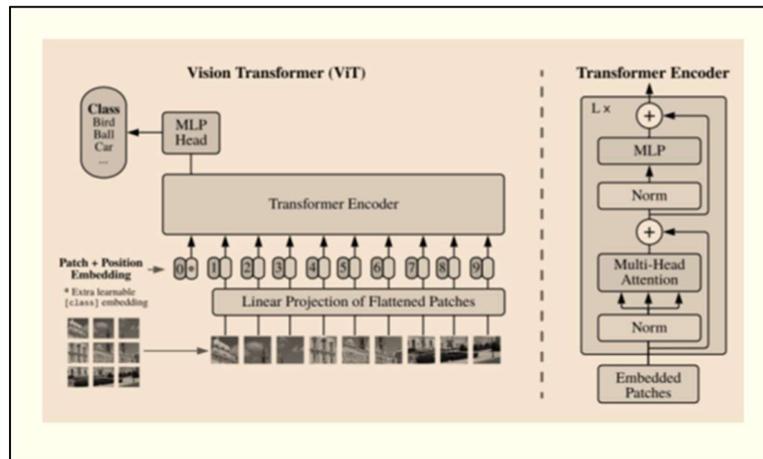


Figura 20: Arquitectura del ViT, en la imagen se puede apreciar como la entrada es dividida en sub-imágenes para luego conseguir una proyección lineal a través de una capa de convolución. (figura obtenida y modificada de[50]).

Como se puede apreciar en (Figura 20), el ViT no es un modelo seqToSeq propiamente, en su lugar, añade una entrada de clase en el proceso de *embedding*, la cual se utiliza posteriormente como parte de una clasificación categórica.

Finalmente, el enfoque en [51] llamado HSLT (de sus siglas en inglés Hierarchical Spatial learning Transformer), fue utilizado como base para el presente trabajo; este utiliza el modelo ViT como unidad de procesamiento de los PSD para diferentes conjuntos de señales EEG, creados a partir de la segmentación según regiones del cerebro específicas (ver Figura 21). Este enfoque tiene aspectos para destacar, como el hecho de que el resultado de la aplicación de un método de análisis espectral deja como resultado un arreglo 2D que es más compatible con un modelo basado en imágenes; por otra parte, la segmentación en diferentes conjuntos de los canales provenientes del EEG le añade una relevancia espacial al modelo.

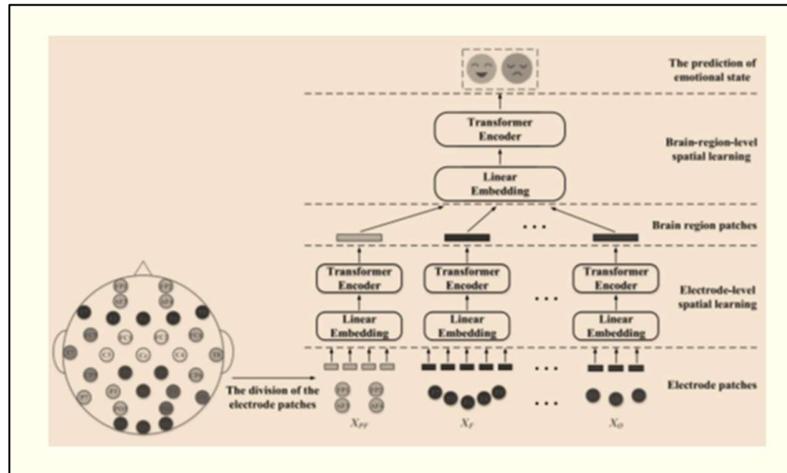


Figura 21: "The overview of the HSLT", en la imagen se expone un modelo que utiliza presunciones basadas en la neurociencia; para su diseño, en este caso se puede observar la segmentación del EEG en conjuntos según las regiones del de cerebro que se describieron anteriormente, también añade el concepto de lateralización al dividir dichas regiones en izquierda y derecha. (figura obtenida y modificada de [51])

4 Metodología

En el presente trabajo se abordará una notación desde el punto de vista de tensores, para representar los resultados de las capas principales del modelo; sin embargo, vale la pena aclarar el que no se pretende abordar el concepto de tensores desde el punto de vista de la definición físico-matemática propiamente, sino que se realiza una aproximación desde el ámbito computacional, tal como lo llega a exponer[52].

4.1 Preprocesamiento

Los estudios relacionados con EEG generan conjuntos de datos con tres características principales, un número de canales n , un número de ensayos p , una frecuencia de muestreo sf y una duración de experimento t . Teniendo en cuenta lo anterior, un solo ensayo EEG se define según:

$$x_t = [x_1, x_2, x_3, \dots, x_n] \in \mathbb{R}^{n \times t \times sf} \quad (4-1)$$

Donde, $t \times sf$ hace referencia a la dimensión temporal que dependerá tanto de la duración del experimento como de la frecuencia de muestreo configurada en la fase de adquisición del EEG, vale la pena aclarar que en algunos conjuntos de datos, la duración en unidades de segundos no es la misma entre participantes; por lo que, para efectos del presente trabajo, se discriminará este aspecto en cuando a la definición metodológica, ya que en la fase de desarrollo se realizarán los respectivos ajustes para contar con una longitud homogénea entre las diferentes mediciones. También se retirará las líneas base en los experimentos donde se haya definido.

Teniendo como insumo (4-1) se aplicará el cálculo del espectrograma de modulación para cada uno de los canales, dando como resultado un objeto definido por:

$$X_f = [x_1, x_2, x_3, \dots, x_n] \in \mathbb{R}^{n \times cf \times mf} \quad (4-2)$$

En este caso cada espectrograma de modulación tiene una dimensión $cf \times mf$. Donde la longitud de cf está definida por las frecuencias convencionales obtenidas de la señal original, cabe destacar que como dicha señal proviene de un EEG dichas frecuencias estarán delimitadas por bandas reconocidas en la actividad neurológica $b = [\delta, \theta, \alpha, \beta, \gamma]$. Es decir, que en la práctica cf estará limitado por el rango $(0, \sim 60) Hz$. La longitud de mf estará asociada a la duración de la señal en la fase de adquisición y a su respectiva frecuencia de muestreo.

Por otra parte, utilizando la propiedad de los espectrogramas de modulación mencionada en [41], se aplicará una extracción de un área específica del espectrograma originalmente obtenido $cf \times mf$, obteniendo, $cf' \times mf'$ tal que $cf' \times mf' < cf \times mf$. Con esto se consigue descartar

de facto los artefactos que estén por encima de 60 Hz, como por ejemplo el ruido ambiental y parte de los EMG, lo anterior tomando como ejemplo un rango para cf' de [0-60].

Un espectrograma de modulación es un arreglo de números complejos, aunque ya hay estudios que utilizan la técnica de *Transformers* bajo este conjunto numérico [53], en el presente trabajo se limitará al procesamiento de arreglos de números reales, por lo que se trabajará con la potencia calculada a partir de la técnica de espectrogramas.

Un conjunto completo de datos se compondrá de múltiples arreglos X_f de la siguiente forma:

$$S_f = [X_f^1, X_f^2, \dots, X_f^p] \in \mathbb{R}^{p \times n \times cf \times mf} \quad (4-3)$$

Donde n y p se habían descrito anteriormente como el número de canales y el número de ensayos respectivamente. Como alternativa a esta parte del proceso se diseñó una simplificación del espectrograma, la cual se describe en detalle en la fase de implementación. En este caso se redefinirían las ecuaciones (4-2) y (4-3) de la siguiente manera:

$$X_f \in \mathbb{R}^{n \times 5 \times mf} \quad (4-4)$$

$$S_f = [X_f^1, X_f^2, \dots, X_f^p] \in \mathbb{R}^{p \times n \times (5 \times mf)} \quad (4-5)$$

En este caso 5 representa el número de bandas $b = [\delta, \theta, \alpha, \beta, \gamma]$, en las que se divide horizontalmente el espectrograma de modulación.

Por otra parte, en cuanto a las etiquetas para la tarea de clasificación, estas dependerán del conjunto de datos, no obstante, se generarán utilizando una codificación estilo *one-hot* en la representación de cada una de estas clases, dando como resultado una estructura de la siguiente forma:

$$y_{cls} = [y_1, y_2, \dots, y_c] \in \mathbb{R}^c \quad (4-6)$$

Donde c es el número de clases. Por otro lado, de la misma forma que (4-3), el conjunto completo de etiquetas se puede describir de la siguiente forma, tomando p como la cantidad de ensayos realizados.

$$S_{cls} = [y_{cls}^1, y_{cls}^2, \dots, y_{cls}^p] \in \mathbb{R}^{p \times c} \quad (4-7)$$

4.2 Modelo general (MST)

Al modelo general se le da el nombre MST (de sus siglas en inglés *Modulation Spectrogram Transformer*). Para este trabajo se construirá un modelo general basado en [50], [51]. En este caso se reutilizará la idea de segmentación de los canales del EEG en subconjuntos según las regiones del cerebro:

- PF: "prefrontal"
- F: "frontal"
- CL: "central izquierda"
- CR: "central derecha"
- TL: "temporal izquierdo"
- TD: "temporal derecho"
- PL: "parietal izquierda"
- PD: "parietal derecha"
- OL: "occipital izquierdo"
- OD: "occipital derecho"

En algunos análisis será necesario prescindir de algunas lateralizaciones y conjugarlas en una sola región, por ejemplo: $C = CL + CR$, esto debido al bajo número de electrodos utilizados en la fase de adquisición. Cada uno de estos subconjuntos define una entrada de *Transformer*, luego cada uno de estos resultados confluyen hacia una última capa de transformación; por lo que, el modelo general se puede representar como una sucesión de capas funcionales (ver [Figura 22](#)).

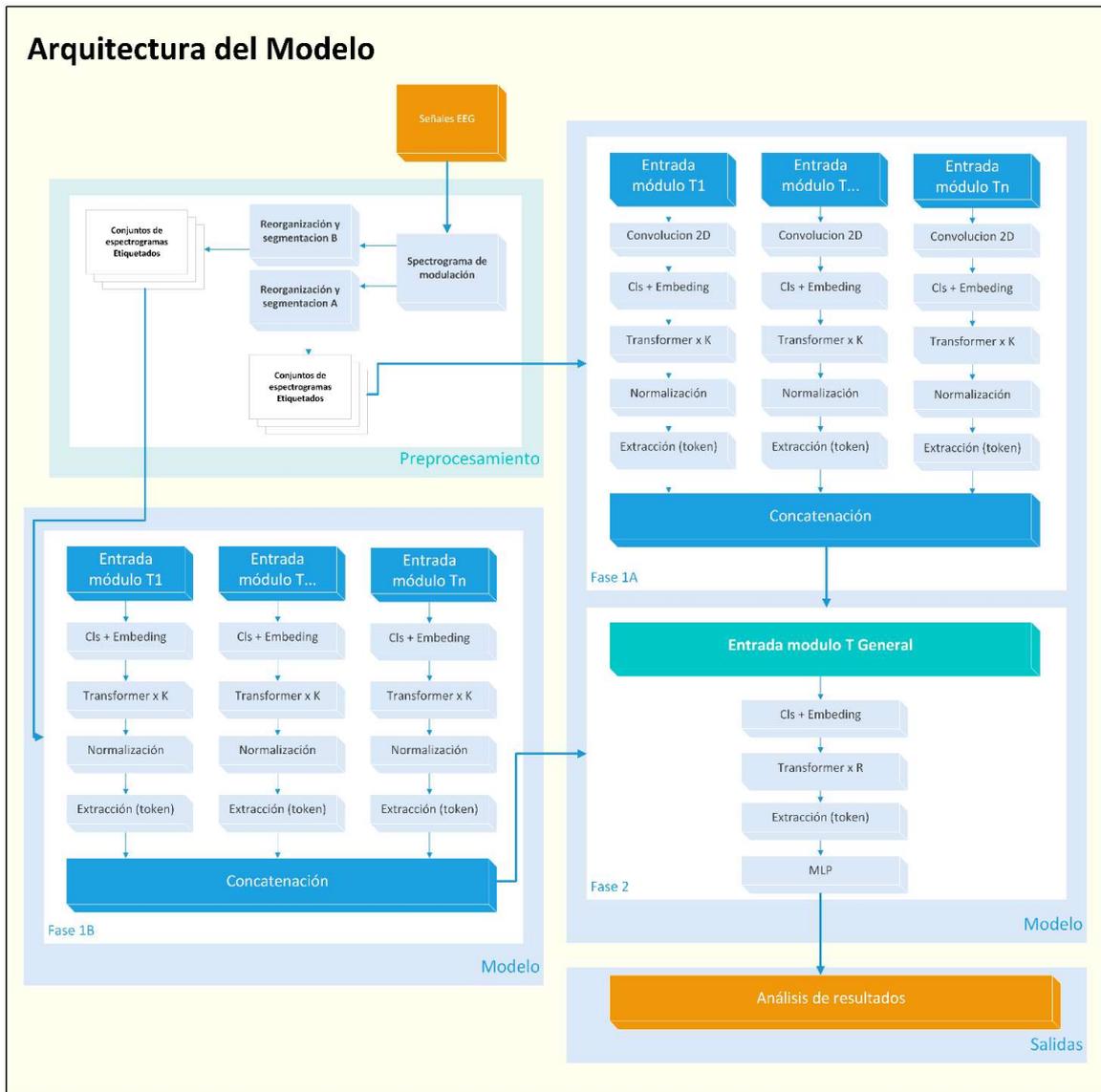


Figura 22: Modelo General (MST). Su implementación se compone de dos fases principalmente; sin embargo, vale la pena indicar que para la primera fase se tienen dos versiones, A y B las cuales se explicarán más adelante. (fuente propia).

Capa de transformación de entradas (Fase 1A): En este caso se realizará una reorganización del conjunto de datos de entrada (4-3), según:

$$X_R = [X_R^1, X_R^2, \dots, X_R^N] \in \mathbb{R}^{N \times cf \times mf} \quad (4-8)$$

Donde N es el número de canales incluidos en el subconjunto, como se aprecia en la imagen anterior (ver *Figura 22*), la primera parte del proceso realiza una convolución 2D sobre los espectrogramas

con el fin de capturar las características destacables de cada uno de ellos, esto se realiza con base en [50], dejando una estructura de la forma:

$$L_R = [L_R^1, L_R^2, \dots, L_R^N] \in \mathbb{R}^{N \times h \times w \times s} \quad (4-9)$$

En este caso $\mathbb{R}^{N \times h \times w \times s}$ está definido por la parametrización del kernel de convolución $h \times w$ (*patch_size*) y la parametrización de filtros s (*hidden_size*) propia de la capa de KERAS® *tf.keras.layers.Conv2D* (ver Anexo 9.J). El paso siguiente estará determinado por la aplicación de un proceso de *embedding* posicional. La cual añadirá una nueva entrada en la estructura definida en (4-9), quedando de esta forma:

$$E_R = [i_R^{cls}, i_R^1 L_R, i_R^2 L_R, \dots, i_R^N L_R] \in \mathbb{R}^{(N+1) \times h \times w \times s} \quad (4-10)$$

Luego de la aplicación del de la capa del *Transformer* estará determinada por:

$$X_T = [Z_T^{cls}, Z_T^1, Z_T^2, \dots, Z_T^N] \in \mathbb{R}^{(N+1) \times s} \quad (4-11)$$

Después de una capa de normalización se realizará la extracción de Z_T^{cls} para cada uno de los subconjuntos, para la siguiente fase del modelo.

Capa de concatenación de las subclases: Los resultados de cada uno de los procesos de transformación en cada uno de los subconjuntos EEG, se concatenarán para luego ser sometidas nuevamente a un procesamiento por *Transformer*, definiendo una nueva capa de entrada, la cual se describe a continuación:

$$X_d = [Z_T^{PE}, Z_T^F, Z_T^{CL}, Z_T^{CR}, Z_T^{TL}, Z_T^{TD}, Z_T^{PL}, Z_T^{PD}, Z_T^{OL}, Z_T^{OD}] \in \mathbb{R}^{N \times s} \quad (4-12)$$

Capa de transformación de entradas (Fase 1B): Esta fase del modelo es más sencilla que la definida para 1A, debido a que como se aplicó una simplificación del SM previamente, ya no será necesario utilizar la capa de convolución (ver Anexo 9.K), para este caso solo bastaría con reformular (4-8) y tener en cuenta que para este caso $h \times w \times s = 5 * mf$. El resto del proceso continuará igual (ver *Figura 22*).

$$X_R = [X_R^1, X_R^2, \dots, X_R^N] \in \mathbb{R}^{N \times 5 * mf} \quad (4-13)$$

Capa de transformación de subclases (Fase 2): El proceso de transformación de subclases, es muy similar a la capa de transformación de entradas de la Fase A, solo difiere la parte inicial del procesamiento, debido a que en este caso las entradas no son sometidas a una capa de convolución, las salidas representadas en (4-12) entran a una capa de *embedding* posicional similar a la definida en (4-10), para luego entrar a una serie de capas de *Transformer*, dejando como resultado una salida de la siguiente forma:

$$Z_d = [Z_d^{cls}, Z_d^{PE}, Z_d^F, Z_d^{CL}, Z_d^{CR}, Z_d^{TL}, Z_d^{TD}, Z_d^{PL}, Z_d^{PD}, Z_d^{OL}, Z_d^{OD}] \in \mathbb{R}^{(N+1) \times s} \quad (4-14)$$

De la salida reflejada en (4-14), se elegirá la clase Z_d^{cls} para la capa de salida.

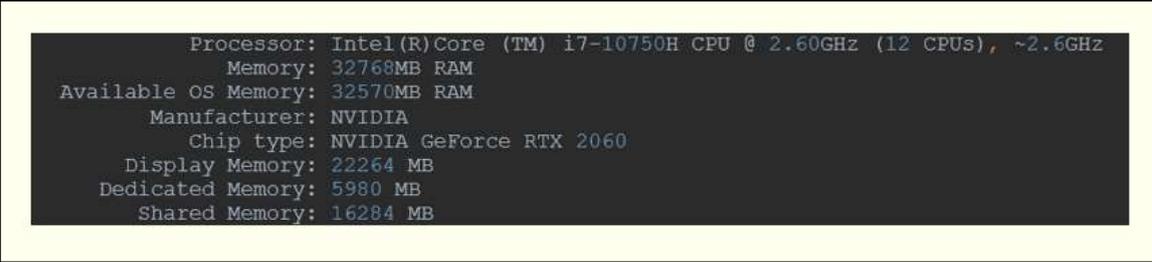
Capa de salida: Estará compuesta por una MLP (de sus siglas en inglés *Multi-Layer Perceptron*). Como se dijo anteriormente, la entrada está determinada por Z_T^{cls} y la última capa de esta red MLP se define como una capa *softmax* de longitud c según (4-7):

$$\hat{y}_{cls} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_c] \in \mathbb{R}^c \quad (4-15)$$

5 Desarrollo

5.1 Herramientas de hardware utilizadas

Para el entrenamiento del modelo se utilizó:

A screenshot of a terminal window displaying system hardware specifications. The text is as follows:

```
Processor: Intel(R)Core (TM) i7-10750H CPU @ 2.60GHz (12 CPUs), ~2.6GHz
Memory: 32768MB RAM
Available OS Memory: 32570MB RAM
Manufacturer: NVIDIA
Chip type: NVIDIA GeForce RTX 2060
Display Memory: 22264 MB
Dedicated Memory: 5980 MB
Shared Memory: 16284 MB
```

Figura 23: Detalles del hardware utilizado, (fuente propia).

5.2 Herramientas de software utilizadas

La fase de implementación se realizó a través del lenguaje de programación Python®[54]. En complemento con el siguiente conjunto de bibliotecas:

Tensorflow®[55]: Biblioteca base de la implementación del proyecto que proporciona el soporte necesario para el uso de tensores, además del soporte de las fases de entrenamiento, validación y cálculo de métricas de desempeño.

Keras®[56]: Biblioteca empleada en la implementación de alto nivel de cada una de las capas que componen al modelo.

AMA® (de sus siglas en inglés *Amplitude-Modulation-Analysis-module*)[57]: Librería de código abierto que da soporte al cálculo de los espectrogramas de modulación.

ViT (de sus siglas en inglés *visión Transformer*)[58]: Librería de código abierto que se utilizó en la implementación de la fase de transformación de entradas.

5.3 Conjuntos de datos seleccionados

La siguiente fase radica en verificar el desempeño del modelo general en la clasificación de las señales EEG para los siguientes procesos cerebrales:

- Clasificación de emociones.
- Diferenciación de estados en el desarrollo de cálculos matemáticos.
- Diferenciación del efecto Flicker[8], en relación a diferentes frecuencias de estímulo.

Conjunto de datos DEAP [59]: Elegido para evaluar el rendimiento del modelo en la clasificación de emociones, este es un conjunto de datos de acceso público de uso controlado, lo que significa que se debe realizar una solicitud al propietario. Este conjunto fue creado en el año de 2011 con la motivación inicial de identificar y clasificar las emociones provocadas en diferentes sujetos de prueba al momento de observar videos musicales, la pretensión inicial consistía en usar dicha clasificación para implementar un modelo de recomendación adaptativo [60]. No obstante, este conjunto de datos ha sido usado en múltiples trabajos con diferentes finalidades, como por ejemplo: en [5], convirtiéndose en uno de los conjuntos de datos referentes en estudios relacionados con la clasificación de emociones a través de EEG. Las características del conjunto de datos de cara al presente estudio son las siguientes:

Tabla 5-1: Descripción del conjunto de datos DEAP.

Característica	Valor
Número de participantes	32
Dimensiones de los datos	[valencia, excitación, dominio, gusto]
Categorización	Escalas con valores continuos
Datos preprocesados	Si
Número de canales EEG	32
Número de muestras por participante	40
Número total de muestras EEG extraíbles	1280
Frecuencia de muestreo	128 Hz
Duración de la muestra	63 segundos
Formato Archivo	Pickle (Python)

Conjunto de datos EEGMAT[61]: Elegido para evaluar el rendimiento del modelo en la clasificación de los estados durante la realización de tareas matemáticas, se encuentra disponible en el portal de Physionet [62]. En este conjunto se someten los participantes a dos momentos específicos, en el primero se adquieren las señales EEG cuando no se está realizando ninguna tarea matemática, en el segundo momento se realiza la adquisición cuando el participante está llevando a cabo un ejercicio matemático. Adicionalmente, los participantes se dividen en dos grupos; el primero llamado G, que corresponde a los participantes que realizan un buen conteo matemático y el segundo grupo llamado B, donde los participantes realizan un conteo matemático de forma deficiente. El presente estudio generará una etiquetación binaria a partir de este conjunto de datos; en este caso, una primera etiqueta indica que el participante está realizando cálculos matemáticos, la segunda se atribuye a un estado de reposo. Las características del conjunto de datos de cara al presente estudio son las siguientes:

Tabla 5-2: Descripción del conjunto de datos EEGMAT.

Característica	Comentario
Número de participantes	36
Categorización de los datos	[cálculo, reposo]
Datos preprocesados	Si
Número de canales EEG	19
Número de muestras por participante	2

Número total de muestras EEG extraíbles	214
Frecuencia de muestreo	500 Hz
Duración de la muestra	182 segundos
Formato Archivo	EDF/EDF+

Conjunto de datos MSSVEPDB [63]: Elegido para evaluar el rendimiento del modelo en la clasificación del efecto *Flicker*[8] bajo diferentes frecuencias, se encuentra disponible en el portal de Physionet [62]. En este experimento se estimula a un grupo de participantes con luces parpadeantes a diferentes frecuencias en cortos segmentos de tiempo, los cuales están separados en momentos de estimulación-descanso, que se dan de forma alternada. La etapa de experimentación se distribuye en tres etapas: reposo, adaptación y estimulación; por otra parte, el conjunto de datos completo consta de tres ciclos; En este caso, para el presente estudio solo utilizarán los dos primeros, debido a diferencias en la metodología aplicada con respecto al tercer ciclo, que hace dispendiosa la homologación en el preprocesamiento. Las características del conjunto de datos de cara al presente estudio son las siguientes:

Tabla 5-3: Descripción del conjunto de datos MSSVEPDB.

Característica	Comentario
Número de participantes	11
Categorización de los datos	[6.66Hz, 7.5Hz, 8.57Hz 10Hz, 12Hz, reposo]
Datos preprocesados	Si
Número de canales EEG	256
Número de muestras participante	23
Número total de muestras EEG extraíbles	5202
Frecuencia de muestreo	250 Hz
Duración de la muestra	5 segundos
Formato Archivo	EDF/EDF+

Por último, para los conjuntos de datos EEGMAT y MSSVEPDB se realizó una generación de etiquetas con base en la descripción de los experimentos, en el caso de EEGMAT se creó una etiquetación binaria que indica si el sujeto está o no realizando operaciones matemáticas. En el caso del MSSVEPDB se generó una etiqueta por cada una de las frecuencias de estímulo más una etiqueta cero para los estados de reposo.

5.4 Exploración de los conjuntos de datos

Antes de emprender la construcción del modelo, se realizaron exploraciones de los conjuntos de datos con la ayuda de las herramientas mencionadas anteriormente, como [57] y otras adicionales tales como MNE®[13]; en este caso, la motivación principal era realizar una inspección visual sobre los diferentes canales del EEG. Esta visualización facilitó la detección de algunas características físicas de las señales originales (ver Anexo 9.C), también proporcionó información acerca del estado de dichas señales después de las transformaciones. Finalmente, esta información se utilizó como base para intuir algunos detalles en la elaboración del modelo (concepto conocido en ML como *insights*).

Adicionalmente, como paso previo a la elección de la técnica de preprocesamiento basada en espectrogramas de modulación, también se revisaron otras técnicas, como los ERP [22], los cuales ayudan a simplificar el ruido de la señal bajo el dominio del tiempo, proyectando las repuestas cerebrales de los estímulos dados. Aunque su uso es amplio y está presente en otros estudios; la idea de la utilización de ERP fue descartada debido a que los análisis no están basados en un bloqueo de tiempo (del concepto en inglés *time-locked*); por lo tanto, la actividad registrada se pierde en el promedio del ERP [34], dato a que principalmente a que su cálculo depende de la media de un electrodo específico (ver Anexo 9.D).

Otras visualizaciones previas de los conjuntos de datos utilizaron mapas topográficos (ver Anexo 9.E); sin embargo, no se avanzó sobre este tipo de técnicas de visualización ya que el estudio no iba encaminado a realizar análisis espaciales como tales como los mostrados en [64].

La fase de preprocesamiento también contó con su etapa de exploración en el cálculo de los espectrogramas, en este caso se utilizó la librería [57]. El método consistió en calcular espectrogramas con las técnicas de STFT y CWT utilizando diferentes parametrizaciones, con el fin de verificar aspectos como la resolución temporal y las resoluciones a nivel de frecuencia.

Tabla 5-4: Consideraciones en la parametrización de los espectrogramas.

Tipo	Técnica	Parámetro	Comentario
Espectrograma	FFT	win_function	Función de la ventana aplicada, valor por defecto <i>Hamming</i> , para los análisis se utilizó la función <i>boxcar</i> .
		win_size	Tamaño de la ventana, Un segmento de ventana corto ofrece una mayor precisión a nivel temporal, una menor precisión a nivel de frecuencias y una menor resolución en estas. Un segmento de ventana largo ofrece una mayor resolución y precisión en frecuencias, pero una baja precisión temporal. utilizó un tamaño de ventana según la frecuencia de muestreo según conjunto de datos... ejemplo: w_size = 1 * EEGMAT.fs
		win_shift	Porcentaje de traslape de la aplicación de la ventana, en el caso del análisis EEG mitiga los artefactos de borde. Se utilizó un 50%
	Wavelet	n_cycles	Número de ciclos wavelet, se utilizó un n=6

En la exploración de la parametrización de los espectrogramas se buscó optimizar la precisión a nivel de frecuencias. Para tal motivo, se desarrollaron una serie de *scripts* con el fin de llevar a cabo el proceso de visualización de los EEG como un conjunto y no de forma individual (ver *Figura 24*), lo anterior basado en la naturaleza del EEG como un conjunto de señales con características espaciales, espectrales y temporales. La visualización en conjunto también ayudó en la búsqueda de patrones que más adelante fueron tomados en cuenta en el desarrollo del modelo.

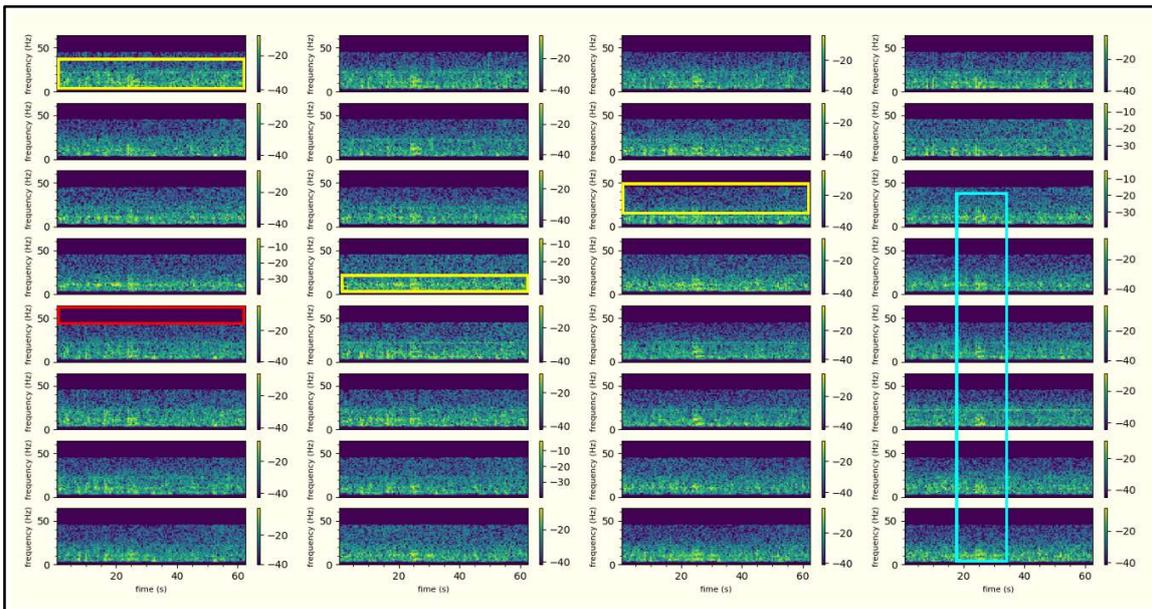


Figura 24: Visualización de espectrogramas basados en FFT para un trial del DEAP. Al revisar los espectrogramas de todos los electrodos del EEG en conjunto se pueden apreciar algunas características generales para las señales, por ejemplo, la ausencia de actividad en las frecuencias por encima de $\approx 50\text{Hz}$ (rojo) esto se debe a que se tomó como insumo señales preprocesadas, las cuales cuentan con la aplicación de un filtro pasa banda $0.4 - 45\text{ Hz}$. Esta visualización también permite realizar una inspección de la actividad en cada una de las bandas agrupando visualmente las frecuencias de forma horizontal (amarillo). Finalmente, se también se puede reconocer algunos eventos en el ámbito temporal (cian). (fuente propia).

Para los espectrogramas de modulación se utilizó la misma técnica y los mismos parámetros en la segunda transformada (ver *Figura 25*). En este caso los espectrogramas de modulación reflejan las características de la señal EEG en regiones visuales muy pequeñas [23]; en consecuencia, se presenta una información que no es legible fácilmente y por ende, no permite ver eventos o particularidades de la señal de forma evidente; en estos casos, la evaluación visual del espectrograma está reservada para el ojo entrenado, como por ejemplo el de un neurólogo. En el presente estudio se avanzó con prudencia en este aspecto con el fin de no caer en interpretaciones inadecuadas, relacionadas con la actividad cerebral y la hipótesis inicial del experimento que generó el conjunto de datos. Por otra parte, se debe tener en cuenta que la resolución del gráfico también puede afectar estas interpretaciones visuales.

Aunque no se realizó una interpretación directa del espectrograma de modulación, esta tarea de análisis sirvió para generar algunas conjeturas con respecto a la elección del clasificador y del modelo de ML a utilizar. También se evaluaron posibles dificultades en el proceso de clasificación ya que, si las diferencias son muy sutiles y cada una de las pruebas son parecidos entre sí, el modelo no será capaz de diferenciar un ejemplo de otro. Finalmente, en la exploración de las técnicas de espectrogramas se evaluaron los tipos wavelet, tanto en espectrograma convencional (ver Anexo 9.G) como en espectrogramas de modulación (ver Anexo 9.H). Para el estudio se descartaron los espectrogramas basados en wavelet como técnica de preprocesamiento, ya que después de varios ensayos, estos no reflejaban diferencias sustanciales que pudiesen ser de utilidad en las fases posteriores, lo anterior con el fin de evitar tendencias se subajuste en el modelo ML. Por otro lado, al realizar cálculos masivos de los espectrogramas tipo wavelet para un conjunto de datos EEG

completo, estos representaban un incremento en el tiempo de ejecución de la fase de preprocesamiento a nivel general; sin embargo, vale la pena aclarar que el estudio no se incluye un análisis comparativo de los tiempos de ejecución entre estas dos técnicas.

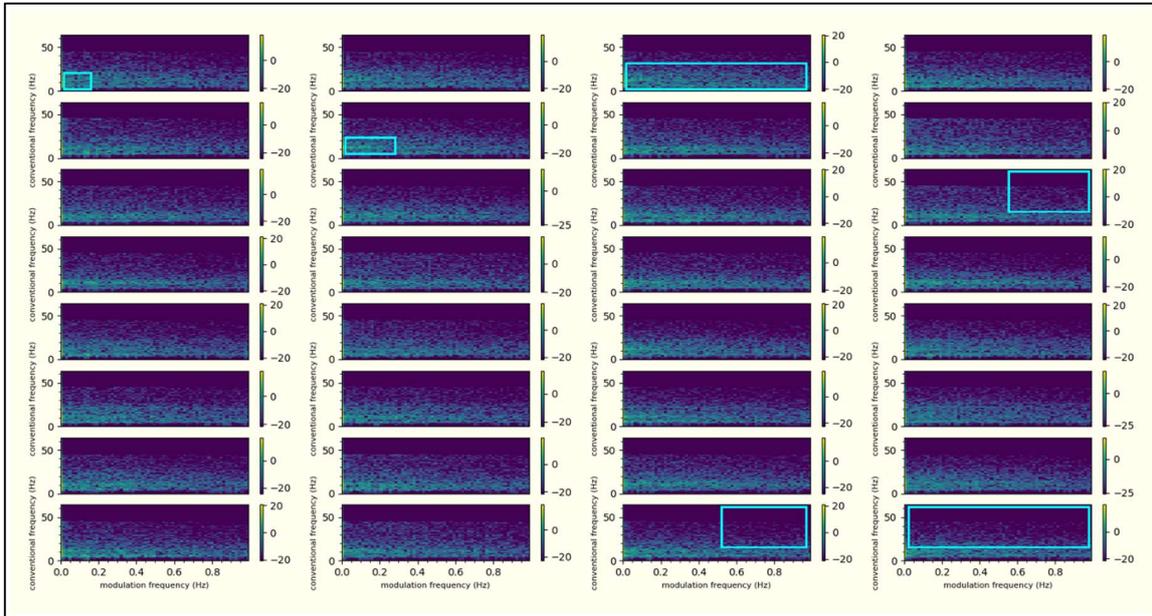


Figura 25: Visualización de espectrogramas de modulación basados en FFT para un trial del DEAP. Se puede apreciar como la escala de la frecuencia de modulación es menor que la escala de las frecuencias convencionales y más pequeña. En la imagen se destacan algunas regiones que llaman la atención (cian), no obstante, sería necesario una comparación entre pruebas y etiquetas para verificar si en realidad estas regiones expresan factores diferenciadores de utilidad para la clasificación. (fuente propia).

Siendo el EEG un conjunto de señales que también tiene características espaciales, se optó por revisar aspectos relacionados con la neurociencia tales como la lateralidad, por tal motivo se realizó una reorganización de los espectrogramas según el posicionamiento 10-20 (ver *Figura 26*). En este se pueden apreciar algunos aspectos que son visibles de forma muy sutil, como la actividad en los electrodos [O1,Oz,O2] lo que da indicios acerca de estímulos visuales, lo cual es verificable en la descripción del experimento del propio conjunto de datos; también se puede discernir una actividad en los lóbulos frontales, lo cual sumado a un segmentación aproximada de 4 a 8 Hz en el eje de la frecuencia convencional permite ubicar una tenue actividad en la banda theta, la cual, está asociada a procesos cognitivos y se cree que está relacionada a estímulos placenteros [38]. Lo anterior también es corroborable desde la descripción del experimento.

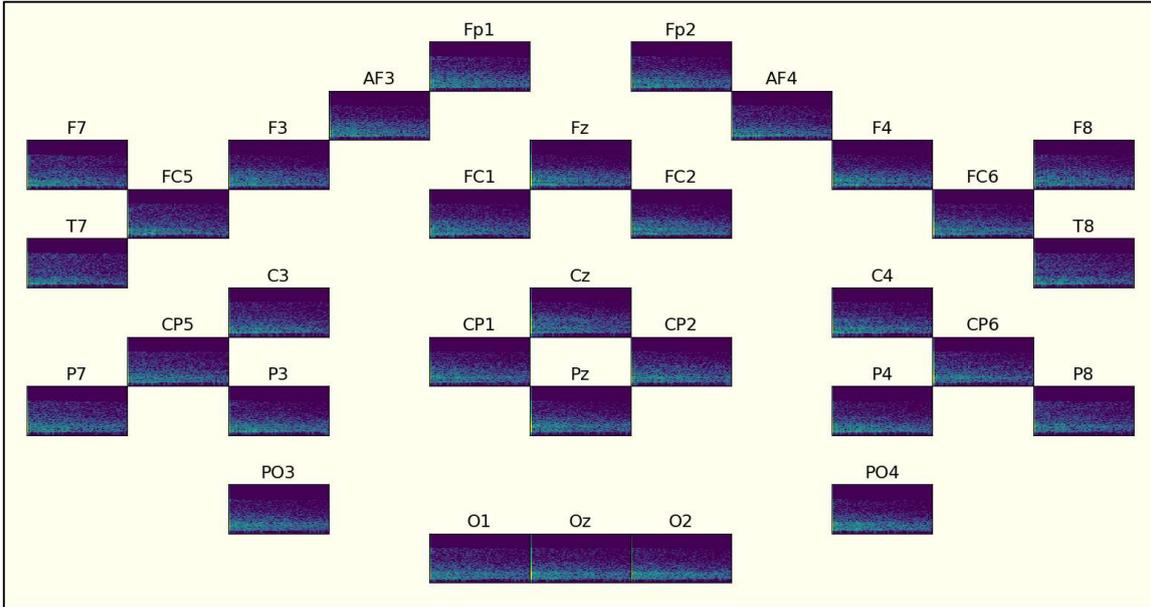


Figura 26: Visualización de los espectrogramas de modulación distribuidos de forma espacial, en este caso se toma como datos de referencia una prueba del conjunto de datos DEAP y a continuación se distribuyen los electrodos según su posición en el sistema 10-20. (fuente propia).

Como se describe en [23], los espectrogramas de modulación permiten realizar segmentaciones funcionales de los componentes espectrales; por ejemplo: la separación entre el sonido del pulmón y sonido del corazón. Desde esta idea principal, se trabajó en el desarrollo de un preprocesamiento en el presente estudio; el cual, está orientado a utilizarse como insumo para modelos ML basados en redes neuronales (ver *Figura 27*).

Cada espectrograma arroja como resultado un arreglo numérico 2D que puede ser dividido según las bandas de frecuencia EEG, cada una de estas bandas constituye otro arreglo 2D; con el fin de simplificar el conjunto de datos de entrada al Transformer se aplica una operación por columnas que consiste en una función agregada tipo media aritmética; no obstante, el método queda abierto para cualquier otro tipo de función, ejemplo: un máximo. Se aclara que en presente estudio se utilizó la media aritmética.

Las operaciones con la función agregada para cada una de las bandas dejan como resultado cinco arreglos 1D, los cuáles se integran en uno solo, luego se realiza una concatenación desde la banda de más baja frecuencia hacia la más alta (ver *Figura 27*). vale la pena aclarar que no hay una razón específica para mantener el orden de las bandas, ya que en teoría, desde el punto del presente trabajo las bandas se analizan de forma independiente según su aspecto funcional, por lo que se podría considerar que no habría por que utilizar un orden específico, caso contrario del Vit[50] donde varios bordes colindantes de las sub-imágenes pueden contener el objeto de interés.

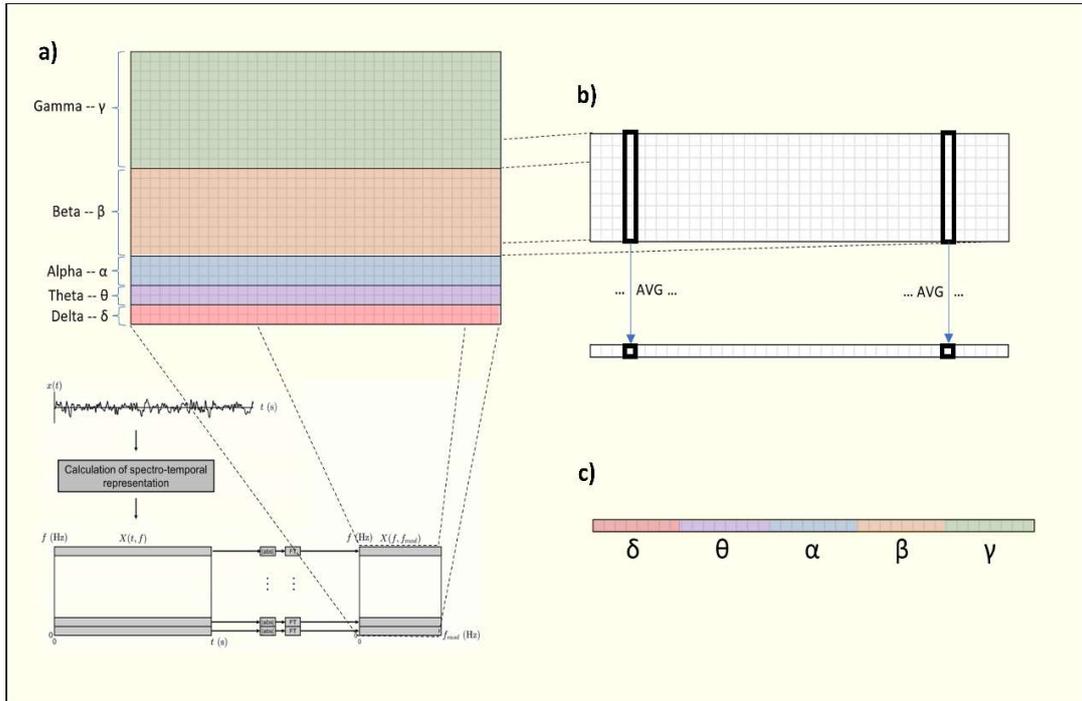


Figura 27: Reorganización del espectrograma de modulación. a) tomada y modificada de [23] muestra cómo se calcula el espectrograma de modulación para cada fragmento de la franja de frecuencias, b) cada banda de frecuencia EEG es un sub-arreglo 2D del espectrograma inicial, al cual se le aplica una función agregada (promedio). c) reorganización de los resultados de b en cada una de las bandas concatenados en un solo arreglo 1D. (fuente propia).

Se entiende que una formalización del método presentado anteriormente, el cual busca reducir el espectrograma de modulación a un arreglo de una dimensión; requiere una profundización y argumentación desde el campo estadístico, por ejemplo: los cálculos de medias deben ir acompañados de análisis de varianzas para garantizar que los datos no estén sesgados por valores extremos; nótese que en la explicación del método también se elude el concepto de reducción de dimensionalidad, ya que lo propuesto en el presente trabajo no constituye claramente la aplicación de alguna de estas técnicas. Por otra parte, en la etapa de revisión teórica del presente estudio se analizaron algunas técnicas de reducción de dimensionalidad, tales como: LLE, ICA, PCA. No obstante, el estudio no aplicó ninguna de las técnicas mencionadas anteriormente debido a que la versión del ViT [50] implementada, realiza una proyección lineal del MS a través de una CNN, dejando cubierto este aspecto.

Esta parte de la simplificación se desarrolló de una forma intuitiva, se trabajó bajo el supuesto de que la potencia no debería presentar grandes fluctuaciones en segmento de banda analizado, por lo que la volatilidad de los valores extremos no supone un problema. Lo anterior se corrobora a través de un script implementado como parte de la exploración de datos (ver *Figura 28*).

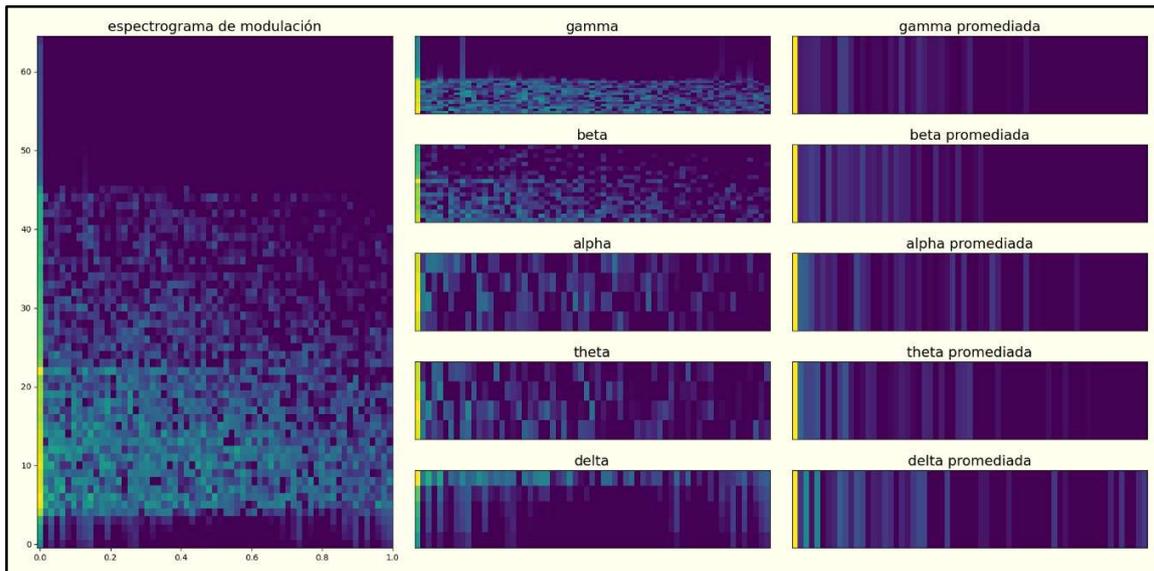


Figura 28: Método de simplificación del espectrograma de modulación, se divide el espectrograma de modulación en subregiones, luego se aplica una función agregada como por ejemplo la media. Se puede apreciar que los arreglos de la parte derecha de la gráfica conservan similitudes en algunos patrones que se visualizan en las columnas del medio. (fuente propia).

La una división en la línea de tiempo de la señal original y posterior cálculo de los espectrogramas de modulación no reflejó variaciones notorias (ver *Figura 29*), por lo que esta técnica no generó valor agregado como estrategia en el aumento del tamaño del conjunto de datos para la fase de entrenamiento del modelo. La única segmentación temporal aplicada fue para separar los diferentes segmentos de estímulos en los conjuntos de datos EEGMAT y MSSVEPDB.

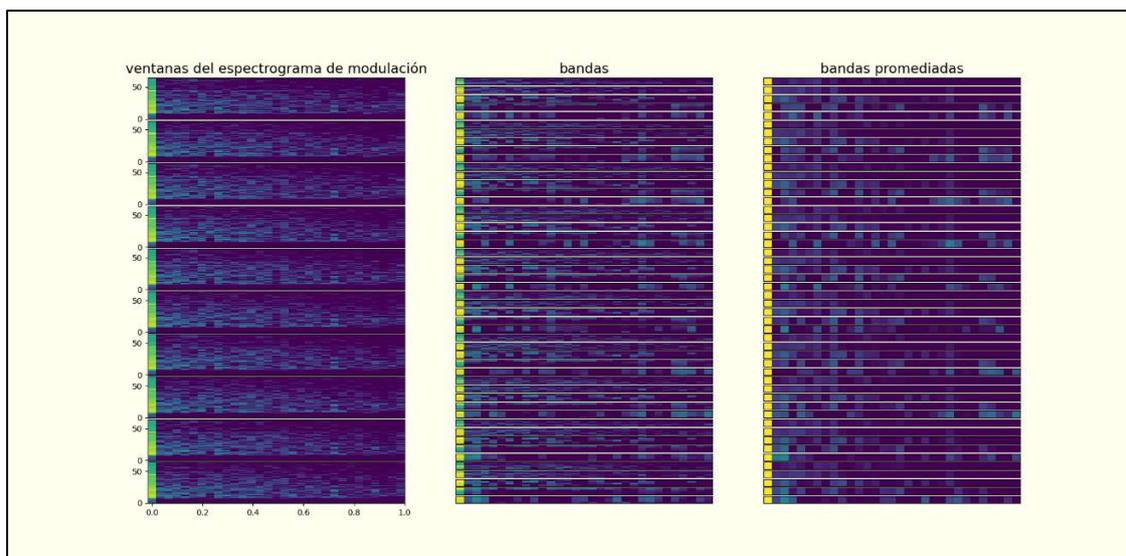


Figura 29: Segmentación de la señal y cálculo de los espectrogramas de modulación, esta técnica no reflejó ventajas al momento de la exploración del conjunto de datos y tampoco en la fase de entrenamiento, por lo que se descartó de la implementación. (fuente propia).

Las etiquetas de cada uno de los conjuntos de datos se codificaron utilizando un formato *one hot*, por el motivo de que, al implementar un modelo de clasificación categórica, la última capa debe ser tipo *softmax*, para una fase posterior de aplicación de resultados también se podría optar por el uso de una capa *argmax*. En el caso del conjunto de datos DEAP ya se contaba con un etiquetado descrito en [60], lo que se realizó en este caso fue adaptar este etiquetado a cuatro categorías según el modelo de clasificación de emociones circumplex (ver *Figura 3*).

5.5 Implementación del MST

La implementación a nivel de software fue diseñada con dos componentes independientes que se intercomunican a través de lotes de datos (ver *Figura 30*), los cuales se dividen entre la fase de preprocesamiento y la fase de procesamiento de datos o MST. Este desacople del componente de preprocesamiento permite conservar los espectrogramas calculados en disco, para luego ser reutilizados con diferentes configuraciones del MST, lo cual reduce el tiempo de ejecución de las pruebas a nivel general. La construcción se realizó en lenguaje de programación Python, como plataforma de ML se utilizó Tensorflow®[55] y el API Keras®[56], la cual dio soporte a la implementación de las diferentes capas, incluidas las de los *Transformer*. A continuación, se presenta una lista de las capas utilizadas en la implementación del modelo:

- `tf.keras.layers.Input`
- `tf.keras.layers.LayerNormalization`
- `tf.keras.layers.Layer`
- `tf.keras.layers.Dense`
- `tf.keras.layers.Lambda`
- `tf.keras.layers.Dropout`
- `tf.keras.Sequential`
- `tf.keras.layers.MultiHeadAttention`

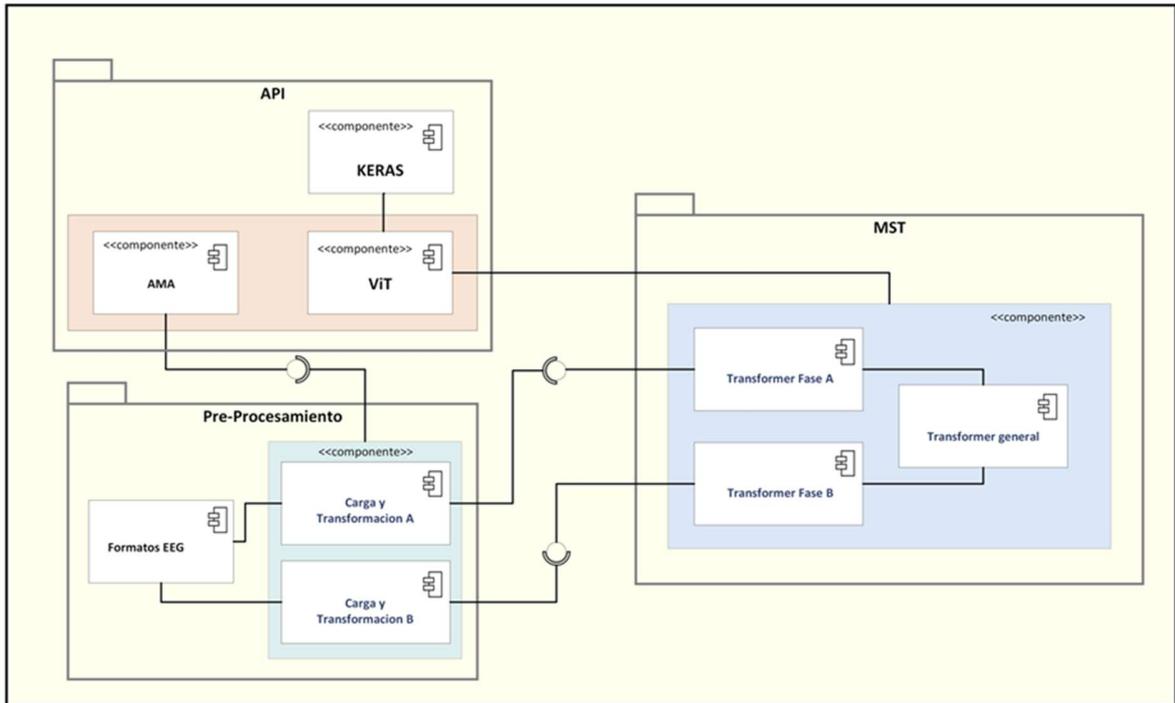


Figura 30: Diagrama de componentes UML para la implementación del MST, muestra la interrelación de los componentes de software comprendidos en la implementación. (fuente propia).

Para el componente de preprocesamiento es posible parametrizar con cuál de los métodos de cálculo de espectrogramas se desea trabajar (Fourier/Wavelet), adicionalmente existe un parámetro que permite definir una región específica de extracción, en el caso de ser requerido. Finalmente, el ultimo parámetro adapta el guardado de los resultados del preprocesamiento. El caso del MST se permiten las siguientes parametrizaciones:

Tabla 5-5: parámetros del MST

Fase	Nombre del parámetro	Descripción
Transformación entrada (fase A/fase B)	patch_size	tamaño del kernel para la convolución según ViT
	hidden_size	número de filtros según ViT
	num_layers	número de capas de Transformer según ViT
	num_heads_in	número de cabezas para el Transformer para el Transformer de entrada
Transformación salida	num_layers_out	número de capas de Transformer para la salida
	num_heads_out	número de cabezas para el Transformer para el Transformer de salida
	ff_dim	dimensión de la <i>feed forward network</i> para el Transformer de salida

Entrenamiento	num_clases	numero de clases configuradas para las etiquetas
	list_patches	lista de áreas del cerebro utilizadas para la discriminación de los electrodos, ejemplo: [pre_frontal, frontal, central, tempo_left, tempo_right, parietal, occipital]
	batch_size	tamaño del lote en la fase de entrenamiento
	epochs	número de épocas en el modelo
	train_val_pct	porcentajes de distribución de los datos para los conjuntos de entrenamiento y validación

Existen otros parámetros fijos que no se modificaron entre cada uno de los ensayos:

Tabla 5-6: Parámetros fijos del MST

Parámetro	Valor
Función de optimización	Adam
Función de pérdida	Categorical Crossentropy
Función de activación interna (MLP)	Linear
Función de activación salida (MLP)	Softmax
Dropout	0.5
Métrica	Categorical

5.6 Entrenamiento y resultados

Para cada uno de los experimentos se realizaron varios ciclos de entrenamiento y validación; a continuación, se presentan los resultados más relevantes con sus respectivas configuraciones:

5.6.1 Clasificación de emociones

Tabla 5-7: Resumen del experimento 1A, este experimento utiliza la fase A del MST

Conjunto de datos	DEAP
Dimensión del conjunto de datos:	(1280, 32, 270)
Dimensión del conjunto de etiquetas	(1280, 4)
Fase	A
Parámetros utilizados	50,230,300
Exactitud en entrenamiento	0.3965
Exactitud en validación	0.4219
Segmentos EEG	"frontal", "parietal", "occipital", "frontal_left", "frontal_right", "parietal_left", "parietal_right", "occipital_left", "occipital_right"

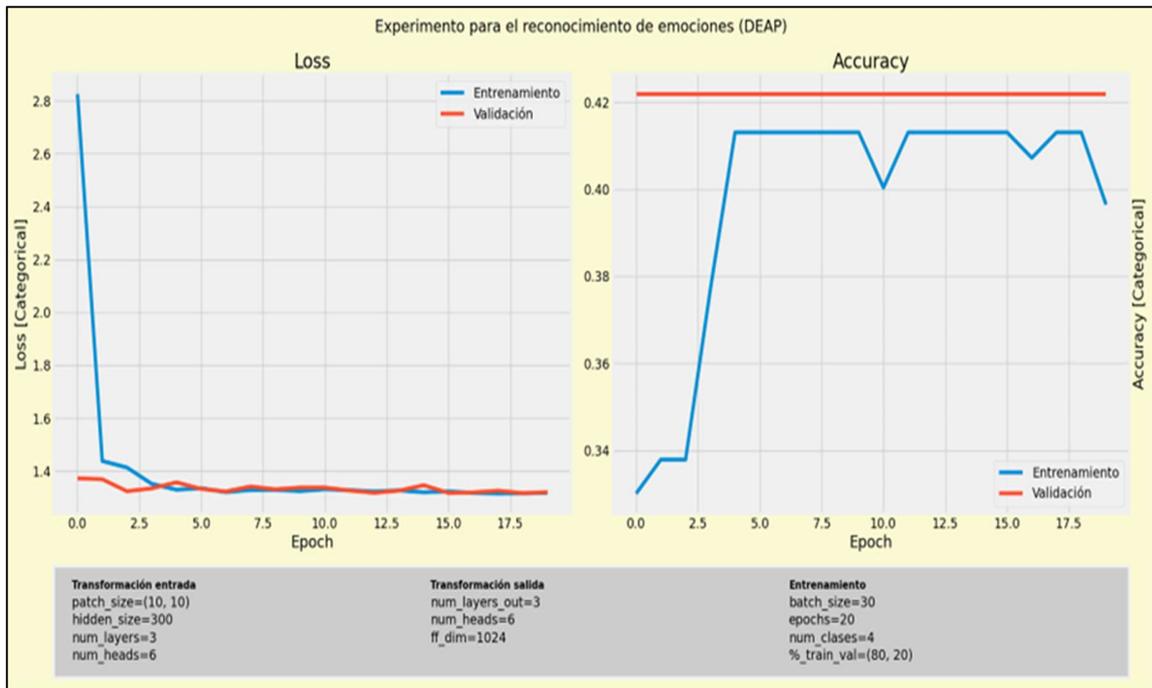


Figura 31: Resultados y parámetros DEAP, experimento utilizando la fase A. (fuente propia).

Tabla 5-8: Resumen del experimento 1B, este experimento utiliza la fase B del MST

Conjunto de datos	DEAP
Dimensión del conjunto de datos:	(1280, 32, 270)
Dimensión del conjunto de etiquetas	(1280, 4)
Fase	B
Parámetros utilizados	14,274,782
Exactitud en entrenamiento	0.4131
Exactitud en validación	0.4219
Segmentos EEG	"frontal", "parietal", "occipital", "frontal_left", "frontal_right", "parietal_left", "parietal_right", "occipital_left", "occipital_right"

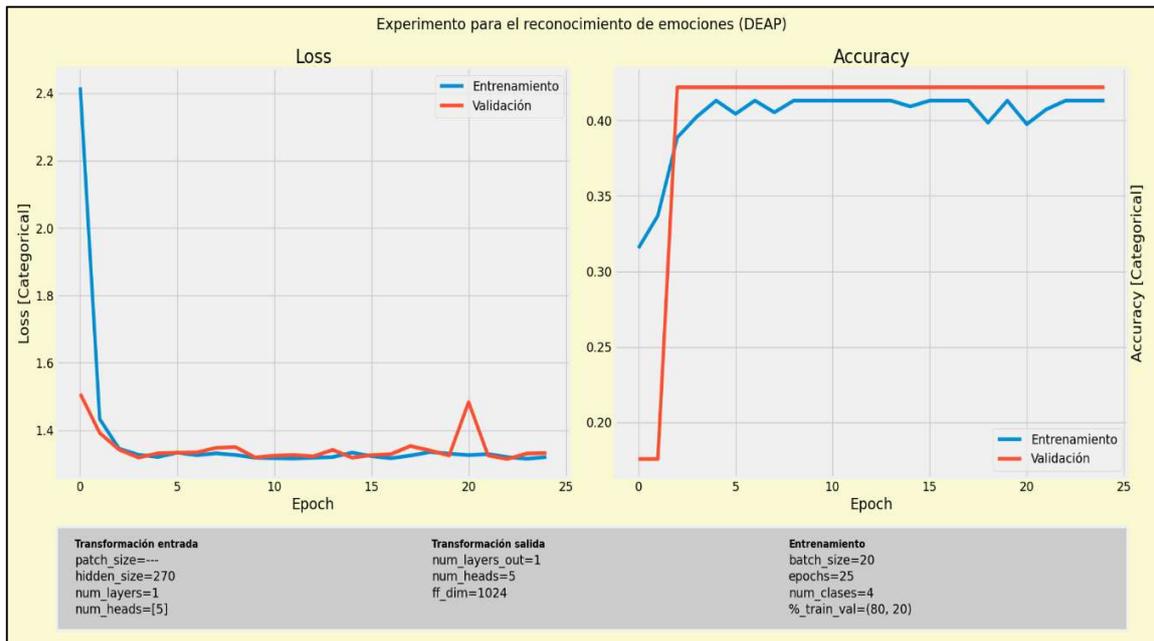


Figura 32: Resultados y parámetros DEAP, experimento utilizando la fase B. (fuente propia).

5.6.2 Clasificación de tarea cognitiva (cálculos)

Tabla 5-9: Resumen del experimento 2A, este experimento utiliza la fase A del MST

Conjunto de datos	EEGMAT
Dimensiones del conjunto de datos de entrada:	(214, 19, 80, 30)
Dimensiones del conjunto de etiquetas:	(214, 2)
Fase usada en el preprocesamiento	A
Parámetros utilizados	143,738,006
Exactitud en entrenamiento	0.6725
Exactitud en Validación	0.6677
Segmentos EEG	"pre_frontal", "frontal", "central", "tempo_left", "tempo_right", "parietal", "occipital"

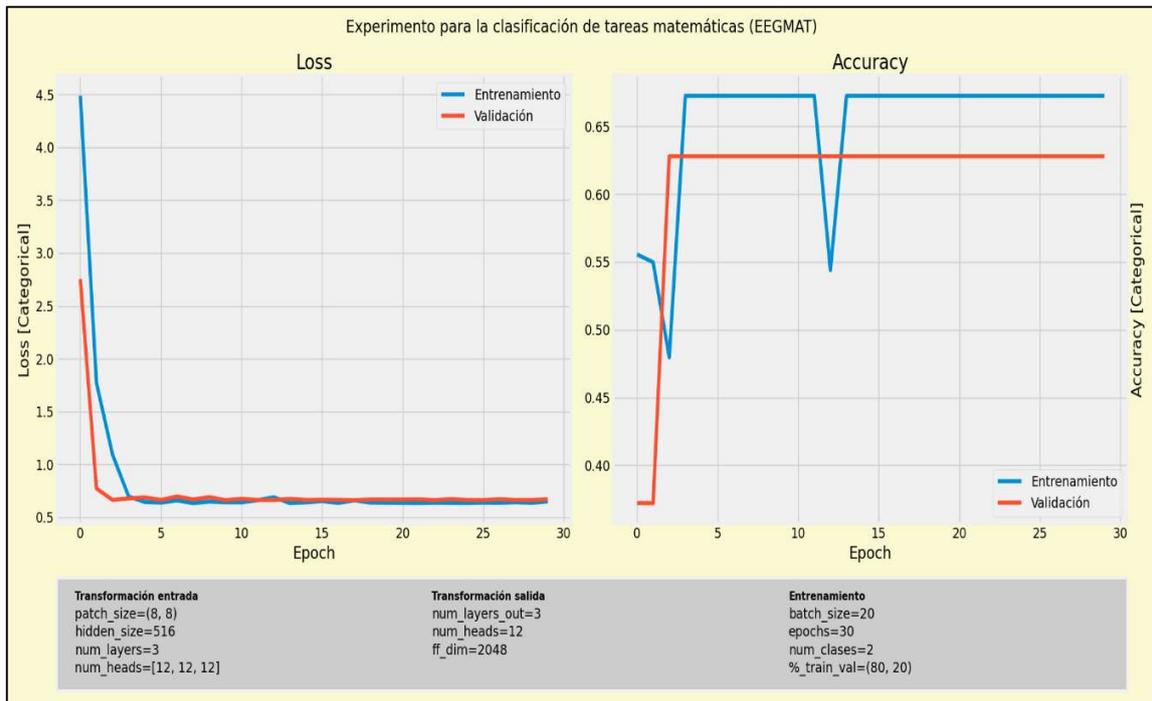


Figura 33: Resultados y parámetros EEGMAT, experimento utilizando la fase A. (fuente propia).

Tabla 5-10: Resumen del experimento 2B, este experimento utiliza la fase B del MST

Conjunto de datos	EEGMAT
Dimensiones del conjunto de datos de entrada:	(214, 19, 150)
Dimensiones del conjunto de etiquetas:	(214, 2)
Fase usada en el preprocesamiento	B
Parámetros utilizados	29,408,894
Exactitud en entrenamiento	0.6725
Exactitud en Validación	0.6279
Segmentos EEG	"pre_frontal", "frontal", "central", "tempo_left", "tempo_right", "parietal", "occipital"

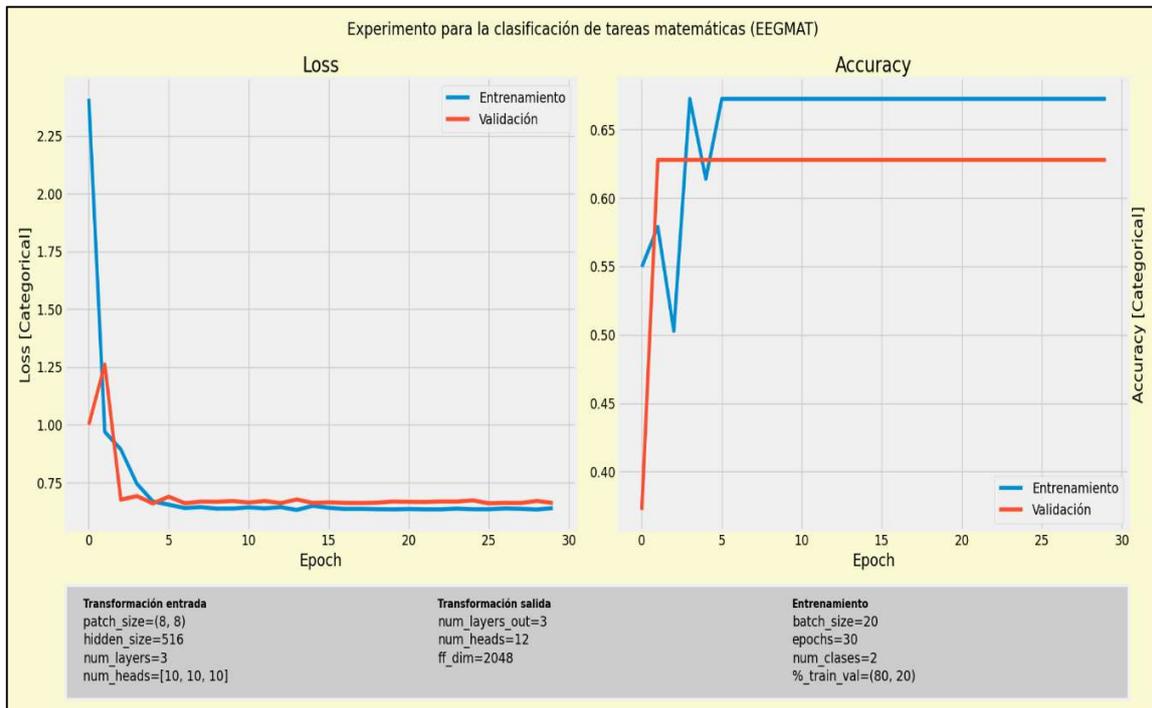


Figura 34: Resultados y parámetros EEGMAT, experimento utilizando la fase B. (fuente propia).

5.6.3 Clasificación efecto Flicker

Tabla 5-11: Resumen del experimento 3A, este experimento utiliza la fase A del MST

Conjunto de datos	MSSVEPDB
Dimensiones del conjunto de datos de entrada:	(5205, 256, 80, 3)
Dimensiones del conjunto de etiquetas:	(5205, 6)
Fase usada en el preprocesamiento	A
Parámetros utilizados	76,427,526
Exactitud en entrenamiento	0.6326
Exactitud en Validación	0.5293
Segmentos EEG	"frontal", "tempo_left", "tempo_right", "parietal", "occipital"

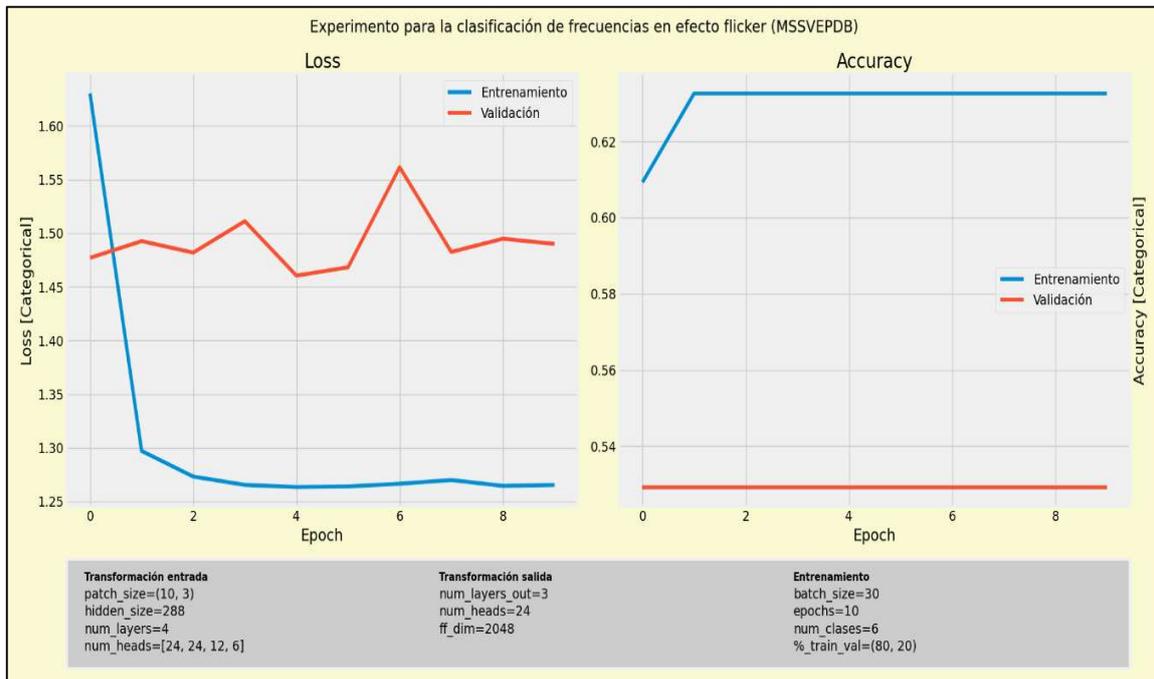


Figura 35: Resultados y parámetros MSSVEPDB, experimento utilizando la fase A. (fuente propia).

Experimento #3B: Clasificación EEG para el efecto Flicker

Tabla 5-12: Resumen del experimento 3B, este experimento utiliza la fase B del MST

Conjunto de datos	MSSVEPDB
Dimensiones del conjunto de datos de entrada:	(5205, 256, 15)
Dimensiones del conjunto de etiquetas:	(5205, 6)
Fase usada en el preprocesamiento	B
Parámetros utilizados	2,055,738
Exactitud en entrenamiento	0.6326
Exactitud en Validación	0.5293
Segmentos EEG	"frontal", "tempo_left", "tempo_right", "parietal", "occipital"

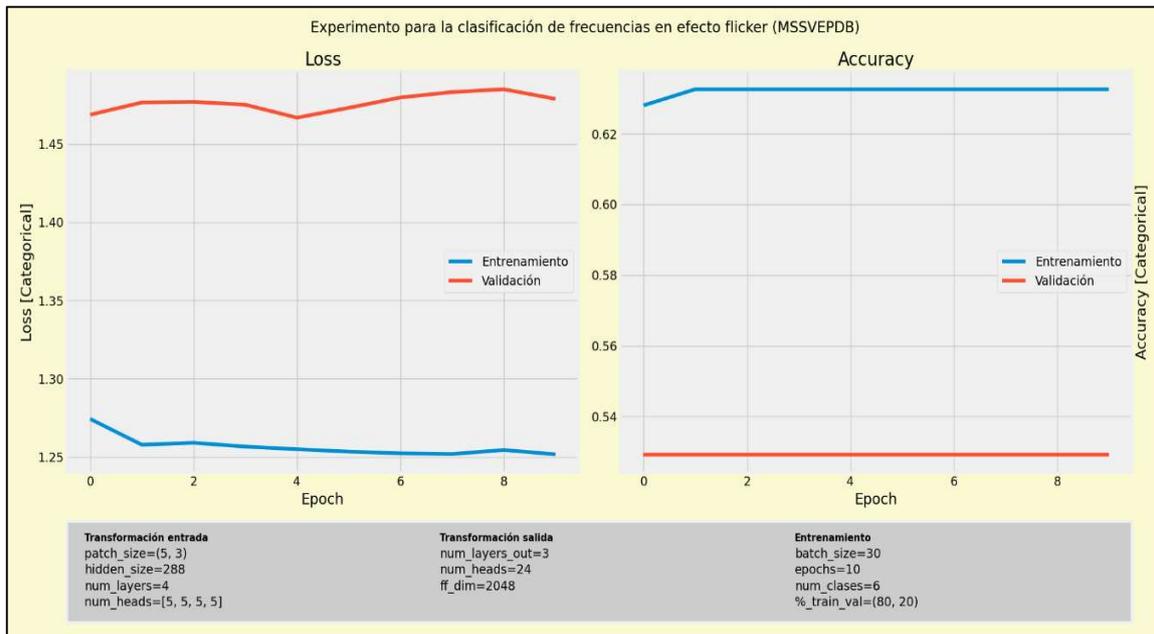


Figura 36: Resultados y parámetros MSSVEPDB, experimento utilizando la fase B. (fuente propia).

Resumen de resultados:

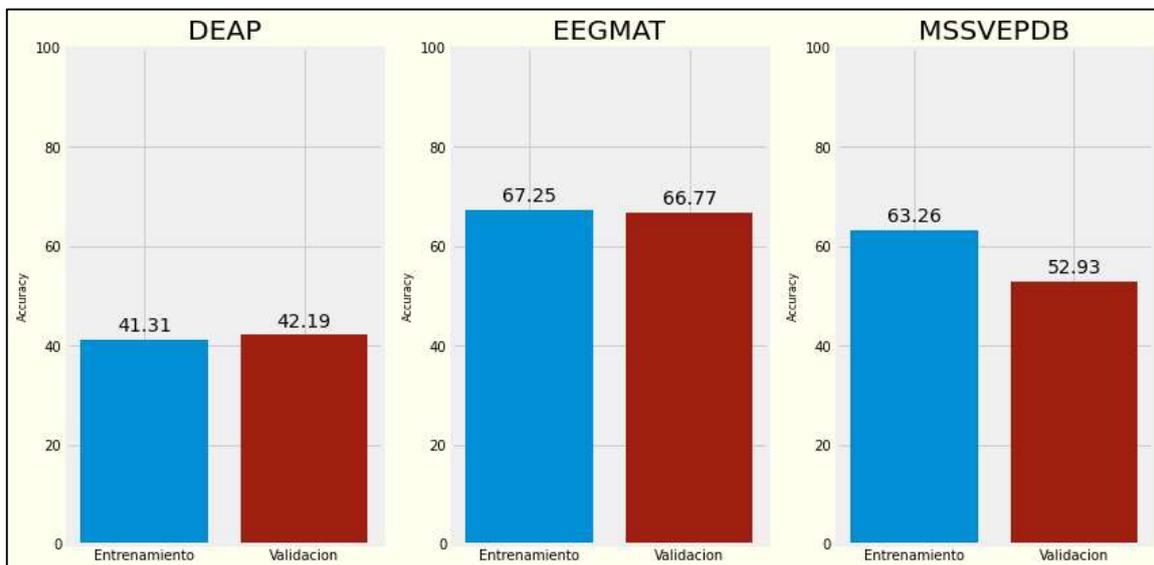
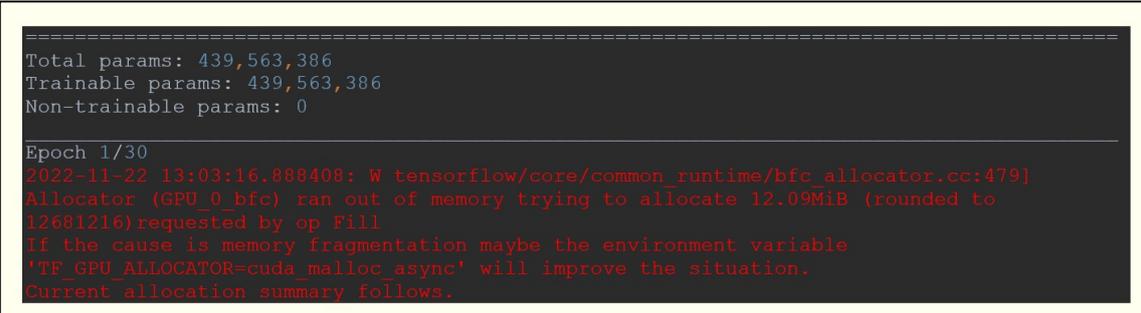


Figura 37: Resumen de resultados: Resumen de los resultados de los tres experimentos. (fuente propia).

5.6.4 Desafíos de la investigación

A nivel de hardware se tuvieron limitantes en cuanto a la cantidad de parámetros a utilizar en la construcción de los modelos de prueba, mientras que [50] indica una cantidad de parámetros entre 307 y 623 millones, en sus versiones larga y extralarga. Nuestro modelo no alcanzaba una cantidad de parámetros límite de 290 millones sin que se corriera el riesgo de tener problemas de asignación de memoria. Esto básicamente afectó la fase A de los experimentos, ya que no se podía contar con una mejor resolución en la capa de convolución y tampoco se podía aumentar el tamaño de las capas internas (ver *Figura 38*).



```
Total params: 439,563,386
Trainable params: 439,563,386
Non-trainable params: 0

Epoch 1/30
2022-11-22 13:03:16.808408: W tensorflow/core/common_runtime/bfc_allocator.cc:479]
Allocator (GPU_0_bfc) ran out of memory trying to allocate 12.09MiB (rounded to
12681216) requested by op Fill
If the cause is memory fragmentation maybe the environment variable
'TF_GPU_ALLOCATOR=cuda_malloc_async' will improve the situation.
Current allocation summary follows.
```

Figura 38: Ejemplo de problemas de entrenamiento con el experimento EEGMAT, en este caso, se redujo el tamaño del kernel de convolución, $patch_size = (4,4)$ y se aumentó el parámetro de los filtros, $hidden_size = 1032$. (fuente propia).

Otra dificultad mencionada es la poca cantidad de datos EEG disponibles para este método, mientras que IMAGENET cuenta con alrededor de 1.4 millones de imágenes, nuestro estudio consiguió cifras de 4066 y 40960 en los conjuntos de datos EEGMAT y DEAP respectivamente. En este caso se discutió el hecho de que división de las señales EEG limitaba el cálculo del MS, ya que un espacio de señal reducido producirá un MS de corta longitud en el eje de las frecuencias de modulación debido al teorema de Nyquist. Por ejemplo, para el conjunto de datos MSSVEPDB calcula MS del orden (80×3) , lo cual se debe a la longitud del estímulo (≈ 5 segundos), se considera que estas longitudes de MS podrían no tener la suficiente representación de la señal para una correcta diferenciación de las etiquetas dentro del entrenamiento del modelo.

6 Discusión

Los estudios relacionados con el análisis de señales EEG a través de *Transformers* han ido floreciendo en los últimos dos años, no obstante, al ser una técnica reciente, hay pocos puntos de comparación entre los diferentes experimentos realizados y otros estudios. En este caso, para el experimento de clasificación de emociones relacionan los resultados de [51], que corresponden a una clasificación de las 4 clases para el conjunto de datos DEAP:

Tabla 6-1: Comparación del accuracy en la clasificación de HSTL vs MST

Técnica	DEAP (P_{acc})
HSLT	56.93
MST	42.19

De todos los modelos entrenados con el modelo MST, el experimento de clasificación de emociones es el que presenta el *accuracy* más bajo. Sin embargo, la clasificación de emociones es una tarea compleja, debido a que la elicitación de una emoción se presenta en un corto periodo de tiempo. Adicionalmente, salirse de una clasificación binaria para entrar a un modelo de mínimo 4 etiquetas añade un mayor reto. Por último, para este experimento no se lograron comprobar las hipótesis de lateralización emocional con MS, debido a que pruebas realizadas de forma adicional con segmentaciones diferentes del EEG no arrojaron resultados significativos.

Los resultados para los experimentos EEGMAT y MSSVEPDB obtuvieron resultados un poco más prometedores; sin embargo, se detectó una tendencia de los resultados del entrenamiento hacia el subajuste, lo anterior se presentó por que aparentemente los datos inyectados al modelo no tenían diferencias significativas entre sí. En síntesis, se cree que es posible que los MS no generen una representación diferenciadora entre cada una de las etiquetas empleadas en la clasificación.

Tabla 6-2: Comparación del accuracy entre los experimentos EEGMAT y MSSVEPDB.

Técnica	EEGMAT (P_{acc})	MSSVEPDB (P_{acc})
MST	66.77%	52.93%

Por último, es destacable que la técnica utilizada en la fase B de los experimentos, alcance una exactitud similar pero con menos parámetros que los utilizados por a técnica de la fase A, inspirada en [50], [51]. Lo que hace que el modelo necesite menos capacidad de cómputo en la fase de entrenamiento.

Tabla 6-3: comparación de los parámetros entre la fase A y la fase B del modelo MST.

Conjunto de datos	Parámetros Fase A	Parámetros Fase B
DEAP	50.230.300	14.274.782
EEGMAT	143.738.006	29.408.894
MSSVEPDB	76.427.526	2.055.738

Trabajos futuros

Uno de los aspectos principales del presente trabajo, es que realiza una clasificación categórica mediante la técnica de *Transformers*. Queda como trabajo futuro probar con otros tipos de predicciones o transformaciones sobre la señal, por ejemplo, los ofrecidos por los modelos secuencia a secuencia o *seqToSeq*. También se podría trabajar una convolución fragmentada en la fase A del MST.

En este trabajo solo se manejó la definición del *Transformer* estándar del ViT, existen otras configuraciones de las capas de atención, como las ofrecidas en [53], [65], las cuales se proponen como punto de exploración para trabajos subsecuentes.

Finalmente, se propone agregar entradas de *Transformer* adicional para espectrogramas provenientes de otros tipos de señales que generalmente acompañan los experimentos EEG, tales como EOG, EMG y ECG que dependen del tipo del experimento realizado, un ejemplo de lo anterior sería incluir algunos de los canales adicionales especificados en el conjunto de datos[60]

7 Conclusiones

Se analizaron los conjuntos de datos seleccionados DEAP, EEGMAT y MSSVEPDB. En este caso, el modelo implementado MST no refleja un gran performance a nivel de exactitud, en comparación con otras técnicas de aprendizaje de máquina utilizadas en el análisis de señales EEG. En el presente trabajo, el mejor resultado obtenido fue de un 66.77%, conseguido en la clasificación de la actividad cerebral al momento de realizar cálculos matemáticos EEGMAT; particularmente, este experimento pertenece a una clasificación de solo dos categorías; lo que, según otros estudios analizados, mejora las posibilidades de obtener una mejor exactitud en la clasificación. No obstante, vale la pena mencionar que otros trabajos relacionados con Transformers y EEG presentan resultados similares al presente estudio. Finalmente, cabe destacar que el presente trabajo sienta una base para estudios posteriores que busquen el mejoramiento de esta técnica.

8 Referencias

- [1] D. Purves, Ed., *Neuroscience*, Sixth edition. New York: Oxford University Press, 2018.
- [2] P. Dayan y L. F. Abbott, *Theoretical neuroscience: computational and mathematical modeling of neural systems*. en *Computational neuroscience*. Cambridge, Mass: Massachusetts Institute of Technology Press, 2001.
- [3] R. Reisenzein, «21A Short History of Psychological Perspectives on Emotion», en *The Oxford Handbook of Affective Computing*, R. Calvo, S. D’Mello, J. Gratch, y A. Kappas, Eds., Oxford University Press, 2015, p. 0. doi: 10.1093/oxfordhb/9780199942237.013.014.
- [4] J. Gratch y S. C. Marsella, «54Appraisal Models», en *The Oxford Handbook of Affective Computing*, R. Calvo, S. D’Mello, J. Gratch, y A. Kappas, Eds., Oxford University Press, 2015, p. 0. doi: 10.1093/oxfordhb/9780199942237.013.015.
- [5] H. Dabas, C. Sethi, C. Dua, M. Dalawat, y D. Sethia, «Emotion Classification Using EEG Signals», en *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence - CSAI '18*, Shenzhen, China: ACM Press, 2018, pp. 380-384. doi: 10.1145/3297156.3297177.
- [6] A. C. (Kia) Nobre y M. S. Stokes, «Memory and Attention: The Back and Forth», en *The Cognitive Neurosciences*, D. Poeppel, G. R. Mangun, y M. S. Gazzaniga, Eds., 6.^a ed. The MIT Press, 2020, pp. 291-300. doi: 10.7551/mitpress/11442.003.0035.
- [7] J. F. Cantlon, «The Nature of Human Mathematical Cognition», en *The Cognitive Neurosciences*, D. Poeppel, G. R. Mangun, y M. S. Gazzaniga, Eds., 6.^a ed. The MIT Press, 2020, pp. 817-826. doi: 10.7551/mitpress/11442.003.0090.
- [8] «Overview of Time-Domain EEG Analyses», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0013.
- [9] E. Niedermeyer, D. L. Schomer, y F. H. Lopes da Silva, Eds., *Niedermeyer’s electroencephalography: basic principles, clinical applications, and related fields*, 6. ed. Philadelphia: Wolters Kluwer, Lippincott Williams & Wilkins, 2011.
- [10] «Introduction to the Physiological Bases of EEG», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0008.
- [11] D. Bansal y R. Mahajan, *EEG-based brain-computer interfaces: cognitive analysis and control applications*. London: Elsevier/Academic Press, 2019.
- [12] L. V. Marcuse, M. C. Fields, y J. (Jenna) Yoo, «1 - Origin and technical aspects of the EEG», en *Rowan’s Primer of EEG (Second Edition)*, L. V. Marcuse, M. C. Fields, y J. (Jenna) Yoo, Eds., Second Edition. London: Elsevier, 2016, pp. 1-37. doi: <https://doi.org/10.1016/B978-0-323-35387-8.00001-9>.
- [13] «MNE — MNE 1.2.1 documentation». <https://mne.tools/stable/index.html> (accedido 2 de noviembre de 2022).
- [14] «EEGLAB». <https://sccn.ucsd.edu/eeglab/index.php> (accedido 11 de abril de 2023).
- [15] A. V. Oppenheim, A. S. Willsky, y S. H. Nawab, *Signals & systems*, 2nd ed. en Prentice-Hall signal processing series. Upper Saddle River, N.J: Prentice Hall, 1997.
- [16] C. S. Lessard, «Basis Functions and Signal Representation», en *Signal Processing of Random Physiological Signals*, en *Synthesis Lectures on Biomedical Engineering*. Cham: Springer International Publishing, 2006, pp. 19-29. doi: 10.1007/978-3-031-01610-3_4.
- [17] «The Discrete Time Fourier Transform, the FFT, and the Convolution Theorem», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0016.

- [18] C. S. Lessard, «Classification of Signals», en *Signal Processing of Random Physiological Signals*, Cham: Springer International Publishing, 2006, pp. 11-18. doi: 10.1007/978-3-031-01610-3_3.
- [19] A. Grami, «Chapter 3 - Signals, Systems, and Spectral Analysis», en *Introduction to Digital Communications*, A. Grami, Ed., Boston: Academic Press, 2016, pp. 41-150. doi: <https://doi.org/10.1016/B978-0-12-407682-2.00003-X>.
- [20] D. L. Schomer, C. M. Epstein, S. T. Herman, D. Maus, y B. J. Fisch, *Recording Principles*, vol. 1. Oxford University Press, 2017. doi: 10.1093/med/9780190228484.003.0005.
- [21] L. V. Marcuse, M. C. Fields, y J. (Jenna) Yoo, «2 - The normal adult EEG», en *Rowan's Primer of EEG (Second Edition)*, L. V. Marcuse, M. C. Fields, y J. (Jenna) Yoo, Eds., Second Edition. London: Elsevier, 2016, pp. 39-66. doi: <https://doi.org/10.1016/B978-0-323-35387-8.00002-0>.
- [22] N. Kamel y A. S. Malik, Eds., *EEG/ERP analysis: methods and applications*. Boca Raton London New York: CRC Press, 2017.
- [23] R. Cassani y T. H. Falk, «Spectrotemporal Modeling of Biomedical Signals: Theoretical Foundation and Applications», en *Encyclopedia of Biomedical Engineering*, Elsevier, 2019, pp. 144-163. doi: 10.1016/B978-0-12-801238-3.99993-8.
- [24] «Bandpass Filtering and the Hilbert Transform», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0019.
- [25] L. V. Marcuse, M. C. Fields, J. Yoo, y A. J. Rowan, *Rowan's primer of EEG*. 2016. Accedido: 2 de abril de 2021. [En línea]. Disponible en: <http://www.sciencedirect.com/science/book/9780323353878>
- [26] «European Data Format (EDF)». <https://www.edfplus.info/> (accedido 8 de noviembre de 2022).
- [27] «Biosemi EEG ECG EMG BSPM NEURO amplifiers systems». https://www.biosemi.com/faq/file_format.htm (accedido 8 de noviembre de 2022).
- [28] D. Bansal y R. Mahajan, «Chapter 2 - EEG-Based Brain-Computer Interfacing (BCI)», en *EEG-Based Brain-Computer Interfaces*, D. Bansal y R. Mahajan, Eds., Academic Press, 2019, pp. 21-71. doi: <https://doi.org/10.1016/B978-0-12-814687-3.00002-8>.
- [29] «Practicalities of EEG Measurement and Experiment Design», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0009.
- [30] «Advantages and Limitations of Time- and Time-Frequency-Domain Analyses», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0005.
- [31] R. A. Calvo, S. D'Mello, J. Gratch, y A. Kappas, Eds., *The Oxford handbook of affective computing*. Oxford ; New York: Oxford University Press, 2015.
- [32] «Preprocessing Steps Necessary and Useful for Advanced Data Analysis», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0011.
- [33] «EEG Artifacts: Their Detection, Influence, and Removal», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0012.
- [34] «Advantages and Limitations of Different Statistical Procedures», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0040.
- [35] E. S. Pane, A. D. Wibawa, y M. H. Purnomo, «Improving the accuracy of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters», *Cogn. Process.*, vol. 20, n.º 4, Art. n.º 4, nov. 2019, doi: 10.1007/s10339-019-00924-z.
- [36] Z. Liang, S. Oba, y S. Ishii, «An unsupervised EEG decoding system for human emotion recognition», *Neural Netw.*, vol. 116, pp. 257-268, ago. 2019, doi: 10.1016/j.neunet.2019.04.003.
- [37] D. P. Tobón Vallejo y A. El Saddik, «Emotional States Detection Approaches Based on Physiological Signals for Healthcare Applications: A Review», en *Connected Health in Smart*

- Cities*, A. El Saddik, M. S. Hossain, y B. Kantarci, Eds., Cham: Springer International Publishing, 2020, pp. 47-74. doi: 10.1007/978-3-030-27844-1_4.
- [38] R. Calvo *et al.*, «Affective Brain-Computer Interfaces», en *The Oxford Handbook of Affective Computing*, R. Calvo, S. D’Mello, J. Gratch, y A. Kappas, Eds., Oxford University Press, 2015. doi: 10.1093/oxfordhb/9780199942237.013.024.
- [39] A. Vaswani *et al.*, «Attention Is All You Need». arXiv, 5 de diciembre de 2017. Accedido: 6 de octubre de 2022. [En línea]. Disponible en: <http://arxiv.org/abs/1706.03762>
- [40] pubmeddev, «Home - PubMed - NCBI», 21 de abril de 2019. <https://www.ncbi.nlm.nih.gov/pubmed/> (accedido 21 de abril de 2019).
- [41] E. M. dos Santos, R. Cassani, T. H. Falk, y F. J. Fraga, «Improved motor imagery brain-computer interface performance via adaptive modulation filtering and two-stage classification», *Biomed. Signal Process. Control*, vol. 57, p. 101812, mar. 2020, doi: 10.1016/j.bspc.2019.101812.
- [42] M. Ramzan y S. Dawn, «Learning-based classification of valence emotion from electroencephalography», *Int. J. Neurosci.*, vol. 129, n.º 11, pp. 1085-1093, nov. 2019, doi: 10.1080/00207454.2019.1634070.
- [43] E. S. Pane, A. D. Wibawa, y M. H. Purnomo, «Improving the accuracy of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters», *Cogn. Process.*, vol. 20, n.º 4, Art. n.º 4, nov. 2019, doi: 10.1007/s10339-019-00924-z.
- [44] Y. Hou y S. Chen, «Distinguishing Different Emotions Evoked by Music via Electroencephalographic Signals», *Comput. Intell. Neurosci.*, vol. 2019, pp. 1-18, mar. 2019, doi: 10.1155/2019/3191903.
- [45] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, y R. Boots, «Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain Computer Interface». arXiv, 10 de junio de 2021. Accedido: 17 de octubre de 2022. [En línea]. Disponible en: <http://arxiv.org/abs/1708.06578>
- [46] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, y B. J. Lance, «EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces». 15 de mayo de 2018. doi: 10.1088/1741-2552/aace8c.
- [47] L. Yue *et al.*, «Exploring BCI Control in Smart Environments: Intention Recognition Via EEG Representation Enhancement Learning», *ACM Trans. Knowl. Discov. Data*, vol. 15, n.º 5, pp. 1-20, oct. 2021, doi: 10.1145/3450449.
- [48] Y. Tao *et al.*, «Gated Transformer for Decoding Human Brain EEG Signals», en *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Mexico: IEEE, nov. 2021, pp. 125-130. doi: 10.1109/EMBC46164.2021.9630210.
- [49] J. Liu, Y. Zhao, H. Wu, y D. Jiang, «Positional-Spectral-Temporal Attention in 3D Convolutional Neural Networks for EEG Emotion Recognition». arXiv, 7 de noviembre de 2021. Accedido: 18 de octubre de 2022. [En línea]. Disponible en: <http://arxiv.org/abs/2110.09955>
- [50] A. Dosovitskiy *et al.*, «An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale». arXiv, 3 de junio de 2021. Accedido: 20 de octubre de 2022. [En línea]. Disponible en: <http://arxiv.org/abs/2010.11929>
- [51] Z. Wang, Y. Wang, C. Hu, Z. Yin, y Y. Song, «Transformers for EEG-Based Emotion Recognition: A Hierarchical Spatial Information Learning Model», *IEEE Sens. J.*, vol. 22, n.º 5, pp. 4359-4368, mar. 2022, doi: 10.1109/JSEN.2022.3144317.
- [52] Y. Liu, J. Liu, Z. Long, y C. Zhu, «Tensor Computation», en *Tensor Computation for Data Analysis*, Cham: Springer International Publishing, 2022, pp. 1-17. doi: 10.1007/978-3-030-74386-4_1.

- [53] M. Yang, M. Q. Ma, D. Li, Y.-H. H. Tsai, y R. Salakhutdinov, «Complex Transformer: A Framework for Modeling Complex-Valued Sequence». arXiv, 6 de agosto de 2021. Accedido: 23 de octubre de 2022. [En línea]. Disponible en: <http://arxiv.org/abs/1910.10202>
- [54] «Welcome to Python.org», *Python.org*. <https://www.python.org/> (accedido 25 de octubre de 2022).
- [55] «TensorFlow». <https://www.tensorflow.org/> (accedido 24 de octubre de 2022).
- [56] «Keras: the Python deep learning API». <https://keras.io/> (accedido 24 de octubre de 2022).
- [57] «GitHub - MuSAELab/amplitude-modulation-analysis-module: Amplitude Modulation Analysis Module for Python». <https://github.com/MuSAELab/amplitude-modulation-analysis-module> (accedido 24 de octubre de 2022).
- [58] «Vision Transformer and MLP-Mixer Architectures». Google Research, 24 de octubre de 2022. Accedido: 24 de octubre de 2022. [En línea]. Disponible en: https://github.com/google-research/vision_transformer
- [59] «DEAP: A Dataset for Emotion Analysis using Physiological and Audiovisual Signals». <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html> (accedido 23 de agosto de 2021).
- [60] S. Koelstra *et al.*, «DEAP: A Database for Emotion Analysis ;Using Physiological Signals», *IEEE Trans. Affect. Comput.*, vol. 3, n.º 1, pp. 18-31, ene. 2012, doi: 10.1109/T-AFFC.2011.15.
- [61] I. Zyma, S. Tukaev, y I. Seleznov, «EEG During Mental Arithmetic Tasks». *physionet.org*, 2018. doi: 10.13026/C2JQ1P.
- [62] «PhysioNet Databases». <https://physionet.org/about/database/> (accedido 21 de noviembre de 2022).
- [63] K. Georgiadis *et al.*, «MAMEM Steady State Visually Evoked Potential EEG Database». *physionet.org*, 2016. doi: 10.13026/C29591.
- [64] «Surface Laplacian», en *Analyzing Neural Time Series Data*, The MIT Press, 2014. doi: 10.7551/mitpress/9609.003.0028.
- [65] J. Lee-Thorp, J. Ainslie, I. Eckstein, y S. Ontanon, «FNet: Mixing Tokens with Fourier Transforms». arXiv, 26 de mayo de 2022. doi: 10.48550/arXiv.2105.03824.
- [66] «Signals and Systems | Supplemental Resources», *MIT OpenCourseWare*. <https://ocw.mit.edu/courses/res-6-007-signals-and-systems-spring-2011/> (accedido 30 de noviembre de 2022).
- [67] M. X. Cohen, *Analyzing neural time series data: theory and practice*. en *Issues in clinical and cognitive neuropsychology*. Cambridge, Massachusetts: The MIT Press, 2014.

9 Anexos

Anexo 9.A – Kits BCI

Tabla 9-1: Dispositivos BCI disponibles en el mercado, Algunos son mencionados en la literatura revisada.

Fabricante	Referencia (Serie)	Canales (máximo)	Frecuencia muestreo (MAX) (SPS-Hz)
Geodesic EEG System	GES 400	256	8000
EMOTIV	EPOC+	14	128
OPEN BCI*	Cyton	8	250
OPEN BCI*	Cyton + Daisy	16	250
NeuroScan	SynAmps RT	256	20000
NeuroScan	Neuvo	256	10000
NeuroScan	Grael 4K PSG	32	4096
BioSemi	AIB+AD-Box	256	2000
Brain Products	actiCHamp Plus	160	100000
Brain Products	LiveAmp	64	1000
NeuroSky	ThinkGear AM (TGAM)	1	512
Advanced Brain Monitoring	Stat X	20	256
G.tec	NAUTILUS PRO	32	500
ANT Neuro	eego	256	16000
Neuroelectronics	Enobio	32	500
Interaxon	MUSE	7	256
Cognionics	Quick	30	1000
Cognionics	Mobile	128	1000
mBrainTrain	mobi	24	500

Anexo 9.B – Señales

El presente anexo es una breve descripción de algunos aspectos relacionados con el procesamiento de señales, en el marco fundamental del cálculo de espectrogramas. Su principal fuente es el curso OCW del MIT [66] y su libro base[15].

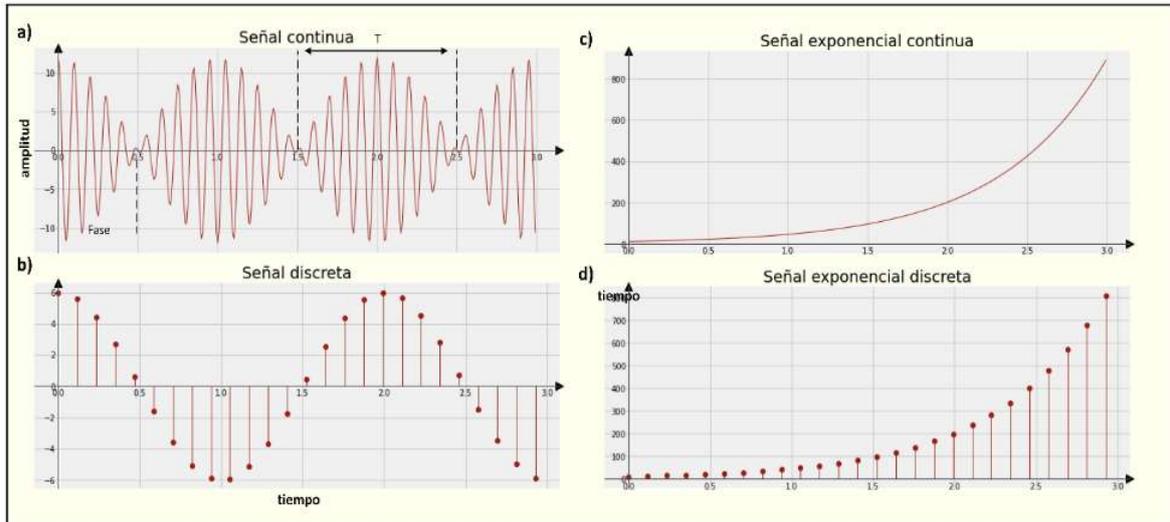


Figura 39: Visualización de diferentes tipos de señales. (figura obtenida y modificada de[15]).

Señales continuas

- Una señal de la forma $x(t) = A \cos \omega_0 t$ es una señal par debido a que $x(t) = x(-t)$
- Una señal de la forma $x(t) = A \sin \omega_0 t$ es una señal impar debido a que: $x(t) = -x(-t)$
- Todas las ondas sinusoidales continuas son periódicas $x(t) = x(t + T_0)$
- En las ondas sinusoidales continuas, el *time shifting* es equivalente al cambio de fase y viceversa
- El periodo de una onda continua corresponde a $periodo = \frac{2\pi}{\omega_0}$

Señales discretas

- Una señal de la forma $x[n] = A \cos \Omega_0 n$ es una señal par debido a que $x[n] = x[-n]$
- Una señal de la forma $x[n] = A \sin \Omega_0 n$ es una señal impar debido a que $x[n] = -x[-n]$
- El periodo para una onda discreta corresponde a $N = \frac{2\pi m}{\Omega_0}$, donde N y m son enteros

Señales exponenciales

- Una señal de la forma $x(t) = Ce^{at}$ es una señal exponencial continua.
- Una señal de la forma $x[n] = Ce^{\beta n} = C\alpha^n$ es una señal exponencial continua.
- Una señal de la forma $x(t) = Ce^{at}$ donde C y a son números complejos talque $C = Ce^{j\theta}$ y en este caso $a = r + j\omega_0$ entonces: $x(t) = |C|e^{rt}e^{j(\omega_0 t + \theta)}$ es una señal exponencial compleja
- Teniendo presente la relación de Euler $\cos(\omega_0 t + \theta) + j\sin(\omega_0 t + \theta) = e^{j(\omega_0 t + \theta)}$ entonces $x(t) = |C|e^{rt}\cos(\omega_0 t + \theta) + j|C|e^{rt}\sin(\omega_0 t + \theta)$

Función escalón e impulso unitario.

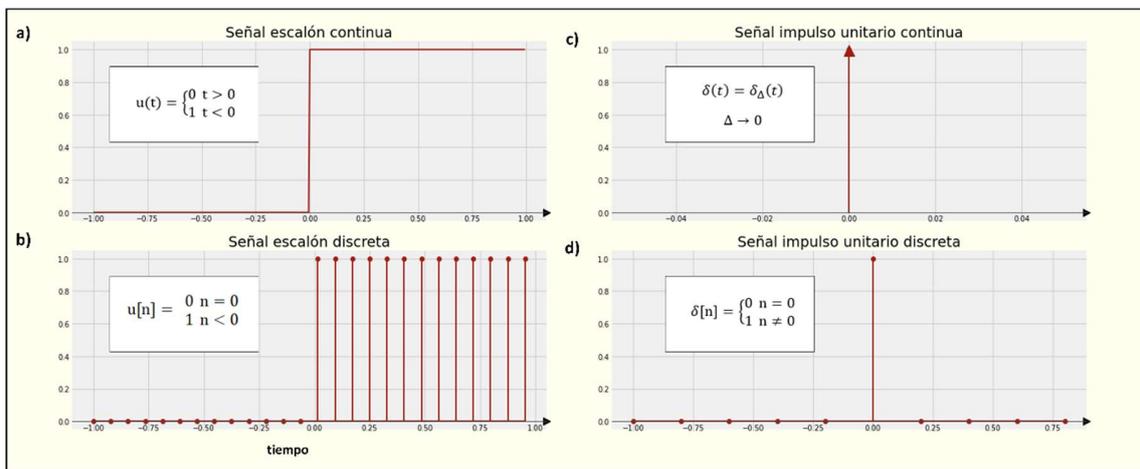


Figura 40: Visualización de diferentes tipos de señales. (figura obtenida y modificada de[15]).

- La relación entre la función de impulso unitario y la función escalón está dada por $\delta[n] = u[n] - u[n - 1]$, esto se conoce como: *first difference*.
- La secuencia de una función escalón es la “running sum” de la función de impulso unitario
- La definición de la función de impulso unitario continua es análoga a su contraparte discreta, en este caso se define como la primera derivada en el rango cero-delta cuando el delta tiende a cero $\delta(t) = \frac{du(t)}{dt}$; $\delta(t) = \frac{du_{\Delta}(t)}{dt}$; $\delta(t) = \delta_{\Delta}(t)$ cuando $\Delta \rightarrow 0$ En este caso cada vez que presente un delta más pequeño, se incrementará la altura, por lo tanto, el resultado de dicha área equivale a 1.

Sistemas

Un sistema es una transformación de una señal de entrada en una señal de salida:

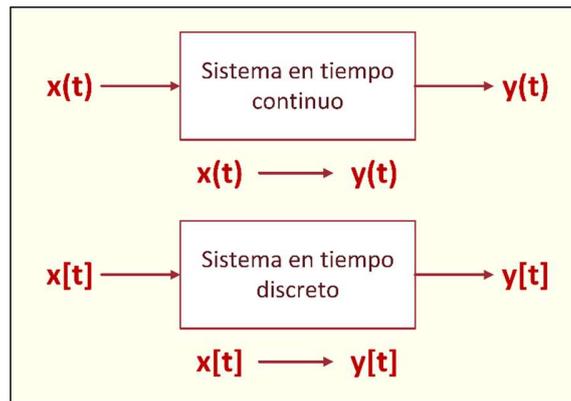


Figura 41: Descripción gráfica de un sistema, se puede apreciar la notación para sistemas continuos y discretos con respecto a la variable tiempo (figura obtenida y modificada de[15])

Interconexiones de los sistemas

Existen sistemas interconectados en serie o cascada y sistemas interconectados en paralelo, en los sistemas en serie el orden cuenta, por otro lado, para los sistemas en paralelo el orden no es relevante.

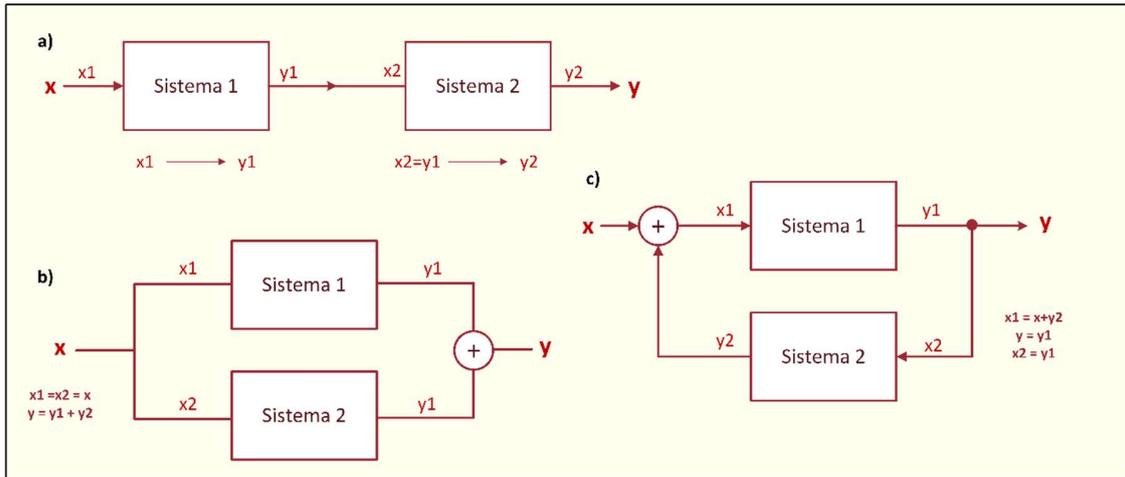


Figura 42: Tipos de sistemas según su interconexión, a) sistemas en serie, b) sistema en paralelo, c) sistema con retroalimentación. (figura obtenida y modificada de[15])

Propiedades de los sistemas

Sin memoria: Si la salida depende únicamente de la entrada en mismo instante de tiempo

Invertible: Si diferentes entradas conducen a diferentes salidas o en otras palabras si es posible inferir el valor de la entrada a partir del resultado

Causalidad: Un sistema es causal si la salida en un determinado t depende solo de valores en el presente y del pasado.

Todos los sistemas sin memoria son sistemas causales.

Estabilidad: Un sistema estable es donde las pequeñas entradas no producen resultados o salidas divergentes. Si una entrada acotada produce una salida acotada entonces es un sistema estable. Las interconexiones tipo retroalimentación son importantes en la estabilización de un sistema inestable

Invariante en el tiempo: Un sistema es “invariante en el tiempo” si un desplazamiento en el z de la entrada produce un desplazamiento z en la salida

Linealidad:

- Una entrada cero produce una salida cero
- Si $\{x_1(t) = y_1(t)\} \wedge \{x_2(t) = y_2(t)\} \rightarrow \{x_1(t) + x_2(t) = y_1(t) + y_2(t)\}$

En teoría si las entradas constituyen una combinación lineal entonces, las salidas deberían constituir una combinación lineal

Convolución

Recordemos que las señales se pueden representar como una combinación entre una función y la función impulso unitario, en ambos casos discretos y continuos

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]$$

$$x(t) = \int_{-\infty}^{+\infty} x(\tau)\delta(t-\tau)d\tau$$

En el caso para función las entradas se pueden expresar como una combinación lineal entre x . δ donde δ presenta un *time shift*. Representando el sistema de la siguiente manera:

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k]h_k[n]$$

Donde, la función $h[n]$ es arbitraria, entonces combinando las definiciones:

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]$$

Sistema lineal:

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h_k[n] \quad ; \delta[n-k] \rightarrow h_k[n]$$

En caso de invariante en el tiempo:

$$\delta[n-k] \rightarrow h_k[n]$$

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h[n-k]$$

Propiedades de la convolución:

La convolución es conmutativa: $x[t] * h[t] = h[t] * x[t]$

La convolución es asociativa: $x[t] * (h_1[t] * h_2[t]) = (x[t] * h_1[t]) * h_2[t]$

La convolución es distributiva: $x[t] * (h_1 + h_2[t]) = x[t] * h_1[t] + x[t] * h_2[t]$

Teorema de la convolución:

Este teorema establece que la convolución en el dominio del tiempo es lo mismo que la transformada inversa de la multiplicación en el dominio de la frecuencia.

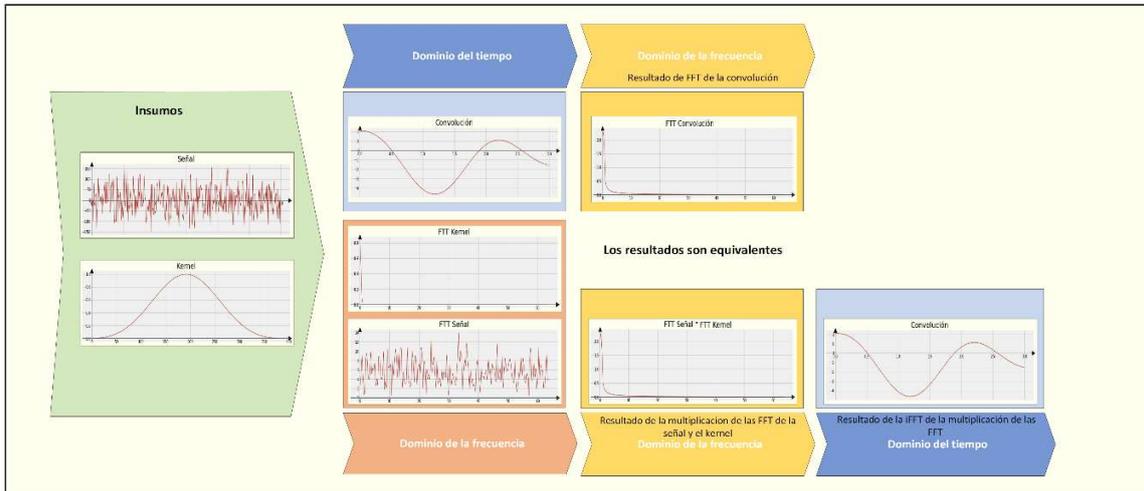


Figura 43: Teorema de convolución, La convolución de una señal en el dominio del tiempo también se puede obtener calculando la transformada inversa de la multiplicación punto por punto de las FFT de la señal y el kernel (figura obtenida y modificada de [67])

Anexo 9.C – Preprocesamiento (Visualización EEG)

La herramienta MNE®[13], es compatible con el lenguaje Python® y coloca a disposición funcionalidades útiles para la elaboración de scripts dedicados al procesamiento de señales EEG, en el siguiente ejemplo se muestra una inspección general de un sujeto de pruebas para uno de los conjuntos de datos. Aunque no se cuenta con la pericia de un ojo experto, es importante verificar las señales en su fase inicial para detectar anomalías o desperfectos evidentes aún para la mirada inexperta

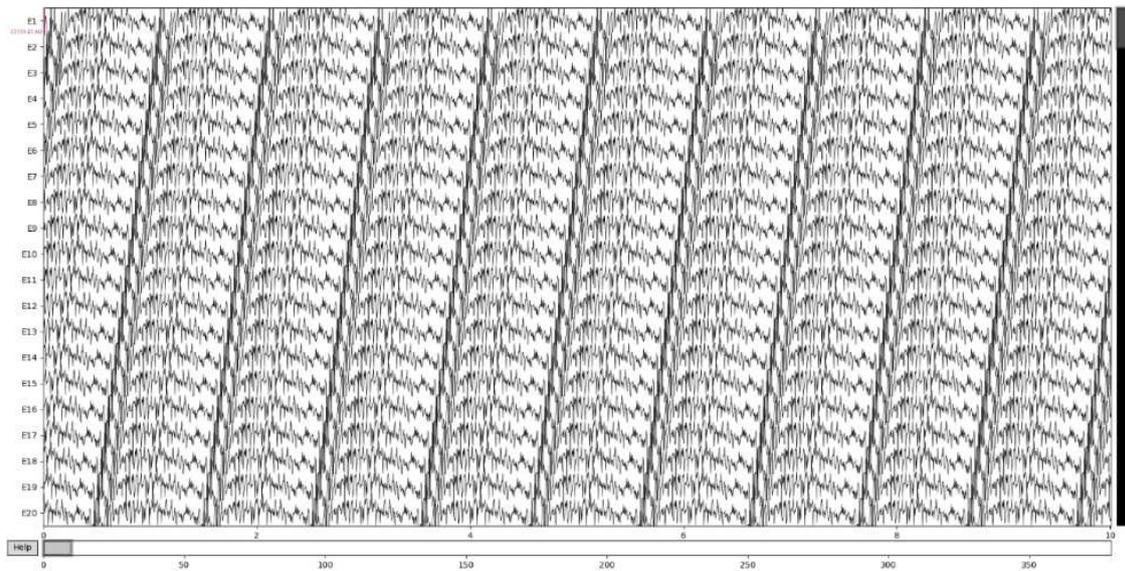


Figura 44: Visualización las señales EEG, pertenecientes a un sujeto del conjunto de datos MSSVEPDB

Anexo 9.D – Preprocesamiento (ERP)

Los ERP son de gran ayuda en la visualización del comportamiento de la actividad cerebral, ante un estímulo determinado, sin embargo, no son objeto del presente estudio.

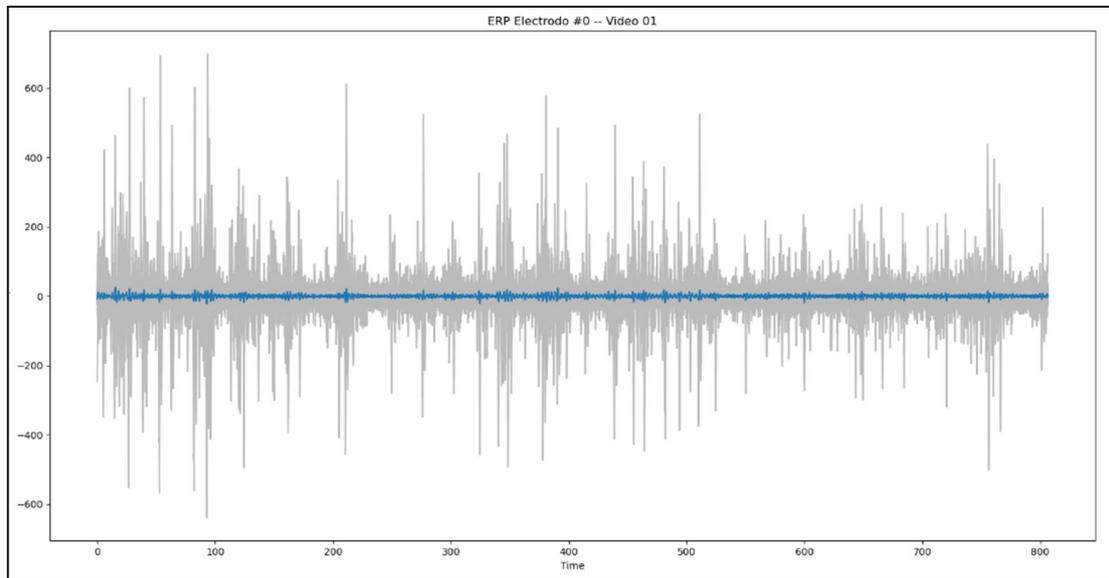


Figura 45: Cálculo y visualización de un ERP para el conjunto de datos DEAP, el ERP utiliza los datos del electrodo 0 para todos los participantes que observaron el video 1, se puede apreciar la línea azul que refleja las componentes de la... Estos son posible bajo el supuesto de la aleatoriedad del ruido, visto como una variable aleatoria normal

Anexo 9.E – Visualización Topográfica

En la etapa de preprocesamiento, también se realizaron inspecciones visuales de las señales EEG a nivel topográfico, esto con el fin de trabajar algunos insights espaciales.

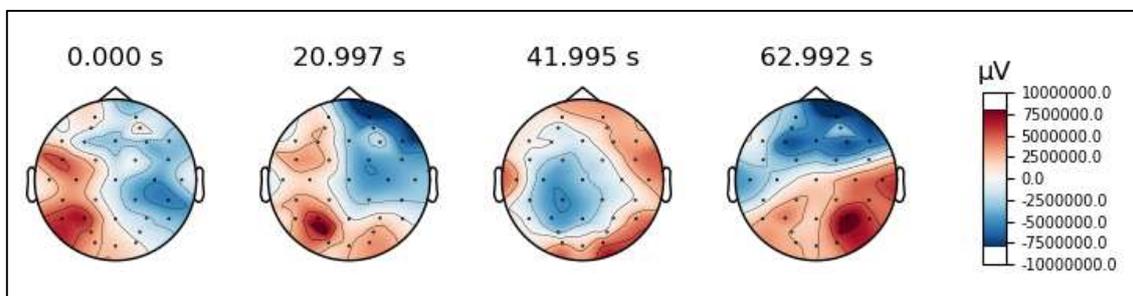


Figura 46: Visualización topográfica de un EEG perteneciente al conjunto de datos DEAP. Este tipo de visualizaciones ayuda a detectar centros de actividad cerebral (color rojo), también ofrece visibilidad de dicha actividad a través del tiempo.

Anexo 9.F – Preprocesamiento (Visualización EEG completo)

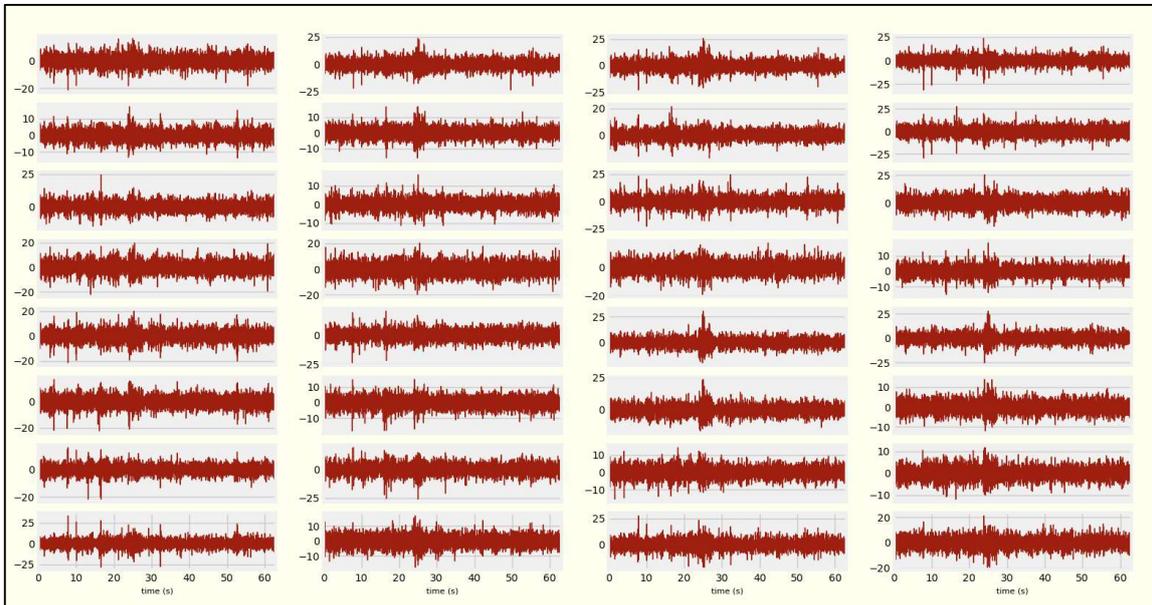


Figura 47: Visualización de un conjunto de señales EEG para uno de los trials del DEAP. Este tipo de visualización ayuda a verificar el estado inicial de las señales, con el fin de detectar anomalías a simple vista, como por ejemplo un electrodo en estado inactivo. (fuente propia).

Anexo 9.G – Preprocesamiento (Visualización de espectrogramas wavelet)

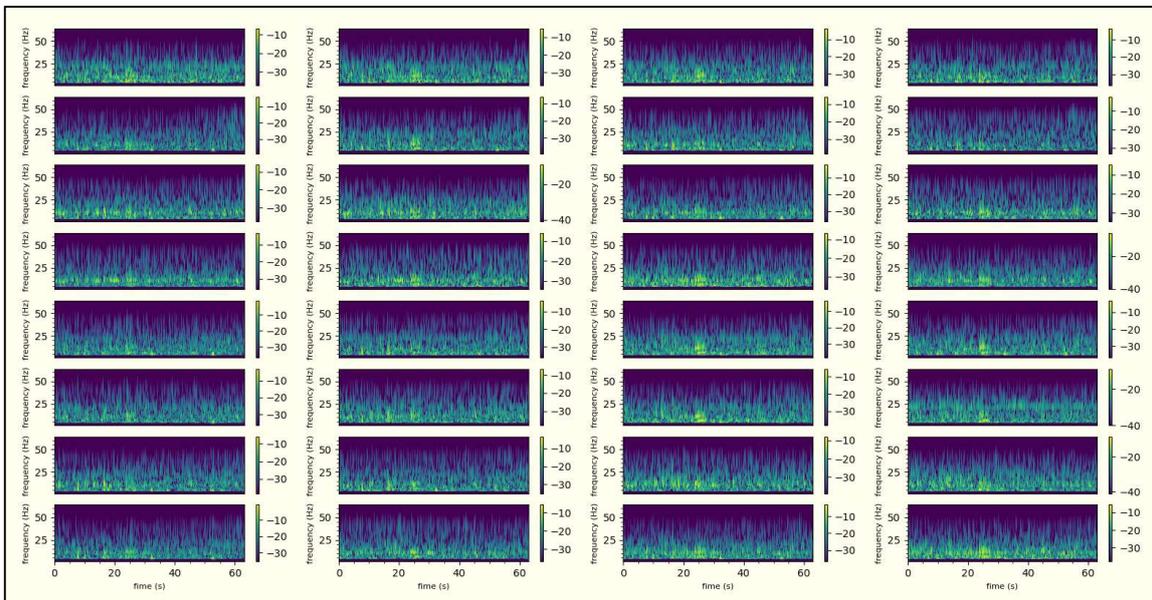


Figura 48: Visualización de un conjunto de espectrogramas basados en Wavelet para un trial del DEAP, se utilizaron 6 ciclos en el diseño de la wavelet tipo “Morlet”, nótese que estos espectrogramas ofrecen una mejor resolución a nivel temporal ya que a mayor número de ciclos mejora la resolución de las frecuencias en detrimento de la resolución temporal.

Anexo 9.H – Preprocesamiento (Visualización de espectrogramas de modulación wavelet)

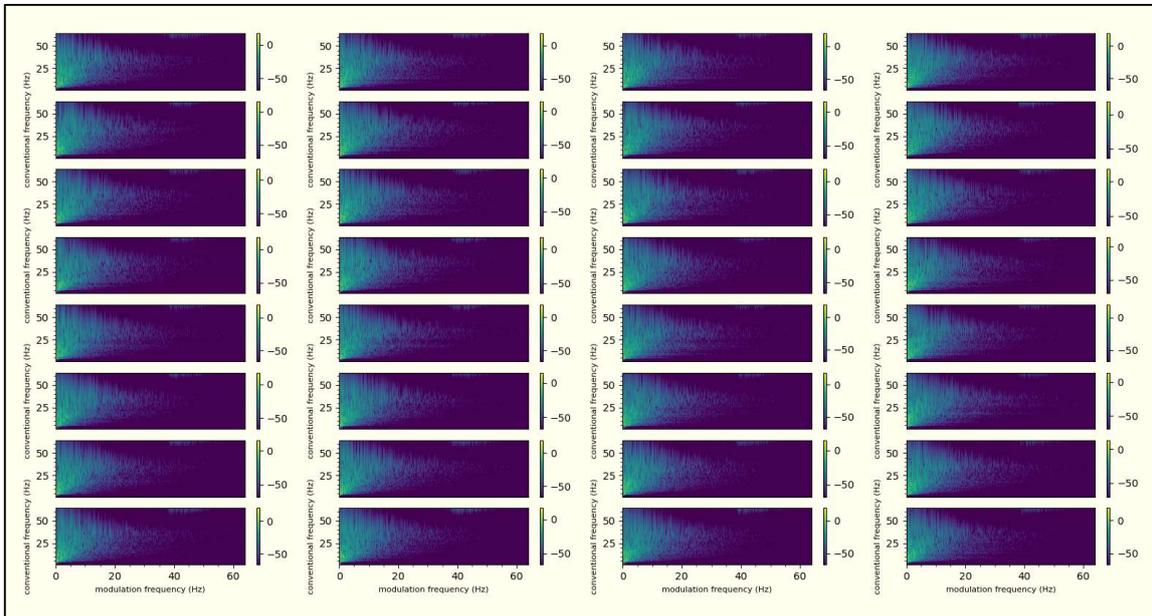


Figura 49: Visualización de un conjunto de espectrogramas de modulación basados en Wavelet para un trial del DEAP, se utilizaron 6 ciclos en el diseño de la wavelet para ambas transformadas, no obstante, para la segunda transformada se utilizó una ondícula tipo “boxcar”, en este caso visualmente se aprecian pocas diferencias a nivel visual.

Anexo 9.1 – Glosario

BCI:	<i>Brain Computer Interface (interfaz cerebro computador)</i>
Benchmark:	Bajo el contexto del presente trabajo, se refiere a un estudio comparativo.
Dataset:	Conjunto de datos pertenecientes a un experimento o una recopilación.
Dongle:	Dispositivo periférico que cumple una funcionalidad específica, ejemplo: comunicación o almacenamiento.
ECG:	Electrocardiograma.
EEG:	Electroencefalograma.
Embedding:	Proceso de reorganización de un conjunto de datos, puede implicar procesos de codificación y de reducción de dimensionalidad.
EMG:	Electromiograma
EOG:	Electrooculograma
Insight:	En ciencia de datos constituye una idea clave que ayuda al entendimiento de características particulares asociadas al problema-contexto de donde provienen los datos
Kernel:	En el procesamiento de señales, específicamente en la operación de convolución el kernel es la función ventana que se utiliza para llevar a cabo dicha operación
One-hot:	Proceso de codificación utilizado en ciencia de datos y aprendizaje de máquina, el cual consiste en codificar cada uno de los elementos de un conjunto de categorías en una secuencia única de [0-1]. Esto puede implicar tanto operaciones a nivel de bits como operaciones a nivel de arreglos
PSD:	Power Spectral Density.
RF:	Random Forest.
Script:	Secuencia de comandos de software desarrollada para llevar a cabo la realización una funcionalidad específica.

- seqToseq:** Sirve para referenciar modelos de traducción o codificación, básicamente una secuencia transformada en otra secuencia
- SVM:** Support Vector Machine
- Toolbox:** En el contexto del software, constituye un conjunto de librerías que están diseñadas para facilitar el proceso de implementación de un programa bajo un contexto específico
- Transformer:** Es una técnica de aprendizaje profundo para la resolución de diferentes tareas como la codificación, la autogeneración.
- ViT:** Vision Transformer
- Wearable:** Dispositivo de electrónico, el cual funciona como una prenda de vestir. Estos dispositivos generalmente tienen la capacidad de adquirir bioseñales provenientes del usuario.

Anexo 9.J – Procesamiento (MST con Fase A)

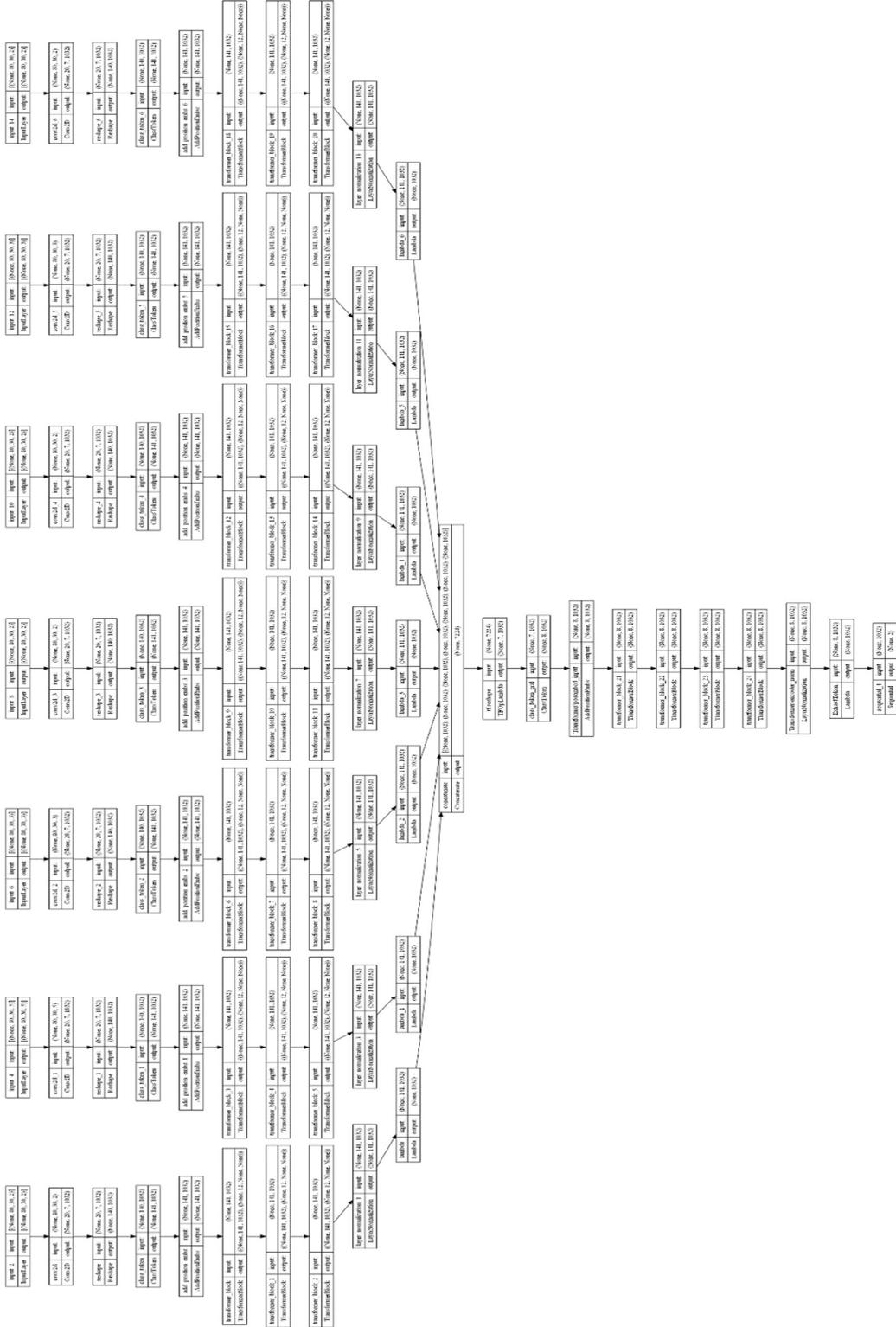


Figura 50: Ejemplo de un modelo MST con fase A compilado en Tensorflow. (fuente propia)

Anexo 9.K – Procesamiento (MST con Fase B)

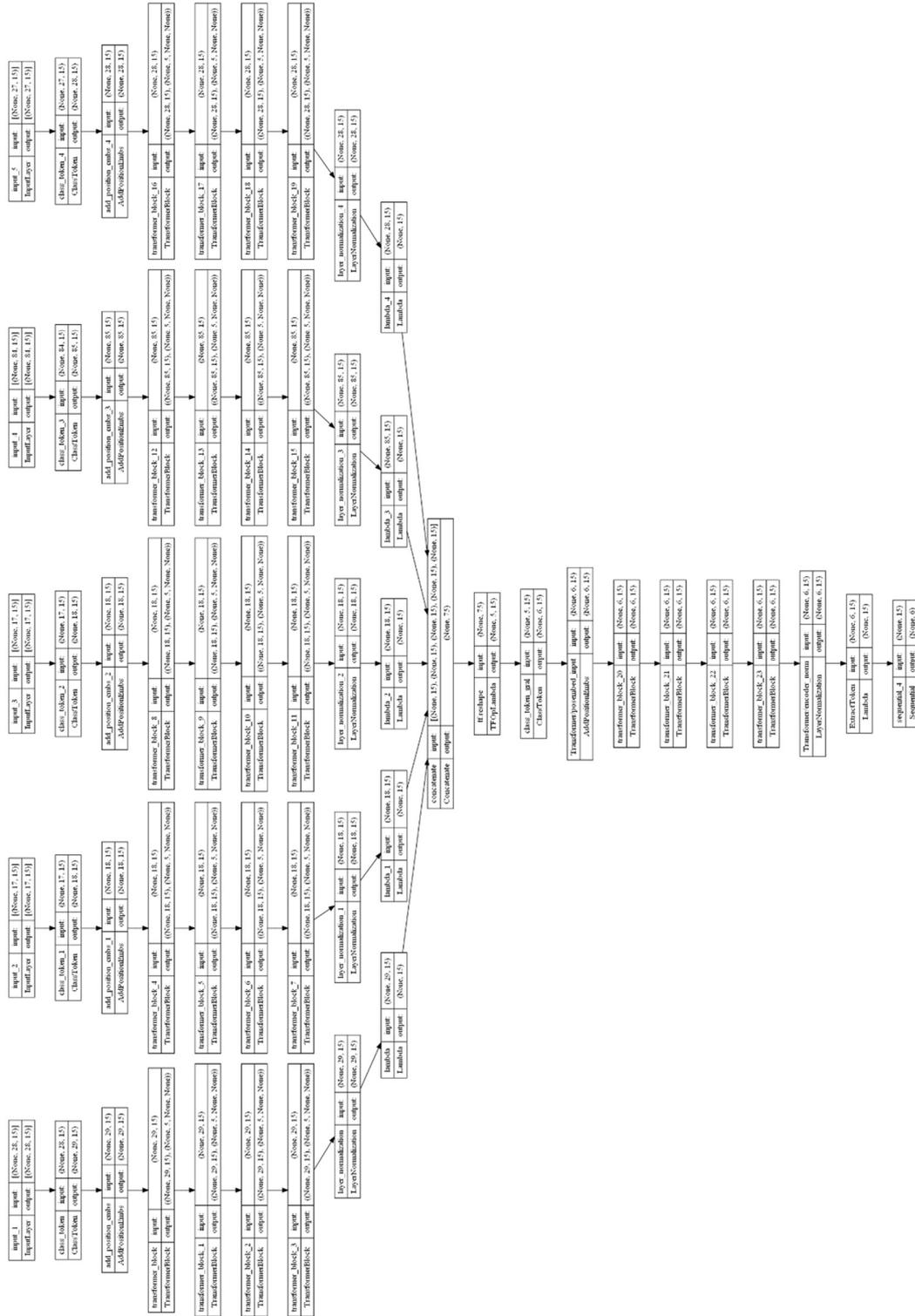
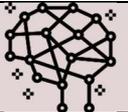
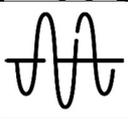
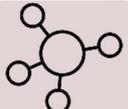


Figura 51: Ejemplo de un modelo MST con fase B compilado en Tensorflow. (fuente propia)

Anexo 9.L – Referencias de imágenes gratuitas descargadas de internet

Imagen	Atribución
	Brain icons created by Eucalyp - Flaticon
	Signal icons created by Vitaly Gorbachev - Flaticon
	Data mining icons created by Becris - Flaticon
	Digital transformation icons created by Freepik - Flaticon
	Data science icons created by surang - Flaticon
	Heartbeat icons created by Freepik - Flaticon
	Application icons created by Nhor Phai - Flaticon
	Electrode icons created by Freepik - Flaticon
	Estructura iconos creados por Vectoricons - Flaticon