Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales
Departamento de Computación

Analizando cerebros con 500 dólares: Validación y estudio de redes de conectividad para un equipo de EEG portátil y económico

Tesis de Licenciatura en Ciencias de la Computación

Mathias Gatti

Directores: Juan Kamienkowski y Daniel Fraiman

Buenos Aires, 2019

Resumen Los estudios de electroencefalografía (EEG) han mostrado ser una herramienta fundamental como método complementario para el diagnóstico clínico y la comprensión de los procesos cognitivos humanos. Los registros de EEG usualmente se restringen a laboratorios o ambientes similares, lo cual supone una limitación para estudiar poblaciones que son difícilmente trasladables al laboratorio o sufren de estrés o comportamiento poco natural al cambiar de entorno. Recientemente, surgieron equipos de EEG de un costo considerablemente menor y mayor portabilidad con un objetivo más recreativo. Sin embargo, el uso de estos equipos podría a largo plazo abrir los horizontes de aplicación de distintas protocolos tanto clínicos como de investigación. En esta tesis investigamos los límites de estos aparatos, buscando determinar la aplicabilidad en los protocolos usuales de medición de actividad cerebral tanto a nivel clínico como a nivel de investigación.

Una manera de analizar los registros de EEG es a través de redes funcionales. Estas redes se construyen a partir de la similaridad entre los distintos electrodos. Pero existen muchas formas de definir esta similaridad. En la presente tesis evaluamos las diferentes medidas de similaridad tanto en términos de su robustez como de su capacidad de predicción de distintos atributos de los sujetos. Este análisis lo realizamos para dos tipos de aparatos de EEG (bajo y alto costo) buscando determinar las diferencias.

Entonces, en primer lugar, se ha desarrollado un entorno de experimentación que posibilitó la toma de datos en el laboratorio y, de forma masiva, en escuelas. Se analizaron dos conjuntos de datos, el primero en adultos utilizando tanto el EEG de bajo costo y como el de alta resolución, y el segundo en chicos de preescolar en las escuelas sólo con el EEG de bajo costo.

Los resultados mostraron que existen medidas de similaridad suficientemente robustas para realizar análisis de conectividad cerebral. En particular, se realizaron distintos experimentaciones utilizando el conjunto de datos de adultos, de las cuales las medidas de PLV (*Phase Locking value*) y correlación (tanto de Spearman como de Pearson) dieron consistentemente los mejores resultados. Luego se analizaron los datos de chicos de preescolar utilizando estas medidas, y se lograron predecir correctamente (AUC$>$0.67) algunas variables individuales como el nivel de sueño, sexo y distinción entre ojos abiertos y cerrados. Basados en nuestros resultados y en los protocolos analizados podemos afirmar que ambos equipos tienen una *performance* muy parecida. Como paso final se publicó todo el código relevante permitiendo que cualquiera pueda replicar la toma de datos y análisis de conectividad con el sistema Emotiv.

Esta tesis prepara el camino para el registro masivo de actividad cerebral con distintos paradigmas y, en particular, el estudio de los cambios en la conec-

tividad cerebral a lo largo del desarrollo, en ambientes más ecológicos.

Abstract Electroencephalography (EEG) studies have shown to be a fundamental tool in the understanding of human cognitive processes and as a complementary method for clinical diagnosis. EEG recordings are usually restricted to the lab's environment, which is an important limitation to the study of populations that are difficult to move to the lab or that this environment causes an effect by itself (such as little naturalness or stress). Recently, new EEG equipment at a significant lower cost and greater portability emerged mainly for gaming. However, the use of this equipment could open the possibility in the long term of application of different clinical and research protocols in more ecologically valid environments.

It is possible to build networks from the EEG recordings by looking at the similarity between the different electrodes, but there are many ways to define this similarity. In this thesis, we aimed to compare the resulting network by applying different measures, in terms of their robustness and their ability to predict different attributes of the subjects.

Thus, in the first place, we developed an experimentation environment that allowed both data collection in the lab and massive data collection in the schools. Two data sets were analyzed, the first including the low-cost EEG and the high-resolution EEG in adults, and the second with only the low-cost EEG in preschool children at the schools.

The results showed that some similarity measures are robust enough to analyze brain activity at the level of networks using the Emotiv system. In particular, we performed different experiments using the adult dataset, of which the PLV (*Phase Locking value*) and the correlation measures (both Spearman and Pearson) consistently gave the best results. Then, the preschool children dataset were analyzed using these measures and some individual variables, such as sleep level, sex and the distinction between open and closed eyes, were successfully predicted (AUC$>$0.67). As a final step, the code was released online, allowing anyone to replicate the data collection and connectivity analysis with the Emotiv system.

This thesis paves the way for the massive recording of brain activity with different paradigms and, in particular, the study of changes in brain connectivity throughout development, in more ecologically valid environments.

Índice

1. Introducción **6**
1.1. Redes en el Cerebro 6
1.1.1. Medidas de Grafos 8
1.1.2. Predicción a partir de redes cerebrales 10
1.2. Electroencefalografía (EEG) 10
1.2.1. Ondas en el cerebro 12
1.2.2. Equipamiento en EEG 13

2. Objetivo **16**

3. Métodos **18**
3.1. Desarrollo de los experimentos 18
3.1.1. Emotiv Bindings 18
3.1.2. Go-NoGo 18
3.1.3. Resting 19
3.2. Procedimiento 19
3.3. Electroencefalógrafos utilizados 20
3.4. Preprocesamiento de las señales 21
3.5. Conectividad Cerebral 22
3.5.1. Phase Locking Value 23
3.5.2. Phase Locking Index 23
3.5.3. Coherence 23
3.5.4. Spearman Correlation 23
3.5.5. Pearson Correlation 24
3.5.6. Otros 24
3.6. Estadística de redes 24
3.6.1. Test de Pareo 24
3.6.2. Análisis de los grafos subyacentes 26
3.6.3. Relación entre modularidad y densidad de las aristas . . 26
3.6.4. Test de clasificación 27
3.6.5. Random Forest 28
3.7. Métodos de validación de la clasificación 29
3.7.1. Cross Validation 29
3.7.2. Curva ROC y el AUC 29
3.7.3. Tests Estadísticos 30

4. Búsqueda de Métricas Robustas entre equipos en Adultos **31**
4.1. Desarrollo de un entorno para interacción con Emotiv en Python 31
4.2. Sensibilidad de las redes funcionales 31
4.3. Sensibilidad de los equipos de medición 32
4.3.1. Algunos Ejemplos 33
4.3.2. Ejemplos de Conectividad 35
4.4. Comparación entre Equipos 36
4.5. Clasificación de ojos abiertos y cerrados 38
4.6. Conclusión 40

5. Predicción de Características Individuales en Niños **41**
5.1. Clasificación de Ojos Abiertos: Atributos Principales 42
5.2. Predicción de sexo 42
5.2.1. Curva ROC 43
5.2.2. Matrices de Confusión 43
5.2.3. Atributos Principales 46
5.3. Predicción de cantidad de sueño 47
5.3.1. Curva ROC 48
5.3.2. Matrices de Confusión 48
5.3.3. Atributos Principales 51
5.4. Predicción de nivel educativo del tutor 52
5.4.1. Conclusión 53

6. Conclusiones **54**
6.1. Trabajo a Futuro 55

A. Apéndice **57**
A.1. Cantidad de samples por grabación 57
A.2. Electrodos de Emotiv y Biosemi 57

1. Introducción

El cerebro es capaz de procesar una gran cantidad de información de fuentes y características muy variadas, integrándola para dar lugar a comportamientos coherentes. Para ello posee una serie de áreas especializadas (según el tipo de información de entrada y el tipo de comportamiento de salida por ejemplo) y una exquisita coordinación entre ellas. De esta manera se da lugar a la noción de redes a gran escala en el cerebro, combinando regiones extensas del mismo, a diferencia de las redes locales que se asocian al comportamiento de conjuntos de neuronas [Spo10]. A lo largo de esta tesis se evaluará el uso de distintas técnicas para medir estas redes en un equipo de electroencefalografía (EEG) portátil y de bajo costo, así como también la capacidad de predicción de las mismas sobre distintas características de los participantes. A partir de lo cual se puedan realizar y analizar experimentos por ejemplo en las escuelas, lo cual constituye una línea de investigación muy activa en el Laboratorio de Inteligencia Artificial Aplicada (Instituto de Ciencias de la Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET).

1.1. Redes en el Cerebro

La *conectividad funcional* se define, en términos generales, a partir de la correlación entre eventos distantes en el cerebro. En ese sentido, la conectividad funcional implica que dos áreas actúan de forma coordinada, pero no dice nada respecto a cómo se comunican. A partir de medidas de conectividad funcional se construyen redes que pueden ser abordadas, por ejemplo, con los métodos usuales de teoría de redes o grafos [Spo10; Fra+14; Cer+17; Amo+17] o comparándolos directamente [Cer+17; FF18].

La conectividad funcional ha sido estudiada con una gran variedad de técnicas de medición, como la resonancia magnética funcional (fMRI), magnetoencefalografía (MEG), electroencefalografía (EEG), etc. La técnica fMRI tiene la virtud de tener una gran resolución espacial, lo cual es muy provechoso para el estudio de redes ya que puede asignarse cada nodo a una región anatómica precisa. Sin embargo, métodos como el MEG o el EEG poseen una resolución temporal mucho mayor permitiendo estudiar mejor las correlaciones temporales de la señal, en particular en distintas bandas de frecuencia. Finalmente, dentro de estos equipos el EEG es el que posee mayor portabilidad, lo cual resultará relevante para el proyecto.

En la figura 1 se puede ver un ejemplo de los pasos por los que pasan los datos hasta su análisis final. Las señales tomadas por el EEG filtran bandas

de frecuencia que están por fuera del rango de interés. Luego, dependiendo el tipo de experimento, o bien se identifican las regiones donde ocurrieron eventos relevantes y se recortan estas series temporales o se procesa todo sin particionar, esto último ocurre por ejemplo cuando se realiza un experimento de resting donde no hay estímulos o eventos particulares que sean de interés. Acto seguido se eliminan fuentes de ruido como contracciones musculares. Después, dependiendo el tipo de experimento, o bien se genera una matriz de correlación analizando la similitud entre distintos electrodos o se genera una serie temporal promedio de la cual se estudia su forma de onda. A partir de estos resultados se realizan tests estadísticos para intentar verificar si el fenómeno esperado se observa de forma significativa.

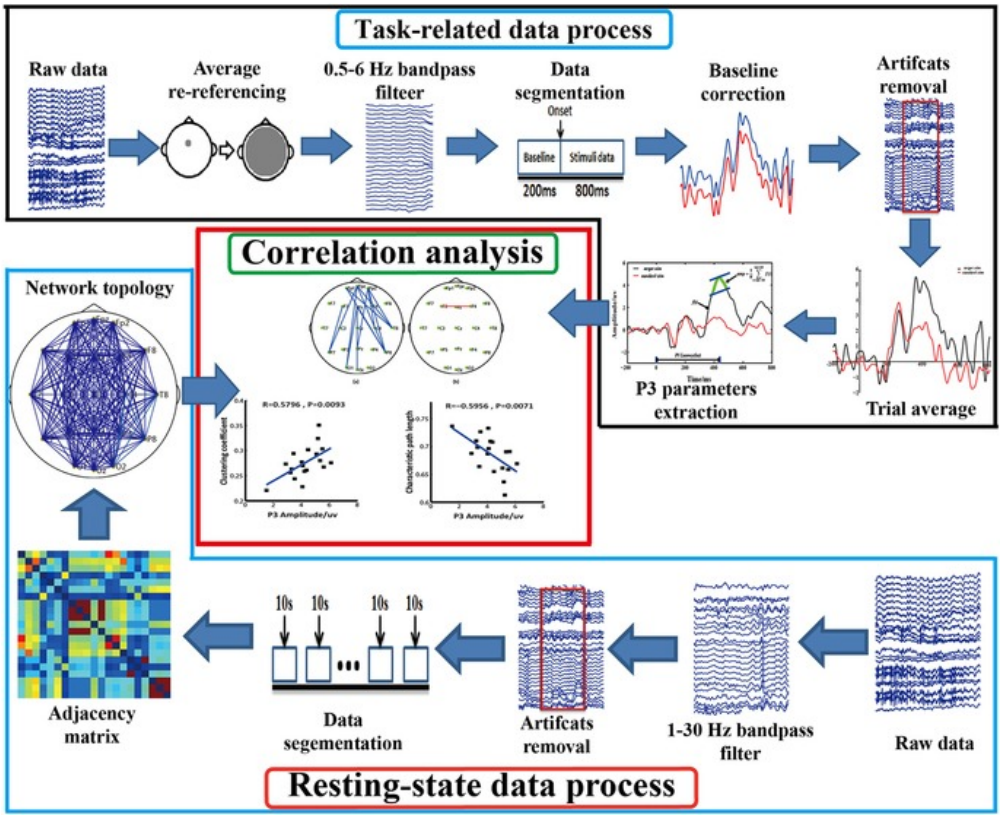

Figura 1: Ejemplo de flujo de los datos, desde los datos crudos del EEG hasta el análisis del grafo. Imagen extraída de https://www.ncbi.nlm.nih.gov/pubmed/28938855.

Los estudios de conectividad funcional han mostrado ser de suma utilidad. Al día de hoy se utilizan en algunos centros de neurocirugía para poder valorar el potencial impacto de cierta operación y así tomar la mejor decisión quirúrgica. También existen muchos esfuerzos que buscan mostrar que la información dada por la conectividad funcional contribuye al diagnóstico temprano de ciertas disfunciones o trastornos cerebrales[GC+15], como así también a la

detección y caracterización de distintos estados de conciencia [CH+14]. Estudios de conectividad funcional también han sido aplicados en una gran variedad de poblaciones para mostrar su potencial como herramienta de análisis. Existen desde estudios que muestran cómo se modifican las redes funcionales por el entrenamiento [Amo+17; Ast+15], hasta estudios que muestran las diferencias en las redes funcionales de hombres y mujeres heterosexuales [Ing+14], pasando por estudios que muestran cómo se evolucionan las redes con la maduración del sujeto [Dos+10].

Una condición de estudio muy común para analizar las redes cerebrales funcionales es la de estado de reposo (*resting state*). En estos estudios se le pide a los participantes que se mantengan quietos, sin dormirse, durante aproximadamente 5 minutos. Luego, se construye la red funcional a partir de las correlaciones (u otra medida de similaridad) entre las series temporales de las distintas regiones o electrodos. Estos estudios implican típicamente que los participantes se acerquen al centro de investigación, y el protocolo suele ser costoso y tedioso por lo que limita tanto el número de participantes como las poblaciones que se pueden alcanzar.

1.1.1. Medidas de Grafos

Una de las herramientas fundamentales para el análisis de redes cerebrales es la teoría de grafos. Esta consiste en el estudio matemático de redes [Bar+16a] a través de abstracciones llamadas grafos. Mientras que algunos grafos son puramente matemáticos y no tienen aplicaciones directas, el uso de esta rama de la matemática es invaluable para el análisis de problemas del mundo real. Algunos ejemplos clásicos son el de Kirchhoff, un físico del siglo XIX que formuló varios teoremas fundacionales de la teoría de circuitos moderna partiendo de la representación de los mismos a través de grafos, o Cayley quien aplicó la teoría de grafos para enumerar isómeros químicos (moléculas que contienen el mismo número de átomos unidos en diferentes formas).

Durante las última dos décadas el estudio de redes se ha expandido para incluir la descripción estadística de sistemas mucho más grandes. Esto permitió analizar redes mucho mas grandes, permitiendo aplicarlo a estudios que van desde las redes sociales hasta las redes de conectividad cerebral.

Un grafo consiste de un conjunto de nodos y aristas. Los nodos representan los elementos fundamentales del sistema (personas, o regiones del cerebro por ejemplo). Las aristas representan conexión entre pares de nodos (amistad, o comunicación entre dos partes del cerebro). Los grafos pueden representarse con matrices cuadradas de tamaño $n \times n$ donde n es la cantidad de nodos. En cada coordenada (i,j) de la matriz pueden haber valores, normalmente 0 o 1 lo

cual indica si hay o no una conexión entre el nodo i y el nodo j.

A esta representación del grafo se le pueden aplicar fácilmente distintas medidas para estudiar sus propiedades. Por ejemplo la cantidad de conexiones o el camino más eficiente entre un par de nodos o el coeficiente de *clustering* que indica cuán conectados están los vecinos de un nodo.

Una medida de grafos muy popular es la de modularidad. Nociones de modularidad suelen utilizarse para analizar la estructura de las redes neuronales [Tag+13; Meu+09]. La idea intuitiva es intentar representar qué tan interconectado está un sistema, o qué tan fácilmente sus componentes podrían ser separados y recombinadas. En términos del cerebro, esto nos habla de regiones que se coordinan, o no, de acuerdo a la función que representan.

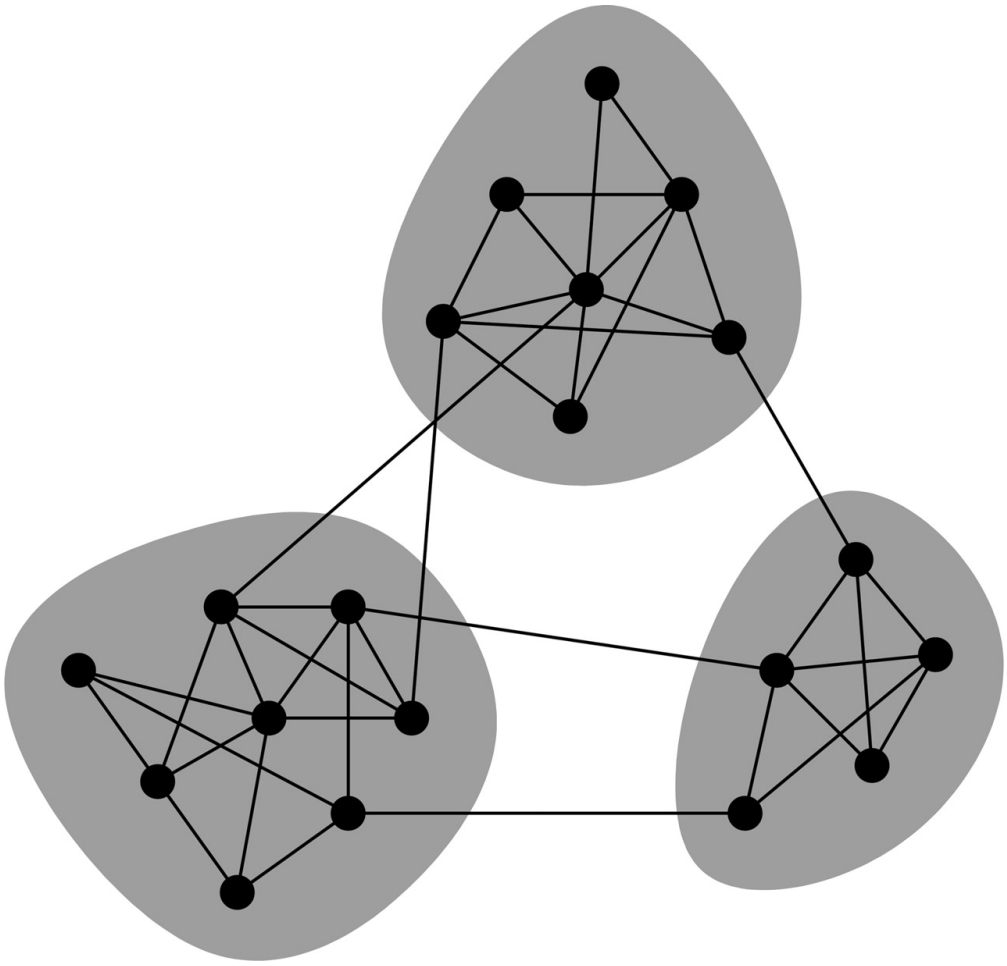

Figura 2: Imagen extraída de Modularity and community structure in networks por M. E. J. Newman

Las unidades neuronales o regiones cerebrales que forman un *cluster* o módulo densamente conectado constituyen un sistema cerebral funcionalmente coherente. Por el contrario, unidades neuronales que pertenecen a diferentes

clusters o módulos no comparten tanta información y permanecen funcionalmente separados unos de otros. Por lo tanto, medidas de modularidad resaltan un aspecto particular de la organización funcional del cerebro, su tendencia a formar subsistemas segregados con propiedades funcionales especializadas (ver figura 2). La identificación de módulos en las redes cerebrales es un primer paso importante hacia la caracterización de estos subsistemas.

1.1.2. Predicción a partir de redes cerebrales

A partir de las redes cerebrales es posible inferir ciertas acciones que puede estar haciendo un individuo en un momento dado, también es posible predecir algunas características personales con cierto nivel de éxito. Utilizando técnicas de *machine learning* se pueden desarrollar clasificadores que predicen variables de interés a partir de redes de conectividad.

Estudios de clasificación han sido utilizados para predecir el desempeño de chicos en tareas perceptuales y de memoria de trabajo [Bar+16b; Kuo+16], e inclusive, a partir de la conectividad, predecir variables individuales como la edad [Dos+10].

1.2. Electroencefalografía (EEG)

Como ya se mencionó en este proyecto se trabajará a partir de registros con electroencefalógrafos, los cuales permiten medir la actividad eléctrica del cerebro de forma no invasiva. Para ello, se apoyan electrodos sobre el cuero cabelludo conectados a un amplificador. En pocas palabras, la señal del EEG es el voltaje asociado a la corriente resultante de la sincronización de un conjunto de neuronas en una región de la corteza dada. Estos dispositivos suelen tener múltiples electrodos distribuidos en el cuero cabelludo y permiten registrar la actividad eléctrica espontánea del cerebro a lo largo de un período de tiempo. Se puede ver un ejemplo de un EEG en la figura 3.

Figura 3: Ejemplo de una electroencefalografía utilizando un EEG clásico en un laboratorio. Imagen extraída de (https://www.biosemi.com/products_activeone.htm)

En contextos clínicos se lo utiliza para ayudar a diagnosticar, entre otras cosas, trastornos de sueño, tumores, daño cerebral y epilepsia[1].

Algunas de las técnicas más comunes para el análisis de las mediciones son los relacionados a potenciales evocados por eventos, análisis del espectro y estudios de conectividad entre regiones cerebrales como los ya mencionados. Brevemente, el primero estudia los cambios en el potencial eléctrico asociado a un evento, tales como apretar un botón o recibir algún estimulo. Mientras que el segundo estudia las popularmente llamadas *brain waves*, los rangos de frecuencias en los que opera el cerebro, cada rango de frecuencia tiene distintas características y estados de conciencia asociados. Por último, como ya se mencionó, los estudios de conectividad analizan cómo distintas partes del cerebro interactúan, cuales regiones del cerebro parecen comportarse sincronizadamente y cuales no. Esto ayuda a entender y analizar la dinámica del cerebro y el comportamiento y eventos asociados a los mismos.

A pesar de su limitada resolución espacial, los EEGs continúan siendo una herramienta muy valiosa para la investigación y diagnóstico, ya que es uno de las pocos dispositivos que ofrecen resolución temporal en el rango de los milisegundos, lo cual no es posible con técnicas como MRI (*Magnetic Resonance Imaging*), PET (*Positron Emission Tomography*) o CT (*Computed Tomography*).

[1] https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875

1.2.1. Ondas en el cerebro

Mucha de la información en el cerebro se encuentra codificada en distintas frecuencias, tanto dentro de un área como globalmente en el cerebro [BD04; Buz06]. Así, las oscilaciones suelen dividirse por bandas de frecuencias en *Delta* (δ), *Theta* (θ), *Alfa* (α), *Beta* (β) y *Gamma* (γ). Cada una de estas bandas de frecuencia se manifiestan con mayor intensidad en situaciones particulares. A continuación se describen las características de cada una.

Delta Las ondas *Delta* (δ) son las que se vinculan al sueño profundo, siendo las que tradicionalmente se han vinculado al descanso y reparación de los tejidos nerviosos. Están en el rango de frecuencias que va de 0.1 a 3.9 Hz.

Theta Las ondas *Theta* (θ) se observan al igual que las alfa cuando estamos relajados, pero en este caso son más frecuentes en momentos en que además de relajados estamos somnolientos, siendo el tipo de onda más predominante durante la fase dos del sueño no REM. Están en el rango de frecuencias que va de 4 a 7.5 Hz.

Alfa Las ondas *Alfa* (α) aparecen en momentos de relajación o ante tareas que no nos requieren concentración o esfuerzo. Están en el rango de frecuencias que va de 8 a 15 Hz.

Beta Las ondas *Beta* (β) suelen reflejar la realización de un esfuerzo mental intenso, apareciendo generalmente mientras estamos despiertos o bien durante el sueño REM. Están en el rango de frecuencias que va de 15 a 25 Hz.

Gamma Por arriba de los 25Hz se define la banda *Gamma* (γ). Esta banda está asociada con actividad cerebral producida en tareas de alta demanda cognitiva. Aunque al día de hoy no hay completo acuerdo sobre su rol. A lo largo de esta tesis estudiaremos la actividad cerebral en las tres bandas del medio (θ, α, y β).

Es importante destacar que la banda *Gamma* suele no utilizarse en EEG, ya que puede estar muy contaminada por artefactos musculares [YG+08]. Por otro lado, en el caso particular de los equipos de bajo costo, el ancho de banda con el que se registra y transmite la señal no es suficiente para trabajar con esta señal. Por otro lado, hay que considerar que los límites de estas bandas de frecuencia cambian con la edad. En particular, el pico alrededor de los 10Hz

(α) que suele observarse en adultos, en chicos se observa desplazado hacia las bajas frecuencias, llegando hasta los 6Hz para infantes[Bas+97].

1.2.2. Equipamiento en EEG

Los electroencefalógrafos o equipos de EEG pueden poseer distintas capacidades, virtudes y defectos. A continuación se describen algunas características o variantes más comunes.

Conectividad: Nuevos EEGs Inalámbricos Un inconveniente con los EEGs tradicionales es la utilización de cables que conectan a cada electrodo, esto hace que dichos dispositivos sean menos portátiles y más complicados de colocar. En los últimos años han surgido dispositivos inalámbricos, como el Emotiv Epoc que se utiliza en este trabajo, los cuales permiten mayor portabilidad y comodidad.

Costo: EEGs Económicos Los electroencefalógrafos suelen ser dispositivos costosos (por encima de los USD 10.000 y hasta USD 80.000), en los últimos años han surgido algunos de precio considerablemente inferior normalmente con finalidades recreativas (desde USD 500). A medida que pasan los años los precios continúan bajando y, aunque la calidad de los mismos va retrasada, va mejorando año a año. Esto hace que nos preguntemos hasta donde podemos llegar con ellos, parte de este trabajo será contestar esa pregunta.

En la figura 4 se puede ver un ejemplo de EEG portátil, inalámbrico y económico (Emotiv Epoc+) que será utilizado durante este trabajo.

Figura 4: Imagen de muestra del Emotiv Epoc+, un EEG portatil, inalámbrico y económico. Fuente: https://www.emotiv.com/epoc/

Unos de los componentes mas importantes de los EEGs son sus electrodos, estos son fundamentales para mantener una buena conectividad con el cuero cabelludo y medir la actividad eléctrica cerebral con la mejor calidad posible. Hay muchos tipos de electrodos, cada tipo tiene sus ventajas y desventajas, algunos permiten mejor manejo del ruido, otros mayor agilidad para la colocación del casco. A continuación se describen los distintos tipos y sus características.

Electrodos activos y pasivos Los electrodos más comunes son los electrodos pasivos que se componen de una terminal de metal (típicamente plata u oro) conectados a través de un cable al amplificador. Una mejora que se ha realizado en equipos más modernos, o de mayor costo, son los electrodos activos, o $actiCAP$. Estos poseen un pre-amplificador en el sitio de registro y luego la señal viaja amplificada hasta el amplificador. De esta manera se disminuye mucho el impacto del ruido ambiente, interferencias y artefactos, que pudiesen entrar por ejemplo por el cable.

Contacto: Electrodos Secos, Esponja, con Gel Sea pasivo o activo, entre el electrodo y el cuero cabello debe introducirse un elemento conductor como pasta (en el caso de los electrodos pasivos de copa) o gel (por ejemplo para el sistema Biosemi). Pero existen electrodos que no requieren de un gel o

líquido conductor para medir la actividad cerebral. Si bien para poder tomar mediciones de buena calidad sobre un gran rango de frecuencias los electrodos con gel son considerados el estándar, preparar los electrodos y el casco suele ser una tarea que toma tiempo. A su vez también la cabeza debe limpiarse del gel luego de la toma de datos. Los electrodos secos suelen utilizar electrodos activos y cascos flexibles con electrodos con forma especial para garantizar un contacto óptimo con el cuero cabelludo. Esto permite ahorrar un tiempo significativo al no necesitar tiempo para agregar el gel ni para sacarlo al final de la toma de datos. Un inconveniente de estos EEGs es que hay un único tamaño de casco lo cual podría no adaptarse correctamente a todas las distintas dimensiones y formas de cráneo. Un ejemplo de estos electrodos puede encontrarse también en un sistema de bajo costo como el OpenEEG [2].

Por último, existen electrodos de esponja o *R-Net*, los cuales consisten en electrodos pasivos puestos en una red flexible y duradera. Este método tampoco requiere gel lo cual lo hace ideal para casos donde el tiempo esta restringido.

[2] http://openeeg.sourceforge.net/doc/

2. Objetivo

El estudio de la actividad cerebral ha mostrado ser una herramienta fundamental en la comprensión de los procesos cognitivos humanos, y un motor para el desarrollo de métodos de registro y análisis de datos en diversas áreas.

Recientemente, surgieron equipos de EEG de un costo considerablemente menor y mayor portabilidad que los normalmente utilizados en clínica e investigación. Estos equipos suelen ser construidos por distintas empresas con un objetivo más recreativo, siendo usualmente utilizados como controladores de videojuegos y demás sistemas de interacción cerebro computadora (*brain computer interface*, BCI). Sin embargo, el uso de estos equipos podría a largo plazo abrir los horizontes de aplicación de distintas protocolos tanto clínicos como de investigación. Sin ir más lejos, el LIAA participa de un proyecto de investigación en el que se estudia el desarrollo cognitivo en contextos de bajo nivel socioecónomico, para lo cual la posibilidad de realizar registros en las escuelas es un condición casi excluyente.

En esta tesis nos planteamos dos objetivos generales:

- Buscar un balance entre la calidad de la señal obtenida y la portabilidad de los equipos, poniendo a prueba estos nuevos EEGs. Para ello, se realizará una validación comparando los mismos contra equipos de alta resolución en adultos en el laboratorio, y luego se analizará un caso de estudio en el campo.

- Determinar cuáles son las métricas más adecuadas (robustas) para estudiar redes funcionales cerebrales en equipos de bajo costo.

En particular, nos proponemos:

- Desarrollar el entorno de experimentación para la toma de datos. Colaborando en el diseño de los experimentos e implementando programas interactivos que enuncien las instrucciones de la tarea a los sujetos, presenten los estímulos, y adquieran respuestas. Todo esto al mismo tiempo que recolectan su actividad cerebral y dejan marcas temporales, indicadoras de distintos eventos relevantes, para su posterior sincronización y análisis.

- Estudio de los datos obtenidos a partir de dichos paradigmas, midiendo la conectividad cerebral con distintas métricas. Los datos fueron adquiridos por Marcos Luis Pietto en el marco de su tesis doctoral [Pie+18a; Pie+18b].

- Desarrollo de métodos de predicción de características individuales de la conectividad cerebral, utilizando las técnicas aprendidas en el punto anterior para obtener las redes mas realistas posibles.

3. Métodos

3.1. Desarrollo de los experimentos

El software propietario del Emotiv Epoc+ es bastante limitado y no permite por ejemplo interactuar con interfaces de experimentación desarrolladas en lenguajes de programación populares en ambientes científicos como, por ejemplo, en Python. Por lo que, antes de comenzar con la toma de datos se colaboró en el desarrollo de un entorno que interactúe con Python para realizar dichos experimentos. Este trabajo dió lugar a dos publicaciones con los datos que se presentan en esta tesis (en el marco de la tesis doctoral de Marcos Luis Pietto[Pie+18a; Pie+18b]). Este también está siendo utilizado en la tesis de licenciatura de Gustavo Juantorena en el laboratorio, y en otros laboratorios. El código fue liberado para que pueda ser utilizado por la comunidad [3].

3.1.1. Emotiv Bindings

Para desarrollar estas actividades se utilizaron librerías científicas clásicas como Psychopy [4], además de esto, para agregar marcas a las mediciones de Emotiv se utilizaron bindings de Python hechos originalmente por Federico Raimondo.

Se colaboró con la adaptación de los mismos a la plataforma experimental que se iba a utilizar en la escuela, también se extendió y mejoró la documentación y el código de las herramientas utilizadas.

3.1.2. Go-NoGo

En psicología las pruebas go/no-go son usadas para medir la capacidad de los participantes para mantener una atención selectiva y control de respuesta. Por ejemplo, una prueba go/no-go que requiera a los participantes realizar una acción dado un cierto estímulo (e.g. presionar el botón con sabor dulce - Go) e inhibir esa acción bajo un conjunto diferente de estímulos sabor amargo o calor (e.g., no presionar el mismo botón - No-Go).

Para desarrollar la interfaz gráfica y manejar de forma precisa los tiempos de exposición de cada imagen se utilizó la librería Psychopy. El programa agregaba múltiples marcas a los datos de la actividad cerebral generados por Emotiv, por ejemplo para indicar el momento en que se muestra un estímulo y cuando se

[3] https://github.com/mathigatti/Emotiv-Experiments
[4] https://www.psychopy.org/

saca, cuando comienza una sección de descanso y cuando empieza de nuevo la siguiente sección del juego, entre otras.

A diferencia de resting esta actividad no terminó siendo utilizada para estudios de conectividad ya que por su naturaleza resultó más interesante (Al menos en una primer étapa) utilizar estos datos para el análisis de ERP comparando los casos de Go contra lo de NoGo.

3.1.3. Resting

Esta actividad consistía simplemente en que los sujetos vean un video de 5 minutos de duración. El programa le agregaba a los datos de Emotiv marcas de inicio y final de la reproducción del corto. Si bien la tarea de mirar un video puede no ser considerada de resting, a lo largo de esta tesis la denominaremos resting ya que no existe una demanda explícita de atención o de toma de decisión en momentos determinados por el investigador.

3.2. Procedimiento

Adultos Para la primer parte de este trabajo se tomaron datos de 15 adultos, estos tuvieron que concurrir 2 días distintos para realizar la misma actividad pero con un EEG distinto en cada ocasión. Un día se utiliza el Emotiv Epoc+ y en la otra oportunidad se utiliza Biosemi. La actividad consistió en ver un video (5) de 7 minutos de duración con el objetivo de captar la conectividad subyacente al estado de *resting* o reposo. Luego para contrastar se graba al sujeto en reposo durante 2 minutos con los ojos cerrados. Los datos fueron registrados en colaboración con Marcos L. Pietto [Pie+18a].

Figura 5: Video infantil utilizado como entretenimiento durante experimento de resting.

Niños Se utilizaron conjuntos de datos obtenidos a partir del trabajo liderado por Juan Kamienkowski y el estudiante de doctorado Marcos Pietto los cuales realizan estudios neurocientíficos [Pie+18b] en centros educativos de bajos recursos.

Este dataset contiene la actividad cerebral medida durante un experimento de resting y diversos datos demográficos de niños de entre 5 y 6 años provenientes de escuelas de bajos recursos. Ya que no se cuenta con la información completa de todos los niños se utilizaron 90 sujetos para desarrollar un clasificador de sexo y solo 60 para clasificadores de otras variables de interés (Nivel de sueño y nivel educativo de la madre).

A diferencia del caso anterior solo se realizaron mediciones con Emotiv.

La tarea es la misma que con adultos, deben permanecer quietos mientras miran el video y la actividad cerebral es registrada.[Pie+18a]

3.3. Electroencefalógrafos utilizados

Las prestaciones de un electroencefalógrafo pueden variar inmensamente. Características como la cantidad de electrodos que poseen, su distribución y la frecuencia de muestreo suelen ser distintas de un dispositivo a otro e influyen en el precio y calidad del mismo.

En este trabajo utilizaremos principalmente el electroencefalografo portatil *Emotiv Epoc+*[5] de 14 electrodos. Este cuenta con una frecuencia de muestreo de 128 muestras por segundo, ancho de banda de 0.16 Hz a 43 Hz con filtro digital interno de 50Hz a 60Hz. Es completamente inalámbrico para permitir su portabilidad y rápida puesta en marcha. En menos de 3 minutos el aparato está listo para tomar registros de actividad cerebral. Y tiene un precio en el mercado de aproximadamente USD 500.

Para los experimentos de validación se utilizará BioSemi[6] el cual es un EEG profesional de 128 electrodos. En este caso todo el proceso previo a la toma de datos puede tardar 25 minutos, y al no ser inalámbrico los experimentos que se pueden realizar son más limitados. El precio de un equipo de estas características es de aproximadamente USD 100.000.

3.4. Preprocesamiento de las señales

Las señales obtenidas de los aparatos de EEG típicamente tienen un alto nivel de ruido y señales espurias suelen colarse en los registros. Es por este motivo que es siempre necesario contar con una etapa de preprocesamiento previa a analizar las señales.

En primer lugar, se filtraron las señales con un filtro pasa banda de 0.5 a 30 Hz.

Luego, se eliminaron manualmente segmentos ruidosos y se aplicó Análisis de Componentes Independientes (*Independent Component Analysis*, ICA) para limpiar las señales, eliminando fuentes de ruido como contracciones musculares y señales de dispositivos eléctricos, e intentando mantener únicamente con las señales propias de la actividad cerebral. ICA es un método computacional que sirve para separar una señal multivariante en subcomponentes aditivas suponiendo que la señal de origen tiene una independencia estadística y es no-Gausiana. Éste es un caso especial de separación ciega de las señales. ICA se encuentra muy relacionado al problema de la separación ciega de fuentes (inglés Blind Signal Separation, BSS) el cual consiste en determinar a través de un arreglo de transductores las señales de las fuentes originales que intervienen en una mezcla, sin información alguna de las señales originales ni de las ponderaciones de la mezcla. Éste es un problema clásico de procesamiento de señales. Si bien es una técnica clásica de preprocesamiento de EEG, no es común utilizarla en dispositivos de bajo costo como el Emotiv donde normalmente solo se realiza un filtrado de las bandas de frecuencia. En este caso, evaluaremos su aplica-

[5] https://www.emotiv.com/product/emotiv-epoc-14-channel-mobile-eeg/tab-description
[6] https://www.biosemi.com/

ción descartando las componentes que representaban contracciones musculares, señales de dispositivos eléctricos y demás.

Por último, se referenció todo al promedio de los electrodos.

En el caso de Biosemi, además se seleccionaron los 14 electrodos que se correspondían con los del Emotiv (Sistema 10-20) y se subsampleo a 128Hz dejando ambos conjuntos de datos equivalentes.

Datos descartados Algunos datos tuvieron que ser descartados por ser muy ruidosos o por contar con la grabación de un solo tipo de EEG faltando su contraparte. Finalmente quedaron 4 datasets, emotiv-ojos-cerrados, emotiv-ojos-abiertos, biosemi-ojos-cerrados, biosemi-ojos-abiertos, cada uno contiene entre 11 y 15 grabaciones utilizables.

Outliers Debido a que hay ciertas grabaciones mucho más largas que la media se decidió recortar los largos en los extremos haciendo que todos (excepto los que son muy cortos) queden con 50.000 samples en total. En el apéndice 1 se encuentra una tabla que muestra, a modo de ejemplo, la cantidad de samples por *DataSet* para ojos abiertos.

3.5. Conectividad Cerebral

Para estudiar las redes de conectividad entre electrodos necesitamos alguna métrica que nos defina esta idea de conexión. Con este objetivo utilizamos medidas que tienen normalmente orígenes en la física y análisis de señales, y actualmente se utilizan cada vez más para la comparación de señales de EEGs. Cabe destacar que este tipo de análisis de conectividad no han sido realizados en EEGs de bajo costo como son el EMOTIC Epoc+ hasta la fecha.

El siguiente es el procedimiento estándar para generar estas matrices de conectividad:

- Se procesa la señal aplicando un filtro pasa banda de Butterworth para quedarse solo con el rango de frecuencias sobre el que se quiere trabajar.

  Las bandas utilizadas son las siguientes.

  - θ (De 4 a 7.5 Hz)
  - α (De 8 a 14 Hz)
  - β (De 15 a 25 Hz)
  - θ+α+β (De 4 a 25 Hz)

- A partir de una función de conectividad o comparación entre señales se arma la matriz de conectividad cerebral, calculando esta medida de a pares de electrodos.
- Se normalizan los datos con z-score.
- Para comparar correctamente Emotiv con Biosemi se realiza un mapeo de electrodos.

A continuación describimos brevemente las métricas de conectividad utilizadas:

3.5.1. Phase Locking Value

Esta métrica toma valores en el intervalo [0,1], donde 0 representa el caso donde no hay sincronicidad de fase y 1 representa que la fase relativa entre dos señales es idéntica en todas las muestras. PLV puede por lo tanto verse como una medida de la variabilidad muestra a muestra en las fases relativas de dos señales.[APL13]

3.5.2. Phase Locking Index

Esta métrica toma valores en el intervalo [0,1] y es 0 si la distribución de fase relativa es simétrica sobre 0 o π. En la práctica PLI y PLV son típicamente estimados a partir de promediar sobre varias muestras o en función del tiempo.[APL13]

3.5.3. Coherence

La coherencia espectral es una métrica estadística que puede ser usada para examinar la relación entre dos señales o conjuntos de datos. Es usada normalmente para estimar el poder de transferencia entre la entrada y salida de un sistema lineal. Esta métrica toma valores en el intervalo [0,1].

3.5.4. Spearman Correlation

En estadística, el coeficiente de correlación de Spearman, ρ, es una medida de la correlación (la asociación o interdependencia) entre dos variables aleatorias continuas.

3.5.5. Pearson Correlation

En estadística, el coeficiente de correlación de Pearson es una medida de la relación lineal entre dos variables aleatorias cuantitativas. A diferencia de la covarianza, la correlación de Pearson es independiente de la escala de medida de las variables.

3.5.6. Otros

A lo largo de este trabajo se mencionan algunas métricas mas las cuales fueron utilizadas para ciertos experimentos pero no terminaron tomando tanta relevancia. Estas son principalmente pequeñas variantes de otras métricas, como *Weighted PLI Debiased* y *PLI Unbiased* que son variantes de PLI o *Imaginary Coherence* que es una variante de la coherencia espectral.

3.6. Estadística de redes

3.6.1. Test de Pareo

Para medir qué tan bueno es Emotiv planteamos un experimento el cual parte de una idea simple, dada una persona a la cual se le mide la actividad cerebral con 2 EEGs distintos estos deberían dar resultados similares o al menos deberían ser más parecidos que si los comparamos con las mediciones que se le hicieron a otra persona.

Más específicamente dado un conjunto de personas P, las cuales fueron medidas con Emotiv e y Biosemi b y una función de comparación entre matrices de conectividad f, se espera que para cualquier $p_i, p_j \in P$ se cumpla que $f(e_{pi}, b_{pj}) \leq f(e_{pi}, b_{pi})$

Definiremos una métrica la cual consiste en calcular para la matriz de conectividad de Emotiv de un sujeto dado la distancia a todas las matrices de Biosemi de los sujetos que realizaron el experimento. Luego ordenaremos estos valores y verificaremos en que posición (o ranking) quedó la matriz de Biosemi del sujeto de interés.

La idea será realizar este procedimiento para todos los sujetos y luego calcular la media. La media de los rankings de todos los sujetos, $\overline{R} = \frac{1}{14}\sum_{i=1}^{14} R_k$, será el estadístico para decidir si el *matching* es bueno o el resultado obtenido solamente es producto del azar. Donde $R_k \in 0, 1, \ldots, 13$ y representa el ranking del sujeto k, o sea, en que posición apareció al realizar el test de pareo buscando con Emotiv su contraparte en Biosemi. Para determinar el p-valor del test, realizamos el procedimiento de *bootstrap* ya que $\overline{R}$ no tiene distribución

Normal, la distribución es asimétrica ya que hay varias formas de realizar un pareo totalmente incorrecto entre los sujetos pero solo una de realizar un pareo perfecto. Se puede ver esta asimetría en la figura (6).

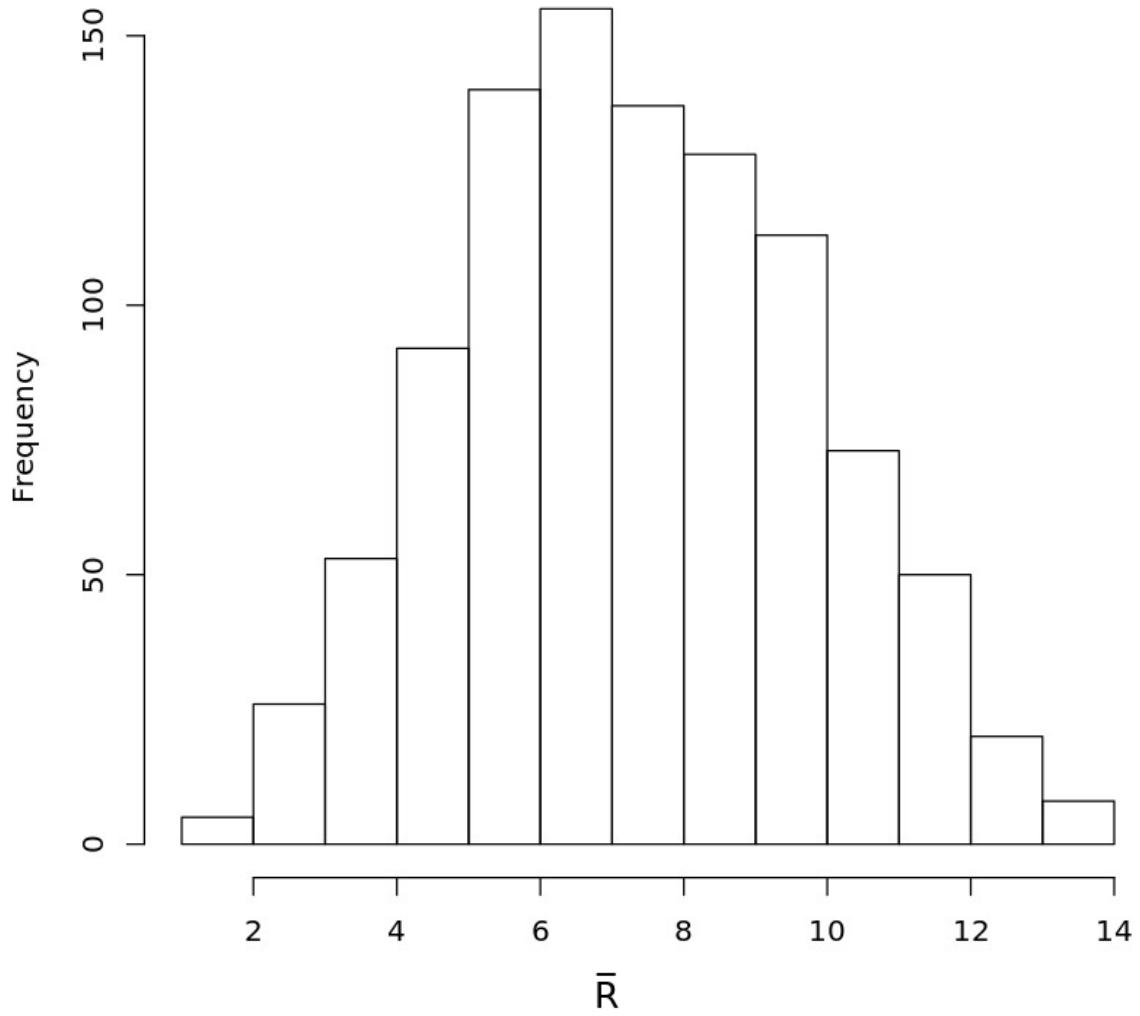

Figura 6: Histograma de 1000 instancias de puntaje promedio del test de *matching* sobre 14 sujetos con matrices de conectividad generadas a partir de aplicar PLI sobre series temporales generadas azarosamente con una distribución normal.

Si existe una relación entre las matrices de los diferentes grupos, entonces se cumplirá que $\overline{R}$ será chico (entre 0 y un valor crítico). El p-valor del test y los valores crítico salen del procedimiento estándar de *bootstrap*. A modo de ejemplo el valor crítico a nivel de significación del 5 % es 3,71.

3.6.2. Análisis de los grafos subyacentes

Una forma de comparar la calidad de las mediciones de Emotiv es ver si las matrices de conectividad tienen una estructura similar, a partir de los grafos subyacentes a las mismas intentaremos usar distintas métricas de teoría de grafos para comparar las mismas con las generadas por Biosemi.

3.6.3. Relación entre modularidad y densidad de las aristas

Intuitivamente, un módulo es un subconjunto de nodos en una red los cuales están más densamente conectados entre si que con el resto de la red. Una partición de la red es definida como una asignación de los nodos en conjuntos disjuntos de nodos. Hay muchas definiciones matemáticas para medir la calidad de dicha partición en términos de separación modular, i.e. cuán bien una partición diferencia subconjuntos de nodos altamente conectados entre sí. Una medida muy popular es introducida por Newman (Newman and Girvan, 2004). El valor de modularidad asignado a una partición usando dicha medida esta dado por la siguiente expresión.

$Q = \frac{1}{2L} \sum_{i,j} \delta(i,j) \left[A_{ij} - \frac{k_i k_j}{2L} \right]$, donde A_{ij} vale 1 si los nodos (electrodos) i y j tienen un enlace, y 0 sino. El valor k_l corresponde a la cantidad de enlaces que tiene el nodo l (grado del nodo).

A partir de esta idea se puede analizar el índice de modularidad dependiendo la cantidad de aristas (y lo mismo se puede hacer para el número de módulos), a partir de este análisis se generan curvas como las de la figura 7. Para variar la densidad de aristas se toma la matriz de conectividad y se discretizan las aristas eliminándolas o manteniéndolas dependiendo si están por arriba o por debajo de cierto valor.

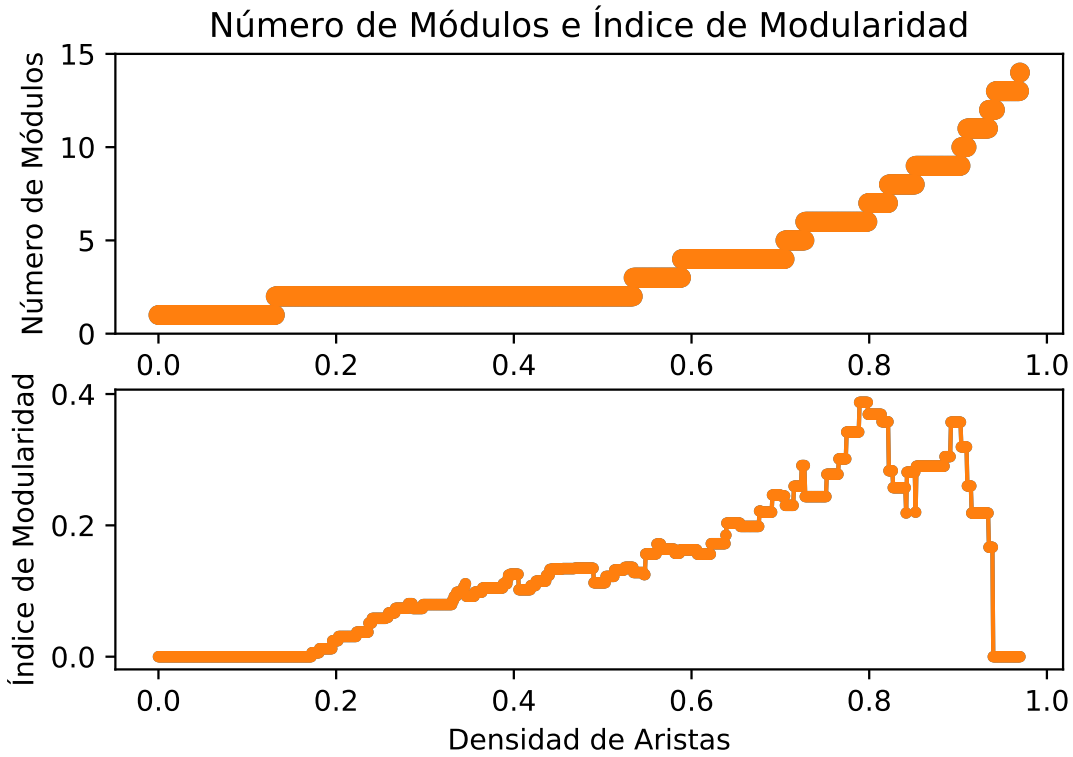

Figura 7: Número de módulos e Índice de modularidad (Q) en función de la densidad de aristas para el caso particular de la matriz de conectividad calculada con correlación de Spearman en la banda *Beta* utilizando Biosemi.

Como era de esperarse el número de módulos es una función incremental en función de la densidad de aristas. Es importante observar que esto no necesariamente ocurre con el índice de modularidad donde la función sí puede decrecer.

3.6.4. Test de clasificación

Diversos trabajos de clasificación se han hecho a partir de redes (Tagliazucchi, E. et. al. (2014) [TL14], Amoruso, L. (2017) [Amo+17]) incluso en niños (Uddin, L. Q. et. al (2013) [Udd+13]). Basándonos en estos trabajos pero aplicándolo por primera vez en un EEG de pocos electrodos y bajo costo intentaremos ver si podemos desarrollar clasificadores exitosos siguiendo las prácticas utilizadas en trabajos previos por otros investigadores.

De esta forma, realizaremos tareas de clasificación utilizando arboles de decisión sobre los conjuntos de datos de Emotiv los cuales tienen dos versiones, uno con ojos abiertos viendo un video y otro con ojos cerrados en estado de reposo. Se intentará ver con qué métricas diferenciamos mejor esos estados entre sí. Se medirá cuan bien da cada métrica con el área bajo la curva (*area under*

the curve, AUC) de la curva *receiving operator curve* (ROC) sobre sets de testeo armados a partir de *cross-validation* con 6 folds.

Se utilizarán las distintas redes generadas y se intentará inferir distintas variables, como pueden ser el sexo o nivel educativo en un set de datos de niños de entre 5 y 6 años a los cuales se les grabó la actividad cerebral con el Emotiv Epoc+ mientras veían un video de 7 minutos.

3.6.5. Random Forest

Para la clasificación se utilizó *Random Forest*, también conocido en castellano como Bosques Aleatorios, es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. Es una modificación sustancial de *bagging* que construye una larga colección de árboles no correlacionados y luego los promedia.

La idea esencial del *bagging* es promediar muchos modelos ruidosos pero aproximadamente imparciales, y por tanto reducir la variación. Los árboles son los candidatos ideales para el *bagging*, dado que ellos pueden registrar estructuras de interacción compleja en los datos, y si crecen suficientemente profundo, tienen relativamente baja parcialidad. Producto de que los árboles son notoriamente ruidosos, ellos se benefician enormemente al promediar.

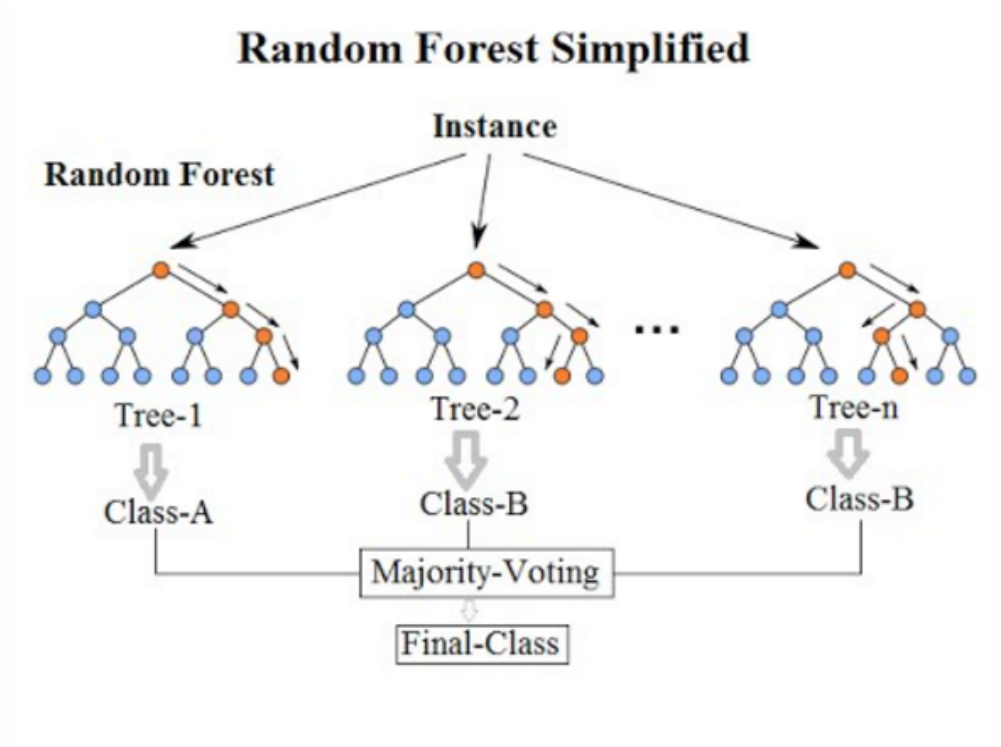

Figura 8: Representación visual del algoritmo *Random Forest*.

Una ventaja de este algoritmo es que es fácil averiguar cuales son las características (o *features*) que mejor resultan para realizar el trabajo de clasificación, esto permite luego inferir las diferencias claves entre las distintas clases que se intentan predecir.

Esto será útil para nuestro trabajo ayudándonos por ejemplo a entender mejor las diferencias principales entre las redes de conectividad de un individuo con ojos cerrados y abiertos.

3.7. Métodos de validación de la clasificación

3.7.1. Cross Validation

Para entrenar los modelos y medir la efectividad de los clasificadores se utilizó *cross-validation*, esta es una técnica clásica utilizada en el entrenamiento de modelos de clasificación.

La validación cruzada o *cross-validation* evalúa los resultados de un análisis estadístico y garantiza que son independientes de la partición entre datos de entrenamiento y prueba. Consiste en calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones del conjunto de datos. Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar la precisión de un modelo. Es una técnica muy utilizada en proyectos de inteligencia artificial para validar modelos generados.

3.7.2. Curva ROC y el AUC

Para visualizar y medir el desempeño de los clasificadores utilizamos la curva ROC y su AUC (*Area Under the Curve*).

Las curvas ROC (*Receiver Operating Characteristic*) presentan la sensibilidad[7] de una prueba diagnóstica que produce resultados continuos, en función de los falsos positivos (complementario de la especificidad), para distintos puntos de corte.

Por ejemplo, para la clasificación entre los grupos de enfermos y sanos representa la capacidad discriminatoria de una prueba para clasificar sanos como sanos y enfermos como enfermos. Un parámetro para evaluar la bondad de una prueba diagnóstica que produce resultados continuos es el área bajo la curva (AUC). Este área puede interpretarse como la probabilidad de que ante un par de individuos, uno enfermo y el otro sano, la prueba los clasifique correctamente.

[7] https://en.wikipedia.org/wiki/Sensitivity_and_specificity

3.7.3. Tests Estadísticos

Wilcoxon En experimentos que analizan la modularidad en distintas redes de conectividad (9) utilizamos el test no paramétrico de Wilcoxon. Este permite comparar series de valores y evaluar si provienen de distribuciones similares.

4. Búsqueda de Métricas Robustas entre equipos en Adultos

4.1. Desarrollo de un entorno para interacción con Emotiv en Python

Se cumplió con éxito el primer resultado, dando lugar a la publicación de dos artículos hasta la fecha [Pie+18b; Pie+18a], dentro del marco de la tesis doctoral de Marcos Pietto, y dos tesis de licenciatura, la presente y la de Gustavo Juantorena con paradigmas de lenguaje. Estos experimentos fueron en parte recolectados masivamente en las escuelas con un tamaño muestral que no se hubiese alcanzado sin este desarrollo (4 sesiones de 1 hora de 150 chicos y chicas de 5 años).

4.2. Sensibilidad de las redes funcionales

A continuación mostramos que las redes funcionales son capaces de distinguir distintos estados como ser ojos cerrados y abiertos en ambos equipos de medición. En esta sección se presentan los resultados para el análisis del número de módulos e índice de modularidad. La descripción detallada de estos conceptos puede encontrarse en la introducción 2.

En las celdas de la tabla se muestra el p-valor del test de Wilcoxonel cual compara los valores de la red promedio de ojos abiertos contra la red promedio de ojos cerrados para una banda y dispositivo dado. Lo que el p-valor indicará es si las redes de ojos abiertos tienen una modularidad similar a la de ojos cerrados. Será esperable de esto que si Emotiv tiene un p-valor menor a 0.05 entonces Biosemi también y viceversa.

	Theta				Alfa				Beta			
	N° de Modulos		Modularidad		N° de Modulos		Modularidad		N° de Modulos		Modularidad	
Métrica	Emotiv	Biosemi	Emotiv	Biosemi	Emotiv	Biosemi	Emotiv	Biosemi	Emotiv	Biosemi	Emotiv	Biosemi
Coherencia	0	0	0,142	0	0	0,0128	0	0,0638	0	0	0,2543	0
Correlación	0	0	0,0444	0	0	0	0,0026	0	0	0	0,0002	0
Coherencia Imaginaria	0	0,0103	0	0	0,0015	0	0	0,0192	0,9213	0	0	0
PLI	0	0,0076	0	0,1027	0,5368	0	0,123	0,65	0,0742	0	0	0
PLI2 Unbiased	0	0	0,3295	0	0	0	0,2018	0,0011	0	0	0	0
PLV	0	0,0008	0	0	0	0	0	0	0	0	0,075	0
PPC	0	0	0	0	0	0	0	0	0	0,0001	0,2066	0
Spearman	0,0003	0	0,0015	0	0	0	0	0	0,0233	0,0011	0,0346	0
WPLI2 Debiased	0	0	0	0	0	0	0	0	0	0	0	0

Figura 9: Tabla compuesta por los p-valores del test de wilcoxon entre ojos abiertos y cerrados para una banda y EEG dado. Se indica en color rojo cuando, para un mismo test, el p-valor de Emotiv tiene una discrepancia con el de Biosemi, o sea cuando uno da significativo y el otro no.

Se puede ver en la tabla 9 que hay algunos valores marcados en rojo, indicando que hay una discrepancia en la modularidad de las redes de Emotiv y Biosemi (Uno dio por encima de 0.05 y el otro no). Las métricas de *Correlación* y *WPLI2 Debiased* son las únicas que dieron siempre consistentemente. Por otro lado *PLI* y *Coherencia* son las que presentan más inconsistencias. *Spearman*, *PPC* y *PLV* tienen una sola inconsistencia, todas en el índice de modularidad de la banda beta.

4.3. Sensibilidad de los equipos de medición

A fin de analizar la robustez de los equipos de medición, y de nuestros propios métodos se realizó como primera medida una prueba de *matching* simple. El registro de cada sujeto se separó en dos (de 0-1 minuto registro 1 y de 1-2 registro 2) y se los consideró en forma independiente. Se construyeron 14 redes funcionales para el primer tramo temporal y otras 14 para el segundo (Una por cada sujeto). Luego, utilizando el procedimiento de *matching* descrito anteriormente se buscó identificar qué red de un grupo correspondía con la del otro grupo. Esto se realizó cuando tenían los ojos abiertos (Tabla 10) y también cuando los tenían cerrados (Tabla 11).

	Bandas											
	Theta				Alfa				Beta			
	Emotiv		Biosemi		Emotiv		Biosemi		Emotiv		Biosemi	
Metrica	**Media**	**p-value**	**Media**	**p-value**	**Media**	**p-value**	**Media**	**p-value**	**Media**	**p-value**	**Media**	**p-value**
Coherencia	0,2	0,000	0,1	0,000	0,0	0,000	0,1	0,000	0,0	0,000	0,0	0,000
Correlación	0,0	0,000	0,9	0,000	0,0	0,000	0,9	0,000	0,0	0,000	0,7	0,000
Coherencia Imaginaria	3,7	0,000	3,5	0,000	3,4	0,000	3,9	0,000	4,4	0,002	3,0	0,000
PLI	6,1	0,162	5,6	0,072	5,6	0,064	4,6	0,005	5,0	0,016	7,1	0,542
PLI2 Unbiased	6,6	0,340	6,3	0,220	5,8	0,093	4,4	0,002	4,4	0,002	7,1	0,572
PLV	0,7	0,000	0,6	0,000	0,6	0,000	0,3	0,000	0,1	0,000	0,0	0,000
PPC	0,8	0,000	0,2	0,000	0,9	0,000	0,2	0,000	0,6	0,000	0,1	0,000
Spearman	0,0	0,000	0,0	0,000	0,0	0,000	0,0	0,000	0,0	0,000	0,0	0,000
WPLI2 Debiased	7,2	0,597	5,9	0,117	5,7	0,082	4,6	0,004	5,4	0,043	6,6	0,340

Figura 10: Tabla compuesta por los p-valores del test de *matching* con ojos abiertos para una banda y EEG dado. Se indican en amarillo los p valores entre 0,1 y 0,05 y en rojo los mayores a 0,1.

	Bandas											
	Theta				Alfa				Beta			
	Emotiv		Biosemi		Emotiv		Biosemi		Emotiv		Biosemi	
Metrica	**Media**	**p-value**	**Media**	**p-value**	**Media**	**p-value**	**Media**	**p-value**	**Media**	**p-value**	**Media**	**p-value**
Coherencia	0,1	0,000	0,2	0,000	0,1	0,000	0,7	0,000	0,7	0,000	0,3	0,000
Correlación	0,5	0,000	0,0	0,000	0,0	0,000	0,0	0,000	0,1	0,000	0,0	0,000
Coherencia Imaginaria	5,7	0,201	4,1	0,005	3,5	0,000	3,1	0,000	4,3	0,007	4,3	0,006
PLI	5,4	0,107	3,5	0,001	4,7	0,023	4,2	0,005	6,7	0,566	5,4	0,069
PLI2 Unbiased	5,8	0,225	3,9	0,003	4,5	0,013	4,3	0,006	6,1	0,327	4,6	0,020
PLV	0,5	0,000	0,0	0,000	0,1	0,000	1,4	0,000	1,1	0,000	0,8	0,000
PPC	0,3	0,000	0,0	0,000	0,3	0,000	1,2	0,000	1,5	0,000	0,8	0,000
Spearman	0,0	0,000	0,0	0,000	0,0	0,000	0,0	0,000	0,0	0,000	0,0	0,000
WPLI2 Debiased	5,3	0,094	3,5	0,001	4,3	0,007	3,8	0,003	5,8	0,225	3,7	0,002

Figura 11: Tabla compuesta por los p-valores del test de *matching* con ojos cerrados para una banda y EEG dado. Se indican en amarillo los p valores entre 0,1 y 0,05 y en rojo los mayores a 0,1.

Los resultados de las tablas anteriores muestran que ambos equipos tienen consistencia interna (o variabilidad inter-sujetos mayor a la variabilidad intra-sujetos). Pero es interesante notar que esta consistencia no se observa para todas las métricas de conectividad. La falta de consistencia indicaría que la medida no es adecuada para caracterizar la dinámica cerebral. Las métricas resaltadas en rojo en las tablas son las que no presentan consistencia. Como se puede observar PLI y sus derivados obtienen resultados bastante pobres. Algo a tener en cuenta es que esta inconsistencia podría deberse a un error en la implementación de la métrica en la librería de MNE, la cual utilizamos.[8]. Esto último lo hemos descubierto luego de haber realizado los análisis.

4.3.1. Algunos Ejemplos

A continuación contrastamos los casos de métricas que dieron correctamente en comparación con casos que dieron mal.

[8]https://martinos.org/mne/stable/generated/mne.connectivity.spectral$_{c}onnectivity.html$

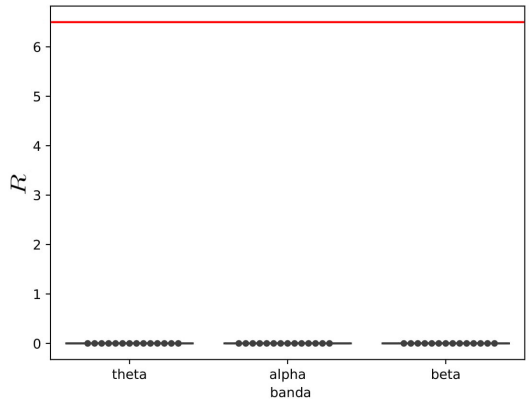

Figura 12: Valores de R al utilizar correlación de Pearson en condición de ojos Abiertos con test de *matching*, Emotiv vs. Emotiv.

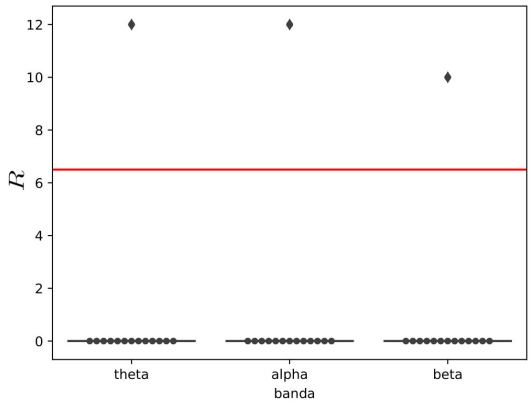

Figura 13: Valores de R al utilizar correlación de Pearson en condición de Ojos Abiertos con test de *matching*, Biosemi vs. Biosemi.

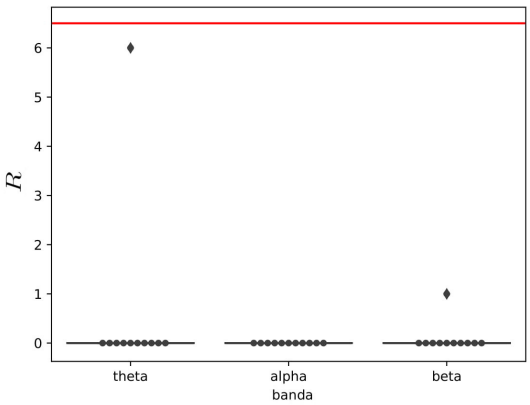

Figura 14: Valores de R al utilizar correlación de Pearson en condición de Ojos Cerrados, con test de *matching* Emotiv vs. Emotiv.

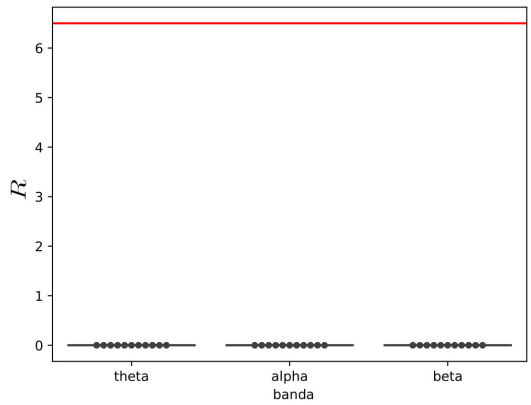

Figura 15: Valores de R al utilizar correlación de Pearson en condición de Ojos Cerrados con test de *matching*, Biosemi vs. Biosemi.

Como se puede observar para todos los casos la métrica de Pearson tiene resultados casi perfectos ($\overline{R} = 0$ cada uno se corresponde con su par adecuado), solo posee un *outlier* que se puede observar en los gráficos 14 y 13.

Por otro lado se puede ver como PLI tiene sus datos mucho más dispersos, con alta varianza y media en varios casos mas arriba que la línea roja que indica el punto de azar.

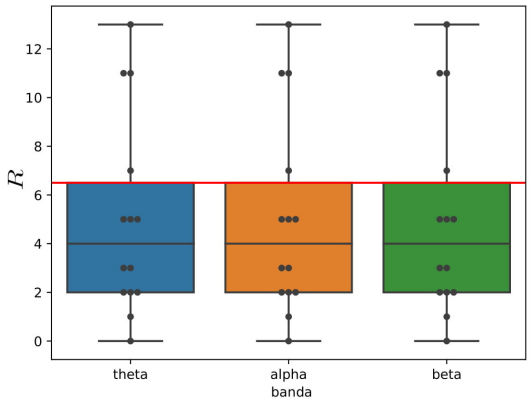

Figura 16: PLI Ojos abiertos con test de *matching*, Emotiv vs. Emotiv.

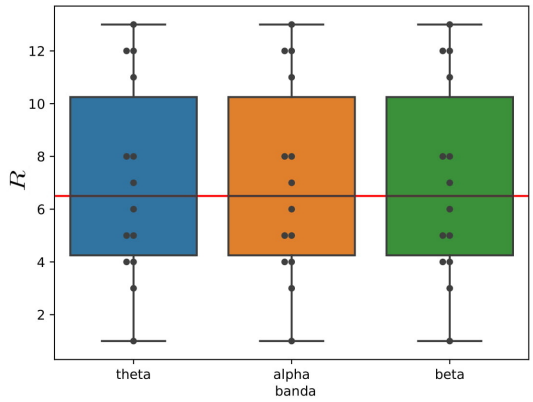

Figura 17: PLI Ojos abiertos con test de *matching*, Biosemi vs. Biosemi.

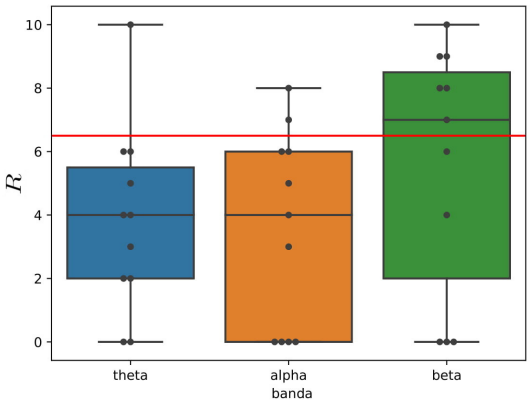

Figura 18: PLI Ojos Cerrados con test de *matching*, Emotiv vs. Emotiv.

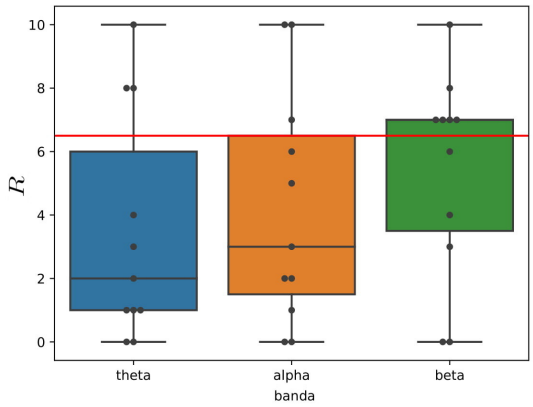

Figura 19: PLI Ojos Cerrados con test de *matching*, Biosemi vs. Biosemi.

Solo da ligeramente mejor que el azar para ojos cerrados con ondas θ (Ver figuras 18 y 19).

4.3.2. Ejemplos de Conectividad

Finalmente para poder visualizar y comparar la conectividad calculada a partir de las mediciones de EEG presentamos algunos gráficos de las redes pesadas generadas para distintos sujetos. El número representado en cada arista corresponde a la correlación entre las series temporales del par de electrodos del enlace.

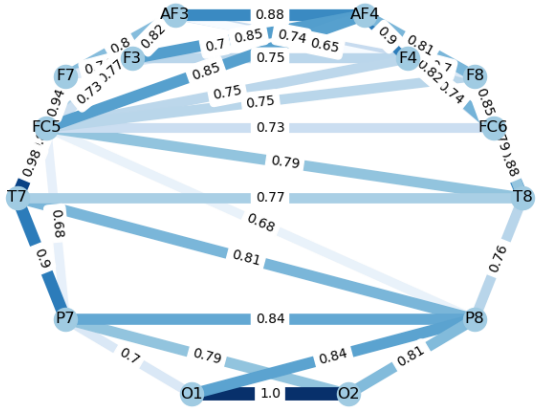

Figura 20: Gráfico de conectividad del sujeto A.K. con Emotiv utilizando los primeros minutos de su grabación.

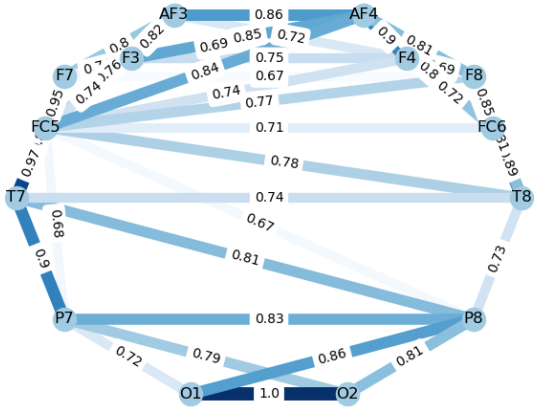

Figura 21: Gráfico de conectividad del sujeto A.K. con Emotiv utilizando los últimos minutos de su grabación.

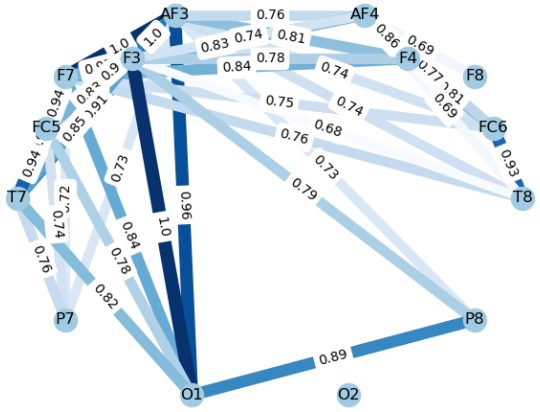

Figura 22: Gráfico de conectividad del sujeto E.G. con Emotiv utilizando los primeros minutos de su grabación.

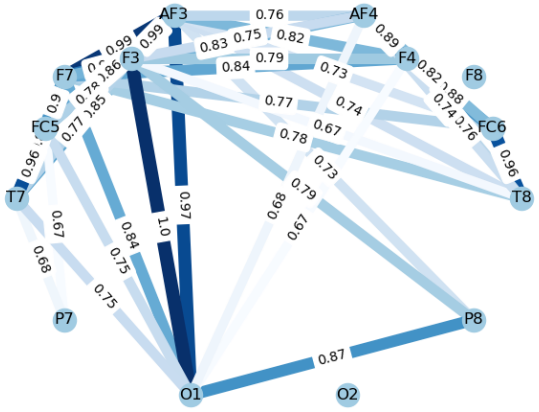

Figura 23: Gráfico de conectividad del sujeto E.G. con Emotiv utilizando los últimos minutos de su grabación.

Como se puede observar las redes de cada sujeto son diferentes entre sí y estas no suelen cambiar mucho por mas que se usen los primeros o últimos minutos de grabación. Por ejemplo la red del gráfico 20 es muy similar a la de 21 al mismo tiempo que parece tener una conectividad muy distinta a la de 22 y 23.

4.4. Comparación entre Equipos

A continuación se muestran los resultados del análisis de pareo entre Emotiv y Biosemi.

El experimento es complejo ya que no solo se trata de dos dispositivos distintos sino que también fueron análisis hechos en días distintos, variables como nivel de sueño pueden modificar en gran medida los resultados obtenidos.

Probablemente debido a esto los resultados en general no fueron buenos aunque algunas métricas dieron considerablemente mejor que otras.

Al igual que en el caso anterior calculamos para cada métrica el puntaje medio de pareo entre todos los sujetos y luego aplicamos *bootstrapping* para calcular el p-valor. Luego corroboramos si el *matching* promedio ($\overline{R}$) presenta evidencia a favor de que los equipos miden similarmente la actividad cerebral.

En términos simples, los resultados con fondo blanco o sin resaltar en las Tablas 24 y 25 indican que el *matching* es bueno y por lo tanto los equipos miden muy parecido. En cambio los resultados presentados en color rojo representan un *matching* malo. Ambas condiciones de ojos cerrados y abierto fueron analizados en forma separada.

	Bandas					
	Theta		Alfa		Beta	
Metrica	Media	p-value	Media	p-value	Media	p-value
Coherencia	7,1	0,512	8,8	0,951	7,3	0,576
Correlación	6,2	0,229	6,6	0,390	6,2	0,229
Coherencia Imaginaria	8,1	0,817	8,0	0,794	8,6	0,931
PLI	6,9	0,481	8,3	0,858	7,2	0,545
PLI2 Unbiased	8,1	0,817	7,7	0,746	5,9	0,185
PLV	8,2	0,838	8,7	0,942	7,3	0,576
PPC	8,1	0,817	9,0	0,959	7,2	0,545
Spearman	6,5	0,335	7,0	0,481	6,5	0,335
WPLI2 Debiased	8,4	0,874	7,7	0,746	7,8	0,772

Figura 24: Tabla compuesta por los p-valores del test de *matching* con ojos cerrados para una banda y EEG dado. Se indican en rojo los p-valores mayores a 0,1.

	Bandas					
	Theta		Alfa		Beta	
Metrica	Media	p-value	Media	p-value	Media	p-value
Coherencia	6,0	0,129	5,9	0,117	6,6	0,340
Correlación	6,4	0,245	6,8	0,395	6,6	0,314
Coherencia Imaginaria	5,3	0,038	6,4	0,245	7,3	0,627
PLI	6,8	0,395	5,5	0,056	6,6	0,340
PLI2 Unbiased	6,6	0,340	5,9	0,103	6,4	0,268
PLV	6,4	0,245	6,6	0,314	7,8	0,800
PPC	6,4	0,268	6,9	0,426	7,4	0,682
Spearman	5,6	0,072	6,6	0,314	6,5	0,292
WPLI2 Debiased	6,7	0,367	6,6	0,340	6,4	0,268

Figura 25: Tabla compuesta por los p-valores del test de *matching* con ojos abiertos para una banda y EEG dado. Se indican en amarillo los p valores entre 0,1 y 0,05 y en rojo los mayores a 0,1.

Los resultados tanto para ojos abiertos 25 como cerrados 24 son bastante pobres, los p-valores exceden en su mayoría el umbral de 0,1. Para el caso de pareo de ojos abiertos los resultados son levemente mejores, se pueden ver en la tabla 25 que el p-valor tiende a ser mas bajo incluso habiendo algunos valores por debajo de 0,1. Es importante destacar que la muestra es chica, hemos comparado solamente 14 sujetos, con lo cual es difícil tener una alta potencia en el test (y obtener p-valores chicos).

4.5. Clasificación de ojos abiertos y cerrados

A continuación se muestran los resultados de la tarea de clasificación sobre los conjuntos de datos de adultos de Emotiv de ojos abiertos y cerrados. Se intentó ver si podemos diferenciar esos estados entre sí, lo cual logramos con distintos niveles de efectividad dependiendo la banda y la métrica utilizada como se puede ver en el siguiente gráfico.

Utilizamos el área bajo la curva de la curva ROC como puntaje de la clasificación.

	Bandas				
Métrica	**Theta**	**Alfa**	**Beta**	**Todas**	**Promedio**
Coherencia	0,87	0,86	0,86	0,88	0,87
Correlación	0,8	0,9	0,91	0,88	0,87
Coherencia Imaginaria	0,43	0,6	0,69	0,54	0,57
PLI	0,42	0,63	0,66	0,69	0,60
PLI2 Unbiased	0,52	0,76	0,72	0,75	0,69
PLV	0,81	0,94	0,93	0,92	0,90
PPC	0,75	0,86	0,82	0,85	0,82
Spearman	0,85	0,84	0,86	0,89	0,86
WPLI2 Debiased	0,45	0,61	0,61	0,58	0,56

Figura 26: Tabla compuesta por el puntaje AUC de la curva ROC para cada métrica en cada banda.

En este experimento se ven resultados similares a los vistos en el experimento simple de *matching*. Métricas como PLI e Imcoh (Coherencia Imaginaria) (Resultados en rojo en la tabla 26) dan consistentemente mal, quedando entre los últimos puestos mientras métricas que otras como coherencia, Spearman, Correlation y PLV (resaltadas en verde en la tabla 26) tienden a dar resultados considerablemente mejores.

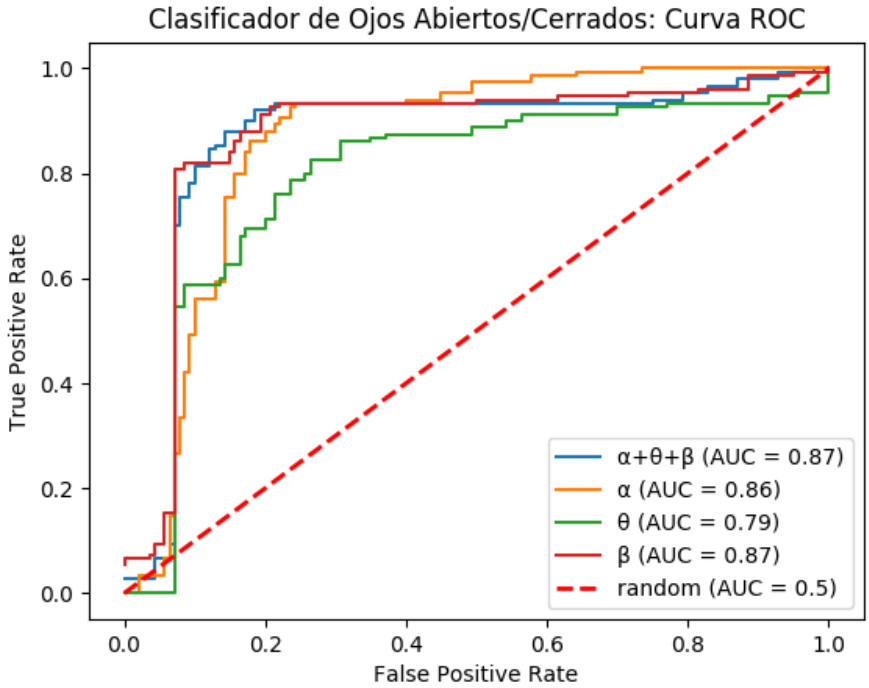

Figura 27: Detalle de la métrica *correlation*, se calcula el AUC por cada banda. Se resaltan con verde las 4 mejores métricas y en rojo las 4 peores.

4.6. Conclusión

Viendo los distintos resultados se nota que métricas como correlación de Pearson y Spearman suelen dar mejor que las demás. Esto se vio principalmente en el experimento de clasificación de ojos abiertos/cerrados, donde otras métricas como PPC y PLV también obtuvieron muy buenos resultados.

Métricas como PLI o coherencia imaginaria dieron en repetidas oportunidades resultados pobres, tanto para clasificación como en el análisis de sensibilidad de los equipos, donde se comparaban los equipos contra sí mismos. Como se mencionó previamente quizás sea un problema en la implementación de estas métricas.

Métricas como Pearson y Spearman dieron p-valores bajos en el análisis de sensibilidad de los equipos, incluso a pesar de que la muestra era pequeño y solo contenía unos 14 sujetos.

Por otro lado es interesante observar cuál es la mejor banda, esto depende de lo que se quiera hacer, por ejemplo en el experimento de *matching* se vio una mayor robustez con las ondas *Theta* para ojos abiertos pero para ojos cerrados los resultados más robustos aparecen en las ondas *Beta*. Por otro lado los resultados de modularidad mostraron que la banda *Beta* no es la más adecuada para distinguir los estados ojos abiertos de cerrados.

También pudimos ver que ninguna métrica dio mucho mejor que el azar en el experimento de *matching*, esto es razonable ya que el experimento no sólo tenía como dificultad reconocer sujetos medidos con distintos EEGs sino que también en distintos días y posiblemente en distintos estados, haciendo demasiado complicada la prueba con la cantidad de sujetos que contamos.

5. Predicción de Características Individuales en Niños

A continuación se describen los resultados obtenidos al intentar aplicar las herramientas desarrolladas en la sección anterior para inferir distintos atributos de los niños a partir de sus redes de conectividad.

En esta sección para la generación de las matrices de conectividad se utilizó como métrica la correlación de Pearson la cual dio consistentemente buenos resultados en la sección anterior. Esto no quita, de todas maneras, que resultados iguales o incluso mejores se puedan obtener con otras métricas que también dieron buenos resultados como PLV o Correlación de Spearman.

En todos los casos se utiliza la librería sklearn[9] de Python con la cual se entrena un modelo de *Random Forest* sobre los datos de entrenamiento aplicando validación cruzada con 6 *folds*. En exploraciones preliminares se utilizaron otras técnicas y combinaciones de ellas, pero se descartó este enfoque porque no mejoraba significativamente y no aportaba a la comprensión del problema.

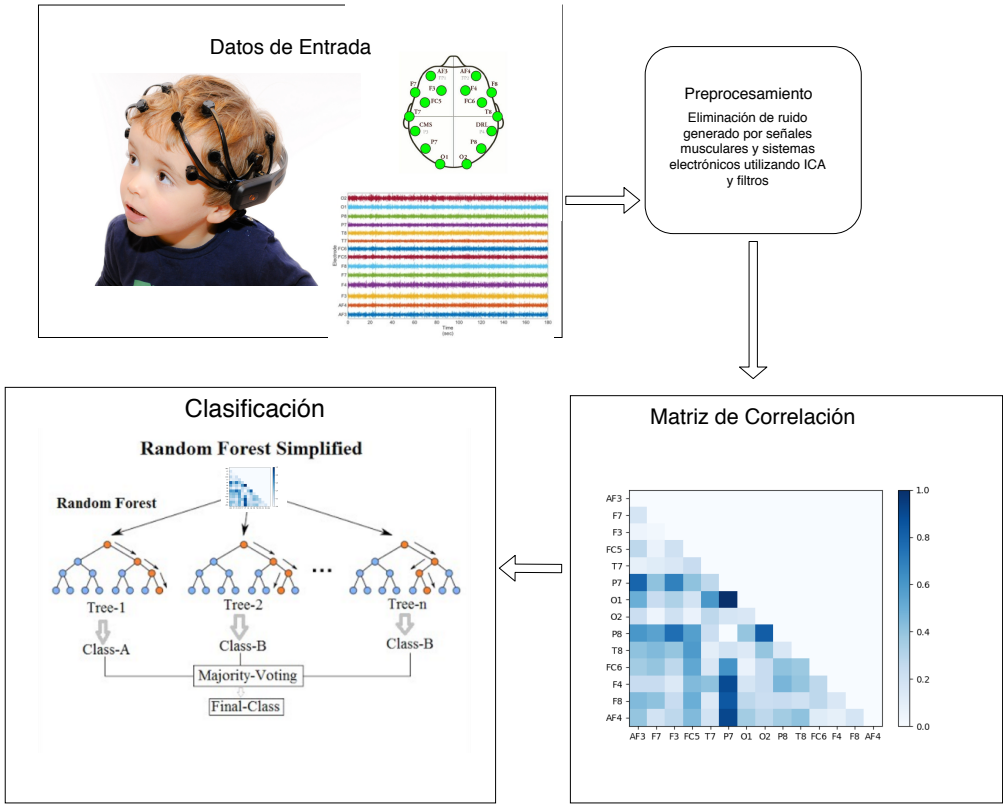

Figura 28: Proceso de clasificación.

En la imagen 28 se puede ver a modo ilustrativo como es el proceso desde

[9] https://scikit-learn.org/stable/

la toma de datos hasta la elaboración del modelo de clasificación.

5.1. Clasificación de Ojos Abiertos: Atributos Principales

A continuación se pueden observar los atributos principales de los clasificadores de ojos abiertos.

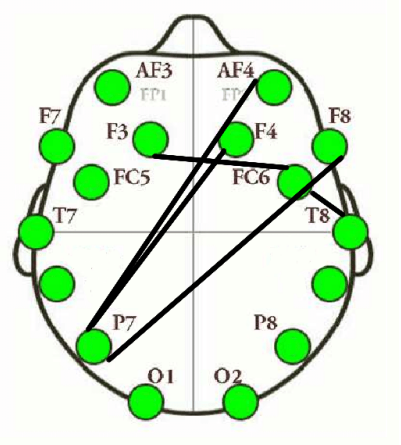

Figura 29: Atributos principales para el clasificador de ojos abiertos/cerrados con banda *Theta* (θ)

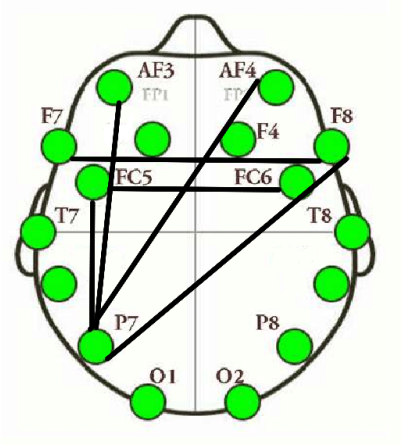

Figura 30: Atributos principales para el clasificador de ojos abiertos/cerrados con banda *Alpha* (α)

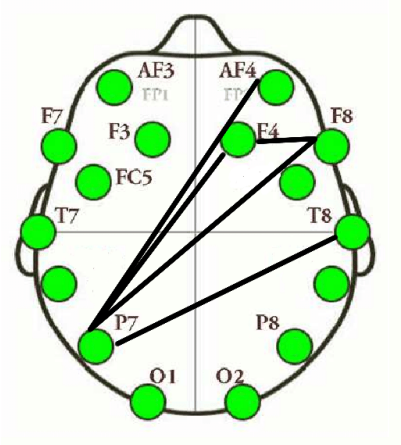

Figura 31: Atributos principales para el clasificador de ojos abiertos/cerrados con banda *Beta* (β)

Parece ser útil la dinámica que se da entre la zona izquierda del lóbulo parietal y el lóbulo frontal derecho. Esto se mantiene de forma bastante consistente entre todas las bandas y coincide con lo mencionado en el trabajo de Xu. et al (2014) [Xu+14] donde ven un incremento en la conectividad entre dichas regiones al tener los ojos abiertos.

5.2. Predicción de sexo

En años recientes algunos autores, como Tomesku et. al (2018) [Tom+18], han encontrado diferencias en el la actividad medida en el EEG en función del sexo de una persona. En estos estudios los autores suelen utilizar EEGs de varios electrodos, por ejemplo 204 en el caso de Tomesku. En esta sección nos proponemos replicar los resultados utilizando Emotiv de 14 electrodos. Elegimos estudiar el sexo porque era una variable que estaba bien medida para todos los sujetos, y la primera que recibimos.

Se desarrolló un clasificador para diferenciar sexo (Masculino/Femenino) a partir de la conectividad cerebral, lo cual ya se ha investigado con EEGs de

alta densidad [Tom+18]. En este caso intentamos recrear el experimento con nuestro conjunto de datos de niños y utilizando Emotiv.

5.2.1. Curva ROC

A pesar de las limitaciones a continuación se puede ver en la figura 32 como en el mejor de los casos conseguimos un AUC de la curva ROC de 0.82 para la banda *Beta* (β).

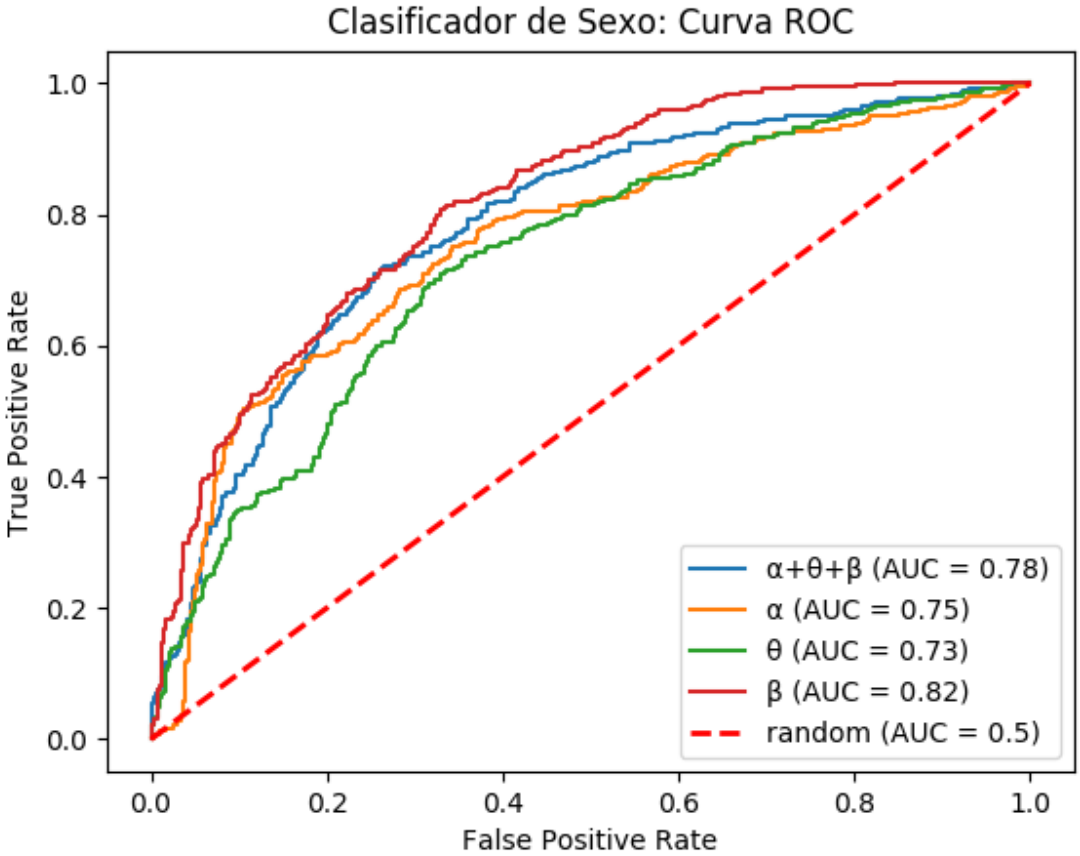

Figura 32: Curva ROC para clasificador de sexo.

En la figura 32 se observa como la banda *Beta* (β) es la mejor banda de frecuencias para distinguir los grupos, mientras que *Theta* (θ) y *Alpha* (α) dan un área bajo la curva (AUC) que es casi un diez por ciento menos.

5.2.2. Matrices de Confusión

A continuación se muestran las matrices de confusión tomando un umbral que pondere de forma lo más equitativa posible los falsos positivos y negativos.

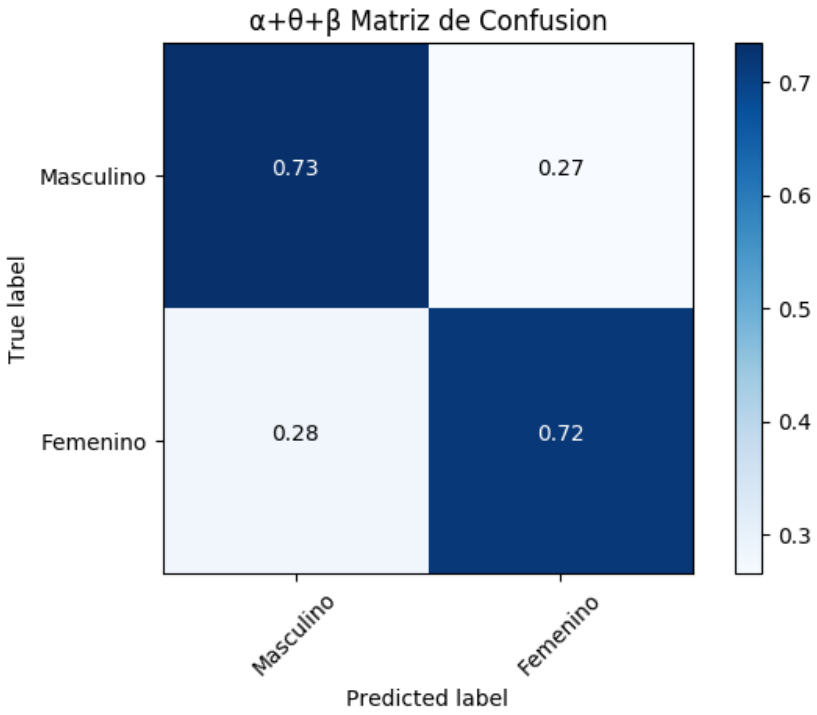

Figura 33: Matriz de confusión de la banda $\alpha + \beta + \theta$

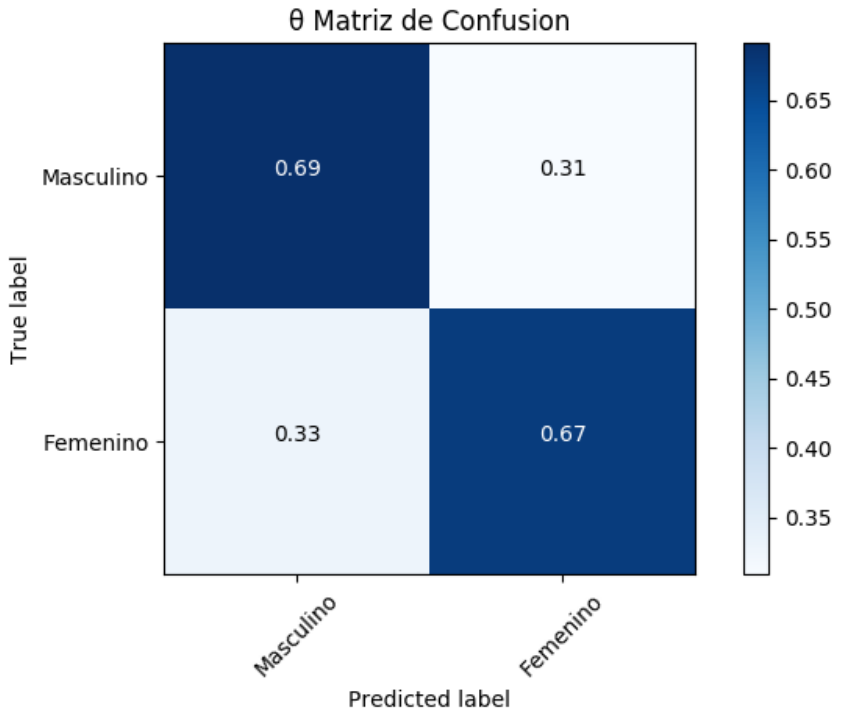

Figura 34: Matriz de confusión de la banda *Theta* (θ)

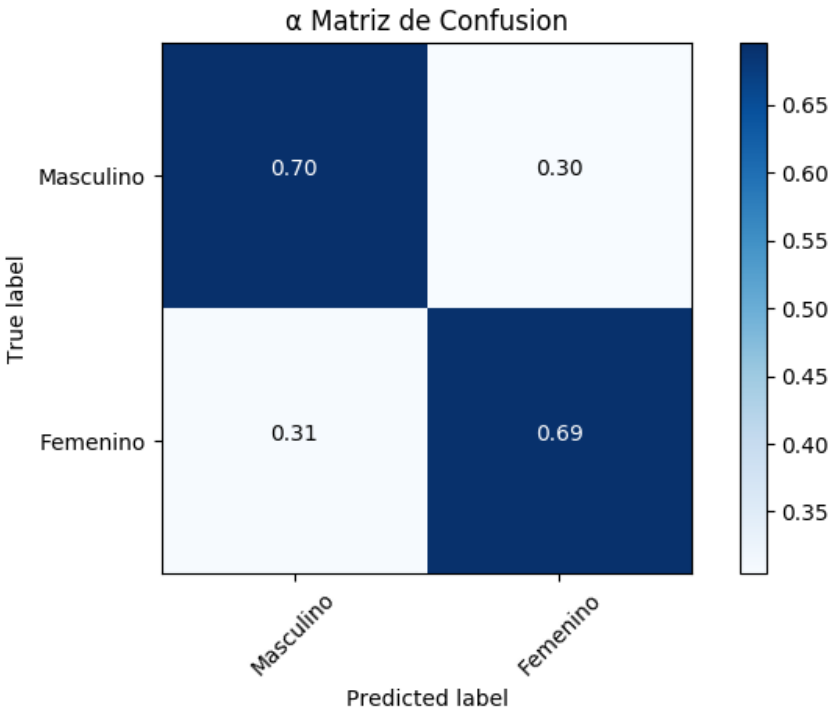

Figura 35: Matriz de confusión de la banda *Alpha* (α)

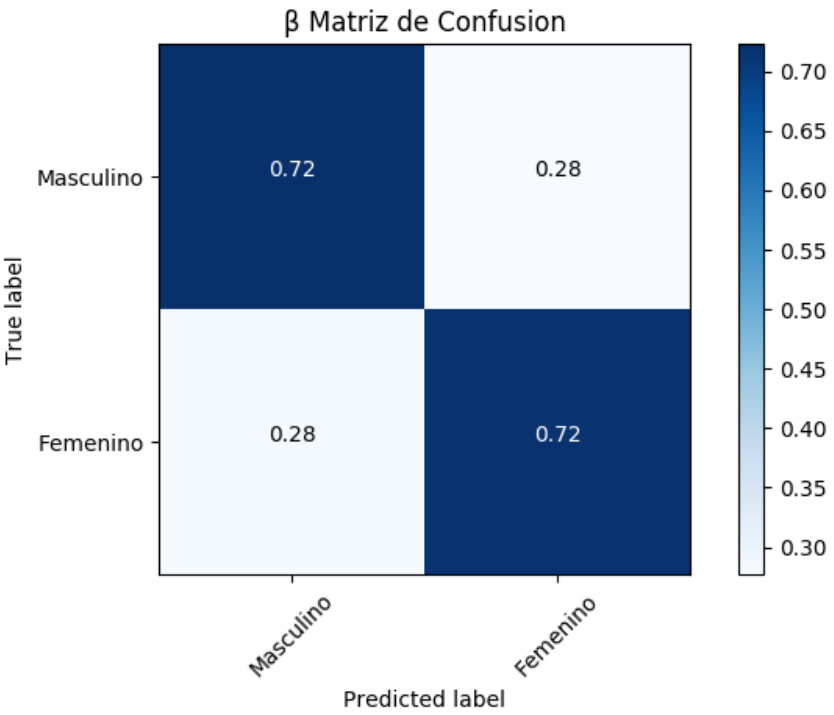

Figura 36: Matriz de confusión de la banda *Beta* (β)

Como se puede ver en las figuras 33 y 36 se puede obtener una precisión y

recuperación de aproximadamente 72 %.

5.2.3. Atributos Principales

A continuación se pueden observar los atributos principales de los clasificadores de sexo.

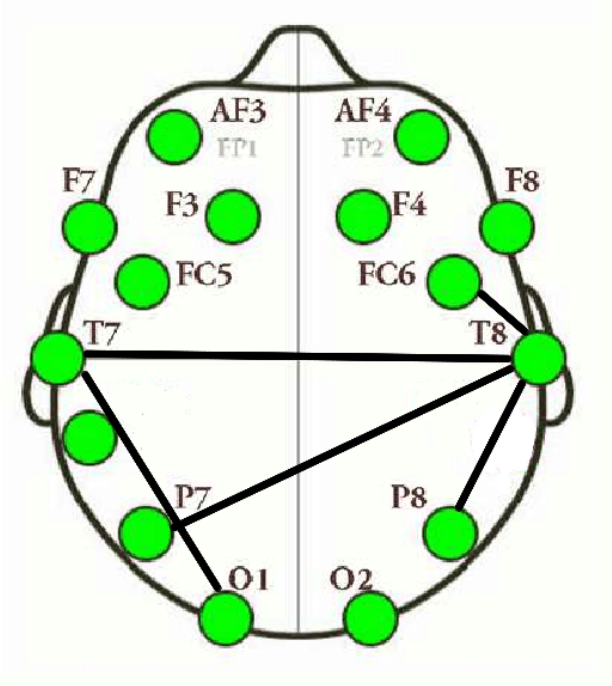

Figura 37: Atributos principales para el clasificador de sexo que utiliza las tres bandas juntas

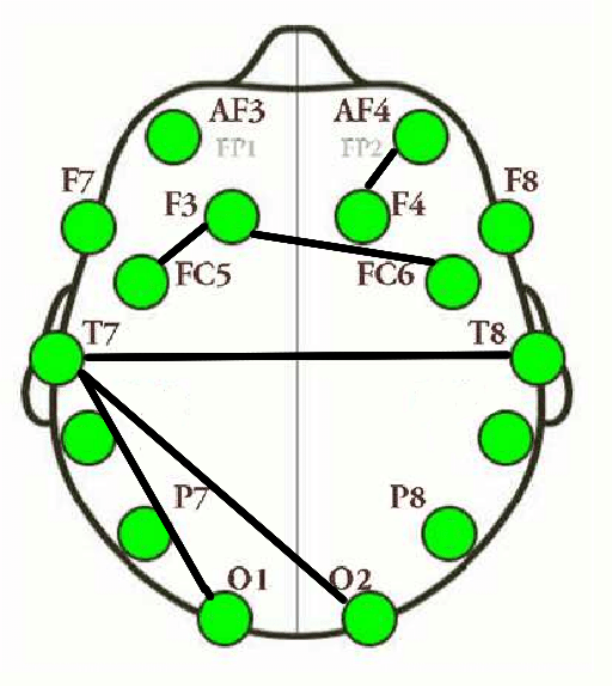

Figura 38: Atributos principales para el clasificador de sexo con banda *Theta* (θ)

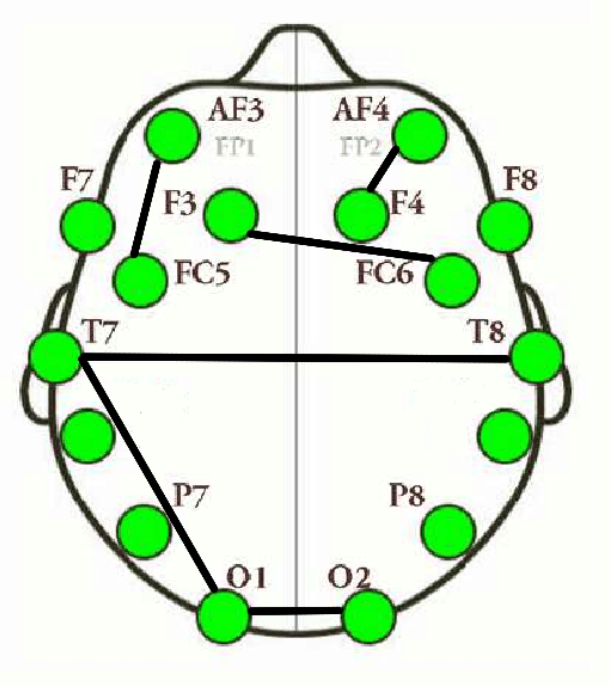

Figura 39: Atributos principales para el clasificador de sexo con banda *Alpha* (α)

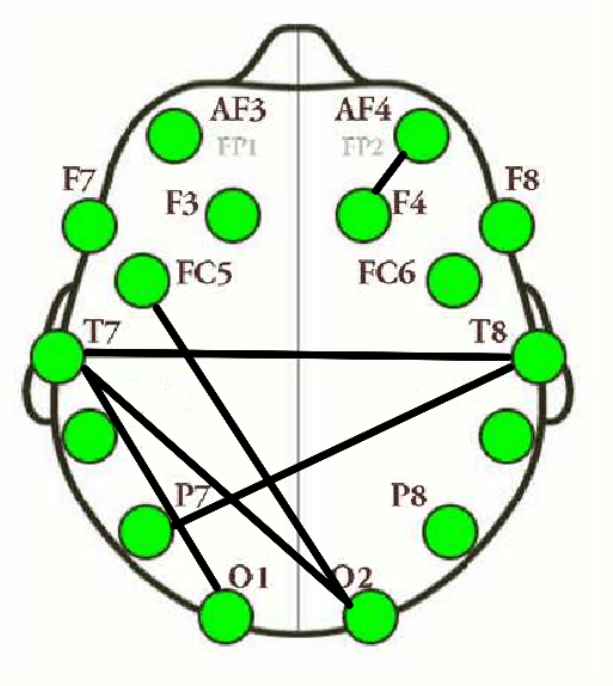

Figura 40: Atributos principales para el clasificador de sexo con banda *Beta* (β)

La conectividad entre lóbulos temporales parece ser crítica para la clasificación apareciendo en los clasificadores de todas las bandas. También la conexión entre lóbulo occipital-izquierdo y temporal-izquierdo parece ser importante. Esto podría estar relacionado con el contraste entre la conectividad entre hemisferios que tienen los hombres y las mujeres, como mencionan en su estudio Ingalhalikar et. al (2014) [Ing+14], las mujeres suelen tener mayor conectividad entre hemisferios que los hombres, y los hombres suelen tener mayor conectividad intra hemisferio.

Hay que mencionar de todas maneras que también se puede estar observando una diferencia en las redes debido a características secundarias del sexo como es el estilo del cabello. Las mediciones del EEG pueden ser afectadas por el cabello largo, lo cual suele ser más común en mujeres que en hombres. Así como también pueden afectar otras variables no controladas en el caso de nuestra muestra que no estaba diseñada para responder esta pregunta

5.3. Predicción de cantidad de sueño

Utilizando información sobre la cantidad de horas que duermen los niños por día se intentó ver si podíamos detectar a los niños que dormían menos, en este caso se intenta ver si podemos diferenciar a los niños que estén en el 25 percentil en cantidad de horas de sueño por noche. En muestras de gran

tamaño y utilizando datos de fMRI se ha observado que es posible distinguir la cantidad de sueño de los adultos [FF18; TL14]. Queremos ver si esto es posible para una muestra pequeña más pequeña y utilizando el equipo Emotiv.

5.3.1. Curva ROC

Se consiguieron resultados considerablemente mejores que el azar, principalmente con las ondas *Alpha* (α) y *Theta* (θ) como se puede ver en la figura 41.

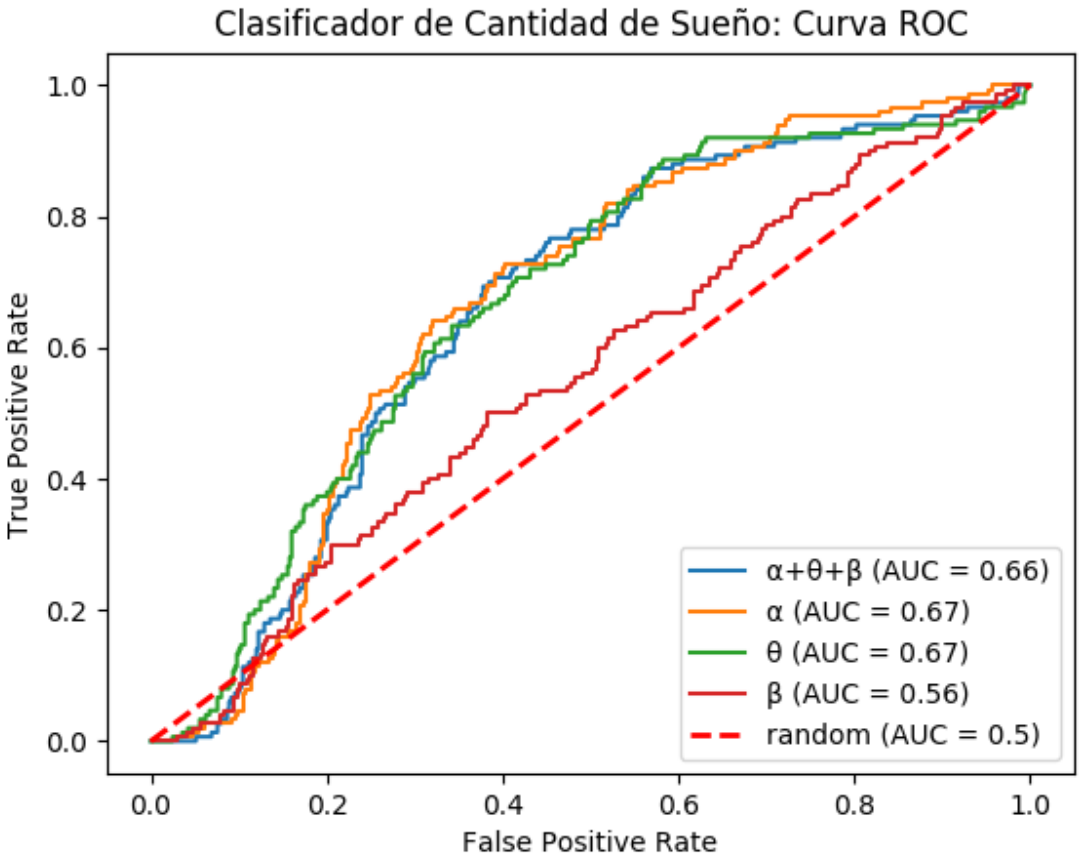

Figura 41: Curva ROC utilizando como métrica de conectividad la correlación de Pearson.

5.3.2. Matrices de Confusión

Las matrices de confusión decidiendo un umbral que pondere de forma lo mas equitativa posible los falsos positivos y negativos fue el siguiente.

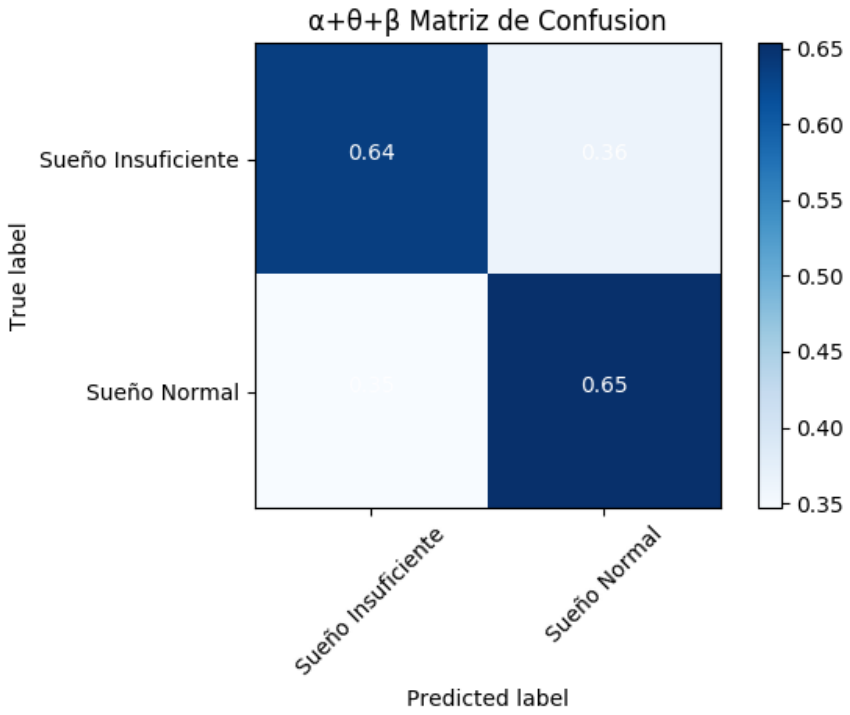

Figura 42: Matriz de confusión de la banda $\alpha + \beta + \theta$

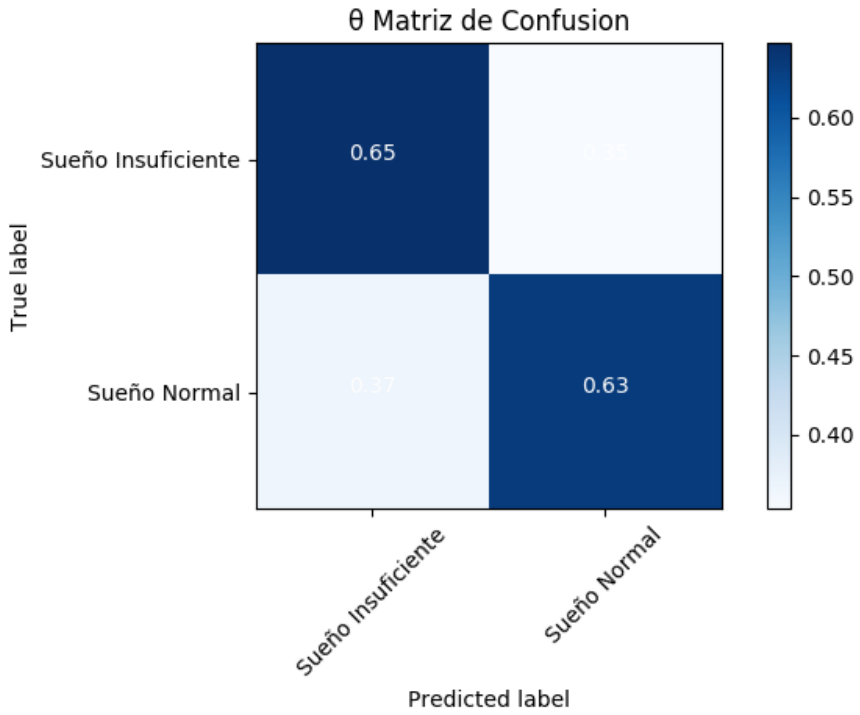

Figura 43: Matriz de confusión de la banda *Theta* (θ)

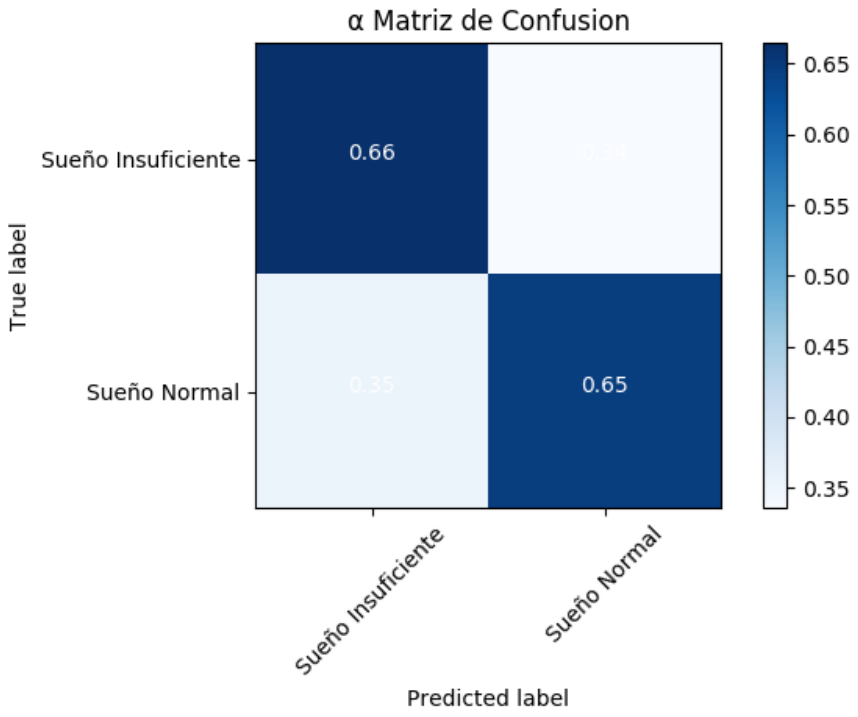

Figura 44: Matriz de confusión de la banda *Alpha* (α)

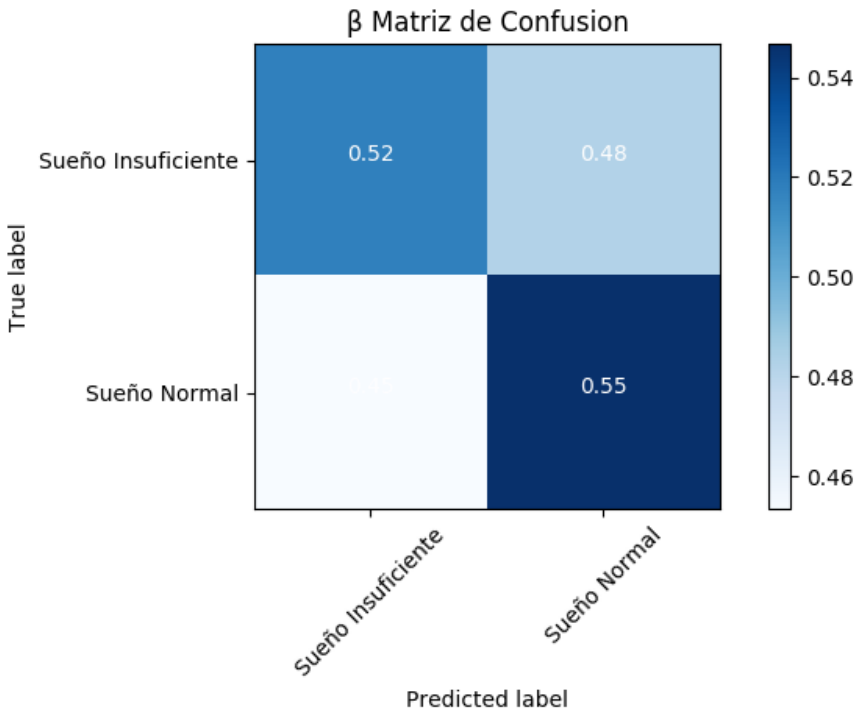

Figura 45: Matriz de confusión de la banda *Beta* (β)

Como se puede ver en las figuras 44 y 42 se puede obtener una precisión y

recuperación de aproximadamente 65 %.

5.3.3. Atributos Principales

A continuación se pueden observar los atributos principales de los clasificadores de sueño.

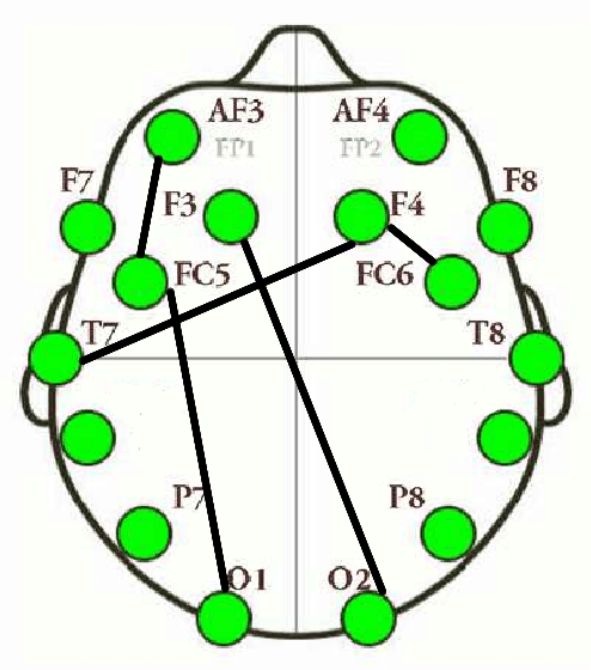

Figura 46: Atributos principales para el clasificador de sueño que utiliza las tres bandas juntas

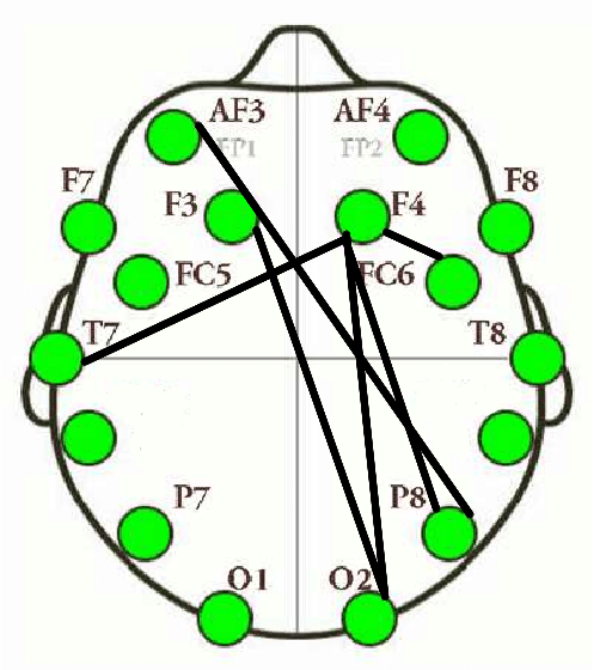

Figura 47: Atributos principales para el clasificador de sueño con banda *Theta* (θ)

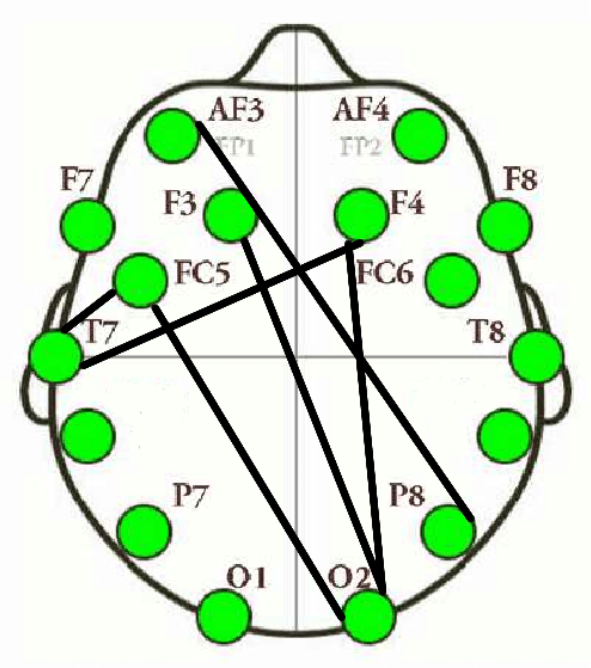

Figura 48: Atributos principales para el clasificador de sueño con banda *Alpha* (α)

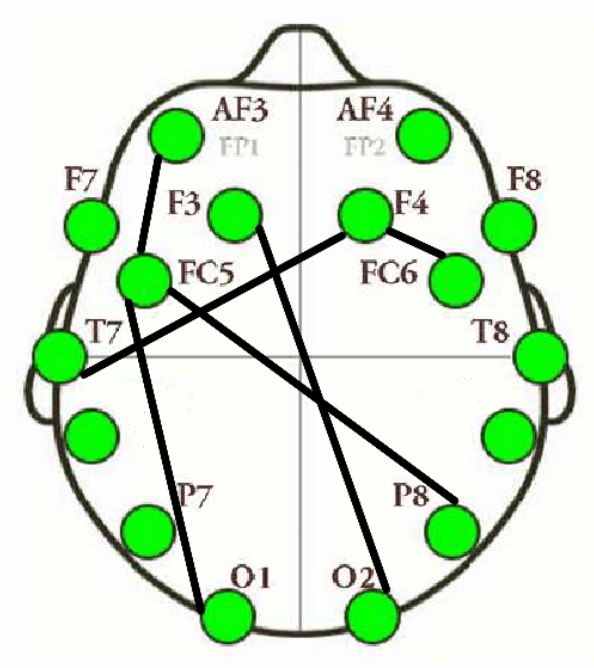

Figura 49: Atributos principales para el clasificador de sueño con banda *Beta* (β)

Como era de esperarse las principales componentes tienen siempre uno de sus extremos en alguna parte del lóbulo frontal, lo cual está asociada con la actividad normal durante el estado de vigilia y su cese esta relacionado al sueño. [TL14]

5.4. Predicción de nivel educativo del tutor

Contamos a su vez con información del nivel educativo de la madre o tutor encargado. Este dato se compone de 7 categorías: Primario incompleto, primario completo, secundario incompleto, secundario completo, universitario incompleto, terciario incompleto, terciario completo.

Es imposible pensar que a partir de los datos de EEG del chico se pueda predecir el nivel educativo de la madre o el tutor del niño. A fines de corroborar esto realizamos el mismo procedimiento de clasificación para este caso.

La siguiente figura muestra la curva ROC al intentar clasificar los datos divididos en solo dos categorías, *Con estudios terciarios* (Las primeras 4 opciones) y *Sin estudios terciarios* (Las últimas 3 opciones). Como era de esperar estas curvas muestran que no es posible clasificar a los tutores a partir de la actividad cerebral de los niños.

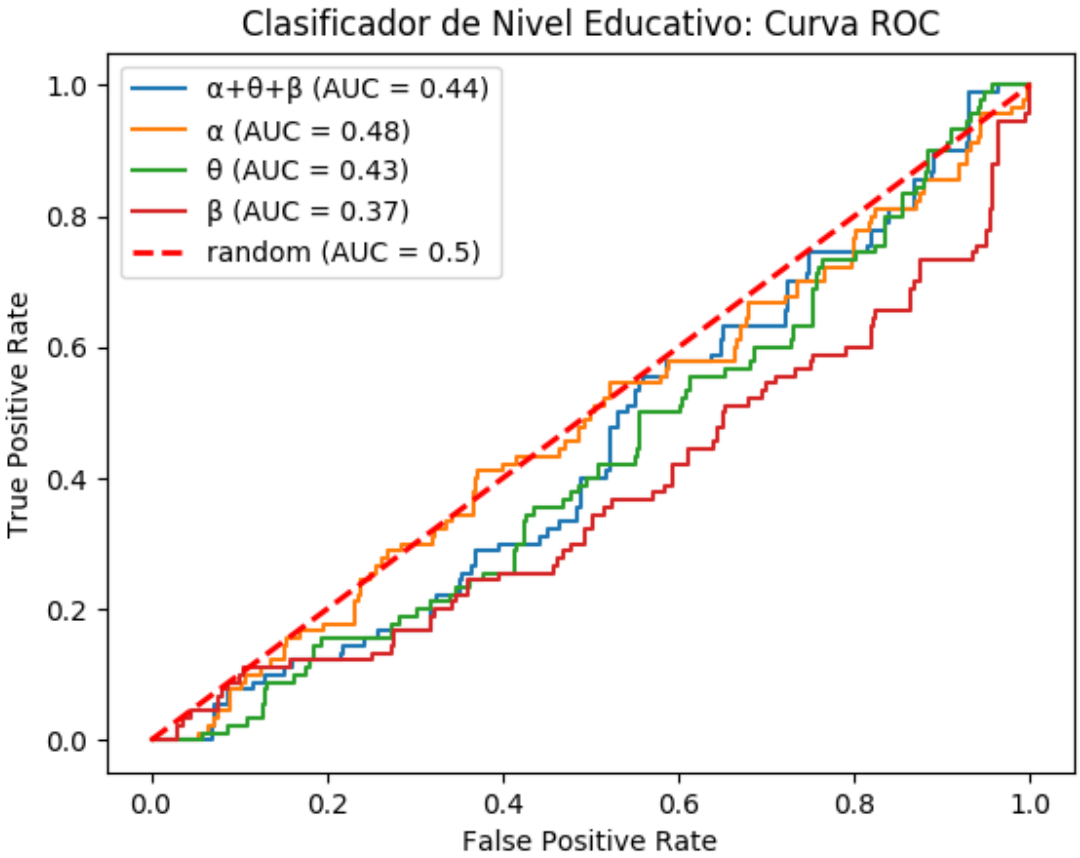

Figura 50: Curva ROC.

5.4.1. Conclusión

En esta etapa del trabajo se lograron desarrollar sistemas predictivos que tuvieron distintos niveles de éxito. En algunos casos como para la predicción de sexo y nivel de sueño se obtuvieron resultados considerablemente mejores que el azar, 0.82 y 0.67 de AUC respectivamente.

Al utilizar árboles de decisión fue fácil luego estudiar cuales eran los enlaces cerebrales que mayor peso tenían a la hora de intentar clasificar a un individuo, a partir de esto pudimos ver como los resultados obtenidos en varios casos coinciden con resultados de estudios anteriores. Por ejemplo para la clasificación de sexo se vio un gran peso en las conexiones entre hemisferios lo cual parece coincidir con resultados obtenidos al comparar la conectividad cerebral entre mujeres y hombres. Por otro lado para predecir nivel de sueño se vio una prevalencia por conexiones que conectan al lóbulo frontal el cual suele tener menor actividad a medida que entramos en niveles mas profundos de sueño y relajación.

6. Conclusiones

Esta tesis busca dar luz sobre las posibilidades de medición de actividad cerebral con aparatos de bajo costo. Como primer paso hemos continuado el trabajo del LIAA sobre cómo poder realizar las mediciones usuales de estos aparatos, ya que no están diseñados con fines de investigación ni con fines de aplicación en la clínica. Este trabajo incluyó paradigmas de experimentación, y la generación de marcas temporales que permitieron medir los primeros Potenciales Evocados con estos aparatos. Estos primeros resultados derivaron en la publicación [Pie+18a] y en otro trabajo del grupo [Pie+18b]. El código desarrollado se encuentra en un repositorio de código abierto[10] para que cualquiera pueda utilizarlo o incluso contribuir al mismo.

Una vez conscientes de que habíamos logrado buenos resultados a nivel potenciales evocados, decidimos profundizar en el tema buscando los límites de medición de este aparato de bajo costo. Nos planteamos entonces analizar las redes funcionales que se obtienen de este aparato como el de uno de mejor resolución (y mucho mayor costo). Mostramos que las métricas de correlación (Tanto de Spearman como de Pearson) y PLV son las más robustas tanto para el Emotiv EPOC+ como para Biosemi. Este análisis fue realizado en las distintas bandas de frecuencia, obteniendo los resultados más robustos en las bandas *Beta* y *Alpha*. Si bien ambas dieron buenos resultados en general, es probable que dependiendo el experimento sea mas razonable utilizar una que otra, dependiendo la naturaleza del fenómeno estudiado, por ejemplo, *Alpha* dió los mejores resultados para el experimento de cantidad de sueño mientras que *Beta* dió los mejores resultados en los experimentos de sexo.
Durante los análisis de validación en los que comparamos Emotiv con Biosemi, si bien existieron algunas pequeñas diferencias dependiendo del contexto del experimento, estas diferencias creemos que se deben a que el tamaño muestral es chico. Es importante destacar que para algunas comparaciones contábamos solamente con 14 sujetos y aún con ese tamaño muestral hemos obtenido, creemos, buenos resultados. Es de destacar que nuestros resultados muestran que ambos aparatos tienen una performance parecida.

Luego, también utilizando redes, analizamos algunos otros protocolos para ver la sensibilidad del aparato Emotiv. Se logró predecir con distintos niveles de éxito parámetros como el sexo, nivel de sueño y ojos abiertos/cerrados validando la efectividad y utilidad del dispositivo de bajo costo. Desafortunadamente, en este caso no contamos con los mismos registros pero con el aparato Biosemi. Por lo tanto, no podemos decir nada respecto de la *performance* en

[10] https://github.com/mathigatti/Emotiv-Experiments

este caso. Sin embargo si comparamos nuestros resultados de predicción con trabajos similares vemos similitudes en la conexiones cerebrales escogidas como indicadoras para inferir las categorías, para clasificación de sexo vimos como nuestro programa aprovechaba el aparente contraste entre la conectividad entre hemisferios del sexo masculino y femenino, lo cual coincide con resultados de otros autores [Tom+18; Ing+14], para predicción de nivel de sueño se vió un incapié en la conectividad con el lóbulo frontal, lo cual se suele asociar con la actividad normal durante el estado de vigilia mientras que su baja actividad esta relacionado al sueño [TL14]. Es importante resaltar, de todas maneras, que en algunos de estos casos se utilizaron distintas técnicas de clasificación o análisis y los datos fueron tomados en algunos cosas con fMRI en vez de EEG.

Si bien queda mucho por hacer, creemos haber contribuido en concientizar a la comunidad de que estos aparatos pueden ser de gran utilidad mediante un tratamiento adecuado de los datos. Creemos que nuestro trabajo puede abrir múltiples proyectos de investigación a futuro, ya que, como se dijo antes, la facilidad de este EEG para utilizarse y transportarse permite experimentos rápidos y a gran escala.

El código utilizado para la generación de las matrices de conectividad y para las tareas de clasificación publicó en Github [11] y posee una breve documentación para su fácil puesta en marcha.

6.1. Trabajo a Futuro

A continuación se enumeran los puntos principales sobre los cuales se puede continuar este trabajo.

- **Incorporar el conjunto de datos extendido creado por Marcos Pietto:** Sobre el conjunto de datos existente en chicos [Pie+18b] hoy ya se ha ampliado la muestra y se han medido una gran cantidad de variables, tanto de desempeño en distintas tareas como de contexto socio-económico. Estos chicos son también parte de una intervención en la que fueron entrenados con juegos de computadora en distintas capacidades de control cognitivo. Como siguiente paso se tomarán las medidas estudiadas en esta tesis para comenzar a analizar distintas variables relevantes al desarrollo de los chicos, en conjunto con Marcos L. Pietto. Estos resultados serán incluidos en una publicación en el futuro próximo.

- **Estudiar poblaciones en nuevos ecosistemas:** Como ya se mencionó, el alcance de este trabajo es extenso y si bien se ha aplicado inicialmente

[11] https://github.com/mathigatti/EmotivClassifier

con conjuntos de datos de chicos en escuelas esta no tiene porque ser la única aplicación. Quedará como trabajo a futuro tomar datos en nuevos ecosistemas como pueden ser universidades, parques, hogares o cualquier otro espacio de interés.

- **Generar grandes conjuntos de datos:** Será crucial aprovechar la portabilidad y facilidad de uso de Emotiv para poder generar fácilmente conjuntos de datos grandes de cientos de personas.

- **Realizar análisis de conectividad sobre nuevas variables y nuevos ecosistemas:** El estudio en nuevos entornos también abriría la posibilidad a nuevas variables de análisis, desde estudios que analicen la relación entre la conectividad cerebral y el desempeño académico en distintas áreas como matemática[DR15; Pre+10] o literatura[FG+14] hasta predicción de estados emocionales[Kho+10] o trastornos mentales como la depresión [Ana+05].

- **Tomar los datos realizando otras actividades:** Será interesante estudiar la conectividad en contextos y estados distintos al obtenido con *resting* al ver un video. Será útil para esto retomar el experimento de Go-NoGo (3.1.2) el cuál fue desarrollado como parte del trabajo de la tesis pero finalmente no fue utilizado para estudios de conectividad.

A. Apéndice

A.1. Cantidad de samples por grabación

Dispositivo	Nombre	Samples
Emotiv	ml2_rest_open	115584
Emotiv	if2_rest_open	112128
Emotiv	ak2_rest_open	107760
Biosemi	mv1_rest_open	69716
Biosemi	mf12_rest_open	66650
Biosemi	lt1_rest_open	66008
Biosemi	mdp1_rest_open	61559
Biosemi	ml1_rest_open	59315
Emotiv	mv2_rest_open	59164
Biosemi	gs1_rest_open	59136
Biosemi	mrv1_rest_open	58709
Emotiv	mdp2_rest_open	57972
Biosemi	ak1_rest_open	57488
Biosemi	eg1_rest_open	56093
Emotiv	mrv2_rest_open	51293
Emotiv	lt2_rest_open	50244
Emotiv	gs2_rest_open	49189
Emotiv	mf2_rest_open	49005
Emotiv	vg2_rest_open	48903
Biosemi	vg1_rest_open	30424
Biosemi	if1_rest_open	22088
Emotiv	eg2_rest_open	15232

A.2. Electrodos de Emotiv y Biosemi

A continuación se presenta la posición de los electrodos de nuestros dos sistemas utilizados.

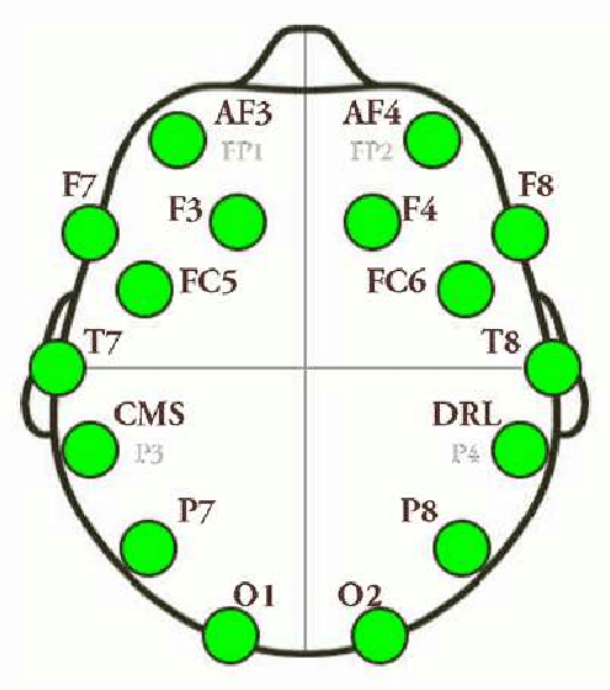

Figura 51: Gráfico orientativo de los 14 electrodos que posee el EEG Emotiv.

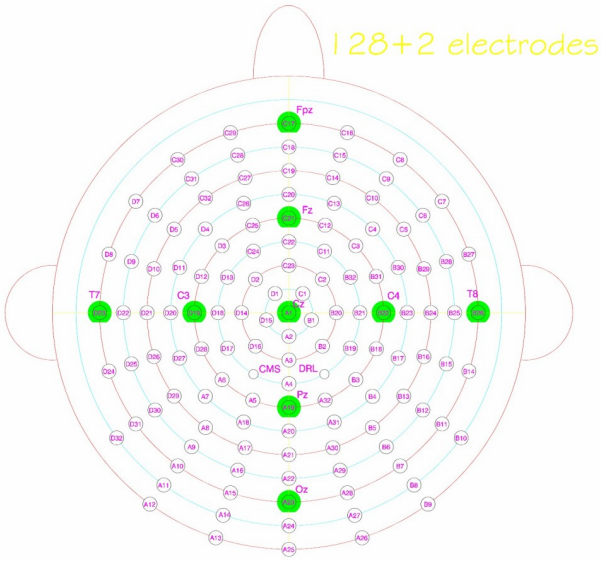

Figura 52: Gráfico orientativo de los 128 electrodos que posee el EEG Biosemi.

Para comparar los electrodos de los dos modelos de EEG utilizados realizamos una selección de 14 de los 128 electrodos de Biosemi buscando que estén en una ubicación lo más similar posible a su equivalente en Emotiv. Los electrodos seleccionados fueron C26, D6, C32, D11, D23, D31, A15, A28, B11, B26, B30, C10, C6, C13.

Referencias

[Amo+17] Lucía Amoruso y col. "Variability in functional brain networks predicts expertise during action observation". En: *NeuroImage* 146 (2017), págs. 690-700.

[Ana+05] Amit Anand y col. "Activity and connectivity of brain mood regulating circuit in depression: a functional magnetic resonance study". En: *Biological psychiatry* 57.10 (2005), págs. 1079-1088.

[APL13] Sergul Aydore, Dimitrios Pantazis y Richard M Leahy. "A note on the phase locking value and its properties". En: *Neuroimage* 74 (2013), págs. 231-244.

[Ast+15] Duncan E Astle y col. "Cognitive training enhances intrinsic brain connectivity in childhood". En: *Journal of Neuroscience* 35.16 (2015), págs. 6277-6283.

[Bar+16a] Albert-László Barabási y col. *Network science*. Cambridge university press, 2016.

[Bar+16b] Jessica J Barnes y col. "Electrophysiological measures of resting state functional connectivity and their relationship with working memory capacity in childhood". En: *Developmental science* 19.1 (2016), págs. 19-31.

[Bas+97] Erol Bas y col. "Is the alpha rhythm a control parameter for brain responses?" En: *Biological Cybernetics* 76.6 (1997), págs. 471-480.

[BD04] György Buzsáki y Andreas Draguhn. "Neuronal oscillations in cortical networks". En: *science* 304.5679 (2004), págs. 1926-1929.

[Buz06] Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford University Press, 2006.

[Cer+17] Andressa Cerqueira y col. "A test of hypotheses for random graph distributions built from EEG data". En: *IEEE Transactions on Network Science and Engineering* 4.2 (2017), págs. 75-82.

[CH+14] Robin Lester Carhart-Harris y col. "The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs". En: *Frontiers in human neuroscience* 8 (2014), pág. 20.

[Dos+10] Nico UF Dosenbach y col. "Prediction of individual brain maturity using fMRI". En: *Science* 329.5997 (2010), págs. 1358-1361.

[DR15] Scott L Decker y Alycia M Roberts. "Specific cognitive predictors of early math problem solving". En: *Psychology in the Schools* 52.5 (2015), págs. 477-488.

[FF18] Daniel Fraiman y Ricardo Fraiman. "An ANOVA approach for statistical comparisons of brain networks". En: *Scientific reports* 8.1 (2018), pág. 4746.

[FG+14] Elda Fischi-Gómez y col. "Structural brain connectivity in school-age preterm infants provides evidence for impaired networks relevant for higher order cognitive skills and social cognition". En: *Cerebral cortex* 25.9 (2014), págs. 2793-2805.

[Fra+14] Daniel Fraiman y col. "Biological motion coding in the brain: analysis of visually driven EEG functional networks". En: *PLoS One* 9.1 (2014), e84612.

[GC+15] Indira García-Cordero y col. "Stroke and neurodegeneration induce different connectivity aberrations in the insula". En: *Stroke* 46.9 (2015), págs. 2673-2677.

[Ing+14] Madhura Ingalhalikar y col. "Sex differences in the structural connectome of the human brain". En: *Proceedings of the National Academy of Sciences* 111.2 (2014), págs. 823-828.

[Kho+10] Reza Khosrowabadi y col. "The dynamic emotion recognition system based on functional connectivity of brain regions". En: *2010 IEEE Intelligent Vehicles Symposium*. IEEE. 2010, págs. 377-381.

[Kuo+16] Bo-Cheng Kuo y col. "Top–down activation of spatiotopic sensory codes in perceptual and working memory search". En: *Journal of cognitive neuroscience* 28.7 (2016), págs. 996-1009.

[Meu+09] David Meunier y col. "Age-related changes in modular organization of human brain functional networks". En: *Neuroimage* 44.3 (2009), págs. 715-723.

[Pie+18a] Marcos L Pietto y col. "Electrophysiological approaches in the study of cognitive development outside the lab". En: *PloS one* 13.11 (2018), e0206983.

[Pie+18b] Marcos Luis Pietto y col. "Enhancement of inhibitory control in a sample of preschoolers from poor homes after cognitive training in a kindergarten setting: Cognitive and ERP evidence". En: *Trends in neuroscience and education* 13 (2018), págs. 34-42.

[Pre+10] James Prescott y col. "Enhanced brain connectivity in math-gifted adolescents: An fMRI study using mental rotation". En: *Cognitive Neuroscience* 1.4 (2010), págs. 277-288.

[Spo10] Olaf Sporns. *Networks of the Brain*. MIT press, 2010.

[Tag+13] Enzo Tagliazucchi y col. "Large-scale brain functional modularity is reflected in slow electroencephalographic rhythms across the human non-rapid eye movement sleep cycle". En: *Neuroimage* 70 (2013), págs. 327-339.

[TL14] Enzo Tagliazucchi y Helmut Laufs. "Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep". En: *Neuron* 82.3 (2014), págs. 695-708.

[Tom+18] MI Tomescu y col. "From swing to cane: Sex differences of EEG resting-state temporal patterns during maturation and aging". En: *Developmental cognitive neuroscience* 31 (2018), págs. 58-66.

[Udd+13] Lucina Q Uddin y col. "Salience network–based classification and prediction of symptom severity in children with autism". En: *JAMA psychiatry* 70.8 (2013), págs. 869-879.

[Xu+14] Pengfei Xu y col. "Different topological organization of human brain functional networks with eyes open versus eyes closed". En: *Neuroimage* 90 (2014), págs. 246-255.

[YG+08] Shlomit Yuval-Greenberg y col. "Transient induced gamma-band response in EEG as a manifestation of miniature saccades". En: *Neuron* 58.3 (2008), págs. 429-441.