



Universidad Pública de Navarra

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA Y BIOCENCIAS
GRADO EN CIENCIA DE DATOS

APLICACIÓN DEL ESTADO DEL ARTE DEL APRENDIZAJE PROFUNDO EN LAS INTERFACES CEREBRO-MÁQUINA

Trabajo de fin de Grado realizado por

Xabier González García

El presente trabajo ha sido desarrollado bajo la dirección de

Francisco Javier Fernandez Fernandez

Javier Fumanal Idocin

Agradecimiento

Este trabajo no hubiese sido posible sin el apoyo e implicación de los profesores del departamento de Estadística, Informática y Matemáticas los cuales me han formado académicamente y profesionalmente. Especialmente, quiero agradecer al investigador pre-doctoral Javier Fumanal Idocin y al Dr. Francisco Javier Fernández Fernández por guiarme durante todos estos meses en el desarrollo del proyecto.

También quiero agradecer a aquellos que me han hecho ser mejor persona y a mi familia, que ha depositado toda su confianza, tiempo y recursos para que reciba la educación que ellos en su día no pudieron tener.

Por último, como estudiante de primera promoción, me gustaría reconocer el gran esfuerzo de los profesores por sacar este grado hacia adelante. La universidad Pública de Navarra ha cumplido y superado todas mis expectativas.

Índice

1. Introducción	4
1.1. Interfaz cerebro-máquina	5
1.2. Logros y Retos	6
1.3. Aprendizaje profundo	7
2. Estado del arte	9
2.1. Metodología clásica	9
2.2. Actualidad	10
2.2.1. Redes neuronales convolucionales profundas	10
2.2.2. Redes neuronales recurrentes	11
2.2.3. Redes de creencias profundas	12
2.2.4. Autoencoder	12
3. Métodos	13
3.1. Red Neuronal Multi-escala	13
3.1.1. Bloque de representación espectro-temporal	13
3.1.2. Bloque de características espaciales	14
3.1.3. Bloque de clasificación	14
3.2. Shallow ConvNet	15
4. Datos	16
4.1. Gix034	16
4.2. IV BCI Competition Dataset 2a	20
5. Experimentación	22
5.1. Configuración	22
5.2. Resultados	23
5.2.1. Resultados: BCI IV Competition Dataset 2a	23
5.2.2. Resultados: Gix034	25
6. Conclusiones	29
7. Futuras mejoras	29

RESUMEN

Los avances tecnológicos han facilitado la tarea de captar actividad bioeléctrica cerebral mediante dispositivos electrónicos. Las interfaces cerebro-máquina son las encargadas de integrar la actividad captada con ordenadores para su posterior análisis. No obstante, obtener datos de calidad partir de las señales cerebrales no es una tarea sencilla. Diversas técnicas han sido estudiados durante décadas en la comunidad científica, dentro de las cuales destacan, por su rendimiento y resultados prometedores, los modelos de aprendizaje automático.

El presente trabajo tiene como objetivo implementar y experimentar con técnicas de aprendizaje profundo del estado del arte para así descifrar la información que traen este tipo de señales, haciendo uso de la Tecnología de Interfaz cerebro-máquina.

Keywords Interfaz cerebro-máquina · Señales cerebrales · Aprendizaje Profundo · Aprendizaje Automático · Estado del arte

ABSTRACT

Technological advances have made it easier to capture bioelectrical brain activity using electronic devices. Brain-machine interfaces are responsible for integrating the captured activity with computers for further analysis. However, obtaining quality data from brain signals is not a simple task. Several techniques have been studied for decades in the scientific community, among which machine learning models stand out for their performance and promising results.

The present work aims to implement and experiment with state-of-the-art deep learning techniques in order to decipher the information brought by this type of signals, making use of Brain-Computer Interface Technology.

Keywords Brain-Computer Interface · Brain Signals · Deep learning · Machine learning · State of the art

1. Introducción

Las neuronas son como misteriosas mariposas del alma, cuyo batir de alas quién sabe si esclarecerá algún día el secreto de la vida mental.

Santiago Ramón y Cajal

El sistema nervioso es conocido por tener la capacidad de comunicar y coordinar los demás elementos del individuo, desde los órganos internos, hasta los movimientos y acciones. Está formado por redes celulares especializadas en transmitir señales e impulsos electroquímicos, los cuales se originan en el cerebro y se transmiten mediante la sinapsis [Nieuwenhuys, 2009].

Se denomina neurona a la unidad celular del sistema nervioso cuyo principal órgano es el cerebro. Las neuronas tienen la capacidad de transmitir señales eléctricas mediante potenciales de acción, los cuales modifican su distribución de carga eléctrica. Además, están interconectadas formando redes neuronales que funcionan como un canal de comunicación en el individuo. La carga eléctrica emitida se denomina actividad cerebral, la cual, se encarga de coordinar desde acciones dinámicas hasta conceptos complejos como la memoria y el razonamiento. Sin embargo, a día de hoy, pese a conocer la fisiología y los mecanismos internos del cerebro humano, aún son muchas las incógnitas acerca de este órgano tan importante [LeDoux, 1998].

A finales del siglo XIX, el catedrático y premio nobel Santiago Ramón y Cajal, conocido como el padre de la neurociencia, propuso la doctrina de la neurona [López-Muñoz et al., 2006]. Esta teoría, vigente a día de hoy, introdujo el concepto de neurona, como unidad celular individual del sistema nervioso. Hasta entonces, se pensaba que el tejido cerebral estaba formado por una única red continua.

En el año 1924, se dio a conocer una nueva disciplina, la electroencefalografía. Esta, fue impulsada por el neurólogo Hans Berger [Ínce et al., 2021], quien consiguió registrar la actividad cerebral humana mediante un electroencefalograma, más comúnmente denominado como EEG. Estos dispositivos están compuestos de sensores colocados estratégicamente para medir la carga eléctrica de distintas regiones cerebrales. La finalidad del estudio consistía en conocer la relación entre los eventos físicos cerebrales y los procesos psíquicos. Años más tarde, el científico fue reconocido internacionalmente por su descubrimiento.

Desde entonces, la electroencefalografía ha evolucionado en base al desarrollo de la neurociencia, las matemáticas y tecnología, adquiriendo una mayor relevancia en estudios cognitivos. Por ejemplo, técnicas de procesamiento de señal como la transformada rápida de Fourier (FFT) [Bracewell and Bracewell, 1986], permitieron analizar la composición espectral de señales de una manera rápida y eficaz. En el siglo XXI se ha conseguido integrar el cerebro con las máquinas.

1.1. Interfaz cerebro-máquina

La Interfaz cerebro-máquina, a partir de ahora ICM, es un sistema de medición, procesamiento y traducción de la actividad bioeléctrica. En función del problema a resolver y los objetivos a alcanzar, la medición se realizará con distintos electrogramas [Reilly and Lee, 2010]. Entre los más conocidos destacan la electroencefalografía (EEG), la resonancia magnética (MRI), la electromiografía (EMG) y la electrocorticografía (ECoG). El preprocesamiento es el apartado mediante el cual se extraen características relevantes a partir de diversas técnicas. Finalmente, se hace uso de la información para satisfacer las necesidades propuestas. Por lo tanto, ICMs son una potente herramienta de conocimiento ya que, permiten cuantificar pensamientos, emociones y acciones, además de su contribución a la neurociencia sobre el funcionamiento del cerebro humano/animal.

El presente trabajo estará enfocado exclusivamente a las señales de imaginería motora (IM) registradas mediante electroencefalografía [Lotze and Halsband, 2006]. La imaginería motora, es un proceso en el que un individuo simula mentalmente una acción motora, por ejemplo, pensar en levantar una mano [Masoomi and Khadem, 2015]. En la IM, se provoca un cambio en la actividad cerebral para realizar el evento, por lo tanto son procesos espontáneos [Busch and VanRullen, 2010]. Este tipo de actividad cerebral, adopta la forma de ondas rítmicas, las cuales son comúnmente conocidas como ondas cerebrales. Son complejas de analizar debido a la cantidad de información que circula cada milisegundo por la red neuronal y la diversidad que se da entre individuos en un mismo entorno experimental [Saha and Baumert, 2020].

Una de las posibles aplicaciones en las que se requiere analizar estas ondas es en la clasificación de eventos, donde la finalidad es predecir el tipo de evento que lleva la señal. Para ello, se han de conocer los tipos de ondas existentes. Las ondas cerebrales se clasifican en base a sus patrones y dominios de frecuencia [Woaswi and Hamzah, 2018].

- Ondas γ : $> 27Hz$ Se relacionan con el **aprendizaje**.
- Ondas β : $[12, 27]Hz$ Suelen estar correlacionados con la **concentración**.
- Ondas α : $[8, 12]Hz$. Varían entre individuos. Se asignan a estados de **relajación, visualización**.
- Ondas θ : $[3, 7]Hz$. Situaciones de **relajación profunda**.
- Ondas δ : $[0, 5, 3]Hz$ Situaciones de **sueño profundo**.

Por lo tanto, las ondas cerebrales son portadoras de información a diferentes rangos de frecuencias.

1.2. Logros y Retos

Los ICM, tienen un amplio espectro de aplicaciones.

Su principal desarrollo se da en el área de la medicina y la neurorehabilitación. Sin embargo, existen múltiples sectores que también se benefician de los avances en esta tecnología. En el sector del entretenimiento, se han desarrollado múltiples prototipos de videojuegos en los que las acciones se toman mentalmente [Marshall et al., 2013]. Por otro lado, la sinergia entre la realidad virtual y los ICMs revolucionarán las aplicaciones comerciales, permitiendo a los usuarios sumergirse en entornos y experiencias mucho más realistas [Kerous et al., 2018].

De la misma forma, el sector del marketing ha desarrollado nuevas metodologías para conseguir conectar a los usuarios con los productos de valor utilizando técnicas de neuromarketing [Lee et al., 2007]. Se ha de mencionar que la neurotecnología no solo se relaciona con procesos y habilidades mentales, si no que puede ser utilizada para mejorar acciones motoras que pueden ser de utilidad en la industria militar y del deporte.

Con la finalidad de mostrar la velocidad de los avances en esta tecnología, a continuación se mostrarán los proyectos ganadores del BCI Award en los últimos años.

- Año 2011. ¿Cuáles son las causas neurofisiológicas de las variaciones de rendimiento en ICM?. [Grosse-Wentrup, 2011]
- Año 2019. Decodificación del habla a partir de conjuntos de multielectrodos intracorticales en las áreas del brazo/mano. [Stavisky et al., 2018].
- Año 2016. Un BCI implantado para el control cortical en tiempo real de los movimientos funcionales de la muñeca y los dedos en un individuos con tetraplejia. [Bouton et al., 2016]
- Año 2020. Comunicación cerebro-texto de alto rendimiento a través de la escritura a mano [Willett et al., 2021]. El proyecto consistía en predecir letras dibujadas mentalmente con la finalidad de permitir la comunicación vía texto en personas discapacitadas.

Principalmente son dos factores los que frenan la investigación en ICM. Por un lado, la tecnología actual no ha conseguido desarrollar dispositivos no-invasivos lo suficientemente precisos como para que el ruido en las ondas no suponga un problema. Por otro lado, la naturaleza de las señales cerebrales dificulta su preproceso y compresión, ya que son datos de gran dimensionalidad y el coste para obtener muestras es elevado.

La neurotecnología no es algo que se asocie exclusivamente con la investigación y el mundo académico. El mercado global de productos *NeuroTech* tendrá un crecimiento de un 65 % según la fuente NeuroTech Reports. En el entorno nacional, en 2020 se impulsó la iniciativa SpainNeurotech, con la que se pretende innovar haciendo uso de la inteligencia artificial en la nanotecnología. No obstante, para implementar en la sociedad aplicaciones que hagan uso de este tipo de tecnología, se requiere un desarrollo previo del aspecto moral. De ahí la relevancia de la neuroética [Klein and Nam, 2016] [Munyon, 2018] [Nijboer et al., 2013].

1.3. Aprendizaje profundo

El aprendizaje profundo es una rama dentro de el aprendizaje automático especializada en modelos de redes neuronales artificiales.

Una red neuronal artificial es un mecanismo matemático que busca imitar la manera en la que las neuronas transmiten y almacenan la información. Se entrenan con datos, los cuales actúan como estímulos que ajustan los parámetros internos del modelo para posteriormente predecir valores desconocidos en base a lo aprendido. Se pueden utilizar en todo tipo de problemas relacionados con el aprendizaje automático, problemas de aprendizaje supervisado, no supervisado, semi-supervisado, aprendizaje por refuerzo... Sin embargo, su principal ventaja es la capacidad de abstracción que tienen siendo capaces de detectar las características más relevantes de los datos. Para ello, previamente se ha de definir la arquitectura interna de la red.

Se define como perceptrón a la unidad básica de una red neuronal artificial estándar. El concepto perceptron fue mencionado por primera vez en en 1958 por el psicólogo Frank Rosenblatt [Rosenblatt, 1958] y su propósito es simular la funcionalidad de las neuronas cerebrales. Recibe una entrada de n elementos, y los agrega mediante una suma ponderada a la que se le suele añadir un término de sesgo. Para una entrada \vec{X} :

$$\vec{X}_{out} = f\left(\theta_0 + \sum_{i=0}^n \theta_i X_i\right)$$

Donde los parámetros θ son internos y exclusivos al perceptron y f es una función de activación no lineal [Nwankpa et al., 2018]. El vector de salida de perceptron puede ser captado por otros perceptrones, los cuales generarán nuevas salidas en base a sus parámetros. De esta manera, se pueden concatenar perceptrones formando una red neuronal artificial multicapa: MLP.

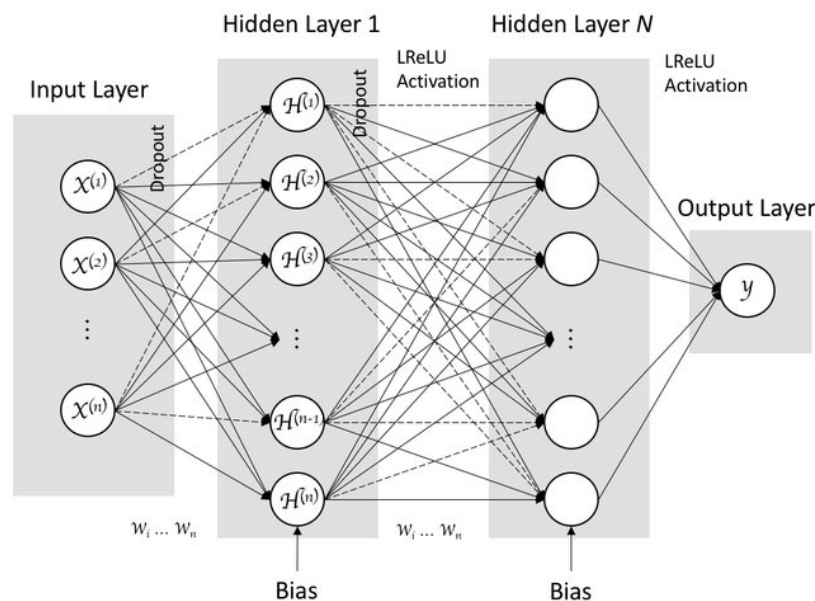


Figura 1. Arquitectura de una red neuronal multicapa. Imagen tomada del artículo [Loukas et al., 2017]

Se denomina capa al conjunto de perceptrones que reciben los mismos datos de entrada. La última capa es la encargada de determinar la salida real correspondiente al vector inicial de datos. Por ejemplo, en un problema de detección de cáncer (clasificación binaria supervisada), teniendo una imagen de radiografía como entrada, la capa de salida determinaría si el paciente tiene cáncer o no. Para ello, el modelo sigue un algoritmo de entrenamiento en el que se ajustan los parámetros de cada neurona. En problemas de aprendizaje supervisado el algoritmo se divide en varias fases:

1. Inicialización de los parámetros: La forma más común de realizarlo es de forma aleatoria. No obstante, hay toda una literatura acerca de cómo inicializar los parámetros de una red neuronal artificial para evitar problemas como el de la *simetría rota*.

Iterar:

2. Ejecutar el paso hacia delante: Calcular las salidas del modelo en base a los datos de entrada mediante la transmisión secuencial entre capas. Dada una configuración de N capas y siendo k el número de neuronas en la última capa oculta, se tiene que la que la salida del ejemplo j es \hat{y}^j .

$$\hat{y}^j = f\left(\theta_0 + \sum_{i=0}^k \theta_{iN} X_{i(N-1)}^j\right)$$

3. Calcular el error cometido. Dependiendo del tipo de problema se utilizarán diversas funciones de coste E .

$$\varepsilon^j : E(y^j, \hat{y}^j)$$

4. Aplicar el algoritmo de retro-propagación [Rumelhart et al., 1986] para calcular las derivadas parciales del error respecto a los parámetros.

$$\frac{\partial E}{\partial \theta_{il}}$$

5. Actualizar los parámetros mediante un algoritmo de optimización iterativo con el fin de minimizar ε . El descenso por gradiente [Robbins and Monro, 1951] es uno de los algoritmos más utilizados.

$$\theta(t)_{il} = \theta(t-1)_{il} - \alpha \frac{\partial E}{\partial \theta_{il}} \quad (1)$$

Donde t hace referencia al número de la iteración en la que se encuentra el algoritmo y α es un parámetro conocido como *ratio de aprendizaje* y es fijado por el programador previamente. De este modo, en cada iteración el modelo disminuirá el error cometido, siendo capaz posteriormente de interpretar datos similares a los que se ha enfrentado en su fase de entrenamiento. A pesar de ello, no es oro todo lo que reluce. La complejidad y el nivel de abstracción de los modelos de aprendizaje profundo dificultan la interpretación de los resultados y la manera en la que aprenden. Cuanto mayor sea el número mayor de parámetros y capas, más laborioso resultará descifrar el por qué de las salidas. Es por ello que este tipo de algoritmos se etiquetan como *cajas negras*.

El modelo estándar de red neuronal profunda, ha ido evolucionando con los años para adaptarse a distintos tipos de problemas. Nuevos paradigmas han surgido en el aprendizaje profundo desde entonces como las arquitecturas de redes convolucionales, utilizadas en tareas de imagen y vídeo, o las redes recurrentes, cuyo principales campos de aplicación son las series temporales y el procesamiento de lenguaje natural. Ahora bien, existen muchas más variantes y arquitecturas.

En lo que a los ICM respecta, las técnicas de aprendizaje profundo han sido utilizadas en múltiples problemas. A continuación, se revisará el estado del arte de técnicas más utilizadas.

2. Estado del arte

La revisión del estado del arte se basará en los siguientes artículos [Lotte et al., 2018] [Al-Saegh et al., 2021]. Principalmente se pueden dividir en dos los grupos de técnicas. Por un lado se hará un breve repaso a las técnicas clásicas de procesamiento y clasificación de ICM y por otro lado, los modelos más utilizados de aprendizaje profundo de la actualidad.

2.1. Metodología clásica

En el artículo [Lotte et al., 2018], se resumen las técnicas más utilizadas y relevantes hasta el 2018. Al comienzo del desarrollo de los ICM, los modelos implementaban algoritmos de extracción de características como el *common spatial patterns*, con las siglas CSP.

El CSP es un método similar al análisis de componentes principales, en el que se buscan los filtros espaciales (componentes) que maximicen la varianza entre clases a la vez que se minimiza la varianza interna de cada clase. De esta manera, siendo X_c la matriz de señales $N \times T$ de la clase c , donde N hace referencia al número de canales y T a la cantidad de observaciones temporales, se buscan M componentes w tal que:

$$CSP(X_c(t)) = W^T X_c$$

con W una matriz $N \times M$. Para ello se resuelve un problema de optimización que se reduce a encontrar los vectores propios w_i , que resuelvan la siguiente ecuación (dadas dos clases).

$$R_i = \frac{1}{N} X_i^T X_i$$

$$R_1 w = \lambda R_2 w$$

Donde el vector λ lo conforman los valores propios de la matriz de autocorrelación R_j .

De esta manera, tras aplicar el CSP, las nuevas componentes son clasificadas mediante modelos deterministas adaptativos capaces de encontrar el hiper-plano que mejor separare las clases. Entre ellos, destacan el análisis discriminante lineal/cuadrático (LDA/QDA), las máquinas de vectores de soporte (SVM) y clasificadores los Gaussianos [Vidaurre et al., 2007], [Li et al., 2013].

Existen técnicas más avanzadas relacionadas con el CSP. Un ejemplo de estas sería el algoritmo de mínima distancia Riemanniana a la media, RMDM [Barachant et al., 2011]. Como se ha podido comprobar en el CSP, las matrices de correlación y covarianza portan información valiosa acerca del

carácter de la señales. Además, se encuentran en el espacio de las matrices simétricas semidefinidas-positivas y, por tanto, caen dentro del dominio de la geometría de Riemann.

Utilizando un marco de geometría diferencial, se pueden mapear los datos de entrada a un espacio geométrico de Riemann mediante una métrica (aplicación) específica. Una vez mapeados los datos, se procede a utilizar un clasificador.

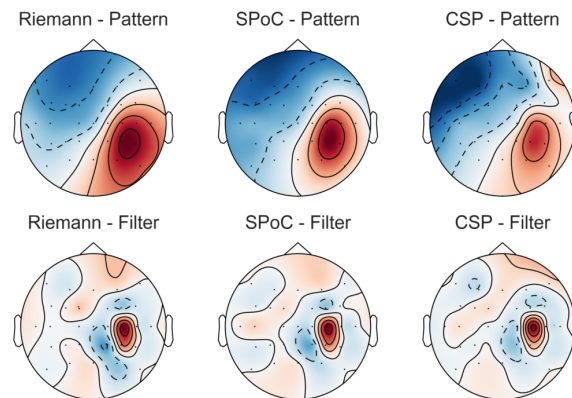


Figura 2. Características espacial-espectrales extraídas tras el uso de varios algoritmos. [barachant et al., 2016]

En problemas de aprendizaje no supervisado era común el uso de clustering con variantes del KNN para resolver la misma tarea [Yazdani et al., 2009].

2.2. Actualidad

Finalmente aparecen las redes neuronales. Es cierto que desde un principio han estado presentes las clásicas arquitecturas en este tipo de problemas, sin embargo, los nuevos paradigmas del aprendizaje profundo, hacen que tomen una mayor importancia. Las técnicas de aprendizaje profundo, permiten plantear un nuevo marco en el que el preproceso de la señal y el clasificador las realiza un mismo modelo. Lo cual no marca la conveniencia o no de preprocesar la señal mediante métodos tradicionales.

2.2.1. Redes neuronales convolucionales profundas

La Red Neuronal Convolucional (CNN) nace a partir del operador de convolución. La idea es que datos de entrada se convolucionan (convolución discreta) con los parámetros de la red, que pasan a tomar el nombre de filtro. A diferencia de los parámetros en los perceptrones, los filtros son propios de una capa convolucional y pueden tomar hasta tres dimensiones.

En el caso de los filtros unidimensionales, el operador convolución actúa como una media móvil ponderada por los pesos del filtro. Al introducir más dimensiones, la convolución proyecta el valor que representa una región de los datos. El entrenamiento de la red se realiza, al igual que en los MLP, ajustando los valores de los filtros en cada iteración.

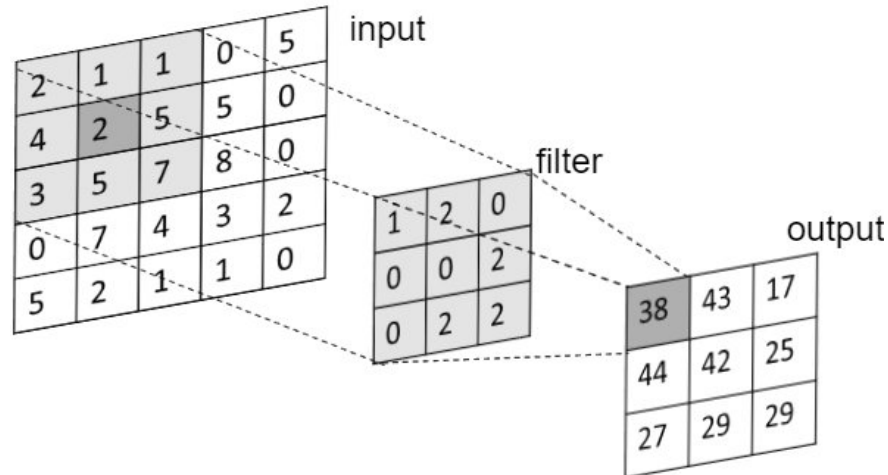


Figura 3. Operador convolución: Forma en la que un filtro actúa sobre una matriz de datos 2-dimensional. Imagen tomada del artículo [Aminu Bello et al., 2020]

Es muy frecuente el uso de este tipo de arquitecturas en problemas de visión, puesto que permiten seleccionar las características relevantes de una imagen y captar la interacción espacial en los datos. La idea es que los filtros de las primeras capas de la red, sean capaces de desarrollar un preprocesamiento de los datos, para que posteriormente, las capas de clasificación determinen su naturaleza. En ese sentido, las CNN toman un papel importante en los EEG. Las ondas recogidas por un electroencefalograma, cuentan con tres componentes que deberán de ser analizadas en conjunto:

- La componente **espacial**, se referencia con la posición del electrodo. Dependiendo del evento, la actividad cerebral diferirá en las diferentes zonas cerebrales.
- La componente **temporal** de las ondas es la que define el estado en el que se encuentra un evento.
- Los valores que toman las ondas en un espacio-tiempo específico referencian a la componente **espectral**

Precisamente, la utilidad de las CNN viene a ser su capacidad de captar las relaciones espacio-temporales mediante los filtros 2-d. No obstante, a mayor volumen de parámetros, la cantidad de datos requerida es mucho mayor, dificultando su uso en problemas de ICM.

2.2.2. Redes neuronales recurrentes

Las redes neuronales recurrentes, RNN, se componen de células recurrentes las cuales, almacenan un estado interno (en la figura 4 denotado como h) que evoluciona con el tiempo. El objetivo de este tipo de redes es explotar las características temporales de los datos, de ahí que su principal uso se da en problemas relacionados con secuencias.

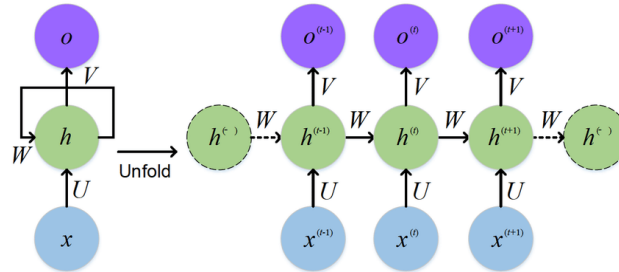


Figura 4. Célula estándar de una RNN. Imagen tomada del artículo [Feng et al., 2017]

El mecanismo recurrente de las RNN permite extraer información de la relación de orden de los elementos temporales, mediante el uso de transformaciones lineales. Éstas se aplican secuencialmente al vector de entrada y provocan la actualización del estado interno de cara a los próximos elementos de la serie. En su versión básica, suelen tener problemas de memoria a largo plazo perdiendo así la interacción entre elementos distantes. Sin embargo, existen arquitecturas específicas como la *Long short-term memory* [Wang et al., 2018], capaces de resolver parcialmente este tipo de problemas.

El principal inconveniente de los modelos recurrentes es que no son capaces de trabajar con la dimensión espacial de los datos. Para ello, existen modelos híbridos, como el CNN–LSTM en el que se combinan ambas arquitecturas.

2.2.3. Redes de creencias profundas

Este tipo de redes, sirven principalmente para problemas de aprendizaje no supervisado y se basan en modelos generativos. Las redes de creencias profundas, DBN, tratan de determinar las distribuciones probabilísticas de los datos de entrada mediante máquinas de Boltzmann restringidas (RBM) concatenadas [Xu and Plataniotis, 2016] [Hinton, 2012].

2.2.4. Autoencoder

Los autoencoders son otro tipo de modelos especializados en problemas de aprendizaje no supervisado. En la fase de codificación, se reduce la dimensionalidad de la entrada para posteriormente decodificarla obteniendo un nuevo conjunto de datos reconstruido. La idea de la reconstrucción es similar al proceso que llevan a cabo las RBM mencionadas [Chen et al., 2020].

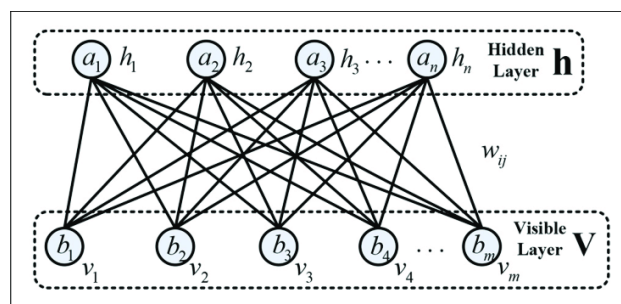


Figura 5. Estructura estándar de una máquina de Boltzmann restringida. Imagen tomada del artículo [Chu et al., 2018]

3. Métodos

El presente trabajo tiene como objetivo el implementar y analizar una red que explote las características espacio-temporal-espectrales de señales EEG con la finalidad de clasificar eventos de IM. Así pues, estamos ante un problema de clasificación de aprendizaje supervisado. Se plantea el uso de dos modelos, con el fin de comparar los resultados entre ellos. Una novedosa arquitectura de red multi-escala y por otro lado una red neuronal convolucional poco profunda.

3.1. Red Neuronal Multi-escala

La red neuronal multi-escala *MSNN* planteada en el artículo [Ko et al., 2021], permite expresar las relaciones espacio-temporal-espectrales las señales EEG aprovechando la propiedad multi-frecuencial que estas tienen. Para ello, se distinguen tres principales bloques dentro de la red.

3.1.1. Bloque de representación espectro-temporal

Supongamos que los datos de entrada tienen la siguiente dimensionalidad $[1, N_{electrodos}, N_{tiempo}]$. En un primer paso, se realiza una convolución temporal por canales en el vector de datos inicial con filtros $1 \times \frac{S_r}{2}$, donde S_r es la frecuencia de muestreo. Los autores justifican el hecho de utilizar filtros de tamaño $\frac{S_r}{2}$, basándose en el teorema de muestreo de Nyquist-Shannon [Shannon, 1949]. De esta manera, utilizando $N_{canales}$ filtros estaremos generando nuevos mapas de características a diferentes frecuencias y sin pérdidas de información, e.d., una salida de dimensionalidad $[N_{canales}, N_{electrodos}, N'_{tiempo}]$.

Posteriormente, se realizan tres convoluciones separables. Una convolución separable está formada por dos sub-convoluciones secuenciales. La primera, denominada convolución en profundidad se aplica a cada canal independientemente. Por lo tanto se utilizan filtros $[N_{electrodos} \times N'_{tiempo}]$ a cada nivel de profundidad. A continuación, la denominada convolución por puntos agrega, elemento a elemento los canales de la entrada haciendo uso de filtros 1×1 .

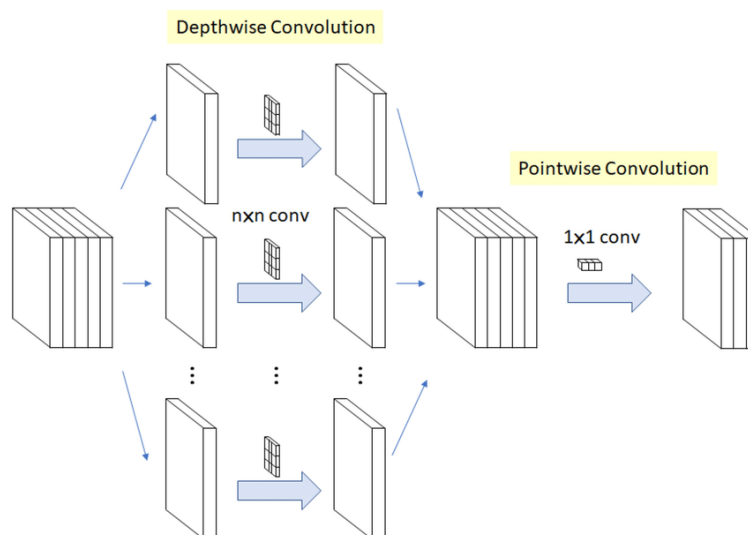


Figura 6. Convolución separable. Imagen tomada del artículo [Junejo and Ahmed, 2021]

Por lo tanto, si se seleccionan M filtros, las salidas de la convolución separable será de dimensión $[M, N_{electrodos}, N'_{tiempo}]$. Las ventajas de este tipo de convoluciones en problemas de ICM, tienen que ver con la manera en la que las características espectral-temporales son filtradas independientemente para posteriormente ser agregadas óptimamente [Ko et al., 2021] [Chollet, 2016].

3.1.2. Bloque de características espaciales

Tras cada convolución separable, los mapas de características se convolucionan (en profundidad) en la dimensión espacial con filtros $[N_{electrodos} \times 1]$, agregando así la componente espacial y obteniendo M vectores temporales. De esta manera, se extraen características espaciales representativas de cada rango de frecuencias. Posteriormente, se concatenan las salidas de cada convolución espacial formando un *embedding* de dimensión $[M', 1, N'_{tiempo}]$ siendo $M' = \sum_{i=1}^3 M_i$.

M_i hace referencia al número de filtros seleccionados en la convolución separable i -ésima.

3.1.3. Bloque de clasificación

El *embedding* se agrega mediante la media, para así, obtener un único canal que represente toda la información. La salida de esta operación, devuelve una matriz con una sola componente, la temporal $[1, 1, N'_{tiempo}]$. Finalmente, se aplican una o varias capas lineales y se obtienen las probabilidades de cada clase aplicando una activación *SoftMax*.

$$SoftMax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{Nclases} e^{z_j}} \quad \text{para } i = 1, 2, \dots, Nclases$$

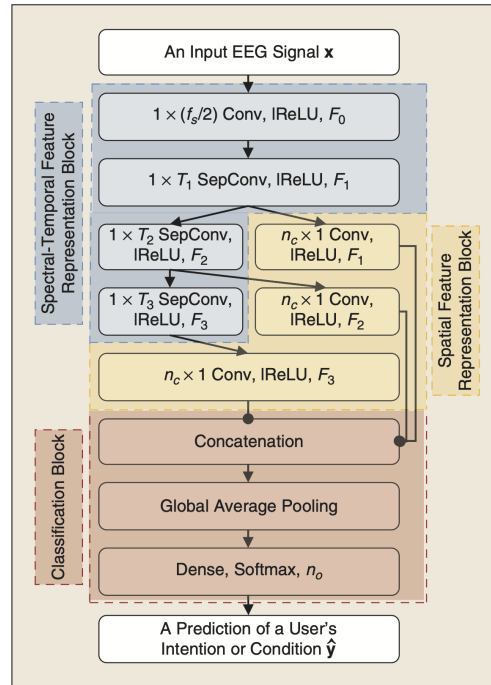


Figura 7. Esquema de la arquitectura MSNN. Imagen tomada del artículo [Ko et al., 2021]

3.2. Shallow ConvNet

El término *Shallow* en una arquitectura de red hace referencia a su poca profundidad. Lo cual implica, que tendrá pocas capas de convolución internas. A priori, puede parecer que el rendimiento de estas redes será menor respecto al de las red profundas. Esto se debe a que el mecanismo de capas internas es el que permite especializarse a cada neurona en características de interés. Al contar con menos neuronas, se puede estar perdiendo información relevante. Ahora bien, si se adaptan al problema, las redes *Shallow* pueden ser un acierto al ser mucho más simples y rápidas de entrenar. Siguiendo el *principio de la parsimonia*, en una situación en la que dos modelos tengan el mismo comportamiento, se deberá de escoger el más simple de ellos.

En el presente trabajo se trabajará con la configuración de red propuesta en el artículo [Schirrneister et al., 2017]. La novedad de esta variante es que las capas están diseñadas para llevar a cabo un proceso similar a las transformaciones que se realizan en el algoritmo *Filter Bank Common Spatial Pattern* (FBCSP), el cual combina filtrados de paso banda a diferentes frecuencias con el algoritmo CSP mencionado previamente. La motivación de utilizar la Shallow ConvNet (SCN) en vez de el propio algoritmo FBCSP, viene dada por la capacidad de adaptación y flexibilidad que tienen los filtros. Además, el proceso será mucho más rápido, escalable y se podrá optimizar.

Para ello, se realizan dos convoluciones consecutivas, sin utilizar funciones de activación de por medio. De esta manera, se fuerza a realizar las transformaciones lineales por separado con filtros unidimensionales. Inicialmente, una convolución temporal simulará filtros paso banda eliminando las frecuencias irrelevantes para la clasificación de la señal. Seguidamente las señales se convolucionan espacialmente, redimensionando los datos y obteniendo nuevas componentes espaciales tal y como ocurre en el CSP.

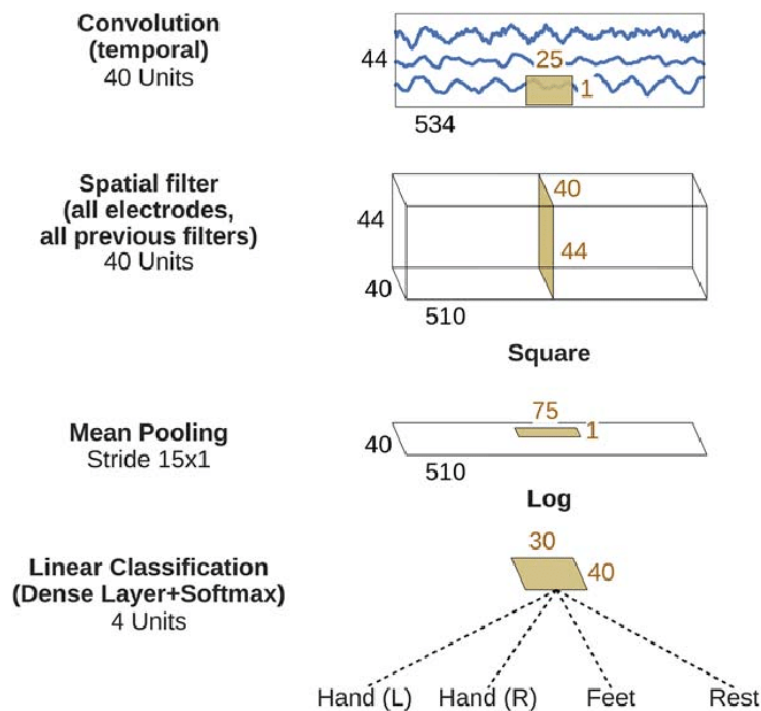


Figura 8. Esquema de la arquitectura SCN. Imagen tomada del artículo [Schirrneister et al., 2017]

Posteriormente, se utiliza una función de activación no-lineal elevando al cuadrado cada elemento y se agregan las componentes espaciales por la media, para obtener una sola componente que representa a todos los electrodos. Esto sucede de la misma manera que en la MSNN. Finalmente, se activa con el logaritmo natural y se utiliza una capa lineal para obtener las probabilidades finales con la función SoftMax.

Estos pasos, son análogos al cálculo de la varianza en el algoritmo FBCSP según el artículo. El cuadrado se toma de tal manera que los elementos más pequeños tiendan a 0 y los mayores se amplifiquen. Por otro lado, permite el uso posterior del logaritmo.

4. Datos

Debido a la dificultad de obtener señales cerebrales, ya sea por falta de recursos, aparatos electrónicos, personal experto y tiempo, se utilizarán datasets estructurados de IM para realizar los experimentos. Existen múltiples repositorios cedidos a la investigación en ICM, entre ellos, se encuentran el dataset Gix034 y los datos de la BCI Competition IV.

4.1. Gix034

El dataset Gix034 [Cho et al., 2017] pertenece a la base de Datos GigaDB. Entre otras cosas, cuenta con la información de 52 sujetos los cuales realizan 100-120 ensayos de IM por cada eventos/clase: Levantar la mano izquierda o la mano derecha. Las señales fueron recogidas por un EEG llamado *Biosemi ActiveTwo system* con 64 electrodos de AgCl (cloruro de plata). La distribución de los electrodos, conocida en la literatura como montaje, se basa en el sistema internacional 10-10. Lo cual significa que la distancia entre electrodos adyacentes es un 10% de la distancia total del cráneo, tal y como muestra la figura 9. Cada electrodo, está colocado estratégicamente para leer el lóbulo (área del cerebral) de interés y estos se nombran para referenciar el lóbulo asignado.

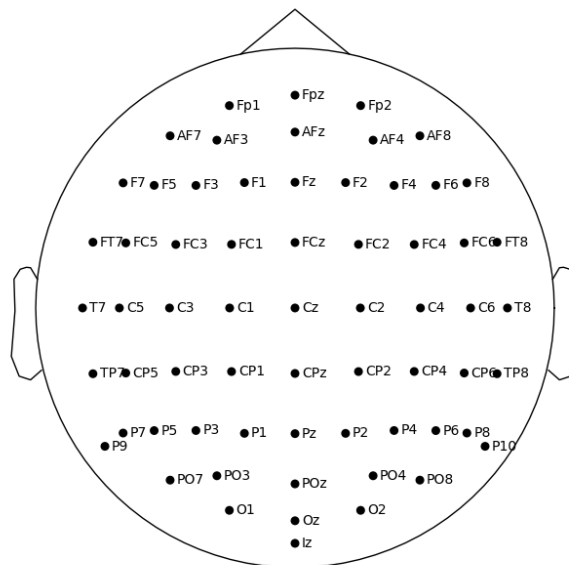


Figura 9. Montaje estándar 10-10 de 64 canales EEG.

Guía de etiquetas en EEG	
Etiqueta	Lóbulo
F	Frontal
T	Temporal
C	Central
P	Parietal
O	Occipital
A	Oreja
N	Nasion
I	Inion
Z	Linea media
IMPAR	Hemisferio derecho
PAR	Hemisferio izquierdo

Las señales fueron recogidas a 512Hz y la experimentación IM se realizó de la siguiente manera:

- Entre ensayo y ensayo, se dejan dos segundos de descanso.
- Posteriormente el sujeto realiza el evento de la clase correspondiente.
- El ensayo finaliza a los 5 segundos de haber comenzado el evento.

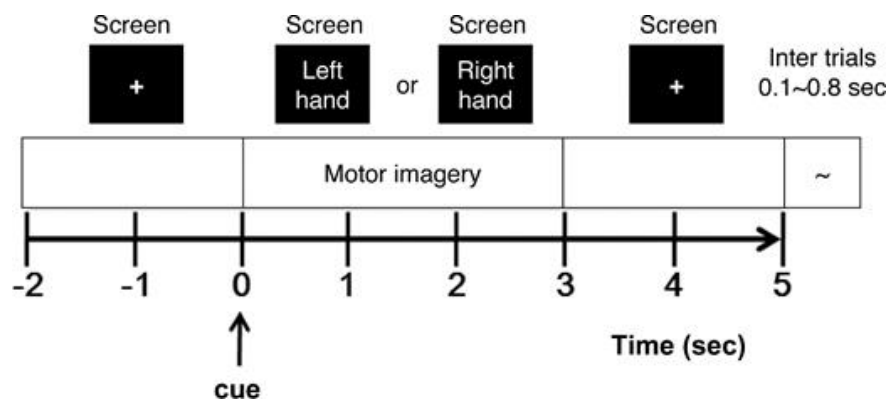


Figura 10. Esquema de un ensayo del experimento. Dataset gix034. . Imagen tomada del artículo [Cho et al., 2017]

Estos datos se utilizarán estos datos para plantear un problema de clasificación binaria intra-sujeto, lo cual significa que se entrenarán modelos independientes para cada uno. Tras cargar los datos, se calcula el ensayo promedio de cada evento para así poder visualizar la calidad de las señales y facilitar el preprocesamiento. Con la finalidad de disminuir la dimensionalidad, tan solo se seleccionarán los tres segundos de IM, de cada ensayo.

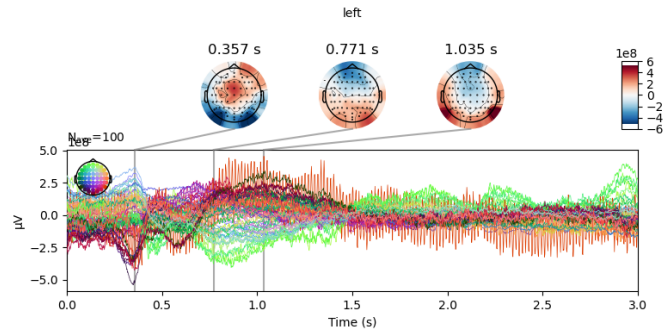


Figura 11. Gráfica del ensayo promedio **no** preprocesado del Sujeto01 del dataset Gix034 dado el evento IM: Mano izquierda.

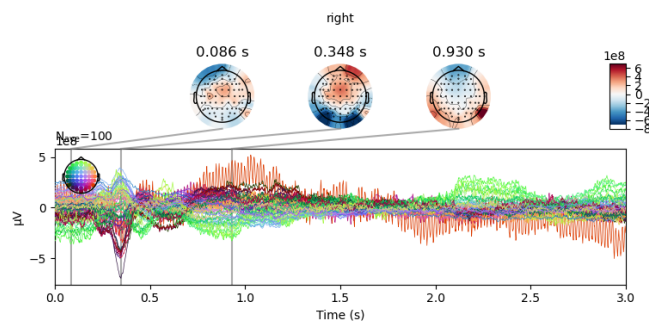


Figura 12. Gráfica del ensayo promedio **no** del Sujeto01 del dataset Gix034 dado el evento IM: Mano derecha.

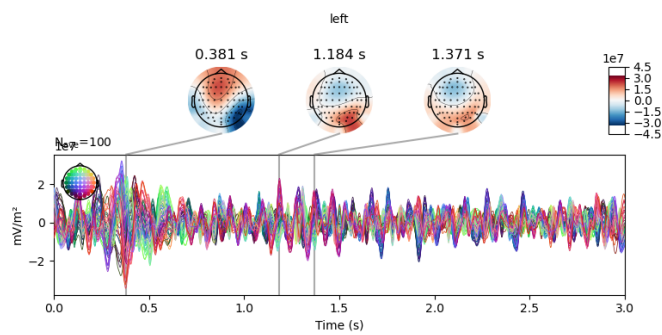


Figura 13. Gráfica del ensayo promedio preprocesado del Sujeto01 del dataset Gix034 dado el evento IM: Mano izquierda.

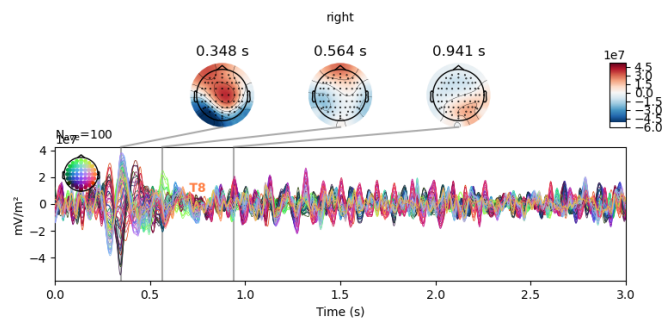


Figura 14. Gráfica del ensayo promedio preprocesado del Sujeto01 del dataset Gix034 dado el evento IM: Mano derecha.

Las figuras 11 y 12, muestran que el ruido está presente en la señales. Por otro lado, se puede llegar a observar que colores similares siguen trayectorias parecidas, lo cual significa que electrodos adyacentes captan una actividad cerebral similar. Para llevar a cabo el preproceso de los datos se aplica un filtrado paso banda *Butterworth* de orden 4 entre 4 y 40 Hz. De esta manera, filtramos las ondas θ y δ las cuales se relacionan con momentos de relajación profunda y sueño como se ha mencionado previamente. El uso del filtro *Butterworth* [Butterworth, 1930], permite que la salida de la señal permanezca lo plana más plana posible hasta la frecuencia de corte 15.

Seguidamente, se aplica el operador laplaciano de superficie discreto, el cual computa la segunda derivada espacial del voltaje registrado para cada electrodo [Kayser and Tenke, 2015], [McFarland et al., 1997]. Por cada electrodo del montaje, se genera una ventana 3x3 de los voltajes en cada instante temporal, y se convolucionan con el siguiente filtro 2-dimensional:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

De esta manera, se evitan cambios bruscos entre canales adyacentes al tener en cuenta la información del vecindario para cada electrodo. Finalmente, se normalizan las señales. Tras procesar las señales, se observa 13, 14 que se han eliminado frecuencias bajas y altas y se han homogeneizado las trayectorias por regiones espaciales.

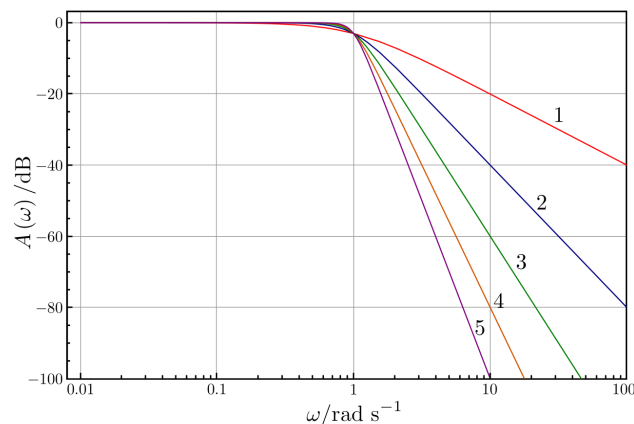


Figura 15. Filtro Butterworth de diferentes órdenes. Imagen tomada [via Wikimedia Commons, 2007]

4.2. IV BCI Competition Dataset 2a

El dataset 2a de la IV competición BCI [Brunner et al., 2008] presenta nuevamente un problema de IM pero esta vez de 4 clases: levantar la Mano izquierda/derecha, mover la lengua y ambos pies. En el experimento participaron 9 sujetos, los cuales realizaron 288 ensayos cada uno (72 por cada evento). Las señales fueron recogidas a 250 Hz en 22 electrodos Ag/AgCl. El montaje utilizado del EEG, se corresponde a una versión del sistema internacional 10-20. Además, en la versión utilizada de este dataset, las señales vienen preprocesadas. A priori, no hace falta aplicar ninguna técnica de limpieza en él.

Así pues, la idea de utilizar este dataset nace a partir de la hipótesis de que las redes son capaces de encontrar la manera de optimizar la limpieza de la señal. En un primer caso, los datos son procesados con un procedimiento generalista y superficial, sin entrar en mayor profundidad. En este segundo caso, se asegura una correcta limpieza de la señal. De esta manera, también se evaluará también la importancia del apartado de procesamiento de señales en los ICM.

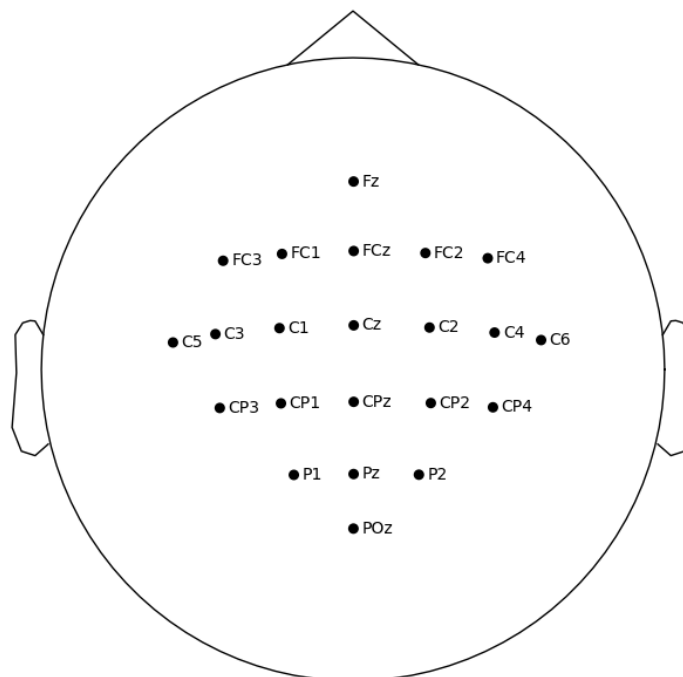


Figura 16. Montaje internacional 10-20 de 22 canales EEG.

A continuación, se muestran las señales promedio por electrodo del conjunto de datos para los distintos eventos. Los tiempos observados, se corresponden al momento del evento IM en el ensayo. Se observa un comportamiento distinto a las señales limpias del dataset Gix034. En este caso, las frecuencias altas parecen tener más presencia.

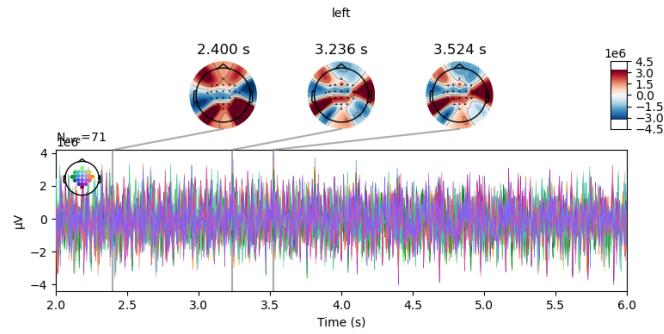


Figura 17. Gráfica del ensayo promedio preprocesado del Sujeto01 del V BCI Competition Dataset 2a dado el evento IM: Mano izquierda.

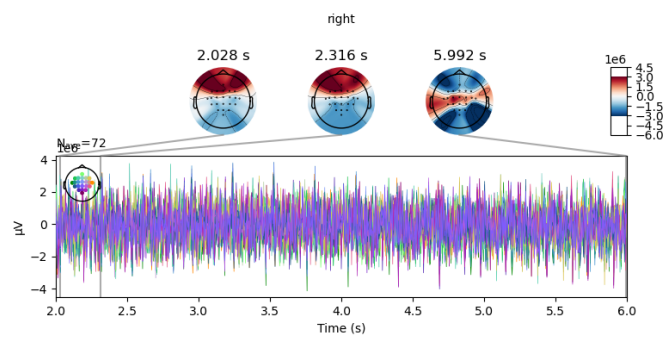


Figura 18. Gráfica del ensayo promedio preprocesado del Sujeto01 del V BCI Competition Dataset 2a dado el evento IM: Mano derecha.

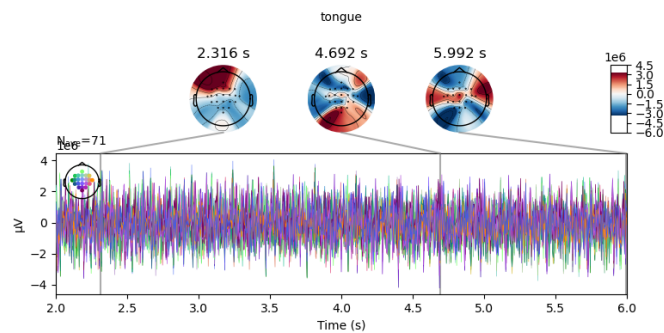


Figura 19. Gráfica del ensayo promedio preprocesado del Sujeto01 del V BCI Competition Dataset 2a dado el evento IM: Lengua.

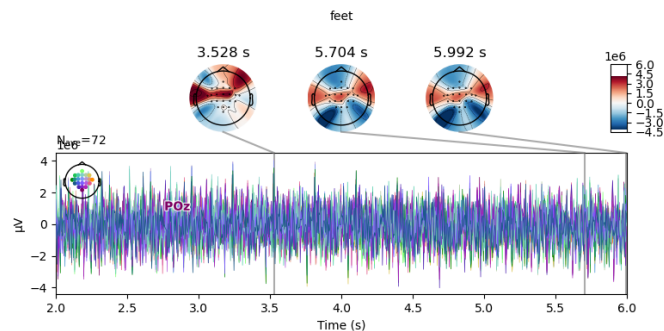


Figura 20. Gráfica del ensayo promedio preprocesado del Sujeto01 del V BCI Competition Dataset 2a dado el evento IM: Ambos pies.

5. Experimentación

El bloque de experimentación se compone de dos ramas. Se comenzará por definir la configuración de los experimentos para posteriormente analizar los resultados.

5.1. Configuración

Para realizar los experimentos en un mismo entorno, se ha decidido utilizar la misma configuración para ambos conjuntos de datos y modelos. Así, podremos comparar de manera realista su rendimiento.

Se ha decidido realizar una división aleatoria de los datos en dos subconjuntos, el de entrenamiento y el de validación. Al subconjunto de entrenamiento se le han asignado el 30% de los datos, y con ello, la red deberá de evaluarse con el 70% restante. Pese a que la división es aleatoria, se mantendrá en proporción el porcentaje de aparición de cada clase. Como en ambos conjuntos de datos se tienen el mismo número de clases, estarán totalmente balanceados. Es por ello, que la métrica a utilizar será el porcentaje de acierto total, más conocido como *Accuracy*. En el BCI Competition IV, el número de clases asciende a 4, por lo que se plantea también el uso de la métrica F1 para analizar si existe alguna clase silenciada por las demás.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1 = \frac{2Precision * Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

En cuanto a las redes, en ambos casos se utiliza el descenso por gradiente como optimizador para el algoritmo de la red, con una tasa de aprendizaje inicial de 10^{-4} y un parámetro de *momentum* de 0,9 para evitar extremos locales. La tasa de aprendizaje además será adaptativa. Para ello se programa un planificador del coeficiente, encargado de aumentar y disminuir el parámetro conforme avanzan las iteraciones. La tasa de aprendizaje evolucionará tal y como lo hace un coseno, oscilando y completando ciclos a una frecuencia fijada. En este caso cada 10 iteraciones o *epoch*. Este planificador es conocido en la literatura como *Cosine Annealing* [Lu et al., 2019]. La función de coste escogida es la entropía cruzada para N clases, la cual, encaja perfectamente con los problemas de clasificación a los que nos enfrentamos.

$$-\sum_{c=1}^N y_{i,c} \log(p_{i,c})$$

Se realizarán 4 experimentos por cada red en ambos conjuntos de datos. De esta manera, los resultados serán más fiables y se medirá la robustez de los clasificadores. En principio el único elemento aleatorio que interfiere en la red es la inicialización de los pesos, los demás factores serán exactamente iguales en cada experimento. Por otra parte, el número total de iteraciones de entrenamiento será de 100, de tal manera que se logre observar las diferentes fases del entrenamiento conforme avanza el tiempo. Para evitar problemas de sobre-ajuste, se han utilizan mecanismos como el *Dropout*, el cual silencia aleatoriamente elementos de las capas y la regularización L2.

5.2. Resultados

A continuación se mostrarán los resultados más relevantes obtenidos en los múltiples experimentos realizados. Para que un modelo sea exitoso, no tan solo deberá de obtener buenas métricas de Accuracy y F1, si no que también se evaluarán otros aspectos como la cantidad de datos entrenamiento requeridos, número de iteraciones que en los que converge el modelo, velocidad de las iteraciones... Además, se tendrá en cuenta también la homogeneidad de resultados y rendimiento frente a los diferentes sujetos. De la manera en la que se ha mencionado previamente, la variabilidad entre sujetos es muy alta, por lo que, a priori habrá sujetos en los que se dificulte el aprendizaje y viceversa.

5.2.1. Resultados: BCI IV Competition Dataset 2a

Comenzando con el conjunto de datos BCI Competition IV, se ha de galardonar a la red multi-escala MSNN, la cual obtiene un promedio de aciertos totales en el conjunto de validación de más del 99%. Teniendo en cuenta que tan solo se utilizan el 30% de los datos para entrenar, este es un resultado excepcional. La SCN por su parte, ha mostrado un correcto desempeño en la tarea obteniendo un promedio de aciertos del 85.05% manteniendo un F1 acorde. No obstante, los resultados no se acercan a la perfección de la MSNN.

Métricas en los datos de validación del conjunto BCI Competition IV2a				
	MSNN		SCN	
Sujeto	ACC	F1	ACC	F1
S01	1.0	1.0	0.8965	0.8472
S02	0.9954	0.9914	0.8390	0.7692
S03	0.9977	0.9969	0.8597	0.7902
S04	0.9954	0.9932	0.8620	0.7969
S05	1.0	1.0	0.9034	0.8482
S06	0.9977	0.9969	0.8137	0.7401
S07	0.9908	0.9877	0.8528	0.7946
S08	1.0	1.0	0.8666	0.7992
S09	0.9954	0.9944	0.8781	0.8213

En cuanto a la evolución del entrenamiento, se visualizan tanto el transcurso del Accuracy en cada paso, como el error cometido en el tiempo. Las ejecuciones han sido realizadas haciendo uso de la arquitectura CUDA con una tarjeta gráfica RTX 3080.

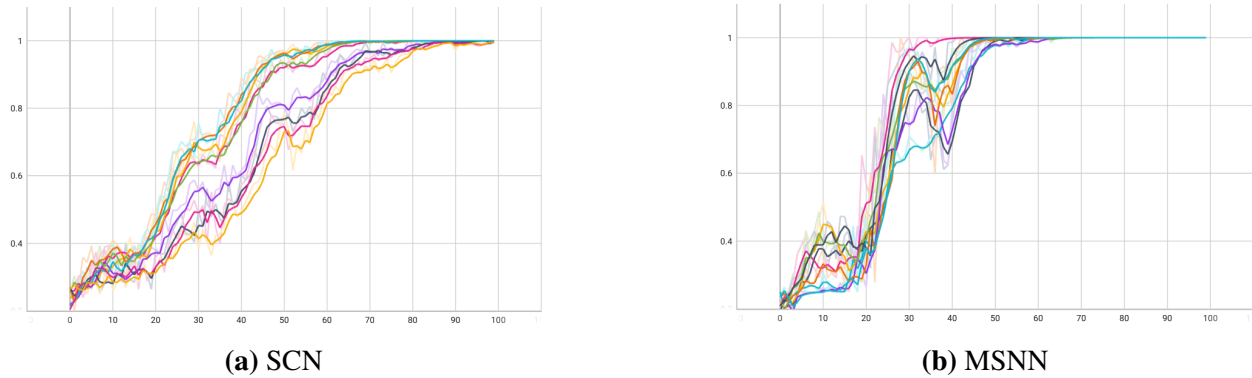


Figura 21. Evolución del Accuracy en un experimento en el dataset BCI Competition IV2a. Cada línea representa el entrenamiento de un sujeto.

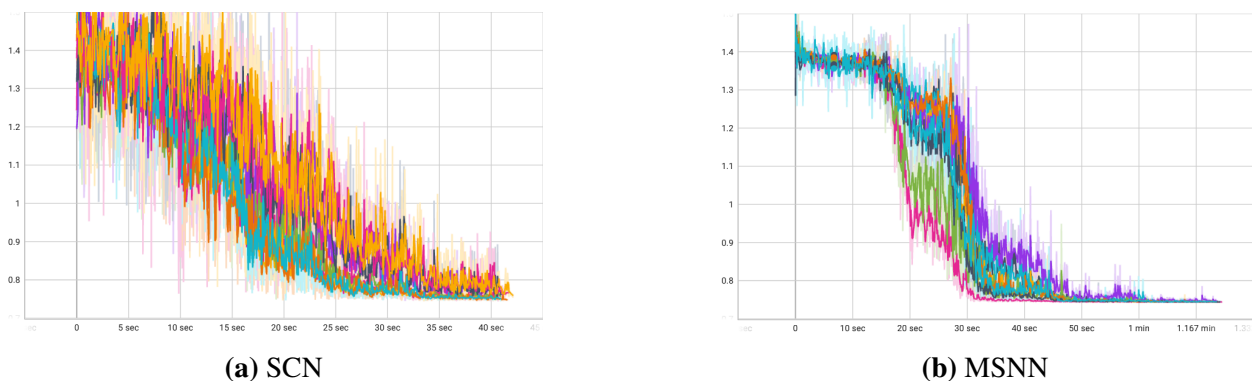


Figura 22. Evolución del Loss en un experimento en el dataset BCI Competition IV2a. Cada línea representa el entrenamiento de un sujeto distinto.

El modelo CSN es más rápido (40 segundos frente a 1.2 minutos), no obstante la MSNN converge en un número menor de iteraciones por lo que realmente no sería necesario completar las 100 si se incluyese un criterio de parada basado en tolerancia del error. Además, ambas redes alcanzan el 100% de acierto en el conjunto de datos de entrenamiento 21a, 21b.

La pendiente de descenso del error del MSNN es más pronunciada una vez supera el periodo de calentamiento en el que se mantiene estable 22b. La SCN por su parte, pese a mantener la tendencia bajista, las oscilaciones en una misma iteración muestran una mayor inestabilidad e incertidumbre 22a.

Es interesante también observar la evolución de las probabilidades de salida. Para ello, se generan histogramas 3d en los que la profundidad se corresponde al instante temporal del modelo. Contra más cercano sea el histograma, más avanzada será la iteración. La probabilidad 0 significa que el ejemplo actual no pertenece a la clase correspondiente y la 1 si lo pertenece. Esto se realiza por cada una de las 4 clases. En un primer caso se observa como la MSNN comienza asignando probabilidades entorno a 25% lo cual implica indecisión total. Sin embargo, esto se soluciona con el tiempo. Al final del entrenamiento todos los individuos se clasifican con probabilidades muy cercanas a los extremos 23b. En la SCN se tiene cierta incertidumbre en algunos ejemplos de las últimas iteraciones 23a.

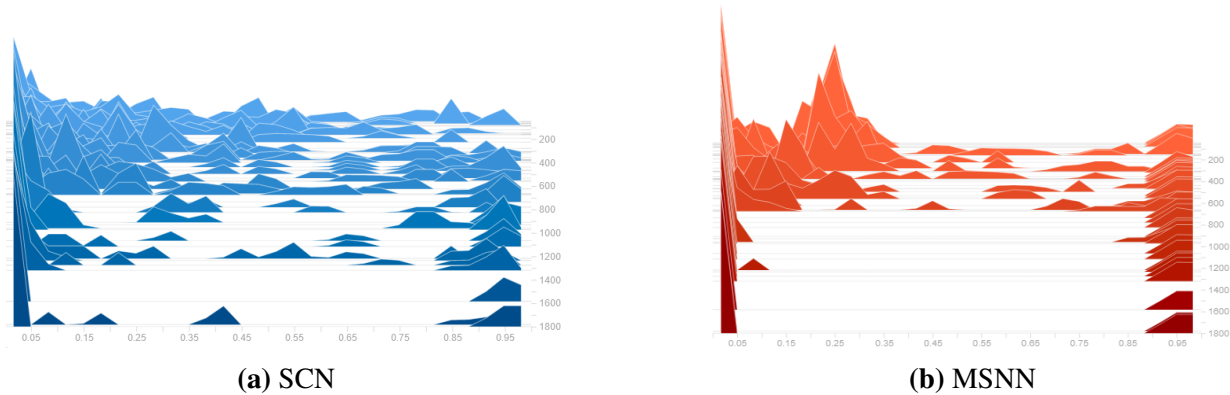


Figura 23. Evolución de las probabilidades de salida en un experimento en el dataset BCI Competition IV2a. El eje vertical, representa la evolución conforme avanza el entrenamiento.

5.2.2. Resultados: Gix034

Los resultados en el dataset Gix034, han sido totalmente opuestos a los del conjunto de datos anterior. Recordar, que en este caso se tienen dos posibles clases por lo que el peor modelo tendrá una tasa de acierto del 50 %.

Ambos modelos obtienen unos resultados parecidos pero el rendimiento en la MSNN es algo mejor. En la mayoría de sujetos el Accuracy está entorno al 45-55 % lo cual implica un resultado nefasto. No obstante, esto no sucede para los sujetos S03,S05 y S06. En estos casos se obtienen salidas significativamente mejores.

Métricas en los datos de validación del conjunto Gix034				
	MSNN		SCN	
Sujeto	ACC	F1	ACC	F1
S01	0.5285	0.4296	0.5433	0.4244
S02	0.4644	0.4843	0.4933	0.3986
S03	0.7722	0.7248	0.6433	0.5603
S04	0.4722	0.4026	0.5166	0.3880
S05	0.875	0.83685	0.8	0.7436
S06	0.7944	0.7207	0.7928	0.7326
S07	0.5714	0.5285	0.4388	0.3535
S08	0.5833	0.4969	0.5533	0.4822
S09	0.4629	0.4033	0.5	0.3434

El transcurso del entrenamiento asegura que en la MSNN existe sobre-ajuste en los datos de entrenamiento 24b. Esto puede estar relacionado con el echo de que al entrenar con el 30% de los datos, la variabilidad entre los eventos de un sujeto provocan el que el aprendizaje no sea extrapolable. Cabe destacar el caso del S05 al obtener un porcentaje de acierto del 87.5 % mientras que, en el experimento mostrado se mantiene constante entorno al 50 % (se corresponde a la línea morada). Además, su error también oscila entorno al valor inicial. Tras observar esto y, se realizaron varios experimentos en los que se retocaron los parámetros iniciales como la regularización, el optimizador, la capa dropout... pero los resultados no cambiaron demasiado.

En la SCN se distinguen tres principales patrones de entrenamiento^{24a}. Por un lado se muestran sujetos en el que el aprendizaje es exitoso. Por otro lado hay casos en los que el error no consigue disminuir pese que el Accuracy oscile, e.d. la red no es capaz de determinar la diferencia entre clases. Por último, encontramos ejemplos parecidos al del S05 en la MSNN, en los que la red no aprende nada.

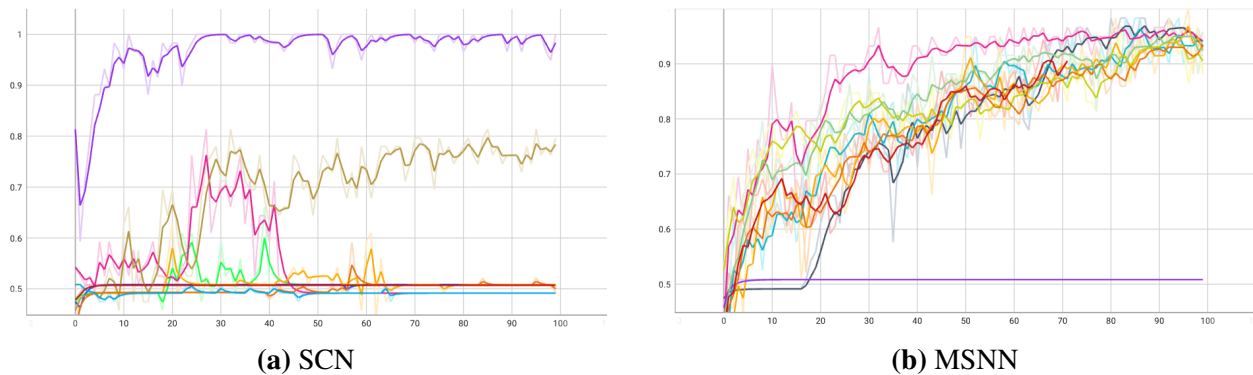


Figura 24. Evolución del Accuracy en un experimento en el dataset Gix034. Cada línea representa el entrenamiento de un sujeto distinto.

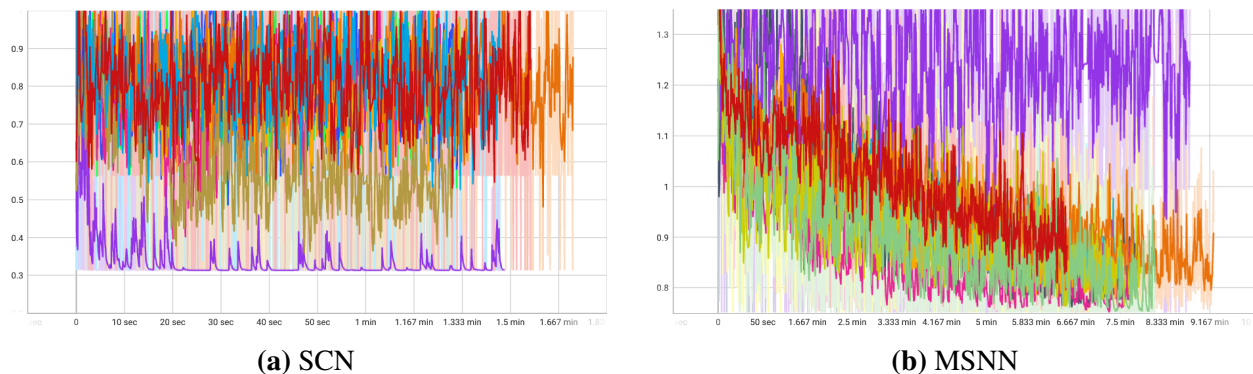


Figura 25. Evolución del Loss en un experimento en el dataset Gix034. Cada línea representa el entrenamiento de un sujeto distinto.

Nuevamente, se muestran la distribuciones de las probabilidades, en este caso, para tratar de averiguar el por qué de la diferencia entre sujetos. En la red MSNN a simple vista se encuentra un patrón natural de aprendizaje^{26b}. A diferencia del dataset anterior, en las iteraciones finales se encuentran ejemplos en los que la red duda. En la SCN, se encuentran casos en los que desde un principio, la red no cambia sus probabilidades^{26a}.

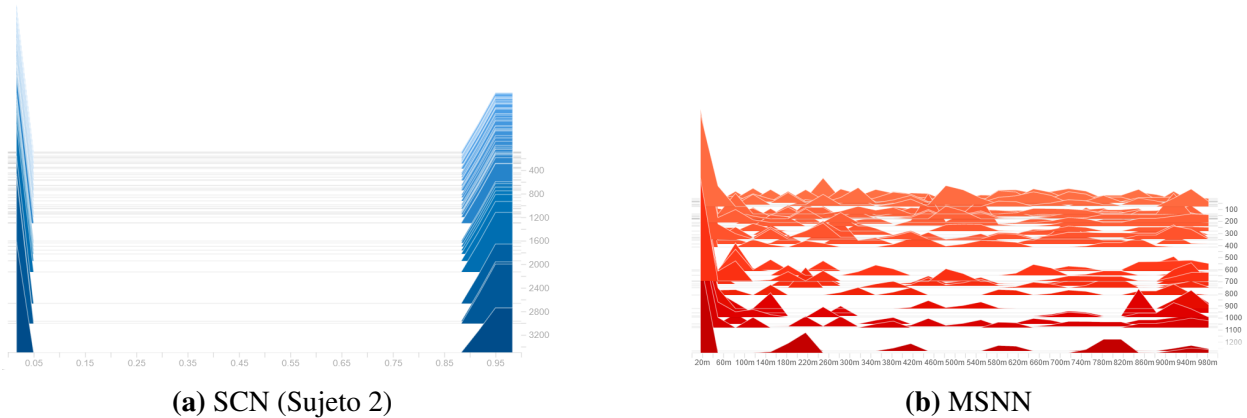


Figura 26. Evolución de las probabilidades de salida en un experimento en el dataset Gix034

Con la finalidad de encontrar la raíz del problema, se decide analizar a fondo el caso del S05, por ser el más extraño. Se ha de recalcar que la configuración de red es la misma en ambos datasets a si que, la única y exclusiva diferencia se deberá a los datos de entrada. Las siguientes imágenes 27 muestran en detalle los mapas topográficos de las señales del sujeto S05 entre los segundos [0,1] por ser la franja temporal donde mayor actividad se observa 13, 14. A simple vista se encuentran diferencias espacio-espectrales en diferentes instantes temporales. Este echo puede ser el motivo por el cual las redes son capaces de encontrar las características relevantes en este sujeto. En cualquier caso, esta imagen representa el evento medio, lo cual no implica que todos los eventos se comporten de esta manera.

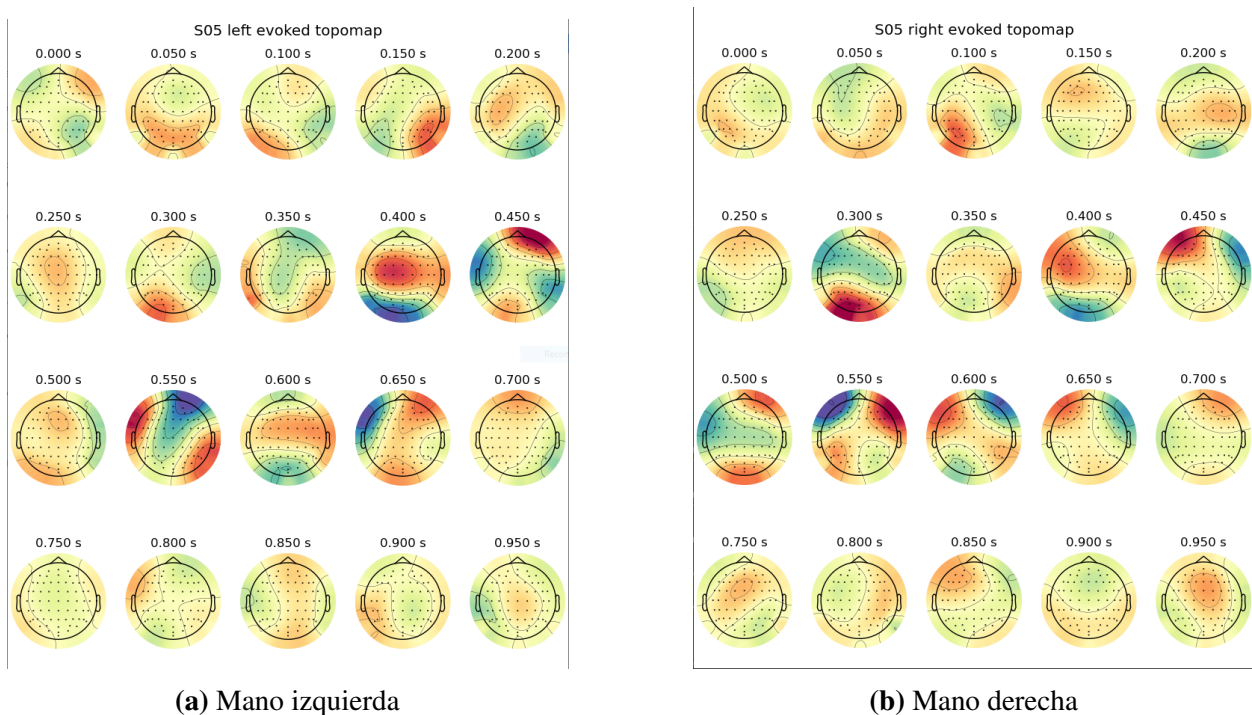


Figura 27. Evolución de los mapas topográficos del ensayo promedio del Sujeto05 del dataset Gix034.

Al analizar por separado los resultados del S05 en los experimentos de la SCN, se observa que en la mitad de experimentos el modelo no aprende mientras que en la otra mitad de experimentos, en cuestión de pocas iteraciones obtiene la manera de acertar el total de datos. Lo mismo ocurre en la MSNN para este sujeto. Ante esta situación, se plantean tres posibles escenarios.

El primero es que el funcionamiento de las redes está definida en su totalidad por la inicialización aleatoria de los pesos iniciales, los cuales son el único elemento de aleatoriedad en toda la configuración experimental [Sutskever et al., 2013]. A priori esta hipótesis no tiene mucho sentido por si sola, ya que, los pesos iniciales siguen la misma distribución y las diferencias en los resultados son exorbitantes.

El segundo escenario plantea que debido a un error de programación, los datos de entrenamiento y validación son los mismos. Esto último, se descarta al comprobar que el marco experimental ha sido el mismo en ambos dataset y redes. Además la comprobación manual mediante técnicas de *debugging*, han rechazado esta hipótesis también.

Por último, el problema puede ser debido a que los gradientes de la red se anulen en iteraciones iniciales de forma que, los parámetros dejen de ajustarse 1. Esto sucede cuando la variación del error respecto al cambio en un parámetro tiende a 0, provocando que el parámetro quede fijado. Pese a introducir mecánicas como el *momentum*, puede que el optimizador quede estancado en un punto silla en las primeras iteraciones [Sutskever et al., 2013]. Para ver si esto ocurre, se realiza un nuevo entrenamiento con la SCN para este sujeto. Tras recopilar los gradientes de las múltiples capas durante el entrenamiento, se genera un histograma global 28.

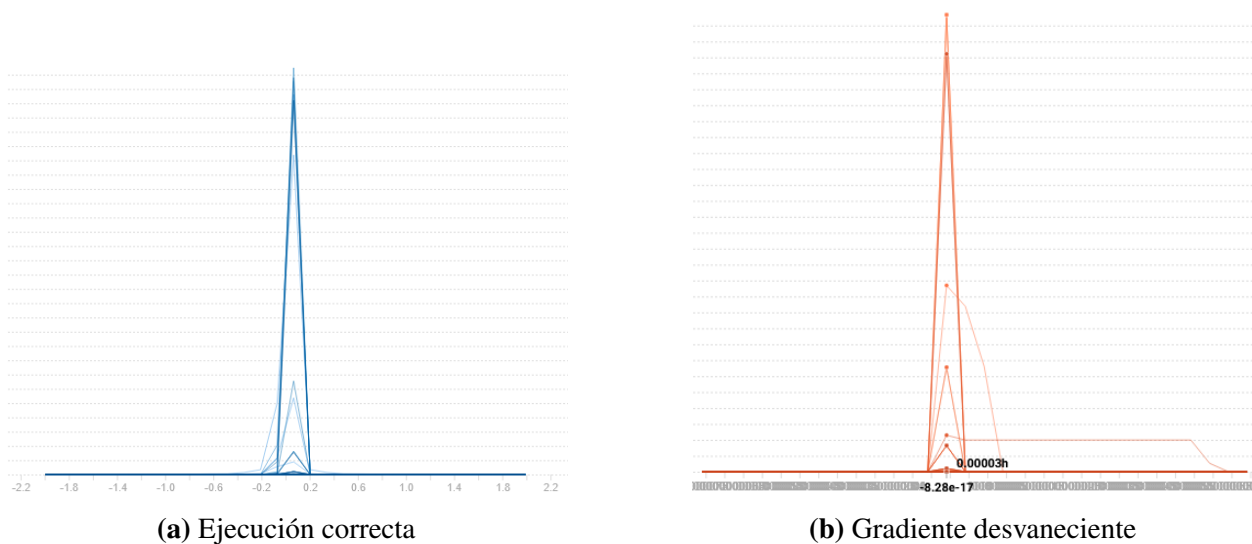


Figura 28. Histograma de los gradientes en el dataset Gix034

Efectivamente, hay experimentos en los que se da este suceso y el entrenamiento se inhabilita. En el ejemplo de la izquierda 28a, los gradientes están entorno al intervalo $[-2,2]$ centrados en 0, lo cual es totalmente lógico, debido a que la mayoría de parámetros de la red acabarán fijados tras un número de iteraciones. Por el contrario, en la imagen de la derecha se observa que todos los gradientes toman valores entorno a $[-10^{-16}, 10^{-16}]$. Por lo tanto, no hay entrenamiento 28b.

6. Conclusiones

El presente trabajo ha pretendido demostrar la importancia y relevancia que tienen las ICM en nuestra sociedad. El cerebro es una máquina que genera cantidades ingentes de información y los modelos de aprendizaje profundo han resultado ser capaces de solventar con precisión las tareas de clasificación de IM propuestas. Sin embargo, la calidad de estos modelos ha estado relacionada en su totalidad con la calidad de los datos.

En el caso del dataset BCI Competition IV2a, ambas redes han sabido extrapolar el conocimiento de los datos de entrenamiento a los de validación obteniendo resultados excepcionales. La MSNN por su parte, ha mostrado mayor robustez, eficacia y el sobre-aprendizaje es inexistente. Por ende, es la ganadora en este experimento. Además la variación entre sujetos no ha resultado un inconveniente.

No ha ocurrido lo mismo en el conjunto de datos Gix034. En la mayoría de sujetos ambas redes han tenido problemas en el entrenamiento y el sobre-aprendizaje es mayor que en el caso anterior. Se concluye que la causa del desvanecimiento de los gradientes es, que la inicialización aleatoria de los pesos en algunos casos conlleva a la red a una solución estacionaria en las primeras iteraciones. Las buenas prácticas como el uso de la tasa de aprendizaje adaptativa, incluir *momentum* en el optimizador, probar diferentes optimizadores etc., no han sido suficientes para solucionar el problema. Por lo tanto, se propone introducir conexiones residuales en el modelo al igual que en la arquitectura ResNet [Szegedy et al., 2017]. También convendría analizar en profundidad la generación de pesos del modelo. Hay toda una rama en la literatura que estudia el efecto de la inicialización de parámetros [Lee et al., 1991]. A su vez, el preprocesamiento utilizado ha podido disminuir la calidad de los datos. Al fin y al cabo, en este tipo de desarrollos, intervienen múltiples procesos y es la combinación de todos los que influyen en los resultados. A pesar de todo hay sujetos en los que se tienen buenos resultados confirmando la eficiencia de los modelos.

7. Futuras mejoras

Por último, se abren puertas a nuevos nuevos estudios y mejoras. Las arquitecturas han sido utilizadas en sus versiones estándar obviando las posibilidades de adaptación que tienen. Sin ir más lejos, se podrían hacer uso de técnicas heurísticas para optimizar parámetros de la red [Cummins et al., 2017], sustituir las agregaciones globales propuestas por otras mejor adaptadas [Dias et al., 2018] y hacer uso de técnicas de transferencia de conocimiento [Torrey and Shavlik, 2010], entre otras.

Referencias

- [Al-Saegh et al., 2021] Al-Saegh, A., Dawwd, S. A., and Abdul-Jabbar, J. M. (2021). Deep learning for motor imagery eeg-based classification: A review. *Biomedical Signal Processing and Control*, 63:102172.
- [Aminu Bello et al., 2020] Aminu Bello, S., Yu, S., Wang, C., Adam, J., and Li, J. (2020). Review: Deep learning on 3d point clouds. *Remote Sensing*, 12:1729.
- [Barachant et al., 2011] Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2011). Multi-class brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928.
- [barachant et al., 2016] barachant, a., Carmel, J. B., Friel, K. M., and Gupta, D. (2016). Extraction of motor patterns from joint EEG/EMG recording: A Riemannian Geometry approach. In *6th International Brain-Computer Interface Meeting*, Monterey, United States.
- [Bouton et al., 2016] Bouton, C. E., Shaikhouni, A., Annetta, N. V., Bockbrader, M. A., Friedenberg, D. A., Nielson, D. M., Sharma, G., Sederberg, P. B., Glenn, B. C., Mysiw, W. J., Morgan, A. G., Deogaonkar, M., and Rezai, A. R. (2016). Restoring cortical control of functional movement in a human with quadriplegia. *Nature*, 533(7602):247–250.
- [Bracewell and Bracewell, 1986] Bracewell, R. N. and Bracewell, R. N. (1986). *The Fourier transform and its applications*, volume 31999. McGraw-Hill New York.
- [Brunner et al., 2008] Brunner, C., Leeb, R., Müller-Putz, G., Schlögl, A., and Pfurtscheller, G. (2008). Bci competition 2008–graz data set a. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, 16:1–6.
- [Busch and VanRullen, 2010] Busch, N. A. and VanRullen, R. (2010). Spontaneous eeg oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, 107(37):16048–16053.
- [Butterworth, 1930] Butterworth, S. (1930). On the Theory of Filter Amplifiers. *Experimental Wireless & the Wireless Engineer*, 7:536–541.
- [Chen et al., 2020] Chen, J., Yu, Z., and Gu, Z. (2020). Semi-supervised deep learning in motor imagery-based brain-computer interfaces with stacked variational autoencoder. In *Journal of Physics: Conference Series*, volume 1631, page 012007. IOP Publishing.
- [Cho et al., 2017] Cho, H., Ahn, M., Ahn, S., Kwon, M., and Jun, S. C. (2017). EEG datasets for motor imagery brain–computer interface. *GigaScience*, 6(7). gix034.
- [Chollet, 2016] Chollet, F. (2016). Xception: Deep learning with depthwise separable convolutions.
- [Chu et al., 2018] Chu, Y., Zhao, X., Zou, Y., Xu, P., Han, J., and Zhao, Y. (2018). A decoding scheme for incomplete motor imagery eeg with deep belief network. *Frontiers in Neuroscience*, 12:680.

- [Cummins et al., 2017] Cummins, C., Petoumenos, P., Wang, Z., and Leather, H. (2017). End-to-end deep learning of optimization heuristics. In *2017 26th International Conference on Parallel Architectures and Compilation Techniques (PACT)*, pages 219–232. IEEE.
- [Dias et al., 2018] Dias, C. A., Bueno, J., Borges, E. N., Botelho, S. S., Dimuro, G. P., Lucca, G., Fernández, J., Bustince, H., and Drews Junior, P. L. J. (2018). Using the choquet integral in the pooling layer in deep learning networks. In *North american fuzzy information processing society annual conference*, pages 144–154. Springer.
- [Feng et al., 2017] Feng, W., Guan, N., Li, Y., Zhang, X., and Luo, Z. (2017). Audio visual speech recognition with multimodal recurrent neural networks. pages 681–688.
- [Grosse-Wentrup, 2011] Grosse-Wentrup, M. (2011). What are the causes of performance variation in brain-computer interfacing? *International Journal of Bioelectromagnetism*, 13(3):115–116.
- [Hinton, 2012] Hinton, G. E. (2012). A practical guide to training restricted boltzmann machines. In *Neural networks: Tricks of the trade*, pages 599–619. Springer.
- [Junejo and Ahmed, 2021] Junejo, I. and Ahmed, N. (2021). Depthwise separable convolutional neural networks for pedestrian attribute recognition. *SN Computer Science*, 2.
- [Kayser and Tenke, 2015] Kayser, J. and Tenke, C. E. (2015). On the benefits of using surface laplacian (current source density) methodology in electrophysiology. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 97(3):171–173. 26071227[pmid].
- [Kerous et al., 2018] Kerous, B., Skola, F., and Liarokapis, F. (2018). Eeg-based bci and video games: a progress report. *Virtual Reality*, 22(2):119–135.
- [Klein and Nam, 2016] Klein, E. and Nam, C. S. (2016). Neuroethics and brain-computer interfaces (bcis). *Brain-Computer Interfaces*, 3(3):123–125.
- [Ko et al., 2021] Ko, W., Jeon, E., Jeong, S., and Suk, H.-I. (2021). Multi-scale neural network for eeg representation learning in bci. *IEEE Computational Intelligence Magazine*, 16(2):31–45.
- [LeDoux, 1998] LeDoux, J. (1998). *The emotional brain: The mysterious underpinnings of emotional life*. Simon and Schuster.
- [Lee et al., 2007] Lee, N., Broderick, A. J., and Chamberlain, L. (2007). What is ‘neuromarketing’? a discussion and agenda for future research. *International Journal of Psychophysiology*, 63(2):199–204. Cognitive Neuroscience: Contributions from Psychophysiology.
- [Lee et al., 1991] Lee, Y., Oh, S.-H., and Kim, M. W. (1991). The effect of initial weights on premature saturation in back-propagation learning. In *IJCNN-91-Seattle international joint conference on neural networks*, volume 1, pages 765–770. IEEE.
- [Li et al., 2013] Li, S., Zhou, W., Yuan, Q., Geng, S., and Cai, D. (2013). Feature extraction and recognition of ictal eeg using emd and svm. *Computers in biology and medicine*, 43(7):807–816.

- [López-Muñoz et al., 2006] López-Muñoz, F., Boya, J., and Alamo, C. (2006). Neuron theory, the cornerstone of neuroscience, on the centenary of the nobel prize award to santiago ramón y cajal. *Brain research bulletin*, 70(4-6):391–405.
- [Lotte et al., 2018] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F. (2018). A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005.
- [Lotze and Halsband, 2006] Lotze, M. and Halsband, U. (2006). Motor imagery. *Journal of Physiology-paris*, 99(4-6):386–395.
- [Loukas et al., 2017] Loukas, G., Vuong, T., Heartfield, R., Sakellari, G., Yoon, Y., and Gan, D. (2017). Cloud-based cyber-physical intrusion detection for vehicles using deep learning. *IEEE Access*, 6:3491–3508.
- [Lu et al., 2019] Lu, Z., Wang, J., and Song, J. (2019). Multi-resolution csi feedback with deep learning in massive mimo system.
- [Marshall et al., 2013] Marshall, D., Coyle, D., Wilson, S., and Callaghan, M. (2013). Games, gameplay, and bci: The state of the art. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):82–99.
- [Masoomi and Khadem, 2015] Masoomi, R. and Khadem, A. (2015). Enhancing lda-based discrimination of left and right hand motor imagery: Outperforming the winner of bci competition ii. In *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pages 392–398. IEEE.
- [McFarland et al., 1997] McFarland, D. J., McCane, L. M., David, S. V., and Wolpaw, J. R. (1997). Spatial filter selection for eeg-based communication. *Electroencephalography and Clinical Neurophysiology*, 103(3):386–394.
- [Munyon, 2018] Munyon, C. N. (2018). Neuroethics of non-primary brain computer interface: Focus on potential military applications. *Frontiers in Neuroscience*, 12.
- [Nieuwenhuys, 2009] Nieuwenhuys, R. (2009). *El sistema nervioso central humano*, volume 2. Editorial Médica Panamericana SA.
- [Nijboer et al., 2013] Nijboer, F., Clausen, J., Allison, B. Z., and Haselager, P. (2013). The asilomar survey: Stakeholders’ opinions on ethical issues related to brain-computer interfacing. *Neuroethics*, 6(3):541–578.
- [Nwankpa et al., 2018] Nwankpa, C., Ijomah, W., Gachagan, A., and Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *CoRR*, abs/1811.03378.
- [Reilly and Lee, 2010] Reilly, R. B. and Lee, T. C. (2010). Electrograms (ecg, eeg, emg, eog). *Technology and Health Care*, 18(6):443–458.

- [Robbins and Monro, 1951] Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407. Full publication date: Sep., 1951.
- [Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.
- [Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- [Saha and Baumert, 2020] Saha, S. and Baumert, M. (2020). Intra-and inter-subject variability in eeg-based sensorimotor brain computer interface: a review. *Frontiers in computational neuroscience*, page 87.
- [Schirrmester et al., 2017] Schirrmester, R., Springenberg, J., Fiederer, L., Glasstetter, M., Eggenberger, K., Tangermann, M., Hutter, F., Burgard, W., and Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization: Convolutional neural networks in eeg analysis. *Human Brain Mapping*, 38.
- [Shannon, 1949] Shannon, C. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21.
- [Stavisky et al., 2018] Stavisky, S. D., Rezaii, P., Willett, F. R., Hochberg, L. R., Shenoy, K. V., and Henderson, J. M. (2018). Decoding speech from intracortical multielectrode arrays in dorsal “arm/hand areas” of human motor cortex. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 93–97.
- [Sutskever et al., 2013] Sutskever, I., Martens, J., Dahl, G., and Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR.
- [Szegedy et al., 2017] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.
- [Torrey and Shavlik, 2010] Torrey, L. and Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global.
- [via Wikimedia Commons, 2007] via Wikimedia Commons, O. (2007). Butterworth filter of several orders.
- [Vidaurre et al., 2007] Vidaurre, C., Schlogl, A., Cabeza, R., Scherer, R., and Pfurtscheller, G. (2007). Study of on-line adaptive discriminant analysis for eeg-based brain computer interfaces. *IEEE transactions on biomedical engineering*, 54(3):550–556.
- [Wang et al., 2018] Wang, P., Jiang, A., Liu, X., Shang, J., and Zhang, L. (2018). Lstm-based eeg classification in motor imagery tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(11):2086–2095.

- [Willett et al., 2021] Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., and Shenoy, K. V. (2021). High-performance brain-to-text communication via handwriting. *Nature*, 593(7858):249–254.
- [Woaswi and Hamzah, 2018] Woaswi, W. and Hamzah, N. (2018). Human emotion detection via brain waves study by using electroencephalogram (eeg).
- [Xu and Plataniotis, 2016] Xu, H. and Plataniotis, K. N. (2016). Eeg-based affect states classification using deep belief networks. In *2016 Digital Media Industry & Academic Forum (DMIAF)*, pages 148–153. IEEE.
- [Yazdani et al., 2009] Yazdani, A., Ebrahimi, T., and Hoffmann, U. (2009). Classification of eeg signals using dempster shafer theory and a k-nearest neighbor classifier. In *2009 4th International IEEE/EMBS Conference on Neural Engineering*, pages 327–330. IEEE.
- [İnce et al., 2021] İnce, R., Adanır, S. S., and Sevmez, F. (2021). The inventor of electroencephalography (eeg): Hans berger (1873–1941). *Child's Nervous System*, 37(9):2723–2724.