



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

INTEGRACIÓN MULTIMODAL PARA LA DETECCIÓN DE ZONA EPILEPTÓGENA
USANDO MECANISMO DE CO-ATENCIÓN

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL ELÉCTRICA

SOFIA PAZ CALLEJAS DIAZ

PROFESOR GUÍA:
PABLO ESTEVEZ VALENCIA

MIEMBROS DE LA COMISIÓN:
JHON INTRIAGO CHICHANDA
FRANCISCO RIVERA SERRANO

Este trabajo ha sido parcialmente financiado por ANID Chile a través del Instituto Milenio en Ingeniería e Inteligencia Artificial para la Salud, ICN2021-004 y Proyecto Fondecyt 1220829.

SANTIAGO DE CHILE
2024

Resumen

El objetivo de esta memoria de título es desarrollar y aplicar un algoritmo supervisado de inteligencia artificial para la co-localización de la alteración estructural en pacientes con epilepsia refractaria, mediante la integración de datos de electroencefalograma (EEG) y resonancia magnética funcional (fMRI), idealmente se desea identificar del foco epileptogénico.

En este estudio se utilizaron cuatro bases de datos distintas. La primera proviene del Children's Hospital Boston y comprende registros de electroencefalograma de pacientes pediátricos con convulsiones intratables. La segunda base de datos, aprobada por la Junta de Revisión Institucional (IRB) de la Universidad de Iowa, incluye registros de pacientes humanos sometidos a estimulación eléctrica durante resonancia magnética funcional. La tercera se compone de datos de pacientes del Child Mind Institute, que cuentan con información de EEG y fMRI para cada individuo. La cuarta base de datos se obtuvo en colaboración con el Instituto de Neurología Dr. Alfonso Asenjo (INCA), y se compone de registros de pacientes con epilepsia refractaria que incluyen EEG y fMRI para cada caso.

El algoritmo propuesto se basa en el aprendizaje conjunto de dos redes neuronales denominadas *Transformers*. Una de ellas se especializa en el análisis del electroencefalograma, que mide la actividad eléctrica cerebral, mientras que la otra se centra en la resonancia magnética funcional en reposo (rfMRI). La clave de este enfoque radica en la implementación de la co-atención, donde ambas redes trabajan de manera conjunta.

Al utilizar el modelo de co-atención para fMRI y EEG, se obtuvo una *focal accuracy* del 81.3%, superando el umbral del 80% y demostrando un rendimiento prometedor en esta fase preliminar. Esta investigación se realiza con cooperación con el Instituto Milenio en Ingeniería e Inteligencia Artificial para la Salud y el Instituto de Neurocirugía Dr Alfonso Asenjo.

Le dedico esta memoria de título a mis padres, gracias a ellos he llegado tan lejos.

Agradecimientos

Me gustaría comenzar expresando mi sincero agradecimiento a mis padres, Carmen y José, quienes han sido un apoyo incondicional en cada etapa de mi vida, incluso en los momentos más difíciles. Gracias a su constante respaldo, he tenido la oportunidad de recibir esta formación académica.

También quiero expresar mi profundo agradecimiento a mi amiga de toda la vida, Rafaela Justiniano. Ella ha sido un pilar fundamental en mi camino, agradeciéndole especialmente por su inquebrantable voluntad y paciencia a lo largo de los años. Su apoyo ha sido invaluable y ha dejado una marca significativa en mi vida.

Además, me gustaría expresar mi agradecimiento al profesor guía, Pablo Estevez, por sus sabios consejos y valiosas enseñanzas. Su orientación ha sido fundamental para mi desarrollo académico y personal.

Agradezco también al equipo del laboratorio, en particular a Jhon Intriago y Giovanni Castiglioni, cuyas habilidades y conocimientos fueron determinantes para mi continuo progreso.

Asimismo, deseo expresar mi agradecimiento al Instituto de Neurocirugía Dr. Alfonso Asenjo, con un reconocimiento especial a los doctores Cecilia Okuma y Enzo Brunetti.

Esta investigación cuenta con la colaboración del Instituto Milenio en Ingeniería e Inteligencia Artificial para la Salud, al cual agradezco por el acceso a su cluster, el cual desempeñó un papel crucial en mi avance y progreso. Su apoyo tecnológico ha sido esencial para el desarrollo exitoso de este proyecto. Expreso mi sincero agradecimiento al Fondecyt 1220829 por su generoso respaldo financiero a mi investigación. Su apoyo ha sido fundamental para llevar a cabo este estudio, permitiendo avances significativos que contribuirán al conocimiento científico y al bienestar de la sociedad.

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Identificación del problema	2
1.3. Objetivo General	2
1.4. Objetivo Específicos	2
2. Marco Teórico y Estado del Arte	3
2.1. Epilepsia	3
2.2. Electroencefalograma	4
2.2.1. Crisis epilépticas	5
2.2.2. División de la señal de EEG en diferentes etapas de la epilepsia	5
2.3. Resonancia Magnética Funcional	7
2.3.1. Señales BOLD	8
2.3.2. Organización de la corteza cerebral humana con FMRI	8
2.4. <i>Transformers</i>	10
2.5. Co-atención	12
2.6. Medidas de Desempeño Utilizadas	12
2.7. Estado del arte	14
2.8. Investigaciones importantes	15
2.8.1. Modelo Conformer	15
2.8.2. Predicción de Convulsiones basada en <i>Transformer</i> utilizando Electroencefalograma	16

2.8.3.	Atención Autodirigida Dual con Redes de Preguntas Dispersas para Responder Preguntas Visuales	18
2.8.4.	Fusión Multimodal con Redes de Co-Atención para la Detección de Noticias Falsas	19
2.8.5.	Atención alternante	20
3.	Algoritmo de predicción de crisis epilépticas usando EEG	22
3.1.	Base de datos	22
3.2.	Pre-procesamiento de datos	24
3.2.1.	Filtrado de señales	24
3.2.2.	Transformada corta de Fourier	25
3.3.	Modelos utilizados	25
3.3.1.	Modelo de Conformer espectral	26
3.3.2.	Modelo de Co-atención usando espectros de Fourier	27
3.3.3.	Modelo de Co-atención usando la serie temporal	30
3.4.	Resultados	32
3.4.1.	Resultados del modelo de Conformer espectral	32
3.4.2.	Resultados del modelo de Co-atención espectral	32
3.4.3.	Resultados del modelo de Co-atención temporal	33
3.4.4.	Análisis de resultados	34
4.	Algoritmo de predicción del foco epileptógeno usando EEG	38
4.1.	Base de datos	38
4.1.1.	Pre-procesamiento de datos	39
4.1.2.	Aumento de datos (Data augmentation)	39
4.2.	Modelos utilizados	39
4.2.1.	Modelo unimodal temporal	39
4.2.2.	Modelo unimodal canales	41
4.3.	Resultados	41

4.4.	Análisis de resultados	42
4.5.	Matriz de atención	42
4.5.1.	sub-007	43
4.5.2.	sub-010	43
4.5.3.	sub-011	45
5.	Algoritmo de clasificación de pacientes con epilepsia usando fMRI	49
5.1.	Base de datos	49
5.1.1.	Preprocesamiento de datos	49
5.2.	Modelos utilizados	51
5.2.1.	Modelo de Co-atención usando espectros de Fourier	51
5.2.2.	Co-atención usando la serie temporal para FMRI	52
5.3.	Resultados	53
5.3.1.	Resultados de modelo de co-atención espectral	53
5.3.2.	Resultados de modelo de co-atención temporal	55
5.4.	Análisis de resultados	55
6.	Algoritmo de predicción del foco epileptógeno usando FMRI	57
6.1.	Base de datos	58
6.1.1.	Pre-procesamiento de datos	58
6.1.2.	Aumento de datos (Data augmentation)	58
6.2.	Modelos utilizados	58
6.2.1.	Modelo unimodal temporal	59
6.2.2.	Modelo unimodal canales	59
6.3.	Resultados	59
6.4.	Análisis de resultados	61
6.5.	Matriz de atención: Búsqueda del foco epileptógeno	62
6.5.1.	sub-005	63

6.5.2.	sub-007	63
6.5.3.	sub-010	63
6.5.4.	sub-011	66
6.5.5.	sub-013	67
6.5.6.	sub-015	67
6.5.7.	sub-017	67
7.	Algoritmo de co-atención utilizando EEG y FMRI, para la predicción de crisis epilépticas	70
7.1.	Base de datos	70
7.1.1.	Pre-procesamiento de datos	70
7.1.2.	Aumento de datos (Data augmentation)	72
7.2.	Arquitectura del modelo de co-atención usando EEG y FMRI	72
7.3.	Resultados	73
7.4.	Análisis de Resultados	76
7.5.	Matriz de atención: Búsqueda del foco epiléptico	78
7.5.1.	sub-007	79
7.5.2.	sub-010	79
7.5.3.	sub-011	80
8.	Conclusiones	82
	Bibliografía	87
	Anexo	88

Índice de Tablas

2.1.	Convolución temporal y espacial. La primera capa utiliza k núcleos de tamaño $(1, 25)$ con un intervalo de $(1, 1)$ para realizar la convolución en la dimensión temporal. La segunda capa conserva k <i>kernels</i> de tamaño $(ch, 1)$ con un intervalo de $(1, 1)$, donde ch representa el número de canales del EEG. La tercera capa es un promedio <i>pooling</i> a lo largo de la dimensión temporal con un tamaño de núcleo de $(1, 75)$ y un paso de $(1, 15)$	16
3.1.	La base de datos EEG CHMBIT contiene información de género, edad y la cantidad de crisis en cada registro extraída de [1].	23
3.2.	Convolución Conformer Espectral	27
3.3.	Distribución de las ventanas de 5 segundos en función de los pacientes.	33
3.4.	Distribución de los ventanas de un segundo en función de los pacientes.	36
4.1.	Distribución de los pacientes del INCA en función del identificador, sexo, edad y cantidad de crisis.	38
4.2.	Distribución de ventanas ictales e interictales de los cinco pacientes del INCA.	41
4.3.	La cantidad de parámetros, tamaño del lote (batch), tasa de aprendizaje (LR) y número de capas (N) para los modelos unimodal temporal de EEG, unimodal de los canales de EEG y co-atención temporal de EEG.	41
4.4.	Resultados del algoritmo de co-atención temporal usando EEG	42
4.5.	Resultados unimodal temporal usando EEG	42
4.6.	Resultados unimodal de los canales usando EEG	42
5.1.	Descripción de pacientes con epilepsia refractaria base de datos FMRI de Iowa	50
5.2.	Distribución de los datos de entrenamiento, compuestos por 47 pacientes con condiciones de salud normales (HC) y 15 pacientes con epilepsia refractaria (DRE).	54

5.3.	Distribución de los datos de prueba, compuestos por 4 pacientes con condiciones de salud normales (HC) y 2 pacientes con epilepsia refractaria (DRE).	54
6.1.	Distribución de los datos de los pacientes del INCA con epilepsia refractaria.	57
6.2.	Distribución de los datos de los pacientes del Child Mind Institute con epilepsia refractaria.	58
6.3.	Distribución de ventanas para los modelos de co-atención temporal, unimodal del tiempo y unimodal de los canales.	60
6.4.	Cantidad de parámetros, tamaño del lote (batch), tasa de aprendizaje (LR) y número de capas (N) para los modelos unimodal temporal de FMRI, unimodal de los canales de FMRI y co-atención temporal de FMRI.	61
6.5.	Resultados modelo de co-atención temporal para los datos de FMRI.	61
6.6.	Resultados modelo de unimodal de los canales para los datos de FMRI.	61
6.7.	Resultados modelo unimodal del tiempo para los datos de FMRI.	62
7.1.	Distribución de los datos para el modelo de predicción de crisis epilépticas, utilizando EEG y FMRI. Pacientes del INCA con epilepsia refractaria.	71
7.2.	Distribución de los datos para el modelo de predicción de crisis epilépticas utilizando EEG y FMRI. Pacientes del Child Mind Institute sin epilepsia refractaria.	71
7.3.	Distribución de ventanas para los modelos de co-atención temporal de FMRI para 18 canales.	74
7.4.	Matrices de co-atención temporal aplicado a los datos FMRI de 18 canales.	74
7.5.	Resultados del modelo unimodal temporal aplicado a los datos FMRI de 18 canales.	74
7.6.	Resultados del modelo unimodal de los canales aplicado a los datos FMRI de 18 canales.	75
7.7.	La distribución de ventanas ictales e interictales tanto para los pacientes de control como en los del INCA.	75
7.8.	Matrices de desempeño de los pacientes del INCA mediante el algoritmo de co-atención utilizando EEG y FMRI.	76

Índice de Ilustraciones

2.1.	Arriba, se exhibe la representación visual de la epilepsia focal. Abajo, se presenta visualmente la epilepsia generalizada, extraída de [11].	4
2.2.	Izquierda, Imagen de una persona con electrodos colocados en el cuero cabelludo para la distribución de la actividad eléctrica cerebral. Derecha, Señales de la actividad eléctrica registradas por los electrodos. Extraída de [8]. . . .	4
2.3.	Momento de crisis epilépticas de un paciente del INCA: la sección blanca representa la etapa inmediatamente anterior a la crisis (preictal), mientras que la porción azul indica el inicio mismo de la crisis (ictal).	5
2.4.	Esquema que ilustra las crisis epilépticas: Inicialmente, se observa una señal de alarma, seguida por la entrada al período de Suspensión Preictal (SPH), que tiene una duración de 5 minutos. Posteriormente, se transita al período de Suspensión Ictogénica (SOP), marcando así el comienzo de la crisis epiléptica.	6
2.5.	Ilustración de las tres fases de la epilepsia en un EEG: ictal, interictal y preictal, junto con su distribución temporal. Extraída de [46].	6
2.6.	Máquina de resonancia magnética del Instituto Milenio en Ingeniería e Inteligencia Artificial para la Salud.	7
2.7.	El gráfico ilustra la respuesta hemodinámica a un estímulo. Durante los primeros 5 segundos después de la aplicación del estímulo, se observa un descenso en la actividad hemodinámica. Posteriormente, se experimenta un ascenso gradual que alcanza un pico máximo en algún punto, seguido por un descenso. Este patrón de respuesta refleja las variaciones en el flujo sanguíneo y la actividad hemodinámica asociada con la estimulación, mostrando una compleja interacción entre la respuesta fisiológica y el tiempo transcurrido desde el inicio del estímulo.	8
2.8.	Distribución de las principales zonas del fMRI según el modelo descrito en [40].	9
2.9.	En las dos ilustraciones presentadas, se visualiza la totalidad del cerebro, destacando la distribución de las 7 redes segmentadas según lo propuesto en [40].	9
2.10.	En las tres representaciones gráficas del cerebro, se exhibe de manera detallada la distribución de las 39 redes identificadas según el estudio [14].	10

2.11. Ejemplo del cálculo de pesos para una señal, donde se selecciona un punto específico en la señal y se procede a calcular el producto punto entre dicho punto y todos los demás puntos de la señal. Este proceso permite determinar el nivel de importancia relativa de ese punto en relación con los demás puntos de la señal.	12
2.12. Desarrollo de algoritmos Transformer específicos para EEG y FMRI, seguidos de su integración mediante co-atención. Este enfoque colaborativo potencia la representación cerebral al fusionar información temporal y espacial.	13
2.13. Matriz de confusión. En la esquina superior izquierda, están los Verdaderos Positivos (TP), en la esquina superior derecha los Falsos Positivos (FP), en la inferior izquierda los Falsos Negativos (FN) y en la inferior derecha los Verdaderos Negativos (TN).	14
2.14. Esquema que muestra las tres etapas fundamentales, convolución para la extracción de características temporales y espaciales, <i>transformer</i> para modelar relaciones complejas, y un clasificador para la tarea específica, extraída de [44].	16
2.15. Se utiliza la Transformada Corta de Fourier para derivar tres vectores. Estos vectores son sometidos a la operación de un <i>Transformer</i> y una red neuronal. Posteriormente, se implementa un mecanismo de puertas ponderadoras (<i>gating</i>) y un clasificador para el análisis final. Extraída de [20].	17
2.16. En la representación de la izquierda, observamos una autoatención donde únicamente se considera la información original sin la presencia de otro tipo de dato adicional. En cambio, en el esquema de la derecha, ilustramos la atención dirigida, donde se considera la información adicional Y [43].	19
2.17. Ilustra la co-atención multimodal entre video, audio y texto. Se han implementado 4 capas de co-atención, y cada capa consiste en dos bloques de atención dirigida en paralelo, cuyas salidas se concatenan. Finalmente, se aplica un clasificador para realizar la tarea específica, que es predecir noticias falsas [45].	20
2.18. Ilustración del proceso de atención alternante. En el primer bloque, la pregunta Q se resume en un vector único s mediante auto-atención. En el segundo bloque, se realiza una atención dirigida entre el vector s y la imagen V , obteniendo así el vector v . Finalmente, en el tercer bloque, se utiliza la característica de la imagen atendida v como guía para dirigir nuevamente la atención a la pregunta Q , resultando en el vector final q [34].	20
3.1. En esta memoria se utiliza 18 canales con distribución 20-10, abarcando regiones como Frontal (Fp1, Fp2), Central (C3, C4, Cz), Parietal (P3, P4, P7, P8), Occipital (O1, O2), y Temporal (T3, T4, T5, T6).	24
3.2. En la representación superior se visualiza una señal temporal con una duración de 5 segundos, mientras que en la representación inferior se exhibe el correspondiente espectrograma asociado a dicha señal.	26

3.3.	La arquitectura del modelo de Conformer espectral, organizada en tres etapas distintas. La primera etapa abarca la convolución, seguida por la segunda etapa que incorpora un transformador. La última etapa consiste en la capa de clasificación, encargada de predecir entre estados ictales e interictales.	27
3.4.	En esta representación gráfica se exhibe el Encoder utilizado, cuya primera etapa implica el mecanismo de <i>self-attention</i> . Este proceso permite que cada posición en la secuencia evalúe la relevancia de otras posiciones, generando así una representación contextualizada. Posteriormente, se incorpora una capa <i>feed-forward</i> para introducir complejidad no lineal a las representaciones. La normalización por capas se aplica después de cada subcapa con el propósito de mantener la estabilidad en la distribución de activaciones. Además, las conexiones residuales, que suman la entrada original a la salida de cada subcapa, facilitan el flujo de información y gradiente, mejorando significativamente la estabilidad y la capacidad de aprendizaje del modelo.	28
3.5.	Ilustración del modelo de co-atención espectral que se compone de dos partes fundamentales. En la primera fase, los tres vectores, que encapsulan la información temporal, de los canales y de la frecuencia, son sometidos individualmente a un proceso de codificación mediante un <i>Encoder</i> y una proyección lineal. En la segunda etapa, se implementan cuatro capas de atención dirigida, que involucran interacciones específicas entre las dimensiones de tiempo, canales y frecuencia. Finalmente, el modelo concluye con una capa de clasificación diseñada para discernir entre ventanas ictales e intraictales.	29
3.6.	Diagrama del modelo de co-atención temporal, el cual se compone de dos etapas fundamentales. En la primera fase, se aplica un <i>Encoder</i> tanto al vector temporal como al vector de canales. Posteriormente, se lleva a cabo una co-atención entre ambas representaciones, permitiendo la obtención de una relación espacio-temporal. Finalmente, se emplea una capa de clasificación. .	31
3.7.	Puntuación F1 del modelo Conformer espectral en función de los pacientes 1, 2 y 3	34
3.8.	Puntuación F1 del modelo Co-atención espectral en función de los pacientes 1, 2 y 3	35
3.9.	Matrices de confusión del algoritmo co-atención espectral para los pacientes 1,2 y 3.	35
3.10.	Puntuación F1 del modelo Co-atención temporal en función de los pacientes 1, 2 y 3	35
3.11.	Matrices de confusión del algoritmo co-atención temporal	36
3.12.	Resultados de <i>f1-score</i> , en función de los pacientes 1,2 y 3 para los tres algoritmos: Conformer espectral, Co-atención espectral y Co-atención temporal. .	37

4.1.	El modelo unimodal temporal se compone de un <i>Encoder</i> seguido por una capa de clasificación. Su objetivo es discernir entre ventanas que contienen crisis y aquellas que son normales.	40
4.2.	El modelo unimodal de los canales se compone de un <i>Encoder</i> seguido por una capa de clasificación. Su objetivo es discernir entre ventanas que contienen crisis y aquellas que son normales.	40
4.3.	Los resultados de <i>f1-score</i> para los tres modelos: co-atención temporal utilizando EEG, unimodal temporal y unimodal de los canales. Estos resultados están representados en función de los cinco pacientes.	43
4.4.	La parte superior izquierda de la figura representa la ubicación donde el paciente sub-007 experimenta la crisis, mientras que la parte inferior ilustra la distribución espacial de la atención durante los primeros 6 segundos.	44
4.5.	Promedio espacial de la atención durante los primeros 6 segundos para el paciente sub-007, centralizada en la región fronto-central izquierda.	44
4.6.	La parte superior izquierda de la figura representa la ubicación donde el paciente sub-010 experimenta la crisis, mientras que la parte inferior ilustra la distribución espacial de la atención durante los primeros 6 segundos.	45
4.7.	Promedio espacial de la atención durante los primeros 6 segundos para el paciente sub-010.	46
4.8.	Promedio espacial de la atención durante los primeros 6 segundos para el paciente sub-011.	46
4.9.	La parte superior izquierda de la figura representa la ubicación donde el paciente sub-011 experimenta la crisis, mientras que la parte inferior ilustra la distribución espacial de la atención durante los primeros 6 segundos.	47
5.1.	Curva de loss del algoritmo de co-atención espectral con FMRI	55
5.2.	Curva de <i>loss</i> del algoritmo de co-atención temporal con FMRI	56
6.1.	El diagrama representa el modelo unimodal temporal, el cual consta de un encoder basado en la arquitectura <i>transformer</i> y una capa de clasificación.	59
6.2.	El diagrama representa el modelo unimodal canales, el cual al igual que el modelo unimodal temporal posee un <i>transformer</i> y una capa de clasificación.	60
6.3.	Resultados de <i>f1-score</i> para los tres modelos, co-atención temporal, unimodal del tiempo y unimodal de los canales, en función de los pacientes	62
6.4.	En la figura a) se visualiza la ubicación de la lesión en el paciente sub-005, mientras que en b) se presenta la región a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.	63

6.5.	En la figura a) se visualiza la ubicación de la lesión en el paciente sub-007, mientras que en b) se presenta la región a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.	64
6.6.	En la figura a) se visualiza la ubicación de la lesión en el paciente sub-010, mientras que en b) se presenta la región a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.	65
6.7.	En la figura a) se visualiza la ubicación de la lesión en el paciente sub-011, mientras que en b) se presenta la región a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.	66
6.8.	En la figura a) se visualiza la ubicación de la lesión en el paciente sub-013, mientras en b) se presentan las regiones a las que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.	68
6.9.	En la figura de arriba se visualiza la ubicación de la lesión en el paciente sub-015, mientras que en las figuras de abajo presentan las regiones a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.	69
6.10.	En la figura de arriba se visualiza la ubicación de la lesión en el paciente sub-017, mientras que en las figuras de abajo presentan las regiones a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.	69
7.1.	La representación gráfica de los dos <i>encoders</i> correspondientes a cada modalidad, junto con la separación de las dos etapas en el modelo de co-atención alternada. La secuencia concluye con la incorporación de una capa de clasificación.	73
7.2.	Curva de pérdida del modelo de co-atención usando EEG y FMRI, para el paciente sub-011.	77
7.3.	Métricas de <i>f1-score</i> para los modelos unimodales de tiempo, canales y de co-atención temporal empleando FMRI con 18 canales.	77
7.4.	Métricas de <i>f1-score</i> para los modelos unimodales de tiempo y canales de EEG, así como para los modelos de co-atención EEG y co-atención EEG junto con FMRI.	78
7.5.	Representación visual de la matriz de atención del paciente sub-007, la que revela los puntos de interés identificados por el modelo de co-atención de EEG y FMRI. En la representación, los puntos destacados en color naranja indican las áreas en las que el modelo enfoca su atención conjunta. La zona verde corresponde al diagnóstico obtenido a través de EEG, mientras que la región resaltada en rojo refleja el diagnóstico proporcionado por FMRI.	79

- 7.6. Representación visual de la matriz de atención del paciente sub-010, la que revela los puntos de interés identificados por el modelo de co-atención de EEG y FMRI. En la representación, los puntos destacados en color naranja indican las áreas en las que el modelo enfoca su atención conjunta. La zona verde corresponde al diagnóstico obtenido a través de EEG, mientras que la región resaltada en rojo refleja el diagnóstico proporcionado por FMRI. 80

- 7.7. Representación visual de la matriz de atención del paciente sub-011, la que revela los puntos de interés identificados por el modelo de co-atención de EEG y FMRI. En la representación, los puntos destacados en color naranja indican las áreas en las que el modelo enfoca su atención conjunta. La zona verde corresponde al diagnóstico obtenido a través de EEG, mientras que la región resaltada en rojo refleja el diagnóstico proporcionado por FMRI. 81

Capítulo 1

Introducción

1.1. Motivación

La aplicación de inteligencia artificial (IA) en medicina ha revolucionado la atención médica. Desde la detección temprana de cáncer de piel mediante redes neuronales [13] hasta la predicción precisa de enfermedades cardíacas con aprendizaje automático [21], la IA ha mejorado significativamente diagnósticos y tratamientos. Además, la personalización de terapias basadas en análisis genético [22] y la gestión eficiente de epidemias, como la predicción de la propagación de COVID-19 [39], resaltan la impactante convergencia entre la tecnología y la medicina, ofreciendo avances transformadores para el cuidado de la salud.

Entre estos avances se encuentra la epilepsia, una enfermedad neurológica crónica que afecta a millones de personas en todo el mundo, y su diagnóstico y tratamiento siguen siendo desafíos clínicos significativos. En el marco de estos avances, es notable destacar la predicción de crisis epilépticas, empleando tanto el aprendizaje de máquinas como las redes neuronales, según lo evidencian varios estudios [37, 27, 30, 5, 18, 36].

En el ámbito de la epilepsia, la identificación precisa del foco epileptogénico es crucial, es decir, la región del cerebro donde se origina la actividad eléctrica anómala que provoca las crisis epilépticas. La co-localización de esta alteración estructural es esencial para determinar las opciones más adecuadas y brindar una atención personalizada a los pacientes [33]. No obstante, la detección precisa del foco epiléptico requiere dos evaluaciones distintas. En primer lugar, se realiza un electroencefalograma (EEG) para medir la actividad eléctrica de la corteza cerebral [16]. Seguidamente, se lleva a cabo una resonancia magnética funcional en reposo (fMRI) para analizar las variaciones en el flujo sanguíneo cerebral [9]. La combinación y análisis conjunto de estos resultados resulta esencial para obtener un diagnóstico preciso del foco epiléptico. Hasta la fecha, no se ha desarrollado un algoritmo de inteligencia artificial que integre ambas modalidades para la predicción del foco epileptogénico.

1.2. Identificación del problema

Dado que la identificación del foco epileptogénico requiere el análisis conjunto de dos exámenes, EEG y fMRI, surge la necesidad de implementar un modelo multimodal. En este contexto, la presente memoria de título se enfoca en el desarrollo y aplicación de un algoritmo de inteligencia artificial supervisado denominado co-atención. Este algoritmo, basado en *transformers*, como se describen en [4], destaca por su capacidad para procesar y comprender patrones complejos en datos secuenciales, siendo esencial para enfoques multimodales. El algoritmo de co-atención, descrito en [43], se basa en una red neuronal capaz de procesar dos tipos de datos distintos, en este caso, electroencefalograma (EEG) y resonancia magnética funcional (fMRI). El propósito principal de esta implementación es mejorar la precisión en la co-localización del foco epileptogénico, aprovechando la complementariedad de la información proporcionada por ambas modalidades.

1.3. Objetivo General

El objetivo de este trabajo es fusionar datos del EEG y fMRI mediante el algoritmo de co-atención para combinar la información de ambas fuentes y aumentar la precisión de la identificación del foco epileptogénico en pacientes con epilepsia refractaria.

1.4. Objetivo Específicos

- Desarrollar, evaluar y comparar algoritmos de *transformer* para EEG y fMRI en la predicción de crisis epilépticas o clasificación de pacientes con epilepsia refractaria. Se empleará la base de datos recopilada en el Children's Hospital Boston para EEG [1] y la base de datos referida en [41] para fMRI.
- Desarrollar, evaluar y comparar algoritmos de EEG y fMRI para co-localizar el foco epileptogénico, utilizando pacientes del INCA.
- Desarrollar, evaluar y comparar la co-atención entre EEG y FMRI para la co-localización del foco epileptogénico, utilizando pacientes del INCA.

Capítulo 2

Marco Teórico y Estado del Arte

En este capítulo se describen los siguientes conceptos:

- Epilepsia
- Electroencefalograma
- Crisis Epilépticas
- *Transformer*
- Co-atención

2.1. Epilepsia

La epilepsia es un trastorno cerebral en el cual una persona tiene convulsiones repetidas durante un tiempo. Las convulsiones son episodios de actividad descontrolada y anormal de las neuronas que pueden causar cambios en la atención o el comportamiento como movimientos corporales [31]. Las convulsiones epilépticas se pueden presentar a cualquier edad. Sin embargo, son más comunes en niños pequeños y adultos mayores de 65 años. Las crisis se clasifican en dos tipos [10]:

- Las convulsiones generalizadas que afectan a ambos lados del cerebro.
- Las convulsiones focales que se localizan en una sola área del cerebro. Estas convulsiones también se conocen como convulsiones parciales.

En la figura 2.1 se observa la distinción en los patrones del EEG entre la epilepsia focal y la generalizada. El enfoque combinado del EEG-FMRI proporciona una perspectiva multidimensional para localizar el foco epiléptico. Esta metodología resulta especialmente valiosa en pacientes con epilepsia refractaria, aquellos que no responden a tratamientos farmacológicos convencionales [23].

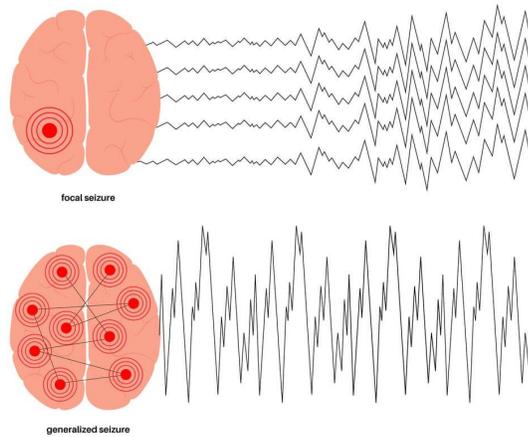


Figura 2.1: Arriba, se exhibe la representación visual de la epilepsia focal. Abajo, se presenta visualmente la epilepsia generalizada, extraída de [11].

2.2. Electroencefalograma

El Electroencefalograma (EEG) es un registro en el que se pueden detectar anomalías en las ondas cerebrales, o actividad eléctrica del encéfalo [16]. Los canales del EEG, tal como se muestra en la figura 2.2, se refieren a las ubicaciones específicas en el cuero cabelludo donde se colocan los electrodos para captar la actividad eléctrica del cerebro. Cada canal representa una derivación que registra la diferencia de potencial eléctrico entre dos electrodos o entre un electrodo y una referencia común. La disposición y el número de canales en un EEG pueden variar, pero su elección depende de los objetivos del registro y de la información deseada.

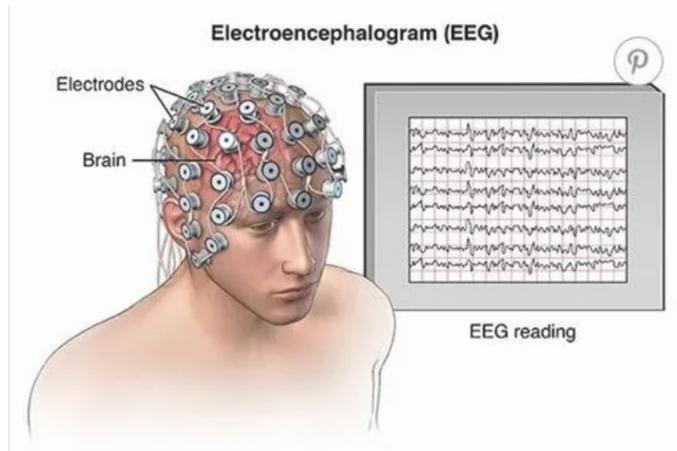


Figura 2.2: Izquierda, Imagen de una persona con electrodos colocados en el cuero cabelludo para la distribución de la actividad eléctrica cerebral. Derecha, Señales de la actividad eléctrica registradas por los electrodos. Extraída de [8].

2.2.1. Crisis epilépticas

Según la Fundación Internacional de la Epilepsia [12], una crisis epiléptica consiste en ráfagas súbitas y temporales de actividad eléctrica en el cerebro que alteran o interrumpen la forma en que se envían los mensajes entre las células cerebrales. Además, es importante tener en cuenta que las crisis en sí mismas no constituyen una enfermedad; más bien, es el conjunto de estas crisis lo que conforma la epilepsia. En la figura 2.3 se muestra una crisis generalizada de un paciente con epilepsia. La sección azul representa la fase de la crisis, mientras que la sección blanca indica la etapa preictal.

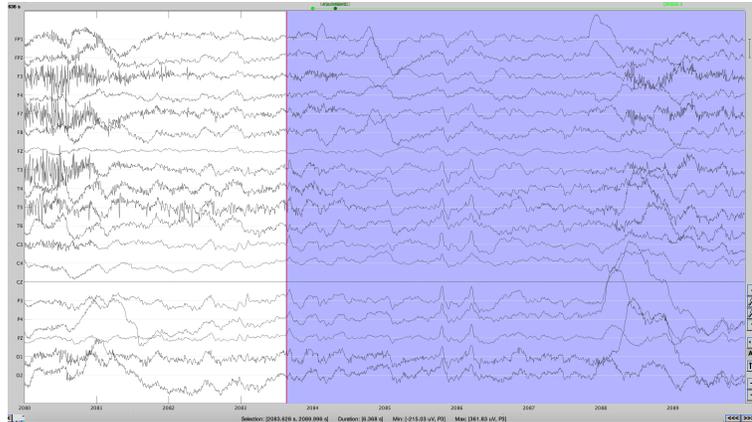


Figura 2.3: Momento de crisis epilépticas de un paciente del INCA: la sección blanca representa la etapa inmediatamente anterior a la crisis (preictal), mientras que la porción azul indica el inicio mismo de la crisis (ictal).

2.2.2. División de la señal de EEG en diferentes etapas de la epilepsia

En diversas investigaciones sobre la predicción de crisis epilépticas [32], se emplearon las siguientes definiciones que también serán utilizadas en esta memoria.

- **Período de ocurrencia de convulsiones (SOP)** se define como el período de tiempo durante el cual se realiza la predicción de las crisis, tiene una duración de 30 minutos.
- **Horizonte de predicción de convulsiones (SPH)** se refiere al período de tiempo desde la activación de la alarma hasta el inicio del propio SOP, tiene una duración de 3-5 minutos.

En la figura 2.4, se observa una conexión entre la fase SPH (estado de alarma) y la fase SOP durante una crisis epiléptica. Se puede afirmar que cuando se activa una señal de alarma y ésta persiste durante un tiempo superior a 5 minutos, se entra en el período SOP, lo que indica la posibilidad de que ocurra una crisis epiléptica. A partir de las definiciones previamente establecidas, se derivan los siguientes términos [46].

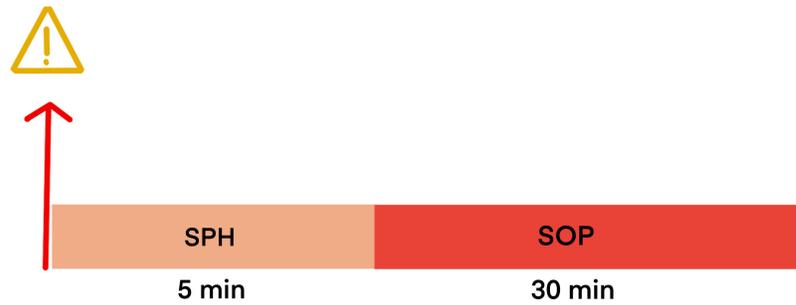


Figura 2.4: Esquema que ilustra las crisis epilépticas: Inicialmente, se observa una señal de alarma, seguida por la entrada al período de Suspensión Preictal (SPH), que tiene una duración de 5 minutos. Posteriormente, se transita al período de Suspensión Ictogénica (SOP), marcando así el comienzo de la crisis epiléptica.

- **Interictal:** Este período se refiere al tiempo entre convulsiones en la epilepsia, es decir, cuando la persona no está experimentando una crisis epiléptica. Durante este tiempo, la persona se encuentra en un estado de normalidad y no experimenta los síntomas característicos de una convulsión. Es un período de relativa estabilidad en el que la actividad cerebral es típica de una persona sin epilepsia.
- **Preictal:** Este término se utiliza para describir el tiempo justo antes de que ocurra una convulsión en personas con epilepsia. Durante este período, algunas personas pueden experimentar signos o síntomas que actúan como una especie de advertencia de que una crisis está por comenzar. Estos síntomas preictales pueden variar de una persona a otra y pueden incluir sensaciones extrañas, cambios en el estado de ánimo o incluso síntomas físicos como temblores o mareos.
- **Ictal:** Este es el estado durante el cual ocurre una convulsión o episodio epiléptico. Durante este período, la persona experimenta los síntomas típicos de una crisis epiléptica, que pueden incluir movimientos involuntarios, pérdida de conocimiento, convulsiones y otros síntomas neurológicos.

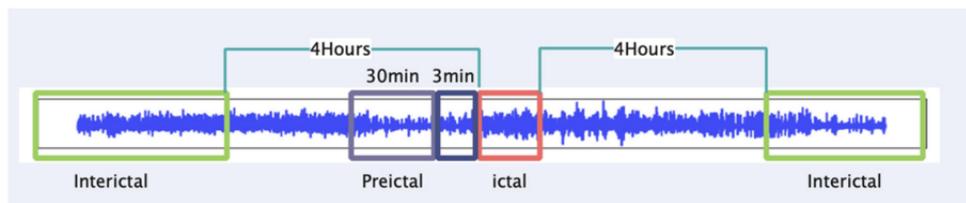


Figura 2.5: Ilustración de las tres fases de la epilepsia en un EEG: ictal, interictal y pre-ictal, junto con su distribución temporal. Extraída de [46].

2.3. Resonancia Magnética Funcional

La Resonancia Magnética Funcional (fMRI) mide la actividad cerebral detectando cambios en el flujo sanguíneo y los niveles de oxigenación en respuesta a la actividad neural [9]. Es una técnica no invasiva que utiliza los mismos principios básicos que la resonancia magnética estándar, pero se enfoca en la dinámica de la función cerebral en lugar de solo en su estructura. Esto permite a los investigadores observar qué partes del cerebro están activas durante tareas específicas o bajo ciertas condiciones. Los datos obtenidos de los escaneos fMRI proporcionan una valiosa información sobre la función cerebral y se utilizan en diversos campos, como la neurociencia, la psicología y la medicina. Una resonancia magnética estándar utiliza un imán, ondas de radio y procesamiento informático para generar imágenes tridimensionales altamente detalladas del interior del cuerpo.

La fMRI funciona de manera muy similar a una resonancia magnética estándar. Durante el procedimiento, el paciente se acuesta en una mesa que puede mover su cuerpo hacia la máquina de resonancia magnética. La mayoría de las veces, la fMRI utiliza un escáner de resonancia magnética en forma de anillo porque las máquinas de resonancia magnética abiertas generalmente no tienen un imán lo suficientemente potente, ver figura 2.6.



Figura 2.6: Máquina de resonancia magnética del Instituto Milenio en Ingeniería e Inteligencia Artificial para la Salud.

Durante una fMRI, la cabeza del paciente estará dentro de un dispositivo especial en forma de casco llamado bobina para la cabeza. La bobina ayuda a producir imágenes más claras. También llevará auriculares especiales que cancelan el sonido proveniente de la máquina de resonancia magnética. Estos auriculares permiten que los proveedores de atención médica, como tecnólogos de resonancia magnética o radiólogos, hablen durante el escaneo y le den instrucciones. Los profesionales médicos pueden solicitar al paciente que realice tareas específicas, tales como hablar, tocar un dedo u otras acciones simples. Además, existe una variante llamada resting fMRI o fMRI en reposo. En este caso, el paciente no realiza tareas específicas, sino que simplemente descansa mientras se toman las imágenes. En el contexto de la epilepsia, la fMRI en reposo es particularmente valiosa. Esto ayuda a los médicos a identificar las áreas del cerebro que pueden tener actividad anormal incluso en ausencia de tareas específicas. Este enfoque es útil para mapear la actividad cerebral basal y detectar patrones anómalos que podrían estar relacionados con la epilepsia.

2.3.1. Señales BOLD

La resonancia magnética funcional es la única forma de ver secuencias de imágenes sobre cómo funciona la mente humana. Dependiente del nivel de oxígeno en sangre (BOLD, blood-oxygen-level-dependent) es la señal que se detecta en la fMRI. Esta refleja desoxihemoglobina impulsada por cambios localizados en el flujo sanguíneo cerebral y la oxigenación de la sangre, que se acoplan a la actividad neuronal subyacente mediante un proceso denominado acoplamiento neurovascular [17]. El gráfico de la señal BOLD muestra la respuesta hemodinámica de una región cerebral activada, ver figura 2.7 . La señal BOLD se representa generalmente como una curva en la que el eje horizontal representa el tiempo y el eje vertical representa la amplitud o intensidad de la señal BOLD. Cuando una tarea se realiza o se produce un estímulo, se observa un aumento inicial en la señal BOLD en la región cerebral correspondiente. Esta señal alcanza su punto máximo y luego disminuye a medida que el cerebro regresa a su estado basal.

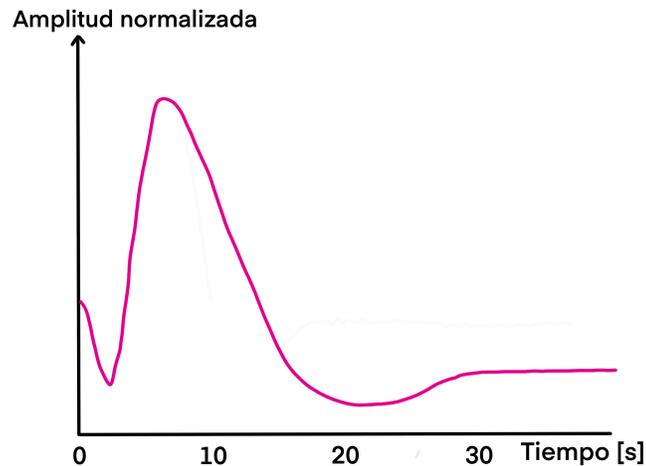


Figura 2.7: El gráfico ilustra la respuesta hemodinámica a un estímulo. Durante los primeros 5 segundos después de la aplicación del estímulo, se observa un descenso en la actividad hemodinámica. Posteriormente, se experimenta un ascenso gradual que alcanza un pico máximo en algún punto, seguido por un descenso. Este patrón de respuesta refleja las variaciones en el flujo sanguíneo y la actividad hemodinámica asociada con la estimulación, mostrando una compleja interacción entre la respuesta fisiológica y el tiempo transcurrido desde el inicio del estímulo.

2.3.2. Organización de la corteza cerebral humana con fMRI

En el marco de esta investigación, se han empleado tres atlas para la organización de la corteza cerebral. La característica clave que los distingue es la variación en la cantidad de subdivisiones utilizadas para delinear la estructura de la corteza cerebral.

1. Organización de la corteza cerebral humana estimada por la conectividad funcional intrínseca.

Este atlas [40], se construyó a partir de datos recopilados de 1000 sujetos. Estos datos fueron registrados mediante un proceso de alineamiento a una superficie esférica, permitiendo así una representación tridimensional de la corteza cerebral. Una vez obtenida esta superficie esférica, se aplicó un método de clusterización para segmentar y definir las distintas regiones de la corteza cerebral. Los resultados de esta investigación han llevado a la conclusión de que la corteza cerebral puede ser segmentada en siete redes principales, y cada una de estas se descompone en áreas más pequeñas, totalizando así 100 partes. En la figuras 2.8 y 2.9, se muestran las 7 redes mencionadas con su distribución espacial en la corteza cerebral.

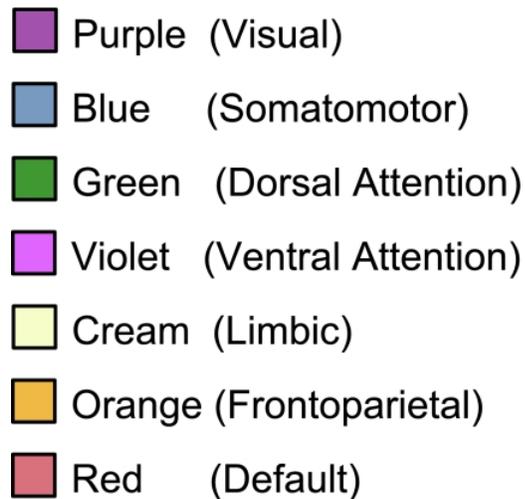


Figura 2.8: Distribución de las principales zonas del fMRI según el modelo descrito en [40].

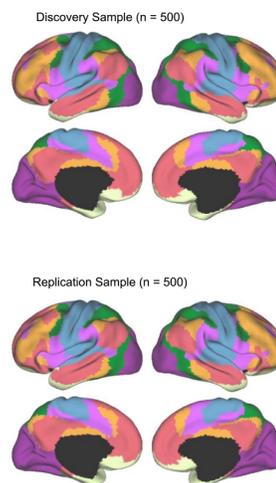


Figura 2.9: En las dos ilustraciones presentadas, se visualiza la totalidad del cerebro, destacando la distribución de las 7 redes segmentadas según lo propuesto en [40].

2. Aprendizaje de diccionarios multi-sujeto para segmentar un atlas de la actividad cerebral espontánea.

Este atlas [14] es un modelo probabilístico jerárquico para la organización funcional intrínseca del cerebro mediante las fluctuaciones continuas en su actividad. Este enfoque, basado en el aprendizaje de diccionarios, para abordar eficientemente este desafío, utiliza un algoritmo basado en técnicas de optimización convexa con penalizaciones no suaves, diseñado especialmente para la reducción de ruido en imágenes. Los resultados obtenidos en datos simulados destacan la capacidad del modelo para recuperar mapas a nivel poblacional y capturar especificidades individuales. Al aplicar este enfoque a datos de resonancia magnética funcional en estado de reposo, consiguieron extraer el primer atlas de la actividad cerebral espontánea. Este atlas revela una parcelación funcional específica del sujeto en regiones localizadas, proporcionando una valiosa comprensión de la variabilidad individual en la organización funcional del cerebro, identificando un total de 39 zonas cerebrales distintas.

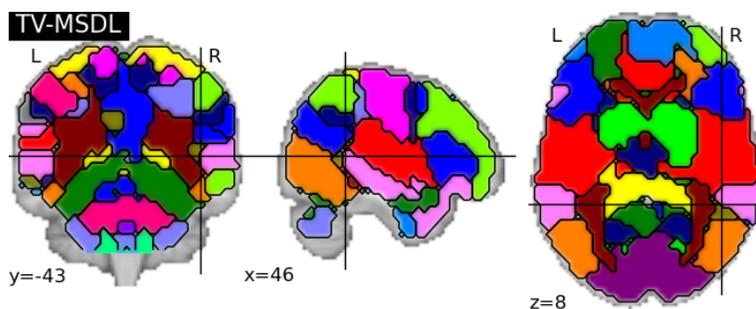


Figura 2.10: En las tres representaciones gráficas del cerebro, se exhibe de manera detallada la distribución de las 39 redes identificadas según el estudio [14].

3. Atlas fMRI-EEG

Considerando la naturaleza multimodal de la investigación, diseñamos un atlas de fMRI que integra los 18 canales EEG específicos utilizados en este estudio, como se muestra en la figura 3.1. La creación de estos canales se realizó mediante una máscara que posee la ubicación de cada uno, y esta máscara se aplicó para calcular el promedio de la señal fMRI en áreas correspondientes. Este atlas juega un papel crucial en la coherencia de la investigación, ya que posibilita el análisis conjunto de las mismas zonas en fMRI y EEG, contribuyendo así a una comprensión más integral de la actividad cerebral relacionada con el foco epiléptico.

2.4. *Transformers*

En esta memoria de título, se utiliza un algoritmo de tipo supervisado para hacer la predicción. En particular, se usa un tipo de red neuronal llamada *Transformer* [4]. El *Transformer* o modelo de auto-atención [4] es un modelo de aprendizaje profundo que ha revolucionado el campo de la inteligencia artificial, en particular en el procesamiento del lenguaje natural. Su innovación clave radica en la utilización de la auto-atención, un mecanismo que permite

al modelo asignar diferentes niveles de importancia a cada parte de los datos de entrada, incluyendo la salida recursiva. En lugar de depender de estructuras de secuencia recurrentes, como las redes neuronales recurrentes (RNN), el *Transformer* utiliza mecanismos de atención para capturar las relaciones entre las diferentes partes de la secuencia de entrada.

La auto-atención en el *Transformer* es fundamental para su funcionamiento. Este mecanismo permite al modelo procesar y contextualizar eficientemente la información en función de su importancia. Cada elemento de la secuencia se pondera en función de su importancia en relación con otros elementos, lo que permite al *Transformer* enfocarse en las partes más relevantes y obtener una representación más precisa de los datos de entrada.

En la estructura del algoritmo se utilizan las *Query*, *Keys* y *Values*, en español Consultas, Claves y Valores. Estas se definen a continuación:

- **Consultas (Q):** Es una matriz donde cada fila representa una consulta para cada posición en la secuencia de entrada.
- **Claves (K):** Es una matriz donde cada fila representa una clave asociada a cada posición en la secuencia de entrada.
- **Valores (V):** Es una matriz V donde cada fila representa un valor asociado a cada posición en la secuencia de entrada.

Para poder calcular la atención se realiza el producto punto entre las consultas y las claves, seguido de una normalización. Se aplica, una función *softmax* para obtener los pesos de atención normalizados. Finalmente, esta matriz llamada pesos de atención se pondera mediante una combinación lineal con los valores, obteniendo la salida de la capa de atención. La fórmula corresponde a la siguiente:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V, \quad (2.1)$$

donde:

- Q , K , y V son matrices que representan las consultas, las claves y los valores respectivamente.
- d_k es la dimensión de las claves.
- La operación QK^T calcula la similitud entre las consultas y las claves.
- El resultado se normaliza utilizando la función *softmax* para obtener pesos de atención.

En la figura 2.11, se muestra un ejemplo del cálculo de la matriz de pesos de atención con una señal de EEG. En este proceso, se analiza un punto específico de la señal EEG y se calcula el producto punto que indica la importancia relativa de cada otro punto de la señal en relación con ese punto en particular.

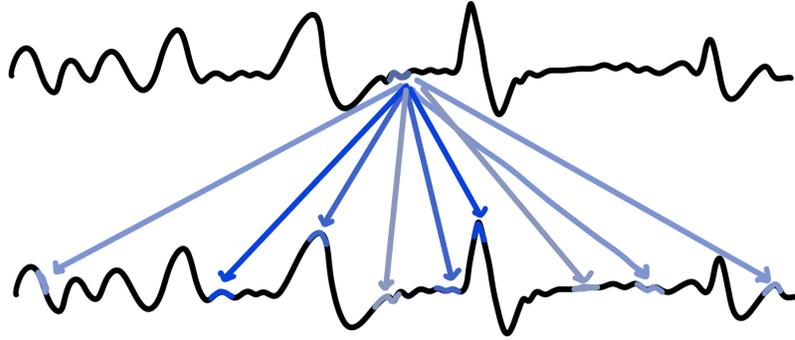


Figura 2.11: Ejemplo del cálculo de pesos para una señal, donde se selecciona un punto específico en la señal y se procede a calcular el producto punto entre dicho punto y todos los demás puntos de la señal. Este proceso permite determinar el nivel de importancia relativa de ese punto en relación con los demás puntos de la señal.

2.5. Co-atención

Un modelo de co-atención [43], también conocido como *co-attention model* en inglés, es un tipo de arquitectura de red neuronal, de un *transformer* o una auto-atención, diseñado para combinar y procesar dos conjuntos de datos diferentes, como por ejemplo secuencias de texto o imágenes, de manera conjunta y contextual. El propósito principal de un modelo de co-atención es capturar las relaciones y las interacciones entre los dos conjuntos de datos para obtener una representación más completa de la información. Esta técnica se utiliza en una variedad de aplicaciones, como el procesamiento del lenguaje natural y la visión por computadora [43] [26] [15] [42] [3] [35] [38] [6].

En un modelo de co-atención, la primera etapa es la creación de la matriz de pesos de atención. Aquí, cada conjunto (o *token*) en la secuencia de entrada tiene asociada una ponderación que determina su importancia en relación con otros *tokens*. Esto se logra mediante diversas técnicas, como el cálculo de similitudes o afinidades entre *tokens*. Una vez que se obtienen los pesos de atención para cada conjunto, se procede a combinar la información de acuerdo con estos pesos. Una manera común de hacerlo es mediante la operación de producto punto entre los pesos de atención de cada dato. Sin embargo, existen variantes en la combinación de los pesos de atención de los conjuntos. Algunas técnicas incluyen el uso de redes neuronales para aprender combinaciones más complejas o el uso de funciones de atención más elaboradas, como la atención ponderada o la atención múltiple [43]. En resumen, un modelo de co-atención es una arquitectura de red neuronal que permite combinar y procesar dos o más tipos de datos diferentes de manera conjunta y contextual. La arquitectura de co-atención propuesta para esta memoria de título se ilustra en la figura 2.12.

2.6. Medidas de Desempeño Utilizadas

- **Matriz de confusión** es una matriz que compara las predicciones de un modelo con las clases reales de un conjunto de datos de prueba. En la matriz, las filas representan las clases reales, mientras que las columnas representan las clases predichas por el modelo.

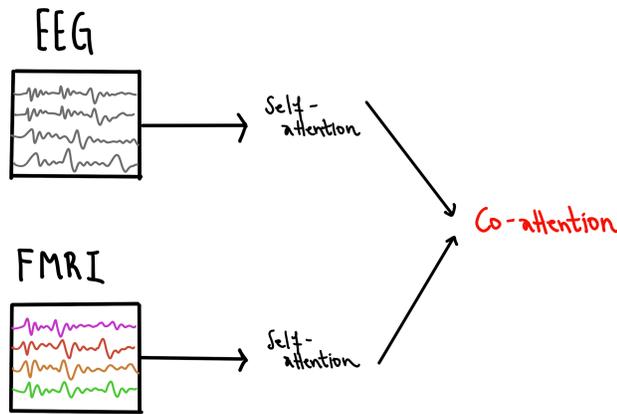


Figura 2.12: Desarrollo de algoritmos Transformer específicos para EEG y FMRI, seguidos de su integración mediante co-atención. Este enfoque colaborativo potencia la representación cerebral al fusionar información temporal y espacial.

La matriz de confusión posee estos 4 elementos:

- **Verdaderos Positivos (TP)**: Instancias correctamente clasificadas como positivas.
- **Verdaderos Negativos (TN)**: Instancias correctamente clasificadas como negativas.
- **Falsos Positivos (FP)**: Instancias incorrectamente clasificadas como positivas.
- **Falsos Negativos (FN)**: Instancias incorrectamente clasificadas como negativas.

La distribución de las celdas de la matriz de confusión se pueden observar en la figura 2.13.

- **Accuracy** se define como la proporción de predicciones correctas sobre el total de predicciones realizadas. Se expresa como:

$$\text{Accuracy} = \frac{\text{Verdaderos positivos} + \text{Verdaderos negativos}}{\text{Total de muestras}} \quad (2.2)$$

- **Recall** mide la proporción de positivos reales que fueron identificados correctamente por el modelo. Se calcula como:

$$\text{Recall} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}} \quad (2.3)$$

- **Precision** mide la proporción de predicciones positivas hechas por el modelo que son realmente positivas. Se calcula como:

$$\text{Precision} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}} \quad (2.4)$$

- **F1-score** es una medida que combina la precisión y el recall en un solo número. Se calcula como la media armónica de precisión y recall:

$$F1 = 2 \times \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}} \quad (2.5)$$

		Predicho	
		Actual	
Actual	Actual	TP	FP
	Actual	FN	TN

Figura 2.13: Matriz de confusión. En la esquina superior izquierda, están los Verdaderos Positivos (TP), en la esquina superior derecha los Falsos Positivos (FP), en la inferior izquierda los Falsos Negativos (FN) y en la inferior derecha los Verdaderos Negativos (TN).

2.7. Estado del arte

Actualmente, existen numerosas investigaciones de predicción de crisis epilépticas en la literatura, la que varía entre las etapas de la epilepsia. Nos enfocaremos en mencionar aquellas que han utilizado la base de datos recopilada en el Children's Hospital Boston [1] y que usen técnicas de *machine learning* o *deep learning*. Estos estudios son:

- **Predicción de crisis epilépticas, usando métodos de aprendizaje de máquinas (2017)** [37]. Predice entre lo ictal y interictal usando técnicas de *machine learning*, obteniendo resultados de 92 % de sensibilidad.
- **Predicción de crisis epilépticas, usando métodos de aprendizaje de máquinas y *deep learning* (2022)**, [27] Predice entre lo ictal y preictal, obteniendo resultados de 84 % de *recall*.
- **Predicción de crisis epilépticas, usando encefalograma (2021)**, [30] Predice entre lo ictal y preictal utilizando redes neuronales convolucionales, obteniendo resultados de 93 % de *recall*.
- **Predicción de crisis epilépticas, usando técnicas de *deep learning* (2020)**, [5] Predice entre lo ictal y preictal utilizando redes convolucionales, usando para la clasificación *support vector machine*, obteniendo resultados de 92.7 % de *recall*.
- **Predicción de crisis epilépticas, basado en la redes convolucionales de multiresolución (2023)**, [18] Predice entre lo ictal y preictal utilizando redes neuronales convolucionales de multiresolución, obteniendo resultados de 82 % de *recall*.
- **EEG basado en la predicción de crisis epilépticas usando transformadores temporales multicanal (2022)**, [36] Predice entre lo ictal y preictal utilizando *transformers*, obteniendo resultados de 96.46 % de *recall*.

En 2022, se publicó una investigación [25] que utiliza resonancia magnética funcional (fMRI) para predecir el foco epiléptico en pacientes con epilepsia del lóbulo temporal. Mediante redes neuronales, el modelo clasificó correctamente el hemisferio de la zona de inicio de la convulsión con una precisión del 90.6 %. La muestra del estudio comprendió 2.132 sujetos sanos como grupo de control y 32 pacientes preoperatorios con epilepsia del lóbulo temporal.

En la multimodalidad, se han realizado estudios matemáticos [28] y médicos [19] que exploran el análisis conjunto del uso de resonancia magnética funcional (fMRI) y electroencefalografía (EEG). Estos estudios buscan integrar y aprovechar la información proporcionada por ambas modalidades para obtener una comprensión más completa y detallada de la actividad cerebral.

Actualmente, detectar este foco requiere dos especialistas: uno en la interpretación de encefalogramas y otro en el análisis de imágenes de fMRI. Por lo tanto, esta herramienta integrada sería de gran ayuda para los neurólogos para simplificar este proceso. Este proyecto tiene un contexto médico y de ingeniería civil eléctrica. En particular, esta última se enfoca en el procesamiento de señales de EEG y fMRI para implementar un algoritmo capaz de predecir el foco epiléptico. Esta combinación entre el ámbito médico y la aplicación de conocimientos de ingeniería es esencial para desarrollar una herramienta más eficiente y precisa que pueda mejorar el diagnóstico y tratamiento de la epilepsia.

2.8. Investigaciones importantes

En la elaboración de esta memoria de título, nos hemos basado en investigaciones previas tales como [44] [20] [43] [45], ya que estamos introduciendo por primera vez el desarrollo y aplicación de un algoritmo de co-atención para predecir el foco epiléptico. Este enfoque técnico implica adaptar métodos ya existentes, utilizando la información recopilada de investigaciones anteriores, con el objetivo específico de mejorar la precisión en la identificación del foco epiléptico.

2.8.1. Modelo Conformer

En un estudio reciente sobre la predicción de señales de EEG [44], se afrontó el desafío empleando una combinación de dos modelos distintos. En primer lugar, se recurrió a redes convolucionales para realizar predicciones de manera más global. Además, se integró un *transformer* en el análisis, lo cual permitió enfocarse en relaciones más específicas, explorando cómo los diferentes segmentos de las señales interactúan entre sí. En el modelo Conformer, a pesar de que las señales de EEG son esencialmente una secuencia unidimensional a lo largo del tiempo, se aplican convoluciones en dos dimensiones (2D), dado que descompone el operador de convolución en dos capas unidimensionales para procesar tanto la información temporal como multicanal de los datos de EEG. En la tabla 2.1 y en la figura 2.14 se puede ver con más detalle la convolución impuesta.

Se llevaron a cabo diversos experimentos utilizando el conjunto de datos BCI Competition

Tabla 2.1: Convolución temporal y espacial. La primera capa utiliza k núcleos de tamaño $(1, 25)$ con un intervalo de $(1, 1)$ para realizar la convolución en la dimensión temporal. La segunda capa conserva k *kernels* de tamaño $(ch, 1)$ con un intervalo de $(1, 1)$, donde ch representa el número de canales del EEG. La tercera capa es un promedio *pooling* a lo largo de la dimensión temporal con un tamaño de núcleo de $(1, 75)$ y un paso de $(1, 15)$.

Capa	Entrada	Salida	Kernel	Paso
Conv. Temporal	1	k	$(1, 25)$	$(1, 1)$
Conv. Espacial	k	k	$(ch, 1)$	$(1, 1)$
Promedio Pooling	k	k	$(1, 75)$	$(1, 15)$

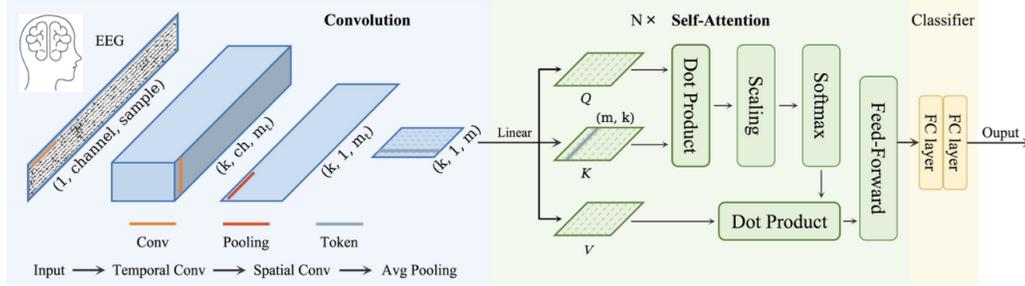


Figura 2.14: Esquema que muestra las tres etapas fundamentales, convolución para la extracción de características temporales y espaciales, *transformer* para modelar relaciones complejas, y un clasificador para la tarea específica, extraída de [44].

IV Dataset 2a, proporcionado por la Universidad de Tecnología de Graz, que incluye datos de EEG de 9 sujetos. Estos experimentos se centraron en cuatro tareas de imaginación motora: mover la mano izquierda, la mano derecha, ambos pies y la lengua. Dentro de estos estudios, se logró un excelente rendimiento con un nivel de *precision* del 81 %.

2.8.2. Predicción de Convulsiones basada en *Transformer* utilizando Electroencefalograma

El objetivo del estudio [20] fue desarrollar un modelo de predicción de convulsiones para auxiliar a los médicos en la toma de decisiones clínicas. Esta investigación se enfocó en predecir la transición entre períodos preictales e interictales, presentando un modelo de predicción fundamentado en la tecnología de *transformers*, simplificado en la figura 2.15.

Primero, se extrajeron las características de tiempo-frecuencia de las señales de electroencefalograma (EEG) mediante la transformada de Fourier de corto tiempo a todos los canales teniendo finalmente una matriz tridimensional (canales, frecuencia, tiempo), que se aplanan en tres vectores:

$$\mathbf{x}_{ch} \in \mathbb{R}^{C \times (F \cdot T)} \quad (2.6)$$

$$\mathbf{x}_f \in \mathbb{R}^{F \times (C \cdot T)} \quad (2.7)$$

$$\mathbf{x}_t \in \mathbb{R}^{T \times (C \cdot F)}, \quad (2.8)$$

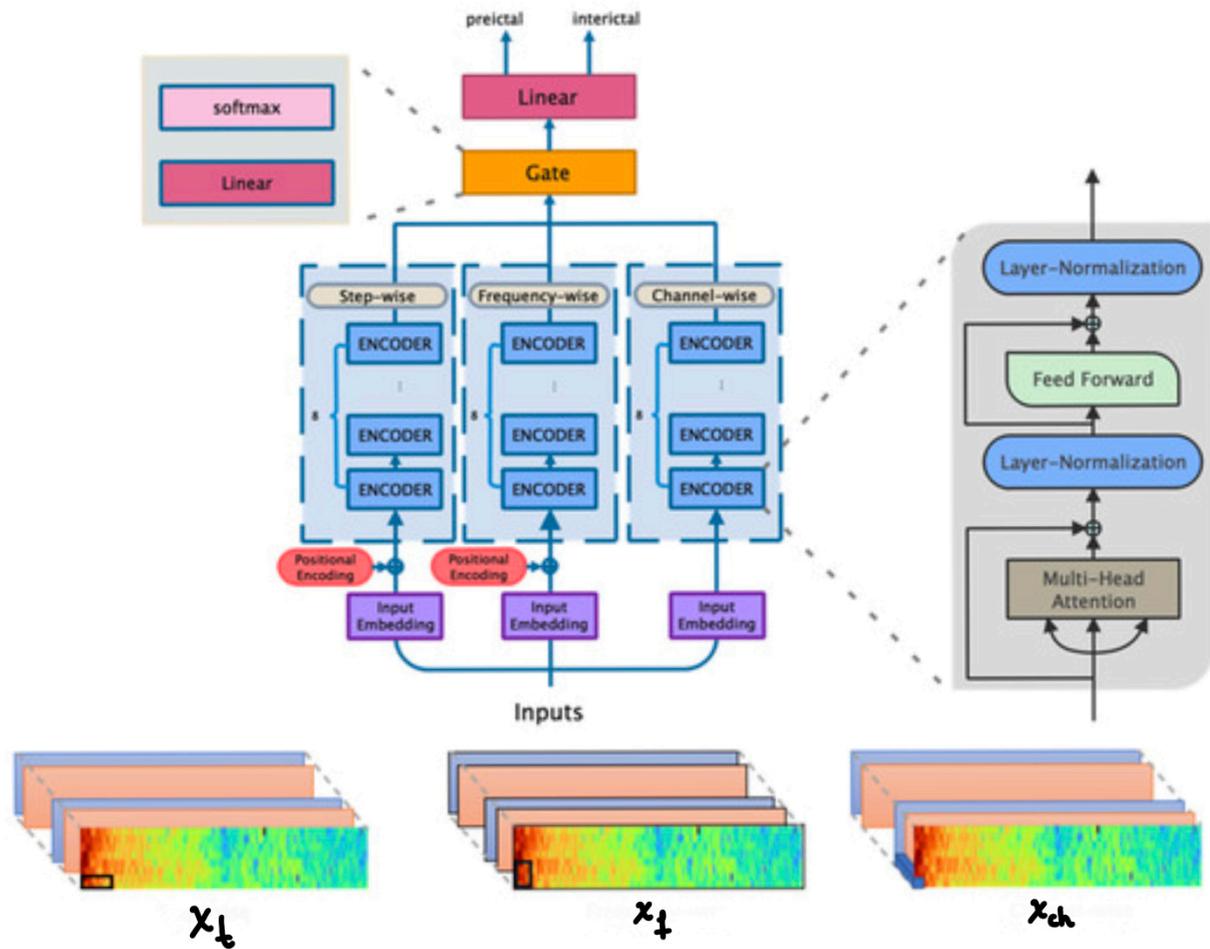


Figura 2.15: Se utiliza la Transformada Corta de Fourier para derivar tres vectores. Estos vectores son sometidos a la operación de un *Transformer* y una red neuronal. Posteriormente, se implementa un mecanismo de puertas ponderadoras (*gating*) y un clasificador para el análisis final. Extraída de [20].

donde,
 Ch = canales
 F = frecuencia
 T = tiempo

Después, se aplican tres *encoders* a los vectores, cada uno compuesto por un *transformer* y una red neuronal. Estos *encoders* procesan la información y generan representaciones más complejas de los vectores originales. Luego, el vector resultante de los *encoders* se somete a una estructura de *gating*. Esta estructura es un mecanismo que asigna pesos adaptables a diferentes partes de la información de entrada. En este contexto, se busca integrar características de tiempo, frecuencia y canal. La combinación de estas características se realiza proyectando los vectores en uno solo a través de una capa de proyección lineal. Lo que se expresa en la formula siguiente:

$$\mathbf{H} = \mathbf{W} \cdot \text{Concat}(\mathbf{Ch}, \mathbf{S}, \mathbf{F}) + \mathbf{b}, \quad (2.9)$$

donde W y b son los parámetros de la capa de proyección lineal, y concatenar significa unir los vectores de tiempo, frecuencia y canal. Después, se aplica la función *softmax* para obtener pesos normalizados, expresados en la fórmula siguiente:

$$\mathbf{g}_t, \mathbf{g}_{ch}, \mathbf{g}_f = \text{Softmax}(\mathbf{H}), \quad (2.10)$$

Donde, g_t = pesos temporales, g_{ch} = pesos de los canales, g_f = pesos frecuenciales.

Estos pesos representan la importancia relativa de cada parte de la información. Finalmente, se obtiene el vector resultante mediante la combinación ponderada de los vectores originales y sus pesos correspondientes, como sigue:

$$\mathbf{y} = \text{Concat}(\mathbf{x}_t \cdot \mathbf{g}_t, \mathbf{x}_{ch} \cdot \mathbf{g}_{ch}, \mathbf{x}_f \cdot \mathbf{g}_f). \quad (2.11)$$

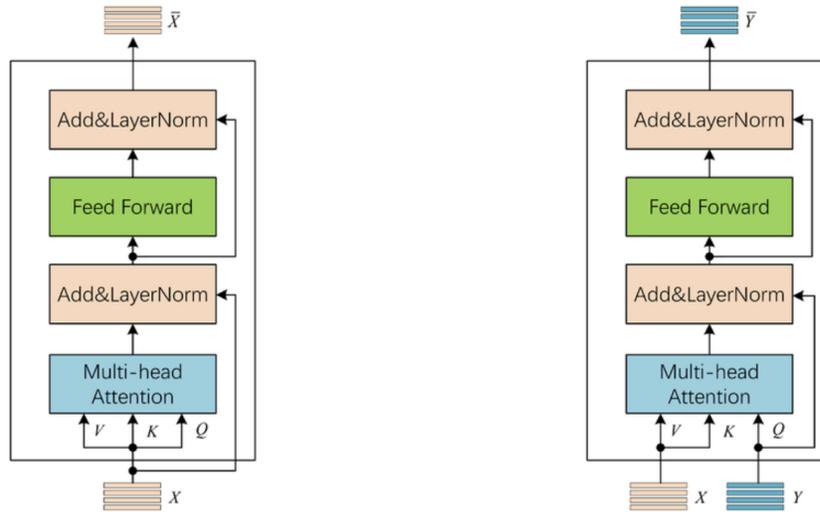
Al final, el vector resultante se somete a un clasificador para realizar la tarea específica de predecir entre estados preictales e intraictales. Se llevaron a cabo experimentos utilizando conjuntos de datos del Children’s Hospital Boston y del Instituto de Tecnología de Massachusetts para evaluar el rendimiento de este modelo. Obteniendo en promedio de todos los pacientes que fueron 22, el *recall* y *accuracy* promedio de 96.01 % y 96.23 %, respectivamente.

2.8.3. Atención Autodirigida Dual con Redes de Preguntas Dispersas para Responder Preguntas Visuales

En [43] se propone un método de co-atención que aborda el desafío de responder preguntas basadas en imágenes. El enfoque principal se centra en la autodirección de la atención, con el propósito de conservar los pesos más significativos cuando la pregunta se somete a atención propia. La atención dirigida utiliza información externa para guiar el enfoque entre elementos en una secuencia, mientras que la autoatención permite que cada elemento influya en todos los demás sin depender de información externa, diferencia mostrada en la figura 2.16. La atención dirigida se calcula de la siguiente forma:

$$\text{Attention}(Q_y, K, V) = \text{softmax} \left(\frac{Q_y K^T}{\sqrt{d_k}} \right) \quad (2.12)$$

- Q_y vector que contiene la información externa que deseamos dirigir mediante la atención dirigida.



(a) self-attention unit.

(b) guided-attention unit.

Figura 2.16: En la representación de la izquierda, observamos una autoatención donde únicamente se considera la información original sin la presencia de otro tipo de dato adicional. En cambio, en el esquema de la derecha, ilustramos la atención dirigida, donde se considera la información adicional Y [43].

- Q_y , K , y V son matrices que representan las consultas, las claves y los valores respectivamente, donde en este caso la consultas son reemplazadas a lo que se quiere dirigir.
- d_k es la dimensión de las claves.
- La operación $Q_y K^T$ calcula la similitud entre las consultas y las claves.
- El resultado se normaliza utilizando la función softmax para obtener pesos de atención.

2.8.4. Fusión Multimodal con Redes de Co-Atención para la Detección de Noticias Falsas

El objetivo de la investigación [45] es predecir noticias falsas utilizando información textual, imágenes y audio. Para lograrlo, emplearon un modelo de triple co-atención multimodal. El bloque de fusión consta de cuatro capas de co-atención diseñadas para combinar información proveniente de imágenes, audio y texto. Una capa implementa dos bloques de atención dirigida en paralelo, según se muestra en la figura 2.17. La salida resultante de estos bloques de atención dirigida se concatena y se utiliza como entrada para la siguiente capa de co-atención en el proceso. Después de completar este proceso cuatro veces, se incorpora un clasificador para abordar la tarea específica, que en este caso es predecir la veracidad de las noticias.

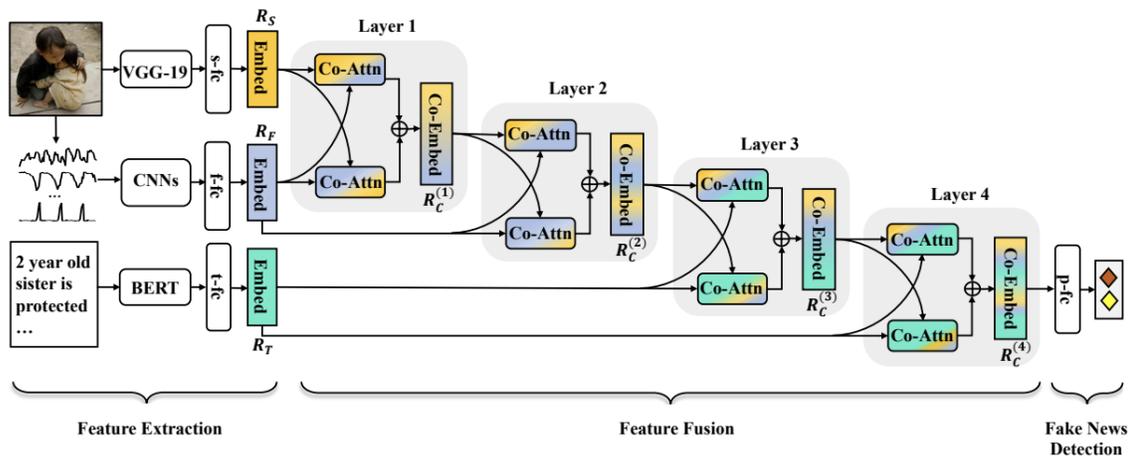


Figura 2.17: Ilustra la co-atención multimodal entre video, audio y texto. Se han implementado 4 capas de co-atención, y cada capa consiste en dos bloques de atención dirigida en paralelo, cuyas salidas se concatenan. Finalmente, se aplica un clasificador para realizar la tarea específica, que es predecir noticias falsas [45].

2.8.5. Atención alternante

Este mecanismo de atención se utilizó para resolver el problema de *Visual Question Answering* (VQA), donde el objetivo es responder preguntas sobre una imagen. En el estudio [34], este mecanismo alterna de manera secuencial la atención entre la imagen y la pregunta. El proceso consta de tres pasos, ilustrados en la figura 2.18.

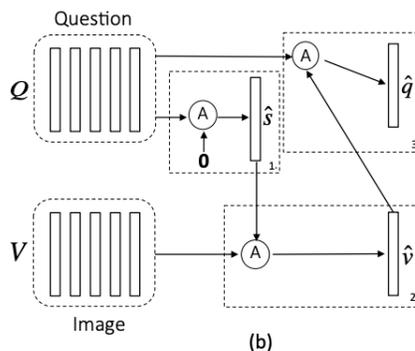


Figura 2.18: Ilustración del proceso de atención alternante. En el primer bloque, la pregunta Q se resume en un vector único s mediante auto-atención. En el segundo bloque, se realiza una atención dirigida entre el vector s y la imagen V , obteniendo así el vector v . Finalmente, en el tercer bloque, se utiliza la característica de la imagen atendida v como guía para dirigir nuevamente la atención a la pregunta Q , resultando en el vector final q [34].

1. **Resumir la pregunta en un vector único** En este paso, la pregunta se resume en un vector único, representado como ‘ q ’. Este vector captura la esencia o el significado clave de la pregunta.
2. **Atender a la imagen basándose en el resumen de la pregunta** Utilizando el resumen obtenido en el paso anterior (‘ q ’), se dirige la atención hacia la imagen. Esto

implica asignar pesos a diferentes partes de la imagen en función de la relevancia para la pregunta resumida.

3. Atender a la pregunta basándose en la característica de la imagen atendida

En este último paso, se atiende a la pregunta en función de la información extraída de la imagen en el paso anterior. La atención se enfoca en aspectos específicos de la pregunta que son relevantes en el contexto de la imagen ya atendida.

Concretamente, definimos una operación de atención $x = A(X; g)$, que toma las características de la imagen (o pregunta) X y la guía de atención g derivada de la pregunta (o imagen) como entradas, y produce el vector de la imagen atendida (o pregunta atendida). Esta operación se expresa en los siguientes pasos:

$$H = \tanh(W_x X + (W_g g)^T) \quad (2.13)$$

$$a_x = \text{softmax}(w_{hx} H) \quad (2.14)$$

$$x = X a_x, \quad (2.15)$$

donde $W_x, W_g \in \mathbb{R}^{k \times d}$ y $w_{hx} \in \mathbb{R}^k$ son parámetros. a_x es el peso de atención de la característica X .

Capítulo 3

Algoritmo de predicción de crisis epilépticas usando EEG

En este capítulo se desarrolla un modelo de clasificación basado en *transformers* que pueda distinguir entre las fases ictales e interictales en electroencefalogramas (EEG).

3.1. Base de datos

La base de datos [1] fue recopilada en el Children’s Hospital Boston y se compone de registros de EEG (Electroencefalograma) obtenidos de pacientes pediátricos que padecen convulsiones intratables. Hay 23 registros obtenidos de un total de 22 sujetos, de los cuales 5 son varones con edades que varían entre 3 y 22 años, y 17 son mujeres con edades comprendidas entre 1.5 y 19 años. Las señales de EEG se muestrearon a una frecuencia de 256 muestras por segundo. Además, cada muestra de señal de EEG se registró con una resolución de 16 bits, lo que garantiza una representación precisa de la amplitud de la señal.

Esta base de datos es una valiosa fuente de información para investigaciones relacionadas con la epilepsia pediátrica, ya que permite el análisis detallado de la actividad cerebral en un grupo diverso de pacientes con convulsiones intratables. La cantidad de canales seleccionados puede variar entre pacientes, por lo que se optó por elegir 18 canales, que representa la cantidad mínima encontrada en un paciente en particular. Estos canales son FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ y CZ-PZ. En la figura 3.1 se muestra la distribución espacial de los canales de EEG seleccionados para esta investigación. En la tabla 3.1 se indica la edad y la cantidad de crisis de los 22 pacientes pediátricos.

Tabla 3.1: La base de datos EEG CHMBIT contiene información de género, edad y la cantidad de crisis en cada registro extraída de [1].

Paciente	Género	Edad	Cantidad de crisis
chb01	F	11	7
chb02	M	11	2
chb03	F	14	7
chb04	M	22	3
chb05	F	7	5
chb06	F	1.5	7
chb07	F	14.5	3
chb08	M	3.5	4
chb09	F	10	3
chb10	M	3	7
chb11	F	12	3
chb12	F	2	13
chb13	F	3	8
chb14	F	9	6
chb15	M	16	13
chb16	F	7	5
chb17	F	12	3
chb18	F	18	6
chb19	F	19	3
chb20	F	6	5
chb21	F	13	4
chb22	F	9	3
chb23	F	6	3

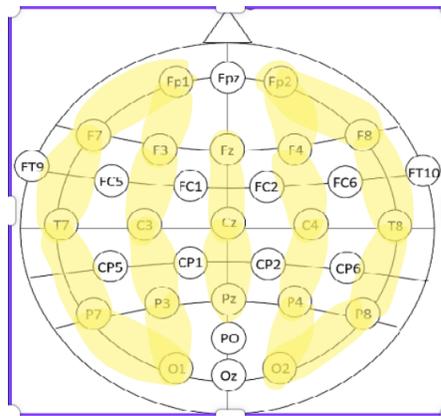


Figura 3.1: En esta memoria se utiliza 18 canales con distribución 20-10, abarcando regiones como Frontal (Fp1, Fp2), Central (C3, C4, Cz), Parietal (P3, P4, P7, P8), Occipital (O1, O2), y Temporal (T3, T4, T5, T6).

3.2. Pre-procesamiento de datos

La elección de ventanas de 5 segundos [20] no está ligada directamente a la frecuencia de muestreo de 256 Hz. Esta duración se basa en la literatura existente, permitiendo analizar segmentos de datos de 1280 puntos independientemente de la frecuencia de muestreo utilizada. Por ende, cada ventana tiene una dimensión de (1280, 18), refiriéndose a los canales de EEG.

3.2.1. Filtrado de señales

Los filtros son esenciales en el procesamiento de señales de EEG para dos propósitos principales: eliminar el ruido no deseado y garantizar que el algoritmo no se vea afectado por interferencias.

- **Filtro de rechazo de banda** (Bandstop Filter)
 - **57 Hz a 63 Hz:** Este filtro se utiliza para eliminar o atenuar selectivamente las frecuencias dentro de ese rango específico (57 Hz a 63 Hz) de la señal de EEG. En general, estas frecuencias son propensas a interferencias de línea eléctrica, artefactos de equipos o movimientos musculares, por lo que filtrarlas permite eliminar el ruido no deseado y mantener la integridad de la señal relacionada con la actividad cerebral.
 - **117 Hz a 123 Hz:** Similar al filtro anterior, pero aquí se apunta a eliminar o reducir frecuencias específicas en un rango diferente (117 Hz a 123 Hz).
- **Filtro pasa alto** (High-Pass Filter):
 - **1 Hz:** Este filtro permite el paso de frecuencias por encima de 1 Hz y atenúa o elimina las frecuencias más bajas que esa cifra. Las señales EEG pueden tener componentes de baja frecuencia causados por artefactos o características no deseadas. El filtro de ayuda a eliminar este ruido o estas componentes de baja frecuencia, preservando las partes de la señal que son más relevantes.

3.2.2. Transformada corta de Fourier

La Transformada de Fourier de Tiempo Reducido (STFT, por sus siglas en inglés Short-Time Fourier Transform) es una técnica que permite analizar cambios en las propiedades de frecuencia de una señal a lo largo del tiempo. En lugar de aplicar la Transformada de Fourier a toda la señal de una vez, la STFT divide la señal en segmentos más pequeños y aplica la Transformada de Fourier a cada uno de estos segmentos individuales. Esto se logra mediante el uso de una función de ventana que selecciona cada segmento de la señal. Al aplicar la Transformada de Fourier a segmentos superpuestos a lo largo del tiempo, se obtiene una representación de cómo varían las frecuencias en la señal a lo largo del tiempo. Esta técnica es útil para el análisis de señales no estacionarias, donde las propiedades de frecuencia pueden cambiar con el tiempo, como en el caso de señales de audio o señales biomédicas como el EEG, donde se pueden identificar patrones de actividad cerebral en diferentes momentos. Se aplicó la Transformada de Fourier de Tiempo Reducido (STFT) a ventanas de 5 segundos. Los parámetros utilizados, basados en la literatura [20] son:

- **Duración de la ventana:** 5 segundos.
- **Número de puntos de la Transformada de Fourier por ventana (nfft):** 1280 puntos de señal, es decir un 5 segundos.
- **Frecuencia de muestreo (fs):** 256 [Hz].
- **Superposición entre segmentos (noverlap):** 128 puntos de señal, es decir medio segundo.

Como resultado, las ventanas generadas tienen una dimensión de (9, 114, 18), donde 9 representa los intervalos de tiempo, 114 indica los intervalos de frecuencia. Por último, los canales que son 18. En la figura 3.2 se presenta un intervalo de 5 segundos de la señal preictal del paciente 4, junto con su correspondiente espectro calculado utilizando los parámetros previamente descritos.

3.3. Modelos utilizados

El objetivo principal es distinguir entre las etapas ictales e interictales. Para alcanzar este propósito, se han creado tres algoritmos supervisado basados en *transformers*.

- Conformer espectral
- Co-atención espectral
- Co-atención temporal

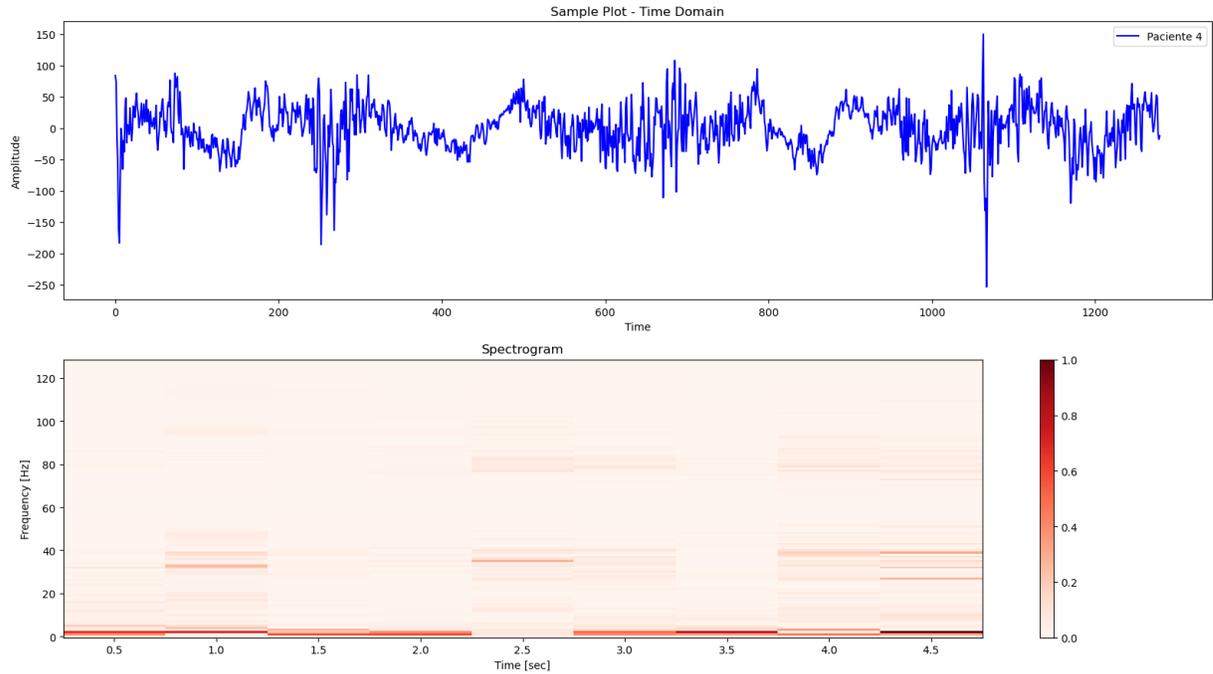


Figura 3.2: En la representación superior se visualiza una señal temporal con una duración de 5 segundos, mientras que en la representación inferior se exhibe el correspondiente espectrograma asociado a dicha señal.

3.3.1. Modelo de Conformer espectral

El modelo Conformer, según la investigación [44], emplea tanto la serie temporal como los canales de EEG. No obstante, durante la observación, se notó una adaptación excesiva a los datos de entrenamiento, sugiriendo la posibilidad de sobreajuste. En respuesta a esta situación, se tomó la decisión de utilizar los espectros de Fourier como alternativa.

Arquitectura de Conformer espectral

La entrada a esta arquitectura son los espectros de Fourier de la señal de EEG, que tienen como dimensión (9, 114, 18), donde 9 representa los intervalos de tiempo, 114 indica los intervalos de frecuencia. Por último, los canales que son 18.

Este modelo se divide en tres partes:

1. **Convolución:** La arquitectura de convolución consiste en una capa de Convolución temporal, con una entrada de dimensión 9 y una salida de 40, utilizando un kernel de (1,25) y un paso de (1,1). Además, incluye una capa de Convolución Espacial, con una entrada y salida de 40, empleando un kernel de (18,1) para los canales y un paso de (1,1). Por último, se integra una capa de *pooling* que utiliza el promedio, con un kernel de (1,75) y un paso de (1,15). La distinción entre el Conformer [44] y Conformer espectral radica en la inclusión de espectrogramas de dimensión 2 (tiempo y frecuencia)

Tabla 3.2: Convolución Conformer Espectral

Capa	Entrada	Salida	Kernel	Paso
Conv. Temporal	9	40	(1,25)	(1,1)
Conv. Espacial	40	40	(18,1)	(1,1)
Promedio Pooling	k	k	(1,75)	(1,15)

en la convolución espacial, convolución explicada en 3.2. Mientras que en el Conformer estándar, solo se considera la señal temporal.

2. **Transformer:** Este Transformer tiene una particularidad que no requiere el uso de codificación posicional, ya que la propia convolución otorga un orden temporal a la señal de EEG. No obstante, en este caso, se implementó exactamente el mismo transformer que se utiliza en el Conformer convencional.
3. **Clasificación:** Estas capas componen una secuencia de transformaciones lineales seguidas de funciones de activación no lineales (ELU), así como capas de *dropout* para introducir regularización. Esta secuencia culmina en una capa de salida final de un vector de dimensión 2, es decir ictal e interictal. En contraste, el diseño del modelo Conformer estaba orientado a abordar tareas relacionadas con la predicción motora.

En la Tabla 3.2, se evidencia la convolución en el modelo Conformer espectral, el cual se distingue del Conformer convencional al incorporar tanto el tiempo como la frecuencia, es decir, dos dimensiones. En la figura 3.3, se muestra la arquitectura completa del modelo de Conformer espectral usado para la predicción de crisis epilépticas usando EEG.

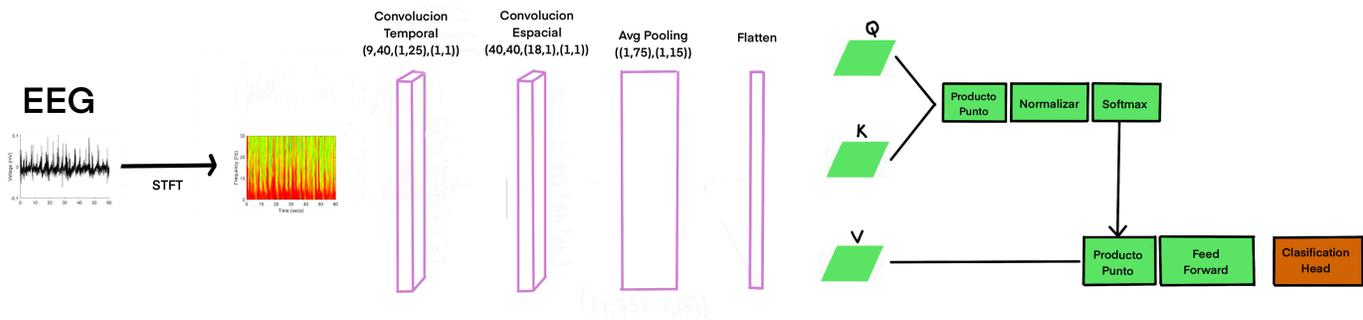


Figura 3.3: La arquitectura del modelo de Conformer espectral, organizada en tres etapas distintas. La primera etapa abarca la convolución, seguida por la segunda etapa que incorpora un transformador. La última etapa consiste en la capa de clasificación, encargada de predecir entre estados ictales e interictales.

3.3.2. Modelo de Co-atención usando espectros de Fourier

Este modelo se basó en dos investigaciones para su desarrollo: la primera, descrita en [20], influyó en la concepción de la primera parte de la arquitectura. En cambio, la segunda investigación, presentada en [45], fue utilizada como referencia para la implementación de la segunda parte de la arquitectura.

Arquitectura de Co-atención espectral

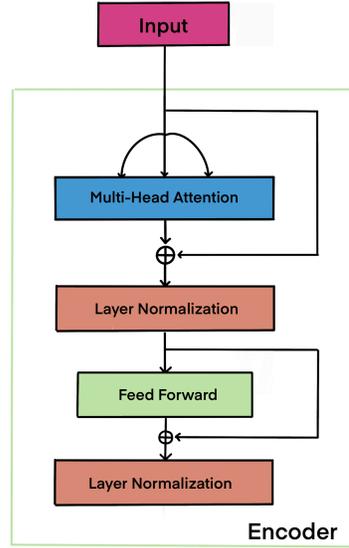


Figura 3.4: En esta representación gráfica se exhibe el Encoder utilizado, cuya primera etapa implica el mecanismo de *self-attention*. Este proceso permite que cada posición en la secuencia evalúe la relevancia de otras posiciones, generando así una representación contextualizada. Posteriormente, se incorpora una capa *feed-forward* para introducir complejidad no lineal a las representaciones. La normalización por capas se aplica después de cada subcapa con el propósito de mantener la estabilidad en la distribución de activaciones. Además, las conexiones residuales, que suman la entrada original a la salida de cada subcapa, facilitan el flujo de información y gradiente, mejorando significativamente la estabilidad y la capacidad de aprendizaje del modelo.

Primera parte

En este caso, al igual que en el modelo de Conformer espectral, se emplean los espectros de Fourier, los cuales tienen una dimensión de (9, 114, 18). En la fase inicial, se procede a la creación de tres vectores a partir de la matriz formada por estos espectros expresados en como sigue:

$$\mathbf{x}_c \in \mathbb{R}^{18 \times (114 \cdot 18)} \quad (3.1)$$

$$\mathbf{x}_f \in \mathbb{R}^{114 \times (18 \cdot 9)} \quad (3.2)$$

$$\mathbf{x}_t \in \mathbb{R}^{9 \times (18 \cdot 114)} \quad (3.3)$$

donde,
c = canales

f = frecuencia
t = tiempo

Después de obtener los tres vectores que representan los canales, la frecuencia y el tiempo, se aplica a cada uno un Encoder ver figura 3.4 que consiste en un *transformer* con un mecanismo de autoatención, junto con una red neuronal. Es importante destacar que únicamente se agrega un *positional encoding* al vector que representa el tiempo, con el propósito de proporcionarle un orden específico en el proceso de codificación.

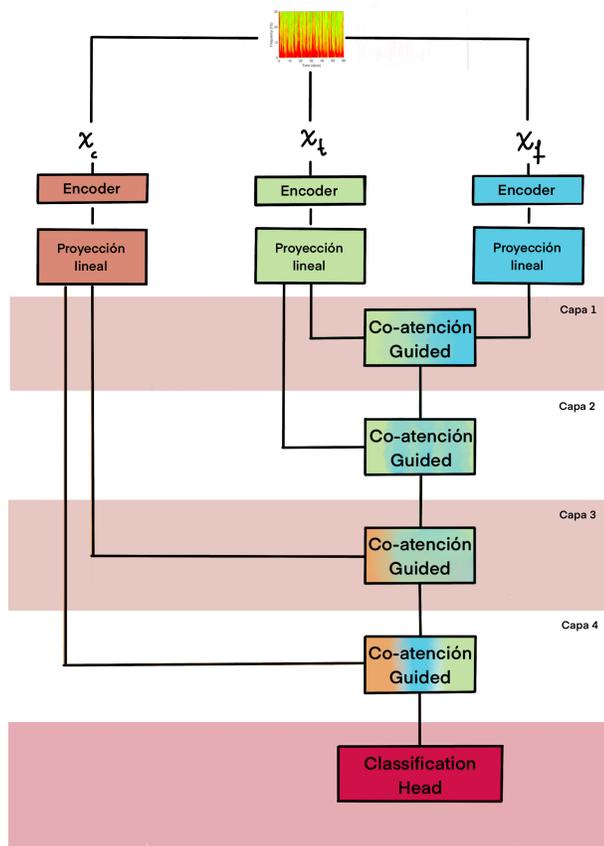


Figura 3.5: Ilustración del modelo de co-atención espectral que se compone de dos partes fundamentales. En la primera fase, los tres vectores, que encapsulan la información temporal, de los canales y de la frecuencia, son sometidos individualmente a un proceso de codificación mediante un *Encoder* y una proyección lineal. En la segunda etapa, se implementan cuatro capas de atención dirigida, que involucran interacciones específicas entre las dimensiones de tiempo, canales y frecuencia. Finalmente, el modelo concluye con una capa de clasificación diseñada para discernir entre ventanas ictales e intraictales.

Segunda parte

El objetivo principal radica en promover la interacción entre los canales y el tiempo, buscando la integración eficiente de información tanto espacial como temporal. En esta dirección, se han incorporado 4 capas de atención dirigidas a las salidas de los *encoders*:

1. **Capa 1:** Implementa atención dirigida entre la salida del *Encoder* temporal y la frecuencia, dirigida por la dimensión temporal.
2. **Capa 2:** Realiza atención dirigida entre la salida de la co-atención de la Capa 1 y la dimensión temporal, bajo la dirección de la Capa 1.
3. **Capa 3:** Ejecuta atención dirigida entre la salida de la co-atención de la Capa 2 y la dimensión de canal, siendo guiada por las características de los canales.
4. **Capa 4:** Establece atención dirigida entre la salida de la co-atención de la Capa 3 y la dimensión de canal, bajo la dirección de la Capa 3.

El proceso concluye con una Capa de Clasificación para discernir entre ventanas ictales e interictales. La figura 3.5 representa el algoritmo completo, donde los tres vectores son procesados por los encoders con la proyección lineal necesaria para realizar las cuatro capas de atención dirigida subsiguientes.

3.3.3. Modelo de Co-atención usando la serie temporal

Este modelo, al igual que la co-atención espectral, se divide en dos partes distintas. La primera parte está inspirada en la investigación [20], pero en lugar de emplear espectros, utiliza la misma serie temporal y los canales de EEG. La segunda parte se centra en la integración de la información temporal y espacial mediante un mecanismo de atención dirigida [43], que vincula los canales y el tiempo.

Aumento de datos (Data augmentation)

Para evitar que el modelo se sobreajuste demasiado a los datos de entrenamiento y garantizar su capacidad para generalizar a nuevas muestras, se implementa una técnica de aumento de datos para electrocardiogramas descritas en la investigación [29]. A continuación, se detallan algunas de las estrategias específicas de aumento de datos utilizadas en el estudio.

- **Transformada de Permutación:** Permuta segmentos de EEG de manera aleatoria para introducir variabilidad temporal.
- **Transformada de Inversión Temporal:** Invierte la señal de EEG con una probabilidad dada, alterando el orden de las muestras.
- **Transformada de Negación:** Niega la señal de EEG (multiplica por -1) con una probabilidad dada, cambiando la polaridad de la señal.
- **Escala:** Aleatoriza la escala de una de las 18 señales de EEG de los canales multiplicándola por un factor aleatorio, variando esta operación según una probabilidad determinada.

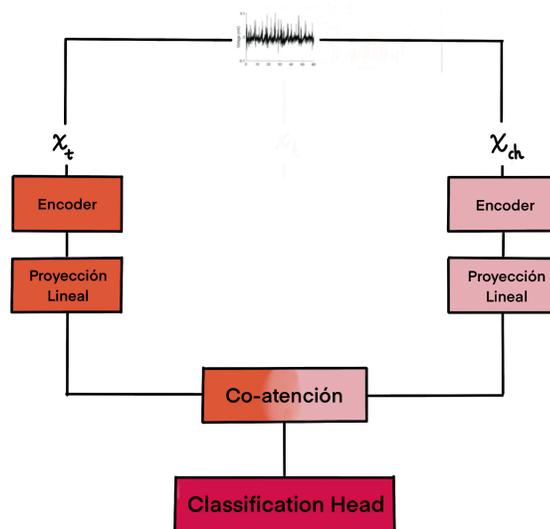


Figura 3.6: Diagrama del modelo de co-atención temporal, el cual se compone de dos etapas fundamentales. En la primera fase, se aplica un *Encoder* tanto al vector temporal como al vector de canales. Posteriormente, se lleva a cabo una co-atención entre ambas representaciones, permitiendo la obtención de una relación espacio-temporal. Finalmente, se emplea una capa de clasificación.

- **Agregar Ruido:** Introduce aleatoriamente ruido *gaussiano* en una de las 18 señales de EEG de los canales, variando esta incorporación según una probabilidad específica dada.

Arquitectura de Co-atención temporal

Primera parte

En este modelo, se emplean ventanas temporales de un segundo de duración, representadas como matrices con dimensiones $(18, 256)$, donde 18 refiere a los canales de EEG utilizados, mientras que 256 representa la cantidad de puntos de datos en un segundo de señal de EEG. Se crean dos vectores a partir de la matriz expresados como sigue:

$$\mathbf{x}_c \in \mathbb{R}^{18 \times (256)} \quad (3.4)$$

$$\mathbf{x}_t \in \mathbb{R}^{256 \times (18)} \quad (3.5)$$

Una vez creadas, las ventanas temporales se someten a un proceso de codificación mediante un *Encoder*, ilustrado en la figura 3.4, donde solo se aplica un *positional encoder* a la componente temporal, seguido de una proyección lineal. Este procedimiento se lleva a cabo con la finalidad de homogeneizar las dimensiones.

Segunda parte

El objetivo central es facilitar la interacción entre los dos vectores mediante un proceso de co-atención, con el fin de establecer una relación espacio-temporal, aspecto crucial en la investigación. En este proceso, se implementa una atención dirigida entre los canales y el tiempo, siendo esta dirigida por los canales. La fase final comprende la clasificación entre estados ictales e interictales. En la figura 3.6 se presenta el diagrama del modelo de co-atención temporal, donde se visualizan la primera y segunda fase del modelo.

3.4. Resultados

Dado que se dispone de un conjunto de datos limitado que constaba únicamente de 22 pacientes, tomamos la decisión de realizar una prueba individual con tres pacientes, reservando el resto para el entrenamiento de los modelos mencionados anteriormente: Conformer espectral, Co-atención espectral y temporal.

3.4.1. Resultados del modelo de Conformer espectral

La tabla 3.3 muestra la cantidad de ventanas de 5 segundos ictales e interictales correspondientes a cada paciente.

Los parámetros del modelo Conformer espectral fueron los siguientes:

- **Tamaño del lote** (Batch size): 100
- **Tasa de aprendizaje** (Learning rate): $2e-3$
- **Número de capas**: 6
- **Número de parámetros**: 1.088.308

En la figura 3.7 se muestra la distribución del *f1-score* correspondiente a los pacientes 1,2 y 3 con el modelo de Conformer espectral. Donde, en promedio el *f1-score* es 0,81 .

3.4.2. Resultados del modelo de Co-atención espectral

Los parámetros del modelo co-atención espectral fueron los siguientes:

- **Tamaño del lote** (Batch size): 32
- **Tasa de aprendizaje** (Learning rate): $1e-4$
- **Dimensión del modelo**: 256

Tabla 3.3: Distribución de las ventanas de 5 segundos en función de los pacientes.

Paciente	Ictal	Interictal
1	1866	625
2	2880	875
3	3162	953
4	6088	3162
5	2882	1775
6	5230	319
7	4798	2836
8	724	743
9	10839	4100
10	5704	1958
11	2879	4404
13	2880	659
14	1953	183
15	1130	612
16	2623	58
17	1542	820
18	2880	792
19	2160	914
20	2880	740
21	2242	821
22	1238	815

- **Número de capas:** 1
- **Número de parámetros:** 4.613.232

En la figura 3.8 se muestra la distribución de *f1-score* del modelo de Co-atención espectral correspondiente a los pacientes 1,2 y 3. Obteniendo, en promedio de *f1-score* de 0,85. En las figuras 3.9 se muestran las matrices de confusión de los tres pacientes.

3.4.3. Resultados del modelo de Co-atención temporal

La tabla 4.2 muestra la cantidad de ventanas de un segundo ictales e interictales correspondientes a cada paciente.

Los parámetros del modelo de co-atención temporal fueron los siguientes:

- **Tamaño del lote** (Batch size): 32
- **Tasa de aprendizaje** (Learning rate): 1e-4
- **Dimensión del modelo:** 256
- **Número de capas:** 1

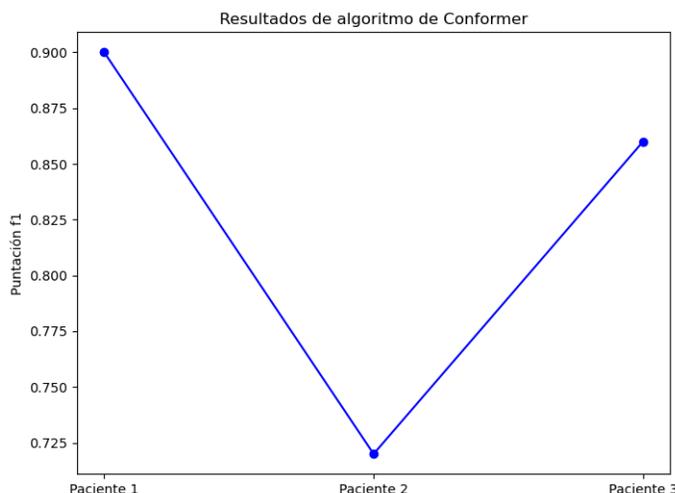


Figura 3.7: Puntuación F1 del modelo Conformer espectral en función de los pacientes 1, 2 y 3

- **Número de parámetros:** 2.913.133

Al considerar ventanas de un segundo, se observa una distribución diferente, detallada en la tabla 4.2. En la figura 3.10 se muestra la distribución de *f1-score* del modelo Co-atención temporal correspondiente a los pacientes 1,2 y 3. Obteniendo, en promedio de *f1-score* de 0,86. En las figuras 3.11, se muestran los resultados de las matrices de confusión para los pacientes 1, 2 y 3.

3.4.4. Análisis de resultados

El análisis del rendimiento basado en el *f1-score* resalta la Co-atención temporal, seguida por la variante espectral y, en última instancia, el modelo Conformer espectral, ver figura 3.12. Es de notar que a pesar de su menor cantidad de parámetros en comparación con los otros modelos, el modelo conformer demuestra un buen rendimiento. Es importante señalar que estos hallazgos son preliminares y que se necesitan más pruebas para obtener una evaluación más exhaustiva de la efectividad del algoritmo de co-atención temporal. Una observación es la capacidad del modelo de co-atención temporal en el reconocimiento de ventanas ictales en comparación con los otros algoritmos. Este punto resulta crucial, ya que los demás modelos tienden a tener dificultades en esta tarea específica.

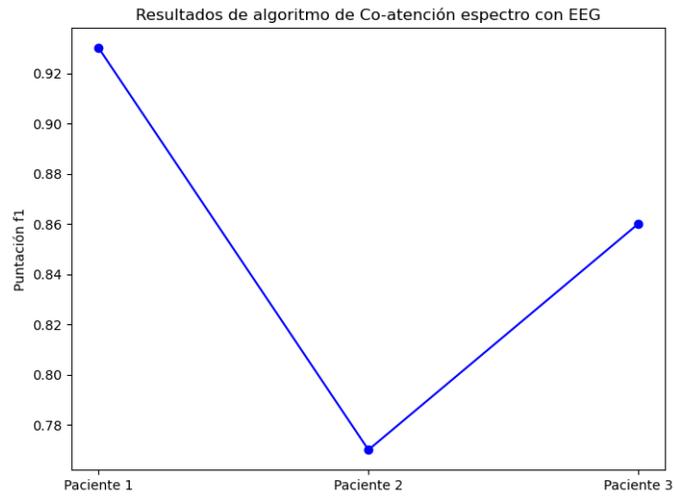


Figura 3.8: Puntuación F1 del modelo Co-atención espectral en función de los pacientes 1, 2 y 3

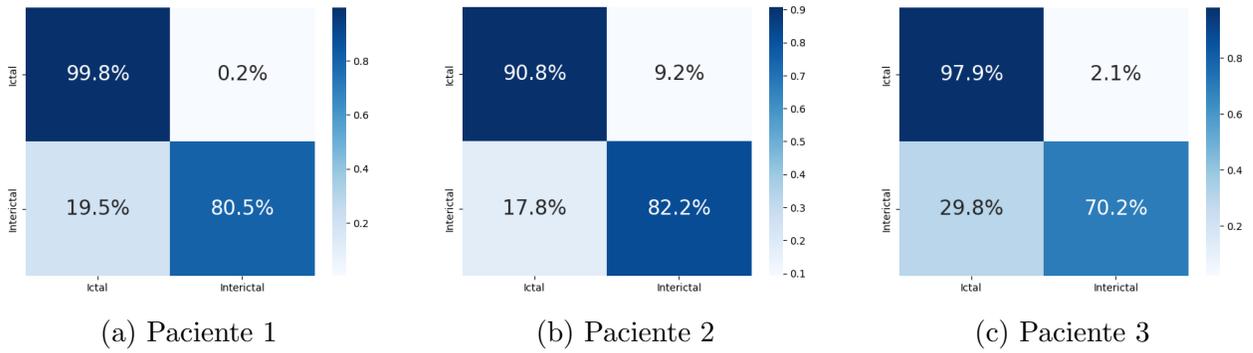


Figura 3.9: Matrices de confusión del algoritmo co-atención espectral para los pacientes 1,2 y 3.

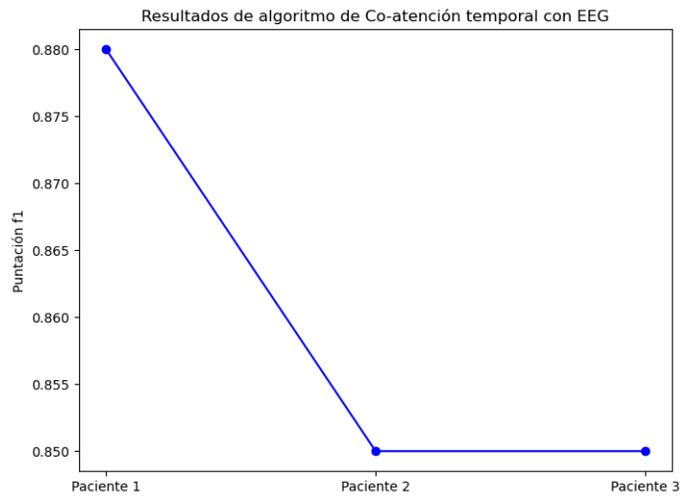
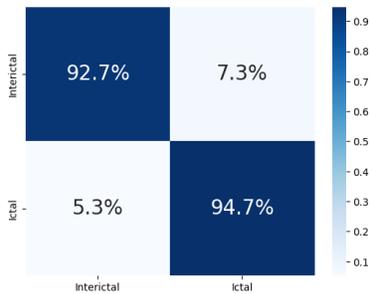


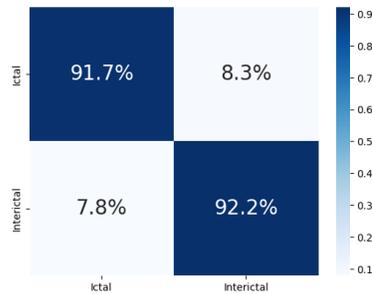
Figura 3.10: Puntuación F1 del modelo Co-atención temporal en función de los pacientes 1, 2 y 3

Tabla 3.4: Distribución de los ventanas de un segundo en función de los pacientes.

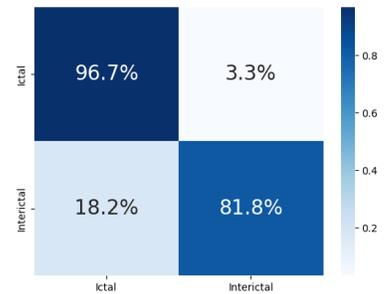
Paciente	Ictal	Interictal
1	9330	3125
2	14400	4375
3	15810	4765
4	6088	3162
5	14410	8875
6	26150	1595
7	23990	14180
8	3620	3715
9	54195	20500
10	28520	9790
11	14365	22020
13	14400	3295
14	9765	915
15	5650	3060
16	13115	290
17	7710	4100
18	14400	3960
19	10800	4570
20	14400	3700
21	11210	4105



(a) Paciente 1



(b) Paciente 2



(c) Paciente 3

Figura 3.11: Matrices de confusión del algoritmo co-atención temporal

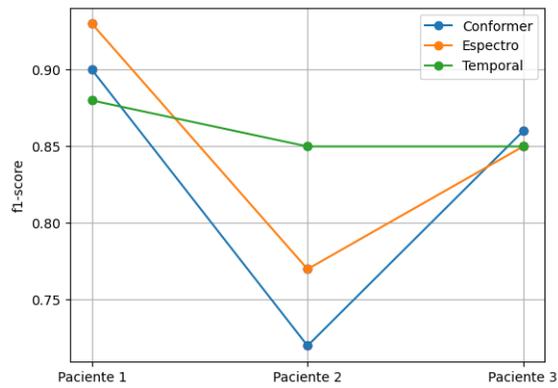


Figura 3.12: Resultados de $f1$ -score, en función de los pacientes 1,2 y 3 para los tres algoritmos: Conformer espectral, Co-atención espectral y Co-atención temporal.

Capítulo 4

Algoritmo de predicción del foco epileptógeno usando EEG

Basándonos en los modelos desarrollados, se evidenció que el modelo de co-atención temporal mostró un rendimiento superior. Para validar de manera sólida la eficacia de la co-atención temporal y espacial en la mejora de la predicción, se llevará a cabo una comparación con los modelos unimodales. En relación con la predicción del foco epileptógeno, la matriz de atención desempeñará un papel clave.

4.1. Base de datos

La base de datos EEG CHMBIT [1] es empleada en algoritmos de predicción de crisis epilépticas. En la tabla 3.1 se presenta la distribución de pacientes según sexo, edad y cantidad de crisis. Sin embargo, con el fin de predecir el foco epiléptico, se incorporaron al mismo algoritmo datos de cinco pacientes pediátricos del INCA, donde se tiene información precisa sobre la ubicación del foco epiléptico, ver tabla 4.1. En la siguiente tabla se presenta la distribución de pacientes según sexo, edad y cantidad de crisis.

Tabla 4.1: Distribución de los pacientes del INCA en función del identificador, sexo, edad y cantidad de crisis.

id paciente	sexo	edad	crisis
sub-007	F	6	4
sub-010	F	10	1
sub-011	F	16	11
sub-015	M	14	1
sub-042	M	2	6

4.1.1. Pre-procesamiento de datos

En este caso, la elección del tamaño de ventana es de un segundo, considerando que la frecuencia de muestreo es de 256 Hz. Por lo tanto, cada ventana consta de 256 puntos.

Filtrado de señales

Los filtros son esenciales en el procesamiento de señales de EEG para dos propósitos principales: eliminar el ruido no deseado y garantizar que el algoritmo no se vea afectado por interferencias. Cuyos filtros son los mismos utilizados en la sección 3.2.1.

4.1.2. Aumento de datos (Data augmentation)

Para evitar que el modelo se adapte demasiado a los datos de entrenamiento y garantizar su capacidad para generalizar a nuevas muestras, se implementa una técnica de aumento de datos para electrocardiogramas descritas en la siguiente investigación [29]. A continuación, se detallan algunas de las estrategias específicas de aumento de datos utilizadas en el estudio. Donde, las funciones utilizadas son las mismas que en la sección 3.3.3.

4.2. Modelos utilizados

En la predicción del foco, se emplea el modelo de co-atención temporal. Con el propósito de evaluar la utilidad de la co-atención, se realizará una comparación con modelos unimodales basados en el tiempo y los canales de manera independiente. Es crucial destacar que el enfoque principal se centra en la relación espacio-temporal, vinculando la información de los canales con el tiempo, lo cual constituye la aportación clave de la co-atención. Este análisis permitirá determinar si la implementación de la co-atención efectivamente mejora el rendimiento en la predicción del foco epileptógeno.

4.2.1. Modelo unimodal temporal

La entrada consiste en un vector temporal con dimensiones (256, 18), donde 256 representa el eje temporal y 18 los canales. Este vector se somete a un proceso de codificación mediante un *Encoder* basado en un *transformer*. Posteriormente, se aplica una capa de clasificación con el objetivo de distinguir entre ventanas que contienen crisis y aquellas que no. En la figura 4.1 se presenta el modelo completo, detallando el *encoder*.

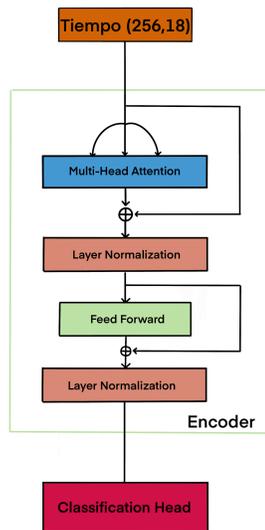


Figura 4.1: El modelo unimodal temporal se compone de un *Encoder* seguido por una capa de clasificación. Su objetivo es discernir entre ventanas que contienen crisis y aquellas que son normales.

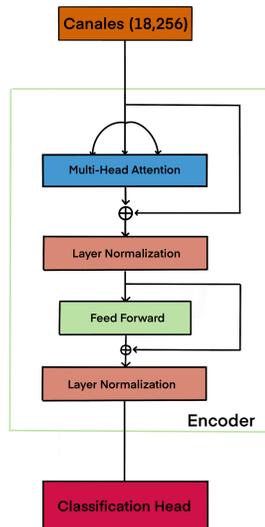


Figura 4.2: El modelo unimodal de los canales se compone de un *Encoder* seguido por una capa de clasificación. Su objetivo es discernir entre ventanas que contienen crisis y aquellas que son normales.

Tabla 4.2: Distribución de ventanas ictales e interictales de los cinco pacientes del INCA.

paciente	interictal	ictal
sub-007	1000	144
sub-010	1000	51
sub-011	2000	573
sub-015	1000	19
sub-042	3000	424

Tabla 4.3: La cantidad de parámetros, tamaño del lote (batch), tasa de aprendizaje (LR) y número de capas (N) para los modelos unimodal temporal de EEG, unimodal de los canales de EEG y co-atención temporal de EEG.

Modelo	parametros	batch	LR	N
Unimodal temporal de EEG	3809837	32	1e-4	1
Unimodal de los canales de EEG	3793453	32	1e-4	1
Co-atención temporal de EEG	3915309	32	1e-4	1

4.2.2. Modelo unimodal canales

A diferencia del modelo unimodal temporal, en este caso, el vector de entrada posee dimensiones (18, 256), donde 18 representa los canales y 256 el eje temporal. Similar al modelo anterior, se implementa un *Encoder* basado en el mismo *transformer*. Posteriormente, la salida del *encoder* se dirige a través de una capa de clasificación con el propósito de diferenciar entre ventanas que contienen crisis y aquellas que no. En la figura 4.2 se presenta de manera detallada el modelo completo, destacando la configuración específica del *encoder* y la capa de clasificación para esta estructura unimodal.

4.3. Resultados

La distribución de las ventanas ictales e interictales de la base de datos CHMBIT se presenta en la tabla 3.3. Para la base de datos del INCA, esta distribución se muestra en la tabla 4.2. En la tabla 4.3 se exhiben los parámetros de cada modelo, el tamaño del lote (batch), la tasa de aprendizaje (LR) y el número de capas (N).

Debido a la limitada cantidad de pacientes disponibles para la formación de los conjuntos de prueba y entrenamiento, hemos decidido evaluar el rendimiento mediante una validación cruzada particular. En cada iteración de prueba, se elige un paciente del INCA, ya que nuestro enfoque se centra en la búsqueda del foco epiléptico, y estos pacientes cuentan con el diagnóstico. En las tablas 4.4, 4.5 y 4.6 se muestran los resultados de *f1-score*, *precision*, *recall* y *accuracy* para los modelos de co-atención temporal utilizando EEG, unimodal temporal y unimodal de los canales, respectivamente.

Tabla 4.4: Resultados del algoritmo de co-atención temporal usando EEG

	sub-007	sub-010	sub-011	sub-042	sub-015
f1-score	0.952	0.965	0.876	0.741	0.733
precision	1	0.932	0.956	0.910	1
recall	0.90	0.968	0.808	0.625	0.578
accuracy	0.98	0.934	0.941	0.945	0.980

Tabla 4.5: Resultados unimodal temporal usando EEG

	sub-007	sub-010	sub-011	sub-042	sub-015
f1-score	0.814	0.879	0.751	0.537	0.480
precision	1	1	1	1	1
recall	0.687	0.784	0.606	0.367	0.315
accuracy	0.961	0.989	0.919	0.921	0.968

4.4. Análisis de resultados

Los resultados del *f1-score* muestran que el modelo de co-atención temporal utilizando EEG posee el mejor rendimiento, con una puntuación promedio de $0,853 \pm 0,111$. Le sigue el modelo de unimodalidad temporal con $0,69 \pm 0,17$, y en último lugar, el modelo de unimodalidad de canales con $0,499 \pm 0,179$. La figura 4.3 ilustra el rendimiento de *f1-score* en función de los pacientes y los canales, evidenciando claramente que el modelo de co-atención temporal tiene el mejor rendimiento. Estos hallazgos respaldan la conclusión de que la implementación de co-atención espacio-temporal, especialmente con la inclusión de datos EEG, mejora significativamente el rendimiento en comparación con el enfoque de utilizar el *encoder transformer* por separado para cada modalidad, ya sea temporal o espacial.

4.5. Matriz de atención

La matriz de atención es una herramienta fundamental en los modelos basados en *Transformer*. En la búsqueda del foco epileptógeno, la matriz de atención puede revelar qué partes de la señal EEG y qué canales son cruciales para la identificación del foco. La interpretación de esta matriz puede proporcionar información valiosa sobre cómo el modelo está procesando la información temporal y espacial, lo que a su vez puede orientar la exploración del foco epileptógeno. En este análisis, se seleccionaron los pacientes sub-007, sub-010 y sub-011, ya que

Tabla 4.6: Resultados unimodal de los canales usando EEG

	sub-007	sub-010	sub-011	sub-042	sub-015
f1-score	0.731	0.247	0.458	0.471	0.592
precision	0.819	0.140	0.365	0.583	1
recall	0.659	0.900	0.614	0.393	0.421
accuracy	0.938	0.731	0.676	0.891	0.973

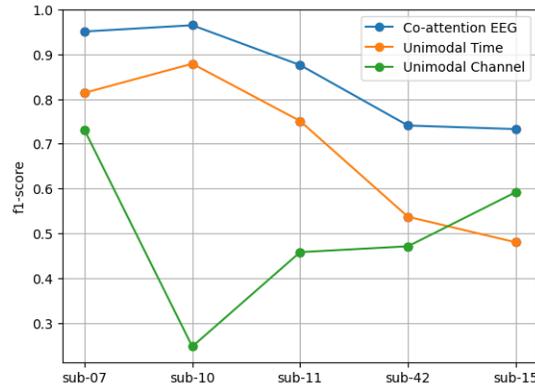


Figura 4.3: Los resultados de $f1$ -score para los tres modelos: co-atención temporal utilizando EEG, unimodal temporal y unimodal de los canales. Estos resultados están representados en función de los cinco pacientes.

el análisis de los pacientes sub-042 y sub-015 no proporciona claridad suficiente. Para cada paciente, se llevarán a cabo dos análisis. El primero se centra en examinar la distribución de la atención espacial, es decir, de los canales, durante los primeros 6 segundos. Este enfoque se justifica dado que, para la predicción del foco epileptógeno, los médicos suelen utilizar los primeros segundos, considerando que una crisis epiléptica es una red que se propaga en el cerebro. El segundo análisis se enfocará en la atención espacial promedio durante los primeros 6 segundos, con el objetivo de identificar las áreas donde el modelo presta mayor atención durante este periodo inicial. Este enfoque permitirá comprender mejor cómo se distribuye la atención espacial en los momentos iniciales de la señal EEG y puede ofrecer perspectivas valiosas para la localización del foco epileptógeno.

4.5.1. sub-007

Este paciente posee una crisis focal en la región fronto-central izquierda, lo que indica que la afectación se localiza en el centro izquierdo del cerebro. En la figura 4.4, se evidencia que durante el primer segundo de la crisis, el modelo dirige su atención de manera destacada hacia la zona afectada. A medida que la crisis se desarrolla hasta el segundo 4, se observa una propagación seguida de una re-focalización posterior, indicando la persistencia del mismo foco de la lesión. En la figura 4.5, se destaca claramente que el foco de atención se encuentra en la zona afectada (Fp1, f7 y f3). Este análisis respalda la conclusión de que el modelo de co-atención temporal responde de manera efectiva para el paciente sub-007, lo que se traduce en una predicción precisa de la actividad epileptógena en este caso particular.

4.5.2. sub-010

Este paciente experimenta una crisis generalizada en el hemisferio izquierdo. En la figura 4.6, durante el primer segundo se observa una clara atención en el hemisferio izquierdo, específicamente en las regiones parietal y occipital. Luego, la crisis empieza a expandirse hasta el segundo 4, donde se focaliza en el hemisferio izquierdo en la parte frontal, extendiéndose en

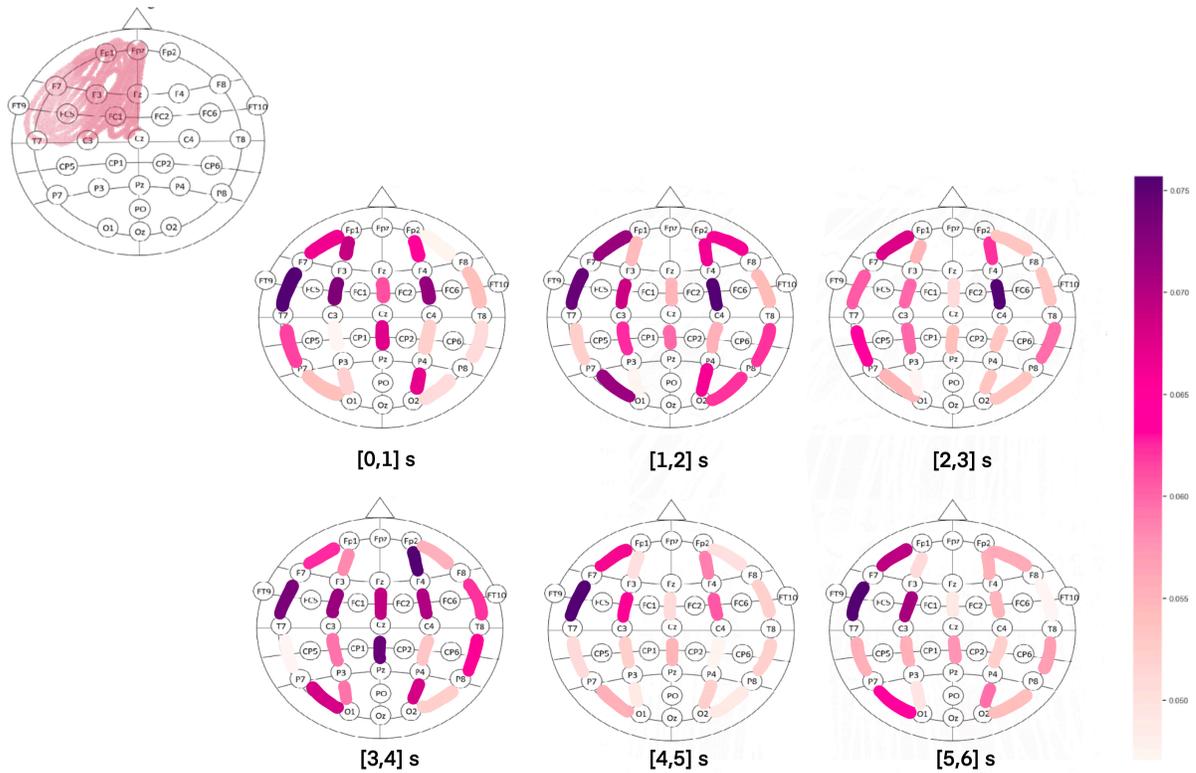


Figura 4.4: La parte superior izquierda de la figura representa la ubicación donde el paciente sub-007 experimenta la crisis, mientras que la parte inferior ilustra la distribución espacial de la atención durante los primeros 6 segundos.

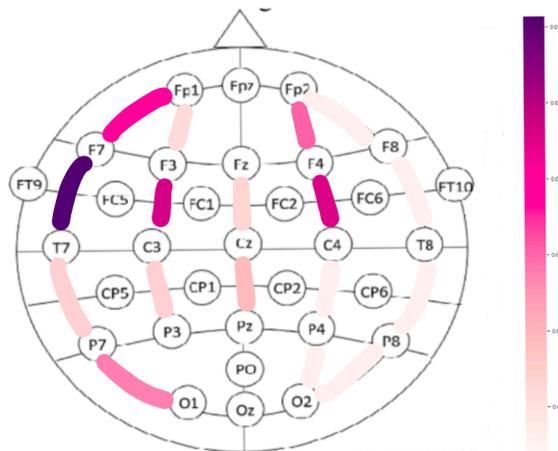


Figura 4.5: Promedio espacial de la atención durante los primeros 6 segundos para el paciente sub-007, centralizada en la región fronto-central izquierda.

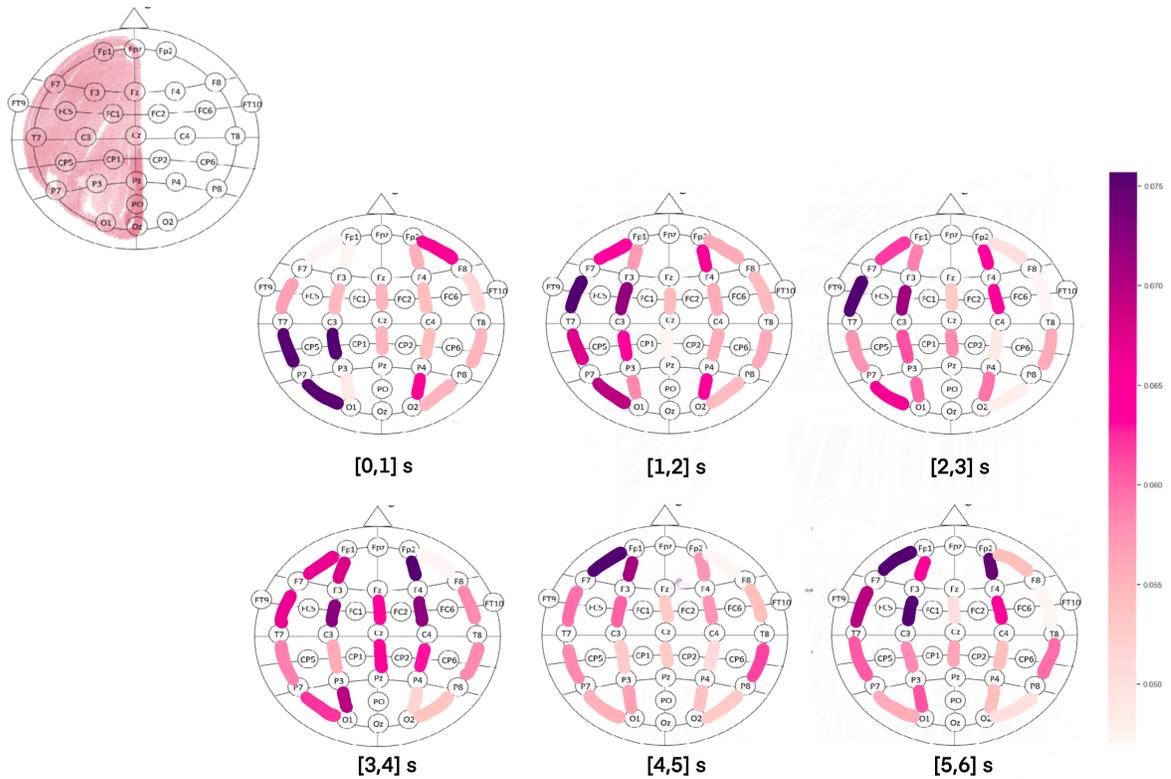


Figura 4.6: La parte superior izquierda de la figura representa la ubicación donde el paciente sub-010 experimenta la crisis, mientras que la parte inferior ilustra la distribución espacial de la atención durante los primeros 6 segundos.

el segundo 5. Aunque no se pueda identificar un foco específico, se aprecia que la crisis está más concentrada en el hemisferio izquierdo. La figura 4.7, muestra que durante los primeros 6 segundos hay una clara atención en el hemisferio izquierdo en comparación con el derecho. En conclusión, el algoritmo de co-atención temporal ha demostrado ser capaz de detectar de manera efectiva la zona afectada para el paciente sub-010.

4.5.3. sub-011

En este paciente se observa una actividad epileptiforme frontal bilateral sincrónica, lo que indica la presencia de descargas epileptiformes en la región frontal de ambos hemisferios cerebrales. La figura 4.9 muestra que esta actividad comienza en el segundo 1, afectando la zona frontal en ambos hemisferios, y se propaga hasta el segundo 3, momento en el cual se focaliza en el hemisferio frontal izquierdo. Posteriormente, se extiende nuevamente hacia el hemisferio izquierdo. Aunque no se visualiza actividad en la zona frontal derecha del cerebro en la figura 4.9, sí se observa en la izquierda. Es importante destacar que este análisis se basa en una única crisis, y es plausible que esta crisis específica tenga una mayor actividad en la zona izquierda. A pesar de este detalle, el algoritmo demuestra su eficacia al identificar

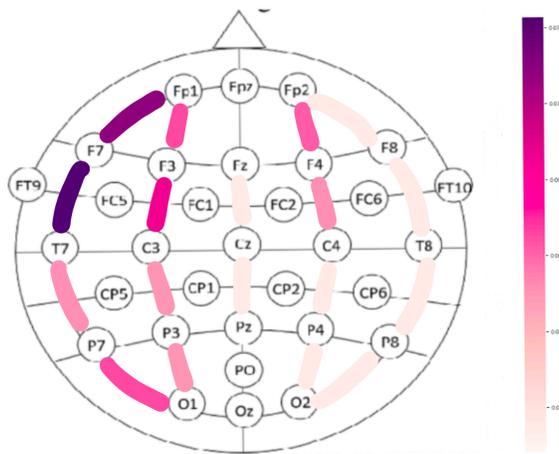


Figura 4.7: Promedio espacial de la atención durante los primeros 6 segundos para el paciente sub-010.

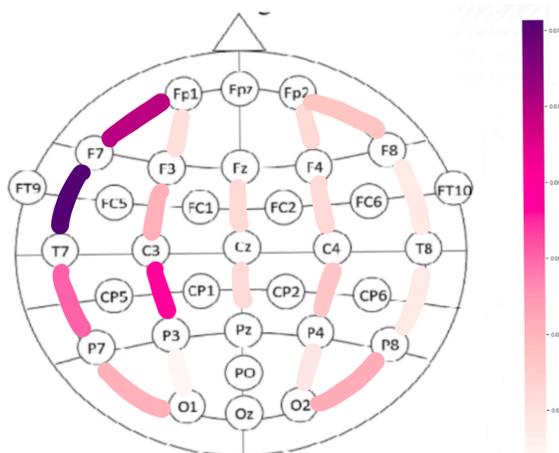


Figura 4.8: Promedio espacial de la atención durante los primeros 6 segundos para el paciente sub-011.

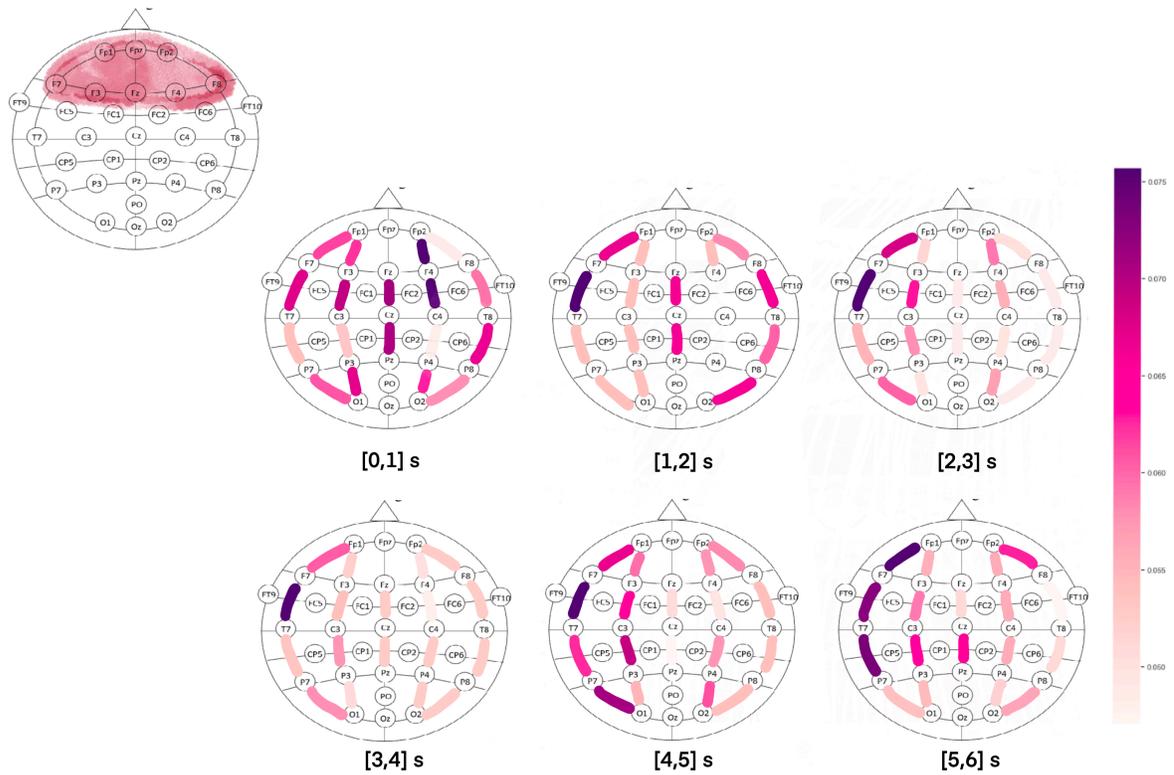


Figura 4.9: La parte superior izquierda de la figura representa la ubicación donde el paciente sub-011 experimenta la crisis, mientras que la parte inferior ilustra la distribución espacial de la atención durante los primeros 6 segundos.

y seguir la evolución de la actividad epileptógena, proporcionando información valiosa sobre la región focal de la crisis en el paciente sub-011.

Capítulo 5

Algoritmo de clasificación de pacientes con epilepsia usando fMRI

En el uso de la resonancia magnética funcional (fMRI), los sujetos de estudio son exclusivamente pacientes que sufren de epilepsia refractaria, es decir, aquellos en quienes los tratamientos farmacológicos no han logrado un efecto adecuado. Estos pacientes son sometidos a este examen, ya que se encuentran en la etapa preoperatoria para evaluar la viabilidad de intervenciones quirúrgicas como tratamiento.

5.1. Base de datos

Para los pacientes que poseen epilepsia refractaria, se usó la base de datos [41], aprobada por la Junta de Revisión Institucional (IRB) de la Universidad de Iowa. Estos datos corresponden a pacientes humanos que fueron sometidos a estimulación eléctrica durante resonancia magnética funcional. La tabla 5.1 presenta la descripción detallada de los pacientes en relación con su edad y género.

5.1.1. Preprocesamiento de datos

1. **Selección del Atlas.** Al emplear FMRI, se necesita asignar la señal a una región cerebral específica lo que implica asociar con un canal a una parte particular de la corteza cerebral. En este caso, se empleó un atlas [40] que divide la FMRI en 100 canales, subdivididos a su vez en 7 partes distintas.
2. **Selección de filtros.** La FMRI utiliza el filtrado Butterworth para procesar señales. Esto implica aplicar un filtro pasa bajo a 0.1 Hz, que permite el paso de frecuencias por debajo de este umbral, así como un filtro pasa alto a 0.01 Hz, que retiene frecuencias por encima de dicho límite.
3. **Eliminación de variables.** Las variables, como *trans_x*, *trans_y*, *trans_z*, *rot_x*, *rot_y*, *rot_z*, *global_signal*, *white_matter* y *csf*, se emplean para eliminar o corregir factores no

Tabla 5.1: Descripción de pacientes con epilepsia refractaria base de datos FMRI de Iowa

Paciente	Género	Edad
1	F	50
2	F	47
3	F	34
4	F	30
5	F	31
6	F	50
7	M	43
8	M	35
9	M	39
10	M	31
11	M	31
12	M	34
13	F	48
14	F	22
15	M	59
16	M	19
17	M	22

deseados.

Las variables de traslación ($trans_x$, $trans_y$, $trans_z$) y rotación (rot_x , rot_y , rot_z) representan el movimiento del sujeto durante la adquisición de las imágenes. La eliminación de estas variables ayuda a corregir los artefactos causados por los movimientos involuntarios.

El término *global_signal* hace referencia a la señal promedio en todo el cerebro, y su eliminación puede contribuir a controlar la variabilidad global de la señal.

White_matter y *csf* representan señales extraídas de la materia blanca y el líquido cefalorraquídeo, respectivamente. La eliminación de estas señales puede ayudar a reducir el ruido proveniente de estas áreas.

4. **Normalización.** La normalización de la señal FMRI es un proceso crucial para ajustar la intensidad de la señal cerebral. Esto es importante porque la intensidad de la señal puede variar entre diferentes momentos o individuos debido a varios factores técnicos o biológicos.

Transformada corta de Fourier

A la señal filtrada y normalizada de la FMRI se le aplicó la transformada corta de Fourier. Los parámetros son:

- **Duración de la ventana:** 30 segundos
- **Frecuencia de muestreo (fs):** 0.5 [Hz], es decir cada dos segundos hay un muestra.

- **Largo de segmento (nperseg)** Se impuso una ventana de largo de 7 segundos de discretación de la Transformada de Fourier.

5.2. Modelos utilizados

Se usarán los siguientes modelos:

- Co-atención espectral
- Co-atención temporal.

5.2.1. Modelo de Co-atención usando espectros de Fourier

Este modelo es una replicación exacta del enfoque de Co-atención espectral utilizado para señales de EEG. Se compone de dos partes, la primera inspirada en [20], y la segunda adopta la metodología presentada en [45].

Primera parte

A diferencia del modelo de EEG, los espectros de Fourier para una señal de FMRI tienen dimensión (4,4,100), correspondientes a la frecuencia, tiempo y canales respectivamente. Entonces, se procede a la creación de tres vectores a partir de la matriz formada por estos espectros expresados como sigue:

$$\mathbf{x}_{ch} \in \mathbb{R}^{100 \times (4 \cdot 4)} \quad (5.1)$$

$$\mathbf{x}_f \in \mathbb{R}^{4 \times (4 \cdot 100)} \quad (5.2)$$

$$\mathbf{x}_t \in \mathbb{R}^{4 \times (100 \cdot 4)} \quad (5.3)$$

donde,
 ch = canales
 f = frecuencia
 t = tiempo

Después de obtener los tres vectores que representan los canales, la frecuencia y el tiempo, cada uno de ellos es sometido a un proceso de codificación a través de un *Encoder*, ver figura 3.4. Sin embargo, únicamente al vector de tiempo se le aplica un *positional encoder* para establecer un orden temporal.

Segunda parte

Dado que se persigue el mismo objetivo que el modelo de co-atención espectral para EEG, centrado en la integración espacio-temporal, se implementan cuatro capas de atención

orientadas hacia las salidas de los *encoders*:

1. **Capa 1:** Implementa atención dirigida entre la salida del *Encoder* temporal y la frecuencia, dirigida por la dimensión temporal.
2. **Capa 2:** Realiza atención dirigida entre la salida de la co-atención de la Capa 1 y la dimensión temporal, bajo la dirección de la Capa 1.
3. **Capa 3:** Ejecuta atención dirigida entre la salida de la co-atención de la Capa 2 y la dimensión de canal, siendo guiada por las características de los canales.
4. **Capa 4:** Establece atención dirigida entre la salida de la co-atención de la Capa 3 y la dimensión de canal, bajo la dirección de la Capa 3.

Al final del proceso, se emplea una Capa de Clasificación para discernir entre pacientes que poseen epilepsia y no. En la figura 3.5 se muestran ambas etapas del algoritmo.

5.2.2. Co-atención usando la serie temporal para FMRI

Este modelo comparte similitudes con la co-atención temporal aplicada a señales de EEG. Se estructura en dos partes, siendo la primera influenciada por la investigación [20]. La diferencia clave radica en el uso de la serie temporal en lugar de los espectros de Fourier. La segunda parte incorpora la atención dirigida entre los canales y el tiempo, siguiendo la metodología propuesta en [43].

Aumento de datos (Data augmentation)

Se ha aplicado una estrategia de aumento de datos inspirada en las técnicas descritas en [29], que, aunque originalmente concebidas para electrocardiogramas, son prácticas comunes en el procesamiento de señales para incrementar la diversidad y robustez del conjunto de datos.

- **Transformada de Permutación:** Permuta segmentos de la señal de FMRI de manera aleatoria para introducir variabilidad temporal.
- **Transformada de Inversión Temporal:** Invierte la señal de FMRI con una probabilidad dada, alterando el orden de las muestras.
- **Transformada de Negación:** Niega la señal de FMRI (multiplica por -1) con una probabilidad dada, cambiando la polaridad de la señal.
- **Escala:** Aleatoriza la escala de una de las 100 señales de los canales multiplicándola por un factor aleatorio, variando esta operación según una probabilidad determinada.
- **Agregar Ruido:** Introduce aleatoriamente ruido gaussiano en una de las 100 señales de los canales, variando esta incorporación según una probabilidad específica dada.

Arquitectura de co-atención temporal para FMRI

Primera parte

Se emplearon ventanas de 30 segundos, por la duración típica de una respuesta hemodinámica. Las ventanas de la señal FMRI tienen dimensión de $(100, 15)$, con un solapamiento de 5 puntos entre cada ventana. Entonces se crearon dos vectores a partir de la matriz expresados como sigue:

$$\mathbf{x}_c \in \mathbb{R}^{100 \times (15)} \quad (5.4)$$

$$\mathbf{x}_t \in \mathbb{R}^{15 \times (100)} \quad (5.5)$$

Una vez creados estos dos vectores, se someten a un proceso de codificación mediante un *Encoder*, ver figura 3.4. Es relevante destacar que solo el vector temporal experimenta la aplicación de un *positional encoder*.

Segunda parte

En sintonía con el modelo de co-atención temporal diseñado para señales de EEG, se establece una relación espacio-temporal mediante la aplicación de atención dirigida entre el vector temporal y el vector de canales. Como paso final, se incorpora una capa de clasificación destinada a discernir entre pacientes con epilepsia refractaria y pacientes sanos. En la figura 3.6 se presenta el diagrama del modelo de co-atención temporal, donde se visualizan la primera y segunda fase del modelo.

5.3. Resultados

Los datos de FMRI se dividieron en dos grupos: pacientes sanos y pacientes con epilepsia refractaria. Se emplearon datos de 61 pacientes para el entrenamiento y datos de 6 pacientes para las pruebas. Las distribuciones de estos conjuntos se muestran en las tablas 5.2 y 5.3, respectivamente, donde DRE es paciente con el epilepsia refractaria y HC sin epilepsia.

5.3.1. Resultados de modelo de co-atención espectral

Los parámetros del modelo fueron los siguientes:

- **Tamaño del lote** (Batch size): 32
- **Tasa de aprendizaje** (Learning rate): $1e-4$
- **Dimensión del modelo**: 256
- **Número de capas**: 1

Tabla 5.2: Distribución de los datos de entrenamiento, compuestos por 47 pacientes con condiciones de salud normales (HC) y 15 pacientes con epilepsia refractaria (DRE).

Paciente	DRE/HC	Número de ventanas
5-52	HC	43
5	DRE	62
6	DRE	55
7	DRE	105
8	DRE	105
9	DRE	105
10	DRE	105
11	DRE	105
12	DRE	84
13	DRE	105
14	DRE	105
15	DRE	63
16	DRE	63
17	DRE	63
18	DRE	63
19	DRE	105

Tabla 5.3: Distribución de los datos de prueba, compuestos por 4 pacientes con condiciones de salud normales (HC) y 2 pacientes con epilepsia refractaria (DRE).

Paciente	DRE/HC	Número de ventanas
1	DRE	105
2	DRE	105
1	HC	43
2	HC	43
3	HC	43
4	HC	43

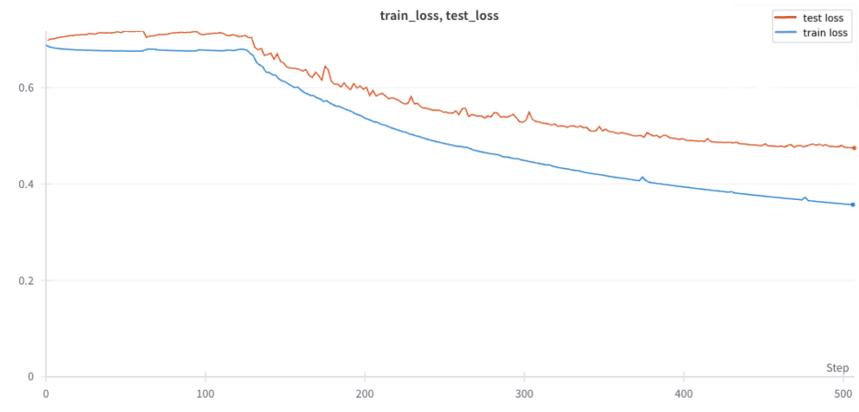


Figura 5.1: Curva de loss del algoritmo de co-atención espectral con FMRI

- **Número de parámetros:** 3.546.473.

Este modelo fue entrenado 5 veces, obteniendo un *f1-score* promedio de $0,87 \pm 0,05$.

En la Figura 5.1, se aprecia la curva de pérdida del modelo que alcanzó el mayor valor de *f1-score*, el cual fue de 0,88.

5.3.2. Resultados de modelo de co-atención temporal

Los parámetros del modelo fueron los siguientes:

- **Tamaño del lote (Batch size):** 32
- **Tasa de aprendizaje (Learning rate):** $1e-4$
- **Dimensión del modelo:** 256
- **Número de capas:** 1
- **Número de parámetros:** 2.843.423.

Este modelo fue entrenado 5 veces, obteniendo un *f1-score* promedio de $0,94 \pm 0,12$.

En la figura 5.2, se representa la curva de pérdida correspondiente al modelo que obtuvo la puntuación más alta en *f1-score*, alcanzando un valor de 0,96.

5.4. Análisis de resultados

Se destaca un mejor desempeño del algoritmo de co-atención temporal, evidenciado por su mayor *f1-score*, respaldado además por una mayor cantidad de parámetros. Respecto a la curva de pérdida del modelo espectral, se aprecia cierto sobreajuste, sin embargo, esto

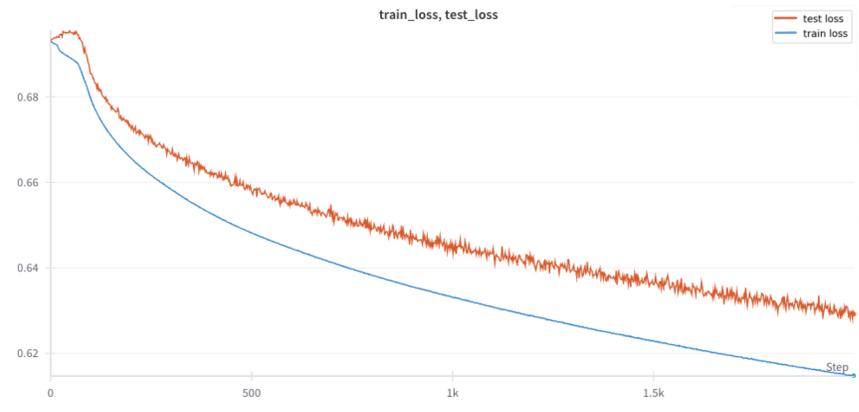


Figura 5.2: Curva de *loss* del algoritmo de co-atención temporal con FMRI

no impacta en los resultados finales. Se aplica la técnica de *early stopping*, deteniendo el algoritmo antes de que ocurra el sobreajuste.

Capítulo 6

Algoritmo de predicción del foco epileptógeno usando FMRI

En el contexto del análisis previo y de los modelos construidos, se ha optado por el algoritmo de co-atención temporal debido a su destacado rendimiento y a la menor complejidad en términos de parámetros. Es fundamental tener presente que el objetivo principal de esta memoria de título es la predicción del foco epileptógeno. Con este objetivo en mente, nos proponemos examinar la relación espacio-temporal del algoritmo, evaluando su influencia en la atención dirigida a los canales y en el tiempo. Para llevar a cabo esta evaluación, se ha realizado una comparación exhaustiva con los algoritmos unimodales. En adición, para la detección del foco epileptógeno, nos centraremos en la matriz de pesos de atención. Esta matriz desempeña un papel crucial al revelar hacia qué aspectos el algoritmo dirige su atención al realizar la predicción. De esta manera, a través del análisis de la matriz de pesos de atención, podremos comprender mejor el comportamiento del algoritmo en términos de la relación temporal y espacial entre los canales.

Tabla 6.1: Distribución de los datos de los pacientes del INCA con epilepsia refractaria.

Paciente	Edad	Sexo
sub-005	10	M
sub-007	6	F
sub-010	10	F
sub-011	16	F
sub-013	8	F
sub-015	14	F
sub-017	10	M

Tabla 6.2: Distribución de los datos de los pacientes del Child Mind Institute con epilepsia refractaria.

Paciente	Edad	Sexo
NDARAH948UF0	5	M
NDARAM675UR8	5	F
NDARAN814UPR	5	F
NDARAP176AD1	6	M
NDARAW179AYF	6	F
NDARBA680RFY	6	F

6.1. Base de datos

Para los pacientes con epilepsia refractaria se usaron 7 pacientes pediátricos del INCA, detallados en la tabla 6.1. Los pacientes de control fueron tomados del estudio [24], que incluyó a 10.000 participantes con edades comprendidas entre los 5 y 21 años. De este grupo, se seleccionaron 6 participantes para lograr un equilibrio con la base de datos del INCA.

6.1.1. Pre-procesamiento de datos

La etapa de preprocesamiento de datos se desglosa en cuatro secciones clave. En primer lugar, se lleva a cabo la selección del atlas, optando por el atlas propuesto en [14], el cual segmenta la fMRI en 39 canales distintos. A continuación, se realiza la selección de filtros, seguida de la eliminación de variables y, por último, se implementa la normalización. Estos procesos se detallan en la sección 5.1.1.

6.1.2. Aumento de datos (Data augmentation)

Se ha aplicado una estrategia de aumento de datos inspirada en las técnicas descritas en la investigación [29], que, aunque originalmente concebidas para electrocardiogramas, son prácticas comunes en el procesamiento de señales para incrementar la diversidad y robustez del conjunto de datos. Funciones explicadas en la sección 5.2.2.

6.2. Modelos utilizados

En el proceso de predicción del foco epileptógeno, se empleará el modelo de co-atención temporal. Este será contrastado con los modelos unimodales de tiempo y canales. El propósito de esta comparación es evaluar si la incorporación de la co-atención realmente contribuye a mejorar los puntajes de rendimiento en comparación con los enfoques unimodales.

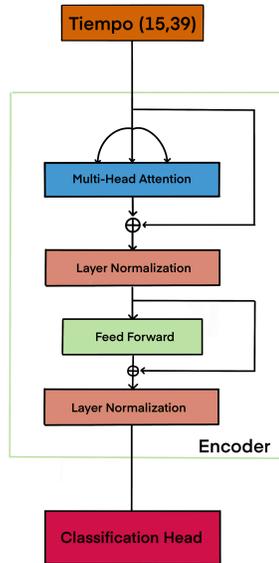


Figura 6.1: El diagrama representa el modelo unimodal temporal, el cual consta de un encoder basado en la arquitectura *transformer* y una capa de clasificación.

6.2.1. Modelo unimodal temporal

La entrada de este modelo consiste en el vector temporal. Posteriormente, dicho vector se somete a un proceso de codificación mediante un *encoder* compuesto por una arquitectura *transformer*. Finalmente, se aplica una capa de clasificación para distinguir entre pacientes con epilepsia y aquellos normales. La Figura 6.1 ilustra de manera completa la estructura de este modelo.

6.2.2. Modelo unimodal canales

La estructura de este modelo comienza con la entrada representada por el vector de los canales. Seguidamente, a dicho vector se le aplica un *Encoder* implementado con la arquitectura *transformer*. Para concluir, se incorpora una capa de clasificación con el fin de diferenciar entre pacientes con epilepsia y aquellos sin dicha condición. La Figura 6.2 detalla de manera integral la configuración completa de este modelo.

6.3. Resultados

Dado que la disponibilidad de pacientes es limitada para la creación de conjuntos de entrenamiento y prueba, se optó por evaluar el rendimiento de este algoritmo utilizando un enfoque de validación cruzada especial. En cada iteración de prueba, se seleccionó un

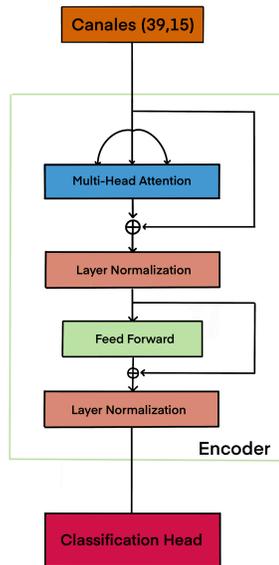


Figura 6.2: El diagrama representa el modelo unimodal canales, el cual al igual que el modelo unimodal temporal posee un transformer y una capa de clasificación.

Tabla 6.3: Distribución de ventanas para los modelos de co-atención temporal, unimodal del tiempo y unimodal de los canales.

paciente	ventanas
NDARAH948UF0	61
NDARAM675UR8	61
NDARAN814UPR	124
NDARAP176AD1	61
NDARAW179AYF	61
NDARBA680RFY	61
sub-005	49
sub-007	49
sub-010	49
sub-011	49
sub-013	49
sub-015	49
sub-017	49

Tabla 6.4: Cantidad de parámetros, tamaño del lote (batch), tasa de aprendizaje (LR) y número de capas (N) para los modelos unimodal temporal de fMRI, unimodal de los canales de fMRI y co-atención temporal de fMRI.

	parametros	batch	LR	N
Unimodal temporal de fMRI	2403973	2	1e-3	1
Unimodal de los canales de fMRI	2403973	2	1e-3	1
Co-atención temporal de fMRI	2423429	2	1e-3	1

Tabla 6.5: Resultados modelo de co-atención temporal para los datos de fMRI.

	sub-05	sub-07	sub-10	sub-11	sub-13	sub-15	sub-17
f1-score	0.94	0.854	0.873	0.923	0.915	0.862	0.912
precision	0.921	0.771	0.831	0.872	0.844	0.830	0.870
recall	0.959	0.959	0.910	0.976	1	0.898	0.959
Accuracy	0.945	0.854	0.880	0.927	0.91	0.872	0.918

par específico de pacientes, uno designado como control sano (HC) y el otro como paciente con epilepsia refractaria (DRE), formando así el conjunto de prueba. Este proceso se repitió para cada paciente con epilepsia, manteniendo fijo al paciente control sano en cada iteración. Este enfoque permitió evaluar exhaustivamente el rendimiento del algoritmo, a pesar de las restricciones en el tamaño del conjunto de datos. En la tabla 6.3 se presenta la distribución de ventanas para todos los pacientes en el conjunto de datos fMRI. En la tabla 6.4 se exhiben los parámetros de cada modelo, el tamaño del lote (batch), la tasa de aprendizaje (LR) y el número de capas (N).

En las tablas 6.5, 6.6 y 6.7, se presentan los resultados de *f1-score*, *precision*, *recall* y *accuracy* para los modelos de co-atención temporal, unimodal del tiempo y unimodal de los canales, respectivamente.

6.4. Análisis de resultados

En los resultados obtenidos, destaca el modelo unimodal temporal como el modelo de mayor rendimiento, según el *f1-score*, con un promedio y desviación estándar en los pacientes de 0.926 ± 0.032 . A continuación, se encuentra el modelo de co-atención temporal utilizando fMRI, con un *f1-score* promedio de 0.897 ± 0.037 . En el último puesto se ubica el modelo unimodal de los canales, con un rendimiento de 0.710 ± 0.059 . Esta evaluación se presenta

Tabla 6.6: Resultados modelo de unimodal de los canales para los datos de fMRI.

	sub-05	sub-07	sub-10	sub-11	sub-13	sub-15	sub-17
f1-score	0.627	0.718	0.667	0.677	0.767	0.8	0.719
precision	0.60	0.58	0.537	0.551	0.682	0.697	0.631
recall	0.653	0.938	0.867	0.877	0.877	0.938	0.83
Accuracy	0.65	0.672	0.6	0.627	0.754	0.79	0.709

Tabla 6.7: Resultados modelo unimodal del tiempo para los datos de FMRI.

	sub-05	sub-07	sub-10	sub-11	sub-13	sub-15	sub-17
f1-score	0.959	0.873	0.916	0.937	0.959	0.887	0.959
precision	0.959	1	0.936	0.957	0.959	0.931	0.959
recall	0.959	0.775	0.898	0.918	0.959	0.836	0.959
Accuracy	0.963	0.9	0.92	0.94	0.963	0.9	0.963

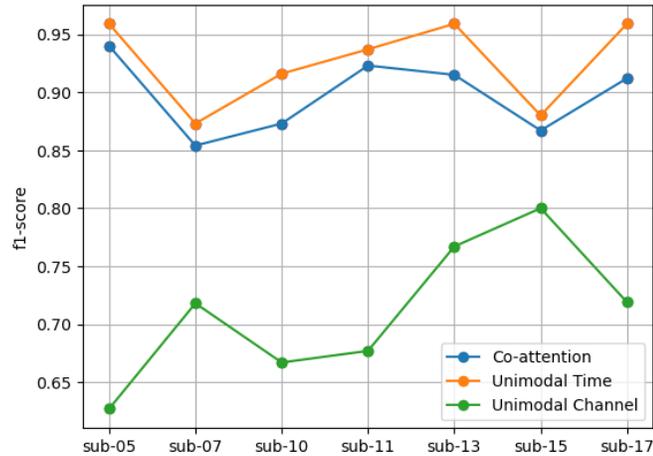


Figura 6.3: Resultados de $f1$ -score para los tres modelos, co-atención temporal, unimodal del tiempo y unimodal de los canales, en función de los pacientes

de manera concisa en la figura 6.3, donde se observa que el modelo temporal ha demostrado ser el más eficaz para todos los pacientes. A pesar de que la co-atención no contribuyó al rendimiento en este caso, esta limitación podría atribuirse al uso de un número reducido de ventanas, especialmente considerando la naturaleza más lenta de la señal temporal del FMRI, con una periodicidad de 2 segundos. No obstante, es fundamental examinar con detenimiento el rendimiento del algoritmo de co-atención temporal utilizando FMRI en la predicción del foco epiléptico, ya que es lo que realmente se busca en esta investigación.

6.5. Matriz de atención: Búsqueda del foco epileptógeno

Siguiendo la misma idea que usamos al analizar el algoritmo para predecir el foco epiléptico con EEG, ahora se utiliza la matriz de atención en el caso de FMRI. Aquí, vamos a calcular los promedios de la matriz de atención, pero mirando sólo las ventanas de tiempo relacionadas con pacientes que tienen epilepsia. Este análisis nos ayudará a entender dónde se concentra el algoritmo al hacer sus predicciones, considerando nuestra idea de que estas áreas específicas están relacionadas con el foco epiléptico.

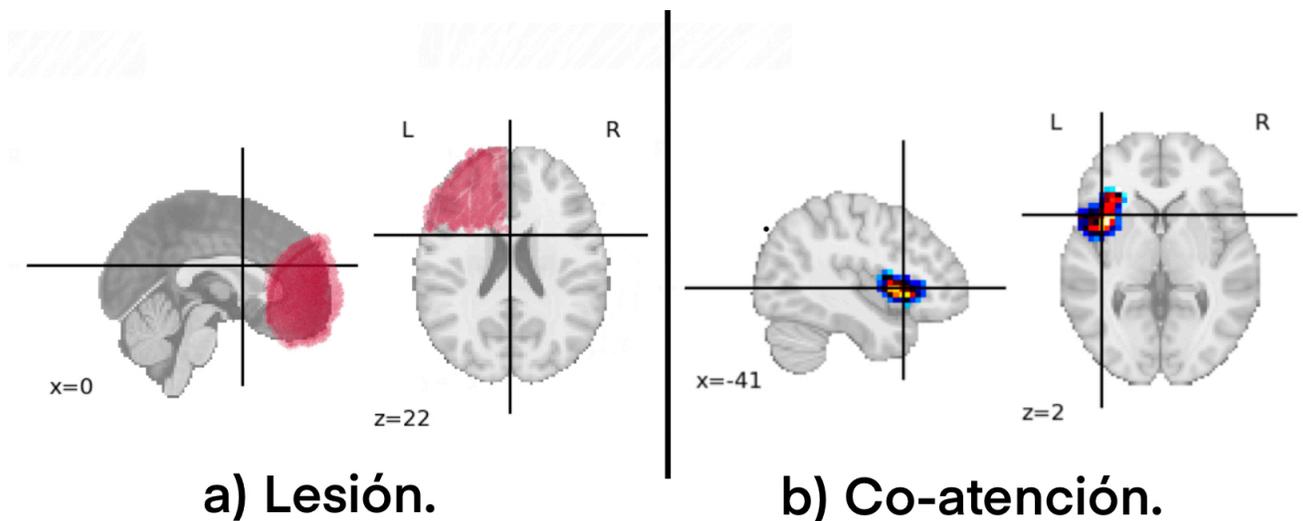


Figura 6.4: En la figura a) se visualiza la ubicación de la lesión en el paciente sub-005, mientras que en b) se presenta la región a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.

6.5.1. sub-005

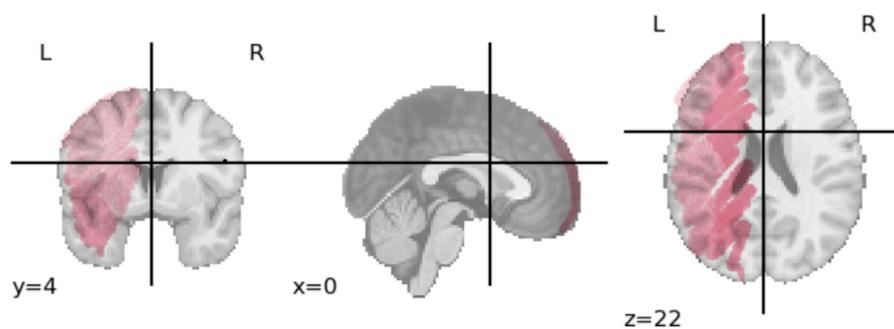
Este paciente presenta una lesión estructural, específicamente un tumor en la región frontal izquierda del cerebro. En la figura 6.4, la parte izquierda representa la zona frontal izquierda, mientras que en la parte derecha se muestra la predicción del algoritmo. Es evidente que la predicción del algoritmo es precisa, ya que se enfoca en esa región, coincidiendo con la ubicación real de la lesión.

6.5.2. sub-007

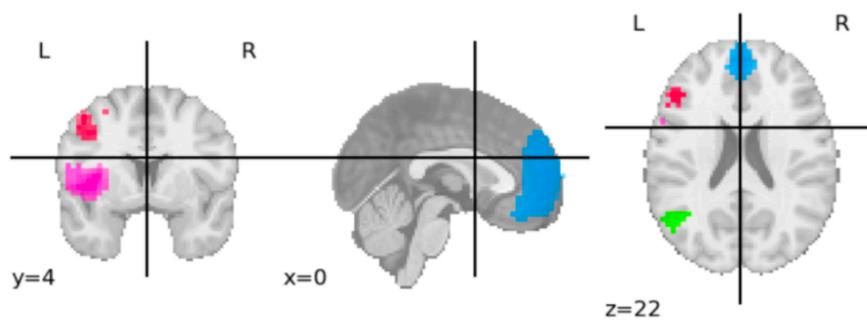
Este paciente presenta una lesión que abarca la totalidad del hemisferio izquierdo del cerebro. Al observar la figura 6.5, se ilustra claramente que en la parte izquierda el algoritmo de atención se focaliza en tres zonas principales, las cuales abarcan la totalidad del hemisferio izquierdo.

6.5.3. sub-010

Este paciente presenta una lesión en la zona frontal izquierda del cerebro. Al examinar la figura 6.6, se nota que en la parte derecha se representa la región a la cual el modelo presta atención. Se observa que el modelo enfoca su atención en dos áreas específicas que abarcan la zona frontal izquierda, coincidiendo con la ubicación de la lesión.

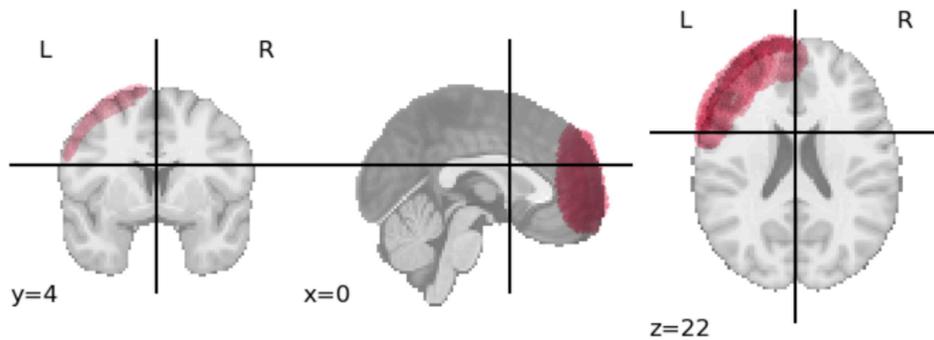


a) Lesión.



b) Co-atención:

Figura 6.5: En la figura a) se visualiza la ubicación de la lesión en el paciente sub-007, mientras que en b) se presenta la región a la que el algoritmo de co-atención temporal, empleando fMRI, presta atención durante la predicción.

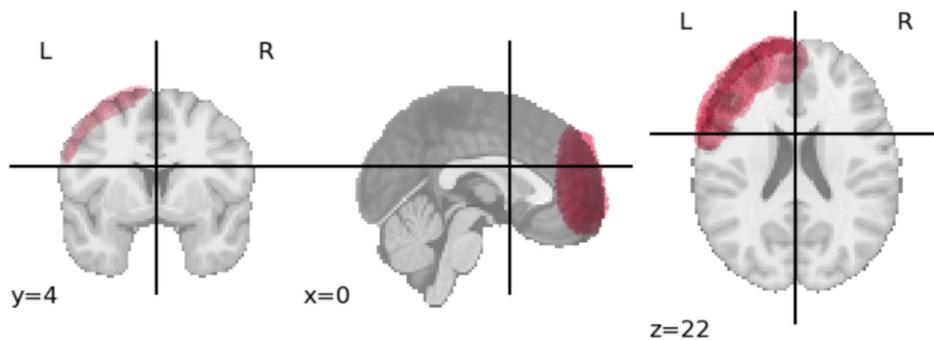


a) Lesión.

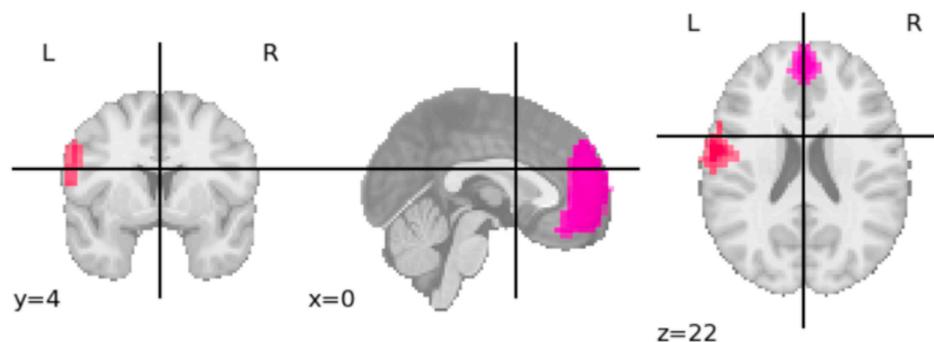


b) Co-atención:

Figura 6.6: En la figura a) se visualiza la ubicación de la lesión en el paciente sub-010, mientras que en b) se presenta la región a la que el algoritmo de co-atención temporal, empleando fMRI, presta atención durante la predicción.



a) Lesión.



b) Co-atención:

Figura 6.7: En la figura a) se visualiza la ubicación de la lesión en el paciente sub-011, mientras que en b) se presenta la región a la que el algoritmo de co-atención temporal, empleando fMRI, presta atención durante la predicción.

6.5.4. sub-011

El paciente sub-011 presenta una lesión en la rodilla y esplenio del cuerpo caloso. En la figura 6.7, la parte a) muestra la ubicación de la lesión, mientras que en b) se indica la región a la cual el algoritmo presta atención. Esta área de atención incluye la ubicación de la lesión, y además, se enfoca en otras dos zonas. Considerando el diagnóstico de EEG que revela la afectación en ambos hemisferios durante las crisis, se postula la hipótesis de que el algoritmo de co-atención temporal usando fMRI centra su atención en esta región debido a la influencia de crisis epilépticas en ambos hemisferios.

6.5.5. sub-013

En este caso, el paciente no presenta una lesión específica, sino que cambios estructurales y neuronales en el hemisferio derecho. En la figura 6.8, en la parte b), se observan las áreas en las cuales el algoritmo de co-atención temporal de FMRI dirige su atención. En las figuras b.2) y b.3), el algoritmo se enfoca en la zona afectada por la lesión estructural, mientras que en las otras b.1) y b.4), se centra en el hemisferio izquierdo. Esto sugiere que el modelo podría tener dificultades para localizar con precisión el foco, dado que este paciente no presenta una lesión estructural definida.

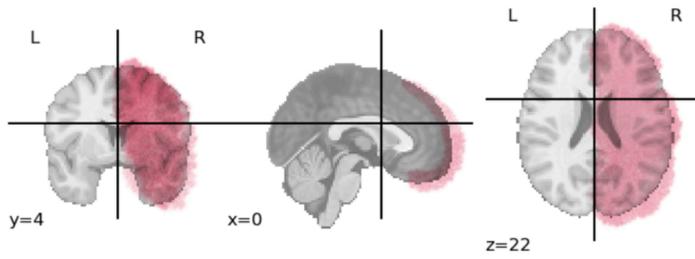
6.5.6. sub-015

Este paciente presenta un tumor en la zona parietal izquierda, como se ilustra en la parte superior de la figura 6.9. En la zona inferior, las marcas indican las áreas a las cuales el algoritmo de co-atención temporal de FMRI dirige su atención, coincidiendo claramente con las zonas afectadas por el tumor del paciente.

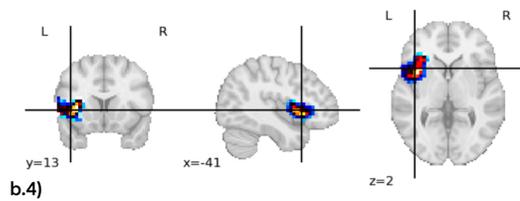
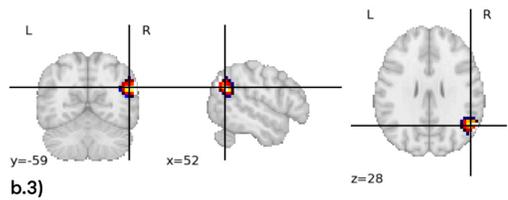
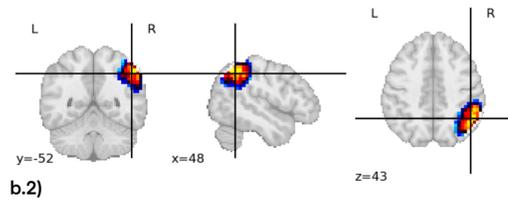
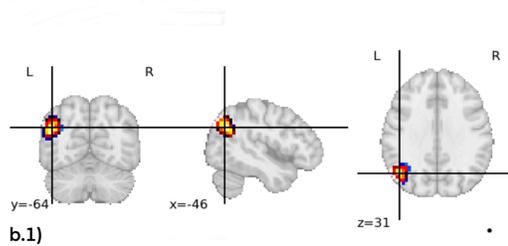
6.5.7. sub-017

Este paciente presenta una lesión en la zona parietal derecha, como se ilustra en la parte superior de la figura 6.10. En la zona inferior, se muestra la región a la cual el modelo de co-atención temporal, utilizando FMRI, dirige su atención, logrando identificar de manera precisa la lesión del paciente.

DMN nodes in MSDL atlas



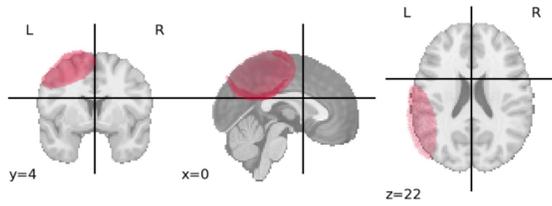
a) Lesión.



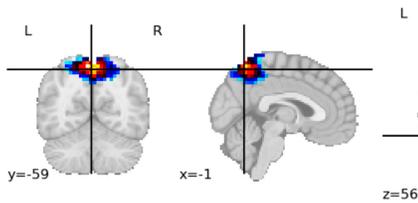
b) Co-atención:

Figura 6.8: En la figura a) se visualiza la ubicación de la lesión en el paciente sub-013, mientras en b) se presentan las regiones a las que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.

DMN nodes in MSDL atlas



DMN nodes in MSDL atlas



DMN nodes in MSDL atlas

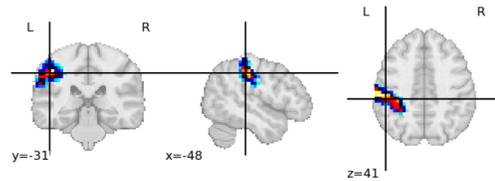
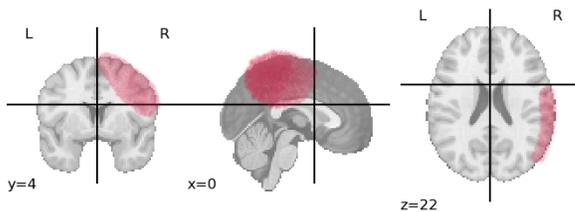
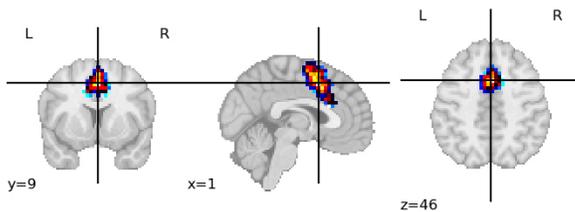


Figura 6.9: En la figura de arriba se visualiza la ubicación de la lesión en el paciente sub-015, mientras que en las figuras de abajo presentan las regiones a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.

DMN nodes in MSDL atlas



DMN nodes in MSDL atlas



DMN nodes in MSDL atlas

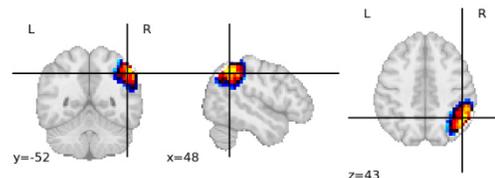


Figura 6.10: En la figura de arriba se visualiza la ubicación de la lesión en el paciente sub-017, mientras que en las figuras de abajo presentan las regiones a la que el algoritmo de co-atención temporal, empleando FMRI, presta atención durante la predicción.

Capítulo 7

Algoritmo de co-atención utilizando EEG y FMRI, para la predicción de crisis epilépticas

El principal desafío de esta investigación radica en la alineación temporal entre las señales de EEG y FMRI, ya que el EEG opera a una frecuencia de 256 Hz, mientras que la FMRI opera a 0.5 Hz, resultando en una discrepancia significativa. Aunque inicialmente se buscaba una sincronización, la realidad clínica en Chile no permite esta alineación debido a la falta de dispositivos. Para superar esta limitación, se adoptó un enfoque de atención alternante [34], procesando las señales de manera secuencial en lugar de en paralelo, lo cual se consideró razonable dada la falta de relación temporal entre ambas señales.

7.1. Base de datos

Se utilizaron datos de cinco pacientes con epilepsia refractaria del Instituto de Neurocirugía Dr. Alfonso Asenjo (INCA) para analizar y desarrollar el modelo de predicción de crisis epilépticas mediante EEG y FMRI. La base de datos del Child Mind Institute [24] se empleó para el grupo de control, compuesto por nueve pacientes, garantizando así una distribución equitativa en las ventanas tanto para las señales de FMRI como para EEG. En las tablas 7.1 y 7.2 se detalla las distribuciones de edad y sexo de los pacientes con epilepsia y del grupo de control, respectivamente.

7.1.1. Pre-procesamiento de datos

FMRI

La etapa de pre-procesamiento de datos se desglosa en cuatro secciones clave. En primer lugar, se lleva a cabo la selección del atlas, para implementar la co-atención entre las señales

Tabla 7.1: Distribución de los datos para el modelo de predicción de crisis epilépticas, utilizando EEG y FMRI. Pacientes del INCA con epilepsia refractaria.

paciente	edad	sexo
sub-007	6	F
sub-010	10	F
sub-011	16	F
sub-015	14	F
sub-017	10	M
sub-042	2	M

Tabla 7.2: Distribución de los datos para el modelo de predicción de crisis epilepticas utilizando EEG y FMRI. Pacientes del Child Mind Institute sin epilepsia refractaria.

paciente	edad	sexo
NDARAH948UF0	5	M
NDARAM675UR8	5	F
NDARAN814UPR	5	F
NDARAP176AD1	6	M
NDARAW179AYF	6	F
NDARBA680RFY	6	F
NDARBF176LPM	7	F
NDARBM839WR5	7	M
NDARBR740NKV	7	M
NDARBU607ZZ3	7	F

de EEG y FMRI, se emplea un atlas creado mediante una máscara, donde cada marcador representa un canal de EEG. En este enfoque, la señal de FMRI se estructura con 18 canales posicionales, coincidiendo con las ubicaciones de los canales de EEG. A continuación, se realiza la selección de filtros, seguida de la eliminación de variables y, por último, se implementa la normalización. Estos procesos se detallan en la sección 5.1.1.

EEG

En este caso, la elección del tamaño de ventana es de un segundo, considerando que la frecuencia de muestreo es de 256 Hz. Por lo tanto, cada ventana consta de 256 puntos.

Filtrado de señales

Los filtros son esenciales en el procesamiento de señales de EEG para dos propósitos principales: eliminar el ruido no deseado y garantizar que el algoritmo no se vea afectado por interferencias. Cuyos filtros son los mismos utilizados en la sección 3.2.1.

7.1.2. Aumento de datos (Data augmentation)

Con el fin de prevenir un ajuste excesivo a los datos de entrenamiento y asegurar la capacidad del modelo para generalizar a nuevas muestras, se emplea una técnica de aumento de datos para electrocardiogramas, detallada en la investigación reciente [29]. Esta estrategia de aumento de datos se aplica de manera consistente tanto a la señal de FMRI como a la señal de EEG. Donde, las funciones utilizadas son las mismas que en la sección 3.3.3.

7.2. Arquitectura del modelo de co-atención usando EEG y FMRI

En el marco teórico, la figura 2.12 expone el modelo de co-atención. Para ambas modalidades, ya sea EEG o FMRI, se implementa un modelo de *transformer*, eligiendo específicamente el modelo de co-atención temporal debido a su destacado rendimiento en esta investigación. La co-atención entre EEG y FMRI no es de tipo paralelo, ya que cada punto en las señales de EEG y FMRI no corresponde al mismo instante de tiempo. Por ende, se opta por una co-atención alternada. Desde esta perspectiva, la señal de FMRI no incluye episodios de crisis, careciendo así de un criterio claro para la clasificación. Su función se limita a proporcionar contexto, por lo que únicamente se aplican las dos primeras etapas de la co-atención alternada:

1. Se comprime la señal completa de FMRI a un solo vector mediante el *encoder* co-atención temporal, con una dimensión de entrada de (300, 18), representando 300 pun-

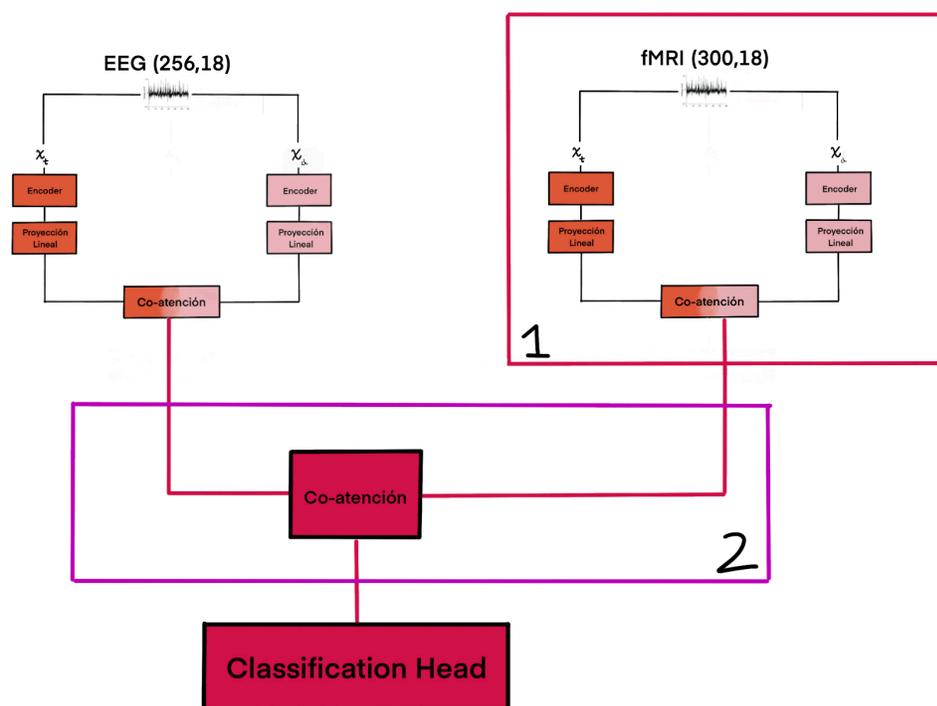


Figura 7.1: La representación gráfica de los dos *encoders* correspondientes a cada modalidad, junto con la separación de las dos etapas en el modelo de co-atención alternada. La secuencia concluye con la incorporación de una capa de clasificación.

tos temporales y 18 canales seleccionados por el atlas FMRI del paciente. Esta etapa se ilustra en la figura 7.1.

2. Se lleva a cabo la atención dirigida entre la salida del modelo de co-atención temporal de EEG y el resumen de la señal de FMRI, generando así un contexto que respalda la predicción en lugar de ser empleado directamente para la clasificación. Esta etapa se visualiza en la figura 7.1.

Después de completar la co-atención alternada entre EEG y FMRI, se implementa una capa de clasificación para discernir entre ventanas de EEG con crisis y sin crisis.

7.3. Resultados

Es necesario evaluar el desempeño del modelo de co-atención temporal de FMRI con 18 canales, dado que representa un cambio con respecto a las configuraciones previas, que fueron evaluadas con 39 o 100 canales. La importancia de esta evaluación radica en garantizar que el algoritmo exhiba un comportamiento efectivo para facilitar la co-atención con EEG y FMRI. En la Tabla 7.3, se detalla la distribución de ventanas según los pacientes para los modelos de co-atención temporal, unimodal temporal y canales, con un enfoque en 18 canales específicos de la señal de FMRI. Los resultados de cada modelo, expresados en métricas como *f1-score*, *precision*, *recall* y *accuracy*, se encuentran también consignados en las tablas 7.4, 7.5 y 7.6.

Tabla 7.3: Distribución de ventanas para los modelos de co-atención temporal de FMRI para 18 canales.

paciente	ventanas
NDARAH948UF0	61
NDARAM675UR8	61
NDARAN814UPR	124
NDARAP176AD1	61
NDARAW179AYF	61
NDARBA680RFY	61
NDARBF176LPM	61
NDARBM839WR5	61
NDARBR740NKV	61
NDARBU607ZZ3	61
sub-007	49
sub-010	49
sub-011	49
sub-015	49
sub-017	49
sub-042	49

Tabla 7.4: Matrices de co-atención temporal aplicado a los datos FMRI de 18 canales.

	f1-score	precision	recall	accuracy
sub-007	0.931	0.903	0.959	0.936
sub-010	0.931	0.903	0.959	0.936
sub-011	0.979	0.979	0.979	0.981
sub-015	0.989	0.980	1	0,990
sub-017	0.959	0.959	0.959	0.963
sub-042	0.970	0.942	1	0.972

Tabla 7.5: Resultados del modelo unimodal temporal aplicado a los datos FMRI de 18 canales.

	f1-score	precision	recall	accuracy
sub-007	0.88	0.862	0.898	0.890
sub-010	0.795	0.971	0.673	0.845
sub-011	0.947	0.978	0.918	0.954
sub-015	0.905	0.842	0.979	0.909
sub-017	0.960	0.924	1	0.954
sub-042	0.875	0.893	0.857	0.890

Tabla 7.6: Resultados del modelo unimodal de los canales aplicado a los datos FMRI de 18 canales.

	f1-score	presicion	recall	accuracy
sub-007	0.898	0.898	0.878	0.909
sub-010	0.526	0.741	0.408	0.672
sub-011	0.9	0.882	0.918	0.9
sub-015	0.738	0.885	0.632	0.8
sub-017	0.738	0.885	0.612	0.790
sub-042	0.898	0.898	0.898	0.900

Tabla 7.7: La distribución de ventanas ictales e interictales tanto para los pacientes de control como en los del INCA.

paciente	interictal	ictal
sub-007	1000	144
sub-010	1000	51
sub-011	2000	573
sub-015	1000	19
sub-017	1000	0
sub-042	3000	424
sub-NDARAH948UF0	687	0
sub-NDARAM675UR8	719	0
sub-NDARAN814UPR	789	0
sub-NDARAP176AD1	717	0
sub-NDARBA680RFY	784	0
sub-NDARBF176LPM	2246	0
sub-NDARBM839WR5	1028	0
sub-NDARBR740NKV	707	0
sub-NDARBU607ZZ3	720	0

Tabla 7.8: Matrices de desempeño de los pacientes del INCA mediante el algoritmo de co-atención utilizando EEG y FMRI.

	f1-score	precision	recall	accuracy
sub-007	1	1	1	1
sub-010	1	1	1	1
sub-011	0.991	1	0.982	0.996
sub-015	1	1	1	1
sub-042	1	1	1	1

En cuanto a la distribución de datos, es importante tener en cuenta que para cada paciente se emplea la totalidad de la señal de fMRI para complementar cada ventana de EEG. La dimensión de la señal de fMRI es de (300, 18), donde 300 representa la dimensión temporal y 18 los canales. Respecto a la señal de EEG, la distribución de ventanas con crisis y sin crisis se detalla en la tabla 7.7. Considerando que el objetivo es predecir crisis epilépticas y, observando la tabla 7.7, se identifica que solo hay 5 pacientes con crisis epilépticas. Dada la limitada cantidad de pacientes, insuficiente para establecer conjuntos de entrenamiento y prueba tradicionales, se optará por una validación cruzada especial. En este enfoque, cada paciente con crisis será designado como el conjunto de prueba, mientras que los demás actuarán como el conjunto de entrenamiento en turnos sucesivos.

Los parámetros del modelo fueron los siguientes:

- **Tamaño del lote** (Batch size): 32
- **Tasa de aprendizaje** (Learning rate): 1e-3
- **Número de capas**: 1
- **Número de parámetros**: 2.517.418

En la tabla 7.8 se muestra los resultados de *f1-score*, *precision*, *recall* y *accuracy* para el modelo de co-atención usando EEG y FMRI. Adicionalmente, se incluirá la representación gráfica de la curva de pérdida (loss) del paciente sub-011, la cual se presenta en la figura 7.2.

7.4. Análisis de Resultados

Al analizar los resultados obtenidos con FMRI utilizando 18 canales, se destaca en la figura 7.3 que el modelo de co-atención temporal exhibe el mejor rendimiento según el *f1-score*. Dado los resultados y considerando que la co-atención temporal-espacial ha demostrado mejor rendimiento que los algoritmos unimodales, surge la posibilidad de utilizar este modelo como encoder en la integración de señales de FMRI en la co-atención conjunta con EEG y FMRI.

Con el fin de realizar un análisis comparativo, los resultados obtenidos se confrontarán con los algoritmos de predicción de foco epiléptico que utilizan EEG, compartiendo cinco pacientes

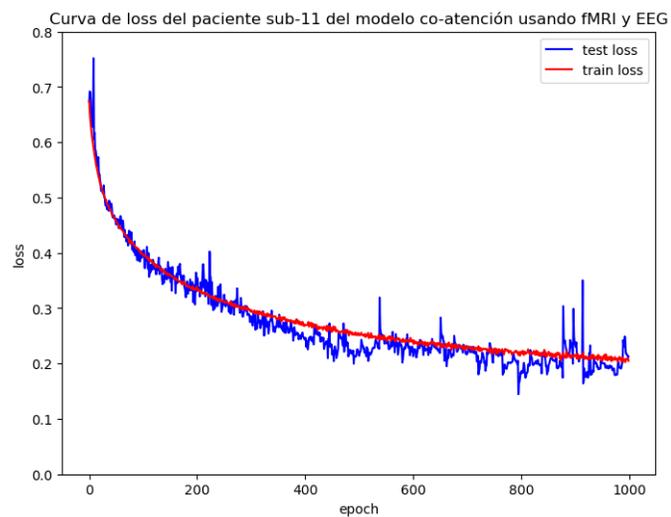


Figura 7.2: Curva de pérdida del modelo de co-atención usando EEG y FMRI, para el paciente sub-011.

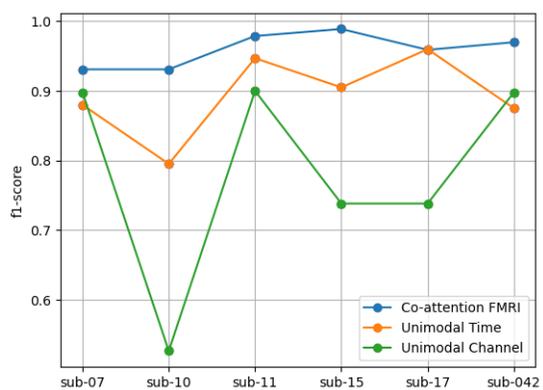


Figura 7.3: Métricas de $f1$ -score para los modelos unimodales de tiempo, canales y de co-atención temporal empleando FMRI con 18 canales.

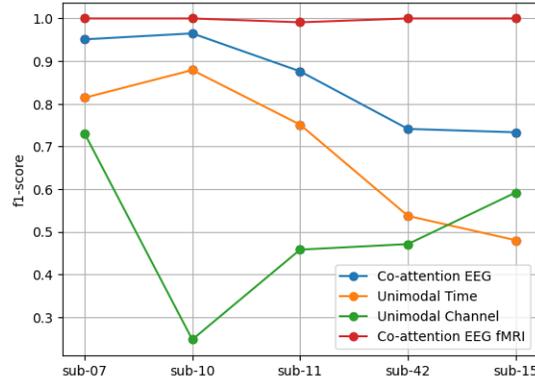


Figura 7.4: Métricas de $f1$ -score para los modelos unimodales de tiempo y canales de EEG, así como para los modelos de co-atención EEG y co-atención EEG junto con fMRI.

en común para llevar a cabo esta evaluación. Con este propósito, se presentan en la figura 7.4 los resultados de $f1$ -score para cada paciente y modelo. El algoritmo de co-atención entre EEG y fMRI fue sometido a un test de hipótesis de permutación. Los resultados revelaron que el valor p obtenido para la comparación entre el rendimiento de la co-atención EEG-fMRI y la co-atención EEG fue de 0,031. Sorprendentemente, este mismo valor p se obtuvo tanto al comparar con el rendimiento de la atención unimodal temporal como con el rendimiento de la atención unimodal de los canales. Este hallazgo subraya la superioridad del modelo de co-atención EEG-fMRI en términos de rendimiento, específicamente demostrado mediante el $f1$ -score.

7.5. Matriz de atención: Búsqueda del foco epiléptico

En forma similar a los modelos anteriores de búsqueda de foco epiléptico en este caso se usa el modelo de co-atención conjunto EEG y fMRI y la matriz de atención para localizar el foco. En este caso, se aplicó un filtro al considerar que los pesos de atención de cada canal se encuentran normalizados entre 0 y 1. Únicamente se tomaron en cuenta los pesos que superaron el umbral de 0.5, descartando aquellos que no cumplían con este criterio. Los pesos seleccionados fueron utilizados en la identificación del foco, mientras que los demás fueron ignorados y no se incorporaron en el análisis. Con el fin de establecer una métrica para evaluar el rendimiento de la predicción del foco, que es la suma de los puntos de atención que se encuentran en la región del foco, dividiendo este valor entre el total de puntos de atención considerados, expresado en la siguiente fórmula.

$$\text{Focal Accuracy (FA)} = \frac{\text{Puntos de atención que están en foco}}{\text{Total de puntos de atención}} \quad (7.1)$$

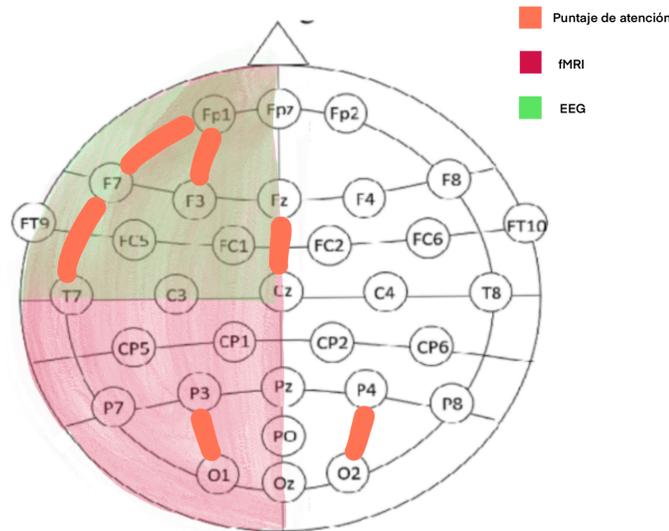


Figura 7.5: Representación visual de la matriz de atención del paciente sub-007, la que revela los puntos de interés identificados por el modelo de co-atención de EEG y FMRI. En la representación, los puntos destacados en color naranja indican las áreas en las que el modelo enfoca su atención conjunta. La zona verde corresponde al diagnóstico obtenido a través de EEG, mientras que la región resaltada en rojo refleja el diagnóstico proporcionado por FMRI.

7.5.1. sub-007

Para el paciente sub-007, el EEG ha confirmado un diagnóstico de epilepsia focal en la zona central izquierda del cerebro, contrastando con el FMRI que muestra una afectación que abarca todo el hemisferio izquierdo, según se ilustra en la figura 7.5. Al examinar las predicciones de la matriz de atención, se destaca una tendencia en las áreas compartidas entre los diagnósticos de FMRI y EEG, especialmente en la región fronto-central izquierda. Sin embargo, se ha observado una atención particular en dos puntos específicos en la zona occipital central (P3-01) y (P4-02). La discrepancia en uno de los puntos de la zona occipital central podría explicarse por la capacidad del modelo de FMRI para señalar más extenso la atención en el hemisferio izquierdo en comparación con el diagnóstico más focal del EEG. Respecto al otro punto de interés en la zona occipital central, es importante reconocer que esta atención puede estar influida por la limitación en el número de pacientes utilizados para entrenar el modelo. A pesar de esta posible restricción en el conjunto de datos de entrenamiento, el modelo demuestra una efectiva localización del foco epileptogénico con un *focal accuracy* del 0,83 %.

7.5.2. sub-010

En el caso del paciente sub-010, se presenta una inversión con respecto al paciente sub-007, ya que el diagnóstico en el EEG indica una epilepsia generalizada en el hemisferio izquierdo, mientras que en el FMRI se observa una lesión en la zona frontal izquierda, según se ilustra

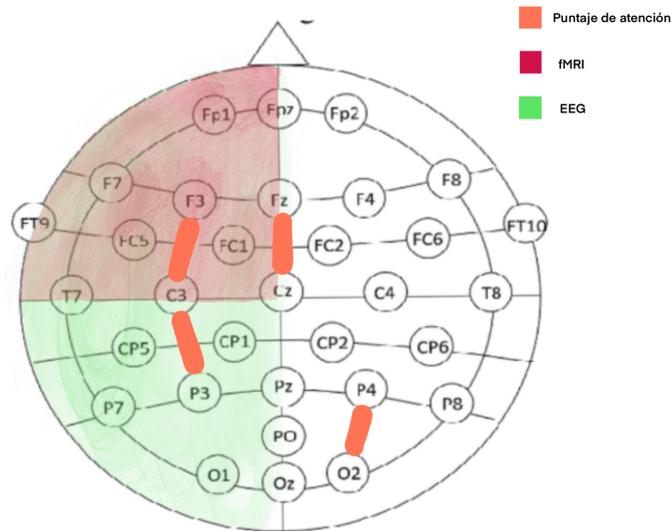


Figura 7.6: Representación visual de la matriz de atención del paciente sub-010, la que revela los puntos de interés identificados por el modelo de co-atención de EEG y FMRI. En la representación, los puntos destacados en color naranja indican las áreas en las que el modelo enfoca su atención conjunta. La zona verde corresponde al diagnóstico obtenido a través de EEG, mientras que la región resaltada en rojo refleja el diagnóstico proporcionado por FMRI.

en la figura 7.6. Al examinar las predicciones del algoritmo de co-atención que utiliza EEG y FMRI, se concluye que la atención se centra en el hemisferio izquierdo, específicamente en las regiones frontal y central. Sin embargo, se destaca un punto de interés en la zona parietal derecha, que podría atribuirse a la limitación en la diversidad de los pacientes utilizados para el entrenamiento del modelo. A pesar de este punto de atención en la zona parietal derecha, la globalidad de las predicciones del algoritmo demuestra una precisión acertada, con un *focal accuracy* del 0,75 %.

7.5.3. sub-011

En el caso del paciente sub-011, se identifica una lesión en el cuerpo calloso, ubicada en la zona central del cerebro, como se muestra en la figura 7.7. Por otro lado, el EEG revela un diagnóstico de epilepsia en la zona frontal, afectando ambos hemisferios. En la predicción del modelo, a través de la matriz de atención, se destaca su capacidad para detectar de manera precisa la lesión presente y también prestar atención al diagnóstico proporcionado por el EEG. Aunque se observa un punto en la predicción que no se alinea completamente con el diagnóstico, es esencial reconocer que este tipo de discrepancias pueden deberse a diversas razones, como la complejidad del caso o las limitaciones en los datos de entrenamiento. A pesar de este detalle, la predicción global del modelo demuestra ser efectiva y concuerda con la presencia de la lesión y el diagnóstico de la epilepsia frontal, con un *focal accuracy* del 0,86 %.

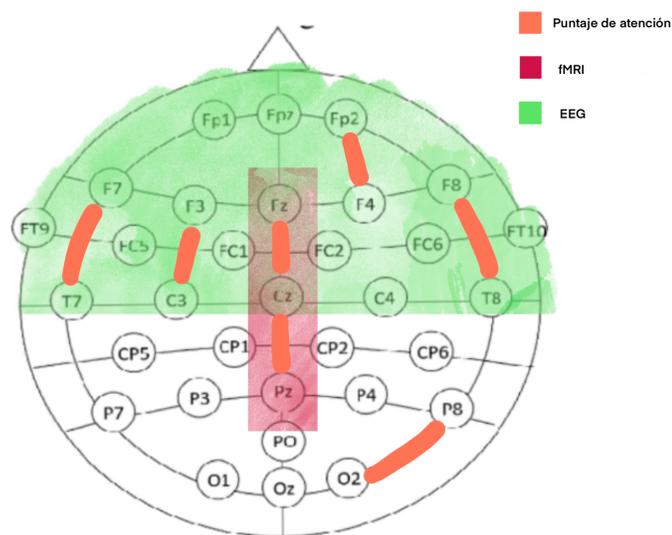


Figura 7.7: Representación visual de la matriz de atención del paciente sub-011, la que revela los puntos de interés identificados por el modelo de co-atención de EEG y FMRI. En la representación, los puntos destacados en color naranja indican las áreas en las que el modelo enfoca su atención conjunta. La zona verde corresponde al diagnóstico obtenido a través de EEG, mientras que la región resaltada en rojo refleja el diagnóstico proporcionado por FMRI.

Capítulo 8

Conclusiones

El propósito de este estudio es desarrollar y aplicar un algoritmo supervisado de inteligencia artificial con el fin de co-localizar alteraciones estructurales en pacientes con epilepsia. La solución propuesta radica en la concepción de un algoritmo de co-atención que facilita la integración multimodal mediante la utilización de datos de electroencefalograma (EEG) y resonancia magnética funcional en estado de reposo (fMRI). Donde se cumplió este objetivo, pero de manera preliminar, dado la poca cantidad de pacientes analizados.

Para ello, se implementaron inicialmente dos modelos de inteligencia artificial. Uno destinado al análisis de electroencefalograma (EEG) para la predicción de momentos de crisis, utilizando la base de datos del Children's Hospital Boston [1]. Este conjunto de datos comprende registros de EEG de pacientes pediátricos con convulsiones intratables, distribuidos en 23 casos con 22 sujetos (5 varones y 17 mujeres), con edades entre 1.5 y 22 años. Se emplearon tres modelos: Conformer espectral, Co-atención espectral utilizando EEG y Co-atención temporal utilizando EEG. Los resultados en $f1$ -score para estos modelos fueron $0,825 \pm 0,09$, $0,856 \pm 0,07$ y $0,84 \pm 0,01$, destacándose como el más efectivo el modelo de co-atención temporal. En una segunda fase, se desarrolló un modelo para la predicción de pacientes con epilepsia refractaria utilizando resonancia magnética funcional (fMRI) con el conjunto de datos [41], aprobado por la Junta de Revisión Institucional (IRB) de la Universidad de Iowa. Este conjunto incluye registros de 26 pacientes humanos sometidos a estimulación eléctrica durante resonancia magnética funcional. Se aplicaron los mismos modelos previamente utilizados en EEG, adaptándolos al fMRI, es decir, co-atención espectral y co-atención temporal para fMRI. Los resultados obtenidos fueron $0,87 \pm 0,05$ y $0,94 \pm 0,12$ en $f1$ -score, respectivamente.

Posteriormente, tras obtener los resultados óptimos con el modelo de co-atención temporal, se procedió a aplicar dicho modelo en pacientes con epilepsia refractaria, donde se conoce el diagnóstico del foco epiléptico. Para este propósito, se utilizaron datos de pacientes del Instituto de Neurocirugía Dr. Alfonso Asenjo. En el caso del modelo de EEG, se replicó la tarea de predicción de crisis epilépticas. Sin embargo, para la identificación del foco epiléptico, se empleó la matriz de atención. Es decir, el algoritmo se enfocó en las áreas señaladas por dicha matriz para localizar el foco. En cuanto a la predicción de crisis, se compararon los resultados de los modelos unimodales temporales y de canales, evaluando si la co-atención temporal y espacial contribuye a mejorar el rendimiento. Los resultados en $f1$ -score fueron

$0,853 \pm 0,111$, $0,690 \pm 0,17$ y $0,499 \pm 0,179$ para los modelos de co-atención temporal y los unimodales temporales y de canales, respectivamente. En relación a la matriz de atención, su aplicación reveló que, para cada paciente, existe una coherencia notable con el diagnóstico propuesto por el médico.

En el fMRI, se emplearon los datos de pacientes provenientes del Instituto de Neurocirugía Dr. Alfonso Asenjo. Al aplicar el modelo de co-atención en conjunto con los modelos unimodales temporales y de canales, se obtuvieron resultados de f1-score de $0,897 \pm 0,037$, $0,926 \pm 0,032$, y $0,710 \pm 0,059$, respectivamente. En cuanto a la predicción del foco, se utilizó la matriz de atención, y su aplicación resultó en un diagnóstico coherente con el propuesto por los médicos para los pacientes del INCA.

El último paso comprendió el desarrollo de la co-atención conjunta entre fMRI y EEG mediante la implementación de un modelo de co-atención alternada. Este enfoque se diseñó considerando que las señales no se registran de manera simultánea, lo que implica la ausencia de coherencia temporal directa. En esta estrategia, la señal de fMRI se empleó para contextualizar la señal de EEG. Los resultados obtenidos demostraron un aumento significativo en el rendimiento, alcanzando un f1-score de $0,998 \pm 0,004$, superando así al modelo de co-atención temporal utilizado anteriormente. En la etapa final, se aplicó la matriz de atención, obteniendo un accuracy de 0,813.

Como trabajo futuro, es crucial tener en cuenta que los resultados obtenidos son preliminares. Aunque se hayan alcanzado los objetivos planteados, es necesario ampliar la muestra de pacientes para obtener un resultado más robusto. En este sentido, se destaca que los modelos se entrenaron exclusivamente con datos pediátricos, por lo que sería beneficioso extender el mismo proceso a datos de EEG de adultos. Adicionalmente, es importante señalar que los pacientes considerados hasta ahora presentan lesiones estructurales. En investigaciones posteriores, se podría explorar la capacidad del modelo para predecir el foco en pacientes sin lesiones, permitiendo así una evaluación más integral de su desempeño y generalización. Dentro del campo de la inteligencia artificial, resulta viable la aplicación de métodos de interpretabilidad, como SHAP o Grad-CAM, con el fin de localizar el foco epileptógeno. Se pueden realizar comparaciones entre ambos métodos, evaluando su rendimiento a través de la métrica de *focal accuracy*.

Bibliografía

- [1] L. Glass J. M. Hausdorff P. C. Ivanov R. G. Mark J. E. Mietus G. B. Moody C. K. Peng H. E. Stanley A. L. Goldberger, L. A. Amaral. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*, 101(23):e215–e220, 2000.
- [2] J. F. Orejuela Zapata A. M. Granados Sánchez. Diagnosis of mesial temporal sclerosis with magnetic resonance imaging. *Neuroradiol J*, 31:701–11, 10 2000.
- [3] G. Kreiman A. Sikarwar. On the efficacy of co-attention transformer layers in visual question answering, 2022.
- [4] N. Parmar J. Uszkoreit L. Jones A. N. Gomez L. Kaiser A. Vaswani, N. Shazeer and I. Polosukhin. Attention is all you need, 2023.
- [5] S. M. Usman; S. Khalid; M. H. Aslam. Epileptic seizures prediction using deep learning techniques. *IEEE Access*, 8:39998–40007, 2020.
- [6] B.-T. Zhang B. Bebensee. Co-attentional transformers for story-based video understanding, 2020.
- [7] L. Rui a Y. Yang a Y. C. b C. Ma a Z. Gao a B. Sun a, J. Lv a. Seizure prediction in scalp eeg based channel attention dual-input convolutional neural network. *Physica A: Statistical Mechanics and its Applications*, 584:126376, 2021.
- [8] Bright Brain Centre. Electroencephalogram (eeg) and brainwaves. <https://www.brightbraincentre.co.uk/electroencephalogram-eeg-brainwaves/>, Año de publicación. Bright Brain Centre Website.
- [9] J. A. Detre. fmri: applications in epilepsy. *Epilepsia*, 45(Suppl 4):26–31, 2004.
- [10] Centers for Disease Control and Prevention. Convulsiones. <https://www.cdc.gov/epilepsy/spanish/basicos/convulsiones.html>, 2023.
- [11] Care Options for Kids. Focal vs. generalized seizures. <https://careoptionsforkids.com/blog/focal-vs-generalized-seizures>, 2023. Care Options for Kids Blog.
- [12] Epilepsy Foundation. Understanding seizures. <https://www.epilepsy.com/what-is-epilepsy/understanding-seizures>, 2023.

- [13] J. Malvey S. Puig G. Argenziano H. Kittler G. Brancaccio, A. Balato. Artificial intelligence in skin cancer diagnosis: A reality check. *Journal of Investigative Dermatology*, 144(3):492–499, 2024.
- [14] F. Pedregosa V. Michel B. Thirion G. Varoquaux, A. Gramfort. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. *Information Processing in Medical Imaging*, 22:562–573, 2011.
- [15] H. Huang X. Yang Z. Wan a X. Wen e H. Zhou, L. Qi. Canet: Co-attention network for rgb-d semantic segmentation. *Pattern Recognition*, 124:108468, 2022.
- [16] Stanford Children’s Health. Electroencefalograma (eeg). <https://www.stanfordchildrens.org/es/topic/default?id=electroencefalogramameeg-92-P09193>, 2023.
- [17] E. M. C. Hillman. Coupling mechanism and significance of the bold signal: a status report. *Annual Review of Neuroscience*, 37:161–181, 2014.
- [18] A. K. Ibrahim, H. Zhuang, E. Tognoli, A. Muhamed Ali, and N. Erdol. Epileptic seizure prediction based on multiresolution convolutional neural networks. *Frontiers in Neuroscience*, 3:8, 2021.
- [19] F. Pittau J. Gotman. Combining eeg and fmri in the study of epileptic discharges. *Epilepsia*, 52(Suppl 4):38–42, Jul 2011.
- [20] H. Xu Y. Yu J. Yan, J. Li and T. Xu. Seizure prediction based on transformer using scalp electroencephalogram. *Appl. Sci. 2022*, 12:4–9, 2023.
- [21] Likitha KN, Nethravathi R, Nithyashree K, Ritika Kumari, Sridhar N, and Venkateswaran K. Heart disease detection using machine learning technique. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pages 1738–1743, 2021.
- [22] Likitha KN, Nethravathi R, Nithyashree K, Ritika Kumari, Sridhar N, and Venkateswaran K. Precision medicine, ai, and the future of personalized health care. *Clin Transl Sci*, 14(1):86–93, 2021.
- [23] U. Javaid Chaudhary L. A. van Graan, L. Lemieux. Methods and utility of eeg-fmri in epilepsy. *Quantitative Imaging in Medicine and Surgery*, 5(2):300–312, April 2015.
- [24] L. Ai C. Andreotti K. Febre A. Mangone N. Vega-Potler N. Langer A. Alexander M. Kovacs S. Litke B. O’Hagan J. Andersen B. Bronstein A. Bui M. Bushey H. Butler V. Castagna N. Camacho E. Chan D. Citera J. Clucas S. Cohen S. Dufek M. P. Milham L. Alexander, J. Escalera. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, 4:170181, 2017.
- [25] Patrick H Lockett, Luigi Maccotta, Jin Joo Lee, Kyung Youn Park, Nico U F Dosenbach, Beau M Ances, Ryan E Hogan, Joshua S Shimony, and Eric C Leuthardt. Deep learning resting state functional magnetic resonance imaging lateralization of temporal lobe epilepsy. *Epilepsia*, 63(6):1542–1552, Jun 2022.

- [26] L. Liu J. Cao Q. Du M. Liu. Co-attention network with label embedding for text classification. *Neurocomputing*, 471:61–69, 2022.
- [27] S. Gite K. Kotecha A. Vidyarthi M. Natu, M. Bachute. Review on epileptic seizure prediction: Machine learning and deep learning approaches. *Comput Math Methods Med*, 2022:7751263, 2022.
- [28] D. Pompili K. Elisevich H. Soltanian-Zadeh M.-P. Hosseini, T. X. Tran. Multimodal data analysis of epileptic eeg and rs-fmri via deep learning and edge computing. *Artificial Intelligence in Medicine*, 104:101813, 2020.
- [29] F. Badilini R. Sassi M. Rahman, M. Walter Rivolta. A systematic survey of data augmentation of eeg signals for ai applications. *Sensors*, 23:5237, 2023.
- [30] Z. Bashir M. Usman, S. Khalid. Epileptic seizure prediction using scalp electroencephalogram signals. *Biocybernetics and Biomedical Engineering*, 41(1):211–220, January–March 2021.
- [31] MedlinePlus. Epilepsia. <https://medlineplus.gov/spanish/ency/article/000694.htm>, 2023.
- [32] L. Kuhlmann M. R. Bonyadi J. Yang O. Kavehei N. D. Truong, A. D. Nguyen. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks*, 105:104–111, 2018.
- [33] World Health Organization. Epilepsy. <https://www.who.int/es/news-room/fact-sheets/detail/epilepsy>, 2023.
- [34] J. Lu J. Yang D. Batra D. Parikh. Hierarchical question-image co-attention for visual question answering. *CoRR*, abs/1606.00061, 2016.
- [35] W. H. Weng T. Y. Chen D. F. Williamson T. Manz M. Shady F. Mahmood R. J. Chen, M. Y. Lu. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3995–4005, 2021.
- [36] P. H. Polegato G. J. G. Lahr R. L. Saute F. N. Nakano H. R. Machado A. C. Sakamoto M. Becker G. A. P. Caurin R. V. Godoy, T. J. S. Reis. Eeg-based epileptic seizure prediction using temporal multi-channel transformers, 2022.
- [37] M. Usman S. M. Usman and S. Fong. Epileptic seizures prediction using machine learning methods. *Comput Math Methods Med*, 2017:9074759, 2017.
- [38] S. L. Thorn b J. A. Onofrey I. H. Melvinsdottir S. Lee J. Langdon A. J. Sinusas J. S. Duncan S. S. Ahn, K. Ta a. Co-attention spatial transformer network for unsupervised motion tracking and cardiac strain analysis in 3d echocardiography. *Medical Image Analysis*, 84:102711, 2023.
- [39] S Sharma, YK Gupta, and AK Mishra. Analysis and prediction of covid-19 multivariate data using deep ensemble learning methods. *Int J Environ Res Public Health*, 20(11):5943, 2023.

- [40] J. Sepulcre M. R. Sabuncu D. Lashkari M. Hollinshead J. L. Roffman J. W. Smoller L. Zöllei J. R. Polimeni B. Fischl H. Liu T. Yeo, F. M. Krienen and R. L. Buckner-corresponding. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165, 2011.
- [41] H. Oya O. Esteban J. M. Shine C. I. Petkov R. A. Poldrack M. Howard R. W. H. Thompson, R. Nair. A data resource from concurrent intracranial stimulation and functional mri of the human brain. *Scientific Data*, 7:258, 2020.
- [42] J. Shen D. Crandall J. Luo X. Lu, W. Wang. Zero-shot video object segmentation with co-attention siamese networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2228–2242, April 2022.
- [43] X. Yu X. Shi. Question answering with co-attention and transformer. *ICME*, 2023.
- [44] B. Liu Y. Song, Q. Zheng and X. Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [45] Y. Zhang L. Wang Zhen Xu1 Y. Wu, P. Zhan. Multimodal fusion with co-attention networks for fake news detection. *Findings of ACL 2021*, pages 3–4, 2023.
- [46] H. Wu J. Zhu M. Sawan Z. Wang, J. Yang. Power efficient refined seizure prediction algorithm based on an enhanced benchmarking. *Scientific Reports*, 11, 2021.

Anexo

Transformada de Fourier de tiempo corto

La Transformada de Fourier de Tiempo Corto (STFT, por sus siglas en inglés) es una técnica que descompone una señal en componentes de frecuencia a lo largo del tiempo, útil para analizar señales no estacionarias.

$$X(t, \omega) = \int_{-\infty}^{\infty} x(\tau) \cdot w(\tau - t) \cdot e^{-j\omega\tau} d\tau$$

Donde

- $X(t, \omega)$ es la representación de la señal en el dominio de la frecuencia-tiempo.
- $x(\tau)$ es la señal original en el tiempo.
- $w(\tau - t)$ es una ventana que limita la señal en un intervalo de tiempo.
- ω es la frecuencia angular.
- j es la unidad imaginaria
- e es la base del logaritmo natural